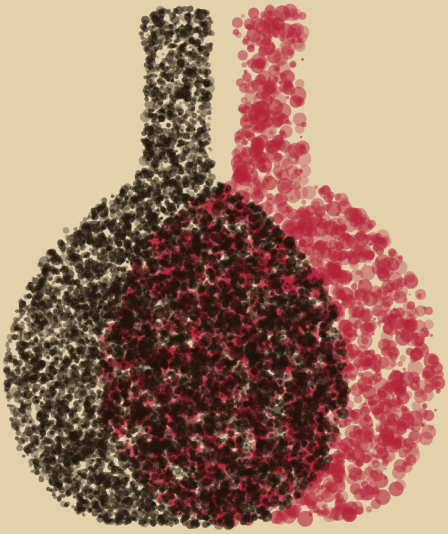


ECML PKDD 2019

Würzburg

Table of Contents

ECML PKDD 2019 Sponsors

Gold Sponsors

Silver Sponsors

Badge Lanyards

Test of Time Award Sponsor

T-Shirts

Supported by

Welcome Message

Welcome to ECML PKDD 2019, the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases 2019, and welcome to Würzburg, Germany!

We offer a rich five-day program with contributions in a variety of formats. Besides the main conference track, we have five keynote speakers, two workshop and tutorial days, a PhD Forum, an industry forum, demonstrations, and discovery challenges.

The main conference consists of two tracks: the *Research Track* and the *Applied Data Science Track*. Papers in both tracks were selected based on the same scientific criteria of quality, novelty and impact, but the focus of the tracks is slightly different. The Applied Data Science Track focuses on novel applications of machine learning and data mining, bridging the gap between practice and current theory, while the Research Track is concerned with the advancement of current theory and methods. From a total of 733 full submissions to the combined tracks, 102 research papers and 28 applied data science papers were selected, which makes for an acceptance rate of almost 18%.
Additionally, the main conference track includes 32 papers from the Journal Track. These papers have been accepted for publication in special issues of either the Springer *Knowledge Discovery & Data Mining* journal or the Springer *Machine Learning* journal.

Our sincere gratitude goes out to all the organizers, program committee members, authors, contributors and participants for making ECML PKDD 2019 possible.

We very much hope that you enjoy the scientific program, the city and the social program!

Élisa Fromont, Arno Knobbe (General Chairs)
Ulf Brefeld, Andreas Hotho, Marloes Maathuis, Céline Robardet (Program Chairs)
Lena Hettinger, Andreas Hotho, Kristof Korwisi, Marc Erich Latoschik (Local Chairs)

Würzburg

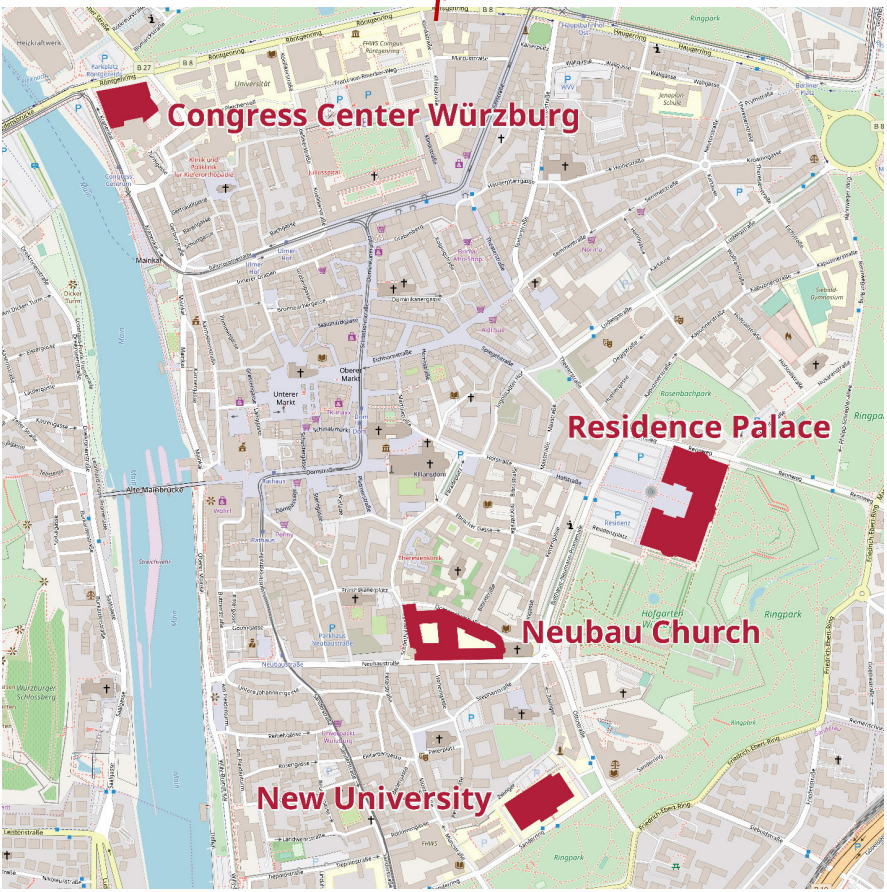
Würzburg is a vibrant town located by the river Main in northern Bavaria, between Frankfurt and Nuremberg. The mix of stunning historical architecture and the young population is what makes the atmosphere so unique, including 35,000 students from three different universities. The mild and sunny climate is ideal to enjoy the many activities Würzburg has to offer: visiting a beer garden next to the river, attending a sporting or cultural event or taking a stroll through one of the parks.
Over the years, Würzburg has been home to many well-known personalities. Artists whose names are inextricably linked to Würzburg are Tilman Riemenschneider and Balthasar Neumann. Wilhelm Conrad Röntgen discovered the X-rays in the local Institute of Physics. Würzburg is also the birthplace of physicist Werner Heisenberg and basketball player Dirk Nowitzki.
Würzburg is the gateway to the Romantic Road, a 350 km long route linking cultural treasures and picturesque landscapes. The Romantic Road connects the beautiful Franconian Wine Country, the world-famous fairy-tale castle Neuschwanstein, the Fugger city of Augsburg and the medieval town of Rothenburg ob der Tauber.
Some highlights we recommend for your visit:

- Visit the Residence Palace (UNESCO World Heritage Site) and the adjoining baroque Court Gardens.
- Enjoy a Brückenschoppen, a glass of Franconian wine served on the Old River Main Bridge.
- Take a trip to the Marienberg fortress for an unforgettable view of the city and its numerous churches.
- Relax in one of Würzburg's parks: the Ringpark which encompasses the old town like a belt, the University's Botanical Garden, or the two former State Garden Show Areas (1990 and 2018).
- Try out Franconian delicacies (such as wine soup, bratwurst or one of the many fish specialties) in a local restaurant or café.

There is something here for everyone! For further information, please visit the website of the tourist information centre: www.wuerzburg.de/en

Downtown Map

Train Station



Local Information

Contacting the **local organisers**: +49 162 1361029 or +49 162 1361081

Available Wifi Networks: @BayernWLAN, eduroam

Currency and Banking: The official currency is Euro (€). ATMs can be found all over the city. Traveller's checks and foreign currency can be exchanged at local banks, but please note that there are no general bank hours in Germany. Some stores and restaurants might not accept credit cards, so it is advised to ask before ordering or to bring cash.

Weather: The weather is relatively mild in September. The temperature ranges from 10 to 20 degrees Celsius (50 to 68 degrees Fahrenheit). Würzburg has less precipitation than almost any other region in Germany.

Electricity: The electrical power supply is 230 V, 50 Hz. Germany uses two plug types, C and F (Euro Plug).

Smoking: Smoking is prohibited inside public buildings, stores, restaurants and cafes.

Taxation: The value-added tax (7 - 19%) is already included in the price, the exact amount is on the receipt. Non-EU residents might get a refund for non-edible purchases.

Conference Venue

The daytime conference events take place in **building Z6** of the university campus **Hubland of the Julius-Maximilians University**, which overlooks the old city of Würzburg. The three-storey building opened in 2011 and contains three large auditoriums and 23 seminar rooms with a total of 2000 seats. The terrace in the upper floor offers a magnificent view of the Marienberg fortress and the city.

Across the main street, and connected by a footbridge, is the newly opened main cafeteria "Hubland Nord". On campus Hubland, you can find most faculties and administrative buildings of the university and parts of the University of Applied Sciences. Since the departure of the US Forces in 2008, the former barracks have been converted to additional campus buildings. The campus area is adjoined by last year's State Garden Show area.

The Julius-Maximilians-University Würzburg is Bavaria's oldest university, founded in 1402. Throughout its history, fourteen Nobel Laureates have researched and taught here. The university offers a wide array of degree courses, both traditional and new ones. Currently, more than 28.000 students are enrolled.

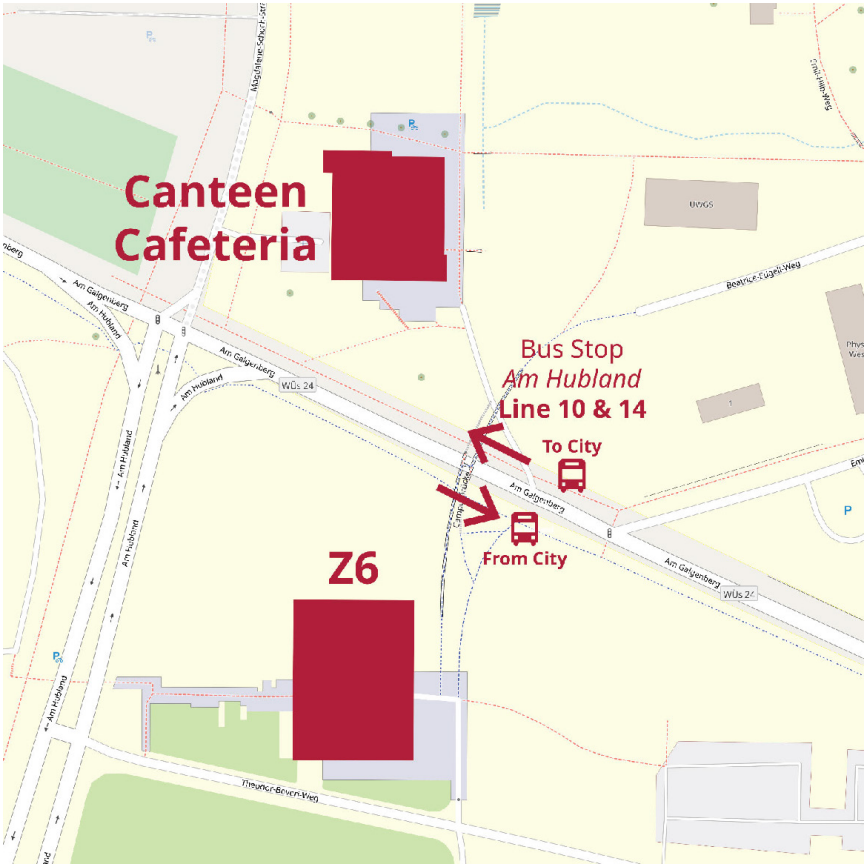
Getting there

From TOWN CENTER, RIVER MAIN, NEW UNIVERSITY
Line 10 or 34 – Stop: Am Hubland

From TRAIN STATION, EAST PART OF THE TOWN CENTER
Line 14, 114, or 219 – Stop: Hubland/Mensa
(**Line 29** – Stop: Philosophisches Institut)

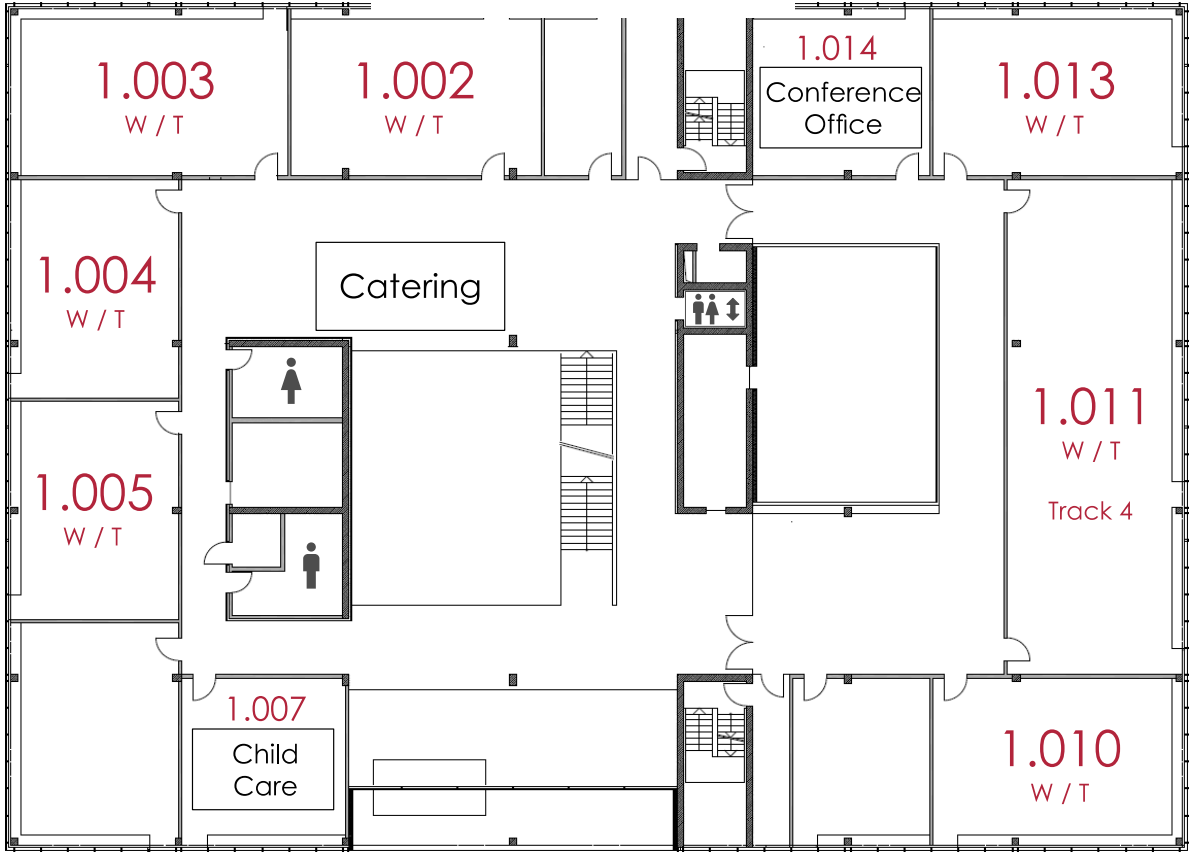
The level entrance and an elevator allow wheelchair accessibility. Parking spaces can be found all over the campus, e.g., the parking area in front of the Philosophische Fakultät.

Campus Map

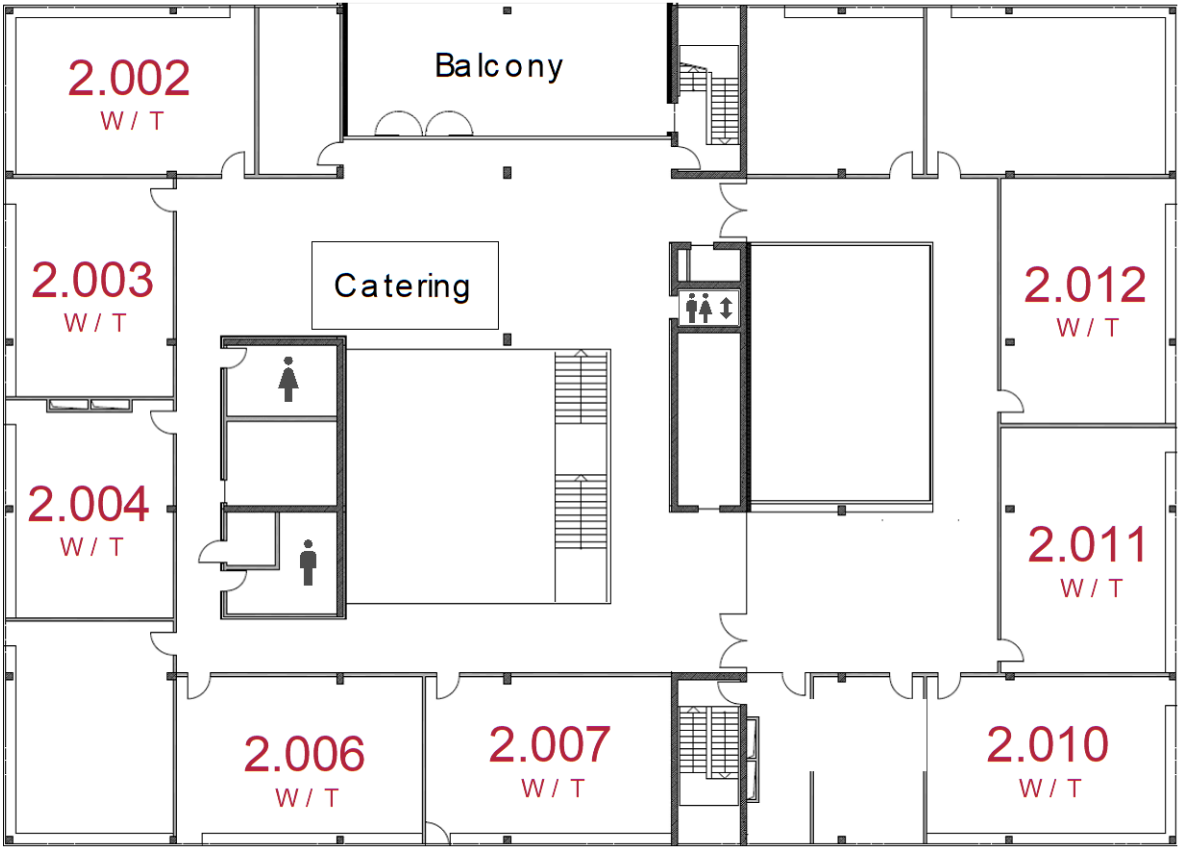

Canteen
Cafeteria
Bus Stop
Am Hubland
Line 10 & 14
To City
From City
Z6
Am Galgenberg
Am Hubland
Theodor-Boveri-Weg
Beatrice-Edgell-Weg
WÜs 24

Z6 Building, Ground Floor

Z6 Building, First Floor



Z6 Building, Second Floor



Venues of the Social Events

The **opening ceremony** on **Monday** takes place at the birthplace of Würzburg's university: the **Neubau Church** (Neubaustraße 9). It was opened in 1591 as the university's church and has undergone numerous renovations and new construction measures in its over 400-year history. Today, it serves as a ceremonial hall for the university.

The **opening reception** on **Monday** takes place in the UNESCO World Heritage Site **Residence Palace** (Residenzplatz 2). The baroque castle is not only famous for its breathtaking grand staircase with the world's largest ceiling fresco by the Venetian artist Tiepolo, but also for the perfect symbiosis of different European architectural influences.

The conference **gala banquet** on **Tuesday** takes place in the **Congress Center Würzburg** (Pleichertorstraße). This building next to the river Main has been renovated in 2015, and a roof level with a panorama view over the city was added. The Congress Center is often used to host concerts, symposia and conferences. The laboratory where Wilhelm Conrad Röntgen discovered the X-rays is a five-minute walk away.

The **poster session** on **Wednesday** takes place in the atrium of the **New University** (Sanderring 2). The New University was built at the end of the 19th century as the university's new main building and still contains the main auditorium and administrative offices. While the exterior façade was built in a new baroque style, the atrium is now home to modern sculptures.

Thursday's poster session takes place in a special environment: the **Staatlicher Hofkeller's** wine cellar beneath the Residence Palace (Residenzplatz 2). The wine cellar belongs to one of the world's oldest vineyards and its history dates back to 1128. Not only the wine is regularly awarded; the cellar also received an architecture award in 2007.

The **farewell party** on **Friday evening**, starting at 19:00, takes place in the MS Zufriedenheit. The Party for the attendees and organisers will be the final event of the conference.

www.neueliebealterhafen.de
MS Zufriedenheit
Oskar-Laredo-Platz 1
97080 Würzburg

Program at a Glance

	Monday	Tuesday	Wednesday	Thursday	Friday
	Registration				
09:00	Workshops, Tutorials & PhD Forum	Keynote	Keynote	Keynote	Keynote
09:30					
10:00		Best Paper ML	Test-of-time Award	Best DM Paper	Coffee Break
10:30	Coffee Break	Coffee Break	Coffee Break	Coffee Break	Discovery Challenges, Workshop & Turorials
11:00	Workshops, Tutorials & PhD Forum	Parallel Talks	Single Track	Parallel Talks	
11:30					
12:00					
12:30	Snack Break	Lunch Break	Lunch Break	Lunch Break	Snack Break
13:00					
13:30					
14:00	Workshops, Tutorials & PhD Forum	Parallel Talks	Parallel Talks	Parallel Talks	Discovery Challenges, Workshop & Turorials
14:30					
15:00	Coffee Break				
15:30	Workshops, Tutorials & PhD Forum				
16:00		Coffee Break	Coffee Break	Coffee Break	Coffee Break
16:30		Parallel Talks	Industry Forum & Parallel Talks	Parallel Talks	Workshop & Tutorials
17:00					
17:30	Opening Ceremony and Keynote		Community Meeting		
18:00		Boat Trip and Veitshöchheim			
18:30					
19:00			Demo and Poster Session	Poster Session	
19:30	Reception				
20:00		Gala Dinner			
20:30					
21:00					
21:30					
22:00					

Program Monday

The following is the generic program for Monday. Please refer to the website of workshops, tutorials or the PhD Forum for more detailed programs.

09:00 – 10:30	Session 1, workshops, tutorials & PhD Forum
10:30 – 11:00	Coffee break
11:00 – 12:40	Session 2, workshops, tutorials & PhD Forum
12:40 – 14:00	Snack break
14:00 – 15:00	Session 3, workshops, tutorials & PhD Forum
15:00 – 15:20	Coffee break
15:20 – 16:20	Session 4, workshops, tutorials & PhD Forum
16:30 – 17:00	Bus transfer from 'Am Hubland': 16:30 - 16:40 - 16:55, or line 10 to 'Sanderring' or line 14 to 'Mainfranken Theater'
17:30 – 19:00	Opening Ceremony & Keynote (Sumit Gulwani – Programming by Examples), Neubau Church
19:30 – 22:00	Welcome Reception, Residence Palace

List of workshops on Monday

- MML 2019: 12th International Workshop on Machine Learning and Music, musml2019.weebly.com, Room 2.012
- Workshop on Multiple-aspect analysis of semantic trajectories (MASTER2019), www.master-project-h2020.eu/workshop-master-2019, Room 2.003
- MIDAS - The 4th Workshop on MIning DAta for financial applicationS, midas.portici.enea.it, Room 2.011
- Second International Workshop on Knowledge Discovery and User Modeling for Smart Cities, umcit-2019.isistan.unicen.edu.ar, Room 2.010
- New Frontiers in Mining Complex Patterns, www.di.uniba.it/~loglisci/NFMCP2019/index.html, Room 1.002
- New Trends in Representation Learning with Knowledge Graphs, sites.google.com/view/kgrlfr-workshop/home, Room 0.002
- Second International Workshop on Energy Efficient Scalable Data Mining and Machine Learning. Green Data Mining, greendatamining.github.io, Room 1.002
- Deep Continuous-Discrete Machine Learning (DeCoDeML 2019), sites.google.com/view/decodeml-workshop-2019, Room 1.003
- Decentralized Machine Learning at the Edge, dmle.iais.fraunhofer.de, Room 1.003

- Applications of Topological Data Analysis, sites.google.com/view/atda2019, Room 1.002
- GEM: Graph Embedding and Mining, gem-ecmlpkdd.github.io, Room 0.001
- Interactive Adaptive Learning (combined with tutorial), p.ies.uni-kassel.de/ial2019/, Room 1.011
- IoT Stream for Data Driven Predictive Maintenance (combined with tutorial), abifet.wixsite.com/iotstream2019, Room 1.010

List of tutorials on Monday

- The ins and outs of reviewing: the good, the bad, and the ugly, Hendrik Blockeel, Jesse Davis, Room 2.006
- Data Mining and Machine Learning using Constraint Programming Languages - An Overview and Future Directions, Ian Davidson, Tias Guns, Siegfried Nijssen, Room 2.007
- Interactive Adaptive Learning (combined with workshop), 9:00 – 12:25, Room 1.011
- IoT Stream for Data Driven Predictive Maintenance (combined with workshop), Room 1.010
- Machines Who Imagine: Going Beyond Data Science, Sridhar Mahadevan, Room 2.006
- Constraint Learning, Luc De Raedt, Andrea Passerini, Stefano Teso, Room 2.007

PhD Forum, Room 1.013 (if not stated otherwise)

09:00 – 10:30	The ins and outs of reviewing: the good, the bad, and the ugly 1 Room 2.006
10:30 – 11:00	Coffee break
11:00 – 12:40	The ins and outs of reviewing: the good, the bad, and the ugly 2 Room 2.006
12:40 – 14:00	Snack break
14:00 – 14:05	Welcome
14:05 – 14:30	Mostly Deep Learning session
14:30 – 14:35	Break
14:35 – 14:55	Mostly Interpretability session
15:00 – 15:20	Coffee break
15:20 – 15:45	Mostly Data Mining session
15:45 – 16:15	Poster session
16:15 – 16:20	Closing

Program Tuesday

09:00 – 10:00	Keynote: Aude Billard – Machine Learning for Robust and Fast Control of Manipulation under Disturbances	0.004 (AOK-Hörsaal)
10:00 – 10:30	Best Machine Learning paper: To be announced	
10:30 – 11:00	Coffee break	
11:00 – 12:40	Parallel sessions	
	1. Deep Learning 1	0.004 (AOK-Hörsaal)
	2. ADS: E-Commerce, Finance, Advertising	0.001
	3. Social Networks & Graphs 1	0.002
	4. Probabilistic Models 1	1.011
12:40 – 14:00	Lunch break	
14:00 – 16:00	Parallel sessions	
	1. Strings and Streams	0.004 (AOK-Hörsaal)
	2. Clustering, Anomaly & Outlier Detection	0.001
	3. Supervised Learning 1	0.002
	4. Ranking	1.011
16:00 – 16:20	Coffee break	
16:20 – 18:00	Parallel sessions	
	1. Autoencoder and Clustering	0.004 (AOK-Hörsaal)
	2. ADS: Healthcare	0.001
	3. Social Networks & Graphs 2	0.002
	4. Decision Trees/Interpretability/Causality	1.011
18:00 – 18:15	Bus transfer	Bus stop Am Hubland
18:15 – 20:00	Boat Trip and Veitshöchheim	River Main & Veitshöchheim
20:00 – 22:00	Gala Dinner	Congress Center Würzburg

Program Wednesday

09:00 – 10:00	Keynote: TBA	
10:00 – 10:30	Test of Time Award: Classifier Chains for Multi-label Classification	0.004 (AOK-Hörsaal)
10:30 – 11:00	Coffee break	
11:00 – 12:40	Plenary session	0.004 (AOK-Hörsaal)
12:40 – 14:00	Lunch break	
14:00 – 16:00	Parallel sessions	
	1. Reinforcement Learning & Bandits 1	0.004 (AOK-Hörsaal)
	2. ADS: Applications 1	0.001
	3. Social Networks & Graphs 3	0.002
	4. Optimization & Learning Theory	1.011
16:00 – 16:20	Coffee break	
16:20 – 17:30	Parallel sessions and industry forum	
	1. Large-Scale Learning	0.004 (AOK-Hörsaal)
	2. ADS: Applications 2	0.001
	3. Privacy & Security	0.002
	4. Industry Forum	1.011
17:30 – 18:30	Community meeting	TBA
19:00 – 22:00	Poster session 1 (incl. demonstrations) Bus transfer from 'Am Hubland' between 17:40 and 18:40, or line 10 to 'Sanderring'	New University (downtown)

Industry Forum

16:20 – 16:40	solvatio AG, Roman Ernst and Valentin Burger – Cost Optimization of Machine Learning based Troubleshooting Systems using Gaussian Processes	1.011
16:45 – 17:05	scoutbee GmbH, Dr. Konstantin Mergenthaler – Supplier search at scale: Leveraging the unstructured content of company websites	1.011

Program Thursday

09:00 – 10:00 Keynote: Maria Florina Balcan, Carnegie Mellon University – Data Driven Algorithm Design — 0.004 (AOK-Hörsaal)

10:00 – 10:30 Best Data Mining paper: To be announced

10:30 – 11:00 Coffee break

11:00 – 12:40 Parallel sessions

1.	Deep Learning 2	0.004 (AOK-Hörsaal)
2.	ADS: Computer Vision & Explanation	0.001
3.	Supervised Learning 2	0.002
4.	Pattern Mining	1.011

12:40 – 14:00 Lunch break

14:00 – 16:00 Parallel sessions

1.	Reinforcement Learning & Bandits 2	0.004 (AOK-Hörsaal)
2.	Clustering	0.001
3.	Natural Language Processing	0.002
4.	Probabilistic Models 2	1.011

16:00 – 16:20 Coffee break

16:20 – 18:00 Parallel sessions

1.	Deep Learning 3	0.004 (AOK-Hörsaal)
2.	ADS: Rich Data	0.001
3.	Multi-Label Learning	0.002
4.	Dimensionality Reduction & Feature Selection	1.011

19:00 – 22:00 Poster session 2 — Residence Palace Wine Cellar

Bus transfer from ‘Am Hubland’ at 18:10, 18:20, 18:25, 18:30, or line 14 to ‘Rennweg’

Program Friday

The following is the generic program for Friday. Please refer to the website of workshops, tutorials or the Discovery Challenges for more detailed programs.

09:00 – 10:00	Keynote: Tinne Tuytelaars – The Quest for the Perfect Image Representation 0.004 (AOK-Hörsaal)
10:00 – 10:30	Coffee break
10:30 – 12:40	Session 1, workshops, tutorials & Discovery Challenges
12:40 – 14:00	Snack break
14:00 – 16:00	Session 2, workshops, tutorials & Discovery Challenges
16:00 – 16:20	Coffee break
16:20 – 18:00	Session 3, workshops, tutorials
19:00 – …	Farewell Party at the MS Zufriedenheit

List of workshops on Friday

- Machine Learning for Cybersecurity (MLCS), mlcs.lasige.di.fc.ul.pt, Room 2.010
- BioASQ: Large-scale biomedical semantic indexing and question answering, www.bioasq.org/workshop, Room 1.004
- 6th Workshop on Sports Analytics: Machine Learning and Data Mining for Sports Analytics (MLSA), dtai.cs.kuleuven.be/events/MLSA19, Room 2.011
- 4th workshop on Advanced Analytics and Learning on Temporal Data, project.inria.fr/aaltd19, Room 0.002
- MACLEAN: MAChine Learning for EArth ObservatioN, mdl4eo.irstea.fr/maclean-machine-learning-for-earth-observation, Room 2.003
- Automating Data Science, sites.google.com/view/autods, Room 1.011
- The Fourth Workshop on Data Science for Social Good, sites.google.com/view/ecmlpkddsogood2019, Room 2.012
- Advances in managing and mining large evolving graphs (third edition), leg-ecmlpkdd19.loria.fr, Room 1.010
- Data and Machine Learning Advances with Multiple Views, damvl.lis-lab.fr, Room 1.002
- Workshop on Data Integration and Applications, sites.google.com/view/dina2019, Room 1.003
- XKDD Tutorial and XKDD - AIMLAI Workshop, kdd.isti.cnr.it/xkdd2019, Room 0.001

List of tutorials on Friday

- Adaptive Influence Maximization, Bogdan Cautis, Silviu Maniu, Nikolaos Tziortziotis, Room 1.013
- Scalable Deep Learning: from theory to practice, Elena Mocanu, Decebal Constantin Mocanu, Room 0.004
- XKDD Tutorial (combined with XKDD – AIMLAI workshop), 9:00 – 13:00, Room 0.001
- On Ordered Sets in Pattern Mining, Aimene Belfodil, Mehdi Kaytoue, Sergei O. Kuznetsov, Amedeo Napoli, Room 1.013
- Machine Learning for Automatic Word Problem Solving, Sowmya S Sundaram, Savitha Sam Abraham, Deepak P, Room 2.002

List of Discovery Challenges

- DC1: Multimodal (Audio, Facial and Gesture) based Emotion Recognition Challenge, Room 2.006 (Morning)
- DC2: SIMAH (SocIaL Media and Harassment): First workshop on categorizing different types of online harassment languages in social media, Room 2.007 (Morning)
- DC3: Correcting Transiting Exoplanet Light Curves for Stellar Spots, Room 2.006 (Afternoon)
- DC4: AutoCV2 Challenge, Room 2.007 (Afternoon)

Keynote Speakers

Monday, 18:00 – 19:00, Neubau Church

Sumit Gulwani, Microsoft Research – Programming by Examples

Sumit Gulwani is a computer scientist seeking connections: between ideas, between research & practice, and with people with varied roles. He is the inventor of many intent-understanding, programming-by-example, and programming-by-natural-language technologies including the popular Flash Fill feature in Excel used by hundreds of millions of people. He founded and currently leads the PROSE research and engineering team that develops APIs for program synthesis and incorporates them into various products. He has published 120+ peer-reviewed papers in top-tier conferences and journals across multiple computer science areas, delivered 45+ keynotes and invited talks at various forums, and authored 50+ patent applications (granted and pending). He was awarded the ACM SIGPLAN Robin Milner Young Researcher Award in 2014 for his pioneering contributions to end-user programming and intelligent tutoring systems. He obtained his PhD in Computer Science from UC-Berkeley, and was awarded the ACM SIGPLAN Outstanding Doctoral Dissertation Award. He obtained his BTech in Computer Science and Engineering from IIT Kanpur, and was awarded the President's Gold Medal.

