ECML PKDD 2014

Sept. 15-19

Nancy France

EUROPEAN CONFERENCE ON MACHINE LEARNING AND
PRINCIPLES AND PRACTICE OF KNOWLEDGE DISCOVERY IN DATABASES

PROGRAMME

PARTNERS AND SUPPORTERS

We would like to thank the following sponsors and supporters

Gold Sponsor

WINTON

Silver Sponsors

Deloitte. xerox

Bronze Sponsors

technicolor EDF orange

YAHOO! LABS CAMBRIDGE UNIVERSITY PRESS

Additional Supporters

Deloitte. HARMONIC PHARMA technicolor

Lanyard

Open for Innovation KNIME

Prize

Springer Deloitte.

Organizing Institutions

Inria cnrs Loria

Institutional Sponsors

ministère éducation nationale enseignement supérieur recherche

La Région Lorraine

Opération réalisée avec le concours financier du conseil regional de Lorraine

GrandNancy COMMUNAUTÉ URBAINE & HUMAINE

CONTENT

CONFERENCE SECRETARIAT

INRIA Nancy-Grand Est
Manifestations scientifiques
615 rue du jardin Botanique
54600 Villers-lès-Nancy, France
E-mail: colloques-ncy@inria.fr

BIENVENUE!

Dear colleagues,

It is our great pleasure to welcome you to the combined ECML/PKDD 2014 and ILP 2014 conferences in Nancy, France, from 14-19 September, 2014.

The local organization committee, the conference chairs, the PC chairs, and all of the other dedicated chairs, have worked in harmony to organize ECML/PKDD + ILP 2014 here in Nancy. All of these people (their names are listed on the conference website) have put all their energy into offering a scientific and social program of very high quality and diversity.

However, a conference is above all made by its authors, invited speakers, and its tutorial and workshop organizers, all of whom we warmly thank for their active and valuable contributions. This year, more than 500 people will again be present to make ECML/PKDD + ILP 2014 a major world conference on machine learning and knowledge discovery.

We hope this booklet will give you all the necessary information to find your way through the scientific and social programs of the conference, and through the city of Nancy as well.

Enjoy your participation in ECML/PKDD + ILP 2014 and your stay in the city of Nancy!

Toon Calders
Floriana Esposito
Eyke Hüllermeier
Rosa Meo
ECML/PKDD Program chairs

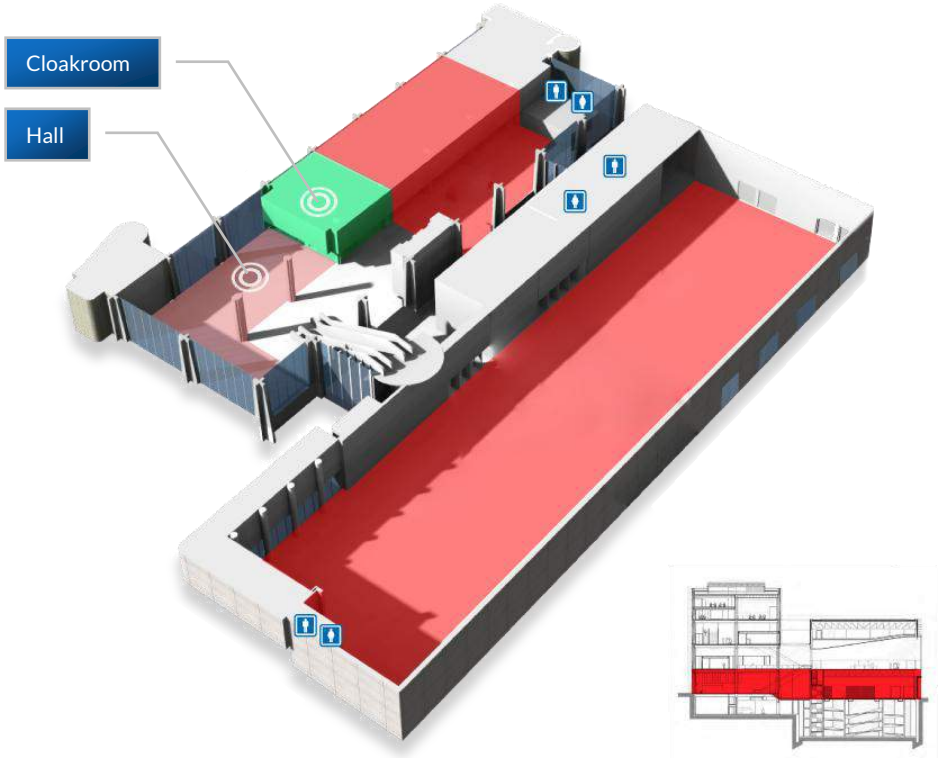
Jesse Davis
Jan Ramon
ILP Program chairs

Amedeo Napoli
Chedy Raïssi
Conference chairs

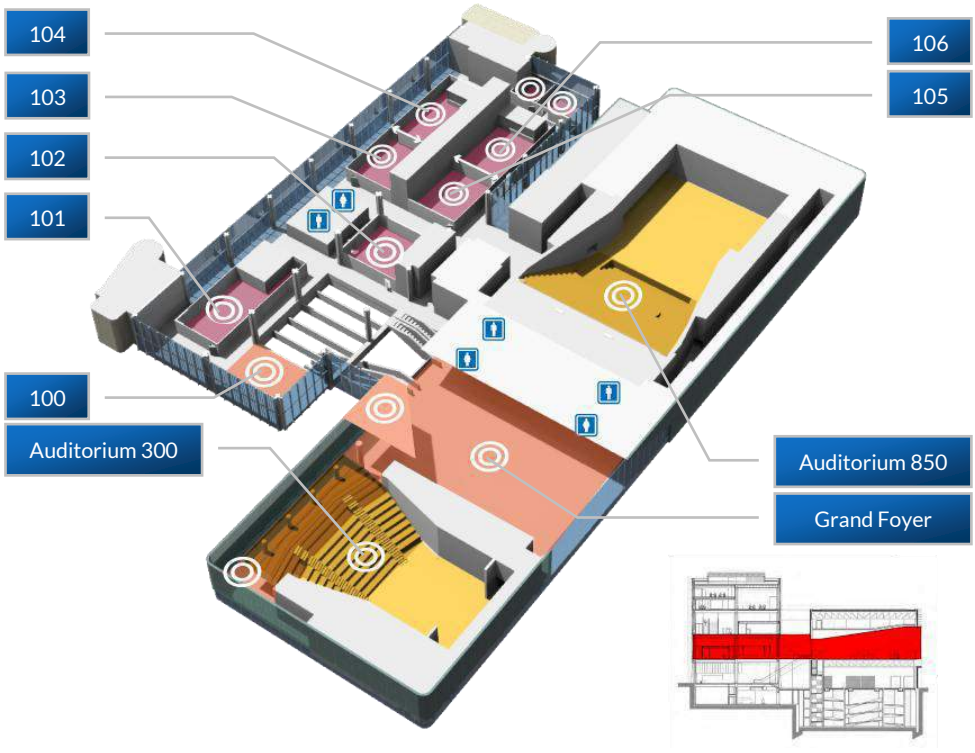
MERCI!

FLOOR PLANS

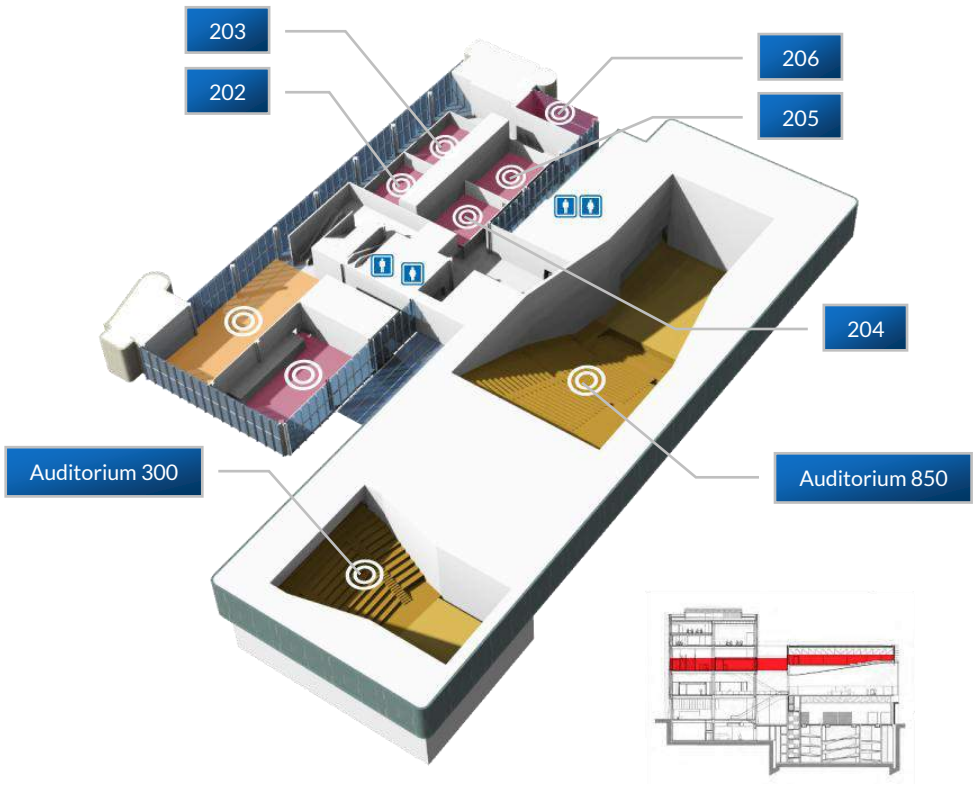
LEVEL 0



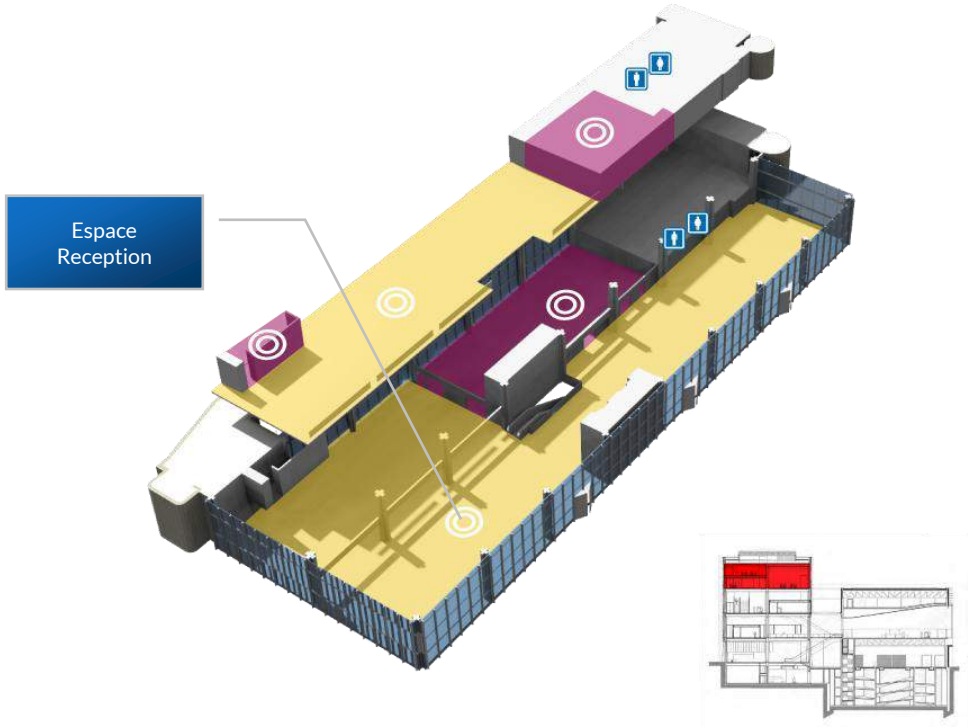
LEVEL 1



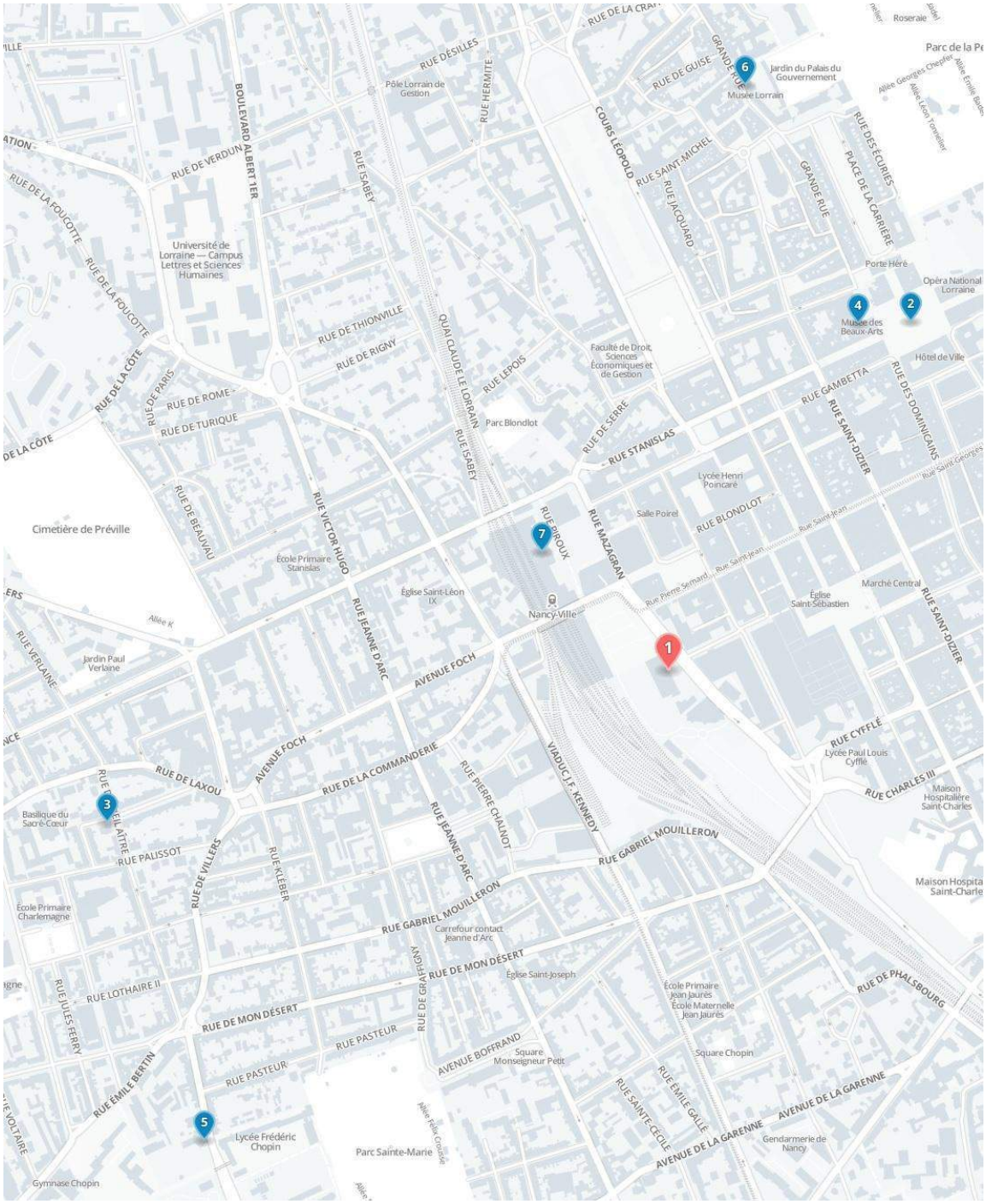
LEVEL 2



LEVEL 3



USEFUL MAPS



1 **Conference Venue**
Centre Prouvé, 8 Boulevard Joffre
2 **Place Stanislas**
3 **Villa Majorelle**, 1 Rue Louis Majorelle
4 **Musée des Beaux-Arts**, 3 Place Stanislas
5 **Musée Ecole de Nancy**, 36-38 Rue Sergent Blandan
6 **Musée Lorrain**, 64 Grande Rue
7 **Gare SNCF**, 3 Place Thiers

REGISTRATION

Opening hours

Sunday, Sept 14, 2014	16:00 – 18:00
Monday, Sept 15, 2014	08:30 – 19:30
Tuesday, Sept 16, 2014	08:30 – 18:30
Wednesday, Sept 17, 2014	08:30 – 18:00
Thursday, Sept 18, 2014	08:30 – 18:30
Friday, Sept 19, 2014	08:30 – 13:00

Registration includes

- Attendance of all workshops and tutorials on Monday and Friday.
- Attendance of the technical program on Tuesday, Wednesday and Thursday.
- Attendance of the social events:
  - a welcome reception (Monday),
  - two catered evening poster sessions (Tuesday and Thursday),
  - a guided tour and a conference banquet (Wednesday).