Tuesday, 09:00 – 10:00, 0.004 (AOK-Hörsaal)

Aude Billard, École polytechnique fédérale de Lausanne – Machine Learning for Robust and Fast Control of Manipulation under Disturbances

Professor Aude Billard is head of the Learning Algorithms and Systems Laboratory (LASA) at the School of Engineering at the EPFL. She received a M.Sc. in Physics from EPFL (1995), a MSc. in Knowledge-based Systems (1996) and a Ph.D. in Artificial Intelligence (1998) from the University of Edinburgh. She was the recipient of the Intel Corporation Teaching award, the Swiss National Science Foundation career award in 2002, the Outstanding Young Person in Science and Innovation from the Swiss Chamber of Commerce and the IEEE-RAS Best Reviewer Award. Aude Billard served as an elected member of the Administrative Committee of the IEEE Robotics and Automation society for two terms (2006-2008 and 2009-2011). She was a plenary speaker at major robotics conferences, (ROMAN, ICRA, Humanoids, HRI) and acted on various positions on the organization committee of more than 15 International Conferences in Robotics. Her research on human-robot interaction and robot learning from human demonstration is featured regularly in premier venues (BBC, IEEE Spectrum, Wired) and received numerous best paper awards at ICRA, IROS and ROMAN, and the 2015 King-Sun Fu Memorial Award for the best 2014 IEEE Transaction in Robotics paper. Professor Billard is active in a number of research organisations in Switzerland and abroad. She is currently a member of the Swiss Science and Technology Council (SNSF) and a member of the Swiss Academy of Engineering Sciences (SATW).

Wednesday, 09:00 – 10:00, 0.004 (AOK-Hörsaal)

To be announced

Thursday, 09:00 – 10:00, 0.004 (AOK-Hörsaal)

Maria Florina Balcan, Carnegie Mellon University – Data Driven Algorithm Design

Maria Florina Balcan is an Associate Professor in the School of Computer Science at Carnegie Mellon University. Her main research interests are machine learning, computational aspects in economics and game theory, and algorithms. Her honors include the CMU SCS Distinguished Dissertation Award, an NSF CAREER Award, a Microsoft Faculty Research Fellowship, a Sloan Research Fellowship, and several paper awards. She was a program committee co-chair for the Conference on Learning Theory in 2014 and for the International Conference on Machine Learning in 2016. She is currently board member of the International Machine Learning Society (since 2011), a Tutorial Chair for ICML 2019, and a Workshop Chair for FOCS 2019.

Friday, 09:00 – 10:00, 0.004 (AOK-Hörsaal)

Tinne Tuytelaars, KU Leuven – The Quest for the Perfect Image Representation

Tinne Tuytelaars is professor at KU Leuven, Belgium, working on computer vision and, in particular, topics related to image representations, vision and language, incremental learning, image generation and more. She has been program chair for ECCV14, general chair for CVPR16, and will again be program chair for CVPR21. She also served as associate-editor-in-chief of the IEEE Transactions on Pattern Analysis and Machine Intelligence over the last four years. She was awarded an ERC Starting Grant in 2009 and received the Koenderink test-of-time award at ECCV16.

Accepted Papers by Session

Best Student Paper Award Winners

To be announced.

Abstracts Tuesday

Deep Learning 1, Tuesday, 11:00 – 12:40

Research Track

Session Chair: Elisa Fromont

Importance Weighted Generative Networks

Reproducible Research

M. Diesendruck (The University of Texas at Austin), E. R. Elenberg (ASAPP, Inc.), R. Sen (Amazon, Inc.), G. W. Cole (The University of Texas at Austin), S. Shakkottai (The University of Texas at Austin), S. A. Williamson (The University of Texas at Austin, CognitiveScale)

While deep generative networks can simulate from complex data distributions, their utility can be hindered by limitations on the data available for training. Specifically, the training data distribution may differ from the target sampling distribution due to sample selection bias, or because the training data comes from a different but related distribution. We present methods to accommodate this difference via importance weighting, which allow us to estimate a loss function with respect to a target distribution even if we cannot access that distribution directly. These estimators, which differentially weight the contribution of data to the loss function, offer theoretical guarantees that heuristic approaches lack, while giving impressive empirical performance in a variety of settings.

Linearly Constrained Weights: Reducing Activation Shift for Faster Training of Neural Networks

T. Kutsuna (Toyota Central R&D Labs)

In this paper, we first identify activation shift, a simple but remarkable phenomenon in a neural network in which the preactivation value of a neuron has non-zero mean that depends on the angle between the weight vector of the neuron and the mean of the activation vector in the previous layer. We then propose linearly constrained weights (LCW) to reduce the activation shift in both fully connected and convolutional layers. The impact of reducing the activation shift in a neural network is studied from the perspective of how the variance of variables in the network changes through layer operations in both forward and backward chains. We also discuss its relationship to the vanishing gradient problem. Experimental results show that LCW enables a deep feedforward network with sigmoid activation functions to be trained efficiently by resolving the vanishing gradient problem. Moreover, combined with batch normalization, LCW improves generalization performance of both feedforward and convolutional networks.

Adversarial Invariant Feature Learning with Accuracy Constraint for Domain Generalization

Reproducible Research

K. Akuzawa, Y. Iwasawa, Y. Matsuo (University of Tokyo)

Learning domain-invariant representation is a dominant approach for domain generalization (DG), where we need to build a classifier that is robust toward domain shifts. However, previous domain-invariance-based methods overlooked the underlying dependency of classes on domains, which is responsible for the trade-off between classification accuracy and domain invariance. Because the primary purpose of DG is to classify unseen domains rather than the invariance itself, the improvement of the invariance can negatively affect DG performance under this trade-off. To overcome the problem, this study first expands the analysis of the trade-off by Xie et al., and provides the notion of accuracy-constrained domain invariance, which means the maximum domain invariance

within a range that does not interfere with accuracy. We then propose a novel method adversarial feature learning with accuracy constraint (AFLAC), which explicitly leads to that invariance on adversarial training. Empirical validations show that the performance of AFLAC is superior to that of domain-invariance-based methods on both synthetic and three real-world datasets, supporting the importance of considering the dependency and the efficacy of the proposed method.

Meta-Learning for Black-box Optimization

V. TV, P. Malhotra, J. Narwariya, L. Vig, G. Shroff (TCS Research, New Delhi)

Recently, neural networks trained as optimizers under the "learning to learn" or meta-learning framework have been shown to be effective for a broad range of optimization tasks including derivative-free black-box function optimization. Recurrent neural networks (RNNs) trained to optimize a diverse set of synthetic non-convex differentiable functions via gradient descent have been effective at optimizing derivative-free black-box functions. In this work, we propose RNN-Opt: an approach for learning RNN-based optimizers for optimizing real-parameter single-objective continuous functions under limited budget constraints. Existing approaches utilize an observed improvement based meta-learning loss function for training such models. We propose training RNN-Opt by using synthetic non-convex functions with known (approximate) optimal values by directly using discounted regret as our meta-learning loss function. We hypothesize that a regret-based loss function mimics typical testing scenarios, and would therefore lead to better optimizers compared to optimizers trained only to propose queries that improve over previous queries. Further, RNN-Opt incorporates simple yet effective enhancements during training and inference procedures to deal with the following practical challenges: (i) Unknown range of possible values for the black-box function to be optimized, and (ii) Practical and domain-knowledge based constraints on the input parameters. We demonstrate the efficacy of RNN-Opt in comparison to existing methods on several synthetic as well as standard benchmark black-box functions along with an anonymized industrial constrained optimization problem.

Training Discrete-Valued Neural Networks with Sign Activations Using Weight Distributions

Reproducible Research

W. Roth (Graz University of Technology), G. Schindler (Ruprecht Karls University, Heidelberg), H. Fröning (Ruprecht Karls University, Heidelberg), F. Pernkopf (Graz University of Technology)

Since resource-constrained devices hardly benefit from the trend towards ever-increasing neural network (NN) structures, there is growing interest in designing more hardware-friendly NNs. In this paper, we consider the training of NNs with discrete-valued weights and sign activation functions that can be implemented more efficiently in terms of inference speed, memory requirements, and power consumption. We build on the framework of probabilistic forward propagations using the local reparameterization trick, where instead of training a single set of NN weights we rather train a distribution over these weights. Using this approach, we can perform gradient-based learning by optimizing the continuous distribution parameters over discrete weights while at the same time perform backpropagation through the sign activation. In our experiments, we investigate the influence of the number of weights on the classification performance on several benchmark datasets, and we show that our method achieves state-of-the-art performance.

E-Commerce, Finance, Advertising, Tuesday, 11:00 – 12:40

ADS Track

Session Chair: Martin Atzmueller

Transfer Learning in Credit Risk

Reproducible Research

H. Suryanto (Rich Data Corporation, Singapore), C. Guan (Rich Data Corporation, Singapore), A. Voumard (Rich Data Corporation, Singapore), G. Beydoun (University of Technology Sydney)

In the credit risk domain, lenders frequently face situations where there is no, or limited historical lending outcome data. This generally results in limited or unaffordable credit for some individuals and small businesses. Transfer learning can potentially reduce this limitation, by leveraging knowledge from related domains, with sufficient outcome data. We investigated the potential for applying transfer learning across various credit domains, for example, from the credit card lending and debt consolidation domain into the small business lending domain.

A Deep Multi-Task Approach for Residual Value Forecasting

A. Rashed (University of Hildesheim), S. Jawed (University of Hildesheim), J. Rehberg (Volkswagen Financial Services AG, Braunschweig), J. Grabocka (University of Hildesheim), L. Schmidt-Thieme (University of Hildesheim), A. Hintsches (Volkswagen Financial Services AG, Braunschweig)

Residual value forecasting plays an important role in many areas, e.g., for vehicles to price leasing contracts. High forecasting accuracy is crucial as any overestimation will lead to lost sales due to customer dissatisfaction, while underestimation will lead to a direct loss in revenue when reselling the car at the end of the leasing contract. Current forecasting models mainly rely on the trend analysis of historical sales records. However, these models require extensive manual steps to filter and preprocess those records which in term limits the frequency at which these models can be updated with new data. To automate, improve and speed up residual value forecasting we propose a multi-task model that utilizes besides the residual value itself as the main target, the actual mileage that has been driven as a co-target. While combining off-the-shelf regression models with careful feature engineering yields already useful models, we show that for residual values further quantities such as the actual driven mileage contains further useful information. As this information is not available when contracts are made and the forecast is due, we model the problem as a multi-task model that uses actual mileage as a training co-target. Experiments on three Volkswagen car models show that the proposed model significantly outperforms the straight-forward modeling as a single-target regression problem.

Shallow Self-Learning for Reject Inference in Credit Scoring

N. Kozodoi (Humboldt University of Berlin; Kreditech, Hamburg), P. Katsas (Kreditech, Hamburg), S. Lessmann (Humboldt University of Berlin), L. Moreira-Matias (Kreditech, Hamburg), K. Papakonstantinou (Kreditech, Hamburg)

Credit scoring models support loan approval decisions in the financial services industry. Lenders train these models on data from previously granted credit applications, where the borrowers' repayment behavior has been observed. This approach creates sample bias. The scoring model is trained on accepted cases only. Applying the model to screen applications from the population of all borrowers degrades its performance. Reject inference comprises techniques to overcome sampling bias through assigning labels to rejected cases. This paper makes two contributions. First, we propose a self-

learning framework for reject inference. The framework is geared toward real-world credit scoring requirements through considering distinct training regimes for labeling and model training. Second, we introduce a new measure to assess the effectiveness of reject inference strategies. Our measure leverages domain knowledge to avoid artificial labeling of rejected cases during evaluation. We demonstrate this approach to offer a robust and operational assessment of reject inference. Experiments on a real-world credit scoring data set confirm the superiority of the suggested self-learning framework over previous reject inference strategies. We also find strong evidence in favor of the proposed evaluation measure assessing reject inference strategies more reliably, raising the performance of the eventual scoring model.

Cold-Start Recommendation for On-Demand Cinemas

B. Li (University of Chinese Academy of Sciences), B. Jin (University of Chinese Academy of Sciences), T. Xue (University of Chinese Academy of Sciences), K. Liu (University of Chinese Academy of Sciences), Q. Zhang (Beijing iQIYI Cinema Management), S. Tian (Beijing iQIYI Cinema Management)

The on-demand cinemas, which has emerged in recent years, provide offline entertainment venues for individuals and small groups. Because of the limitation of network speed and storage space, it is necessary to recommend movies to cinemas, that is, to suggest cinemas to download the recommended movies in advance. This is particularly true for new cinemas. For the new cinema cold-start recommendation, we build a matrix factorization framework and then fuse location categories of cinemas and co-popular relationship between movies in the framework. Specifically, location categories of cinemas are learned through LDA from the type information of POIs around the cinemas and used to approximate cinema latent representations. Moreover, a SPPMI matrix that reflects co-popular relationship between movies is constructed and factorized collectively with the interaction matrix by sharing the same item latent representations. Extensive experiments on real-world data are conducted. The experimental results show that the proposed approach yields significant improvements over state-of-the-art cold-start recommenders in terms of precision, recall and NDCG.

Scalable Bid Landscape Forecasting in Real-time Bidding

A. Ghosh (University of Massachusetts, Amherst), S. Mitra (Adobe Research), S. Sarkhel (Adobe Research), J. Xie (Adobe Advertising Cloud), G. Wu (Adobe Research), V. Swaminathan (Adobe Research)

In programmatic advertising, ad slots are usually sold using second-price (SP) auctions in real-time. The highest bidding advertiser wins but pays only the second highest bid (known as the winning price). In SP, for a single item, the dominant strategy of each bidder is to bid the true value from the bidder's perspective. However, in a practical setting, with budget constraints, bidding the true value is a sub-optimal strategy. Hence, to devise an optimal bidding strategy, it is of utmost importance to learn the winning price distribution accurately. Moreover, a demand-side platform (DSP), which bids on behalf of advertisers, observes the winning price if it wins the auction. For losing auctions, DSPs can only treat its bidding price as the lower bound for the unknown winning price. In literature, typically censored regression is used to model such partially observed data. A common assumption in censored regression is that the winning price is drawn from a fixed variance (homoscedastic) uni-modal distribution (most often Gaussian). However, in reality, these assumptions are often violated. We relax these assumptions and propose a heteroscedastic fully parametric censored regression approach, as well as a mixture density censored network. Our approach not only generalizes censored regression but also provides flexibility to model arbitrarily distributed real-world data. Experimental evaluation on the publicly available dataset for winning price estimation demonstrates the

effectiveness of our method. Furthermore, we evaluate our algorithm on one of the largest demand-side platform and significant improvement has been achieved in comparison with the baseline solutions.

Session Social Networks & Graphs 1, Tuesday, 11:00 – 12:40

Research Track

Session Chair: Tijl de Bie

Triangle Completion Time Prediction using Time-conserving Embedding

Reproducible Research

V. S. Dave and M. Al Hasan (Indiana University Purdue University Indianapolis)

A triangle is an important building block of social networks, so the study of triangle formation in a network is critical for better understanding of the dynamics of such networks. Existing works in this area mainly focus on triangle counting, or generating synthetic networks by matching the prevalence of triangles in real-life networks. While these efforts increase our understanding of triangle's role in a network, they have limited practical utility. In this work we undertake an interesting problem relating to triangle formation in a network, which is, to predict the time by which the third link of a triangle appears in a network. Since the third link completes a triangle, we name this task as Triangle Completion Time Prediction (TCTP). Solution to TCTP problem is valuable for real-life link recommendation in social/e-commerce networks, also it provides vital information for dynamic network analysis and community generation study. An efficient and robust framework (GraNiTE) is proposed for solving the TCTP problem. GraNiTE uses neural networks based approach for learning a representation vector of a triangle completing edge, which is a concatenation of two representation vectors: first one is learnt from graphlet based local topology around that edge and the second one is learnt from time-preserving embedding of the constituting vertices of that edge. A comparison of the proposed solution with several baseline methods shows that the mean absolute error (MAE) of the proposed method is at least one-fourth of that of the best baseline method.

Node Representation Learning for Directed Graphs

Reproducible Research

M. Khosla, J. Leonhardt, W. Nejdl, A. Anand (L3S Resaerch Center, Hannover)

We propose a novel approach for learning node representations in directed graphs, which maintains separate views or embedding spaces for the two distinct node roles induced by the directionality of the edges. We argue that the previous approaches either fail to encode the edge directionality or their encodings cannot be generalized across tasks. With our simple alternating random walk strategy, we generate role specific vertex neighborhoods and train node embeddings in their corresponding source/target roles while fully exploiting the semantics of directed graphs. We also unearth the limitations of evaluations on directed graphs in previous works and propose a clear strategy for evaluating link prediction and graph reconstruction in directed graphs. We conduct extensive experiments to showcase our effectiveness on several real-world datasets on link prediction, node classification and graph reconstruction tasks. We show that the embeddings from our approach are indeed robust, generalizable and well performing across multiple kinds of tasks and graphs. We show that we consistently outperform all baselines for node classification task. In addition to providing a theoretical interpretation of our method we also show that we are considerably more robust than the other directed graph approaches.

SoRecGAT: Leveraging Graph Attention Mechanism for Top-N Social Recommendation

Reproducible Research

V. M, S. Shevade, M. N. Murty (Indian Institute of Science)

Social recommendation systems typically combine extra information like a social network with the user-item interaction network in order to alleviate data sparsity issues. This also helps in making more accurate and personalized recommendations. However, most of the existing systems work under the assumption that all socially connected users have equal influence on each other in a social network, which is not true in practice. Further, estimating the quantum of influence that exists among entities in a user-item interaction network is essential when only implicit ratings are available. This has been ignored even in many recent state-of-the-art models such as SAMN (Social Attentional Memory Network) and DeepSoR (Deep neural network model on Social Relations). Many a time, capturing a complex relationship between the entities (users/items) is essential to boost the performance of a recommendation system. We address these limitations by proposing a novel neural network model, SoRecGAT, which employs multi-head and multi-layer graph attention mechanism. The attention mechanism helps the model learn the influence of entities on each other more accurately. The proposed model also takes care of heterogeneity among the entities seamlessly. SoRecGAT is a general approach and we also validate its suitability when information in the form of a network of co-purchased items is available. Empirical results on eight real-world datasets demonstrate that the proposed model outperforms state-of-the-art models.

Node Classification for Signed Social Networks Using Diffuse Interface Methods

Reproducible Research

P. Mercado (University of Tübingen), J. Bosch (University of British Columbia), M. Stoll (Technische Universität Chemnitz)

Signed networks contain both positive and negative kinds of interactions like friendship and enmity. The task of node classification in non-signed graphs has proven to be beneficial in many real-world applications, yet extensions to signed networks remain largely unexplored. In this paper we introduce the first analysis of node classification in signed social networks via diffuse interface methods based on the Ginzburg-Landau functional together with different extensions of the graph Laplacian to signed networks. We show that blending the information from both positive and negative interactions leads to performance improvement in real signed social networks, consistently outperforming the current state of the art.

Link Prediction via Higher-Order Motif Features

Reproducible Research

G. Abuoda (College of Science and Engineering, HBKU, Doha), G. De Francisci Morales (ISI Foundation, Turin), A. Aboulnaga (Qatar Computing Research Institute, Doha)

Link prediction requires predicting which new links are likely to appear in a graph. In this paper, we present an approach for link prediction that relies on higher-order analysis of the graph topology, well beyond the typical approach which relies on common neighbors. We treat the link prediction problem as a supervised classification problem, and we propose a set of features that depend on the patterns or motifs that a pair of nodes occurs in. By using motifs of sizes 3, 4, and 5, our approach captures a high level of detail about the graph topology. In addition, we propose two optimizations to construct the classification dataset from the graph. First, we propose adding negative examples to the graph as an alternative to the common approach of removing positive ones. Second, we show that it is important to control for the shortest-path distance when sampling pairs of nodes to form negative

examples, since the difficulty of prediction varies with the distance. We experimentally demonstrate that using our proposed motif features in off-the-shelf classifiers results in up to 10 percentage points increase in accuracy over prior topology-based and feature-learning methods.

Probabilistic Models 1, Tuesday, 11:00 – 12:40

Research Track

Session Chair: Hollmen Jaakko

Scalable Large Margin Gaussian Process Classification

M. Wistuba, A. Rawat (IBM Research)

We introduce a new Large Margin Gaussian Process (LMGP) model by formulating a pseudo-likelihood for a generalized multi-class hinge loss. We derive a highly scalable training objective for the proposed model using variational-inference and inducing point approximation. Additionally, we consider the joint learning of LMGP-DNN which combines the proposed model with traditional Deep Learning methods to enable learning for unstructured data. We demonstrate the effectiveness of the Large Margin GP with respect to both training time and accuracy in an extensive classification experiment consisting of 68 structured and two unstructured data sets. Finally, we highlight the key capability and usefulness of our model in yielding prediction uncertainty for classification by demonstrating its effectiveness in the tasks of large-scale active learning and detection of adversarial images.

Integrating Learning and Reasoning with Deep Logic Models

G. Marra (University of Florence; University of Siena), F. Giannini (University of Siena), M. Diligenti (University of Siena), M. Gori (University of Siena)

Deep learning is very effective at jointly learning feature representations and classification models, especially when dealing with high dimensional input patterns. Probabilistic logic reasoning, on the other hand, is capable of take consistent and robust decisions in complex environments. The integration of deep learning and logic reasoning is still an open-research problem and it is considered to be the key for the development of real intelligent agents. This paper presents Deep Logic Models, which are deep graphical models integrating deep learning and logic reasoning both for learning and inference. Deep Logic Models create an end-to-end differentiable architecture, where deep learners are embedded into a network implementing a continuous relaxation of the logic knowledge. The learning process allows to jointly learn the weights of the deep learners and the meta-parameters controlling the high-level reasoning. The experimental results show that the proposed methodology overcomes the limitations of the other approaches that have been proposed to bridge deep learning and reasoning.

Data Association with Gaussian Processes

M. Kaiser (Siemens AG; Technical University of Munich), C. Otte (Siemens AG), T. A. Runkler (Siemens AG; Technical University of Munich), C. H. Ek (University of Bristol)

The data association problem is concerned with separating data coming from different generating processes, for example when data comes from different data sources, contain significant noise, or exhibit multimodality. We present a fully Bayesian approach to this problem. Our model is capable of simultaneously solving the data association problem and the induced supervised learning problem. Underpinning our approach is the use of Gaussian process priors to encode the structure of both the data and the data associations. We present an efficient learning scheme based on doubly stochastic variational inference and discuss how it can be applied to deep Gaussian process priors.

Incorporating Dependencies in Spectral Kernels for Gaussian Processes

Reproducible Research

K. Chen (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences; Shenzhen College of Advanced Technology, University of Chinese Academy of Sciences; Radboud University; Shenzhen Engineering Laboratory of Ocean Environmental Big Data Analysis & Application), T. van Laarhoven (Radboud University; Open University of The Netherlands), J. Chen (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences; Shenzhen College of Advanced Technology, University of Chinese Academy of Sciences; Shenzhen Engineering Laboratory of Ocean Environmental Big Data Analysis & Application), E. Marchiori (Radboud University)

Gaussian processes (GPs) are an elegant Bayesian approach to model an unknown function. The choice of the kernel characterizes one's assumption on how the unknown function autocovaries. It is a core aspect of a GP design, since the posterior distribution can significantly vary for different kernels. The spectral mixture (SM) kernel is derived by modelling a spectral density – the Fourier transform of a kernel – with a linear mixture of Gaussian components. As such, the SM kernel cannot model dependencies between components. In this paper we use cross convolution to model dependencies between components and derive a new kernel called Generalized Convolution Spectral Mixture (GCSM). Experimental analysis of GCSM on synthetic and real-life datasets indicates the benefit of modeling dependencies between components for reducing uncertainty and for improving performance in extrapolation tasks.

Deep convolutional Gaussian processes

Reproducible Research

K. Blomqvist, S. Kaski, M. Heinonen (Aalto University; Helsinki Institute for Information Technology HIIT)

We propose deep convolutional Gaussian processes, a deep Gaussian process architecture with convolutional structure. The model is a principled Bayesian framework for detecting hierarchical combinations of local features for image classification. We demonstrate greatly improved image classification performance compared to current convolutional Gaussian process approaches on the MNIST and CIFAR-10 datasets. In particular, we improve state-of-the-art CIFAR-10 accuracy by over 10 percentage points.

Strings and Streams, Tuesday, 14:00 – 16:00

Research Track

Session Chair: Francois Petitjean

String Sanitization: A Combinatorial Approach

Reproducible Research

G. Bernardini (University of Milano-Bicocca), H. Chen (King's College London), A. Conte (University of Pisa), R. Grossi (University of Pisa; INRIA, Lyon), G. Loukides (King's College London), N. Pisanti (University of Pisa; INRIA, Lyon), S. P. Pissis (INRIA, Lyon; CWI, Amsterdam), G. Rosone (University of Pisa)

String data are often disseminated to support applications such as location-based service provision or DNA sequence analysis. This dissemination, however, may expose sensitive patterns that model confidential knowledge (e.g., trips to mental health clinics from a string representing a user's location history). In this paper, we consider the problem of sanitizing a string by concealing the occurrences of sensitive patterns, while maintaining data utility. First, we propose a time-optimal algorithm, TFS-

ALGO, to construct the shortest string preserving the order of appearance and the frequency of all non-sensitive patterns. Such a string allows accurately performing tasks based on the sequential nature and pattern frequencies of the string. Second, we propose a time-optimal algorithm, PFS-ALGO, which preserves a partial order of appearance of non-sensitive patterns but produces a much shorter string that can be analyzed more efficiently. The strings produced by either of these algorithms may reveal the location of sensitive patterns. In response, we propose a heuristic, MCSR-ALGO, which replaces letters in these strings with carefully selected letters, so that sensitive patterns are not reinstated and occurrences of spurious patterns are prevented. We implemented our sanitization approach that applies TFS-ALGO, PFS-ALGO and then MCSR-ALGO and experimentally show that it is effective and efficient.

Fast likelihood-based change point detection

Reproducible Research

N. Tatti (University of Helsinki)

Change point detection plays a fundamental role in many real-world applications, where the goal is to analyze and monitor the behaviour of a data stream. In this paper, we study change detection in binary streams. To this end, we use a likelihood ratio between two models as a measure for indicating change. The first model is a single bernoulli variable while the second model divides the stored data in two segments, and models each segment with its own bernoulli variable. Finding the optimal split can be done in O(n) time, where n is the number of entries since the last change point. This is too expensive for large n. To combat this we propose an approximation scheme that yields $(1-\varepsilon)$ approximation in $O(\varepsilon^{-1}\log^2 n)$ time. The speed-up consists of several steps: First we reduce the number of possible candidates by adopting a known result from segmentation problems. We then show that for fixed bernoulli parameters we can find the optimal change point in logarithmic time. Finally, we show how to construct a candidate list of size $O(\varepsilon^{-1}\log n)$ for model parameters. We demonstrate empirically the approximation quality and the running time of our algorithm, showing that we can gain a significant speed-up with a minimal average loss in optimality.

Online Linear Models for Edge Computing

H. Sivan (Technion), M. Gabel (University of Toronto), A. Schuster (Technion)

Maintaining an accurate trained model on an infinite data stream is challenging due to concept drifts that render a learned model inaccurate. Updating the model periodically can be expensive, and so traditional approaches for computationally limited devices involve a variation of online or incremental learning, which tend to be less robust. The advent of heterogeneous architectures and Internet-connected devices gives rise to a new opportunity. A weak processor can call upon a stronger processor or a cloud server to perform a complete batch training pass once a concept drift is detected – trading power or network bandwidth for increased accuracy. We capitalize on this opportunity in two steps. We first develop a computationally efficient bound for changes in any linear model with convex, differentiable loss. We then propose a sliding window-based algorithm that uses a small number of batch model computations to maintain an accurate model of the data stream. It uses the bound to continuously evaluate the difference between the parameters of the existing model and a hypothetical optimal model, triggering computation only as needed. Empirical evaluation on real and synthetic datasets shows that our proposed algorithm adapts well to concept drifts and provides a better tradeoff between the number of model computations and model accuracy than classic concept drift detectors. When predicting changes in electricity prices, for example, we achieve 6% better accuracy than the popular EDDM, using only 20 model computations.

Temporal Density Extrapolation using a Dynamic Basis Approach
G. Krempl, D. Lang, V. Hofer

Density estimation is a versatile technique underlying many data mining tasks and techniques, ranging from exploration and presentation of static data, to probabilistic classification, or identifying changes or irregularities in streaming data. With the pervasiveness of embedded systems and digitisation, this latter type of streaming and evolving data becomes more important. Nevertheless, research in density estimation has so far focused on stationary data, leaving the task of extrapolating and predicting density at time points outside a training window an open problem. For this task, temporal density extrapolation (TDX) is proposed. This novel method models and predicts gradual monotonous changes in a distribution. It is based on the expansion of basis functions, whose weights are modelled as functions of compositional data over time by using an isometric log-ratio transformation. Extrapolated density estimates are then obtained by extrapolating the weights to the requested time point, and querying the density from the basis functions with back-transformed weights. Our approach aims for broad applicability by neither being restricted to a specific parametric distribution, nor relying on cluster structure in the data. It requires only two additional extrapolation-specific parameters, for which reasonable defaults exist. Experimental evaluation on various data streams, synthetic as well as from the real-world domains of credit scoring and environmental health, shows that the model manages to capture monotonous drift patterns accurately and better than existing methods. Thereby, it requires not more than 1.5 times the run time of a corresponding static density estimation approach.

A Drift-based Dynamic Ensemble Members Selection using Clustering for Time Series Forecasting

Reproducible Research

A. Saadallah, F. Priebe, K. Morik (TU Dortmund)

Both complex and evolving nature of time series structure make forecasting among one of the most important and challenging tasks in time series analysis. Typical methods for forecasting are designed to model time-evolving dependencies between data observations. However, it is generally accepted that none of them is universally valid for every application. Therefore, methods for learning heterogeneous ensembles by combining a diverse set of forecasts together appear as a promising solution to tackle this task. Hitherto, in classical ML literature, ensemble techniques such as stacking, cascading and voting are mostly restricted to operate in a static manner. To deal with changes in the relative performance of models as well as changes in the data distribution, we propose a drift-aware meta-learning approach for adaptively selecting and combining forecasting models. Our assumption is that different forecasting models have different areas of expertise and a varying relative performance. Our method ensures dynamic selection of initial ensemble base models candidates through a performance drift detection mechanism. Since diversity is a fundamental component in ensemble methods, we propose a second stage selection with clustering that is computed after each drift detection. Predictions of final selected models are combined into a single prediction. An exhaustive empirical testing of the method was performed, evaluating both generalization error and scalability of the approach using time series from several real-world domains. Empirical results show the competitiveness of the method in comparison to state-of-the-art approaches for combining forecasters.

Temporal Pattern Attention for Multivariate Time Series Forecasting
Research Track

Shun-Yao Shih, Fan-Keng Sun, Hung-yi Lee

Forecasting of multivariate time series data, for instance the prediction of electricity consumption, solar power production, and polyphonic piano pieces, has numerous valuable applications. However, complex and non-linear interdependencies between time steps and series complicate this task. To obtain accurate prediction, it is crucial to model long-term dependency in time series data, which can be achieved by recurrent neural networks (RNNs) with an attention mechanism. The typical attention mechanism reviews the information at each previous time step and selects relevant information to help generate the outputs; however, it fails to capture temporal patterns across multiple time steps. In this paper, we propose using a set of filters to extract time-invariant temporal patterns, similar to transforming time series data into its "frequency domain". Then we propose a novel attention mechanism to select relevant time series, and use its frequency domain information for multivariate forecasting. We apply the proposed model on several real-world tasks and achieve state-of-the-art performance in almost all of cases. Our source code is available at https://github.com/gantheory/TPA-LSTM.

Clustering, Anomaly & Outlier Detection, Tuesday, 14:00 – 16:00

Session Chair: Thomas Seidl

Unsupervised and Active Learning using Maximin-based Anomaly Detection
Reproducible Research

Z. Ghafoori, J. C. Bezdek, C. Leckie, S. Karunasekera (University of Melbourne)

Unsupervised anomaly detection is commonly performed using a distance or density-based technique, such as K-Nearest neighbours, Local Outlier Factor or One-class Support Vector Machines. One-class Support Vector Machines reduce the computational cost of testing new data by providing sparse solutions. However, all these techniques have relatively high computational requirements for training. Moreover, identifying anomalies based solely on density or distance is not sufficient when both point (isolated) and cluster anomalies exist in an unlabeled training set. Finally, these unsupervised anomaly detection techniques are not readily adapted for active learning, where the training algorithm should identify examples for which labelling would make a significant impact on the accuracy of the learned model. In this paper, we propose a novel technique called Maximin-based Anomaly Detection that addresses these challenges by selecting a representative subset of data in combination with a kernel-based model construction. We show that the proposed technique (a) provides a statistically significant improvement in the accuracy as well as the computation time required for training and testing compared to several benchmark unsupervised anomaly detection techniques, and (b) effectively uses active learning with a limited budget.

The Elliptical Basis Function Data Descriptor (EBFDD) Network – A One-Class Classification Approach to Anomaly Detection
Reproducible Research

M. H. Z. Bazargani, B. Mac Namee (The Insight Centre for Data Analytics, School of Computer Science, University College Dublin)

This paper introduces the Elliptical Basis Function Data Descriptor (EBFDD) network, a one-class classification approach to anomaly detection based on Radial Basis Function (RBF) neural networks. The EBFDD network uses elliptical basis functions, which allows it to learn sophisticated decision

boundaries while retaining the advantages of a shallow network. We have proposed a novel cost function, whose minimization results in a trained anomaly detector that only requires examples of the normal class at training time. The paper includes a large benchmark experiment that evaluates the performance of EBFDD network and compares it to state of the art one-class classification algorithms including the One-Class Support Vector Machine and the Isolation Forest. The experiments show that, overall, the EBFDD network outperforms the state-of-the-art approaches.

Robust Anomaly Detection in Images using Adversarial Autoencoders

L. Beggel (Bosch Center for Artificial Intelligence, Renningen; Ludwig-Maximilians University Munich), M. Pfeiffer (Bosch Center for Artificial Intelligence, Renningen), B. Bischl (Ludwig-Maximilians University Munich)

Reliably detecting anomalies in a given set of images is a task of high practical relevance for visual quality inspection, surveillance, or medical image analysis. Autoencoder neural networks learn to reconstruct normal images, and hence can classify those images as anomalies, where the reconstruction error exceeds some threshold. Here we analyze a fundamental problem of this approach when the training set is contaminated with a small fraction of outliers. We find that continued training of autoencoders inevitably reduces the reconstruction error of outliers, and hence degrades the anomaly detection performance. In order to counteract this effect, an adversarial autoencoder architecture is adapted, which imposes a prior distribution on the latent representation, typically placing anomalies into low likelihood-regions. Utilizing the likelihood model, potential anomalies can be identified and rejected already during training, which results in an anomaly detector that is significantly more robust to the presence of outliers during training.

Pattern-Based Anomaly Detection in Mixed-Type Time Series

Reproducible Research

L. Feremans (University of Antwerp), V. Vercruyssen (KU Leuven), B. Cule (University of Antwerp), W. Meert (KU Leuven), B. Goethals (University of Antwerp; Monash University)

The present-day accessibility of technology enables easy logging of both sensor values and event logs over extended periods. In this context, detecting abnormal segments in time series data has become an important data mining task. Existing work on anomaly detection focuses either on continuous time series or discrete event logs and not on the combination. However, in many practical applications, the patterns extracted from the event log can reveal contextual and operational conditions of a device that must be taken into account when predicting anomalies in the continuous time series. This paper proposes an anomaly detection method that can handle mixed-type time series. The method leverages frequent pattern mining techniques to construct an embedding of mixed-type time series on which an isolation forest is trained. Experiments on several real-world univariate and multivariate time series, as well as a synthetic mixed-type time series, show that our anomaly detection algorithm outperforms state-of-the-art anomaly detection techniques such as MatrixProfile, Pav, Mifpod and Fpof.

Fast and Parallelizable Ranking with Outliers from Pairwise Comparisons

Reproducible Research

S. Im, M. Montazer Qaem (University of California)

In this paper, we initiate the study of the problem of ordering objects from their pairwise comparison results when allowed to discard up to a certain number of objects as outliers. More specifically, we seek to find an ordering under the popular Kendall tau distance measure, i.e., minimizing the number of pairwise comparison results that are inconsistent with the ordering, with some outliers removed.

The presence of outliers challenges the assumption that a global consistent ordering exists and obscures the measure. This problem does not admit a polynomial time algorithm unless NP $\subseteq$ BPP, and therefore, we develop approximation algorithms with provable guarantees for all inputs. Our algorithms have running time and memory usage that are almost linear in the input size. Further, they are readily adaptable to run on massively parallel platforms such as MapReduce or Spark.

CatchCore: Catching Hierarchical Dense Subtensor

Reproducible Research

W. Feng, S. Liu, X. Cheng (CAS Key Laboratory of Network Data Science & Technology, Institute of Computing Technology, University of Chinese Academy of Sciences)

Dense subtensor detection gains remarkable success in spotting anomaly and fraudulent behaviors for the multi-aspect data (i.e., tensors), like in social media and event streams. Existing methods detect the densest subtensors flatly and separately, with an underlying assumption that those subtensors are exclusive. However, many real-world tensors usually present hierarchical properties, e.g., the core-periphery structure or dynamic communities in networks. In this paper, we propose CatchCore, a novel framework to effectively find the hierarchical dense subtensors. We first design a unified metric for dense subtensor detection, which can be optimized with gradient-based methods. With the proposed metric, \method detects hierarchical dense subtensors through the hierarchy-wise alternative optimization. Finally, we utilize the minimum description length principle to measure the quality of detection result and select the optimal hierarchical dense subtensors. Extensive experiments on synthetic and real-world datasets demonstrate that \method outperforms the top competitors in accuracy for detecting dense subtensors and anomaly patterns. Additionally, CatchCore successfully identified a hierarchical researcher co-authorship group with intense interactions in DBLP dataset. Meanwhile, CatchCore also scales linearly with all aspects of tensors.

Supervised Learning 1, Tuesday, 14:00 – 16:00

Research Track

Session Chair: Peter Flach

Aggregating Algorithm for Prediction of Packs

Dmitry Adamskiy, Anthony Bellotti, Raisa Dzhamtyrova, Yuri Kalnishkan

This paper formulates a protocol for prediction of packs, which is a special case of on-line prediction under delayed feedback. Under the prediction of packs protocol, the learner must make a few predictions without seeing the respective outcomes and then the outcomes are revealed in one go. The paper develops the theory of prediction with expert advice for packs by generalising the concept of mixability. We propose a number of merging algorithms for prediction of packs with tight worst case loss upper bounds similar to those for Vovk's Aggregating Algorithm. Unlike existing algorithms for delayed feedback settings, our algorithms do not depend on the order of outcomes in a pack. Empirical experiments on sports and house price datasets are carried out to study the performance of the new algorithms and compare them against an existing method.

Continual Rare-Class Recognition with Emerging Novel Subclasses

Reproducible Research

H. Nguyen, X. Wang, L. Akoglu (Carnegie Mellon University)

Given a labeled dataset that contains a rare (or minority) class of of-interest instances, as well as a large class of instances that are not of interest, how can we learn to recognize future of-interest instances over a continuous stream? We introduce RaRecognize, which (i) estimates a general

decision boundary between the rare and the majority class, (ii) learns to recognize individual rare subclasses that exist within the training data, as well as (iii) flags instances from previously unseen rare subclasses as newly emerging. The learner in (i) is general in the sense that by construction it is dissimilar to the specialized learners in (ii), thus distinguishes minority from the majority without overly tuning to what is seen in the training data. Thanks to this generality, RaRecognize ignores all future instances that it labels as majority and recognizes the recurrent as well as emerging rare subclasses only. This saves effort at test time as well as ensures that the model size grows moderately over time as it only maintains specialized minority learners. Through extensive experiments, we show that RaRecognize outperforms state-of-the-art baselines on three real-world datasets that contain corporate-risk and disaster documents as rare classes.

Shift Happens: Adjusting Classifiers

Reproducible Research

T. J. T. Heiser, M.-L. Allikivi, M. Kull (University of Tartu)

Minimizing expected loss measured by a proper scoring rule, such as Brier score or log-loss (cross-entropy), is a common objective while training a probabilistic classifier. If the data have experienced dataset shift where the class distributions change post-training, then often the model's performance will decrease, over-estimating the probabilities of some classes while under-estimating the others on average. We propose unbounded and bounded general adjustment (UGA and BGA) methods that transform all predictions to (re-)equalize the average prediction and the class distribution. These methods act differently depending on which proper scoring rule is to be minimized, and we have a theoretical guarantee of reducing loss on test data, if the exact class distribution is known. We also demonstrate experimentally that, when in practice the class distribution is known only approximately, there is often still a reduction in loss depending on the amount of shift and the precision to which the class distribution is known.

Unjustified Classification Regions and Counterfactual Explanations in Machine Learning

Reproducible Research

T. Laugel (Sorbonne Université), M.-J. Lesot (Sorbonne Université), C. Marsala (Sorbonne Université), X. Renard (AXA, Paris), M. Detyniecki (Sorbonne Université, AXA, Paris, Polish Academy of Science)

Post-hoc interpretability approaches, although powerful tools to generate explanations for predictions made by a trained black-box model, have been shown to be vulnerable to issues caused by lack of robustness of the classifier. In particular, this paper focuses on the notion of explanation justification, defined as connectedness to ground-truth data, in the context of counterfactuals. In this work, we explore the extent of the risk of generating unjustified explanations. We propose an empirical study to assess the vulnerability of classifiers and show that the chosen learning algorithm heavily impacts the vulnerability of the model. Additionally, we show that state-of-the-art post-hoc counterfactual approaches can minimize the impact of this risk by generating less local explanations.

Non-parametric Bayesian Isotonic Calibration: Fighting Over-confidence in Binary Classification

Reproducible Research

M.-L. Allikivi, M. Kull (University of Tartu)

Classifiers can often output a score or a probability indicating how sure they are about the predicted class. Classifier calibration methods can map these into calibrated class probabilities, supporting cost-optimal decision making. Isotonic calibration is the standard non-parametric calibration method for binary classifiers, and it can be shown to yield the most likely monotonic calibration map on the given

data, where monotonicity means that instances with higher predicted scores are more likely to be positive. Another non-parametric method is ENIR (ensemble of near-isotonic regression models) which allows for some non-monotonicity, but adds a penalty for it. We first demonstrate that these two methods tend to be over-confident and show that applying label smoothing improves calibration of both methods in more than 90% of studied cases. Unfortunately, label smoothing reduces confidence on the under-confident predictions also, and it does not reduce the raggedness of isotonic calibration. As the main contribution we propose a non-parametric Bayesian isotonic calibration method which has the flexibility of isotonic calibration to fit maps of all monotonic shapes but it adds smoothness and reduces over-confidence without requiring label smoothing. The method introduces a prior over piecewise linear monotonic calibration maps and uses a simple Monte Carlo sampling-based approach to approximate the posterior mean calibration map. Our experiments demonstrate that on average the proposed method results in better calibrated probabilities than the state-of-the-art calibration methods, including isotonic calibration and ENIR.

Deeply Supervised Model for Click-Through Rate Prediction in Sponsored Search
Jelena Gligorijevic, Djordje Gligorijevic, Ivan Stojkovic, Xiao Bai, Amit Goyal, Zoran Obradovic

In sponsored search it is critical to match ads that are relevant to a query and to accurately predict their likelihood of being clicked. Commercial search engines typically use machine learning models for both query-ad relevance matching and click-through-rate (CTR) prediction. However, matching models are based on the similarity between a query and an ad, ignoring the fact that a retrieved ad may not attract clicks, while click models rely on click history, limiting their use for new queries and ads. We propose a deeply supervised architecture that jointly learns the semantic embeddings of a query and an ad as well as their corresponding CTR. We also propose a novel cohort negative sampling technique for learning implicit negative signals. We trained the proposed architecture using one billion query-ad pairs from a major commercial web search engine. This architecture improves the best-performing baseline deep neural architectures by 2% of AUC for CTR prediction and by statistically significant 0.5% of NDCG for query-ad matching.

Ranking, Tuesday, 14:00 – 16:00

Research Track

Session Chair: Eyke Hüllermeier

Pairwise Learning to Rank by Neural Networks Revisited: Reconstruction, Theoretical Analysis and Practical Performance

Reproducible Research

M. Köppel (Johannes Gutenberg-Universität Mainz), A. Segner (Johannes Gutenberg-Universität Mainz), M. Wagener (Johannes Gutenberg-Universität Mainz), L. Pensel (Johannes Gutenberg-Universität Mainz), A. Karwath (University of Birmingham), S. Kramer (Johannes Gutenberg-Universität Mainz)

We present a pairwise learning to rank approach based on a neural net, called DirectRanker, that generalizes the RankNet architecture. We show mathematically that our model is reflexive, antisymmetric, and transitive allowing for simplified training and improved performance. Experimental results on the LETOR MSLR-WEB10K, MQ2007 and MQ2008 datasets show that our model outperforms numerous state-of-the-art methods, while being inherently simpler in structure and using a pairwise approach only.

Learning to Calibrate and Rerank Multi-label Predictions

Reproducible Research

C. Li, V. Pavlu, J. Aslam, B. Wang, K. Qin (Northeastern University)

A multi-label classifier assigns a set of labels to each data object. A natural requirement in many end-use applications is that the classifier also provides a well-calibrated confidence (probability) to indicate the likelihood of the predicted set being correct; for example, an application may automate high-confidence predictions while manually verifying low-confidence predictions. The simplest multi-label classifier, called Binary Relevance (BR), applies one binary classifier to each label independently and takes the product of the individual label probabilities as the overall label-set probability (confidence). Despite its many known drawbacks, such as generating suboptimal predictions and poorly calibrated confidence scores, BR is widely used in practice due to its speed and simplicity. We seek in this work to improve both BR's confidence estimation and prediction through a post calibration and reranking procedure. We take the BR predicted set of labels and its product score as features, extract more features from the prediction itself to capture label constraints, and apply Gradient Boosted Trees (GB) as a calibrator to map these features into a calibrated confidence score. GB not only produces well-calibrated scores (aligned with accuracy and sharp), but also models label interactions, correcting a critical flaw in BR. We further show that reranking label sets by the new calibrated confidence makes accurate set predictions on par with state-of-the-art multi-label classifiers – yet calibrated, simpler, and faster.

A Ranking Model Motivated by Nonnegative Matrix Factorization with Applications to Tennis Tournaments

Reproducible Research

R. Xia (National University of Singapore), V. Y. F. Tan (National University of Singapore), L. Filstroff (IRIT, Université de Toulouse), C. Févotte (IRIT, Université de Toulouse)

We propose a novel ranking model that combines the Bradley-Terry-Luce probability model with a nonnegative matrix factorization framework to model and uncover the presence of latent variables that influence the performance of top tennis players. We derive an efficient, provably convergent, and numerically stable majorization-minimization-based algorithm to maximize the likelihood of datasets under the proposed statistical model. The model is tested on datasets involving the outcomes of matches between 20 top male and female tennis players over 14 major tournaments for men (including the Grand Slams and the ATP Masters 1000) and 16 major tournaments for women over the past 10 years. Our model automatically infers that the surface of the court (e.g., clay or hard court) is a key determinant of the performances of male players, but less so for females. Top players on various surfaces over this longitudinal period are also identified in an objective manner.

A Reduction of Label Ranking to Multiclass Classification

K. Brinker (Hamm-Lippstadt University of Applied Sciences), E. Hüllermeier (Paderborn University)

Label ranking considers the problem of learning a mapping from instances to strict total orders over a predefined set of labels. In this paper, we present a framework for label ranking using a decomposition into a set of multiclass problems. Conceptually, our approach can be seen as a generalization of pairwise preference learning. In contrast to the latter, it allows for controlling the granularity of the decomposition, varying between binary preferences and complete rankings as extreme cases. It is specifically motivated by limitations of pairwise learning with regard to the minimization of certain loss functions. We discuss theoretical properties of the proposed method in terms of accuracy, error correction, and computational complexity. Experimental results are

promising and indicate that improvements upon the special case of pairwise preference decomposition are indeed possible.

Sequential Learning over Implicit Feedback for Robust Large-Scale Recommender Systems

Reproducible Research

A. Burashnikova (Skolkovo Institute of Science and Technology; Université Grenoble Alpes), Y. Maximov (Skolkovo Institute of Science and Technology; Los Alamos National Laboratory), M.-R. Amini (Université Grenoble Alpes)

In this paper, we propose a theoretically founded sequential strategy for training large-scale Recommender Systems (RS) over implicit feedback mainly in the form of clicks. The proposed approach consists in minimizing pairwise ranking loss over blocks of consecutive items constituted by a sequence of non-clicked items followed by a clicked one for each user. Parameter updates are discarded if for a given user the number of sequential blocks is below or above some given thresholds estimated over the distribution of the number of blocks in the training set. This is to prevent from updating the parameters for an abnormally high number of clicks over some targeted items, mainly due to bots; or very few user interactions. Both scenarios affect the decision of RS and imply a shift over the distribution of items that are shown to the users. We provide a proof of convergence of the algorithm to the minimizer of the ranking loss, in the case where the latter is convex. Furthermore, experimental results on five large-scale collections demonstrate the efficiency of the proposed algorithm concerning the state-of-the-art approaches, both regarding different ranking measures and computation time.

Rankboost+: An Improvement to Rankboost

Harold Connamacher

RANKBOOST is a well-known algorithm that iteratively creates and aggregates a collection of “weak rankers” to build an effective ranking procedure. Initial work on RANKBOOST proposed two variants. One variant, that we call RB-D and which is designed for the scenario where all weak rankers have the binary range {0, 1}, has good theoretical properties, but does not perform well in practice. The other, that we call RB-C, has good empirical behavior and is the recommended variation for this binary weak ranker scenario but lacks a theoretical grounding. In this paper, we rectify this situation by proposing an improved RANKBOOST algorithm for the binary weak ranker scenario that we call RANKBOOST+. We prove that this approach is theoretically sound and also show empirically that it outperforms both RANKBOOST variants in practice. Further, the theory behind RANKBOOST+ helps us to explain why RB-D may not perform well in practice, and why RB-C is better behaved in the binary weak ranker scenario, as has been observed in prior work.

Autoencoder and Clustering, Tuesday, 16:20 – 18:00

Research Track

Session Chair: Giuseppe Manco

A Framework for Deep Constrained Clustering - Algorithms and Advances

Reproducible Research

H. Zhang (University of California, Davis), S. Basu (Google Research), I. Davidson (University of California, Davis)

The area of constrained clustering has been extensively explored by researchers and used by practitioners. Constrained clustering formulations exist for popular algorithms such as k-means, mixture models, and spectral clustering but have several limitations. A fundamental strength of deep learning is its flexibility, and here we explore a deep learning framework for constrained clustering and in particular explore how it can extend the field of constrained clustering. We show that our framework can not only handle standard together/apart constraints (without the well documented negative effects reported earlier) generated from labeled side information but more complex constraints generated from new types of side information such as continuous values and high-level domain knowledge.

Agnostic feature selection

G. Doquet and M. Sebag (TAU CNRS - INRIA - LRI - Université Paris-Saclay)

Unsupervised feature selection is mostly assessed along a supervised learning setting, depending on whether the selected features efficiently permit to predict the (unknown) target variable. Another setting is proposed in this paper: the selected features aim to efficiently recover the whole dataset. The proposed algorithm, called AgnoS, combines an AutoEncoder with structural regularizations to sidestep the combinatorial optimization problem at the core of feature selection. The extensive experimental validation of AgnoS on the scikit-feature benchmark suite demonstrates its ability compared to the state of the art, both in terms of supervised learning and data compression.

From abstract items to latent spaces to observed data and back: Compositional Variational Auto-Encoder

Reproducible Research

V. Berger and M. Sebag (TAU, CNRS - INRIA - LRI - Univ. Paris-Saclay)

Conditional Generative Models are now acknowledged an essential tool in Machine Learning. This paper focuses on their control. While many approaches aim at disentangling the data through the coordinate-wise control of their latent representations, another direction is explored in this paper. The proposed CompVAE handles data with a natural multi-ensemblist structure (i.e. that can naturally be decomposed into elements). Derived from Bayesian variational principles, CompVAE learns a latent representation leveraging both observational and symbolic information. A first contribution of the approach is that this latent representation supports a compositional generative model, amenable to multi-ensemblist operations (addition or subtraction of elements in the composition). This compositional ability is enabled by the invariance and generality of the whole framework w.r.t. respectively, the order and number of the elements. The second contribution of the paper is a proof of concept on synthetic 1D and 2D problems, demonstrating the efficiency of the proposed approach.

Deep Collective Matrix Factorization for Augmented Multi-View Learning
Ragunathan Mariappan, Vaibhav Rajan

Learning by integrating multiple heterogeneous data sources is a common requirement in many tasks. Collective Matrix Factorization (CMF) is a technique to learn shared latent representations from arbitrary collections of matrices. It can be used to simultaneously complete one or more matrices, for predicting the unknown entries. Classical CMF methods assume linearity in the interaction of latent factors which can be restrictive and fails to capture complex non-linear interactions. In this paper, we develop the first deep-learning based method, called dCMF, for unsupervised learning of multiple shared representations, that can model such non-linear interactions, from an arbitrary collection of matrices. We address optimization challenges that arise due to dependencies between shared representations through multi-task Bayesian optimization and design an acquisition function adapted for collective learning of hyperparameters. Our experiments show that dCMF significantly outperforms previous CMF algorithms in integrating heterogeneous data for predictive modeling. Further, on two tasks—recommendation and prediction of gene-disease association—dCMF outperforms state-of-the-art matrix completion algorithms that can utilize auxiliary sources of information.

Improving latent variable descriptiveness by modelling rather than ad-hoc factors
Alex Mansbridge, Roberto Fierimonte, Ilya Feige, David Barber

Powerful generative models, particularly in natural language modelling, are commonly trained by maximizing a variational lower bound on the data log likelihood. These models often suffer from poor use of their latent variable, with ad-hoc annealing factors used to encourage retention of information in the latent variable. We discuss an alternative and general approach to latent variable modelling, based on an objective that encourages a perfect reconstruction by tying a stochastic autoencoder with a variational autoencoder (VAE). This ensures by design that the latent variable captures information about the observations, whilst retaining the ability to generate well. Interestingly, although our model is fundamentally different to a VAE, the lower bound attained is identical to the standard VAE bound but with the addition of a simple pre-factor; thus, providing a formal interpretation of the commonly used, ad-hoc pre-factors in training VAEs.

Healthcare, Tuesday, 16:20 – 18:00

ADS Track

Session Chair: TBA

Augmenting Physiological Time Series Data: A Case Study for Sleep Apnea Detection
K. Nikolaidis (University of Oslo), S. Kristiansen (University of Oslo), V. Goebel (University of Oslo), T. Plagemann (University of Oslo), K. Liestøl (University of Oslo), M. Kankanhalli (National University of Singapore)

Supervised machine learning applications in the health domain often face the problem of insufficient training datasets. The quantity of labelled data is small due to privacy concerns and the cost of data acquisition and labelling by a medical expert. Furthermore, it is quite common that collected data are unbalanced and getting enough data to personalize models for individuals is very expensive or even infeasible. This paper addresses these problems by (1) designing a recurrent Generative Adversarial Network to generate realistic synthetic data and to augment the original dataset, (2) enabling the generation of balanced datasets based on a heavily unbalanced dataset, and (3) to control the data generation in such a way that the generated data resembles data from specific individuals. We apply

these solutions for sleep apnea detection and study in the evaluation the performance of four well-known techniques, i.e., K-Nearest Neighbour, Random Forest, Multi-Layer Perceptron, and Support Vector Machine. All classifiers exhibit in the experiments a consistent increase in sensitivity and a kappa statistic increase by between $0.72 \cdot 10^{-2}$ and $18.2 \cdot 10^{-2}$.

Augmenting Semantic Representation of Depressive Language: from Forums to Microblogs

Reproducible Research

N. Farruque, O. Zaiane, R. Goebel (University of Alberta)

We discuss and analyze the process of creating word embedding feature representations specifically designed for a learning task when annotated data is scarce, like depressive language detection from Tweets. We start from rich word embedding pre-trained from a general dataset, then enhance it with embedding learned from a domain specific but relatively much smaller dataset. Our strengthened representation portrays better the domain of depression we are interested in as it combines the semantics learned from the specific domain and word coverage from the general language. We present a comparative analysis of our word embedding representations with a simple bag-of-words model, a well-known sentiment lexicon, a psycholinguistic lexicon, and a general pre-trained word embedding, based on their efficacy in accurately identifying depressive Tweets. We show that our representations achieve a significantly better F1 score than the others when applied to a high-quality dataset.

CASTNet: Community-Attentive Spatio-Temporal Networks for Opioid Overdose Forecasting

Reproducible Research

A. M. Ertugrul (University of Pittsburgh; Middle East Technical University, Ankara), Y.-R. Lin (University of Pittsburgh), T. Taskaya-Temizel (Middle East Technical University, Ankara)

Opioid overdose is a growing public health crisis in the United States. This crisis, recognized as “opioid epidemic”, has widespread societal consequences including the degradation of health, and the increase in crime rates and family problems. To improve the overdose surveillance and to identify the areas in need of prevention effort, in this work, we focus on forecasting opioid overdose using real-time crime dynamics. Previous work identified various types of links between opioid use and criminal activities, such as financial motives and common causes. Motivated by these observations, we propose a novel spatio-temporal predictive model for opioid overdose forecasting by leveraging the spatio-temporal patterns of crime incidents. Our proposed model incorporates multi-head attentional networks to learn different representation subspaces of features. Such deep learning architecture, called “community-attentive” networks, allows the prediction for a given location to be optimized by a mixture of groups (i.e., communities) of regions. In addition, our proposed model allows for interpreting what features, from what communities, have more contributions to predicting local incidents as well as how these communities are captured through forecasting. Our results on two real-world overdose datasets indicate that our model achieves superior forecasting performance and provides meaningful interpretations in terms of spatio-temporal relationships between the dynamics of crime and that of opioid overdose.

Investigating Time Series Classification Techniques for Rapid Pathogen Identification with Single-Cell MALDI-TOF Mass Spectrum Data

C. Papagiannopoulou (Ghent University), R. Parchen (BiosparQ, Leiden), W. Waegeman (Ghent University)

Matrix-assisted laser desorption/ionization-time-of-flight mass spectrometry (MALDI-TOF-MS) is a well-known technology, widely used in species identification. Specifically, MALDI-TOF-MS is applied on samples that usually include bacterial cells, generating representative signals for the various bacterial species. However, for a reliable identification result, a significant amount of biomass is required. For most samples used for diagnostics of infectious diseases, the sample volume is extremely low to obtain the required amount of biomass. Therefore, amplification of the bacterial load is performed by a culturing phase. If the MALDI process could be applied to individual bacteria, it would be possible to circumvent the need for culturing and isolation, accelerating the whole process. In this paper, we briefly describe an implementation of a MALDI-TOF MS procedure in a setting of individual cells and we demonstrate the use of the produced data for the application of pathogen identification. The identification of pathogens (bacterial species) is performed by using machine learning algorithms on the generated single-cell signals. The high predictive performance of the machine learning models indicates that the produced bacterial signatures constitute an informative representation, helpful in distinguishing the different bacterial species. In addition, we reformulate the bacterial species identification problem as a time series classification task by considering the intensity sequences of a given spectrum as time series values. Experimental results show that algorithms originally introduced for time series analysis are beneficial in modelling observations of single-cell MALDI-TOF MS.

Wearable-based Parkinson's Disease Severity Monitoring using Deep Learning

J. Goschenhofer (Ludwig-Maximilians University Munich; ConnectedLife), F. M. J. Pfister (Ludwig-Maximilians University Munich; ConnectedLife), K. A. Yuksel (ConnectedLife), B. Bischl (Ludwig-Maximilians University Munich), U. Fietzek (Dept. of Neurology, Ludwig-Maximilians University Munich; Schoen Clinic Schwabing), J. Thomas (Ludwig-Maximilians University Munich)

One major challenge in the medication of Parkinson's disease is that the severity of the disease, reflected in the patients' motor state, cannot be measured using accessible biomarkers. Therefore, we develop and examine a variety of statistical models to detect the motor state of such patients based on sensor data from a wearable device. We find that deep learning models consistently outperform a classical machine learning model applied on hand-crafted features in this time series classification task. Furthermore, our results suggest that treating this problem as a regression instead of an ordinal regression or a classification task is most appropriate. For consistent model evaluation and training, we adopt the leave-one-subject-out validation scheme to the training of deep learning models. We also employ a class-weighting scheme to successfully mitigate the problem of high multi-class imbalances in this domain. In addition, we propose a customized performance measure that reflects the requirements of the involved medical staff on the model. To solve the problem of limited availability of high-quality training data, we propose a transfer learning technique which helps to improve model performance substantially. Our results suggest that deep learning techniques offer a high potential to autonomously detect motor states of patients with Parkinson's disease.

Research Track

Novel Dense Subgraph Discovery Primitives: Risk Aversion and Exclusion Queries

Reproducible Research

C. E. Tsourakakis (Boston University), T. Chen (Boston University), N. Kakimura (Keio University), J. Pachocki (OpenAI)

In the densest subgraph problem, given an undirected graph $G(V, E, w)$ with non-negative edge weights we are asked to find a set of nodes $S \subseteq V$ that maximizes the degree density $w(S)/|S|$, where $w(S)$ is the sum of the weights of the edges in the graph induced by S. This problem is solvable in polynomial time, and in general is well studied. But what happens when the edge weights are negative? Is the problem still solvable in polynomial time? Also, why should we care about the densest subgraph problem in the presence of negative weights? In this work we answer the aforementioned questions. Specifically, we provide two novel graph mining primitives that are applicable to a wide variety of applications. Our primitives can be used to answer questions such as "how can we find a dense subgraph in Twitter with lots of replies and mentions but no follows?", "how do we extract a dense subgraph with high expected reward and low risk from an uncertain graph"? We formulate both problems mathematically as special instances of dense subgraph discovery in graphs with negative weights. We study the hardness of the problem, and we prove that the problem in general is NP-hard, but we also provide sufficient conditions under which it is poly-time solvable. We design an efficient approximation algorithm that works best in the presence of small negative weights, and an effective heuristic for the more general case. Finally, we perform experiments on various real-world datasets that verify the value of the proposed primitives, and the effectiveness of our proposed algorithms.

Graph Signal Processing for Directed Graphs based on the Hermitian Laplacian

S. Furutani (NTT Secure Platform Laboratories, Tokyo), T. Shibahara (NTT Secure Platform Laboratories, Tokyo), M. Akiyama (NTT Secure Platform Laboratories, Tokyo), K. Hato (NTT Secure Platform Laboratories, Tokyo), M. Aida (Tokyo Metropolitan University)

Graph signal processing is a useful tool for representing, analyzing, and processing the signal lying on a graph, and has attracted attention in several fields including data mining and machine learning. A key to construct the graph signal processing is the graph Fourier transform, which is defined by using eigenvectors of the graph Laplacian of an undirected graph. The orthonormality of eigenvectors gives the graph Fourier transform algebraically desirable properties, and thus the graph signal processing for undirected graphs has been well developed. However, since eigenvectors of the graph Laplacian of a directed graph are generally not orthonormal, it is difficult to simply extend the graph signal processing to directed graphs. In this paper, we present a general framework for extending the graph signal processing to directed graphs. To this end, we introduce the Hermitian Laplacian which is a complex matrix obtained from an extension of the graph Laplacian. The Hermitian Laplacian is defined so as to preserve the edge directionality and Hermitian property and enables the graph signal processing to be straightforwardly extended to directed graphs. Furthermore, the Hermitian Laplacian guarantees some desirable properties, such as non-negative real eigenvalues and the unitarity of the Fourier transform. Finally, experimental results for representation learning and signal denoising of/on directed graphs show the effectiveness of our framework.

Learning Aligned-Spatial Graph Convolutional Networks for Graph Classification

L. Bai (Central University of Finance and Economics, Beijing), Y. Jiao (Central University of Finance and Economics, Beijing), L. Cui (Central University of Finance and Economics, Beijing), E. R. Hancock (University of York)

In this paper, we develop a novel Aligned-Spatial Graph Convolutional Network (ASGCN) model to learn effective features for graph classification. Our idea is to transform arbitrary-sized graphs into fixed-sized aligned grid structures, and define a new spatial graph convolution operation associated with the grid structures. We show that the proposed ASGCN model not only reduces the problems of information loss and imprecise information representation arising in existing spatially-based Graph Convolutional Network (GCN) models, but also bridges the theoretical gap between traditional Convolutional Neural Network (CNN) models and spatially-based GCN models. Moreover, the proposed ASGCN model can adaptively discriminate the importance between specified vertices during the process of spatial graph convolution, explaining the effectiveness of the proposed model. Experiments on standard graph datasets demonstrate the effectiveness of the proposed model.

A Soft Affiliation Graph Model for Scalable Overlapping Community Detection

Reproducible Research

N. Laitonjam, W. Huáng, N. J. Hurley (Insight Centre for Data Analytics, University College Dublin)

We propose an overlapping community model based on the Affiliation Graph Model (AGM), that exhibits the pluralistic homophily property that the probability of a link between nodes increases with increasing number of shared communities. We take inspiration from the Mixed Membership Stochastic Blockmodel (MMSB), in proposing an edgewise community affiliation. This allows decoupling of community affiliations between nodes, opening the way to scalable inference. We show that our model corresponds to an AGM with soft community affiliations and develop a scalable algorithm based on a Stochastic Gradient Riemannian Langevin Dynamics (SGRLD) sampler. Empirical results show that the model can scale to network sizes that are beyond the capabilities of MCMC samplers of the standard AGM. We achieve comparable performance in terms of accuracy and run-time efficiency to scalable MMSB samplers.

node2bits: Compact Time- and Attribute-aware Node Representations for User Stitching

Reproducible Research

D. Jin (University of Michigan), M. Heimann (University of Michigan), R. A. Rossi (Adobe Research), D. Koutra (University of Michigan)

Identity stitching, the task of identifying and matching various online references (e.g., sessions over different devices and timespans) to the same user in real-world web services, is crucial for personalization and recommendations. However, traditional user stitching approaches, such as grouping or blocking, require quadratic pairwise comparisons between a massive number of user activities, thus posing both computational and storage challenges. Recent works, which are often application-specific, heuristically seek to reduce the amount of comparisons, but they suffer from low precision and recall. To solve the problem in an application-independent way, we take a heterogeneous network-based approach in which users (nodes) interact with content (e.g., sessions, websites), and may have attributes (e.g., location). We propose node2bits, an efficient framework that represents multi-dimensional features of node contexts with binary hashcodes. node2bits leverages feature-based temporal walks to encapsulate short- and long-term interactions between nodes in heterogeneous web networks, and adopts SimHash to obtain compact, binary

representations and avoid the quadratic complexity for similarity search. Extensive experiments on large-scale real networks show that node2bits outperforms traditional techniques and existing works that generate real-valued embeddings by up to 5.16% in *F*1 score on user stitching, while taking only up to 1.56% as much storage.