Registration excludes

- Hotel booking.
- Breakfasts, lunches and dinners on Monday and Friday.

Accompanying Person´s Fee Includes

- Welcome reception on Monday, September 15th.
- Guided tour and a conference banquet (Wednesday, September 17, 2014).

INSTRUCTIONS FOR PRESENTERS

Presentation types

- **Main technical track** (both proceedings and journal papers): oral presentation and a poster.
- **Nectar and industrial tracks, plenary talks, tutorials**: only oral presentation.
- **Demo track**: oral spotlight presentation and live demo.
- **Workshops**: oral presentation OR poster + oral spotlight presentation (The authors have been informed about their mode of presentation.).

Oral Presentations

- Presentation time:
  - **Journal** and proceedings papers in the main technical track: **20 min**. incl. 5 min. for discussion,
  - Nectar track presentations: **30 min**. incl. 5 min. for discussion,
  - **Demo** spotlights: **9 min,**
  - **Workshop** papers: per workshop instructions.
- A computer and a **data projector** will be available. Please bring your presentation on **a USB memory stick** or on a **CD**. Accepted file formats are MS-Power Point **PPT** and Adobe **PDF**.
- Please **upload** your presentation in the lecture room where your talk will take place **10 minutes** before the start of your session. **Exception**: for **Friday 10:45 workshops**, please submit your presentation at the Speakers Ready Corner (see below) during the coffee break preceding these workshops.
- When your session is over, your presentation will be deleted from all computers, no copies or backups will be made.
- Speakers can use their **own computers** for the presentation. This option is not recommended if the presentation is a standard PPT or PDF file. An HDMI cable and a 230V power outlet will be available. The setup should be tested in the session room in one of the breaks before the session as early as possible.
- The **Speakers Ready Corner** will be available for speakers requiring assistance with format conversions, file upload etc. The operation hours are from 8:00 AM to 8:00 PM (Mon) / 6:00 PM (Tue, Wed, Thu) / 7:00 PM (Fri).
- Special **PowerPoint** considerations:
  - Please use one of these versions: PP 97-2003 and 95 or 2007, 2010 and save your presentation as a PPT rather than PPS. All videos or animations in the presentation must run automatically.
  - **Fonts**: Only fonts that are included in the basic installation of MS-Windows will be available (English version of Windows). Other fonts can cause a wrong layout of your presentation. The suggested fonts are Arial, Times New Roman. If you insist on using different fonts, these must be embedded into your presentation by choosing the right option when saving your presentation: Click on "File", then "Save As", check the "Tools" menu and select "Embed True Type Fonts."
- **Pictures and Videos**
  - JPG is the preferred file format for images. GIF, TIF or BMP formats will be accepted as well.
  - Images inserted into PowerPoint are embedded into the presentations. Images that are created at a dpi setting higher than 200 dpi are not necessary and will only increase the file size of your presentation.
  - We cannot provide support for embedded videos in your presentation; please test your presentation with the on-site PC several hours before your presentation. Generally, the WMV format should work with no difficulties.
  - In case your video is not embedded in the presentation it is possible to have it in other formats: MPEG 2,4, AVI (code: DivX, XviD, h264) or WMV. Suggested bitrate for all mpeg4 based code is about 1Mbps with SD PAL resolution (1024 x 576pix with square pixels, AR: 16/9).
  - In case of Full HD videos, please let us know before the meeting so that they can be tested.
  - Videos that require additional reading or projection equipment (e.g., VHS) will be not accepted.

POSTERS

Poster Area

- Posters of the **main technical track** will be displayed in the **Grand Foyer** of the congress centre. From Tuesday to Thursday, it will be accessible all day long and besides poster visits, it can be used for meetings and discussions.

Mounting and Removal Times

- Tuesday's main track papers: mount after Tuesday 8:30 AM, remove after Tuesday's poster session.
- Thursday's main track papers: mount after Thursday 8:30 AM, remove after Thursday's poster session.

Format and Mounting

- All posters of the **main technical track** will be mounted on poster **stands** in the poster room.
- The **maximum size** of the poster is 180 cm (height, 71 inches) x 97 cm (width, 38 inches). The **recommended** size is 100 cm (39 inches) x 95 cm (37 inches).
- Fixing material (pins and stickers) will be available. For wall mounting, only stickers can be used.
- Congress staff will be available to assist you during the time of poster mounting.

LIVE DEMOS

- Live demos will be shown during the Tuesday's poster session using the authors' own computers. Tables and 230V power outlets will be available. Please contact the local chair as soon as possible if you need additional assistance.
- A demo may be accompanied by a poster on a poster stand next to the demo table. Demo presenters interested in displaying a poster should write a request to the demo chair as soon as possible. The poster conditions described above apply here as well, with mounting and removal times same as for Tuesday's main track papers.

GENERAL INFORMATION

Currency

The official currency of the France is the Euro (€).
International credit cards are accepted for payments in most hotels, restaurants and shops. ATM machines are easily available throughout the city.

Wi-Fi

Wi-Fi will be available at the Congress Centre Prouvé during the whole conference.

Cloakroom

A cloakroom is located in the hall– see the floor plan. Opening hours correspond with the opening hours of the Registration Desk **(clothing and luggage only)**.

Insurance

The Organizers of the Conference do not accept liability for any injury, loss or damage, arising from accidents or other situations during, or as a consequence of the Conference. Participants are therefore advised to arrange insurance for health and accident prior to travelling to the Conference.

Language

The official language of the Conference is English. Simultaneous interpretation is not provided.

Lunches

ECML/PKDD 2014 does not organize lunches for the participants.
A list of restaurants in the city of Nancy will be distributed at the beginning of the conference.

Message System

The Message Board is located in the Registration Area. There you may leave a message for your friends or colleagues.

Mobile Phones

Delegates are kindly requested to switch off their mobile phones during the sessions.

Program Changes

The organizers cannot assume liability for any changes in the program due to external or unforeseen circumstances.

Smoking Policy

Smoking is not allowed inside the building and in all public places.

Staff

Conference staff will be happy to assist to participants during the Conference.

GOOD TO KNOW

Electricity

France uses a 230 volt 50 Hz system.

Shopping

Most shops in Nancy are open from 9:30 to 19:00, Monday through Saturday.

SOCIAL EVENTS

Welcome Reception

Where: Grands salons, Hôtel de ville of Nancy (Place Stanislas)
When: 20:00 – 21:40 – Monday, September 15th, 2014

Conference Dinner

Where: Espace Reception, Centre Prouvé Congress Center
When: 20:00 – 24:00 – Wednesday, September 17th, 2014

The Gala dinner ticket is mandatory to access to the dinner.

Catered evening poster sessions

Where: Grand Foyer, Centre Prouvé Congress Center
When: 19:30 – 21:30 – Tuesday, September 16th, 2014.
19:30 – 21:30 – Thursday, September 18th, 2014.

GUIDED TOUR

When: 18:00 – 20:00 – Wednesday, September 17th 2014.

Registration at the desk on Monday.

KEYNOTE SPEAKERS

INVITED TALK

Lise Getoor

Scalable Collective Reasoning using Probabilistic Soft Logic

Abstract

One of the challenges in big data analytics is to efficiently learn and reason collectively about extremely large, heterogeneous, incomplete, noisy interlinked data. Collective reasoning requires the ability to exploit both the logical and relational structure in the data and the probabilistic dependencies. In this talk I will overview our recent work on probabilistic soft logic (PSL), a framework for collective, probabilistic reasoning in relational domains. PSL is able to reason holistically about both entity attributes and relationships among the entities. The underlying mathematical framework, which we refer to as a hinge-loss Markov random field, supports extremely efficient, exact inference. This family of graphical models captures logic-like dependencies with convex hinge-loss potentials. I will survey applications of PSL to diverse problems ranging from information extraction to computational social science. Our recent results show that by building on state-of-the-art optimization methods in a distributed implementation, we can solve large-scale problems with millions of random variables orders of magnitude faster than existing approaches.

Bio

In 1995, Lise Getoor decided to return to school to get her PhD in Computer Science at Stanford University. She received a National Physical Sciences Consortium fellowship, which in addition to supporting her for six years, supported a summer internship at Xerox PARC, where she worked with Markus Fromherz and his group. Daphne Koller was her PhD advisor; in addition, she worked closely with Nir Friedman, and many other members of the DAGS group, including Avi Pfeffer, Mehran Sahami, Ben Taskar, Carlos Guestrin,Uri Lerner,Ron Parr, Eran Segal, Simon Tong.
In 2001, Lise Getoor joined the Computer Science Department at the University of Maryland, College Park.

Raymong NG

Big Data for Personalized Medicine: a case study of Biomarker Discovery

Abstract

Personalized medicine has been hailed as one of the main frontiers for medical research in this century. In the first half of the talk, we will give an overview on our projects that use gene expression, proteomics, DNA and clinical features for biomarker discovery. In the second half of the talk, we will describe some of the challenges involved in biomarker discovery. One of the challenges is the lack of quality assessment tools for data generated by ever-evolving genomics platforms. We will conclude the talk by giving an overview of some of the techniques we have developed on data cleansing and pre-processing.

Bio

Dr. Raymond Ng is a professor in Computer Science at the University of British Columbia. His main research area for the past two decades is on data mining, with a specific focus on health informatics and text mining. He has published over 180 peer-reviewed publications on data clustering, outlier detection, OLAP processing, health informatics and text mining. He is the recipient of two best paper awards – from 2001 ACM SIGKDD conference, which is the premier data mining conference worldwide, and the 2005 ACM SIGMOD conference, which is one of the top database conferences worldwide. He was one of the program co-chairs of the 2009 International conference on Data Engineering, and one of the program co-chairs of the 2002 ACM SIGKDD conference. He was also one of the general co-chairs of the 2008 ACM SIGMOD conference.
For the past decade, Dr. Ng has co-led several large scale genomic projects, funded by Genome Canada, Genome BC and industrial collaborators. The total amount of funding of those projects well exceeded $40 million Canadian dollars. He now holds the Chief Informatics Officer position of the PROOF Centre of Excellence, which focuses on biomarker development for end-stage organ failures.

Francis Bach

Beyond stochastic gradient descent for large-scale machine learning

Abstract

Many machine learning and signal processing problems are traditionally cast as convex optimization problems. A common difficulty in solving these problems is the size of the data, where there are many observations ("large n") and each of these is large ("large p"). In this setting, online algorithms such as stochastic gradient descent which pass over the data only once, are usually preferred over batch algorithms, which require multiple passes over the data. In this talk, I will show how the smoothness of loss functions may be used to design novel algorithms with improved behavior, both in theory and practice: in the ideal infinite-data setting, an efficient novel Newton-based stochastic approximation algorithm leads to a convergence rate of O(1/n) without strong convexity assumptions, while in the practical finite-data setting, an appropriate combination of batch and online algorithms leads to unexpected behaviors, such as a linear convergence rate for strongly convex problems, with an iteration cost similar to stochastic gradient descent. (joint work with Nicolas Le Roux, Eric Moulines and Mark Schmidt).

Bio

Francis Bach is a researcher at INRIA, leading since 2011 the SIERRA project-team, which is part of the Computer Science Laboratory at Ecole Normale Superieure. He completed his Ph.D. in Computer Science at U.C. Berkeley, working with Professor Michael Jordan, and spent two years in the Mathematical Morphology group at Ecole des Mines de Paris, then he joined the WILLOW project-team at INRIA/Ecole Normale Superieure from 2007 to 2010. Francis Bach is interested in statistical machine learning, and especially in graphical models, sparse methods, kernel-based learning, convex optimization vision and signal processing.

Tie-Yan Liu

Machine Learning for Search Ranking and Ad Auction

Abstract

In the era of information explosion, search has become an important tool for people to retrieve useful information. Every day, billions of search queries are submitted to commercial search engines. In response to a query, search engines return a list of relevant documents according to a ranking model. In addition, they also return some ads to users, and extract revenue by running an auction among advertisers if users click on these ads. This "search + ads" paradigm has become a key business model in today's Internet industry, and has incubated a few hundred-billion-dollar companies. Recently, machine learning has been widely adopted in search and advertising, mainly due to the availability of huge amount of interaction data between users, advertisers, and search engines. In this talk, we discuss how to use machine learning to build effective ranking models (which we call learning to rank) and to optimize auction mechanisms. (i) The difficulty of learning to rank lies in the interdependency between documents in the ranked list. To tackle it, we propose the so-called listwise ranking algorithms, whose loss functions are defined on the permutations of documents, instead of individual documents or document pairs. We prove the effectiveness of these algorithms by analyzing their generalization ability and statistical consistency, based on the assumption of a two-layer probabilistic sampling procedure for queries and documents, and the characterization of the relationship between their loss functions and the evaluation measures used by search engines (e.g., NDCG and MAP). (ii) The difficulty of learning the optimal auction mechanism lies in that advertisers' behavior data are strategically generated in response to the auction mechanism, but not randomly sampled in an i.i.d. manner. To tackle this challenge, we propose a game-theoretic learning method, which first models the strategic behaviors of advertisers, and then optimizes the auction mechanism by assuming the advertisers to respond to new auction mechanisms according to the learned behavior model. We prove the effectiveness of the proposed method by analyzing the generalization bounds for both behavior learning and auction mechanism learning based on a novel Markov framework.

Bio

Tie-Yan Liu is a senior researcher and research manager at Microsoft Research. His research interests include machine learning (learning to rank, online learning, statistical learning theory, and deep learning), algorithmic game theory, and computational economics. He is well known for his work on learning to rank for information retrieval. He has authored the first book in this area, and published tens of highly-cited papers on both algorithms and theorems of learning to rank. He has also published extensively on other related topics. In particular, his paper won the best student paper award of SIGIR (2008), and the most cited paper award of the Journal of Visual Communication and Image Representation (2004-2006); his group won the research break-through award of Microsoft Research Asia (2012). Tie-Yan is very active in serving the research community. He is a program committee co-chair of ACML (2015), WINE (2014), AIRS (2013), and RIAO (2010), a local co-chair of ICML 2014, a tutorial co-chair of WWW 2014, a demo/exhibit co-chair of KDD (2012), and an area/track chair of many conferences including ACML (2014), SIGIR (2008-2011), AIRS (2009-2011), and WWW (2011). He is an associate editor of ACM Transactions on Information System (TOIS), an editorial board member of Information Retrieval Journal and Foundations and Trends in Information Retrieval. He has given keynote speeches at CCML (2013), CCIR (2011), and PCM (2010), and tutorials at SIGIR (2008, 2010, 2012), WWW (2008, 2009, 2011), and KDD (2012). He is a senior member of the IEEE and the ACM.

Charu Aggarwal

Network Analysis in the Big Data Age: Mining Graph and Social Streams

Abstract

The advent of large interaction-based communication and social networks has led to challenging streaming scenarios in graph and social stream analysis. The graphs that result from such interactions are large, transient, and very often cannot even be stored on disk. In such cases, even simple frequency-based aggregation operations become challenging, whereas traditional mining operations are far more complex. When the graph cannot be explicitly stored on disk, mining algorithms must work with a limited knowledge of the network structure. Social streams add yet another layer of complexity, wherein the streaming content associated with the nodes and edges needs to be incorporated into the mining process. A significant gap exists between the problems that need to be solved, and the techniques that are available for streaming graph analysis. In spite of these challenges, recent years have seen some advances in which carefully chosen synopses of the graph and social streams are leveraged for approximate analysis. This talk will focus on several recent advances in this direction.

Bio

Charu Aggarwal is a Research Scientist at the IBM T. J. Watson Research Center in Yorktown Heights, New York. He completed his B.S. from IIT Kanpur in 1993 and his Ph.D. from Massachusetts Institute of Technology in 1996. His research interest during his Ph.D. years was in combinatorial optimization (network flow algorithms), and his thesis advisor was Professor James B. Orlin. He has since worked in the field of data mining, with particular interests in data streams, privacy, uncertain data and social network analysis. He has published over 200 papers in refereed venues, and has applied for or been granted over 80 patents. Because of the commercial value of the above-mentioned patents, he has received several invention achievement awards and has thrice been designated a Master Inventor at IBM. He is a recipient of an IBM Corporate Award (2003) for his work on bio-terrorist threat detection in data streams, a recipient of the IBM Outstanding Innovation Award (2008) for his scientific contributions to privacy technology, and a recipient of an IBM Research Division Award (2008) for his scientific contributions to data stream research. He has served on the program committees of most major database/data mining conferences, and served as program vice-chairs of the SIAM Conference on Data Mining, 2007, the IEEE ICDM Conference, 2007, the WWW Conference 2009, and the IEEE ICDM Conference, 2009. He served as an associate editor of the IEEE Transactions on Knowledge and Data Engineering Journal from 2004 to 2008. He is an associate editor of the ACM TKDD Journal, an action editor of the Data Mining and Knowledge Discovery Journal, an associate editor of the ACM SIGKDD Explorations, and an associate editor of the Knowledge and Information Systems Journal. He is a fellow of the ACM (2013) and the IEEE (2010) for "contributions to knowledge discovery and data mining techniques".

INDUSTRY INVITED TALK

Arthur von Eschen

Machine Learning and Data Mining in Call of Duty

Abstract

Data science is relatively new to the video game industry, but it has quickly emerged as one of the main resources for ensuring game quality. At Activision, we leverage data science to analyze the behavior of our games and our players to improve in-game algorithms and the player experience. We use machine learning and data mining techniques to influence creative decisions and help inform the game design process. We also build analytic services that support the game in real-time; one example is a cheating detection system which is very similar to fraud detection systems used for credit cards and insurance. This talk will focus on our data science work for Call of Duty, one of the bestselling video games in the world.

Bio

Arthur Von Eschen is Senior Director of Game Analytics at Activision. He and his team are responsible for analytics work that supports video game design on franchises such as Call of Duty and Skylanders. In addition to holding a PhD in Operations Research, Arthur has over 15 years of experience in analytics consulting and R&D with the U.S. Fortune 500. His work has spanned across industries such as banking, financial services, insurance, retail, CPG and now interactive entertainment (video games). Prior to Activision he worked at Fair Isaac Corporation (FICO). Before FICO he ran his own analytics consulting firm for six years.

Georges Hébrail

Making smart metering smarter by applying data analytics

Abstract

New data is being collected from electric smart meters which are deployed in many countries. Electric power meters measure and transmit to a central information system electric power consumption from every individual household or enterprise. The sampling rate may vary from 10 minutes to 24 hours and the latency to reach the central information system may vary from a few minutes to 24h. This generates a large amount of – possibly streaming – data if we consider customers from an entire country (ex. 35 millions in France). This data is collected firstly for billing purposes but can be processed with data analytics tools with several other goals. The first part of the talk will recall the structure of electric power smart metering data and review the different applications which are considered today for applying data analytics to such data. In a second part of the talk, we will focus on a specific problem: spatio-temporal estimation of aggregated electric power consumption from incomplete metering data.

Bio

Georges Hébrail is a senior researcher at EDF Lab, the research centre of Electricité de France, one of the world's leading electric utility. His background is in Business Intelligence covering many aspects from data storage and querying to data analytics. From 2002 to 2010, he was a professor of computer science at Telecom ParisTech, teaching and doing research in the field of information systems and business intelligence, with a focus on time series management, stream processing and mining. His current research interest is on distributed and privacy-preserving data mining on electric power related data.

Alexandre Cotarmanac'h

Ads that matter

Abstract

The advent of realtime bidding and online ad-exchanges has created a new and fast-growing competitive marketplace. In this new setting, media-buyers can make fine-grained decisions for each of the impressions being auctioned taking into account information from the context, the user and his/her past behavior. This new landscape is particularly interesting for online e-commerce players where user actions can also be measured online and thus allow for a complete measure of return on ad-spend.
Despite those benefits, new challenges need to be addressed such as :

- the design of a realtime bidding architecture handling high volumes of queries at low latencies,
- the exploration of a sparse and volatile high-dimensional space
- as well as several statistical modeling problems (e.g. pricing, offer and creative selection).

In this talk, I will present an approach to realtime media buying for online e-commerce from our experience working in the field. I will review the aforementioned challenges and discuss open problems for serving ads that matter.

Bio

Alexandre Cotarmanac'h is Vice-President Distribution & Platform for Twenga.
Twenga is a services and solutions provider generating high value-added leads to online merchants that was founded in 2006.
Originally hired to help launch Twenga's second generation search engine and to manage the optimization of revenue, he launched in 2011 the affinitAD line of business and Twenga's publisher network. Thanks to the advanced contextual analysis which allows for targeting the right audience according to their desire to buy e-commerce goods whilst keeping in line with the content offered, affinitAD brings Twenga's e-commerce expertise to web publishers. Alexandre also oversees Twenga's merchant program and strives to offer Twenga's merchants new services and solutions to improve their acquisition of customers.

Mike Bodkin

Algorithms, Evolution and Network-Based Approaches in Molecular Discovery

Abstract

Drug research generates huge quantities of data around targets, compounds and their effects. Network modelling can be used to describe such relationships with the aim to couple our understanding of disease networks with the changes in small molecule properties. This talk will build off of the data that is routinely captured in drug discovery and describe the methods and tools that we have developed for compound design using predictive modelling, evolutionary algorithms and network-based mining.

Bio

Mike did his PhD in protein de-novo design for Nobel laureate sir James Black before taking up a fellowship in computational drug design at Cambridge University. He moved to AstraZeneca as a computational chemist before joining Eli Lilly in 2000. Since 2003 he was head of the computational drug discovery group at Lilly but recently jumped ship to Evotec to work as the VP for computational chemistry and cheminformatics. His research aims are to continue to develop new algorithms and software in the fields of drug discovery and systems informatics and to deliver and apply current and novel methods as tools for use in drug research.

ECML/PKDD PROGRAM AT A GLANCE

Monday 15/9

09:00 – 12:40
- **Workshops**: DyNaK II; DMNLP; MUSE; SSDM; SensML 2014; MTP; RL 2014
- **Discovery challenge**: C1: Neural Connectomics Workshop - From Imaging to Connectivity
- **Tutorials**: T1: Medical Mining for Clinical Knowledge Discovery; T2: Patterns in Noisy and Multidimensional Relations and Graphs

12:40 – 14:00 Lunch

14:00 – 17:30
- **Workshops**: DyNaK II; DMNLP; MUSE; SSDM; SensML 2014; MTP
- **Discovery challenge**: C1: Neural Connectomics Workshop - From Imaging to Connectivity
- **Tutorials**: T3: The Pervasiveness of Machine Learning in Omics Science; T4: Conformal Predictions for Reliable Machine Learning

18:00 – 18:30 **Opening & Awards** (Auditorium 850)

18:30 – 19:30 **Invited talk**
Lise Getoor
(Auditorium 850)

20:00 – 21:40 **Opening reception**

Tuesday 16/9

09:00 – 10:00 **Invited talk**
Raymond T. Ng
(Auditorium 850)

10:00 – 10:30 **Test of time award talk** (Auditorium 850)

10:30 – 11:00 *Coffee break*

11:00 – 12:40
- **Joint session with S-ILP** (101)
- S1 : Networks (103-104)
- S2 : Data projection and dimensionality reduction (102)
- S3 : Precision and recall in classification (105)

12:40 – 14:00 Lunch

14:00 – 15:40
- **ILP community meeting** (101)
- S4 : Kernel-based learning and structured outputs (102)
- S5 : Classification (105)
- S6 : Feature selection and construction (106)
- S7 : Clustering (103-104)

15:40 – 16:10 *Coffee break*

16:10 – 17:50
- S8 : Decomposition and latent variables (103-104)
- S9 : Multi-task and multi-label learning (102)
- S10 : Applications and social data mining (105)
- S11 : Pattern mining (106)

17:50 – 18:10 **Demos spotlight** (13 demos) (103-104)

19:30 – 21:30
- **Poster session**
Papers in S1 - S11 + S-ILP will be presented
(Grand Foyer)
- **Demo track** (Grand Foyer)

Wednesday 17/9

09:00 – 10:00 **Invited talk**
Francis Bach
(Auditorium 850)

10:10 – 11:10
- **Nectar** NS1 (102)
- S12 : Optimization and approximation (105)
- S13 : Kernel-based learning (106)
- S14 : Community detection (103-104)

11:10 – 11:40 *Coffee break*

11:40 – 13:00
- S15 : Data factorization (102)
- S16 : Bandits (103-104)
- S17 : Spatial and temporal data (105)
- S18 : Text mining (106)

13:00 – 14:20 Lunch

14:20 – 15:50
- **Nectar** NS2 (103-104)
- **Industry invited talks**
Arthur von Eschen
Mike Bodkin
(Auditorium 850)

15:50 – 16:20 *Coffee break*

16:20 – 17:40
- S19 : Neural structures and deep learning (105-106)
- S20 : Reinforcement learning (102)
- S21 : Recommendation systems and dyadic data (103-104)
- S22 : Data stream mining (101)

18:00 – 20:00 **City visit**

20:00 – 21:30 **Conference dinner**
(Espace Reception)

Thursday 18/9

09:00 – 10:00 **Invited talk**
Tie-Yan Liu
(Auditorium 850)

10:10 – 11:10
- **Nectar** NS3 (103-104)
- S23 : Classifier evaluation (102)
- S24 : Data Mining tools and frameworks (105)
- S25 : Spectral learning (106)

11:10 – 11:40 *Coffee break*

11:40 – 13:00
- S26 : Neural networks and deep learning (103-104)
- S27 : Partially and semi-supervised learning (102)
- S28 : Reliable prediction (105)
- S29 : Multi-target and transfer learning (106)

13:00 – 14:20 Lunch

14:20 – 15:50
- **Nectar** NS4 (103-104)
- **Industry invited talks**
Georges Hébrail
Alexandre Cotarmanac'h
(Auditorium 850)

15:50 – 16:20 *Coffee break*

16:20 – 17:40
- S30 : Support vector machines (102)
- S31 : Privacy and anti-discrimination in data mining (105)
- S32 : Probabilistic and Bayesian methods (106)
- S33 : Time-evolving graphs and Dynamic Networks (103-104)

18:10 – 19:10 **Community Meeting** (Auditorium 850)

19:30 – 21:30
- **Poster session**
Papers in S12 - S33 will be presented
(Grand Foyer)

Friday 19/9

09:00 – 10:00 **Invited talk**
Charu Aggarwal
(Auditorium 300)

10:00 – 10:15 *Coffee break*

10:15 – 13:00
- **Workshops**: DARE'14; LD4KD; NFmcp 2014; LEMA 2014; LMCE 2014
- **Tutorials**: T5: The Lunch is Never Free: How Information Theory, MDL, and Statistics are Connected; T9: Deep Learning
- **PhD Session**
- **Discovery challenge**: C2: Predictive Web Analytics

13:00 – 14:00 Lunch

14:00 – 18:30
- **Workshops**: DARE'14; LD4KD; NFmcp 2014; LEMA 2014; LMCE 2014
- **Tutorials**: T6: Information theoretic Methods in Data Mining; T7: Machine Learning with Analogical Proportions; T8: Preference Learning Problems
- **PhD Session**

Monday 15/9

Room:	102	103	104	202	203	204	205	105	106	206
09:00	SensML 2014	MUSE	DyNaK II	RL 2014	DMNLP	SSDM	MTP	Tutorial **T1**	Tutorial **T2**	Discovery Challenge **C1**
12:40	Lunch									
14:00	SensML 2014	MUSE	DyNaK II		DMNLP	SSDM	MTP	Tutorial **T3**	Tutorial **T4**	Discovery Challenge **C1**
17:30										
18:00	**Opening & Awards** (Auditorium 850)									
18:30	**Invited talk** (Auditorium 850) Lise Getoor									
19:30										
20:00	**Opening reception**									
21:40										

T1 *: Medical Mining for Clinical Knowledge Discovery*
T2 *: Patterns in Noisy and Multidimensional Relations and Graphs*
C1 *: Neural Connectomics Workshop – From Imaging to Connectivity*
T3 *: The Pervasiveness of Machine Learning in Omics Science*
T4 *: Conformal Predictions for Reliable Machine Learning*

Friday 19/9

Room:	101	102	202	203	205	103-104	105	106	204
09:00	**Invited talk** (Auditorium 300) Charu Aggarwal								
10:00	Coffee break								
10:15	DARE'14	LEMA 2014	LMCE 2014	NFmcp 2014	LD4KD	Tutorial **T9**	Tutorial **T5**	Discovery Challenge **C2**	PhD Session
12:40	Lunch								
14:00	DARE'14	LEMA 2014	LMCE 2014	NFmcp 2014	LD4KD	Tutorial **T8**	Tutorial **T6**	Tutorial **T7**	PhD Session
18:30									

T5 *: The Lunch is Never Free: How Information Theory, MDL, and Statistics are Connected*
T6 *: Information Theoretic Methods in Data Mining*
T7 *: Machine Learning with Analogical Proportions*
T8 *: Preference Learning Problems*
T9 *: Deep Learning*
C2 *: Predictive Web Analytics*

MONDAY 15 SEPTEMBER 2014

MONDAY INVITED TALK

Scalable Collective Reasoning using Probabilistic Soft Logic

Speaker: Lise Getoor
Time: 18:30 – 19:30
Room: Auditorium 850

Abstract
One of the challenges in big data analytics is to efficiently learn and reason collectively about extremely large, heterogeneous, incomplete, noisy interlinked data. Collective reasoning requires the ability to exploit both the logical and relational structure in the data and the probabilistic dependencies. In this talk I will overview our recent work on probabilistic soft logic (PSL), a framework for collective, probabilistic reasoning in relational domains. PSL is able to reason holistically about both entity attributes and relationships among the entities. The underlying mathematical framework, which we refer to as a hinge-loss Markov random field, supports extremely efficient, exact inference. This family of graphical models captures logic-like dependencies with convex hinge-loss potentials. I will survey applications of PSL to diverse problems ranging from information extraction to computational social science. Our recent results show that by building on state-of-the-art optimization methods in a distributed implementation, we can solve large-scale problems with millions of random variables orders of magnitude faster than existing approaches.