Decision Trees/Interpretability/Causality, Tuesday, 16:20 – 18:00

Research Track

Session Chair: Jilles Vreeken

An Algorithm for Reducing the Number of Distinct Branching Conditions in a Decision Forest

A. Nakamura, K. Sakurada (Hokkaido University)

Given a decision forest, we study a problem of reducing the number of its distinct branching conditions without changing each tree's structure while keeping classification performance. A decision forest with a smaller number of distinct branching conditions can not only have a smaller description length but also be implemented by hardware more efficiently. To force the modified decision forest to keep classification performance, we consider a condition that the decision paths at each branching node do not change for $100\sigma\%$ of the given feature vectors passing through the node for a given $0 \leq \sigma < 1$. Under this condition, we propose an algorithm that minimizes the number of distinct branching conditions by sharing the same condition among multiple branching nodes. According to our experimental results using 13 datasets in UCI machine learning repository, our algorithm succeeded more than 90% reduction on the number of distinct branching conditions for random forests learned from 3 datasets without degrading classification performance. 90% condition reduction was also observed for 7 other datasets within 0.17 degradation of prediction accuracy from the original prediction accuracy at least 0.673.

CaDET: Interpretable Parametric Conditional Density Estimation with Decision Trees and Forests

Cyrus Cousins, Matteo Riondato

We introduce CaDET, an algorithm for *parametric Conditional Density Estimation* (CDE) based on decision trees and random forests. CaDET uses the empirical cross entropy impurity criterion for tree growth, which incentivizes splits that improve predictive accuracy more than the regression criteria or estimated mean-integrated-square-error used in previous works. CaDET also admits more efficient training and query procedures than existing tree-based CDE approaches, and stores only a bounded amount of information at each tree leaf, by using *sufficient statistics* for all computations. Previous tree-based CDE techniques produce complicated uninterpretable distribution objects, whereas CaDET may be instantiated with easily interpretable distribution families, making every part of the model easy to understand. Our experimental evaluation on real datasets shows that CaDET usually learns more accurate, smaller, and more interpretable models, and is less prone to overfitting than existing tree-based CDE approaches.

A Flexible Probabilistic Framework for Large-Margin Mixture of Experts

Archit Sharma, Siddhartha Saxena, Piyush Rai

Mixture-of-Experts (MoE) enable learning highly nonlinear models by combining simple expert models. Each expert handles a small region of the data space, as dictated by the gating network which generates the (soft) assignment of input to the corresponding experts. Despite their flexibility and

renewed interest lately, existing MoE constructions pose several difficulties during model training. Crucially, neither of the two popular gating networks used in MoE, namely the softmax gating network and hierarchical gating network (the latter used in the hierarchical mixture of experts), have efficient inference algorithms. The problem is further exacerbated if the experts do not have conjugate likelihood and lack a naturally probabilistic formulation (e.g., logistic regression or large-margin classifiers such as SVM). To address these issues, we develop novel inference algorithms with closed-form parameter updates, leveraging some of the recent advances in data augmentation techniques. We also present a novel probabilistic framework for MoE, consisting of a range of gating networks with efficient inference made possible through our proposed algorithms. We exploit this framework by using Bayesian linear SVMs as experts on various classification problems (which has a non-conjugate likelihood otherwise generally), providing our final model with attractive large-margin properties. We show that our models are significantly more efficient than other training algorithms for MoE while outperforming other traditional non-linear models like Kernel SVMs and Gaussian Processes on several benchmark datasets.

Adjustment Criteria for Recovering Causal Effects from Missing Data

M. Saadati, J. Tian (Iowa State University)

Confounding bias, missing data, and selection bias are three common obstacles to valid causal inference in the data sciences. Covariate adjustment is the most pervasive technique for recovering casual effects from confounding bias. In this paper we introduce a covariate adjustment formulation for controlling confounding bias in the presence of missing-not-at-random data and develop a necessary and sufficient condition for recovering causal effects using the adjustment. We also introduce an adjustment formulation for controlling both confounding and selection biases in the presence of missing data and develop a necessary and sufficient condition for valid adjustment. Furthermore, we present an algorithm that lists all valid adjustment sets and an algorithm that finds a valid adjustment set containing the minimum number of variables, which are useful for researchers interested in selecting adjustment sets with desired properties.

Shrinkage Estimators for Uplift Regression

K. Rudaś (Warsaw University of Technology; Institute of Computer Science, Polish Academy of Sciences), S. Jaroszewicz (Institute of Computer Science, Polish Academy of Sciences)

Uplift modeling is an approach to machine learning which allows for predicting the net effect of an action (with respect to not taking the action). To achieve this, the training population is divided into two parts: the treatment group, which is subjected to the action, and the control group, on which the action is not taken. Our task is to construct a model which will predict the difference between outcomes in the treatment and control groups conditional on individual objects' features. When the group assignment is random, the model admits a causal interpretation. When we assume linear responses in both groups, the simplest way of estimating the net effect of the action on an individual is to build two separate linear ordinary least squares (OLS) regressions on the treatment and control groups and compute the difference between their predictions. In classical linear models, improvements in accuracy can be achieved through the use of so-called shrinkage estimators such as the well-known James-Stein estimator, which has a provably lower mean squared error than the OLS estimator. In this paper we investigate the use of shrinkage estimators in the uplift modeling problem. Unfortunately, direct generalization of the James-Stein estimator does not lead to improved predictions, nor does shrinking treatment and control models separately.

Therefore, we propose a new uplift shrinkage method where estimators in the treatment and control groups are shrunk jointly so as to minimize the error in the predicted net effect of the action. We prove that the proposed estimator does indeed improve on the double regression estimator.

Abstracts Wednesday

Plenary Track, Wednesday, 11:00 – 12:40

Research Track

Session Chair: Arno Knobbe

Fast Gradient Boosting Decision Trees with Bit-Level Data Structures

Reproducible Research

L. Devos, W. Meert, J. Davis (KU Leuven)

A gradient boosting decision tree model is a powerful machine learning method that iteratively constructs decision trees to form an additive ensemble model. The method uses the gradient of the loss function to improve the model at each iteration step. Inspired by the database literature, we exploit bitset and bitslice data structures in order to improve the run time efficiency of learning the trees. We can use these structures in two ways. First, they can represent the input data itself. Second, they can store the discretized gradient values used by the learning algorithm to construct the trees in the boosting model. Using these bit-level data structures reduces the problem of finding the best split, which involves counting of instances and summing gradient values, to counting one-bits in bit strings. Modern CPUs can efficiently count one-bits using AVX2 SIMD instructions. Empirically, our proposed improvements can result in speed-ups of 2 to up to 10 times on datasets with a large number of categorical features without sacrificing predictive performance.

Sets of Robust Rules, and How to Find Them

Reproducible Research

J. Fischer (Max Planck Institute for Informatics; Saarland University), J. Vreeken (CISPA Helmholtz Center for Information Security)

Association rules are among the most important concepts in data mining. Rules of the form $X \rightarrow Y$ are simple to understand, simple to act upon, yet can model important local dependencies in data. The problem is, however, that there are so many of them. Both traditional and state-of-the-art frameworks typically yield millions of rules, rather than identifying a small set of rules that capture the most important dependencies of the data. In this paper, we define the problem of association rule mining in terms of the Minimum Description Length principle. That is, we identify the best set of rules as the one that most succinctly describes the data. We show that the resulting optimization problem does not lend itself for exact search, and hence propose Grab, a greedy heuristic to efficiently discover good sets of noise-resistant rules directly from data. Through extensive experiments we show that, unlike the state-of-the-art, Grab does reliably recover the ground truth. On real world data we show it finds reasonable numbers of rules, that upon close inspection give clear insight in the local distribution of the data.

Assessing the multi-labelness of multi-label data

L. A. F. Park (Western Sydney University), Y. Guo (Western Sydney University), J. Read (École Polytechnique)

Before constructing a classifier, we should examine the data to gain an understanding of the relationships between the variables, to assist with the design of the classifier. Using multi-label data requires us to examine the association between labels: its multi-labelness. We cannot directly measure association between two labels, since the labels' relationships are confounded with the set of observation variables. A better approach is to fit an analytical model to a label with respect to the observations and remaining labels, but this might present false relationships due to the problem of multicollinearity between the observations and labels. In this article, we examine the utility of

regularised logistic regression and a new form of split logistic regression for assessing Research Track labelness of data. We find that a split analytical model using regularisation is able to pr label relationships when no relationships exist, or if the labels can be partitioned. We also find that if label relationships do exist, logistic regression with l_1 regularisation provides the better measurement of multi-labelness.

Black Box Explanation by Learning Image Exemplars in the Latent Feature Space Reproducible Research

R. Guidotti (ISTI-CNR, Pisa), A. Monreale (University of Pisa), S. Matwin (Dalhousie University; Polish Academy of Sciences), D. Pedreschi (University of Pisa)

We present an approach to explain the decisions of black box models for image classification. While using the black box to label images, our explanation method exploits the latent feature space learned through an adversarial autoencoder. The proposed method first generates exemplar images in the latent feature space and learns a decision tree classifier. Then, it selects and decodes exemplars respecting local decision rules. Finally, it visualizes them in a manner that shows to the user how the exemplars can be modified to either stay within their class, or to become counter-factuals by "morphing" into another class. Since we focus on black box decision systems for image classification, the explanation obtained from the exemplars also provides a saliency map highlighting the areas of the image that contribute to its classification, and areas of the image that push it into another class. We present the results of an experimental evaluation on three datasets and two black box models. Besides providing the most useful and interpretable explanations, we show that the proposed method outperforms existing explainers in terms of fidelity, relevance, coherence, and stability.

TD-Regularized Actor-Critic Methods

Simone Parisi, Voot Tangkaratt, Jan Peters, Mohammad Emtiyaz Khan

Actor-critic methods can achieve incredible performance on difficult reinforcement learning problems, but they are also prone to instability. This is partly due to the interaction between the actor and critic during learning, e.g., an inaccurate step taken by one of them might adversely affect the other and destabilize the learning. To avoid such issues, we propose to regularize the learning objective of the actor by penalizing the temporal difference (TD) error of the critic. This improves stability by avoiding large steps in the actor update whenever the critic is highly inaccurate. The resulting method, which we call the TD-regularized actor-critic method, is a simple plug-and-play approach to improve stability and overall performance of the actor-critic methods. Evaluations on standard benchmarks confirm this. Source code can be found at https://github.com/sparisi/td-reg.

Reinforcement Learning & Bandits 1, Wednesday, 14:00 – 16:00

Session Chair: TBA

Deep Ordinal Reinforcement Learning Reproducible Research

A. Zap, T. Joppen, J. Fürnkranz (TU Darmstadt)

Reinforcement learning usually makes use of numerical rewards, which have nice properties but also come with drawbacks and difficulties. Using rewards on an ordinal scale (ordinal rewards) is an alternative to numerical rewards that has received more attention in recent years. In this paper, a general approach to adapting reinforcement learning problems to the use of ordinal rewards is

presented and motivated. We show how to convert common reinforcement learning algorithms to an ordinal variation by the example of Q-learning and introduce Ordinal Deep Q-Networks, which adapt deep reinforcement learning to ordinal rewards. Additionally, we run evaluations on problems provided by the OpenAI Gym framework, showing that our ordinal variants exhibit a performance that is comparable to the numerical variations for a number of problems. We also give first evidence that our ordinal variant is able to produce better results for problems with less engineered and simpler-to-design reward signals.

Attentive Multi-Task Deep Reinforcement Learning

Reproducible Research

T. Bräm, G. Brunner (ETH Zurich)

Sharing knowledge between tasks is vital for efficient learning in a multi-task setting. However, most research so far has focused on the easier case where knowledge transfer is not harmful, i.e., where knowledge from one task cannot negatively impact the performance on another task. In contrast, we present an approach to multi-task deep reinforcement learning based on attention that does not require any a-priori assumptions about the relationships between tasks. Our attention network automatically groups task knowledge into sub-networks on a state level granularity. It thereby achieves positive knowledge transfer if possible, and avoids negative transfer in cases where tasks interfere. We test our algorithm against two state-of-the-art multi-task/transfer learning approaches and show comparable or superior performance while requiring fewer network parameters.

Sample-Efficient Model-Free Reinforcement Learning with Off-Policy Critics

Reproducible Research

D. Steckelmacher (Vrije Universiteit Brussel), H. Plisnier (Vrije Universiteit Brussel), D. M. Roijers (VU Amsterdam), A. Nowé (Vrije Universiteit Brussel)

Value-based reinforcement-learning algorithms provide state-of-the-art results in model-free discrete-action settings, and tend to outperform actor-critic algorithms. We argue that actor-critic algorithms are limited by their need for an on-policy critic. We propose Bootstrapped Dual Policy Iteration (BDPI), a novel model-free reinforcement-learning algorithm for continuous states and discrete actions, with an actor and several off-policy critics. Off-policy critics are compatible with experience replay, ensuring high sample-efficiency, without the need for off-policy corrections. The actor, by slowly imitating the average greedy policy of the critics, leads to high-quality and state-specific exploration, which we compare to Thompson sampling. Because the actor and critics are fully decoupled, BDPI is remarkably stable, and unusually robust to its hyper-parameters. BDPI is significantly more sample-efficient than Bootstrapped DQN, PPO, and ACKTR, on discrete, continuous and pixel-based tasks.

Policy Prediction Network: Model-Free Behavior Policy with Model-Based Learning in Continuous Action Space

Z. Wellmer (Hong Kong University of Science), J. T. Kwok (Technology)

This paper proposes a novel deep reinforcement learning architecture that was inspired by previous tree structured architectures which were only useable in discrete action spaces. Policy Prediction Network offers a way to improve sample complexity and performance on continuous control problems in exchange for extra computation at training time but at no cost in computation at rollout time. Our approach integrates a mix between model-free and model-based reinforcement learning. Policy Prediction Network is the first to introduce implicit model-based learning to Policy Gradient algorithms for continuous action space and is made possible via the empirically justified clipping

scheme. Our experiments are focused on the MuJoCo environments so that they can be compared with similar work done in this area.

Learning 3D Navigation Protocols on Touch Interfaces with Cooperative Multi-Agent Reinforcement Learning

Q. Debard (Itekube, LIRIS), J. S. Dibangoye (Inria, CITI-Lab), S. Canu (LITIS, INSA-Rouen), C. Wolf (LIRIS, Inria, CITI-Lab, INSA-Lyon)

Using touch devices to navigate in virtual 3D environments such as computer assisted design (CAD) models or geographical information systems (GIS) is inherently difficult for humans, as the 3D operations have to be performed by the user on a 2D touch surface. This ill-posed problem is classically solved with a fixed and handcrafted interaction protocol, which must be learned by the user. We propose to automatically learn a new interaction protocol allowing to map a 2D user input to 3D actions in virtual environments using reinforcement learning (RL). A fundamental problem of RL methods is the vast amount of interactions often required, which are difficult to come by when humans are involved. To overcome this limitation, we make use of two collaborative agents. The first agent models the human by learning to perform the 2D finger trajectories. The second agent acts as the interaction protocol, interpreting and translating to 3D operations the 2D finger trajectories from the first agent. We restrict the learned 2D trajectories to be similar to a training set of collected human gestures by first performing state representation learning, prior to reinforcement learning. This state representation learning is addressed by projecting the gestures into a latent space learned by a variational auto encoder (VAE).

Safe Policy Improvement with Soft Baseline Bootstrapping

Reproducible Research

K. Nadjahi (Télécom Paris), R. Laroche (Microsoft Research Montréal), R. Tachet des Combes (Microsoft Research Montréal)

Batch Reinforcement Learning (Batch RL) consists in training a policy using trajectories collected with another policy, called the behavioural policy. Safe policy improvement (SPI) provides guarantees with high probability that the trained policy performs better than the behavioural policy, also called baseline in this setting. Previous work shows that the SPI objective improves mean performance as compared to using the basic RL objective, which boils down to solving the MDP with maximum likelihood. Here, we build on that work and improve more precisely the SPI with Baseline Bootstrapping algorithm (SPIBB) by allowing the policy search over a wider set of policies. Instead of binarily classifying the state-action pairs into two sets (the uncertain and the safe-to-train-on ones), we adopt a softer strategy that controls the error in the value estimates by constraining the policy change according to the local model uncertainty. The method can take more risks on uncertain actions all the while remaining provably-safe, and is therefore less conservative than the state-of-the-art methods. We propose two algorithms (one optimal and one approximate) to solve this constrained optimization problem and empirically show a significant improvement over existing SPI algorithms both on finite MDPS and on infinite MDPs with a neural network function approximation.

Applications 1, Wednesday, 14:00 – 16:00

ADS Track

Session Chair: Tias Guns

Generative Adversarial Networks for Failure Prediction

S. Zheng, A. Farahat, C. Gupta (Industrial AI Lab, Hitachi America)

Prognostics and Health Management (PHM) is an emerging engineering discipline which is concerned with the analysis and prediction of equipment health and performance. One of the key challenges in PHM is to accurately predict impending failures in the equipment. In recent years, solutions for failure prediction have evolved from building complex physical models to the use of machine learning algorithms that leverage the data generated by the equipment. However, failure prediction problems pose a set of unique challenges that make direct application of traditional classification and prediction algorithms impractical. These challenges include the highly imbalanced training data, the extremely high cost of collecting more failure samples, and the complexity of the failure patterns. Traditional oversampling techniques will not be able to capture such complexity and accordingly result in overfitting the training data. This paper addresses these challenges by proposing a novel algorithm for failure prediction using Generative Adversarial Networks (GAN-FP). GAN-FP first utilizes two GAN networks to simultaneously generate training samples and build an inference network that can be used to predict failures for new samples. GAN-FP first adopts an infoGAN to generate realistic failure and non-failure samples, and initialize the weights of the first few layers of the inference network. The inference network is then tuned by optimizing a weighted loss objective using only real failure and non-failure samples. The inference network is further tuned using a second GAN whose purpose is to guarantee the consistency between the generated samples and corresponding labels. GAN-FP can be used for other imbalanced classification problems as well. Empirical evaluation on several benchmark datasets demonstrates that GAN-FP significantly outperforms existing approaches, including under-sampling, SMOTE, ADASYN, weighted loss, and infoGAN augmented training.

Interpreting atypical conditions in systems with deep conditional Autoencoders: the case of electrical consumption

Reproducible Research

A. Marot (Réseau Transport Électricité R&D), A. Rosin (Réseau Transport Électricité R&D), L. Crochepierre (Réseau Transport Électricité R&D; Université de Lorraine), B. Donnot (Réseau Transport Électricité R&D), P. Pinson (DTU Technical University of Denmark), L. Boudjeloud-Assala (Université de Lorraine)

In this paper, we propose a new method to iteratively and interactively characterize new feature conditions for signals of daily French electrical consumption from our historical database, relying on Conditional Variational Autoencoders. An autoencoder first learn a compressed similarity-based representation of the signals in a latent space, in which one can select and extract well-represented expert features. Then, we successfully condition the model over the set of extracted features, as opposed to simple target label previously, to learn conditionally independent new residual latent representations. Unknown, or previously unselected factors such as atypical conditions now appear well-represented to be detected and further interpreted by experts. By applying it, we recover the appropriate known expert features and eventually discover, through adapted representations, atypical known and unknown conditions such as holidays, fuzzy non-working days and weather events, which were actually related to important events that influenced consumption.

Manufacturing Dispatching using Reinforcement and Transfer Learning
S. Zheng, C. Gupta, S. Serita (Industrial AI Lab, Hitachi America)

Efficient dispatching rule in manufacturing industry is key to ensure product on-time delivery and minimum past-due and inventory cost. Manufacturing, especially in the developed world, is moving towards on-demand manufacturing meaning a high mix, low volume product mix. This requires efficient dispatching that can work in dynamic and stochastic environments, meaning it allows for quick response to new orders received and can work over a disparate set of shop floor settings. In this paper we address this problem of dispatching in manufacturing. Using reinforcement learning (RL), we propose a new design to formulate the shop floor state as a 2-D matrix, incorporate job slack time into state representation, and design lateness and tardiness rewards function for dispatching purpose. However, maintaining a separate RL model for each production line on a manufacturing shop floor is costly and often infeasible. To address this, we enhance our deep RL model with an approach for dispatching policy transfer. This increases policy generalization and saves time and cost for model training and data collection. Experiments show that: (1) our approach performs the best in terms of total discounted reward and average lateness, tardiness, (2) the proposed policy transfer approach reduces training time and increases policy generalization.

An aggregate learning approach for interpretable semi-supervised population prediction and disaggregation using ancillary data
Reproducible Research
G. Derval (ICTEAM, UCLouvain), F. Docquier (IRES, UCLouvain), P. Schaus (ICTEAM, UCLouvain)

Census data provide detailed information about population characteristics at a coarse resolution. Nevertheless, fine-grained, high-resolution mappings of population counts are increasingly needed to characterize population dynamics and to assess the consequences of climate shocks, natural disasters, investments in infrastructure, development policies, etc. Dissagregating these censuses is a complex machine learning, and multiple solutions have been proposed in past research. We propose in this paper to view the problem in the context of the aggregate learning paradigm, where the output value for all training points is not known, but where it is only known for aggregates of the points (i.e. in this context, for regions of pixels where a census is available). We demonstrate with a very simple and interpretable model that this method is on par, and even outperforms on some metrics, the state of the art, despite its simplicity.

Optimizing Neural Networks for Patent Classification
Reproducible Research
L. Abdelgawad (Averbis, Freiburg), P. Kluegl (Averbis, Freiburg), E. Genc (Averbis, Freiburg), S. Falkner (Albert-Ludwigs University of Freiburg), F. Hutter (Albert-Ludwigs University of Freiburg)

A great number of patents is filed every day to the patent offices worldwide. Each of these patents has to be labeled by domain experts with one or many of thousands of categories. This process is not only extremely expensive but also overwhelming for the experts, due to the considerable increase of filed patents over the years and the increasing complexity of the hierarchical categorization structure. Therefore, it is critical to automate the manual classification process using a classification model. In this paper, the automation of the task is carried out based on recent advances in deep learning for NLP and compared to customized approaches. Moreover, an extensive optimization analysis grants insights about hyperparameter importance. Our optimized convolutional neural network achieves a new state-of-the-art performance of 55.02% accuracy on the public Wipo-Alpha dataset.

Research Track

User-Guided Clustering in Heterogeneous Information Networks via Motif-Based Comprehensive Transcription

Reproducible Research

Y. Shi, X. He, N. Zhang, C. Yang, J. Han (University of Illinois at Urbana-Champaign)

Heterogeneous information networks (HINs) with rich semantics are ubiquitous in real-world applications. For a given HIN, many reasonable clustering results with distinct semantic meaning can simultaneously exist. User-guided clustering is hence of great practical value for HINs where users provide labels to a small portion of nodes. To cater to a broad spectrum of user guidance evidenced by different expected clustering results, carefully exploiting the signals residing in the data is potentially useful. Meanwhile, as one type of complex networks, HINs often encapsulate higher-order interactions that reflect the interlocked nature among nodes and edges. Network motifs, sometimes referred to as meta-graphs, have been used as tools to capture such higher-order interactions and reveal the many different semantics. We therefore approach the problem of user-guided clustering in HINs with network motifs. In this process, we identify the utility and importance of directly modeling higher-order interactions without collapsing them to pairwise interactions. To achieve this, we comprehensively transcribe the higher-order interaction signals to a series of tensors via motifs and propose the MoCHIN model based on joint non-negative tensor factorization. This approach applies to arbitrarily many, arbitrary forms of HIN motifs. An inference algorithm with speed-up methods is also proposed to tackle the challenge that tensor size grows exponentially as the number of nodes in a motif increases. We validate the effectiveness of the proposed method on two real-world datasets and three tasks, and MoCHIN outperforms all baselines in three evaluation tasks under three different metrics. Additional experiments demonstrated the utility of motifs and the benefit of directly modeling higher-order information especially when user guidance is limited.

Finding lasting dense graphs

Konstantinos Semertzidis, Evaggelia Pitoura, Evimaria Terzi, Panayiotis Tsaparas

Graphs form a natural model for relationships and interactions between entities, for example, between people in social and cooperation networks, servers in computer networks, or tags and words in documents and tweets. But, which of these relationships or interactions are the most lasting ones? In this paper, we study the following problem: given a set of graph snapshots, which may correspond to the state of an evolving graph at different time instances, identify the set of nodes that are the most densely connected in all snapshots. We call this problem the *Best Friends Forever* (BFF) problem. We provide definitions for density over multiple graph snapshots, that capture different semantics of connectedness over time, and we study the corresponding variants of the BFF problem. We then look at the *On–Off*BFF (O^2BFF) problem that relaxes the requirement of nodes being connected in all snapshots, and asks for the densest set of nodes in at least k of a given set of graph snapshots. We show that this problem is NP-complete for all definitions of density, and we propose a set of efficient algorithms. Finally, we present experiments with synthetic and real datasets that show both the efficiency of our algorithms and the usefulness of the BFF and the O^2BFF problems.

Model-Free Inference of Diffusion Networks using RKHS embeddings
Shoubo Hu, Bogdan Cautis, Zhitang Chen, Laiwan Chan, Yanhui Geng, Xiuqiang He

We revisit in this paper the problem of inferring a diffusion network from information cascades. In our study, we make no assumptions on the underlying diffusion model, in this way obtaining a generic method with broader practical applicability. Our approach exploits the pairwise adoption-time intervals from cascades. Starting from the observation that *different kinds of information spread differently*, these time intervals are interpreted as samples drawn from unknown (conditional) distributions. In order to statistically distinguish them, we propose a novel method using Reproducing Kernel Hilbert Space embeddings. Experiments on both synthetic and real-world data from Twitter and Flixster show that our method significantly outperforms the state-of-the-art methods. We argue that our algorithm can be implemented by parallel batch processing, in this way meeting the needs in terms of efficiency and scalability of real-world applications.

Counts-of-Counts Similarity for Prediction and Search in Relational Data
Manfred Jaeger, Marco Lippi, Giovanni Pellegrini, Andrea Passerini

Defining appropriate distance functions is a crucial aspect of effective and efficient similarity-based prediction and retrieval. Relational data are especially challenging in this regard. By viewing relational data as multi-relational graphs, one can easily see that a distance between a pair of nodes can be defined in terms of a virtually unlimited class of features, including node attributes, attributes of node neighbors, structural aspects of the node neighborhood and arbitrary combinations of these properties. In this paper we propose a rich and flexible class of metrics on graph entities based on earth mover's distance applied to a hierarchy of complex counts-of-counts statistics. We further propose an approximate version of the distance using sums of marginal earth mover's distances. We show that the approximation is correct for many cases of practical interest and allows efficient nearest-neighbor retrieval when combined with a simple metric tree data structure. An experimental evaluation on two real-world scenarios highlights the flexibility of our framework for designing metrics representing different notions of similarity. Substantial improvements in similarity-based prediction are reported when compared to solutions based on state-of-the-art graph kernels.

Robust active attacks on social graphs
Sjouke Mauw, Yunior Ramírez-Cruz, Rolando Trujillo-Rasua

In order to prevent the disclosure of privacy-sensitive data, such as names and relations between users, social network graphs have to be anonymised before publication. Naive anonymisation of social network graphs often consists in deleting all identifying information of the users, while maintaining the original graph structure. Various types of attacks on naively anonymised graphs have been developed. Active attacks form a special type of such privacy attacks, in which the adversary enrols a number of fake users, often called *sybils*, to the social network, allowing the adversary to create unique structural patterns later used to re-identify the sybil nodes and other users after anonymisation. Several studies have shown that adding a small amount of noise to the published graph already suffices to mitigate such active attacks. Consequently, active attacks have been dubbed a negligible threat to privacy-preserving social graph publication. In this paper, we argue that these studies unveil shortcomings of specific attacks, rather than inherent problems of active attacks as a general strategy. In order to support this claim, we develop the notion of a *robust active attack*, which is an active attack that is resilient to small perturbations of the social network graph. We formulate the design of robust active attacks as an optimisation problem and we give definitions of robustness for different stages of the active attack strategy. Moreover, we introduce various heuristics to achieve

these notions of robustness and experimentally show that the new robust attacks are considerably more resilient than the original ones, while remaining at the same level of feasibility.

Optimization & Learning Theory, Wednesday, 14:00 – 16:00

Research Track

Session Chair: Sibylle Hess

A Stochastic Quasi-Newton Method with Nesterov's Accelerated Gradient

S. Indrapriyadarsini (Shizuoka University), S. Mahboubi (Shonan Institute of Technology), H. Ninomiya (Shonan Institute of Technology), H. Asai (Shizuoka University)

Incorporating second order curvature information in gradient based methods have shown to improve convergence drastically despite its computational intensity. In this paper, we propose a stochastic (online) quasi-Newton method with Nesterov's accelerated gradient in both its full and limited memory forms for solving large scale non-convex optimization problems in neural networks. The performance of the proposed algorithm is evaluated in TensorFlow on benchmark classification and regression problems. The results show improved performance compared to the classical second order oBFGS and oLBFGS methods and popular first order stochastic methods such as SGD and Adam. The performance with different momentum rates and batch sizes have also been illustrated.

Efficient learning with robust gradient descent

Matthew J. Holland, Kazushi Ikeda

Minimizing the empirical risk is a popular training strategy, but for learning tasks where the data may be noisy or heavy-tailed, one may require many observations in order to generalize well. To achieve better performance under less stringent requirements, we introduce a procedure which constructs a robust approximation of the risk gradient for use in an iterative learning routine. Using high-probability bounds on the excess risk of this algorithm, we show that our update does not deviate far from the ideal gradient-based update. Empirical tests using both controlled simulations and real-world benchmark data show that in diverse settings, the proposed procedure can learn more efficiently, using less resources (iterations and observations) while generalizing better.

On the analysis of adaptability in multi-source domain adaptation

Ievgen Redko, Amaury Habrard, Marc Sebban

In many real-world applications, it may be desirable to benefit from a classifier trained on a given *source* task from some largely annotated dataset in order to address a different but related *target* task for which only weakly labeled data are available. Domain adaptation (DA) is the framework which aims at leveraging the statistical similarities between the source and target distributions to learn well. Current theoretical results show that the efficiency of DA algorithms depends on (i) their capacity of minimizing the divergence between the source and target domains and (ii) the existence of a good hypothesis that commits few errors in both domains. While most of the work in DA has focused on new divergence measures, the second aspect, often modeled as the *capability term*, remains surprisingly under-investigated. In this paper, we show that the problem of the best joint hypothesis estimation can be reformulated using a Wasserstein distance-based error function in the context of multi-source DA. Based on this idea, we provide a theoretical analysis of the capability term and derive inequalities allowing us to estimate it from finite samples. We empirically illustrate the proposed idea on different data sets.

The Teaching Size: Computable Teachers and Learners for Universal Languages
Jan Arne Telle, José Hernández-Orallo, Cèsar Ferri

The theoretical hardness of machine teaching has usually been analyzed for a range of concept languages under several variants of the teaching dimension: the minimum number of examples that a teacher needs to figure out so that the learner identifies the concept. However, for languages where concepts have structure (and hence size), such as Turing-complete languages, a low teaching dimension can be achieved at the cost of using very large examples, which are hard to process by the learner. In this paper we introduce the *teaching size*, a more intuitive way of assessing the theoretical feasibility of teaching concepts for structured languages. In the most general case of universal languages, we show that focusing on the total size of a witness set rather than its cardinality, we can teach all total functions that are computable within some fixed time bound. We complement the theoretical results with a range of experimental results on a simple Turing-complete language, showing how teaching dimension and teaching size differ in practice. Quite remarkably, we found that witness sets are usually smaller than the programs they identify, which is an illuminating justification of why machine teaching from examples makes sense at all.

On PAC-Bayesian Bounds for Random Forests
Stephan S. Lorenzen, Christian Igel, Yevgeny Seldin

Existing guarantees in terms of rigorous upper bounds on the generalization error for the original random forest algorithm, one of the most frequently used machine learning methods, are unsatisfying. We discuss and evaluate various PAC-Bayesian approaches to derive such bounds. The bounds do not require additional hold-out data, because the out-of-bag samples from the bagging in the training process can be exploited. A random forest predicts by taking a majority vote of an ensemble of decision trees. The first approach is to bound the error of the vote by twice the error of the corresponding Gibbs classifier (classifying with a single member of the ensemble selected at random). However, this approach does not take into account the effect of averaging out of errors of individual classifiers when taking the majority vote. This effect provides a significant boost in performance when the errors are independent or negatively correlated, but when the correlations are strong the advantage from taking the majority vote is small. The second approach based on PAC-Bayesian *C*-bounds takes dependencies between ensemble members into account, but it requires estimating correlations between the errors of the individual classifiers. When the correlations are high or the estimation is poor, the bounds degrade. In our experiments, we compute generalization bounds for random forests on various benchmark data sets. Because the individual decision trees already perform well, their predictions are highly correlated and the *C*-bounds do not lead to satisfactory results. For the same reason, the bounds based on the analysis of Gibbs classifiers are typically superior and often reasonably tight. Bounds based on a validation set coming at the cost of a smaller training set gave better performance guarantees, but worse performance in most experiments.

Nuclear Discrepancy for Single-Shot Batch Active Learning
Tom J. Viering, Jesse H. Krijthe, Marco Loog

Active learning algorithms propose what data should be labeled given a pool of unlabeled data. Instead of selecting randomly what data to annotate, active learning strategies aim to select data so as to get a good predictive model with as little labeled samples as possible. Single-shot batch active learners select all samples to be labeled in a single step, before any labels are observed. We study single-shot active learners that minimize generalization bounds to select a representative sample, such as the maximum mean discrepancy (MMD) active learner. We prove that a related bound, the

discrepancy, provides a tighter worst-case bound. We study these bounds probabilistically, which inspires us to introduce a novel bound, the nuclear discrepancy (ND). The ND bound is tighter for the expected loss under optimistic probabilistic assumptions. Our experiments show that the MMD active learner performs better than the discrepancy in terms of the mean squared error, indicating that tighter worst case bounds do not imply better active learning performance. The proposed active learner improves significantly upon the MMD and discrepancy in the realizable setting and a similar trend is observed in the agnostic setting, showing the benefits of a probabilistic approach to active learning. Our study highlights that assumptions underlying generalization bounds can be equally important as bound-tightness, when it comes to active learning performance. Code for reproducing our experimental results can be found at github.com/tomviering/NuclearDiscrepancy.