Bio
In 1995, Lise Getoor decided to return to school to get her PhD in Computer Science at Stanford University. She received a National Physical Sciences Consortium fellowship, which in addition to supporting her for six years, supported a summer internship at Xerox PARC, where she worked with Markus Fromherz and his group. Daphne Koller was her PhD advisor; in addition, she worked closely with Nir Friedman, and many other members of the DAGS group, including Avi Pfeffer, Mehran Sahami, Ben Taskar, Carlos Guestrin,Uri Lerner,Ron Parr, Eran Segal, Simon Tong. In 2001, Lise Getoor joined the Computer Science Department at the University of Maryland, College Park.

MONDAY WORKSHOPS

DyNaK II: Dynamic Networks and Knowledge Discovery

Rushed Kanawati, Ruggero G. Pensa, Céline Rouveirol
Room: 104

Modeling and analyzing networks is a major emerging topic in different research areas, such as computational biology, social science, document retrieval, etc. Nowadays, the scientific communities have access to huge volumes of network-structured data, such as social networks, gene/proteins/metabolic networks, sensor networks, peer-to-peer networks. Most often, these data are not only static, but they are collected at different time points. This dynamic view of the system allows the time component to play a key role in the comprehension of the evolutionary behavior of the network.

Handling such data is a major challenge for current research in machine learning and data mining, and it has led to the development of recent innovative techniques that consider complex/multi-level networks, time-evolving graphs, heterogeneous information (nodes and links), and requires scalable algorithms that are able to manage huge and complex networks.

DyNaK workshop is motivated by the interest of providing a meeting point for scientists with different backgrounds that are interested in the study of large complex networks and the dynamic aspects of such networks. It aims at attracting contributions from both aspects of networks analysis: large real network analysis and modelling, and knowledge discovery within those networks.

More information: http://lipn.univ-paris13.fr/dynak2/DyNakII/

DMNLP: Interactions between Data Mining and Natural Language Processing

Peggy Cellier, Thierry Charnois, Andreas Hotho, Stan Matwin, Marie-Francine Moens, Yannick Toussaint
Room: 203

On the one hand, in the field of Natural Language Processing (NLP), numerical Machine Learning methods (e.g., SVM, CRF) have been intensively explored and applied. Despite the good results obtained by the numerical methods, one major drawback is that they do not provide a human readable model. A promising direction is the integration of symbolic knowledge. On the other hand, research in Data Mining has progressed significantly in the last decades, through the development of advanced algorithms and techniques to extract knowledge from data in different forms. In particular, for two decades Pattern Mining has been one of the most active field in Knowledge Discovery.

Recently, a new field has emerged taking benefit of both domains: Data Mining and NLP. The objective of DMNLP is thus to provide a forum to discuss how Data Mining can be interesting for NLP tasks, providing symbolic knowledge, but also how NLP can enhance data mining approaches by providing richer and/or more complex information to mine and by integrating linguistics knowledge directly in the mining process.

The workshop aims at bringing together researchers from both communities in order to stimulate discussions about the cross-fertilization of those two research fields. The idea of this workshop is to discuss future directions and new challenges emerging from the cross-fertilization of Data Mining and NLP and in the same time initiate collaborations between researchers of both communities.

More information: http://dmnlp.loria.fr/

MUSE: Mining Ubiquitous and Social Environments

Martin Atzmüller, Christoph Scholz
Room: 103

The emergence of ubiquitous computing has started to create new environments consisting of small, heterogeneous, and distributed devices that foster the social interaction of users in several dimensions. Similarly, the upcoming social web also integrates the user interactions in social networking environments.

In typical ubiquitous settings, the mining system can be implemented inside the small devices and sometimes on central servers, for real-time applications, similar to common mining approaches. However, the characteristics of ubiquitous and social mining in general are quite different from the current mainstream data mining and machine learning. Unlike in traditional data mining scenarios, data does not emerge from a small number of (heterogeneous) data sources, but potentially from hundreds to millions of different sources. Often there is only minimal coordination and thus these sources can overlap or diverge in many possible ways. Steps into this new and exciting application area are the analysis of this new data, the adaptation of well known data mining and machine learning algorithms and finally the development of new algorithms.

Mining big data in ubiquitous and social environments is an emerging area of research focusing on advanced systems for data mining in such distributed and network-organized systems. Therefore, for this workshop, we aim to attract researchers from all over the world working in the field of data mining and machine learning with a special focus on analyzing big data in ubiquitous and social environments.

The goal of this workshop is to promote an interdisciplinary forum for researchers working in the fields of ubiquitous computing, mobile sensing, social web, Web 2.0, and social networks which are interested in utilizing data mining in a ubiquitous setting. The workshop seeks for contributions adopting state-of-the-art mining algorithms on ubiquitous social data. Papers combining aspects of the two fields are especially welcome. In short, we want to accelerate the process of identifying the power of advanced data mining operating on data collected in ubiquitous and social environments, as well as the process of advancing data mining through lessons learned in analyzing these new data.

More information: http://www.kde.cs.uni-kassel.de/ws/muse2014/

SSDM: Statistically Sound Data Mining

Wilhelmiina Hämäläinen, François Petitjean, Geoff Webb
Room: 204

Even if Data Mining has its roots in Statistics, there was a long while when data miners and statisticians walked their own paths. Data miners concentrated on developing efficient algorithms that addressed the practical issues associated with huge data sets, but in doing so may sometimes have paid less attention to the reliability of patterns or even their utility. On the other hand, statisticians continued on their traditional line offering well-founded and sound methods for validating statistically meaningful patterns, but they could not offer computational means to find them. Fortunately, the situation is now changing and both data miners and statisticians are recognizing the need for cooperation.

The main impetus for this new trend is coming from a third party, the application fields. In the computerized world, it is easy to collect large data sets but their analysis is more difficult. Knowing the traditional statistical tests is no more sufficient for scientists, because one should first find the most promising hidden patterns and models to be tested. This means that there is an urgent need for efficient data mining algorithms which are able to find desired patterns, without missing any significant discoveries or producing too many spurious ones. A related problem is to find a statistically justified compromise between underfitted (too generic to catch all important aspects) and overfitted (too specific, holding just due to chance) patterns. However, before any algorithms can be designed, one should first solve many principal problems, like how to define the statistical significance of desired patterns, how to evaluate overfitting, how to interprete the p-values when multiple patterns are tested, and so on. In addition, one should evaluate the existing data mining methods, alternative algorithms and goodness measures to see which of them produce statistically valid results.

As we can see, there are many important problems which should be worked together with people from Data mining, Machine learning, and Statistics as well as application fields. The goal of this workshop is to offer a meeting point for this discussion. We want bring together people from different backgrounds and schools of science, both theoretically and practically oriented, to specify problems, share solutions and brainstorm new ideas.

To encourage real workshopping of actual problems, the workshop is arranged in a novel way, containing an invited lecture and inspiring groupworks in addition to traditional presentations. This means that also the non-author participants can contribute to workshop results and submit a paper to the final proceedings afterwards. If you have relevant problems which you would like to be worked together in the workshop, please send them before the workshop.

More information: http://cs.joensuu.fi/pages/whamalai/SSDM/ssdm14.html

SenseML 2014: Machine Learning for Urban Sensor Data

Frederik Janssen, Immanuel Schweizer
Room: 102

As the focus in the Wireless Sensor Networks and Sensor Systems community is shifting from “How do we collect data?” to “What can we learn from the data and how do the models look like?” we want to bring researchers from this community and the Machine Learning community together. Working with sensor data, machine learning methods become more and more popular (e.g., at the ACM SenSys conference – the major conference in this area – in 2013 the First International Workshop on Sensing and Big Data Mining (SenseMine) took place).
As the applications for machine learning expand into other areas, the need for high-quality machine learning methods constantly grows. Additionally, there is a need for interpretable models as researchers want to grasp the models and get a sense of how the sensor information is combined in the model.

However, sensor data poses a number of unique challenges for machine learning. Ranging from missing values, unreliable measurements, missing calibration to high spatial diversity. Most challenges have not been addressed with a focus on real-world sensor data. It is our belief that a discussion will help foster new results in the intersection of both communities.

More information: https://www.tk.informatik.tu-darmstadt.de/en/senseml-2014/

MTP: Multi-Target Prediction

Willem Waegeman, Krzysztof Dembczynski, Tapio Pahikkala, Antti Airola
Room: 205

Traditional methods in machine learning and statistics provide data-driven models for predicting one-dimensional targets, such as binary outputs in classification and real-valued outputs in regression. multi-task learning In recent years, novel application domains have triggered fundamental research on more complicated problems where multi-target predictions are required. Such problems arise in diverse application domains, such as document categorization, tag recommendation of images, videos and music, information retrieval, medical decision making, drug discovery, marketing, biology, geographical information systems, etc.

According to a general definition, the targets in multi-target prediction problems might be characterized by diverse data types, such as binary, nominal, ordinal and real-valued variables, but also rankings and relational structures, representing different entities of interest. Moreover, they often exhibit specific relationships, in the sense of being structured as a tree-shaped hierarchy or a directed acyclic graph, or being characterized by mutual exclusion, parent-child and other types of relationships. Specific multi-target prediction problems have been studied in a variety of subfields of machine learning and statistics, such as multi-label classification (prediction of multiple binary targets), multivariate regression (prediction of multiple numerical targets), sequence learning (ordered targets of varying length), structured output prediction (targets with inherent structure), preference learning (prediction of a preference relation between multiple targets, as in label ranking), multi-task learning (prediction of multiple targets in different but related domains) and collective learning (prediction for dependent observations).

Despite their commonalities, work on solving problems in the above domains has typically been performed in isolation, without much interaction between the different sub-communities. Moreover, several of the problems have been studied in different communities under different names. Sometimes there is even terminological confusion within the same community. multi-task learning The main goal of the workshop is to present a unifying overview of the above-mentioned subfields of machine learning, by focusing on the simultaneous prediction of multiple, mutually

dependent output variables. In the different subfields of machine learning that cover multi-target prediction it has been acknowledged by many authors that it is important to explicitly model the dependencies between the predicted targets. As a result of numerous discussions with experts in the above-mentioned domains, we are convinced that existing solutions can be brought to a fruitful cross-fertilization.

Despite the encouraging progress that has been made in the last decade, the current understanding of multi-target learning tasks and methods remains shallow. Further communication and education on the fundamental insights for this type of problems is still required. To date, it remains unclear which of the numerous approaches recently proposed performs better and under what assumptions. Therefore, the workshop intends to cover an overview of existing methods, while focussing on cross-domain methodologies.

With the workshop we aim to attract both researchers that are already active in one of the above domains, as well as researchers with little or no prior experience in multi-target prediction. As such, we believe that the workshop will attract ECML attendees from diverse subfields of machine learning and with different background.

More information: http://www.kermit.ugent.be/multi-target-prediction/

RL 2014: Representation Learning

Thierry Artieres, Jun Yan, Jun Guo, Sheng Gao
Room: 202

Representation learning has developed at the crossroad of different disciplines and application domains. It has recently enjoyed enormous success in learning useful representations of data from various application areas such as vision, speech, audio, or natural language processing. It has developed as a research field by itself with several successful workshops at major machine learning conferences, sessions at the main machine learning conferences (e.g., 3 sessions on deep learning at ICML 2013 + related sessions on e.g. tensors or compressed sensing) and with the recent ICLR (International Conference on Learning Representations) whose first edition was in 2013.

We take here a broad view of this field and want to attract researchers concerned with statistical learning of representations, including matrix- and tensor-based latent factor models, probabilistic latent models, metric learning, graphical models and also recent techniques such as deep learning, feature learning, compositional models, and issues concerned with non-linear structured prediction models. The focus of this workshop will be on representation learning approaches, including deep learning, feature learning, metric learning, algebraic and probabilistic latent models, dictionary learning and other compositional models, to problems in real-world data mining. Papers on new models and learning algorithms that combine aspects of the two fields of representation learning and data mining are especially welcome. This one-day workshop will include a mixture of invited talks, and contributed presentations, which will cover a broad range of subjects pertinent to the workshop theme. Besides classical paper presentations, the call also includes demonstration for applications on these topics. We believe this workshop will accelerate the process of identifying the power of representation learning operating on semantic data.

More information: http://conference.bupt.edu.cn/rl2014/

MONDAY DISCOVERY CHALLENGE

C1: Neural Connectomics Workshop – From Imaging to Connectivity

Vincent Lemaire, Demian Battaglia, Isabelle Guyon, Jordi Soriano
Room: 206

Understanding the brain structure and some of its alterations caused by disease, is key to accompany research on the treatment of epilepsy and Alzheimer's disease and other neuropathologies, as well as gaining understanding of the general functioning of the brain and its learning capabilities. At the neural level, recovering the exact wiring of the brain (connectome) including nearly 100 billion neurons, having on average 7000 synaptic connections to other neurons, is a daunting task.