Large-Scale Learning, Wednesday, 16:20 – 17:30

Research Track

Session Chair: TBA

Distributed Learning of Non-Convex Linear Models with One Round of Communication
M. Izbicki (Claremont McKenna College), C. R. Shelton (UC Riverside)

We present the optimal weighted average (OWA) distributed learning algorithm for linear models. OWA achieves statistically optimal learning rates, uses only one round of communication, works on non-convex problems, and supports a fast cross validation procedure. The OWA algorithm first trains local models on each of the compute nodes; then a master machine merges the models using a second round of optimization. This second optimization uses only a small fraction of the data, and so has negligible computational cost. Compared with similar distributed estimators that merge locally trained models, OWA either has stronger statistical guarantees, is applicable to more models, or has a more computationally efficient merging procedure.

SLSGD: Secure and Efficient Distributed On-device Machine Learning
Reproducible Research
C. Xie, O. Koyejo, I. Gupta (University of Illinois at Urbana-Champaign)

We consider distributed on-device learning with limited communication and security requirements. We propose a new robust distributed optimization algorithm with efficient communication and attack tolerance. The proposed algorithm has provable convergence and robustness under non-IID settings. Empirical results show that the proposed algorithm stabilizes the convergence and tolerates data poisoning on a small number of workers.

Trade-offs in Large-Scale Distributed Tuplewise Estimation and Learning
Reproducible Research
R. Vogel (Telecom Paris, LTCI, Institut Polytechnique de Paris; IDEMIA), A. Bellet (INRIA), S. Clémençon (Telecom Paris, LTCI, Institut Polytechnique de Paris), O. Jelassi (Telecom Paris, LTCI, Institut Polytechnique de Paris), G. Papa (Telecom Paris, LTCI, Institut Polytechnique de Paris)

The development of cluster computing frameworks has allowed practitioners to scale out various statistical estimation and machine learning algorithms with minimal programming effort. This is especially true for machine learning problems whose objective function is nicely separable across individual data points, such as classification and regression. In contrast, statistical learning tasks involving pairs (or more generally tuples) of data points - such as metric learning, clustering or ranking - do not lend themselves as easily to data-parallelism and in-memory computing. In this paper, we investigate how to balance between statistical performance and computational efficiency in such

distributed tuplewise statistical problems. We first propose a simple strategy based on occasionally repartitioning data across workers between parallel computation stages, where the number of repartitioning steps rules the trade-off between accuracy and runtime. We then present some theoretical results highlighting the benefits brought by the proposed method in terms of variance reduction, and extend our results to design distributed stochastic gradient descent algorithms for tuplewise empirical risk minimization. Our results are supported by numerical experiments in pairwise statistical estimation and learning on synthetic and real-world datasets.

Applications 2, Wednesday, 16:20 – 17:30

ADS Track

Session Chair: Tassadit Bouadi

The Search for Equations - Learning to Identify Similarities between Mathematical Expressions

Reproducible Research

L. Pfahler, J. Schill, K. Morik (TU Dortmund University)

On your search for scientific articles relevant to your research question, you judge the relevance of a mathematical expression that you stumble upon using extensive background knowledge about the domain, its problems and its notations. We wonder if machine learning can support this process and work toward implementing a search engine for mathematical expressions in scientific publications. Thousands of scientific publications with millions of mathematical expressions or equations are accessible at arXiv.org. We want to use this data to learn about equations, their distribution and their relations in order to find similar equations. To this end we propose an embedding model based on convolutional neural networks that maps bitmap images of equations into a low-dimensional vector-space where similarity is evaluated via dot-product. However, no annotated similarity data is available to train this mapping. We mitigate this by proposing a number of different unsupervised proxy tasks that use available features as weak labels. We evaluate our system using a number of metrics, including results on a small hand-labeled subset of equations. In addition, we show and discuss a number of result-sets for some sample queries. The results show that we are able to automatically identify related mathematical expressions. Our dataset is published at https://whadup.github.io/EquationLearning/ and we invite the community to use it.

Data-driven Policy on Feasibility Determination for the Train Shunting Problem

P. R. de Oliveira da Costa (Eindhoven University of Technology), J. Rhuggenaath (Eindhoven University of Technology), Y. Zhang (Eindhoven University of Technology), A. Akcay (Eindhoven University of Technology), W.-J. Lee (Dutch Railways), U. Kaymak (Eindhoven University of Technology)

Parking, matching, scheduling, and routing are common problems in train maintenance. In particular, train units are commonly maintained and cleaned at dedicated shunting yards. The planning problem that results from such situations is referred to as the Train Unit Shunting Problem (TUSP). This problem involves matching arriving train units to service tasks and determining the schedule for departing trains. The TUSP is an important problem as it is used to determine the capacity of shunting yards and arises as a sub-problem of more general scheduling and planning problems. In this paper, we consider the case of the Dutch Railways (NS) TUSP. As the TUSP is complex, NS currently uses a local search (LS) heuristic to determine if an instance of the TUSP has a feasible solution. Given the number of shunting yards and the size of the planning problems, improving the evaluation speed of the LS brings significant computational gain. In this work, we use a machine learning approach that complements the LS and accelerates the search process. We use a Deep Graph Convolutional Neural Network (DGCNN) model to predict the feasibility of solutions obtained during the run of the LS heuristic. We

use this model to decide whether to continue or abort the search process. In this way, the computation time is used more efficiently as it is spent on instances that are more likely to be feasible. Using simulations based on real-life instances of the TUSP, we show how our approach improves upon the previous method on prediction accuracy and leads to computational gains for the decision-making process.

Automated Data Transformation with Inductive Programming and Dynamic Background Knowledge

Reproducible Research

L. Contreras-Ochando (Universitat Politècnica de València), C. Ferri (Universitat Politècnica de València), J. Hernández-Orallo (Universitat Politècnica de València), F. Martínez-Plumed (Universitat Politècnica de València), M. J. Ramírez-Quintana (Universitat Politècnica de València), S. Katayama (University of Miyazaki)

Data quality is essential for database integration, machine learning and data science in general. Despite the increasing number of tools for data preparation, the most tedious tasks of data wrangling -and feature manipulation in particular- still resist automation partly because the problem strongly depends on domain information. For instance, if the strings “17th of August of 2017” and “2017-08-17” are to be formatted into “08/17/2017” to be properly recognised by a data analytics tool, humans usually process this in two steps: (1) they recognise that this is about dates and (2) they apply conversions that are specific to the date domain. However, the mechanisms to manipulate dates are very different from those to manipulate addresses. This requires huge amounts of background knowledge, which usually becomes a bottleneck as the diversity of domains and formats increases. In this paper we help alleviate this problem by using inductive programming (IP) with a dynamic background knowledge (BK) fueled by a machine learning meta-model that selects the domain, the primitives (or both) from several descriptive features of the data wrangling problem. We illustrate these new alternatives for the automation of data format transformation, which we evaluate on an integrated benchmark and code for data wrangling, which we share publicly for the community.

Privacy & Security, Wednesday, 16:20 – 17:30

Research Track

Session Chair: TBA

A Differentially Private Kernel Two-Sample Test

Reproducible Research

A. Raj (Max Planck Institute for Intelligent Systems, Tubingen), H. C. L. Law (University of Oxford), D. Sejdinovic (University of Oxford), M. Park (Max Planck Institute for Intelligent Systems, Tubingen)

Kernel two-sample testing is a useful statistical tool in determining whether data samples arise from different distributions without imposing any parametric assumptions on those distributions. However, raw data samples can expose sensitive information about individuals who participate in scientific studies, which makes the current tests vulnerable to privacy breaches. Hence, we design a new framework for kernel two-sample testing conforming to differential privacy constraints, in order to guarantee the privacy of subjects in the data. Unlike existing differentially private parametric tests that simply add noise to data, kernel-based testing imposes a challenge due to a complex dependence of test statistics on the raw data, as these statistics correspond to estimators of distances between representations of probability measures in Hilbert spaces. Our approach considers finite dimensional approximations to those representations. As a result, a simple chi-squared test is obtained, where a test statistic depends on a mean and covariance of empirical differences between the samples, which we perturb for a privacy guarantee. We investigate the utility of our framework in two realistic

settings and conclude that our method requires only a relatively modest increase in sample size to achieve a similar level of power to the non-private tests in both settings.

Learning to Signal in the Goldilocks Zone: Improving Adversary Compliance in Security Games

S. Cooney (University of Southern California), K. Wang (University of Southern California), E. Bondi (University of Southern California), T. Nguyen (University of Oregon), P. Vayanos (University of Southern California), H. Winetrobe (University of Southern California), E. A. Cranford (Carnegie Mellon University), C. Gonzalez (Carnegie Mellon University), C. Lebiere (Carnegie Mellon University), M. Tambe (University of Southern California)

Many real-world security scenarios can be modeled via a game-theoretic framework known as a security game in which there is a defender trying to protect potential targets from an attacker. Recent work in security games has shown that deceptive signaling by the defender can convince an attacker to withdraw his attack. For instance, a warning message to commuters indicating speed enforcement is in progress ahead might lead to them driving more slowly, even if it turns out no enforcement is in progress. However, the results of this work are limited by the unrealistic assumption that the attackers will behave with perfect rationality, meaning they always choose an action that gives them the best expected reward. We address the problem of training boundedly rational (human) attackers to comply with signals via repeated interaction with signaling without incurring a loss to the defender, and offer the four following contributions: (i) We learn new decision tree and neural network-based models of attacker compliance with signaling. (ii) Based on these machine learning models of a boundedly rational attacker's response to signaling, we develop a theory of signaling in the Goldilocks zone, a balance of signaling and deception that increases attacker compliance and improves defender utility. (iii) We present game-theoretic algorithms to solve for signaling schemes based on the learned models of attacker compliance with signaling. (iv) We conduct extensive human subject experiments using an online game. The game simulates the scenario of an inside attacker trying to steal sensitive information from company computers, and results show that our algorithms based on learned models of attacker behavior lead to better attacker compliance and improved defender utility compared to the state-of-the-art algorithm for rational attackers with signaling.

Joint Detection of Malicious Domains and Infected Clients

Paul Prasse, René Knaebel, Lukáš Machlica, Tomáš Pevný, Tobias Scheffer

Detection of malware-infected computers and detection of malicious web domains based on their encrypted HTTPS traffic are challenging problems, because only addresses, timestamps, and data volumes are observable. The detection problems are coupled, because infected clients tend to interact with malicious domains. Traffic data can be collected at a large scale, and antivirus tools can be used to identify infected clients in retrospect. Domains, by contrast, have to be labeled individually after forensic analysis. We explore transfer learning based on sluice networks; this allows the detection models to bootstrap each other. In a large-scale experimental study, we find that the model outperforms known reference models and detects previously unknown malware, previously unknown malware families, and previously unknown malicious domains.

Abstracts Thursday

Deep Learning 2, Thursday, 11:00 – 12:40

Research Track

Session Chair: Dino Ienco

L_0-ARM: Network Sparsification via Stochastic Binary Optimization

Y. Li, S. Ji (Georgia State University)

We consider network sparsification as an L_0-norm regularized binary optimization problem, where each unit of a neural network (e.g., weight, neuron, or channel, etc.) is attached with a stochastic binary gate, whose parameters are jointly optimized with original network parameters. The Augment-Reinforce-Merge (ARM), a recently proposed unbiased gradient estimator, is investigated for this binary optimization problem. Compared to the hard concrete gradient estimator from Louizos et al., ARM demonstrates superior performance of pruning network architectures while retaining almost the same accuracies of baseline methods. Similar to the hard concrete estimator, ARM also enables conditional computation during model training but with improved effectiveness due to the exact binary stochasticity. Thanks to the flexibility of ARM, many smooth or non-smooth parametric functions, such as scaled sigmoid or hard sigmoid, can be used to parameterize this binary optimization problem and the unbiasness of the ARM estimator is retained, while the hard concrete estimator has to rely on the hard sigmoid function to achieve conditional computation and thus accelerated training. Extensive experiments on multiple public datasets demonstrate state-of-the-art pruning rates with almost the same accuracies of baseline methods. The resulting algorithm L_0-ARM sparsifies the Wide-ResNet models on CIFAR-10 and CIFAR-100 while the hard concrete estimator cannot.

Learning with Random Learning Rates

Reproducible Research

L. Blier (TAU, LRI, Inria, Université Paris Sud; Facebook AI Research), P. Wolinski (TAU, LRI, Inria, Université Paris Sud), Y. Ollivier (Facebook AI Research)

In neural network optimization, the learning rate of the gradient descent strongly affects performance. This prevents reliable out-of-the-box training of a model on a new problem. We propose the All Learning Rates At Once (Alrao) algorithm for deep learning architectures: each neuron or unit in the network gets its own learning rate, randomly sampled at startup from a distribution spanning several orders of magnitude. The network becomes a mixture of slow and fast learning units. Surprisingly, Alrao performs close to SGD with an optimally tuned learning rate, for various tasks and network architectures. In our experiments, all Alrao runs were able to learn well without any tuning.

Single-Path NAS: Designing Hardware-Efficient ConvNets in less than 4 Hours

Reproducible Research

D. Stamoulis (Carnegie Mellon University), R. Ding (Carnegie Mellon University), D. Wang (Microsoft), D. Lymberopoulos (Microsoft), B. Priyantha (Microsoft), J. Liu (Harbin Institute of Technology), D. Marculescu (Carnegie Mellon University)

Can we automatically design a Convolutional Network (ConvNet) with the highest image classification accuracy under the latency constraint of a mobile device? Neural architecture search (NAS) has revolutionized the design of hardware-efficient ConvNets by automating this process. However, the NAS problem remains challenging due to the combinatorially large design space, causing a significant searching time (at least 200 GPU-hours). To alleviate this complexity, we propose Single-Path NAS, a

novel differentiable NAS method for designing hardware-efficient ConvNets in less than 4 hours. Our contributions are as follows: 1. Single-path search space: Compared to previous differentiable NAS methods, Single-Path NAS uses one single-path over-parameterized ConvNet to encode all architectural decisions with shared convolutional kernel parameters, hence drastically decreasing the number of trainable parameters and the search cost down to few epochs. 2. Hardware-efficient ImageNet classification: Single-Path NAS achieves 74.96% top-1 accuracy on ImageNet with 79ms latency on a Pixel 1 phone, which is state-of-the-art accuracy compared to NAS methods with similar inference latency constraints (≤ 80 ms). 3. NAS efficiency: Single-Path NAS search cost is only 8 epochs (30 TPU-hours), which is up to 5,000x faster compared to prior work. 4. Reproducibility: Unlike all recent mobile-efficient NAS methods which only release pretrained models, we open-source our entire codebase at https://github.com/dstamoulis/single-path-nas.

Multitask Hopfield Networks

M. Frasca, G. Grossi, G. Valentini (Università degli Studi di Milano)

Multitask algorithms typically use task similarity information as a bias to speed up and improve the performance of learning processes. Tasks are learned jointly, sharing information across them, in order to construct models more accurate than those learned separately over single tasks. In this contribution, we present the first multitask model, to our knowledge, based on Hopfield Networks (HNs), named HoMTask. We show that by appropriately building a unique HN embedding all tasks, a more robust and effective classification model can be learned. HoMTask is a transductive semi-supervised parametric HN, that minimizes an energy function extended to all nodes and to all tasks under study. We provide theoretical evidence that the optimal parameters automatically estimated by HoMTask make coherent the model itself with the prior knowledge (connection weights and node labels). The convergence properties of HNs are preserved, and the fixed point reached by the network dynamics gives rise to the prediction of unlabeled nodes. The proposed model improves the classification abilities of single-task HNs on a preliminary benchmark comparison, and achieves competitive performance with state-of-the-art semi-supervised graph-based algorithms.

Deep Eyedentification: Biometric Identification using Micro-Movements of the Eye

Reproducible Research

L. A. Jäger (University of Potsdam), S. Makowski (University of Potsdam), P. Prasse (University of Potsdam), S. Liehr (Independent researcher), M. Seidler (University of Potsdam), T. Scheffer (University of Potsdam)

We study involuntary micro-movements of the eye for biometric identification. While prior studies extract lower-frequency macro-movements from the output of video-based eye-tracking systems and engineer explicit features of these macro-movements, we develop a deep convolutional architecture that processes the raw eye-tracking signal. Compared to prior work, the network attains a lower error rate by one order of magnitude and is faster by two orders of magnitude: it identifies users accurately within seconds.

Computer Vision & Explanation, Thursday, 11:00 – 12:40

ADS Track

Session Chair: Myra Spiliopoulou

Automatic Recognition of Student Engagement using Deep Learning and Facial Expression

Reproducible Research

O. M. Nezami (Macquarie University; CSIRO's Data61), M. Dras (Macquarie University), L. Hamey (Macquarie University), D. Richards (Macquarie University), S. Wan (CSIRO's Data61), C. Paris (CSIRO's Data61)

Engagement is a key indicator of the quality of learning experience, and one that plays a major role in developing intelligent educational interfaces. Any such interface requires the ability to recognise the level of engagement in order to respond appropriately; however, there is very little existing data to learn from, and new data is expensive and difficult to acquire. This paper presents a deep learning model to improve engagement recognition from images that overcomes the data sparsity challenge by pre-training on readily available basic facial expression data, before training on specialised engagement data. In the first of two steps, a facial expression recognition model is trained to provide a rich face representation using deep learning. In the second step, we use the model's weights to initialize our deep learning based model to recognize engagement; we term this the engagement model. We train the model on our new engagement recognition dataset with 4627 engaged and disengaged samples. We find that the engagement model outperforms effective deep learning architectures that we apply for the first time to engagement recognition, as well as approaches using histogram of oriented gradients and support vector machines.

Pushing the Limits of Exoplanet Discovery via Direct Imaging with Deep Learning

Reproducible Research

K. H. Yip (University College London), N. Nikolaou (University College London), P. Coronica (University of Cambridge), A. Tsiaras (University College London), B. Edwards (University College London), Q. Changeat (University College London), M. Morvan (University College London), B. Biller (University of Edinburgh), S. Hinkley (University of Exeter), J. Salmond (University of Cambridge), M. Archer (University of Cambridge), P. Sumption (University of Cambridge), E. Choquet (Aix Marseille Univ), R. Soummer (STScI), L. Pueyo (STScI), I. P. Waldmann (University College London)

Further advances in exoplanet detection and characterisation require sampling a diverse population of extrasolar planets. One technique to detect these distant worlds is through the direct detection of their thermal emission. The so-called direct imaging technique, is suitable for observing young planets far from their star. These are very low signal-to-noise-ratio (SNR) measurements and limited ground truth hinders the use of supervised learning approaches. In this paper, we combine deep generative and discriminative models to bypass the issues arising when directly training on real data. We use a Generative Adversarial Network to obtain a suitable dataset for training Convolutional Neural Network classifiers to detect and locate planets across a wide range of SNRs. Tested on artificial data, our detectors exhibit good predictive performance and robustness across SNRs. To demonstrate the limits of the detectors, we provide maps of the precision and recall of the model per pixel of the input image. On real data, the models can re-confirm bright source detections.

Marine Mammal Species Classification using Convolutional Neural Networks and a Novel Acoustic Representation

M. Thomas (Dalhousie University Faculty of Computer Science), B. Martin (JASCO Applied Sciences), K. Kowarski (JASCO Applied Sciences), B. Gaudet (JASCO Applied Sciences), S. Matwin (Dalhousie University Faculty of Computer Science, Institute of Computer Science Polish Academy of Sciences)

Research into automated systems for detecting and classifying marine mammals in acoustic recordings is expanding internationally due to the necessity to analyze large collections of data for conservation purposes. In this work, we present a Convolutional Neural Network that is capable of classifying the vocalizations of three species of whales, non-biological sources of noise, and a fifth class pertaining to ambient noise. In this way, the classifier is capable of detecting the presence and absence of whale vocalizations in an acoustic recording. Through transfer learning, we show that the classifier is capable of learning high-level representations and can generalize to additional species. We also propose a novel representation of acoustic signals that builds upon the commonly used spectrogram representation by way of interpolating and stacking multiple spectrograms produced using different Short-time Fourier Transform (STFT) parameters. The proposed representation is particularly effective for the task of marine mammal species classification where the acoustic events we are attempting to classify are sensitive to the parameters of the STFT.

Learning Disentangled Representations of Satellite Image Time Series

E. H. Sanchez (IRT Saint Exupéry, Toulouse, IRIT, Université Toulouse III - Paul Sabatier), M. Serrurier (IRT Saint Exupéry, Toulouse, IRIT, Université Toulouse III - Paul Sabatier), M. Ortner (IRT Saint Exupery, Toulouse)

In this paper, we investigate how to learn a suitable representation of satellite image time series in an unsupervised manner by leveraging large amounts of unlabeled data. Additionally, we aim to disentangle the representation of time series into two representations: a shared representation that captures the common information between the images of a time series and an exclusive representation that contains the specific information of each image of the time series. To address these issues, we propose a model that combines a novel component called cross-domain autoencoders with the variational autoencoder (VAE) and generative adversarial network (GAN) methods. In order to learn disentangled representations of time series, our model learns the multimodal image-to-image translation task. We train our model using satellite image time series provided by the Sentinel-2 mission. Several experiments are carried out to evaluate the obtained representations. We show that these disentangled representations can be very useful to perform multiple tasks such as image classification, image retrieval, image segmentation and change detection.

J3R: Joint Multi-task Learning of Ratings and Review Summaries for Explainable Recommendation

A. PVS (Technische Universität Darmstadt), Y. Ren (RMIT University), C. M. Meyer (Technische Universität Darmstadt), J. Chan (RMIT University), Z. Bao (RMIT University), M. Sanderson (RMIT University)

We learn user preferences from ratings and reviews by using multi-task learning (MTL) of rating prediction and summarization of item reviews. Reviews of an item tend to describe detailed user preferences (e.g., the cast, genre, or screenplay of a movie). A summary of such a review or a rating

describes an overall user experience of the item. Our objective is to learn latent vectors which are shared across rating prediction and review summary generation. Additionally, the learned latent vectors and the generated summary act as explanations for the recommendation. Our MTL-based approach J3R uses a multi-layer perceptron for rating prediction, combined with pointer-generator networks with attention mechanism for the summarization component. We provide empirical evidence for joint learning of rating prediction and summary generation being beneficial for recommendation by conducting experiments on the Yelp dataset and six domains of the Amazon 5-core dataset. Additionally, we provide two ways of explanations visualizing (a) the user vectors on different topics of a domain, computed from our J3R approach and (b) a ten-word review summary of a review and the attention highlights generated on the review based on the user-item vectors.

Supervised Learning 2, Thursday, 11:00 – 12:40

Research Track

Session Chair: Thomas Gaertner

Exploiting the Earth's Spherical Geometry to Geolocate Images

Reproducible Research

M. Izbicki (Claremont McKenna College), E. E. Papalexakis (University of California Riverside), V. J. Tsotras (University of California Riverside)

Existing methods for geolocating images use standard classification or image retrieval techniques. These methods have poor theoretical properties because they do not take advantage of the earth's spherical geometry. In some cases, they require training data sets that grow exponentially with the number of feature dimensions. This paper introduces the first image geolocation method that exploits the earth's spherical geometry. Our method is based on the Mixture of von-Mises Fisher (MvMF) distribution, which is a spherical analogue of the popular Gaussian mixture model. We prove that this method requires only a dataset of size linear in the number of feature dimensions, and empirical results show that our method outperforms previous methods with orders of magnitude less training data and computation.

Beyond the Selected Completely At Random Assumption for Learning from Positive and Unlabeled Data

Reproducible Research

J. Bekker, P. Robberechts, J. Davis (KU Leuven)

Most positive and unlabeled data is subject to selection biases. The labeled examples can, for example, be selected from the positive set because they are easier to obtain or more obviously positive. This paper investigates how learning can be enabled in this setting. We propose and theoretically analyze an empirical-risk-based method for incorporating the labeling mechanism. 6Additionally, we investigate under which assumptions learning is possible when the labeling mechanism is not fully understood and propose a practical method to enable this. Our empirical analysis supports the theoretical results and shows that taking into account the possibility of a selection bias, even when the labeling mechanism is unknown, improves the trained classifiers.

Distribution-Free Uncertainty Quantification for Kernel Methods by Gradient Perturbations

Balázs Cs. Csáji, Krisztián B. Kis

We propose a data-driven approach to quantify the uncertainty of models constructed by kernel methods. Our approach minimizes the needed distributional assumptions, hence, instead of working with, for example, Gaussian processes or exponential families, it only requires knowledge about some

mild regularity of the measurement noise, such as it is being symmetric or exchangeable. We show, by building on recent results from finite-sample system identification, that by perturbing the residuals in the gradient of the objective function, information can be extracted about the amount of uncertainty our model has. Particularly, we provide an algorithm to build exact, non-asymptotically guaranteed, distribution-free confidence regions for ideal, noise-free representations of the function we try to estimate. For the typical convex quadratic problems and symmetric noises, the regions are star convex centered around a given nominal estimate, and have efficient ellipsoidal outer approximations. Finally, we illustrate the ideas on typical kernel methods, such as LS-SVC, KRR, ε-SVR and kernelized LASSO.

Cost Sensitive Evaluation of Instance Hardness in Machine Learning

R. B. C. Prudêncio (Universidade Federal de Pernambuco)

Measuring hardness of individual instances in machine learning contributes to a deeper analysis of learning performance. This work proposes instance hardness measures for binary classification in cost-sensitive scenarios. Here cost curves are generated for each instance, defined as the loss observed for a pool of learning models for that instance along the range of cost proportions. Instance hardness is defined as the area under the cost curves and can be seen as an expected loss of difficulty along cost proportions. Different cost curves were proposed by considering common decision threshold choice methods in literature, thus providing alternative views of instance hardness.

Classification with Label Noise: A Markov Chain Sampling Framework

Zijin Zhao, Lingyang Chu, Dacheng Tao, Jian Pei

The effectiveness of classification methods relies largely on the correctness of instance labels. In real applications, however, the labels of instances are often not highly reliable due to the presence of label noise. Training effective classifiers in the presence of label noise is a challenging task that enjoys many real-world applications. In this paper, we propose a Markov chain sampling (MCS) framework that accurately identifies mislabeled instances and robustly learns effective classifiers. MCS builds a Markov chain where each state uniquely represents a set of randomly sampled instances. We show that the Markov chain has a unique stationary distribution, which puts much larger probability weights on the states dominated by correctly labeled instances than the states dominated by mislabeled instances. We propose a Markov Chain Monte Carlo sampling algorithm to approximate the stationary distribution, which is further used to compute the mislabeling probability for each instance, and train noise-resistant classifiers. The MCS framework is highly compatible with a wide spectrum of classifiers that produce probabilistic classification results. Extensive experiments on both real and synthetic data sets demonstrate the superior effectiveness and efficiency of the proposed MCS framework.

Pattern Mining, Thursday, 11:00 – 12:40

Research Track

Session Chair: Bruno Crémilleux

DEvIANT: Discovering Significant Exceptional (Dis-)Agreement Within Groups

Reproducible Research

A. Belfodil (INSA Lyon), W. Duivesteijn (Technische Universiteit Eindhoven), M. Plantevit (Univ Lyon), S. Cazalens (INSA Lyon), P. Lamarre (INSA Lyon)

We strive to find contexts (i.e., subgroups of entities) under which exceptional (dis-)agreement occurs among a group of individuals, in any type of data featuring individuals (e.g., parliamentarians, customers) performing observable actions (e.g., votes, ratings) on entities (e.g., legislative procedures, movies). To this end, we introduce the problem of discovering statistically significant exceptional contextual intra-group agreement patterns. To handle the sparsity inherent to voting and rating data, we use Krippendorff's Alpha measure for assessing the agreement among individuals. We devise a branch-and-bound algorithm, named DEvIANT, to discover such patterns. DEvIANT exploits both closure operators and tight optimistic estimates. We derive analytic approximations for the confidence intervals (CIs) associated with patterns for a computationally efficient significance assessment. We prove that these approximate CIs are nested along specialization of patterns. This allows to incorporate pruning properties in DEvIANT to quickly discard non-significant patterns. Empirical study on several datasets demonstrates the efficiency and the usefulness of DEvIANT.

Maximal Closed Set and Half-Space Separations in Finite Closure Systems

F. Seiffarth (University of Bonn), T. Horváth (University of Bonn; Fraunhofer IAIS, Sankt Augustin; Fraunhofer Center for Machine Learning, Sankt Augustin), S. Wrobel (University of Bonn; Fraunhofer IAIS, Sankt Augustin; Fraunhofer Center for Machine Learning, Sankt Augustin)

Motivated by various binary classification problems in structured data (e.g., graphs or other relational and algebraic structures), we investigate some algorithmic properties of closed set and half-space separation in abstract closure systems. Assuming that the underlying closure system is finite and given by the corresponding closure operator, we formulate some negative and positive complexity results for these two separation problems. In particular, we prove that deciding half-space separability in abstract closure systems is NP-complete in general. On the other hand, for the relaxed problem of maximal closed set separation we propose a simple greedy algorithm and show that it is efficient and has the best possible lower bound on the number of closure operator calls. As a second direction to overcome the negative result above, we consider Kakutani closure systems and show first that our greedy algorithm provides an algorithmic characterization of this kind of set systems. As one of the major potential application fields, we then focus on Kakutani closure systems over graphs and generalize a fundamental characterization result based on the Pasch axiom to graph structure partitioning of finite sets. Though the primary focus of this work is on the generality of the results obtained, we experimentally demonstrate the practical usefulness of our approach on vertex classification in different graph datasets.

Multi-location Visibility Query Processing Using Portion-based Transactional Modeling and Pattern Mining
Lakshmi Gangumalla, P. Krishna Reddy, Anirban Mondal

Visibility computation is critical in spatial databases for realizing various interesting and diverse applications such as defence-related surveillance, identifying interesting spots in tourist places and online warcraft games. Existing works address the problem of identifying individual locations for maximizing the visibility of a given target object. However, in case of many applications, a set of locations may be more effective than just individual locations towards maximizing the visibility of the given target object. In this paper, we introduce the Multi-Location Visibility (MLV) query. An MLV query determines the top-k query locations from which the visibility of a given target object can be maximized. We propose a portion-based transactional framework and coverage pattern mining based algorithm to process MLV queries. Our performance evaluation with real datasets demonstrates the effectiveness of the proposed scheme in terms of query processing time, pruning efficiency and target object visibility w.r.t. a recent existing scheme.

Mining Skypatterns in Fuzzy Tensors
Nicolas Nadisic, Aurélien Coussat, Loïc Cerf

Many data mining tasks rely on pattern mining. To identify the patterns of interest in a dataset, an analyst may define several measures that score, in different ways, the relevance of a pattern. Until recently, most algorithms have only handled constraints in an efficient way, i.e., every measure had to be associated with a user-defined threshold, which can be tricky to determine. *Skypatterns* were introduced to allow analysts to simply define the measures of interest, and to get as a result a set of globally optimal and semantically relevant patterns. Skypatterns are Pareto-optimal patterns: no other pattern scores better on one of the chosen measures and scores at least as well on every remaining measure. This article tackles the search of the skypatterns in a more general context than the 0/1 (aka Boolean) matrix: the *fuzzy tensor*. The proposed solution supports a large class of measures. After explaining why and how their common mathematical property enables a safe pruning of the search space, an algorithm is presented. It builds upon `multidupehack`, a generalist pattern mining framework, which is now able to efficiently list skypatterns in addition to enforcing constraints on them. Experiments on two real-world fuzzy tensors illustrate the versatility of the proposal. Other experiments show it is typically more than one order of magnitude faster than the state-of-the-art algorithms, which can only mine 0/1 matrices.

Research Track

Reproducible Research

Stochastic Activation Actor Critic Methods

W. Shang, H. van Hoof, M. Welling (University of Amsterdam, Bosch, Deltalab)

Stochastic elements in reinforcement learning (RL) have shown promise to improve exploration and handling of uncertainty, such as the utilization of stochastic weights in NoisyNets and stochastic policies in the maximum entropy RL frameworks. Yet effective and general approaches to include such elements in actor-critic models are still lacking. Inspired by the aforementioned techniques, we propose an effective way to inject randomness into actor-critic models to improve general exploratory behavior and reflect environment uncertainty. Specifically, randomness is added at the level of intermediate activations that feed into both policy and value functions to achieve better correlated and more complex perturbations. The proposed framework also features flexibility and simplicity, which allows straightforward adaptation to a variety of tasks. We test several actor-critic models enhanced with stochastic activations and demonstrate their effectiveness in a wide range of Atari 2600 games, a continuous control problem and a car racing task. Lastly, in a qualitative analysis, we present evidence of the proposed model adapting the noise in the policy and value functions to reflect uncertainty and ambiguity in the environment.

Compatible Natural Gradient Policy Search

Joni Pajarinen, Hong Linh Thai, Riad Akrour, Jan Peters, Gerhard Neumann

Trust-region methods have yielded state-of-the-art results in policy search. A common approach is to use KL-divergence to bound the region of trust resulting in a natural gradient policy update. We show that the natural gradient and trust region optimization are equivalent if we use the natural parameterization of a standard exponential policy distribution in combination with compatible value function approximation. Moreover, we show that standard natural gradient updates may reduce the entropy of the policy according to a wrong schedule leading to premature convergence. To control entropy reduction, we introduce a new policy search method called compatible policy search (COPOS) which bounds entropy loss. The experimental results show that COPOS yields state-of-the-art results in challenging continuous control tasks and in discrete partially observable tasks.

Stochastic One-Sided Full-Information Bandit

H. Zhao (Tsinghua University), W. Chen (Microsoft Research, Beijing)

In this paper, we study the stochastic version of the one-sided full information bandit problem, where we have K arms $[K] = \{1,2,\ldots,K\}$, and playing arm i would gain reward from an unknown distribution for arm i while obtaining reward feedback for all arms $j > i$. One-sided full information bandit can model the online repeated second-price auctions, where the auctioneer could select the reserved price in each round and the bidders only reveal their bids when their bids are higher than the reserved price. In this paper, we present an elimination-based algorithm to solve the problem. Our elimination-based algorithm achieves distribution independent regret upper bound $O(\sqrt{T \cdot \log(TK)})$, and distribution dependent bound $O\big((\log T + \log K) f(\Delta)\big)$, where T is the time horizon, Δ is a vector of gaps between the mean reward of arms and the mean reward of the best arm, and $f(\Delta)$ is a formula depending on the gap vector that we will specify in detail.

Our algorithm has the best theoretical regret upper bound so far. We also validate our algorithm empirically against other possible alternatives.

Practical Open-Loop Optimistic Planning

Reproducible Research

E. Leurent (SequeL team, INRIA Lille - Nord Europe, Renault Group), O.-A. Maillard (SequeL team, INRIA Lille - Nord Europe)

We consider the problem of online planning in a Markov Decision Process when given only access to a generative model, restricted to open-loop policies – i.e. sequences of actions – and under budget constraint. In this setting, the Open-Loop Optimistic Planning (OLOP) algorithm enjoys good theoretical guarantees but is overly conservative in practice, as we show in numerical experiments. We propose a modified version of the algorithm with tighter upper-confidence bounds, KL-OLOP, that leads to better practical performances while retaining the sample complexity bound. Finally, we propose an efficient implementation that significantly improves the time complexity of both algorithms.

An Engineered Empirical Bernstein Bound

Reproducible Research

M. A. Burgess (Australian National University), A. C. Chapman (University of Sydney), P. Scott (Australian National University)

We derive a tightened empirical Bernstein bound (EBB) on the variation of the sample mean from the population mean, and show that it improves the performance of upper confidence bound (UCB) methods in multi-armed bandit problems. Like other EBBs, our EBB is a concentration inequality for the variation of the sample mean in terms of the sample variance. Its derivation uses a combination of probability unions and Chernoff bounds for the mean of samples and mean of sample squares. Analysis reveals that our approach can tighten the best existing EBBs by about a third, and thereby halves the distance to a bound constructed with perfect variance information. We illustrate the practical usefulness of our novel EBB by applying it to a multi-armed bandit problem as a component of a UCB method. Our method outperforms existing approaches by producing lower expected regret than variants of UCB employing several other bounds, including state-of-the-art EBBs.

BelMan: An Information-Geometric Approach to Stochastic Bandits

Reproducible Research

D. Basu (Chalmers University of Technology), P. Senellart (DI ENS, ENS, CNRS, PSL University; INRIA), S. Bressan (National University of Singapore)

We propose a Bayesian information-geometric approach to the exploration-exploitation trade-off in stochastic multi-armed bandits. The uncertainty on reward generation and belief is represented using the manifold of joint distributions of rewards and beliefs. Accumulated information is summarised by the barycentre of joint distributions, the pseudobelief-reward. While the pseudobelief-reward facilitates information accumulation through exploration, another mechanism is needed to increase exploitation by gradually focusing on higher rewards, the pseudobelief-focal-reward. Our resulting algorithm, BelMan, alternates between projection of the pseudobelief-focal-reward onto belief-reward distributions to choose the arm to play, and projection of the updated belief-reward distributions onto the pseudobelief-focal-reward. We theoretically prove BelMan to be asymptotically optimal and to incur a sublinear regret growth. We instantiate BelMan to stochastic bandits with Bernoulli and exponential rewards, and to a real-life application of scheduling queueing bandits. Comparative evaluation with the state of the art shows that BelMan is not only competitive for Bernoulli bandits but in many cases also outperforms other approaches for exponential and queueing bandits.

Clustering, Thursday, 14:00 – 16:00

Research Track

Session Chair: Christel Vrain

Holistic Assessment of Structure Discovery Capabilities of Clustering Algorithms

Reproducible Research

F. Höppner, M. Jahnke (Ostfalia University of Applied Sciences)

Existing cluster validity indices often possess a similar bias as the clustering algorithm they were introduced for, e.g. to determine the optimal number of clusters. We suggest an efficient and holistic assessment of the structure discovery capabilities of clustering algorithms based on three criteria. We determine the robustness or stability of cluster assignments and interpret it as the confidence of the clustering algorithm in its result. This information is then used to label the data and evaluate the consistency of the stability-assessment with the notion of a cluster as an area of dense and separated data. The resulting criteria of stability, structure and consistency provide interpretable means to judge the capabilities of clustering algorithms without the typical biases of prominent indices, including the judgment of a clustering tendency.

k is the Magic Number - Inferring the Number of Clusters Through Nonparametric Concentration Inequalities

Reproducible Research

S. Hess, W. Duivesteijn (Technische Universiteit Eindhoven)

Most convex and nonconvex clustering algorithms come with one crucial parameter: the k in k-means. To this day, there is not one generally accepted way to accurately determine this parameter. Popular methods are simple yet theoretically unfounded, such as searching for an elbow in the curve of a given cost measure. In contrast, statistically founded methods often make strict assumptions over the data distribution or come with their own optimization scheme for the clustering objective. This limits either the set of applicable datasets or clustering algorithms. In this paper, we strive to determine the number of clusters by answering a simple question: given two clusters, is it likely that they jointly stem from a single distribution? To this end, we propose a bound on the probability that two clusters originate from the distribution of the unified cluster, specified only by the sample mean and variance. Our method is applicable as a simple wrapper to the result of any clustering method minimizing the objective of k-means, which includes Gaussian mixtures and Spectral Clustering. We focus in our experimental evaluation on an application for nonconvex clustering and demonstrate the suitability of our theoretical results. Our SpecialK clustering algorithm automatically determines the appropriate value for k, without requiring any data transformation or projection, and without assumptions on the data distribution. Additionally, it is capable to decide that the data consists of only a single cluster, which many existing algorithms cannot.

Uncovering Hidden Block Structure for Clustering

Reproducible Research

L. le Gorrec (Université de Toulouse), S. Mouysset (Université de Toulouse), P. A. Knight (STFC Rutherford Appleton Laboratory, CERFACS, Toulouse), I. S. Duff (University of Strathclyde), D. Ruiz (Université de Toulouse)

We present a multistage procedure to cluster directed and undirected weighted graphs by finding the block structure of their adjacency matrices. A central part of the process is to scale the adjacency matrix into a doubly-stochastic form, which permits detection of the whole matrix block structure with minimal spectral information (theoretically a single pair of singular vectors suffices). We present the different stages of our method, namely the impact of the doubly-stochastic scaling on singular

vectors, detection of the block structure by means of these vectors, and details such as cluster refinement and a stopping criterion. Then we test the algorithm's effectiveness by using it on two unsupervised classification tasks: community detection in networks and shape detection in clouds of points in two dimensions. By comparing results of our approach with those of widely used algorithms designed for specific purposes, we observe that our method is competitive (for community detection) if not superior (for shape detection) in comparison with existing methods.

A Framework for Parallelizing Hierarchical Clustering Methods

S. Lattanzi (Google Zürich), T. Lavastida (Carnegie Mellon University), K. Lu (Washington and Lee University), B. Moseley (Carnegie Mellon University)

Hierarchical clustering is a fundamental tool in data mining, machine learning and statistics. Popular hierarchical clustering algorithms include top-down divisive approaches such as bisecting k-means, k-median, and k-center and bottom-up agglomerative approaches such as single-linkage, average-linkage, and centroid-linkage. Unfortunately, only a few scalable hierarchical clustering algorithms are known, mostly based on the single-linkage algorithm. So, as datasets increase in size every day, there is a pressing need to scale other popular methods. We introduce efficient distributed algorithms for bisecting k-means, k-median, and k-center as well as centroid-linkage. In particular, we first formalize a notion of closeness for a hierarchical clustering algorithm, and then we use this notion to design new scalable distributed methods with strong worst-case bounds on the running time and the quality of the solutions. Finally, we show experimentally that the introduced algorithms are efficient and close to their sequential variants in practice.

Heavy-tailed kernels reveal a finer cluster structure in t-SNE visualisations

Reproducible Research

D. Kobak (University of Tubingen), G. Linderman (Yale University), S. Steinerberger (Yale University), Y. Kluger (Yale University, Yale School of Medicine), P. Berens (University of Tubingen)

T-distributed stochastic neighbour embedding (t-SNE) is a widely used data visualisation technique. It differs from its predecessor SNE by the low-dimensional similarity kernel: the Gaussian kernel was replaced by the heavy-tailed Cauchy kernel, solving the 'crowding problem' of SNE. Here, we develop an efficient implementation of t-SNE for a t-distribution kernel with an arbitrary degree of freedom ν, with $\nu \to \infty$ corresponding to SNE and $\nu = 1$ corresponding to the standard t-SNE. Using theoretical analysis and toy examples, we show that $\nu < 1$ can further reduce the crowding problem and reveal finer cluster structure that is invisible in standard t-SNE. We further demonstrate the striking effect of heavier-tailed kernels on large real-life data sets such as MNIST, single-cell RNA-sequencing data, and the HathiTrust library. We use domain knowledge to confirm that the revealed clusters are meaningful. Overall, we argue that modifying the tail heaviness of the t-SNE kernel can yield additional insight into the cluster structure of the data.

Noise-free Latent Block Model for High Dimensional Data

Charlotte Laclau, Vincent Brault

Co-clustering is known to be a very powerful and efficient approach in unsupervised learning because of its ability to partition data based on both the observations and the variables of a given dataset. However, in high-dimensional context co-clustering methods may fail to provide a meaningful result due to the presence of noisy and/or irrelevant features. In this paper, we tackle this issue by proposing a novel co-clustering model which assumes the existence of a noise cluster, that contains all irrelevant features. A variational expectation-maximization-based algorithm is derived for this task, where the

automatic variable selection as well as the joint clustering of objects and variables are achieved via a Bayesian framework. Experimental results on synthetic datasets show the efficiency of our model in the context of high-dimensional noisy data. Finally, we highlight the interest of the approach on two real datasets which goal is to study genetic diversity across the world.

Natural Language Processing, Thursday, 14:00 – 16:00

Research Track

Session Chair: Peggy Cellier

Unsupervised Sentence Embedding Using Document Structure-based Context

T. Lee and Y. Park (IBM T.J. Watson Research Center)

We present a new unsupervised method for learning general-purpose sentence embeddings. Unlike existing methods which rely on local contexts, such as words inside the sentence or immediately neighboring sentences, our method selects, for each target sentence, influential sentences from the entire document based on the document structure. We identify a dependency structure of sentences using metadata and text styles. Additionally, we propose an out-of-vocabulary word handling technique for the neural network outputs to model many domain-specific terms which were mostly discarded by existing sentence embedding training methods. We empirically show that the model relies on the proposed dependencies more than the sequential dependency in many cases. We also validate our model on several NLP tasks showing 23% *F*1-score improvement in coreference resolution in a technical domain and 5% accuracy increase in paraphrase detection compared to baselines.

Copy Mechanism and Tailored Training for Character-based Data-to-text Generation

Reproducible Research

M. Roberti (University of Turin), G. Bonetta (University of Turin), R. Cancelliere (University of Turin), P. Gallinari (Sorbonne Université, Criteo AI Lab)

In the last few years, many different methods have been focusing on using deep recurrent neural networks for natural language generation. The most widely used sequence-to-sequence neural methods are word-based: as such, they need a pre-processing step called delexicalization (conversely, relexicalization) to deal with uncommon or unknown words. These forms of processing, however, give rise to models that depend on the vocabulary used and are not completely neural. In this work, we present an end-to-end sequence-to-sequence model with attention mechanism which reads and generates at a character level, no longer requiring delexicalization, tokenization, nor even lowercasing. Moreover, since characters constitute the common “building blocks” of every text, it also allows a more general approach to text generation, enabling the possibility to exploit transfer learning for training. These skills are obtained thanks to two major features: (*) the possibility to alternate between the standard generation mechanism and a copy one, which allows to directly copy input facts to produce outputs, and (*) the use of an original training pipeline that further improves the quality of the generated texts. We also introduce a new dataset called E2E+, designed to highlight the copying capabilities of character-based models, that is a modified version of the well-known E2E dataset used in the E2E Challenge. We tested our model according to five broadly accepted metrics (including the widely used bleu), showing that it yields competitive performance with respect to both character-based and word-based approaches.

A Semi-discriminative Approach for Sub-sentence Level Topic Classification on a Small Dataset

Reproducible Research

C. Ferner, S. Wegenkittl (Salzburg University of Applied Sciences)

This paper aims at identifying sequences of words related to specific product components in online product reviews. A reliable baseline performance for this topic classification problem is given by a Max Entropy classifier which assumes independence over subsequent topics. However, the reviews exhibit an inherent structure on the document level allowing to frame the task as sequence classification problem. Since more flexible models from the class of Conditional Random Fields were not competitive because of the limited amount of training data available, we propose using a Hidden Markov Model instead and decouple the training of transition and emission probabilities. The discriminating power of the Max Entropy approach is used for the latter. Besides outperforming both standalone methods as well as more generic models such as linear-chain Conditional Random Fields, the combined classifier is able to assign topics on sub-sentence level although labeling in the training data is only available on sentence level.

Generating Black-Box Adversarial Examples for Text Classifiers Using a Deep Reinforced Model

P. Vijayaraghavan, D. Roy (MIT Media Lab)

Recently, generating adversarial examples has become an important means of measuring robustness of a deep learning model. Adversarial examples help us identify the susceptibilities of the model and further counter those vulnerabilities by applying adversarial training techniques. In natural language domain, small perturbations in the form of misspellings or paraphrases can drastically change the semantics of the text. We propose a reinforcement learning based approach towards generating adversarial examples in black-box settings. We demonstrate that our method is able to fool well-trained models for (a) IMDB sentiment classification task and (b) AG's news corpus news categorization task with significantly high success rates. We find that the adversarial examples generated are semantics-preserving perturbations to the original text.

NSEEN: Neural Semantic Embedding for Entity Normalization

Reproducible Research

S. Fakhraei, J. Mathew, J. L. Ambite (University of Southern California)

Much of human knowledge is encoded in text, available in scientific publications, books, and the web. Given the rapid growth of these resources, we need automated methods to extract such knowledge into machine-processable structures, such as knowledge graphs. An important task in this process is entity normalization, which consists of mapping noisy entity mentions in text to canonical entities in well-known reference sets. However, entity normalization is a challenging problem; there often are many textual forms for a canonical entity that may not be captured in the reference set, and entities mentioned in text may include many syntactic variations, or errors. The problem is particularly acute in scientific domains, such as biology. To address this problem, we have developed a general, scalable solution based on a deep Siamese neural network model to embed the semantic information about the entities, as well as their syntactic variations. We use these embeddings for fast mapping of new entities to large reference sets, and empirically show the effectiveness of our framework in challenging bio-entity normalization datasets.

Beyond Bag-of-Concepts: Vectors of Locally Aggregated Concepts

Reproducible Research

M. Grootendorst (Jheronimus Academy of Data Science), J. Vanschoren (Eindhoven University of Technology)

Bag-of-Concepts, a model that counts the frequency of clustered word embeddings (i.e., concepts) in a document, has demonstrated the feasibility of leveraging clustered word embeddings to create features for document representation. However, information is lost as the word embeddings themselves are not used in the resulting feature vector. This paper presents a novel text representation method, Vectors of Locally Aggregated Concepts (VLAC). Like Bag-of-Concepts, it clusters word embeddings for its feature generation. However, instead of counting the frequency of clustered word embeddings, VLAC takes each cluster's sum of residuals with respect to its centroid and concatenates those to create a feature vector. The resulting feature vectors contain more discriminative information than Bag-of-Concepts due to the additional inclusion of these first order statistics. The proposed method is tested on four different data sets for single-label classification and compared with several baselines, including TF-IDF and Bag-of-Concepts. Results indicate that when combining features of VLAC with TF-IDF significant improvements in performance were found regardless of which word embeddings were used.

Probabilistic Models 2, Thursday, 14:00 – 16:00

Research Track

Session Chair: Katharina Morik

FastPoint: Scalable Deep Point Processes

A. C. Türkmen (Bogazici University), Y. Wang (Amazon Research), A. J. Smola (Amazon Research)

We propose FastPoint, a novel multivariate point process that enables fast and accurate learning and inference. FastPoint uses deep recurrent neural networks to capture complex temporal dependency patterns among different marks, while self-excitation dynamics within each mark are modeled with Hawkes processes. This results in substantially more efficient learning and scales to millions of correlated marks with superior predictive accuracy. Our construction also allows for efficient and parallel sequential Monte Carlo sampling for fast predictive inference. FastPoint outperforms baseline methods in prediction tasks on synthetic and real-world high-dimensional event data at a small fraction of the computational cost.

Fine-Grained Explanations using Markov Logic

K. M. Al Farabi (University of Memphis), S. Sarkhel (Adobe Research), S. Dey (University of Illinois at Urbana-Champaign), D. Venugopal (University of Memphis)

Explaining the results of Machine learning algorithms is crucial given the rapid growth and potential applicability of these methods in critical domains including healthcare, defense, autonomous driving, etc. In this paper, we address this problem in the context of Markov Logic Networks (MLNs) which are highly expressive statistical relational models that combine first-order logic with probabilistic graphical models. MLNs in general are known to be interpretable models, i.e., MLNs can be understood more easily by humans as compared to models learned by approaches such as deep learning. However, at the same time, it is not straightforward to obtain human-understandable explanations specific to an observed inference result (e.g. marginal probability estimate). This is because, the MLN provides a lifted interpretation, one that generalizes to all possible worlds/instantiations, which are not query/evidence specific. In this paper, we extract grounded-

explanations, i.e., explanations defined w.r.t specific inference queries and observed evidence. We extract these explanations from importance weights defined over the MLN formulas that encode the contribution of formulas towards the final inference results. We validate our approach in real world problems related to analyzing reviews from Yelp, and show through user-studies that our explanations are richer than state-of-the-art non-relational explainers such as LIME.

Neural Control Variates for Monte Carlo Variance Reduction

R. Wan (Peking University), M. Zhong (University of Lincoln), H. Xiong (Baidu Inc.), Z. Zhu (Peking University, Beijing Institute of Big Data Research)

In statistics and machine learning, approximation of an intractable integration is often achieved by using the unbiased Monte Carlo estimator, but the variances of the estimation are generally high in many applications. Control variates approaches are well-known to reduce the variance of the estimation. These control variates are typically constructed by employing predefined parametric functions or polynomials, determined by using those samples drawn from the relevant distributions. Instead, we propose to construct those control variates by learning neural networks to handle the cases when test functions are complex. In many applications, obtaining a large number of samples for Monte Carlo estimation is expensive, the adoption of the original loss function may result in severe overfitting when training a neural network. This issue was not reported in those literature on control variates with neural networks. We thus further introduce a constrained control variates with neural networks to alleviate the overfitting issue. We apply the proposed control variates to both toy and real data problems, including a synthetic data problem, Bayesian model evidence evaluation and Bayesian neural networks. Experimental results demonstrate that our method can achieve significant variance reduction compared to other methods.

Grouped Gaussian Processes for Solar Power Prediction

Astrid Dahl, Edwin V. Bonilla

We consider multi-task regression models where the observations are assumed to be a linear combination of several latent node functions and weight functions, which are both drawn from Gaussian process priors. Driven by the problem of developing scalable methods for forecasting distributed solar and other renewable power generation, we propose coupled priors over groups of (node or weight) processes to exploit spatial dependence between functions. We estimate forecast models for solar power at multiple distributed sites and ground wind speed at multiple proximate weather stations. Our results show that our approach maintains or improves point-prediction accuracy relative to competing solar benchmarks and improves over wind forecast benchmark models on all measures. Our approach consistently dominates the equivalent model without coupled priors, achieving faster gains in forecast accuracy. At the same time our approach provides better quantification of predictive uncertainties.

Bayesian Generalized Horseshoe Estimation of Generalized Linear Models

Reproducible Research

D. F. Schmidt (Monash University; University of Melbourne), E. Makalic (University of Melbourne)

Bayesian global-local shrinkage estimation with the generalized horseshoe prior represents the state of the art for Gaussian regression models. The extension to non-Gaussian data, such as binary or Student-*t* regression, is usually done by exploiting a scale-mixture-of-normals approach. However, many standard distributions, such as the gamma and the Poisson, do not admit such a representation. We contribute two extensions to global-local shrinkage methodology. The first is an adaption of recent

auxiliary gradient based-sampling schemes to the global-local shrinkage framework, which yields simple algorithms for sampling from generalized linear models. We also introduce two new samplers for the hyperparameters in the generalized horseshoe model, one based on an inverse-gamma mixture of inverse-gamma distributions, and the second a rejection sampler. Results show that these new samplers are highly competitive with the no U-turn sampler for small numbers of predictors, and potentially perform better for larger numbers of predictors. Results for hyperparameter sampling show our new inverse-gamma inverse-gamma based sampling scheme outperforms the standard sampler based on a gamma mixture of gamma distributions.

Stochastic Gradient Hamiltonian Monte Carlo with variance reduction for Bayesian inference

Zhize Li, Tianyi Zhang, Shuyu Cheng, Jun Zhu, Jian Li

Gradient-based Monte Carlo sampling algorithms, like Langevin dynamics and Hamiltonian Monte Carlo, are important methods for Bayesian inference. In large-scale settings, full-gradients are not affordable and thus stochastic gradients evaluated on mini-batches are used as a replacement. In order to reduce the high variance of noisy stochastic gradients, Dubey et al. (in: Advances in neural information processing systems, pp 1154–1162, 2016) applied the standard variance reduction technique on stochastic gradient Langevin dynamics and obtained both theoretical and experimental improvements. In this paper, we apply the variance reduction tricks on Hamiltonian Monte Carlo and achieve better theoretical convergence results compared with the variance-reduced Langevin dynamics. Moreover, we apply the symmetric splitting scheme in our variance-reduced Hamiltonian Monte Carlo algorithms to further improve the theoretical results. The experimental results are also consistent with the theoretical results. As our experiment shows, variance-reduced Hamiltonian Monte Carlo demonstrates better performance than variance-reduced Langevin dynamics in Bayesian regression and classification tasks on real-world datasets.

Deep Learning 3, Thursday, 16:20 – 18:00

Research Track

Session Chair: Romain Tavenard

Sobolev Training with Approximated Derivatives for Black-Box Function Regression with Neural Networks

Reproducible Research

M. Kissel, K. Diepold (Technical University of Munich)

With Sobolev Training, neural networks are trained to fit target output values as well as target derivatives with respect to the inputs. This leads to better generalization and fewer required training examples for certain problems. In this paper, we present a training pipeline that enables Sobolev Training for regression problems where target derivatives are not directly available. Thus, we propose to use a least-squares estimate of the target derivatives based on function values of neighboring training samples. We show for a variety of black-box function regression tasks that our training pipeline achieves smaller test errors compared to the traditional training method. Since our method has no additional requirements on the data collection process, it has great potential to improve the results for various regression tasks.

Hyper-Parameter-Free Generative Modelling with Deep Boltzmann Trees
N. Piatkowski (TU Dortmund)

Deep neural networks achieve state-of-the-art results in various classification and synthetic data generation tasks. However, only little is known about why depth improves a model. We investigate the structure of stochastic deep neural works, also known as Deep Boltzmann Machines, to shed some light on this issue. While the best-known results postulate an exponential dependence between the number of visible units and the depth of the model, we show that the required depth is upper bounded by the longest path in the underlying junction tree, which is at most linear in the number of visible units. Moreover, we show that the conditional independence structure of any categorical Deep Boltzmann Machine contains a sub-tree that allows the consistent estimation of the full joint probability mass function of all visible units. We connect our results to I_1-regularized maximum-likelihood estimation and Chow-Liu trees. Based on our theoretical findings, we present a new tractable version of Deep Boltzmann Machines, namely the Deep Boltzmann Tree (DBT). We provide a hyper-parameter-free algorithm for learning the DBT from data, and propose a new initialization method to enforce convergence to good solutions. Our findings provide some theoretical evidence for why a deep model might be beneficial. Experimental results on benchmark data show, that the DBT is a theoretical sound alternative to likelihood-free generative models.

LYRICS: a General Interface Layer to Integrate Logic Inference and Deep Learning
G. Marra (University of Florence; University of Siena), F. Giannini (University of Siena), M. Diligenti (University of Siena), M. Gori (University of Siena)

In spite of the amazing results obtained by deep learning in many applications, a real intelligent behavior of an agent acting in a complex environment is likely to require some kind of higher-level symbolic inference. Therefore, there is a clear need for the definition of a general and tight integration between low-level tasks, processing sensorial data that can be effectively elaborated using deep learning techniques, and the logic reasoning that allows humans to take decisions in complex environments. This paper presents LYRICS, a generic interface layer for AI, which is implemented in TensorFlow (TF). LYRICS provides an input language that allows to define arbitrary First Order Logic (FOL) background knowledge. The predicates and functions of the FOL knowledge can be bound to any TF computational graph, and the formulas are converted into a set of real-valued constraints, which participate to the overall optimization problem. This allows to learn the weights of the learners, under the constraints imposed by the prior knowledge. The framework is extremely general as it imposes no restrictions in terms of which models or knowledge can be integrated. In this paper, we show the generality of the approach showing some use cases of the presented language, including model checking, supervised learning and collective classification.

Quantile Layers: Statistical Aggregation in Deep Neural Networks for Eye Movement Biometrics
Reproducible Research
A. Abdelwahab, N. Landwehr (Leibniz Institute of Agricultural Engineering and Bioeconomy e.V. (ATB), Potsdam)

Human eye gaze patterns are highly individually characteristic. Gaze patterns observed during the routine access of a user to a device or document can therefore be used to identify subjects unobtrusively, that is, without the need to perform an explicit verification such as entering a password. Existing approaches to biometric identification from gaze patterns segment raw gaze data into short, local patterns called saccades and fixations. Subjects are then identified by characterizing the distribution of these patterns or deriving hand-crafted features for them. In this paper, we follow

a different approach by training deep neural networks directly on the raw gaze data. As the distribution of short, local patterns has been shown to be particularly informative for distinguishing subjects, we introduce a parameterized and end-to-end learnable statistical aggregation layer called the quantile layer that enables the network to explicitly fit the distribution of filter activations in preceding layers. We empirically show that deep neural networks with quantile layers outperform existing probabilistic and feature-based methods for identifying subjects based on eye movements by a large margin.

LSALSA: Accelerated Source Separation via Learned Sparse Coding
Benjamin Cowen, Apoorva Nandini Saridena, Anna Choromanska

We propose an efficient algorithm for the generalized sparse coding (SC) inference problem. The proposed framework applies to both the single dictionary setting, where each data point is represented as a sparse combination of the columns of one dictionary matrix, as well as the multiple dictionary setting as given in morphological component analysis (MCA), where the goal is to separate a signal into additive parts such that each part has distinct sparse representation within an appropriately chosen corresponding dictionary. Both the SC task and its generalization via MCA have been cast as l_1 -regularized optimization problems of minimizing quadratic reconstruction error. In an effort to accelerate traditional acquisition of sparse codes, we propose a deep learning architecture that constitutes a trainable time-unfolded version of the split augmented lagrangian shrinkage algorithm (SALSA), a special case of the alternating direction method of multipliers (ADMM). We empirically validate both variants of the algorithm, that we refer to as learned-SALSA (LSALSA), on image vision tasks and demonstrate that at inference our networks achieve vast improvements in terms of the running time and the quality of estimated sparse codes on both classic SC and MCA problems over more common baselines. We also demonstrate the visual advantage of our technique on the task of source separation. Finally, we present a theoretical framework for analyzing LSALSA network: we show that the proposed approach exactly implements a truncated ADMM applied to a new, learned cost function with curvature modified by one of the learned parameterized matrices. We extend a very recent stochastic alternating optimization analysis framework to show that a gradient descent step along this learned loss landscape is equivalent to a modified gradient descent step along the original loss landscape. In this framework, the acceleration achieved by LSALSA could potentially be explained by the network's ability to learn a correction to the gradient direction of steeper descent.

Rich Data, Thursday, 16:20 – 18:00

ADS Track

Session Chair: Luis Galarraga

LSTM encoder-predictor for short-term train load forecasting
K. Pasini (Université Paris-Est, IRT SystemX), M. Khouadjia (Université Paris-Est), A. Samé (IRT SystemX), F. Ganansia (SNCF- Innovation & Recherche), L. Oukhellou (IRT SystemX)

The increase in the amount of data collected in the transport domain can greatly benefit mobility studies and help to create high value-added mobility services for passengers as well as regulation tools for operators. The research detailed in this paper is related to the development of an advanced machine learning approach with the aim of forecasting the passenger load of trains in public transport. Predicting the crowding level on public transport can indeed be useful for enriching the information available to passengers to enable them to better plan their daily trips. Moreover, operators will increasingly need to assess and predict network passenger load to improve train regulation processes and service quality levels. The main issues to address in this forecasting task are the variability in the

train load series induced by the train schedule and the influence of several contextual factors, such as calendar information. We propose a neural network LSTM encoder-predictor combined with a contextual representation learning to address this problem. Experiments are conducted on a real dataset provided by the French railway company SNCF and collected over a period of one and a half years. The prediction performance provided by the proposed model are compared to those given by historical models and by traditional machine learning models. The obtained results have demonstrated the potential of the proposed LSTM encoder-predictor to address both one-step-ahead and multi-step forecasting and to outperform other models by maintaining robustness in the quality of the forecasts throughout the time horizon.

Player Vectors: Characterizing Soccer Players' Playing Style from Match Event Streams
T. Decroos, J. Davis (KU Leuven)

Transfer fees for soccer players are at an all-time high. To make the most of their budget, soccer clubs need to understand the type of players they have and the type of players that are on the market. Current insights in the playing style of players are mostly based on the opinions of human soccer experts such as trainers and scouts. Unfortunately, their opinions are inherently subjective and thus prone to faults. In this paper, we characterize the playing style of a player in a more rigorous, objective and data-driven manner. We characterize the playing style of a player using a so-called 'player vector' that can be interpreted both by human experts and machine learning systems. We demonstrate the validity of our approach by retrieving player identities from anonymized event stream data and present a number of use cases related to scouting and monitoring player development in top European competitions.

Compact Representation of a Multi-dimensional Combustion Manifold Using Deep Neural Networks
S. Bhalla, M. Yao, J.-P. Hickey, M. Crowley (University of Waterloo)

The computational challenges in turbulent combustion simulations stem from the physical complexities and multi-scale nature of the problem which make it intractable to compute scale-resolving simulations. For most engineering applications, the large-scale separation between the flame (typically sub-millimeter scale) and the characteristic turbulent flow (typically centimeter or meter scale) allows us to evoke simplifying assumptions -such as done for the flamelet model- to pre-compute all the chemical reactions and map them to a low-order manifold. The resulting manifold is then tabulated and looked-up at run-time. As the physical complexity of combustion simulations increases (including radiation, soot formation, pressure variations etc.) the dimensionality of the resulting manifold grows which impedes an efficient tabulation and look-up. In this paper we present a novel approach to model the multi-dimensional combustion manifold. We approximate the combustion manifold using a neural network function approximator and use it to predict the temperature and composition of the reaction. We present a novel training procedure which is developed to generate a smooth output curve for temperature over the course of a reaction. We then evaluate our work against the current approach of tabulation with linear interpolation in combustion simulations. We also provide an ablation study of our training procedure in the context of over-fitting in our model. The combustion dataset used for the modeling of combustion of H_2 and O_2 in this work is released alongside this paper.

A Semi-Supervised and Online Learning Approach for Non-Intrusive Load Monitoring
H. Salem (Institut Mines-Télécom Lille Douai; Manouba University), M. Sayed-Mouchaweh (Institut Mines-Télécom Lille Douai)

Non-Intrusive Load Monitoring (NILM) approaches aim at identifying the consumption of a single appliance from the total load provided by smart meters. Several research works based on Hidden Markov Models (HMM) were developed for NILM where training is performed offline. However, these approaches suffer from different issues: First, they fail to generalize to unseen appliances with different configurations or brands than the ones used for training. Second, obtaining data about all active states of each appliance requires long time, which is impractical for residents. Third, offline training requires storage of huge amount of data, yielding to share resident consumption data with external servers and causing privacy issues. Therefore, in this paper, a new approach is proposed in order to tackle these issues. This approach is based on the use of an HMM conditioned on discriminant contextual features (e.g., time of usage, duration of usage). The conditional HMM (CHMM) is trained online using data related to a single appliance consumption extracted from aggregated load in order to adapt its parameters to the appliance specificity's (e.g., brand, configuration, etc.). Experiments are performed using real data from publicly available data sets and comparative evaluation are performed on a publicly available NILM framework.

Characterization and Early Detection of Evergreen News Articles
Y. Liao (Pennsylvania State University), S. Wang (The Washington Post), E.-H. S. Han (Marriott International), J. Lee (Sungkyunkwan University), D. Lee (Pennsylvania State University)

Although the majority of news articles are only viewed for days or weeks, there are a small fraction of news articles that are read across years, thus named as evergreen news articles. Because evergreen articles maintain a timeless quality and are consistently of interests to the public, understanding their characteristics better has huge implications for news outlets and platforms yet there are few studies that have explicitly investigated on evergreen articles. Addressing this gap, in this paper, we first propose a flexible parameterized definition of evergreen articles to capture their long-term high traffic patterns. Using a real dataset from the Washington Post, then, we unearth several distinctive characteristics of evergreen articles and build an early prediction model with encouraging results. Although less than 1% of news articles were identified as evergreen, our model achieves 0.961 in ROC AUC and 0.172 in PR AUC in 10-fold cross validation.