The goal of this workshop is to bring together researchers in machine learning and neuroscience to discuss progress and remaining challenges in this exciting and rapidly evolving field. We aim to attract machine learning and computer vision specialists interested in learning about a new problem, as well as computational neuroscientists who may be interested in modeling connectivity data. We will discuss also the results of the First ChaLearn Neural Connectomics Challenge.

More information: http://connectomics.chalearn.org/workshop

MONDAY MORNING TUTORIALS

T1: Medical Mining for Clinical Knowledge Discovery

Pedro Pereira Rodrigues, Myra Spiliopoulou, Ernestina Menasalvas
Room: 105

Medical data mining is a mature area of research, characterized by both simple and very elaborate methods, mostly dedicated to solving a concrete problem of disease diagnosis, disease description or success prediction for a treatment. Clinical knowledge discovery encompasses analysis of epidemiological data, and of clinical and administrative data on patients; clinical decision support builds upon findings on these data. We elaborate on how data mining can contribute to such findings, we enumerate challenges of model learning, data availability and data provenance, and identify challenges on Big Medical Data.

Outline

- Self-presentation of the Tutorialists and Overview of the Domain (all)
- Clinical and Administrative Data Mining – Pedro Pereira Rodrigues
- Knowledge Discovery from Volatile Epidemiological Data – Myra Spiliopoulou
- Knowledge Discovery Challenges on Big Medical Data – Ernestina Menasalvas
- Clinical Decision Support – Pedro Pereira Rodrigues
- Concluding Remarks

T2: Patterns in Noisy and Multidimensional Relations and Graphs

Wagner Meira, Loic Cerf
Room: 106

In this tutorial, we will consider generalizations of closed itemset mining toward n-ary relations and toward noise tolerance. Declarative aspects (in particular, how to define "noise") as well as procedural aspects (how to efficiently traverse the pattern space) will be discussed.
Both generalizations worsen two problems that already affect the discovery of itemsets: 1) the number of valid (but rarely relevant) patterns exponentially grows with the size of the dataset and 2) so does the time to extract them all. We will see how both problems can be solved through user-defined relevance constraints that are enforced during the pattern space traversal (search space pruning).
Finally, adaptations of those patterns to graphs will be discussed. More precisely, we will study cross-graph quasi-clique mining (quasi-cliques frequently occurring in a collection of graphs) and correlated pattern mining (quasi-cliques in a graph where every vertex is associated with multiple labels).

Outline

- Patterns in Multidimensional Relations and Graphs: Defining Them
  - Introduction: a bottom-up approach toward an inductive database system
  - From binary relations to n-ary relations, a natural generalization
  - From crisp relations to fuzzy relations, tolerating noise globally or per-element, absolutely or relatively
  - Constraints for readability, quality and efficiency
  - From n-ary relations to collections of graphs through the symmetry constraint
  - Patterns in vertex-multilabeled graphs
  - Conclusion: Summary
- Patterns in Multidimensional Relations and Graphs: Mining Them
  - Introduction: Constraints for both a greater expressiveness and a greater scalability
  - Mining the closed itemsets, a generic algorithm and its extension to one definition of fault-tolerance
  - Classes of constraints, definitions and enforcements
  - Mining patterns in fuzzy n-ary relations
  - Mining patterns in collections of graphs
  - Mining patterns in vertex-multilabeled graphs
  - Conclusion: Summary and perspectives

MONDAY AFTERNOON TUTORIALS

T3: The Pervasiveness of Machine Learning in Omics Science

Ronnie Alves, Claude Pasquier, Nicolas Pasquier
Room: 105

Biology has become an enormously data-rich subject. Data is generated in many flavors and follows particularities of the omics perspective adopted along experimental studies. For instance, genomics is the field of study dealing with genomes and it is mostly associated with the static view (the genes and where they are placed along the genome). The dynamic view is brought from the transcriptomics perspective, so the gene expression and its regulation. Finally, interactomics is usually associated to gene products, proteins, and their interactions. However it could also be seen as a huge graph network with layers of interaction integrating distinct omics perspectives. Omics science applications of unsupervised and/or supervised machine learning (ML) techniques abound in the literature. In this tutorial, we discuss machine learning on omics data, putting the emphasis on (i) mapping and (ii) learning omics patterns. We consider three main omics data: genomics, transcriptomics and interactomics. For each perspective, we first provide, the biological problem, the data mapping (from a biological problem to a machine learning problem), the core ML methods employed and its implementation in the R language.

Outline

- Introduction and overview of the omics science
- Machine learning in genomics data – foundations, methods and applications
- Machine learning in transcriptomics data – foundations, methods and applications
- Machine learning in interactomics data – foundations, methods and applications
- Outlook – summary and future challenges

T4: Conformal Predictions for Reliable Machine Learning

Vineeth N. Balasubramanian, Shen-Shyang Ho, Vladimir Vovk
Room: 106

Reliable estimation of confidence remains a significant challenge as learning algorithms proliferate into challenging real-world pattern recognition applications. The Conformal Predictions framework is a recent development in machine learning to associate reliable measures of confidence with results in classification and regression. This framework is founded on the principles of algorithmic randomness (closely related to Kolmogorov complexity), transductive inference and hypothesis testing, and has several desirable properties for potential use in various real-world applications. This theory is based on the relationship derived between transductive inference and the randomness deficiency of an i.i.d. (identically independently distributed) sequence of data instances. One of the desirable features of this framework is the calibration of the obtained confidence values in an online setting. While probability/confidence values generated by existing approaches can often be unreliable and difficult to interpret, the theory behind the CP framework guarantees that the confidence values obtained using this transductive inference framework manifest as the actual error frequencies in the online setting i.e. they are well-calibrated. Further, this framework can be applied across all existing classification and regression methods (such as neural networks, Support Vector Machines, k-Nearest Neighbors, ridge regression, etc), thus making it a very generalizable approach.
Over the last few years, there has been a growing interest in applying this framework to real-world problems such as clinical decision support, medical diagnosis, sea surveillance, network traffic classification, and face recognition. The promising results have generated in further extensions of the framework to problem settings beyond just classification or regression. The framework has now been extended towards newer settings such as active learning, model selection, feature selection, change detection, outlier detection, and anomaly detection.

Outline

- Expose the audience to the basic theory of the framework
- Demonstrate examples of how the framework can be applied in real world problems
- Provide sample adaptations of the framework to related machine learning problems such as active learning, anomaly detection, feature selection and model selection

TUESDAY 16 SEPTEMBER 2014

TUESDAY INVITED TALK

Big Data for Personalized Medicine: a case study of Biomarker Discovery

Speaker: Raymong NG
Time: 09:00 – 10:00
Room: Auditorium 850

Abstract
Personalized medicine has been hailed as one of the main frontiers for medical research in this century. In the first half of the talk, we will give an overview on our projects that use gene expression, proteomics, DNA and clinical features for biomarker discovery. In the second half of the talk, we will describe some of the challenges involved in biomarker discovery. One of the challenges is the lack of quality assessment tools for data generated by ever-evolving genomics platforms. We will conclude the talk by giving an overview of some of the techniques we have developed on data cleansing and pre-processing.

Bio
Dr. Raymond Ng is a professor in Computer Science at the University of British Columbia. His main research area for the past two decades is on data mining, with a specific focus on health informatics and text mining. He has published over 180 peer-reviewed publications on data clustering, outlier detection, OLAP processing, health informatics and text mining. He is the recipient of two best paper awards – from 2001 ACM SIGKDD conference, which is the premier data mining conference worldwide, and the 2005 ACM SIGMOD conference, which is one of the top database conferences worldwide. He was one of the program co-chairs of the 2009 International conference on Data Engineering, and one of the program co-chairs of the 2002 ACM SIGKDD conference. He was also one of the general co-chairs of the 2008 ACM SIGMOD conference.
For the past decade, Dr. Ng has co-led several large scale genomic projects, funded by Genome Canada, Genome BC and industrial collaborators. The total amount of funding of those projects well exceeded $40 million Canadian dollars. He now holds the Chief Informatics Officer position of the PROOF Centre of Excellence, which focuses on biomarker development for end-stage organ failures.

TEST OF TIME AWARD TALK

Time: 10:00 – 10:30
Room: Auditorium 850

The "Test-of-Time" award is given to a paper that was presented 10 years ago at ECML/PKDD which is now considered to have been the most influential of all papers from that period.

This year, senior program committee members voted for their candidate from a short-list of 11 papers from ECML/PKDD 2004. The winning paper will be announced during the opening ceremony.

TUESDAY SESSIONS AT A GLANCE

Session 1: Networks

Room: 103-104 **Chair: Jilles Vreeken**

11:00 – 11:20 **Link Prediction in Multi-modal Social Networks**
Panagiotis Symeonidis, Christos Perentis

11:20 – 11:40 **Density-Based Subspace Clustering in Heterogeneous Networks**
Brigitte Boden, Martin Ester, Thomas Seidl

11:40 – 12:00 **Interestingness-driven Diffusion Process Summarization in Dynamic Networks**
Qiang Qu, Siyuan Liu, Christian Jensen, Feida Zhu, Christos Faloutsos

12:00 – 12:20 **FLIP: Active Learning for Relational Network Classification**
Tanwistha Saha, Huzefa Rangwala, Carlotta Domeniconi

12:20 – 12:40 **Faster way to agony: discovering hierarchies in directed graphs**
Nikolaj Tatti

Session 2: Data projection and dimensionality reduction

Room: 102 **Chair: Raymond Ng**

11:00 – 11:20 **Flexible Shift-Invariant Locality and Globality Preserving Projections**
Feiping Nie, Xiao Cai, Heng Huang

11:20 – 11:40 **Anomaly Detection with Score Functions based on the Reconstruction Error of the Kernel PCA**
Laetitia Chapel, Chloé Friguet

11:40 – 12:00 **Learning Binary Codes with Bagging PCA**
Cong Leng, Jian Cheng, Ting Yuan, Hanqing Lu

12:00 – 12:20 **A Unified Framework for Probabilistic Component Analysis**
Mihalis Nicolaou, Stefanos Zafeiriou, Maja Pantic

12:20 – 12:40 **Interactive Knowledge-Based Kernel PCA**
Dino Oglic, Daniel Paurat, Thomas Gaertner

Session 3: Precision and recall in classification

Room: 105 **Chair: Krzysztof Dembczyński**

11:00 – 11:20 **Optimal Thresholding of Classifiers to Maximize F1 Measure**
Zachary Lipton, Charles Elkan, Balakrishnan Narayanaswamy

11:20 – 11:40 **Rate-constrained ranking and the rate-weighted AUC**
Louise Millard, Peter Flach, Julian Higgins

11:40 – 12:00 **Rate-oriented point-wise confidence bounds for ROC curves**
Louise Millard, Meelis Kull, Peter Flach

12:00 – 12:20 **The Bane of Skew: Uncertain Ranks and Unrepresentative Precision**
Thomas Lampert, Pierre Gançarski

12:20 – 12:40 **On the null distribution of the precision and recall curve**
Miguel Lopes, Gianluca Bontempi

S-ILP: Joint ILP Session

Room: 101 **Chair: Jesse Davis**

11:00 – 11:20 **Evidence-based Clustering for Scalable Inference in Markov Logic**
Deepak Venugopal, Vibhav Gogate

11:20 – 11:40 **Effective Blending of Two and Three-way Interactions for Modeling Multi-relational Data**
Alberto Garcia-Duran, Antoine Bordes, Nicolas Usunier

11:40 – 12:00 **Towards Automatic Feature Construction for Supervised Classification**
Marc Boullé

12:00 - 12:20 **Fast Learning of Relational Dependency Networks**
Oliver Schulte, Zhensong Qian, Arthur E. Kirkpatrick, Xiaoqian Yin, Yan Sun

12:20 - 12:40 **Complex aggregates over subsets of elements**
Celine Vens, Sofie Van Gassen, Tom Dhaene, Yvan Saeys

Session 4: Kernel-based learning and structured outputs

Room: 102 **Chair: Willem Waegeman**

14:00 – 14:20 **Attributed Graph Kernels Using the Jensen-Tsallis q-Differences**
Lu Bai, Luca Rossi, Horst Bunke, Edwin Hancock

14:20 – 14:40 **Relative Comparison Kernel Learning with Auxiliary Kernels**
Eric Heim, Hamed Valizadegan, Milos Hauskrecht

14:40 – 15:00 **Approximate Consistency: Towards Foundations of Approximate Kernel Selection**
Lizhong Ding, Shizhong Liao

15:00 – 15:20 **Kernel Principal Geodesic Analysis**
Suyash Awate, Yen-Yun Yu, Ross Whitaker

15:20 – 15:40 **FASOLE: Fast Algorithm for Structured Output LEarning**
Vojtech Franc

Session 5: Classification

Room: 105 **Chair: Charles Elkan**

14:00 – 14:20 **Robust Distributed Training of Linear Classifiers Based on Divergence Minimization Principle**
Junpei Komiyama, Hidekazu Oiwa, Hiroshi Nakagawa

14:20 – 14:40 **Learning A Priori Constrained Weighted Majority Votes**
Aurélien Bellet, Amaury Habrard, Emilie Morvant, Marc Sebban

14:40 – 15:00 **Integer Bayesian Network Classifiers**
Sebastian Tschiatschek, Karin Paul, Franz Pernkopf

15:00 – 15:20 **Randomized Operating Point Selection in Adversarial Classification**
Viliam Lisý, Robert Kessl, Tomas Pevny

15:20 – 15:40 **Separating Rule Refinement and Rule Selection Heuristics in Inductive Rule Learning**
Julius Stecher, Frederik Janssen, Johannes Fuernkranz

Session 6: Feature selection and construction

Room: 106 **Chair: Gianluca Bontempi**

14:00 – 14:20 **Deterministic Feature Selection for Regularized Least Squares Classification**
Saurabh Paul, Petros Drineas

14:20 – 14:40 **Automatic design of neuromarkers for OCD characterization**
Oscar Garcia-Hinde, Emilio Parrado-Hernandez, Vanessa Gomez-Verdejo, Manel Martinez-Ramon, Carles Soriano-Mas

14:40 – 15:00 **Unsupervised Feature Selection via Unified Trace Ratio Formulation and K-means Clustering (TRACK)**
De Wang, Feiping Nie, Heng Huang

15:00 – 15:20 **Covariate-Correlated Lasso for Feature Selection**
Bo Jiang, Chris Ding, Bin Luo

15:20 – 15:40 **Unsupervised Interaction-Preserving Discretization of Multivariate Data**
Hoang-Vu Nguyen, Emmanuel Müller, Jilles Vreeken, Klemens Bohm

Session 7: Clustering

Room: 103-104 **Chair: Christel Vrain**

14:00 – 14:20 **Clustering via Mode Seeking by Direct Estimation of the Gradient of a Log-Density**
Hiroaki Sasaki, Aapo Hyvarinen, Masashi Sugiyama