Multi-Label Learning, Thursday, 16:20 – 18:00

Research Track

Session Chair: Hendrik Blockeel

Data scarcity, robustness and extreme multi-label classification
Rohit Babbar, Bernhard Schölkopf

The goal in extreme multi-label classification (XMC) is to learn a classifier which can assign a small subset of relevant labels to an instance from an extremely large set of target labels. The distribution of training instances among labels in XMC exhibits a long tail, implying that a large fraction of labels have a very small number of positive training instances. Detecting tail-labels, which represent diversity of the label space and account for a large fraction (up to 80%) of all the labels, has been a significant research challenge in XMC. In this work, we pose the tail-label detection task in XMC as robust learning in the presence of worst-case perturbations. This viewpoint is motivated by a key observation that

there is a significant change in the distribution of the feature composition of instances of these labels from the training set to test set. For shallow classifiers, our robustness perspective to XMC naturally motivates the well-known l_1-regularized classification. Contrary to the popular belief that Hamming loss is unsuitable for tail-labels detection in XMC, we show that minimizing (convex upper bound on) Hamming loss with appropriate regularization surpasses many state-of-the-art methods. Furthermore, we also highlight the sub-optimality of the co-ordinate descent based solver in the LibLinear package, which, given its ubiquity, is interesting in its own right. We also investigate the spectral properties of label graphs for providing novel insights towards understanding the conditions governing the performance of Hamming loss based one-vs-rest scheme vis-à-vis label embedding methods.

Neural Message Passing for Multi-Label Classification

Reproducible Research

J. Lanchantin, A. Sekhon, Y. Qi (University of Virginia)

Multi-label classification (MLC) is the task of assigning a set of target labels for a given sample. Modeling the combinatorial label interactions in MLC has been a long-haul challenge. We propose Label Message Passing (LaMP) Neural Networks to efficiently model the joint prediction of multiple labels. LaMP treats labels as nodes on a label-interaction graph and computes the hidden representation of each label node conditioned on the input using attention-based neural message passing. Attention enables LaMP to assign different importances to neighbor nodes per label, learning how labels interact (implicitly). The proposed models are simple, accurate, interpretable, structure-agnostic, and applicable for predicting dense labels since LaMP is incredibly parallelizable. We validate the benefits of LaMP on seven real-world MLC datasets, covering a broad spectrum of input/output types and outperforming the state-of-the-art results. Notably, LaMP enables intuitive interpretation of how classifying each label depends on the elements of a sample and at the same time rely on its interaction with other labels.

Synthetic Oversampling of Multi-Label Data based on Local Label Distribution

Reproducible Research

B. Liu, G. Tsoumakas (Aristotle University of Thessaloniki)

Class-imbalance is an inherent characteristic of multi-label data which affects the prediction accuracy of most multi-label learning methods. One efficient strategy to deal with this problem is to employ resampling techniques before training the classifier. Existing multi-label sampling methods alleviate the (global) imbalance of multi-label datasets. However, performance degradation is mainly due to rare sub-concepts and overlapping of classes that could be analysed by looking at the local characteristics of the minority examples, rather than the imbalance of the whole dataset. We propose a new method for synthetic oversampling of multi-label data that focuses on local label distribution to generate more diverse and better labeled instances. Experimental results on 13 multi-label datasets demonstrate the effectiveness of the proposed approach in a variety of evaluation measures, particularly in the case of an ensemble of classifiers trained on repeated samples of the original data.

PP-PLL: Probability Propagation for Partial Label Learning

Reproducible Research

K. Sun, Z. Min, J. Wang (Chongqing University of Posts and Telecommunications)

Partial label learning (PLL) is a weakly supervised learning framework which learns from the data where each example is associated with a set of candidate labels, among which only one is correct. Most existing approaches are based on the disambiguation strategy, which either identifies the valid

label iteratively or treats each candidate label equally based on the averaging strategy. In both cases, the disambiguation strategy shares a common shortcoming that the ground-truth label may be overwhelmed by the false positive candidate labels, especially when the number of candidate labels becomes large. In this paper, a probability propagation method for partial label learning (PP-PLL) is proposed. Specifically, based on the manifold assumption, a biconvex regular function is proposed to model the linear mapping relationships between input features and output true labels. In PP-PLL, the topological relations among training samples are used as additional information to strengthen the mutual exclusiveness among candidate labels, which helps to prevent the ground-truth label from being overwhelmed by a large number of candidate labels. Experimental studies on both artificial and real-world data sets demonstrate that the proposed PP-PLL method can achieve superior or comparable performance against the state-of-the-art methods.

Dynamic Principal Projection for Cost-Sensitive Online Multi-Label Classification
Hong-Min Chu, Kuan-Hao Huang, Hsuan-Tien Lin

We study multi-label classification (MLC) with three important real-world issues: online updating, label space dimension reduction (LSDR), and cost-sensitivity. Current MLC algorithms have not been designed to address these three issues simultaneously. In this paper, we propose a novel algorithm, cost-sensitive dynamic principal projection (CS-DPP) that resolves all three issues. The foundation of CS-DPP is an online LSDR framework derived from a leading LSDR algorithm. In particular, CS-DPP is equipped with an efficient online dimension reducer motivated by matrix stochastic gradient, and establishes its theoretical backbone when coupled with a carefully-designed online regression learner. In addition, CS-DPP embeds the cost information into label weights to achieve cost-sensitivity along with theoretical guarantees. Experimental results verify that CS-DPP achieves better practical performance than current MLC algorithms across different evaluation criteria, and demonstrate the importance of resolving the three issues simultaneously.

Dimensionality Reduction & Feature Selection, Thursday, 16:20 – 18:00

Research Track

Session Chair: TBA

Interpretable Discriminative Dimensionality Reduction and Feature Selection on the Manifold

Reproducible Research

B. Hosseini, B. Hammer (Bielefeld University)

Dimensionality reduction (DR) on the manifold includes effective methods which project the data from an implicit relational space onto a vectorial space. Regardless of the achievements in this area, these algorithms suffer from the lack of interpretation of the projection dimensions. Therefore, it is often difficult to explain the physical meaning behind the embedding dimensions. In this research, we propose the interpretable kernel DR algorithm (I-KDR) as a new algorithm which maps the data from the feature space to a lower dimensional space where the classes are more condensed with less overlapping. Besides, the algorithm creates the dimensions upon local contributions of the data samples, which makes it easier to interpret them by class labels. Additionally, we efficiently fuse the DR with feature selection task to select the most relevant features of the original space to the discriminative objective. Based on the empirical evidence, I-KDR provides better interpretations for embedding dimensions as well as higher discriminative performance in the embedded space compared to the state-of-the-art and popular DR algorithms.

On the Stability of Feature Selection in the Presence of Feature Correlations

Reproducible Research

K. Sechidis (University of Manchester), K. Papangelou (University of Manchester), S. Nogueira (Criteo, Paris), J. Weatherall (Advanced Analytics Centre, Global Medicines Development, AstraZeneca, Cambridge), G. Brown (University of Manchester)

Feature selection is central to modern data science. The 'stability' of a feature selection algorithm refers to the sensitivity of its choices to small changes in training data. This is, in effect, the robustness of the chosen features. This paper considers the estimation of stability when we expect strong pairwise correlations, otherwise known as feature redundancy. We demonstrate that existing measures are inappropriate here, as they systematically underestimate the true stability, giving an overly pessimistic view of a feature set. We propose a new statistical measure which overcomes this issue, and generalises previous work.

Joint Multi-Source Reduction

L. Zhang (Chinese Academy of Sciences), S. Wang (Chinese Academy of Sciences), X. Jin (National Computer Network Emergency Response Technical Team/Coordination Center of China), S. Jia (Chinese Academy of Sciences)

The redundant sources problem in multi-source learning always exists in various real-world applications such as multimedia analysis, information retrieval, and medical diagnosis, in which the heterogeneous representations from different sources always have three-way redundancies. More seriously, the redundancies will cost a lot of storage space, cause high computational time, and degrade the performance of learner. This paper is an attempt to jointly reduce redundant sources. Specifically, a novel Heterogeneous Manifold Smoothness Learning (HMSL) model is proposed to linearly map multi-source data to a low-dimensional feature-isomorphic space, in which the information-correlated representations are close along manifold while the semantic-complementary instances are close in Euclidean distance. Furthermore, to eliminate three-way redundancies, we present a new Correlation-based Multi-source Redundancy Reduction (CMRR) method with 2,1-norm equation and generalized elementary transformation constraints to reduce redundant sources in the learned feature-isomorphic space. Comprehensive empirical investigations are presented that confirm the promise of our proposed framework.

Efficient Feature Selection Using Shrinkage Estimators

Konstantinos Sechidis, Laura Azzimonti, Adam Pocock, Giorgio Corani, James Weatherall, Gavin Brown

Information theoretic feature selection methods quantify the importance of each feature by estimating mutual information terms to capture: the relevancy, the redundancy and the complementarity. These terms are commonly estimated by maximum likelihood, while an under-explored area of research is how to use shrinkage methods instead. Our work suggests a novel shrinkage method for data-efficient estimation of information theoretic terms. The small sample behaviour makes it particularly suitable for estimation of discrete distributions with large number of categories (bins). Using our novel estimators, we derive a framework for generating feature selection criteria that capture any high-order feature interaction for redundancy and complementarity. We perform a thorough empirical study across datasets from diverse sources and using various evaluation measures. Our first finding is that our shrinkage based methods achieve better results, while they keep the same computational cost as the simple maximum likelihood based methods. Furthermore, under our framework we derive efficient novel high-order criteria that outperform state-of-the-art methods in various tasks.

Journal Track Papers

The journal track papers have been, or will be published in two Springer journals, *Data Mining and Knowledge Discovery* and *Machine Learning*.

Data Mining and Knowledge Discovery papers

Noise-free Latent Block Model for High Dimensional Data
Charlotte Laclau, Vincent Brault

Robust active attacks on social graphs
Sjouke Mauw, Yunior Ramírez-Cruz, Rolando Trujillo-Rasua

Classification with Label Noise: A Markov Chain Sampling Framework
Zijin Zhao, Lingyang Chu, Dacheng Tao, Jian Pei

Model-Free Inference of Diffusion Networks using RKHS embeddings
Shoubo Hu, Bogdan Cautis, Zhitang Chen, Laiwan Chan, Yanhui Geng, Xiuqiang He

Counts-of-Counts Similarity for Prediction and Search in Relational Data
Manfred Jaeger, Marco Lippi, Giovanni Pellegrini, Andrea Passerini

Finding lasting dense graphs
Konstantinos Semertzidis, Evaggelia Pitoura, Evimaria Terzi, Panayiotis Tsaparas

Deeply Supervised Model for Click-Through Rate Prediction in Sponsored Search
Jelena Gligorijevic, Djordje Gligorijevic, Ivan Stojkovic, Xiao Bai, Amit Goyal, Zoran Obradovic

Temporal Density Extrapolation using a Dynamic Basis Approach
G. Krempl, D. Lang, V. Hofer

Mining Skypatterns in Fuzzy Tensors
Nicolas Nadisic, Aurélien Coussat, Loïc Cerf

Multi-location Visibility Query Processing Using Portion-based Transactional Modeling and Pattern Mining
Lakshmi Gangumalla, P. Krishna Reddy, Anirban Mondal

Machine Learning papers

Aggregating Algorithm for Prediction of Packs
Dmitry Adamskiy, Anthony Bellotti, Raisa Dzhamtyrova, Yuri Kalnishkan

Dynamic Principal Projection for Cost-Sensitive Online Multi-Label Classification
Hong-Min Chu, Kuan-Hao Huang, Hsuan-Tien Lin

Stochastic Gradient Hamiltonian Monte Carlo with variance reduction for Bayesian inference
Zhize Li, Tianyi Zhang, Shuyu Cheng, Jun Zhu, Jian Li

Rankboost+: An Improvement to Rankboost
Harold Connamacher

Efficient Feature Selection Using Shrinkage Estimators
Konstantinos Sechidis, Laura Azzimonti, Adam Pocock, Giorgio Corani, James Weatherall, Gavin Brown

TD-Regularized Actor-Critic Methods
Simone Parisi, Voot Tangkaratt, Jan Peters, Mohammad Emtiyaz Khan

Joint Detection of Malicious Domains and Infected Clients
Paul Prasse, René Knaebel, Lukáš Machlica, Tomáš Pevný, Tobias Scheffer

Efficient learning with robust gradient descent
Matthew J. Holland, Kazushi Ikeda

Nuclear Discrepancy for Single-Shot Batch Active Learning
Tom J. Viering, Jesse H. Krijthe, Marco Loog

On PAC-Bayesian Bounds for Random Forests
Stephan S. Lorenzen, Christian Igel, Yevgeny Seldin

LSALSA: Accelerated Source Separation via Learned Sparse Coding
Benjamin Cowen, Apoorva Nandini Saridena, Anna Choromanska

Data scarcity, robustness and extreme multi-label classification
Rohit Babbar, Bernhard Schölkopf

CaDET: Interpretable Parametric Conditional Density Estimation with Decision Trees and Forests
Cyrus Cousins, Matteo Riondato

Improving latent variable descriptiveness by modelling rather than ad-hoc factors
Alex Mansbridge, Roberto Fierimonte, Ilya Feige, David Barber

Deep Collective Matrix Factorization for Augmented Multi-View Learning
Ragunathan Mariappan, Vaibhav Rajan

A Flexible Probabilistic Framework for Large-Margin Mixture of Experts
Archit Sharma, Siddhartha Saxena, Piyush Rai

Temporal Pattern Attention for Multivariate Time Series Forecasting
Shun-Yao Shih, Fan-Keng Sun, Hung-yi Lee

Grouped Gaussian Processes for Solar Power Prediction
Astrid Dahl, Edwin V. Bonilla

On the analysis of adaptability in multi-source domain adaptation
Ievgen Redko, Amaury Habrard, Marc Sebban

The Teaching Size: Computable Teachers and Learners for Universal Languages
Jan Arne Telle, José Hernández-Orallo, Cèsar Ferri

Distribution-Free Uncertainty Quantification for Kernel Methods by Gradient Perturbations
Balázs Cs. Csáji, Krisztián B. Kis

Compatible Natural Gradient Policy Search
Joni Pajarinen, Hong Linh Thai, Riad Akrour, Jan Peters, Gerhard Neumann

Awards

Test of Time Award 2019

The Test of Time Award honors the paper from ECML PKDD 2009 with the highest impact in the field. The award is sponsored by European Research Center for Information Systems (ERCIS).

Classifier Chains for Multi-label Classification
Jesse Read, Bernhard Pfahringer, Geoff Holmes, Eibe Frank (The University of Waikato)

The widely known binary relevance method for multi-label classification, which considers each label as an independent binary problem, has been sidelined in the literature due to the perceived inadequacy of its label-independence assumption. Instead, most current methods invest considerable complexity to model interdependencies between labels. This paper shows that binary relevance-based methods have much to offer, especially in terms of scalability to large datasets. We exemplify this with a novel chaining method that can model label correlations while maintaining acceptable computational complexity. Empirical evaluation over a broad range of multi-label datasets with a variety of evaluation metrics demonstrates the competitiveness of our chaining method against related and state-of-the-art methods, both in terms of predictive performance and time complexity.

Praise by the Award Chairs: The well-known transformation of a multi-label problem into a binary problem for each label has the disadvantage that the labels might correlate. Passing label correlation information along a chain of classifiers counteracts this deficiency while maintaining the ease of implementation and understanding. The paper is extremely well recognised by 537 citations of the conference paper and 981 of the subsequent journal version.

Best Student Paper Award (Data Mining)

The Best Student Paper Award (Data Mining) is selected by the Awards Chairs, after nomination from the Program Chairs, from the submitted Data Mining papers that involve at least one student author. The award is sponsored by the Springer *Data Mining and Knowledge Discovery* journal.

To be announced.

Best Student Paper Award (Machine Learning)

The Best Student Paper Award (Machine Learning) is selected by the Awards Chairs, after nomination from the Program Chairs, from the submitted Machine Learning papers that involve at least one student author. The award is sponsored by the Springer *Machine Learning* journal.

To be announced.

Workshops Monday

MML 2019: 12th International Workshop on Machine Learning and Music, Room 2.012

musml2019.weebly.com

Machine learning and artificial intelligence have permeated nearly every area of music informatics, driven by a profusion of recordings available in digital audio formats, steady improvements to the accessibility and quality of symbolic corpora, availability of powerful algorithms, and theoretical advances in machine learning and data mining. As the complexity of the problems investigated by researchers on machine learning and music increases, there is a need to develop new algorithms and methods. The 12th International Workshop on Machine Learning and Music (MML 2019) aims to promote fruitful multidisciplinary collaboration among researchers who are using machine learning techniques in musical applications, by providing the opportunity to discuss ongoing work in the area. MML 2019 welcomes papers in all applications on music and machine learning.

Organisers: Rafael Ramirez (Universitat Pompeu Fabra, Spain), Darrell Conklin (University of the Basque Country, Spain), José Manuel Iñesta (University of Alicante, Spain)

Workshop on Multiple-aspect analysis of semantic trajectories (MASTER2019), Room 2.003

www.master-project-h2020.eu/workshop-master-2019

An ever-increasing number of diverse, real-life applications, ranging from mobile to social media apps and surveillance systems, produce massive amounts of spatio-temporal data representing trajectories of moving objects. The fusion of those trajectories, commonly represented by timestamped location sequence data (e.g. check-ins and GPS traces), with generally available and semantic-rich data resources can result in an enriched set of more comprehensive and semantically significant objects. The analysis of these sets, referred to as "semantic trajectories", can unveil solutions to traditional problems and unlock the challenges for the advent of novel applications and application domains, such as transportation, security, health, environment and even policy modeling. The MASTER 2019 workshop anticipates to solicit high-quality, scientifically sound and innovative contributions on the topic with the intention to promote and follow the developments in the domain.

Organisers: Chiara Renso (CNR, Pisa, Italy), Stan Matwin (Dalhousie University, Halifax, Canada), Konstantinos Tserpes (HUA, Athens, Greece)

MIDAS - The 4th Workshop on MIning DAta for financial applicationS, Room 2.011

midas.portici.enea.it

Like the famous King Midas, popularly remembered in Greek mythology for his ability to turn everything he touched with his hand into gold, we believe that the wealth of data generated by modern technologies, with widespread presence of computers, users and media connected by Internet, is a goldmine for tackling a variety of problems in the financial domain. The MIDAS workshop is aimed at discussing challenges, potentialities, and applications of leveraging data-mining tasks to tackle problems in the financial domain. The workshop provides a premier forum for sharing findings, knowledge, insights, experience and lessons learned from mining data generated in various application domains.

Organisers: Valerio Bitetta (Unicredit, R&D Dept.), Ilaria Bordino (Unicredit, R&D Dept.), Andrea Ferretti (Unicredit, R&D Dept.), Francesco Gullo (Unicredit, R&D Dept.), Stefano Pascolutti (Unicredit, R&D Dept.), Giovanni Ponti (ENEA, Portici Research Center)

Second International Workshop on Knowledge Discovery and User Modeling for Smart Cities, Room 2.010

umcit-2019.isistan.unicen.edu.ar

User modelling and personalization are commonly used in multiple tasks, in which users are characterized only based on explicit information about their knowledge, behaviour, social relations or preferences, aiming at adapting generic systems to the particularities of each user. The ubiquitousness of social networking sites, and mobile and smart-devices offer new information sources, opportunities and challenges for changing the personalization paradigm. The analysis of these new data source offers new research opportunities across a wide variety of disciplines, including media and communication studies, linguistics, sociology, health, psychology, information and computer sciences, or education. This has important implications in the context of inclusive eGovernment and Smart Cities, which could leverage on the user's models to design and tailor services according to the characteristics and needs of each particular citizen. In this context, this workshop targets academy and industrial practitioners leveraging on diverse data mining and machine learning techniques, including content aggregation, content analysis, predictive modeling, deep learning and user embedding for modeling user behavior and analyzing urban data.

Organisers: Rabeah Alzaidy (King Abdullah University of Science and Technology, Thuwal, Saudi Arabia), Marcelo G. Armentano (ISISTAN, CONICET-UNICEN. Buenos Aires, Argentina), Antonela Tommasel (ISISTAN, CONICET-UNICEN. Buenos Aires, Argentina), Ludovico Boratto (Eurecat, Barcelona, Spain), Clyde L. Giles (College of Information Sciences and Technology, Pennsylvania State University)

New Frontiers in Mining Complex Patterns, Room 2.002

www.di.uniba.it/~loglisci/NFMCP2019/index.html

Modern automatic systems are able to collect huge volumes of data often with a complex structure. The massive and complex data pose new challenges for current research in Knowledge Discovery and Data Mining. They require new methods for storing, managing and analyzing them by taking into account various complexity aspects: Complex structures (e.g. multi-relational, time series and sequences, networks, and trees) as input/output of the data mining process; Massive amounts of high dimensional data collections flooding as high-speed streams and requiring (near) real time processing

and model adaptation to concept drifts; New application scenarios involving security issues, interaction with other entities and real-time response to events triggered by sensors. The purpose of the workshop is to bring together researchers and practitioners of data mining and machine learning interested in analysis of complex and massive data sources such as blogs, event or log data, medical data, spatio-temporal data, social networks, mobility data, sensor data and streams.

Organisers: Michelangelo Ceci (University of Bari Aldo Moro, Bari, Italy), Corrado Loglisci (University of Bari Aldo Moro, Bari, Italy), Giuseppe Manco (ICAR-CNR, Rende, Italy), Elio Masciari (ICAR-CNR, Rende, Italy), Zbigniew W. Ras (University of North Carolina, Charlotte, USA & Warsaw University of Technology, Poland)

New Trends in Representation Learning with Knowledge Graphs, Room 0.002

sites.google.com/view/kgrlfr-workshop/home

Knowledge Graphs are becoming the standard for storing, retrieving and querying structured data. In academia and in industry, they are increasingly used to provide background knowledge. Over the last years, several research contributions are made to show machine learning especially representation learning is successfully applied to knowledge graphs enabling inductive inference about facts with unknown truth values. In this workshop we intend to focus on the most exciting new developments in Knowledge Graph learning, bridging the gap to recent developments in different fields. Also, we want to bring together researchers from different disciplines but united by their adoption of earlier mentioned techniques from machine learning.

Organisers: Volker Tresp (Ludwig-Maximilians University and Siemens, Germany), Jens Lehmann (Bonn University and Fraunhofer IAIS, Germany), Aditya Mogadala (Saarland University, Germany), Achim Rettinger (Trier University, Germany), Afshin Sadeghi (Fraunhofer IAIS, Germany), Mehdi Ali (Bonn University and Fraunhofer IAIS, Germany)

Second International Workshop on Energy Efficient Scalable Data Mining and Machine Learning. Green Data Mining, Room 1.002 (Morning only!)

greendatamining.github.io

This workshop aims to bring together people from different areas and backgrounds in data mining and machine learning that have a common interest in scalability, edge computing, and energy efficiency. These fields include, but are not limited to: deep learning, Internet of Things (IoT), scalable machine learning, systems for machine learning, information retrieval systems, and stream mining. The goal is to provide a venue where researchers with heterogeneous backgrounds can debate and discuss challenges related to building energy efficient machine learning algorithms, systems, and hardware platforms. We accept original work, already completed, or in progress. Position papers are also welcomed.

Organisers: Eva Garcia-Martin, Albert Bifet, Crefeda Faviola Rodrigues, Heitor Murilo Gomes

Deep Continuous-Discrete Machine Learning (DeCoDeML 2019), Room 1.003 (Afternoon only!)

sites.google.com/view/decodeml-workshop-2019

Since the beginnings of machine learning – and indeed already hinted at in Alan Turing's groundbreaking 1950 paper "Computing machinery and intelligence" – two opposing approaches have been pursued: On the one hand, approaches that relate learning to knowledge and mostly use "discrete" formalisms of formal logic. On the other hand, approaches which, mostly motivated by biological models, investigate learning in artificial neural networks and predominantly use "continuous" methods from numerical optimization and statistics. The recent successes of deep learning can be attributed to the latter, the "continuous" approach, and are currently opening up new opportunities for computers to "perceive" the world and to act, with far reaching consequences for industry, science and society. The massive success in recognizing "continuous" patterns is the catalyst for a new enthusiasm for artificial intelligence methods. However, today's artificial neural networks are hardly suitable for learning and understanding "discrete" logical structures, and this is one of the major hurdles to further progress. Accordingly, one of the biggest open problems is to clarify the connection between these two learning approaches (logical-discrete, neural-continuous). In particular, the role and benefits of prior knowledge need to be reassessed and clarified. The role of formal logic in ensuring sound reasoning must be related to perception through deep networks. Further, the question of how to use prior knowledge to make the results of deep learning more stable, and to explain and justify them, is to be discussed. The extraction of symbolic knowledge from networks is a topic that needs to be reexamined against the background of the successes of deep learning. Finally, it is an open question if and how the principles responsible for the success of deep learning methods can be transferred to symbolic learning. The First Workshop on Deep Continuous-Discrete Machine Learning (DeCoDeML) at ECML PKDD 2019 intends to attract papers on how recent deep learning methods can be connected to discrete structures and symbolic knowledge, addressing all of the above questions.

Organisers: Kristian Kersting (Technical University Darmstadt), Stefan Kramer (Johannes Gutenberg University Mainz)

Decentralized Machine Learning at the Edge, Room 1.003 (Morning only!)

dmle.iais.fraunhofer.de

Many of today's parallel machine learning algorithms were developed for tightly coupled systems like computing clusters or clouds. However, the volumes of data generated from machine-to-machine interaction, by mobile phones or autonomous vehicles, surpass the amount of data that can be realistically centralized. Thus, traditional cloud computing approaches are rendered infeasible. To scale parallel machine learning to such volumes of data, computation needs to be pushed towards the edge, that is, towards the data generating devices. By learning models directly on the data sources---which often have computational power of their own, for example, mobile phones, smart sensors, and vehicles---network communication can be reduced by orders of magnitude. Moreover, it enables training a central model without centralizing privacy-sensitive data. This workshop is the second edition of the successful DMLE workshop at last year's ECMLPKDD. It aims to foster discussion, discovery, and dissemination of novel ideas and approaches for decentralized machine learning.

Organisers: Michael Kamp, Yamuna Krishnamurthy, Daniel Paurat

Applications of Topological Data Analysis, Room 1.002 (Afternoon only!)
sites.google.com/view/atda2019

The emergent area of Topological data analysis (TDA) aims to uncover hidden structure in a wide variety of data sets, combining methods from algebraic topology and other tools of pure mathematics to study the shape of data. Though the pure mathematical foundation of TDA is a major research topic on its own, TDA has been applied to a wide variety of real-world problems, among which image compression, cancer research, and shape or pattern recognition are only a few of the many examples. As TDA is generally not a well-known topic to the data mining and machine learning community, this workshop aims to address the flow of information between the different communities. By illustrating some of its recent and new applications, we will discuss the potential of TDA to active researchers within the fields of data science and machine learning. Furthermore, this workshop provides new and young TDA researchers a chance to present their work to a new community in an interesting and creative way, emphasizing the many possible applications of TDA in real-world data sets.

Organisers: Robin Vandaele, Tijl De Bie, John Harer

GEM: Graph Embedding and Mining, Room 0.001
gem-ecmlpkdd.github.io

Graphs enable the specification of relational structure between entities, adding significant additional flexibility in comparison to data consisting of unrelated data points. The ability to model and discover knowledge from such network data is therefore fast gaining in importance. This workshop aims to bring together researchers and practitioners in graph mining, modeling, and embedding. We solicit papers that advance the state of the art for particular applications or introduce new algorithms for machine learning, embedding, or pattern mining on graphs.

Organisers: Bo Kang (Ghent University), Remy Cazabet (Université de Lyon), Christine Largeron (Université Jean Monnet), Polo Chau (Georgia Tech), Jefrey Lijffijt (Ghent University), Tijl De Bie (Ghent University)

Interactive Adaptive Learning, Room 1.011
p.ies.uni-kassel.de/ial2019/

The tutorial and workshop aims at discussing techniques and approaches for optimising the whole learning process, including the interaction with human supervisors, processing systems, and includes adaptive, active, semi-supervised, and transfer learning techniques, and combinations thereof in interactive and adaptive machine learning systems. Our objective is to bridge the communities researching and developing these techniques and systems in machine learning and data mining.

Organisers: Georg Krempl (Utrecht University, Netherlands), Vincent Lemaire (Orange Labs, France), Daniel Kottke (University of Kassel, Germany), Andreas Holzinger (Medical University Graz, Austria), Adrian Calma (Darwin, USA)

IoT Stream for Data Driven Predictive Maintenance, Room 1.010

abifet.wixsite.com/iotstream2019

Maintenance is a critical issue in the industrial context for the prevention of high costs or injures. The emerging technologies of Industry 4.0 empowered data production and exchange which lead to new concepts and methodologies exploitation for maintenance. Intensive research effort in data driven Predictive Maintenance (PdM) has been producing encouraged outcomes. Therefore, the main objective of this workshop is to raise awareness of research trends and promote interdisciplinary discussion in this field. This workshop includes a tutorial.

Organisers: Rita Ribeiro (University of Porto), Albert Bifet (Telecom-ParisTech), João Gama (University of Porto), Anders Holst (RISE SICS), Sepideh Pashami (Halmstad University)

Machine Learning for Cybersecurity (MLCS), Room 2.010

mlcs.lasige.di.fc.ul.pt

The last decade has been a critical one regarding cybersecurity, with studies estimating the cost of cybercrime to be up to 0.8 percent of the global GDP. The capability to detect, analyze, and defend against threats in (near) real-time conditions is not possible without employing machine learning techniques and big data infrastructure. This gives rise to cyberthreat intelligence and analytic solutions, such as (informed) machine learning on big data and open-source intelligence, to perceive, reason, learn, and act against cyber adversary techniques and actions. Moreover, organisations' security analysts have to manage and protect systems and deal with the privacy and security of all personal and institutional data under their control. The aim of this workshop is to provide researchers with a forum to exchange and discuss scientific contributions, open challenges and recent achievements in machine learning and their role in the development of secure systems.

Organisers: Annalisa Appice, Battista Biggio, Donato Malerba, Fabio Roli, Ibéria Medeiros, Michael Kamp, Pedro Ferreira

BioASQ: Large-scale biomedical semantic indexing and question answering, Room 1.004

www.bioasq.org/workshop

BioASQ is a series of workshops and challenges (shared tasks) that reward highly precise biomedical information access and machine learning systems. The aim of BioASQ is to push the research frontier towards systems that use the diverse and voluminous information available online to respond directly to the information needs of biomedical scientists.

Organisers: George Paliouras (NCSR "Demokritos", Greece and University of Houston, USA), Anastasia Krithara (NCSR "Demokritos", Greece), Anastasios Nentidis (NCSR "Demokritos", Greece and Aristotle University of Thessaloniki, Greece)

6th Workshop on Sports Analytics: Machine Learning and Data Mining for Sports Analytics (MLSA), Room 2.011

dtai.cs.kuleuven.be/events/MLSA19

Sports Analytics has been a steadily growing and rapidly evolving area over the last decade that has attracted significant attention from professional clubs, researchers, and fans. This area has come to rely on techniques from machine learning and data mining. In terms of tasks, this field is very broad and it encompasses areas such as analyzing match strategy and tactics, valuing and rating players, designing training regimes, preventing injuries, predicting the outcomes of matches, and designing tournaments and schedules. The workshop solicits papers covering both predictive and descriptive Machine Learning, Data Mining, and related approaches to Sports Analytics settings. Adopting a broad definition of sports, the workshop is also open to submissions on electronic sports (i.e., e-sports) as well.

Organisers: Jesse Davis (KU Leuven), Ulf Brefeld (Leuphana University), Jan Van Haaren (SciSports), Albrecht Zimmermann (University of Caen)

4th workshop on Advanced Analytics and Learning on Temporal Data, Room 0.002

project.inria.fr/aaltd19

Temporal data are frequently encountered in a wide range of domains such as bio-informatics, medicine, finance and engineering, among many others. They are naturally present in applications covering language, motion and vision analysis, or more emerging ones as energy efficient building, smart cities, dynamic social media or sensor networks. Contrary to static data, temporal data are of complex nature, they are generally noisy, of high dimensionality, they may be nonstationary (i.e. first order statistics vary with time) and irregular (involving several time granularities), they may have several invariant domain-dependent factors as time delay, translation, scale or tendency effects. The aim of this workshop is to bring together researchers and experts in machine learning, data mining, pattern analysis and statistics to share their challenging issues and advance researches on temporal data analysis. Analysis and learning from temporal data cover a wide scope of tasks including learning metrics, learning representations, unsupervised feature extraction, clustering and classification.

Organisers: Anthony Bagnall (University of East Anglia, England), Alexis Bondu (Orange Labs, France), Pádraig Cunningham (University College Dublin, Ireland), Thomas Guyet (Agrocampus, IRISA France), Vincent Lemaire (Orange Labs, France), Simon Malinowski (Université de Rennes 1, IRISA, France), Romain Tavenard (Université de Rennes 2, COSTEL, France)

MACLEAN: MAChine Learning for EArth ObservatioN, Room 2.003

mdl4eo.irstea.fr/maclean-machine-learning-for-earth-observation

The huge amount of data currently produced by modern Earth Observation (EO) missions has raised up new challenges for the Remote Sensing communities. EO sensors are now able to offer (very) high spatial resolution images with revisit time frequencies never achieved before considering different kind of signals, e.g., multi-(hyper)spectral optical, radar, LiDAR and Digital Surface Models. In this context, modern machine learning techniques can play a crucial role to deal with such amount of heterogeneous, multi-scale and multi-modal data. Some examples of techniques that are gaining attention in this domain include deep learning, domain adaptation, semi-supervised approach, time series analysis and active learning. Even though the use of machine learning and the development of ad-hoc techniques are gaining increasing popularity in the EO domain, we can witness that a significant lack of interaction between domain experts and machine learning researchers still exists. The objective of this workshop is to supply an international forum where machine learning researchers and domain-experts can meet each other, in order to exchange, debate and draw short and long term research objectives around the exploitation and analysis of EO data via Machine Learning techniques. Among the workshop's objectives, we want to give an overview of the current machine learning researches dealing with EO data, and, on the other hand, we want to stimulate concrete discussions to pave the way to new machine learning frameworks especially tailored to deal with such data.