14:20 – 14:40 **Fast Gaussian Pairwise Constrained Spectral Clustering**
David Chatel, Marc Tommasi, Pascal Denis

14:40 – 15:00 **Ratio-based Multiple Kernel Clustering**
Grigorios Tzortzis, Aristidis Likas

15:00 – 15:20 **Boosted Mean Shift Clustering**
Yazhou Ren, Uday Kamath, Carlotta Domeniconi, Guoji Zhang

15:20 – 15:40 **FILTA: Better View Discovery from Collections of Clusterings via Filtering**
Yang Lei, Nguyen Xuan Vinh, Jeffrey Chan, James Bailey

Session 8: Decomposition and latent variables

Room: 103-104 **Chair: Marie-Francine Moens**

16:10 – 16:30 **Pushing-Down Tensor Decompositions over Unions to Promote Reuse of Materialized Decompositions**
Mijung Kim, K. Selcuk Candan

16:30 – 16:50 **Hierarchical Latent Tree Analysis for Topic Detection**
Tengfei LIU, Nevin Zhang, Peixian Chen

16:50 – 17:10 **Causal Clustering for 2-Factor Measurement Models**
Peter Spirtes, Erich Kummerfeld, Joseph Ramsey, Renjie Yang, Richard Scheines

17:10 – 17:30 **On learning matrices with orthogonal columns or disjoint supports**
Kevin Vervier, Pierre Mahé, Alexandre d'Aspremont, Jean-Baptiste Veyrieras, Jean-Philippe Vert

17:30 – 17:50 **Scalable Moment-based Inference for Latent Dirichlet Allocation**
Chi Wang, Xueqing Liu, Yanglei Song, Jiawei Han

Session 9: Multi-task and multi-label learning

Room: 102 **Chair: Stefan Kramer**

16:10 – 16:30 **Distinct Chains for Different Instances: an Effective Strategy for Multi-Label Classifier Chains**
Pablo Silva, Eduardo Gonçalves, Alex Freitas, Alexandre Plastino

16:30 – 16:50 **Bi-Directional Representation Learning for Multi-label Classification**
Xin Li, Yuhong Guo

16:50 – 17:10 **Gaussian Process Multi-task Learning Using Joint Feature Selection**
Srijith P. K., Shirish Shevade

17:10 – 17:30 **Conic Multi-Task Classification**
Cong Li, Michael Georgiopoulos, Georgios Anagnostopoulos

17:30 – 17:50 **Random forests with random projections of the output space for high dimensional multi-label classification**
Arnaud Joly, Pierre Geurts, Louis Wehenkel

Session 10: Applications and social data mining

Room: 105 **Chair: Geoffrey Holmes**

16:10 – 16:30 **Learning about meetings**
Been Kim, Cynthia Rudin

16:30 – 16:50 **Approximating the Crowd**
Seyda Ertekin, Cynthia Rudin, Haym Hirsh

16:50 – 17:10 **Revisit Behavior in Social Media: The Phoenix-R Model and Discoveries**
Flavio Figueiredo, Jussara Almeida, Christos Faloutsos, Bruno Ribeiro, Yasuko Matsubara

17:10 – 17:30 **Conditional Log-linear Models for Mobile Application Usage Prediction**
Jingu Kim, Taneli Mielikäinen

17:30 – 17:50 **Students, Teachers, Exams and MOOCs: Predicting and Optimizing Attainment In Web-Based Education Using A Probabilistic Graphical Model**
Yoram Bachrach, Bar Shalem, John Guiver, Chris Bishop

Session 11: Pattern mining

Room: 106 **Chair: Arno Siebes**

16:10 – 16:30 **Fast estimation of the pattern frequency spectrum**
Matthijs van Leeuwen, Antti Ukkonen

16:30 – 16:50 **A Fast Method of Statistical Assessment for Combinatorial Hypotheses Based on Frequent Itemset Enumeration**
Shin-ichi Minato, Takeaki Uno, Koji Tsuda, Aika Terada, Jun Sese

16:50 – 17:10 **Ranked Tiling**
Thanh Le Van, Matthijs van Leeuwen, Siegfried Nijssen, Ana Carolina Fierro, Kathleen Marchal, Luc De Raedt

17:10 – 17:30 **A Lossless Data Reduction For Mining Constrained Patterns in n-ary Relations**
Gabriel Poesia, Loic Cerf

17:30 – 17:50 **Uncovering the Plot: Detecting Surprising Coalitions of Entities in Multi-Relational Schemas**
Hao Wu, Jilles Vreeken, Nikolaj Tatti, Naren Ramakrishnan

Session: Demo Track

Room: Grand Foyer (spotlight in 103-104)

BestTime: Finding Representatives in Time Series Datasets
Stephan Spiegel, David Schultz, Sahin Albayrak

GrammarViz 2.0: a tool for grammar-based pattern discovery in time series
Pavel Senin, Jessica Lin, Xing Wang, Tim Oates, Manfred Lerner, Arnold Boedihardjo, Crystal Chen, Susan Frankenstein, Sunil Gandhi

KnowNow: a Serendipity-based Educational Tool for Learning Time-Linked Knowledge
Luigi Di Caro, Livio Robaldo, Nicoletta Bersia

Spa: a web-based viewer for text mining in Evidence Based Medicine
Joël Kuiper, Iain Marshall, Byron Wallace, Morris Swertz

Khiops CoViz: a tool for visual exploratory analysis of k-coclustering results
Bruno Guerraz, Marc Boullé, Dominique Gay, Fabrice Clérot

MinUS: Mining User Similarity with Trajectory Patterns
Jun Pang, Xihui Chen, Piotr Kordy, Ruipeng Lu

WebDR: A Web Workbench for Data Reduction
Stefanos Ougiaroglou, Georgios Evangelidis

BMaD – A Boolean Matrix Decomposition Framework
Andrey Tyukin, Stefan Kramer, Jörg Wicker

Branty: a social media ranking tool for brands
Alexandros Arvanitidis, Anna Serafi, Athina Vakali, Grigorios Tsoumakas

Propositionalization Online
Nada Lavrac, Matic Perovšek, Anže Vavpetič

PYTHIA: Employing Lexical and Semantic Features for Sentiment Analysis
Ioannis Katakis, Iraklis Varlamis, George Tsatsaronis

Interactive Medical Miner: Interactively exploring subpopulations in epidemiological datasets
Uli Niemann, Myra Spiliopoulou, Henry Völzke, Jens-Peter Kühn

Insight4News: Connecting News to Relevant Social Conversations
Georgiana Ifrim, Bichen Shi, Neil Hurley

TUESDAY SESSIONS, WITH ABSTRACTS

Session 1: Networks

Room: 103-104 **Chair: Jilles Vreeken**

11:00 – 11:20

Link Prediction in Multi-modal Social Networks

Panagiotis Symeonidis, Christos Perentis

Online social networks like Facebook recommend new friends to users based on an explicit social network that users build by adding each other as friends. The majority of earlier work in link prediction infers new interactions between users by mainly focusing on a single network type. However, users also form several implicit social networks through their daily interactions like commenting on people's posts or rating similarly the same products. Prior work primarily exploited both explicit and implicit social networks to tackle the group/item recommendation problem that recommends to users groups to join or items to buy. In this paper, we show that auxiliary information from the user-item network fruitfully combines with the friendship network to enhance friend recommendations. We transform the well-known Katz algorithm to utilize a multi-modal network and provide friend recommendations. We experimentally show that the proposed method is more accurate in recommending friends when compared with two single source path-based algorithms using both synthetic and real data sets.

11:20 – 11:40

Density-Based Subspace Clustering in Heterogeneous Networks

Brigitte Boden, Martin Ester, Thomas Seidl

Many real-world data sets, like data from social media or bibliographic data, can be represented as heterogeneous networks with several vertex types. Often additional attributes are available for the vertices, such as keywords for a paper. Clustering vertices in such networks, and analyzing the complex interactions between clusters of different types, can provide useful insights into the structure of the data. To exploit the full information content of the data, clustering approaches should consider the connections in the network as well as the vertex attributes. We propose the density-based clustering model TCSC for the detection of clusters in heterogeneous networks that are densely connected in the network as well as in the attribute space. Unlike previous approaches for clustering heterogeneous networks, TCSC enables the detection of clusters that show similarity only in a subset of the attributes, which is more effective in the presence of a large number of attributes.

11:40 – 12:00

Interestingness-driven Diffusion Process Summarization in Dynamic Networks

Qiang Qu, Siyuan Liu, Christian Jensen, Feida Zhu, Christos Faloutsos

The widespread use of social networks enables the rapid diffusion of information, e.g., news, among users in very large communities. It is a substantial challenge to be able to observe and understand such diffusion processes, which may be modeled as networks that are both very large and highly dynamic. A key tool in this regard is data summarization. However, to the best of our knowledge, few existing studies aim to summarize graphs/networks for dynamics. Dynamic network offers new challenges over static settings, including time sensitivity and the needs for online interestingness evaluation and summary traceability, that render existing techniques inapplicable. We study the topic of dynamic network summarization: how to summarize dynamic networks with millions of nodes by only capturing the few most interesting nodes or edges over time, and we address the problem of computing summaries that represent information diffusion processes from start to end. Based on the concepts of diffusion radius and scope, we define interestingness measures in dynamic networks, and we propose OSNet, an online summarization framework for dynamic diffusion process networks.We report on extensive experiments with both synthetic and different large-scale real-life datasets. The study offers insight into the effectiveness, efficiency, and design properties of OSNet.

12:00 – 12:20

FLIP: Active Learning for Relational Network Classification

Tanwistha Saha, Huzefa Rangwala, Carlotta Domeniconi

Active learning in relational networks has gained popularity in recent years, especially for scenarios when the costs of obtaining training samples are very high. We investigate the problem of active learning for both single- and multi-labeled relational network classification in the absence of node features during training. The problem becomes harder when the number of labeled nodes available for training a model is limited due to budget constraints. The inability to use a traditional learning setup for classification of relational data, has motivated researchers to propose Collective Classification algorithms that jointly classifies all the test nodes in a network by exploiting the underlying correlation between the labels of a node and its neighbors. In this paper, we propose active learning algorithms based on different query strategies using a collective classification model where each node in a network can belong to either one class (single-labeled network) or multiple classes (multi-labeled network). We have evaluated our method on both single-labeled and multi-labeled networks, and our results are promising in both the cases for several real world datasets.

12:20 – 12:40

Faster way to agony: discovering hierarchies in directed graphs

Nikolaj Tatti

Many real-world phenomena exhibit strong hierarchical structure. Consequently, in many real-world directed social networks vertices do not play equal role. Instead, vertices form a hierarchy such that the edges appear mainly from upper levels to lower levels. Discovering hierarchies from such graphs is a challenging problem that has gained attention. Formally, given a directed graph, we want to partition vertices into levels such that ideally there are only edges from upper levels to lower levels. From computational point of view, the ideal case is when the underlying directed graph is acyclic. In such case, we can partition the vertices into a hierarchy such that there are only edges from upper levels to lower edges. In practice, graphs are rarely acyclic, hence we need to penalize the edges that violate the hierarchy. One practical approach is agony, where each violating edge is penalized based on the severity of the violation. The fastest algorithm for computing agony requires O(nm^2) time. In the paper we present an algorithm for computing agony that has better theoretical bound, namely O(m^2). We also show that in practice the obtained bound is pessimistic and that we can use our algorithm to compute agony for large datasets. Moreover, our algorithm can be used as any-time algorithm.

Session 2: Data projection and dimensionality reduction

Room: 102 **Chair: Raymond Ng**

11:00 – 11:20

Flexible Shift-Invariant Locality and Globality Preserving Projections

Feiping Nie, Xiao Cai, Heng Huang

In data mining and machine learning, the embedding methods have commonly been used as a principled way to understand the high-dimensional data. To solve the out-of-sample problem, local preserving projection (LPP) was proposed and applied to many applications. However, LPP suffers two crucial deficiencies: 1) the LPP has no shift-invariant property which is an important property of embedding methods; 2) the rigid linear embedding is used as constraint, which often inhibits the optimal manifold structures finding. To overcome these two important problems, we propose a novel flexible shift-invariant locality and globality preserving projection method, which utilizes a newly defined graph Laplacian and the relaxed embedding constraint. The proposed objective is very challenging to solve, hence we derive a new optimization algorithm with rigorously proved global convergence. More importantly, we prove our optimization algorithm is a Newton method with fast quadratic convergence rate. Extensive experiments have been performed on six benchmark data sets. In all empirical results, our method shows promising results.

11:20 – 11:40

Anomaly Detection with Score Functions based on the Reconstruction Error of the Kernel PCA

Laetitia Chapel, Chloé Friguet

We propose a novel non-parametric statistical test that allows the detection of anomalies given a set of (possibly high dimensional) sample points drawn from a nominal probability distribution. Our test statistic is the distance of a query point mapped in a feature space to its projection on the eigen-structure of the kernel matrix computed on the sample points. Indeed, the eigenfunction expansion of a Gram matrix is dependent on the input data density f_0. The resulting statistical test is shown to be uniformly most powerful for a given false alarm level alpha when the alternative density is uniform over the support of the null distribution. The algorithm can be computed in O(n^3 + n^2) and testing a query point only involves matrix vector products. Our method is tested on both artificial and benchmarked real data sets and demonstrates good performances w.r.t. competing methods.

11:40 – 12:00

Learning Binary Codes with Bagging PCA

Cong Leng, Jian Cheng, Ting Yuan, Hanqing Lu

For the eigendecomposition based hashing approaches, the information caught in different dimensions is unbalanced and most of them is typically contained in the top eigenvectors. This often leads to an unexpected phenomenon that longer code does not necessarily yield better performance. This paper attempts to leverage the bootstrap sampling idea and integrate it with PCA, resulting in a new projection method called Bagging PCA, in order to learn effective binary codes. Specifically, a small fraction of the training data is randomly sampled to learn the PCA directions each time and only the top eigenvectors are kept to generate one piece of short code. This process is repeated several times and the obtained short codes are concatenated into one piece of long code. By considering each piece of short code as a “”super-bit””, the whole process is closely connected with the core idea of LSH. Both theoretical and experimental analyses demonstrate the effectiveness of the proposed method.

12:00 – 12:20

A Unified Framework for Probabilistic Component Analysis

Mihalis Nicolaou, Stefanos Zafeiriou, Maja Pantic

We present a unifying framework which reduces the construction of probabilistic component analysis techniques to a mere selection of the latent neighbourhood, thus providing an elegant and principled framework for creating novel component analysis models as well as constructing probabilistic equivalents of deterministic component analysis methods. Under our framework, we unify many very popular and well-studied component analysis algorithms, such as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Locality Preserving Projections (LPP) and Slow Feature Analysis (SFA), some of which have no probabilistic equivalents in literature thus far. We firstly define the Markov Random Fields (MRFs) which encapsulate the latent connectivity of the aforementioned component analysis techniques; subsequently, we show that the projection directions produced by all PCA, LDA, LPP and SFA are also produced by the Maximum Likelihood (ML) solution of a single joint probability density function, composed by selecting one of the defined MRF priors while utilising a simple observation model. Furthermore, we propose novel Expectation Maximization (EM) algorithms, exploiting the proposed joint PDF, while we generalize the proposed methodologies to arbitrary connectivities via parametrizable MRF products. Theoretical analysis and experiments on both simulated and real world data show the usefulness of the proposed framework, by deriving methods which well outperform state-of-the-art equivalents.