Organisers: Thomas Corpetti, Dino Ienco, Roberto Interdonato, Minh-Tan Pham, Sébastien Lefèvre

Automating Data Science, Room 1.011
sites.google.com/view/autods

Data science is concerned with the extraction of knowledge and insight, and ultimately societal or economic value, from data. It complements traditional statistics in that its object is data as it presents itself in the wild (often complex and heterogeneous, noisy, loosely structured, biased, etc.), rather than well-structured data sampled in carefully designed studies. It also has a strong computer science focus, and is related to popular areas such as big data, machine learning, data mining and knowledge discovery. It is therefore highly relevant to the ECMLPKDD community. Data science is becoming increasingly important with the abundance of big data, while the number of skilled data scientists is lagging. This has raised the question as to whether it is possible to automate data science in several contexts. First, from an artificial intelligence perspective, it is interesting to investigate whether (data) science (or portions of it) can be automated, as it is an activity currently requiring high levels of human expertise. Second, the field of machine learning has a long-standing interest in applying machine learning at the meta-level, in order to obtain better machine learning algorithms, yielding recent successes in automated parameter tuning, algorithm configuration and algorithm selection. Third, there is an interest in automating not only the model building process itself (cf. the Automated Statistician) but also in automating the preprocessing steps (data wrangling) and the postprocessing steps (model deployment, monitoring and maintenance).

Organisers: Tijl De Bie (UGent, Belgium), Luc De Raedt (KU Leuven, Belgium), Jose Hernandez-Orallo (Universitat Politecnica de Valencia, Spain)

The Fourth Workshop on Data Science for Social Good, Room 2.012
sites.google.com/view/ecmlpkddsogood2019

The aim of the workshop is to present applications of Data Science to Social Good, or else that take into account social aspects of Data Science methods and techniques. Every domain shall be considered, particularly if aligned with the UN Sustainable Development Goals. We also want to attract proposals of data-related projects that are looking for collaborators

Organisers: Ricard Gavalda (UPC BarcelonaTech, Spain), Irena Koprinska (University of Sydney, Australia),
João Gama (University of Porto, Portugal)

Advances in managing and mining large evolving graphs (third edition), Room 1.010
leg-ecmlpkdd19.loria.fr

The aim of the workshop called Managing and mining Large Evolving Graphs (LEG) is to bring together active scholars and practitioners of dynamic graphs. Graph models and algorithms are ubiquitous of a large number of application domains, ranging from transportation to social networks, semantic web, or data mining. However, many applications require graph models that are time dependent. For example, applications related to urban mobility analysis employ a graph structure of the underlying road network where the traffic density and speed continuously change over time. Therefore, the time a moving object takes to cross a path segment typically depends on the starting instant of time. This dynamicity makes it more challenging to mine temporal and graph patterns, yet this task is essential to study such structures. The same holds in other contexts, such as social networks. In this workshop,

we aim to discuss the problem of mining large evolving graphs, since there are many real-world applications deal with large volumes of such data. Managing and analysing large evolving graphs is very challenging since this requires sophisticated methods and techniques for creating, storing, accessing and processing such graphs in a distributed environment, because centralized approaches do not scale in a Big Data scenario. Contributions will clearly point out answers to one of these challenges focusing on large-scale graphs.

Organisers: Sabeur Aridhi (LORIA/Inria NGE, University of Lorraine, France), José Antonio de Macedo (Universidade Federal do Ceará, Fortaleza, Brazil), Engelbert Mephu Nguifo (LIMOS, University Clermont Auvergne, France), Karine Zeitouni (DAVID, Université de Versailles Saint-Quentin, France)

Data and Machine Learning Advances with Multiple Views, Room 1.002

damvl.lis-lab.fr

Multiview machine learning occurs whenever a model has to be learned to process tasks over data coming from different description spaces. For example, medical diagnosis might use various types (aka views) of examinations (MRI, ECG, blood analysis, etc.) in order to take a decision. These views are supposed to carry different types of information regarding the learning task, which means that these views reveal different types of patterns regularities. The workshop aims at bringing together people interested with multi-view learning, both from dataset providers to researchers in machine learning. Such a way, researchers could easily have the opportunity to inspect the reality of some true learning problems related to multi-view learning, meanwhile providers of natural multi-viewed data could get aware of the many current or potential solutions to address their learning tasks. From this perspective, in addition to a usual session of talks and posters, we propose to organise a tiny challenge – a hackathon – on real multiview data, which could be provided by participants, in order to highlight a synergy during the workshop.

Organisers: Stéphane Ayache, Cécile Capponi, Rémi Emonet, Isabelle Guyon

Workshop on Data Integration and Applications, Room 1.003

sites.google.com/view/dina2019

The goal of the Data Integration and Applications (DINA) workshop is to bring together computer scientists with researchers from other domains and practitioners from businesses and governments to present and discuss current research directions on multi-source data integration and its application. The workshop will provide a forum for original high-quality research papers on record linkage, data integration, population informatics, mining techniques of integrated data, and applications, as well as multidisciplinary research opportunities.

Organisers: Luiza Antonie (University of Guelph, Canada), Peter Christen (The Australian National University, Australia), Erhard Rahm (University of Leipzig, Germany), Osmar Zaïane (University of Alberta, Canada)

XKDD Tutorial and XKDD - AIMLAI Workshop, Room 0.001

kdd.isti.cnr.it/xkdd2019

The purpose of AIMLAI-XKDD (Advances in Interpretable Machine Learning and Artificial Intelligence & eXplainable Knowledge Discovery in Data Mining), is to encourage principled research that will lead to the advancement of explainable, transparent, ethical and fair data mining, machine learning, artificial intelligence. AIMLAI-XKDD is an event organized into two moments: a tutorial to introduce audience to the topic, and a workshop to discuss recent advances in the research field. The tutorial will provide a broad overview of the state of the art and the major applications for explainable and transparent approaches. Likewise, it will highlight the main open challenges. The workshop will seek top-quality submissions addressing uncovered important issues related to explainable and interpretable data mining and machine learning models. Papers should present research results in any of the topics of interest for the workshop as well as application experiences, tools and promising preliminary ideas. AIMLAI-XKDD asks for contributions from researchers, academia and industries, working on topics addressing these challenges primarily from a technical point of view, but also from a legal, ethical or sociological perspective.

Organisers: Riccardo Guidotti (KDD Lab, ISTI-CNR, Italy), Pasquale Minervini (University College London, UK), Anna Monreale (KDD Lab, University of Pisa, Italy), Salvatore Rinzivillo (KDD Lab, ISTI-CNR, Italy), Adrien Bibal (University of Namur, Belgium), Tassadit Bouadi (University of Rennes/IRISA, France), Benoît Frénay (University of Namur, Belgium), Luis Galárraga (Inria/IRISA, France), Stefan Kramer (Universität Mainz, Germany), Ruggero G. Pensa (University of Turin, Italy)

Discovery Challenges

There are four distinct challenges, two in the morning (10:30 – 12:40), and two in the afternoon (14:00 – 16:00)

DC1: Multimodal (Audio, Facial and Gesture) based Emotion Recognition Challenge, Room 2.006 (Morning)
www.merip.imsi.pl

People express emotions through different modalities. Integration of verbal and non-verbal communication channels creates a system in which the message is easier to understand. Expanding the focus to several expression forms can facilitate research one motion recognition as well as human-machine interaction. In this competition, the authors present a Polish emotional database composed of three modalities: facial expressions, body movement and gestures, and speech. The corpora contain recordings registered in studio conditions, acted out by 16 professional actors (8 male and 8 female). The data is labeled with six basic emotions categories, according to Ekman's emotion categories.

The participants will have to analyze all 3 modalities and, based on all 3 modalities, perform the emotion recognition. The participants must submit the code and all dependencies via codalab and the organizer will run the codes. The evaluation would be based on the average correct emotion recognition using each modality as well as all 3 modalities together. In case of equal performance, the processing time will be used in order to indicate the ranking. The Training data will be provided followed by the validation dataset. The test data will be finally launched with no label and it will be used for the evaluation of participants.

Organisers: Dorota Kaminska (Lodz University of Technology), Tomasz Sapiński (Lodz University of Technology), Kamal Nasrollahi (University of Aalborg), Hasan Demirel (Eastern Mediterranean University), Cagri Ozcinar (Trinity College Dublin), Gholamreza Anbarjafari (iCV Lab, University of Tartu)

DC2: SIMAH (SocIaL Media And Harassment): First workshop on categorizing different types of online harassment languages in social media, Room 2.007 (Morning)
sites.google.com/view/simah-2019/home

The proposed competition focusing of online harassment in Twitter in English. It has two related tasks: the first task is a binary classification to classify online harassment tweets versus not_harassment tweets, the second task is multi-class classification of online harassment tweets into three categories of "Indirect harassment", "sexual harassment" and "physical harassment".

Organisers: Sima Sharifirad, Stan Matwin (Dalhousie University, Halifax, Canada)

DC3: Correcting Transiting Exoplanet Light Curves for Stellar Spots, Room 2.006 (Afternoon)
ariel-datachallenge.space

The field of exoplanet discovery and characterisation has been growing rapidly in the last decade. However, several big challenges remain, many of which could be addressed using machine learning and data mining methodology. For instance, the most successful method for detecting exoplanets, transit photometry –measuring the faint decrease in incoming stellar light as an exoplanet passes

between the Earth and a target star– is very sensitive to the presence of stellar spots and faculae. The current approach is to identify the effects of spots visually and correct for them manually or discard the data. As a first step to automate this process, we propose a regular competition on data generated by ArielSim, the simulator of the European Space Agency's upcoming Ariel mission, whose objective is to characterise the atmosphere of 1000 exoplanets. The data consist of pairs of light curves corrupted by stellar spots and the corresponding clean ones, along with auxiliary observation information. The goal is to correct the light curve for the presence of stellar spots (signal denoising). This is a yet unsolved problem in the community. Solving it will mean improving our understanding of the characteristics of currently confirmed exoplanets, potentially recognising false positive / false negative detections and improving our ability to analyse new observations – primarily but not limited to those expected from Ariel– without the need to equip new telescopes with additional instruments with all the extra costs this implies.

Organisers: Nikolaos Nikolaou (UCL, England), Ingo P. Waldmann (UCL, England), Subhajit Sarkar (University of Cardiff, Wales), Angelos Tsiaras (UCL, England), Billy Edwards (UCL, England), Mario Morvan (UCL, England), Kai Hou Yip (UCL, England), Giovanna Tinetti (UCL, England)

DC4: AutoCV2 Challenge, Room 2.007 (Afternoon)
autodl.chalearn.org

As part of the AutoDL challenges, the AutoCV2 challenge aims at finding fully automated solutions for classification tasks in computer vision. Compared to the recent AutoCV challenge, AutoCV2 challenge targets not only image classification tasks, but also video classification tasks. Participants need to make code submissions containing machine learning code that is trained and tested on the CodaLab platform, without human intervention whatsoever, with time and memory limitations. All problems are multi-label classification problems, coming from various domains. Raw data is provided, but formatted in a uniform manner, to encourage participants to submit generic algorithms.

Organisers: Sergio Escalera (U. of Barcelona/Computer Vision Center Barcelona, Spain), Isabelle Guyon (ChaLearn, USA - Inria/Université Paris-Saclay, France), Zhengying Liu (Inria/Université Paris-Saclay, France), Wei-Wei Tu (4Paradigm, China)

Demonstrations

Demonstration papers also appear in the proceedings. The demonstrations will be performed during the first post session on Wednesday (New University).

BK-ADAPT: Dynamic Background Knowledge for Automating Data Transformation

Lidia Contreras-Ochando (Universitat Politècnica de València), Cèsar Ferri (Universitat Politècnica València), Jose Hernandez-Orallo (Polytechnic University of Valencia), Fernando Martínez-Plumed (Technical University of Valencia), M. José Ramírez-Quintana (Technical University of Valencia), Susumu Katayama (University of Miyazaki)

An enormous effort is usually devoted to data wrangling, the tedious process of cleaning, transforming and combining data, such that it is ready for modelling, visualisation or aggregation. Data transformation and formatting is one common task in data wrangling, which is performed by humans in two steps: (1) they recognise the specific domain of data (dates, phones, addresses, etc.) and (2) they apply conversions that are specific to that domain. However, the mechanisms to manipulate one specific domain can be unique and highly different from other domains. In this paper we present bka, a system that uses inductive programming (IP) with a dynamic background knowledge (BK) generated by a machine learning meta-model that selects the domain and/or the primitives from several descriptive features of the data wrangling problem. To show the performance of our method, we have created a web-based tool that allows users to provide a set of inputs and one or more examples of outputs, in such a way that the rest of examples are automatically transformed by the tool.

A Tool for Researchers: Querying Big Scholarly Data through Graph Databases

Fabio Mercorio (University of Milan-Bicocca), Mario Mezzanzanica (University of Milan-Bicocca), Vincenzo Moscato (University of Naples "Federico II"), Antonio Picariello (University of Naples "Federico II"), Giancarlo Sperlì (University of Naples "Federico II")

We demonstrate GraphDBLP, a tool to allow researchers for querying the DBLP bibliography as a graph. The DBLP source data were enriched with semantic similarity relationships computed using word-embeddings. A user can interact with the system either via a Web-based GUI or using a shell-interface, both provided with three parametric and pre-defined queries. GraphDBLP would represent a first graph-database instance of the computer scientist network, that can be improved through new relationships and properties on nodes at any time, and this is the main purpose of the tool, that is freely available on Github. To date, GraphDBLP contains 5+ million nodes and 24+ million relationships.

OCADaMi: One-Class Anomaly Detection and Data Mining toolbox

Andreas Theissler (Aalen University of Applied Sciences)

This paper introduces the modular anomaly detection toolbox OCADaMi that incorporates machine learning and visual analytics. The case often encountered in practice where no or only a non-representative number of anomalies exist beforehand is addressed, which is solved using one-class classification. Target users are developers, engineers, test engineers and operators of technical systems. The users can interactively analyse data and define workflows for the detection of anomalies and visualisation. There is a variety of application-domains, e.g. manufacturing or testing of automotive systems. The functioning of the system is shown for fault detection in real-world automotive data from road trials.

MatrixCalculus.org - Computing Derivatives of Matrix and Tensor Expressions
Sören Laue, Matthias Mitterreiter, Joachim Giesen (Friedrich Schiller University Jena)

Computing derivatives of matrix and tensor expressions is an integral part of developing and implementing optimization algorithms in machine learning. However, it is a time-consuming and error-prone task when done by hand. Hence, we present the first system that performs matrix and tensor calculus automatically.

Towards a Predictive Patent Analytics and Evaluation Platform
Nebula Alam, Khoi-Nguyen Tran, Sue Ann Chen, John Wagner, Josh Andres, Mukesh Mohania (IBM Research)

The importance of patents is well recognised across many regions of the world. Many patent mining systems have been proposed, but with limited predictive capabilities. In this demo, we showcase how predictive algorithms leveraging the state-of-the-art machine learning and deep learning techniques can be used to improve understanding of patents for inventors, patent evaluators, and business analysts alike.

A Virtualized Video Surveillance System for Public Transportation
Talmaj Marinč (Fraunhofer Heinrich Hertz Institute), Serhan Gül (Fraunhofer HHI), Cornelius Hellge (Fraunhofer HHI), Peter Schüßler (DResearch Fahrzeugelektronik GmbH), Thomas Riegel (Siemens Corporate Technology), Peter Amon (Siemens Corporate Technology)

Modern surveillance systems have recently started to employ computer vision algorithms for advanced analysis of the captured video content. Public transportation is one of the domains that may highly benefit from the advances in video analysis. This paper presents a video-based surveillance system that uses a deep neural network-based face verification algorithm to accurately and robustly re-identify a subject person. Our implementation is highly scalable due to its container-based architecture and is easily deployable on a cloud platform to support larger processing loads. During the demo, the users will be able to interactively select a target person from pre-recorded surveillance videos and inspect the results on our web-based visualization platform.

Distributed Algorithms to Find Similar Time Series

Oleksandra Levchenko (INRIA), Boyan Kolev (INRIA), Djamel Edine Yagoubi (INRIA), Dennis Shasha (NYU, USA), Themis Palpanas (Paris Descartes University), Patrick Valduriez (INRIA), Reza Akbarinia (INRIA), Florent Masseglia (INRIA)

As sensors improve in both bandwidth and quantity over time, the need for high performance sensor fusion increases. This requires both better (quasi-linear time if possible) algorithms and parallelism. This demonstration uses financial and seismic data to show how two state-of-the-art algorithms construct indexes and answer similarity queries using Spark. Demo visitors will be able to choose query time series, see how each algorithm approximates nearest neighbors and compare times in a parallel environment.

UnFOOT: Unsupervised Football Analytics Tool

José Carlos Coutinho (University of Twente), João Moreira (INESC TEC), Claudio Rebelo de Sá (University of Twente)

Labelled football (soccer) data is hard to acquire and it usually needs humans to annotate the match events. This process makes it more expensive to be obtained by smaller clubs. UnFOOT (Unsupervised Football Analytics Tool) combines data mining techniques and basic statistics to measure the performance of players and teams from positional data. The capabilities of the tool involve preprocessing the match data, extraction of features, visualization of player and team performance. It also has built-in data mining techniques, such as association rule mining and subgroup discovery.

ISETS: Incremental Shapelet Extraction from Streaming Time Series

Jingwei Zuo, Karine Zeitouni, Yehia Taher (Université de Versailles-St-Quentin)

In recent years, Time Series (TS) analysis has attracted widespread attention in the community of Data Mining due to its special data format and broad application scenarios. An important aspect in TS analysis is Time Series Classification (TSC), which has been applied in medical diagnosis, human activity recognition, industrial troubleshooting, etc. Typically, all TSC work trains a stable model from an off-line TS dataset, without considering potential Concept Drift in streaming con-text. Conventional data stream is considered as independent examples (e.g., row data) coming in real-time, but rarely considers real-valued data coming in a sequential order, called Streaming Time Series. Processing such type of data, requires combining techniques in both communities of Time Series (TS) and Data Streams. To facilitate the users' understanding of this combination, we propose ISETS, a web-based application which allows users to monitor the evolution of interpretable features in Streaming Time Series.

Industrial Event Log Analyzer - Self-ServiceData Mining for Domain Experts

Reuben Borrison (ABB), Benjamin Klöpper (ABB Research), Sunil Saini (ABB)

Industrial applications of machine learning rely heavily ondeep domain knowledge that data scientist and machine learning expertusually do not have. Iterative and time-consuming communicantion between machine learning expert and domain expert are the consequence. In this demo we introduce a functional mock-up that demonstrates thatdomain users can be guided through a machine learning process if thescope of problem and data type is narrowed done.

Conference Organisers

General Chairs

gc_chairs@ecmlpkdd2019.org
Élisa Fromont, IRISA/INRIA, University of Rennes 1, France
Arno Knobbe, Leiden Institute of Advanced Computer Science, the Netherlands

Program Chairs (Research Track and Applied Data Science Track)

pc_chairs@ecmlpkdd2019.org
Ulf Brefeld, Leuphana University of Lüneburg, Germany
Andreas Hotho, University of Würzburg, Germany
Marloes Maathuis, ETH Zürich, Switzerland
Céline Robardet, LIRIS, INSA-Lyon, France

Journal Track Chairs

jt_chairs@ecmlpkdd2019.org
Karsten Borgwardt, D-BSSE, ETH Zürich, Switzerland
Po-Ling Loh, University of Wisconsin, USA
Evimaria Terzi, Boston University, USA
Antti Ukkonen, University of Helsinki, Finland

Local Chairs

local_chairs@ecmlpkdd2019.org
Lena Hettinger, University of Würzburg, Germany
Andreas Hotho, University of Würzburg, Germany
Kristof Korwisi, University of Würzburg, Germany
Marc Erich Latoschik, University of Würzburg, Germany

Proceeding Chairs

proceeding_chairs@ecmlpkdd2019.org
Xin Du, Eindhoven University of Technology, the Netherlands
Wouter Duivesteijn, Eindhoven University of Technology, the Netherlands
Sibylle Hess, Eindhoven University of Technology, the Netherlands

Discovery Challenge Chairs

dc_chairs@ecmlpkdd2019.org
Sergio Escalera, University of Barcelona, Spain
Isabelle Guyon, Paris-Sud University, France

Workshop & Tutorial Chairs

wt_chairs@ecmlpkdd2019.org
Peggy Cellier, IRISA, INSA Rennes, France
Kurt Driessens, Maastricht University, the Netherlands

Demonstration Chairs

demo_chairs@ecmlpkdd2019.org

Martin Atzmüller, Tilburg University, the Netherlands
Emilie Morvant, University Saint-Etienne, France

PhD Forum Chairs

phd_chairs@ecmlpkdd2019.org

Tassadit Bouadi, IRISA, University of Rennes 1, France
Tias Guns, Vrije Universiteit Bruxelles, Belgium

Production, Publicity & Public Relations Chairs

pppr_chairs@ecmlpkdd2019.org

Parisa Kordjamshidi, Tulane University/Florida IHMC, USA
Albrecht Zimmermann, Université de Caen Normandie, France

Awards Committee

awards_committee@ecmlpkdd2019.org

Katharina Morik, TU Dortmund, Germany
Geoff Webb, Monash University, Australia

Sponsorship Chairs

sponsor_chair@ecmlpkdd2019.org

Albert Bifet, LTCI, Télécom ParisTech, France
Heike Trautmann, University of Münster, Germany

Web Chair

web_chairs@ecmlpkdd2019.org

Florian Lautenschlager, University of Würzburg, Germany
Vanessa Breitenbach, University of Würzburg, Germany

Program Committees

Guest Editorial Board, Journal Track

Annalisa Appice
Marta Arias
Martin Atzmueller
Albert Bifet
Hendrik Blockeel
Toon Calders
Michelangelo Ceci
Loïc Cerf
Nicolas Courty
Bruno Crémilleux
Tijl De Bie
Krzysztof Dembczyński
Yagoubi Djamel Edine
Tapio Elomaa
Rémi Emonet
Stefano Ferilli
Joao Gama
Tias Guns
Amaury Habrard
Xiao He
Jaakko Hollmén
Szymon Jaroszewicz
Alipio Jorge
Ajin Joseph
Samuel Kaski
Kristian Kersting
Dragi Kocev
Peer Kröger
Ondrej Kuzelka
Mark Last
Matthijs van Leeuwen
Limin Li
Jessica Lin
Christoph Lippert
Brian Mac Namee
Gjorgji Madjarov
Fabrizio Maria Maggi
Ernestina Menasalvas
Aditya Menon
Animesh Mukherjee
Maximilian Nickel
Siegfried Nijssen
Panče Panov
Panagiotis Papapetrou
Andrea Passerini
Mykola Pechenizkiy
Charlotte Pelletier
Ruggero Pensa
Marc Plantevit
Pascal Poupart
Kai Puolamäki
Chedy Raissi
Bastian Rieck
Matteo Riondato
Marko Robnik-Sikonja
Matthias Schubert
Fabrizio Sebastiani
Thomas Seidl
Gerasimos Spanakis
Myra Spiliopoulou
Jerzy Stefanowski
Mahito Sugiyama
Nikolaj Tatti
Alexandre Termier
Ivor Tsang
Panayiotis Tsaparas
Charalampos Tsourakakis
Nick Vannieuwenhoven
Celine Vens
Antonio Vergari
Christel Vrain
Jilles Vreeken
Byron Wallace
Stefan Wrobel
Yinchong Yang
Ye Zhu
Albrecht Zimmermann
Marinka Zitnik
Indrė Žliobaitė

Area Chairs, Research and Applied Data Science Tracks

Fabrizio Angiulli
Roberto Bayardo
Michael Berthold
Albert Bifet
Hendrik Blockeel
Francesco Bonchi
Toon Calders
Michelangelo Ceci
Nicolas Courty
Bruno Cremilleux
Philippe Cudre-Mauroux
Jesse Davis
Tijl De Bie
Tapio Elomaa
Amir-massoud Farahmand
Paolo Frasconi
Johannes Fürnkranz
Patrick Gallinari
João Gama
Aristides Gionis
Thomas Gärtner
Allan Hanbury
Jaakko Hollmén
Eyke Hüllermeier
Alipio Jorge
Marius Kloft
Nick Koudas
Stefan Kramer
Sébastien Lefèvre
Jörg Lücke
Giuseppe Manco
Pauli Miettinen
Anna Monreale
Katharina Morik
Siegfried Nijssen
Andrea Passerini
Mykola Pechenizkiy
Francois Petitjean
Elmar Rueckert
Tom Schaul
Thomas Seidl
Arno Siebes
Myra Spiliopoulou
Einoshin Suzuki
Marc Tommasi
Celine Vens
Christel Vrain
Jilles Vreeken
Min-Ling Zhang
Herke van Hoof

Program Committee Members, Research and Applied Data Science Tracks

M. Ehsan Abbasnejad
Leman Akoglu
Tristan Allard
Aijun An
Ali Anaissi
Annalisa Appice
Paul Assendorp
Ira Assent
Martin Atzmüller
Alexandre Aussem
Suyash Awate
Antonio Bahamonde
Jaume Baixeries
Jose Balcazar
Sambaran
Bandyopadhyay
Zhifeng Bao
Mitra Baratchi
Gianni Barlacchi
Martin Becker
Srikanta Bedathur
Vaishak Belle
Andras Benczur
Daniel Bengs
Petr Berka
Bertrand Cuissart
Marenglen Biba
Chris Biemann
Battista Biggio
Bernd Bischl
Thomas Bonald
Gianluca Bontempi
Henrik Bostrom
Tassadit Bouadi
Ahcène Boubekki
Zied Bouraoui
Paula Branco
Pavel Brazdil
Dariusz Brzezinski
Sebastian Buschjäger
Narayanan C Krishnan
RicardoJ. G. B. Campello
Brais Cancela
Francisco Casacuberta
Remy Cazabet
Peggy Cellier
Loïc Cerf
Tania Cerquitelli
Ricardo Cerri
Tanmoy Chakraborty
Edward Chang
Jeremy Charlier
Sneha Chaudhari
Abon Chaudhuri
Keke Chen
Weiwei Cheng
Giovanni Chierchia
Silvia Chiusano
Sunav Choudhary
Frans Coenen
Mário Cordeiro
Robson Cordeiro
Roberto Corizzo
Fabrizio Costa
Vitor Santos Costa
Adrien Coulet
Boris Cule
MariaLuisa Damiani
Martine De Cock
Anne Denton
Christian Desrosiers
Nicola Di Mauro
Claudia Diamantini
Jilles S. Dibangoye
Tom Diethe
Wei Ding
Stephan Doerfel
Carlotta Domeniconi
Madalina Drugan
Stefan Duffner
Wouter Duivesteijn
Sebastijan Dumancic
Inês Dutra
Mireille El Gheche
Dora Erdos
Samuel G. Fadel
Ad Feelders
Jing Feng
Stefano Ferilli
Carlos Ferreira
Cèsar Ferri
Matthias Fey
Răzvan V. Florian
Germain Forestier
Eibe Frank
Fabio Fumarola
Amita Gajewar
Esther Galbrun
Dragan Gamberger
Byron Gao
Junbin Gao
Paolo Garza
Konstantinos Georgatzis
Pierre Geurts
Arnaud Giacometti
Rémi Gilleron
Dorota Glowacka
Uwe Glässer
Nico Goernitz
HeitorM Gomes
James Goulding
Cédric Gouy-Pailler
Josif Grabocka
Michael Granitzer
Derek Greene
Quanquan Gu
Riccardo Guidotti
Francesco Gullo
Tias Guns
Xueying Guo
Deepak K Gupta
Thomas Guyet
Vicenç Gómez
Vanessa Gómez-Verdejo
Stephan Günnemann
Maria Halkidi
Barbara Hammer
Jiawei Han
Tom Hanika
Mohammad Hasan
Xiao He
Denis Helic
Andreas Henelius
Daniel Hernandez-Lobato
Jose Hernandez-Orallo
Sibylle Hess
Thanh Lam Hoang
Frank Hoeppner
Arjen Hommersom
Tamas Horvath
Homa Hosseinmardi
Chao Huang
Yuanhua Huang
NeilJoseph Hurley
Dino Ienco
Iñaki Inza
Tomoki Ito

Mahdi Jalili
Szymon Jaroszewicz
Giuseppe Jurman
Anup Kalia
Toshihiro Kamishima
Michael Kamp
Bo Kang
Pinar Karagoz
Konstantinos Karanasos
Sarvnaz Karimi
George Karypis
Mehdi Kaytoue
Mikaela Keller
Latifur Khan
Daiki Kimura
Frank Klawonn
Jiri Klema
Tomas Kliegr
Dragi Kocev
Levente Kocsis
Yun Sing Koh
Effrosyni Kokiopoulou
Alek Kolcz
Irena Koprinska
Frederic Koriche
Lars Kotthoff
Danai Koutra
Polychronis Koutsakis
Tomas Krilavicius
Yamuna Krishnamurthy
Matjaz Kukar
Meelis Kull
Vinod Kumar Kurmi
Ondrej Kuzelka
Sofiane LAGRAA
Nicolas Lachiche
Leo Lahti
Christine Largeron
Christine Largouet
Pedro Larranaga
Niklas Lavesson
BinhThanh Le
Florian Lemmerich
Dagang Li
Jian Li
Jiuyong Li
Limin Li
Jefrey Lijffijt
Tony Lindgren
Marco Lippi
Bing Liu
Corrado Loglisci
Peter Lucas
Sebastian Mair
Arun Maiya
Donato Malerba
Luca Martino
Elio Masciari
Andres Masegosa
Florent Masseglia
Wannes Meert
Corrado Mencar
Saayan Mitra
Atsushi Miyamoto
Dunja Mladenic
Sandy Moens
Miguel Molina-Solana
Nuno Moniz
Hankyu Moon
Joao Moreira
Luis Moreira-Matias
Emilie Morvant
Andreas Mueller
Asim Munawar
Pierre-Alexandre Murena
Vineeth N Balasubramanian
Mohamed Nadif
Jinseok Nam
Mirco Nanni
Amedeo Napoli
Sriraam Natarajan
Benjamin Negrevergne
Benjamin Nguyen
Xia Ning
Kjetil Norvag
Eirini Ntoutsi
Andreas Nurnberger
Luca Oneto
Kok-Leong Ong
Francesco Orsini
Nikunj Oza
Panče Panov
Apostolos N Papadopoulos
Panagiotis Papapetrou
Youngja Park
Ioannis Partalas
Charlotte Pelletier
Jaakko Peltonen
Lukas Pfahler
Nico Piatkowski
Andrea Pietracaprina
Gianvito Pio
Claudia Plant
Marc Plantevit
Pascal Poncelet
Miguel A. Prada
Paul Prasse
Philippe Preux
Patrick Putzky
Buyue Qian
Yongrui Qin
Dimitrios Rafailidis
Jörg Rahnenführer
Chedy Raissi
Sutharshan Rajasegarar
Sai Rajeshwar
Jan Ramon
Santu Rana
Huzefa Rangwala
Chotirat Ratanamahatana
Jan Rauch
Chiara Renso
Achim Rettinger
Christophe Rigotti
Fabrizio Riguzzi
Matteo Riondato
Juan Rodriguez
Fabrice Rossi
Ryan A. Rossi
Céline Rouveirol
Stefan Rueping
Anne Sabourin
Yvan Saeys
Amrita Saha
Lorenza Saitta
Tomoya Sakai
Tetsuya Sakurai
Ansaf Salleb-Aouissi
Somdeb Sarkhel
Claudio Sartori
Luciano Sbaiz
Pierre Schaus
Tobias Scheffer
Ute Schmid
Lars Schmidt-Thieme
Christoph Schommer
Matthias Schubert
Rajat Sen

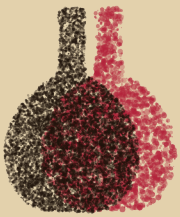
Vinay Setty
Chao Shang
Junming Shao
Bernhard Sick
Nikola Simidjievski
Andrzej Skowron
Dominik Slezak
Gavin Smith
Tomislav Šmuc
Arnaud Soulet
Mauro Sozio
Alessandro Sperduti
Jerzy Stefanowski
Giovanni Stilo
Mahito Sugiyama
Mika Sulkava
Yizhou Sun
Viswanathan
Swaminathan
Stephen Swift
Andrea Tagarelli
Domenico Talia
Letizia Tanca
Jovan Tanevski
Nikolaj Tatti
Maryam Tavakol
Maguelonne Teisseire
Alexandre Termier
Stefano Teso
Ljupco Todorovski
Ricardo Torres
Volker Tresp
Isaac Triguero
Ivor Tsang
Vincent Tseng
Charalampos Tsourakakis
Radu Tudoran
Cigdem Turan
Nikolaos Tziortziotis
Theodoros Tzouramanis
Antti Ukkonen
Daniel V. Smith
Martijn Van Otterlo
Jan Van Rijn
Robin Vandaele
Iraklis Varlamis
Ranga Vatsavai
Julien Velcin
Bruno Veloso
Shankar Vembu
Deepak Venugopal
Ricardo Vigario
Prashanth Vijayaraghavan
HernaL Viktor
Willem Waegeman
Di Wang
Hao Wang
Jianyong Wang
Yuyi Wang
Jeremy C. Weiss
Joerg Wicker
Marco Wiering
Martin Wistuba
Christian Wolf
Gang Wu
Lina Yao
Philip Yu
Bianca Zadrozny
Gerson Zaverucha
Bernard Zenko
Chengkun Zhang
Jianpeng Zhang
Junping Zhang
Shichao Zhang
Yingqian Zhang
Sichen Zhao
Ying Zhao
Shuai Zheng
Arthur Zimek
Albrecht Zimmermann
Indrė Žliobaitė
Tanja Zseb
Claudia d'Amato
Juan Jose del Coz
Bas van Stein

Author Index

Booklet compiled by Ulf Brefeld, Wouter Duivesteijn, Arno Knobbe, Kristof Korwisi, Florian Lautenschlager

Photos: City of Würzburg and University of Würzburg

Maps: Data copyrighted OpenStreetMap contributors and available from https://www.openstreetmap.org



ECMLPKDD

Würzburg 2019