12:20 – 12:40

Interactive Knowledge-Based Kernel PCA

Dino Oglic, Daniel Paurat, Thomas Gaertner

Data understanding is an iterative process in which domain experts combine their knowledge with the data at hand to explore and confirm hypotheses. One important set of tools for exploring hypotheses about data are visualizations. Often, however, traditional, unsupervised dimensionality reduction algorithms are used for visualization. These tools allow for interaction, i.e., exploring different visualizations, only by means of manipulating some technical parameters of the algorithm. Therefore, instead of being able to intuitively interact with the visualization, domain experts have to learn and argue about these technical parameters. In this paper we propose a knowledge-based kernel PCA approach that allows for intuitive interaction with data visualizations. Each embedding direction is given by a non-convex quadratic optimization problem over an ellipsoid and has a globally optimal solution in the kernel feature space. A solution can be found in polynomial time using the algorithm presented in this paper. To facilitate direct feedback, i.e., updating the whole embedding with a sufficiently high frame-rate during interaction, we reduce the computational complexity further by incremental up- and down-dating. Our empirical evaluation demonstrates the flexibility and utility of this approach.

Session 3: Precision and recall in classification

Room: 105 **Chair: Krzysztof Dembczyński**

11:00 – 11:20

Optimal Thresholding of Classifiers to Maximize F1 Measure

Zachary Lipton, Charles Elkan, Balakrishnan Narayanaswamy

This paper provides new insight into maximizing F1 measures in the context of binary classification and also in the context of multilabel classification. The harmonic mean of precision and recall, the F1 measure is widely used to evaluate the success of a binary classifier when one class is rare. Micro average, macro average, and per instance average F1 measures are used in multilabel classification. For any classifier that produces a real-valued output, we derive the relationship between the best achievable F1 value and the decision-making threshold that achieves this optimum. As a special case, if the classifier outputs are well-calibrated conditional probabilities, then the optimal threshold is half the optimal F1 value. As another special case, if the classifier is completely uninformative, then the optimal behavior is to classify all examples as positive. When the actual prevalence of positive examples is low, this behavior can be undesirable. As a case study, we discuss the results, which can be surprising, of maximizing F1 when predicting 26,853 labels for Medline documents.

11:20 – 11:40

Rate-constrained ranking and the rate-weighted AUC

Louise Millard, Peter Flach, Julian Higgins

Ranking tasks, where instances are ranked by a predicted score, are common in machine learning. Often only a proportion of the instances in the ranking can be processed, and this quantity, the predicted positive rate (PPR), may not be known precisely. In this situation, the evaluation of a model's performance needs to account for these imprecise constraints on the PPR, but existing metrics such as the area under the ROC curve (AUC) and early retrieval metrics such as normalised discounted cumulative gain (NDCG) cannot do this. In this paper we introduce a novel metric, the rate-weighted AUC (rAUC), to evaluate ranking models when constraints across the PPR exist, and provide an efficient algorithm to estimate the rAUC using an empirical ROC curve. Our experiments show that rAUC, AUC and NDCG often select different models. We demonstrate the usefulness of rAUC on a practical application: ranking articles for rapid reviews in epidemiology.

11:40 – 12:00

Rate-oriented point-wise confidence bounds for ROC curves

Louise Millard, Meelis Kull, Peter Flach

Common approaches to generating confidence bounds around ROC curves have several shortcomings. We resolve these weaknesses with a new `rate-oriented' approach. We generate confidence bounds composed of a series of confidence intervals for a consensus curve, each at a particular predicted positive rate (PPR), with the aim that each confidence interval contains new samples of this consensus curve with probability 95%. We propose two approaches; a parametric and a bootstrapping approach, which we base on a derivation from first principles. Our method is particularly appropriate with models used for a common type of task that we call rate-constrained, where a certain proportion of examples needs to be classified as positive by the model, such that the operating point will be set at a particular PPR value.

12:00 – 12:20

The Bane of Skew: Uncertain Ranks and Unrepresentative Precision

Thomas Lampert, Pierre Gançarski

While a problem's skew is often assumed to be constant, this paper discusses three settings where this assumption does not hold. Consequently, incorrectly assuming skew to be constant in these contradicting cases results in an over or under estimation of an algorithm's performance. The area under a precision-recall curve (AUCPR) is a common summary measurement used to report the performance of machine learning algorithms. It is well known that precision is dependent upon class skew, which often varies between datasets. In addition to this, it is demonstrated herein that under certain circumstances the relative ranking of an algorithm (as measured by AUCPR) is not constant and is instead also dependent upon skew. The skew at which the performance of two algorithms inverts and the relationship between precision measured at different skews are defined. This is extended to account for temporal skew characteristics and situations in which skew cannot be precisely defined. Formal proofs for these findings are presented, desirable properties are proved and their application demonstrated.

12:20 – 12:40

On the null distribution of the precision and recall curve

Miguel Lopes, Gianluca Bontempi

Precision recall curves (pr-curves) and the associated area under (AUPRC) are commonly used to assess the accuracy of information retrieval (IR) algorithms. An informative baseline is random selection. The associated probability distribution makes it possible to assess pr-curve significancy (as a p-value relative to the null of random). To our knowledge, no analytical expression of the null distribution of empirical pr-curves is available, and the only measure of significancy used in the literature relies on non-parametric Monte Carlo simulations. In this paper, we derive analytically the expected null pr-curve and AUPRC, for different interpolation strategies. The AUPRC variance is also derived, and we use it to propose a continuous approximation to the null AUPRC distribution, based on the beta distribution. Properties of the empirical pr-curve and common interpolation strategies are also discussed.

S-ILP: Joint ILP Session

Room: 101 **Chair: Jesse Davis**

11:00 - 11:20

Evidence-based Clustering for Scalable Inference in Markov Logic

Deepak Venugopal, Vibhav Gogate

Markov Logic is a powerful representation that unifies first-order logic and probabilistic graphical models. However, scaling-up inference in Markov Logic Networks (MLNs) is extremely challenging. Standard graphical model inference algorithms operate on the propositional Markov network obtained by grounding the MLN and do not scale well as the number of objects in the real-world domain increases. On the other hand, algorithms which perform inference directly at the first-order level, namely lifted inference algorithms, although more scalable than propositional algorithms, require the MLN to have specific symmetric structure. Worse still, evidence breaks symmetries, and the performance of lifted inference is the same as propositional inference (or sometimes worse, due to overhead). In this paper, we propose a general method for solving this ""evidence"" problem. The main idea in our method is to approximate the given MLN having, say, objects by an MLN having k objects such that k is much lesser than n and the results obtained by running potentially much faster inference on the smaller MLN are as close as possible to the ones obtained by running inference on the larger MLN. We achieve this by finding clusters of ""similar"" groundings using standard clustering algorithms (e.g., K-means), and replacing all groundings in the cluster by their cluster center. To this end, we develop a novel distance (or similarity) function for measuring the similarity between two groundings, based on the evidence presented to the MLN. We evaluated our approach on many different benchmark MLNs utilizing various clustering and inference algorithms. Our experiments clearly show the generality and scalability of our approach.

11:20 - 11:40

Effective Blending of Two and Three-way Interactions for Modeling Multi-relational Data

Alberto Garcia-Duran, Antoine Bordes, Nicolas Usunier

Much work has been recently proposed to model relational data, especially in the multi-relational case, where different kinds of relationships are used to connect the various data entities. Previous attempts either consist of powerful systems with high capacity to model complex connectivity patterns, which unfortunately usually end up overfitting on rare relationships, or in approaches that trade capacity for simplicity in order to fairly model all relationships, frequent or not. In this paper, we propose a happy medium obtained by complementing a high-capacity model with a simpler one, both pre-trained separately and jointly fine- tuned. We show that our approach outperforms existing models on different types of relationships, and achieves state-of-the-art results on two benchmarks of the literature.

11:40 - 12:00

Towards Automatic Feature Construction for Supervised Classification

Marc Boullé

We suggest an approach to automate variable construction for supervised learning, especially in the multi-relational setting. Domain knowledge is specified by describing the structure of data by the means of variables, tables and links across tables, and choosing construction rules. The space of variables that can be constructed is virtually infinite, which raises both combinatorial and over-fitting problems. We introduce a prior distribution over all the constructed variables, as well as an effective algorithm to draw samples of constructed variables from this distribution. Experiments show that the approach is robust and efficient.

12:00 - 12:20

Fast Learning of Relational Dependency Networks

Oliver Schulte, Zhensong Qian, Arthur E. Kirkpatrick, Xiaoqian Yin, Yan Sun

A Relational Dependency Network (RDN) is a directed graphical model widely used for multi-relational data. These networks allow cyclic dependencies, necessary to represent relational autocorrelations. We describe an approach for learning both the RDN's structure and its parameters, given an input relational database: First learn a Bayesian network (BN), then transform the Bayesian network to an RDN. Thus fast Bayes net learning can provide fast RDN learning. The BN-to-RDN transform comprises a simple, local adjustment of the Bayes net structure and a closed-form transform of the Bayes net parameters. This method can learn an RDN for a dataset with a million tuples in minutes. We empirically compare our approach to state-of-the art RDN learning methods that use functional gradient boosting, on ve benchmark datasets. Learning RDNs via BNs scales much better to large datasets than learning RDNs with boosting, and provides competitive accuracy in predictions.

12:20 - 12:40

Complex aggregates over subsets of elements

Celine Vens, Sofie Van Gassen, Tom Dhaene, Yvan Saeys

Complex aggregates have been proposed as a way to bridge the gap between approaches that handle sets by imposing conditions on specific elements, and approaches that handle them by imposing conditions on aggregated values. A complex aggregate aggregates over a subset of the elements in a set, where this subset is defined by conditions on the attribute values. In this paper, we present a new type of complex aggregate, where this subset is defined to be a cluster of the set. This is useful if subsets that are relevant for the task at hand are difficult to describe in terms of attribute conditions. This work is motivated from the analysis of flow cytometry data, where the sets are cells, and the subsets are cell populations. We describe two approaches to aggregate over clusters, and validate one of them empirically, motivating future research in this direction.

Session 4: Kernel-based learning and structured outputs

Room: 102 **Chair: Willem Waegeman**

14:00 – 14:20

Attributed Graph Kernels Using the Jensen-Tsallis q-Differences

Lu Bai, Luca Rossi, Horst Bunke, Edwin Hancock

We propose a family of attributed graph kernels based on mutual information measures, i.e., the Jensen-Tsallis (JT) q-differences (for q\in[1,2]) between probability distributions over the graphs. To this end, we first assign a probability to each vertex of the graph through a continuous-time quantum walk (CTQW). We then adopt the tree-index approach to strengthen the original vertex labels, and we show how the CTQW can induce a probability distribution over these strengthened labels. We show that our JT kernel (for q=1) overcomes the shortcoming of discarding non-isomorphic substructures arising in the R-convolution kernels. Moreover, we prove that the JT kernels generalize the Jensen-Shannon graph kernel (for q=1) and the classical subtree kernel (for q=2), respectively. Experimental evaluations demonstrate the effectiveness and efficiency of the JT kernels.

14:20 – 14:40

Relative Comparison Kernel Learning with Auxiliary Kernels

Eric Heim, Hamed Valizadegan, Milos Hauskrecht

In this work we consider the problem of learning a positive semidefinite kernel matrix from relative comparisons of the form: “object A is more similar to object B than it is to C”, where comparisons are given by humans. Existing solutions to this problem assume many comparisons are provided to learn a meaningful kernel. However, this can be considered unrealistic for many real-world tasks since a large amount of human input is often costly or difficult to obtain. Because of this, only a limited number of these comparisons may be provided. We propose a new kernel learning approach that supplements the few relative comparisons with “auxiliary” kernels built from more easily extractable features in order to learn a kernel that more completely models the notion of similarity gained from human feedback. Our proposed formulation is a convex optimization problem that adds only minor overhead to methods that use no auxiliary information. Empirical results show that in the presence of few training relative comparisons, our method can learn kernels that generalize to more out-of-sample comparisons than methods that do not utilize auxiliary information, as well as similar metric learning methods.

14:40 – 15:00

Approximate Consistency: Towards Foundations of Approximate Kernel Selection

Lizhong Ding, Shizhong Liao

Kernel selection is critical to kernel methods. Approximate kernel selection is an emerging approach to alleviating the computational burdens of kernel selection by introducing kernel matrix approximation. Theoretical problems faced by approximate kernel selection are how kernel matrix approximation impacts kernel selection and whether this impact can be ignored for large enough examples. In this paper, we introduce the notion of approximate consistency for kernel matrix approximation algorithm to tackle the theoretical problems and establish the preliminary foundations of approximate kernel selection. By analyzing the approximate consistency of kernel matrix approximation algorithms, we can answer the question that, under what conditions, and how, the approximate kernel selection criterion converges to the accurate one. Taking two kernel selection criteria as examples, we analyze the approximate consistency of Nystr¨om approximation and multilevel circulant matrix approximation. Finally, we empirically verify our theoretical findings.

15:00 – 15:20

Kernel Principal Geodesic Analysis

Suyash Awate, Yen-Yun Yu, Ross Whitaker

Kernel principal component analysis (kPCA) has been proposed as a dimensionality-reduction technique that achieves nonlinear, low-dimensional representations of data via the mapping to kernel feature space. Conventionally, kPCA relies on Euclidean statistics in kernel feature space. However, Euclidean analysis can make kPCA inefficient or incorrect for many popular kernels that map input points to a hypersphere in kernel feature space. To address this problem, this paper proposes a novel adaptation of kPCA, namely kernel principal geodesic analysis (kPGA), for hyperspherical statistical analysis in kernel feature space. This paper proposes tools for statistical analyses on the Riemannian manifold of the Hilbert sphere in the reproducing kernel Hilbert space, including algorithms for computing the sample weighted Karcher mean and eigen analysis of the sample weighted Karcher covariance. It then applies these tools to propose novel methods for (i)~dimensionality reduction and (ii)~clustering using mixture-model fitting. The results, on simulated and real-world data, show that kPGA-based methods perform favorably relative to their kPCA-based analogs.

15:20 – 15:40

FASOLE: Fast Algorithm for Structured Output LEarning

Vojtech Franc

This paper proposes a novel Fast Algorithm for Structured Ouput LEarning (FASOLE). FASOLE implements the Sequential Dual Ascent (SDA) algorithm for solving the dual problem of the Structured Output Support Vector Machines (SO-SVM). Unlike existing instances of SDA algorithm applied for SO-SVM, the proposed FASOLE uses a different working set selection strategy which provides nearly maximal improvement of the objective function in each update. FASOLE processes examples in on-line fashion and it provides certificate of optimality. FASOLE is proven to find the $\veps$-optimal solution in $\SO(\frac{1}{\veps^2})$ time in the worst case. In the empirical comparison FASOLE consistently outperforms the existing state-of-the-art solvers, like the Cutting Plane Algorithm or the Block-Coordinate Frank-Wolfe algorithm, achieving up to an order of magnitude speedups while obtaining the same precise solution.

Session 5: Classification

Room: 105 **Chair: Charles Elkan**

14:00 – 14:20

Robust Distributed Training of Linear Classifiers Based on Divergence Minimization Principle

Junpei Komiyama, Hidekazu Oiwa, Hiroshi Nakagawa

We study a distributed training of a linear classifier in which the data is separated into many shards and each worker only has access to its own shard. The goal of this distributed training is to utilize the data of all shards to obtain a well-performing linear classifier. The iterative parameter mixture (IPM) framework (Mann et al., 2009) is a state-of-the-art distributed learning framework that has a strong theoretical guarantee when the data is clean. However, contamination on shards, which sometimes arises in real world environments, largely deteriorates the performances of the distributed training. To remedy the negative effect of the contamination, we propose a divergence minimization principle for the weight determination in IPM. From this principle, we can naturally derive the Beta-IPM scheme, which leverages the power of robust estimation based on the beta divergence. A mistake/loss bound analysis indicates the advantage of our Beta-IPM in contaminated environments. Experiments with various datasets revealed that, even when 80% of the shards are contaminated, Beta-IPM can suppress the influence of the contamination.

14:20 – 14:40

Learning A Priori Constrained Weighted Majority Votes

Aurélien Bellet, Amaury Habrard, Emilie Morvant, Marc Sebban

Weighted majority votes allow one to combine the output of several classifiers or voters. MinCq is a recent algorithm for optimizing the weight of each voter based on the minimization of a theoretical bound over the risk of the vote with elegant PAC-Bayesian generalization guarantees. However, while it has demonstrated good performance when combining weak classifiers, MinCq cannot make use of the useful a priori knowledge that one may have when using a mixture of weak and strong voters. In this paper, we propose P-MinCq, an extension of MinCq that can incorporate such knowledge in the form of a constraint over the distribution of the weights, along with general proofs of convergence that stand in the sample compression setting for data-dependent voters. The approach is applied to a vote of k-NN classifiers with a specific modeling of the voters' performance. P-MinCq significantly outperforms the classic k-NN classifier, a symmetric NN and MinCq using the same voters. We show that it is also competitive with LMNN, a popular metric learning algorithm, and that combining both approaches further reduces the error.

14:40 – 15:00

Integer Bayesian Network Classifiers

Sebastian Tschiatschek, Karin Paul, Franz Pernkopf

This paper introduces integer Bayesian network classifiers (BNCs), i.e. BNCs with discrete valued nodes where parameters are stored as integer numbers. These networks allow for efficient implementation in hardware while maintaining a (partial) probabilistic interpretation under scaling. An algorithm for the computation of margin maximizing integer parameters is presented and its efficiency is demonstrated. The resulting parameters have superior classification performance compared to parameters obtained by simple rounding of double-precision parameters, particularly for very low number of bits.

15:00 – 15:20

Randomized Operating Point Selection in Adversarial Classification

Viliam Lisý, Robert Kessl, Tomas Pevny

Security systems for email spam filtering, network intrusion detection, steganalysis, and watermarking, frequently use classifiers to separate malicious behavior from legitimate. Typically, they use a fixed operating point minimizing the expected cost / error. This allows a rational attacker to deliver invisible attacks just below the detection threshold. We model this situation as a non-zero sum normal form game capturing attacker's expected payoffs for detected and undetected attacks, and detector's costs for false positives and false negatives computed based on the Receiver Operating Characteristic (ROC) curve of the classifier. The analysis of Nash and Stackelberg equilibria reveals that using a randomized strategy over multiple operating points forces the rational attacker to design less efficient attacks and substantially lowers the expected cost of the detector. We present the equilibrium strategies for sample ROC curves from network intrusion detection system and evaluate the corresponding benefits.

15:20 – 15:40

Separating Rule Refinement and Rule Selection Heuristics in Inductive Rule Learning

Julius Stecher, Frederik Janssen, Johannes Fuernkranz

Conventional rule learning algorithms use a single heuristic for evaluating both, rule refinements and rule selection. However, whereas rule selection proceeds in a bottom-up specific-to-general direction, rule refinement typically operates top-down. Hence, we propose in this paper that criteria for evaluating rule refinements should reflect this by operating in an inverted coverage space. We motivate this choice by examples, and show that a suitably adapted rule learning algorithm outperforms its original counter-part on a large set of benchmark problems.

Session 6: Feature selection and construction

Room: 106 **Chair: Gianluca Bontempi**

14:00 – 14:20

Deterministic Feature Selection for Regularized Least Squares Classification

Saurabh Paul, Petros Drineas

We introduce a deterministic sampling based feature selection technique for regularized least squares classification. The method is unsupervised and gives worst-case guarantees of the generalization power of the classification function after feature selection with respect to the classification function obtained using all features. We perform experiments on synthetic and real-world datasets, namely a subset of TechTC-300 datasets, to support our theory. Experimental results indicate that the proposed method performs better than the existing feature selection methods.

14:20 – 14:40

Automatic design of neuromarkers for OCD characterization

Oscar Garcia-Hinde, Emilio Parrado-Hernandez, Vanessa Gomez-Verdejo, Manel Martinez-Ramon, Carles Soriano-Mas

This paper proposes a new paradigm to discover biomarkers capable of characterizing obsessive-compulsive disorder (OCD). These biomarkers, named neuromarkers, will be obtained through the analysis of sets of magnetic resonance images (MRI) of OCD patients and control subjects. The design of the neuromarkers stems from a method for the automatic discovery of clusters of voxels relevant to OCD recently published by the authors. With these clusters as starting point, we will define the neuromarkers as a set of measurements describing features of these individual regions. The principal goal of the project is to come up with a set of about 50 neuromarkers for OCD characterization that are easy to interpret and handle by the psychiatric community.

14:40 – 15:00

Unsupervised Feature Selection via Unified Trace Ratio Formulation and K-means Clustering (TRACK)

De Wang, Feiping Nie, Heng Huang

Feature selection plays a crucial role in scientific research and practical applications. In the real world applications, labeling data is time and labor consuming. Thus, unsupervised feature selection methods are desired for many practical applications. Linear discriminant analysis (LDA) with trace ratio criterion is a supervised dimensionality reduction method that has shown good performance to improve classifications. In this paper, we first propose a unified objective to seamlessly accommodate trace ratio formulation and K-means clustering procedure, such that the trace ratio criterion is extended to unsupervised model. After that, we propose a novel unsupervised feature selection method by integrating unsupervised trace ratio formulation and structured sparsity-inducing norms regularization. The proposed method can harness the discriminant power of trace ratio criterion, thus it tends to select discriminative features. Meanwhile, we also provide two important theorems to guarantee the unsupervised feature selection process. Empirical results on four benchmark data sets show that the proposed method outperforms other sate-of-the-art unsupervised feature selection algorithms in all three clustering evaluation metrics.

15:00 – 15:20

Covariate-Correlated Lasso for Feature Selection

Bo Jiang, Chris Ding, Bin Luo

Lasso-type variable selection has been increasingly adopted in many applications. In this paper, we propose a covariate-correlated Lasso that selects the covariates correlated more strongly with the response variable. We propose an efficient algorithm to solve this Lasso-type optimization and prove its convergence. Experiments on DNA gene expression data sets show that the selected covariates correlate more strongly with the response variable, and the residual values are decreased, indicating better covariate selection. The selected covariates lead to better classification performance.

15:20 – 15:40

Unsupervised Interaction-Preserving Discretization of Multivariate Data

Hoang-Vu Nguyen, Emmanuel Müller, Jilles Vreeken, Klemens Bohm

Discretization is the transformation of continuous data into discrete bins. It is an important and general pre-processing technique, and a critical element of many data mining and data management tasks. The general goal is to obtain data that retains as much information in the continuous original as possible. In general, but in particular for exploratory tasks, a key open question is how to discretize multivariate data such that significant associations and patterns are preserved. That is exactly the problem we study in this paper. We propose IPD, an information-theoretic method for unsupervised discretiza- tion that focuses on preserving multivariate interactions. To this end, when discretiz- ing a dimension, we consider the distribution of the data over all other dimensions. In particular, our method examines consecutive multivariate regions and combines them if (a) their multivariate data distributions are statistically similar, and (b) this merge reduces the MDL encoding cost. To assess the similarities, we propose ID, a novel interaction distance that does not require assuming a distribution and permits computation in closed form. We give an efficient algorithm for finding the optimal bin merge, as well as a fast well-performing heuristic. Empirical evaluation through pattern-based compression, outlier mining, and clas- sification shows that by preserving interactions we consistently outperform the state of the art in both quality and speed.

Session 7: Clustering

Room: 103-104 **Chair: Christel Vrain**

14:00 – 14:20

Clustering via Mode Seeking by Direct Estimation of the Gradient of a Log-Density

Hiroaki Sasaki, Aapo Hyvarinen, Masashi Sugiyama

Mean shift clustering finds the modes of the data probability density by identifying the zero points of the density gradient. Since it does not require to fix the number of clusters in advance, the mean shift has been a popular clustering algorithm in various application fields. A typical implementation of the mean shift is to first estimate the density by kernel density estimation and then compute its gradient. However, since a good density estimation does not necessarily imply an accurate estimation of the density gradient, such an indirect two-step approach is not reliable. In this paper, we propose a method to directly estimate the gradient of the log-density without going through density estimation. The proposed method gives the global solution analytically and thus is computationally efficient. We then develop a mean-shift-like fixed-point algorithm to find the modes of the density for clustering. As in the mean shift, one does not need to set the number of clusters in advance. We experimentally show that the proposed clustering method significantly outperforms the mean shift especially for high-dimensional data.

14:20 – 14:40

Fast Gaussian Pairwise Constrained Spectral Clustering

David Chatel, Marc Tommasi, Pascal Denis

We consider the problem of spectral clustering with partial supervision in the form of must-link and cannot-link constraints. Such pairwise constraints are common in problems like coreference resolution in natural language processing. The approach developed in this paper is to learn a new representation space for the data together with a distance in this new space. The representation space is obtained through a constraint-driven linear transformation of a spectral embedding of the data. Constraints are expressed with a Gaussian function that locally reweights the similarities in the projected space. A global, non-convex optimization objective is then derived and the model is learned via gradient descent techniques. Our algorithm is evaluated on standard datasets and compared with state of the art algorithms, like [Kamvar 2003,Li 2009,Wang and Davidson 2010]. Results on these datasets, as well on the CoNLL-2012 coreference resolution shared task dataset, show that our algorithm significantly outperforms related approaches and is also much more scalable.

14:40 – 15:00

Ratio-based Multiple Kernel Clustering

Grigorios Tzortzis, Aristidis Likas

Maximum margin clustering (MMC) approaches extend the large margin principle of SVM to unsupervised learning with considerable success. In this work, we utilize the ratio between the margin and the intra-cluster variance, to explicitly consider both the separation and the compactness of the clusters in the objective. Moreover, we employ multiple kernel learning (MKL) to jointly learn the kernel and a partitioning of the instances, thus overcoming the kernel selection problem of MMC. Importantly, the margin alone cannot reliably reflect the quality of the learned kernel, as it can be enlarged by a simple scaling of the kernel. In contrast, our ratio-based objective is scale invariant and also invariant to the type of norm constraints on the kernel parameters. Optimization of the objective is performed using an iterative gradient-based algorithm. Comparative clustering experiments on various datasets demonstrate the effectiveness of the proposed formulation.

15:00 – 15:20

Boosted Mean Shift Clustering

Yazhou Ren, Uday Kamath, Carlotta Domeniconi, Guoji Zhang

Mean shift is a nonparametric clustering technique that does not require the number of clusters in input and can find clusters of arbitrary shapes. While appealing, the performance of the mean shift algorithm is sensitive to the selection of the bandwidth, and can fail to capture the correct clustering structure when multiple modes exist in one cluster. DBSCAN is an efficient density based clustering algorithm, but it is also sensitive to its parameters and typically merges overlapping clusters. In this paper we propose Boosted Mean Shift Clustering (BMSC) to address these issues. BMSC partitions the data across a grid and applies mean shift locally on the cells of the grid, each providing a number of intermediate modes (iModes). A mode-boosting technique is proposed to select points in denser regions iteratively, and DBSCAN is utilized to partition the obtained iModes iteratively. Our proposed BMSC can overcome the limitations of mean shift and DBSCAN, while preserving their desirable properties. Complexity analysis shows its potential to deal with large-scale data and extensive experimental results on both synthetic and real benchmark data demonstrate its effectiveness and robustness to parameter settings.

15:20 – 15:40

FILTA: Better View Discovery from Collections of Clusterings via Filtering

Yang Lei, Nguyen Xuan Vinh, Jeffrey Chan, James Bailey

Multiple clustering analysis is a growing area whose aim is to discover multiple high quality and non-redundant clusterings (views) of a dataset. A popular method for this task is meta clustering, which involves generation of a large number of base clusterings, that serve as input for further user navigation and refinement. However, the effectiveness of meta clustering for view discovery is highly dependent on the distribution of the base clusterings and open challenges exist with regard to its stability and noise tolerance. In this paper we propose a simple and effective filtering algorithm (FILTA) that can be flexibly used in conjunction with any meta clustering method.Given a (raw) set of base clusterings, FILTA employs information theoretic criteria to remove those having poor quality or high redundancy, yielding a filtered output set of clusterings. This filtered set is then highly suitable for further exploration, particularly the use of visualization for determining the dominant views that exist in the dataset. We evaluate FILTA on both synthetic and real world datasets, and see how its use can enhance view discovery for complex scenarios.

Session 8: Decomposition and latent variables

Room: 103-104 **Chair: Marie-Francine Moens**

16:10 – 16:30

Pushing-Down Tensor Decompositions over Unions to Promote Reuse of Materialized Decompositions

Mijung Kim, K. Selcuk Candan

From data collection to decision making, the life cycle of data often involves many steps of integration, manipulation, and analysis. To be able to provide end-to-end support for the full data life cycle, today's data management and decision making systems increasingly combine operations for data manipulation, integration as well as data analysis. Tensor-relational model (TRM) is a framework proposed to support both relational algebraic operations (for data manipulation and integration) and tensor algebraic operations (for data analysis). In this paper, we consider joint processing of relational algebraic and tensor analysis operations. In particular, we focus on data processing workflows that involve data integration from multiple sources (through unions) and tensor decomposition tasks. While, in traditional relational algebra, the costliest operation is known to be the join, in a framework that provides both relational and tensor operations, tensor decomposition tends to be the computationally costliest operation. Therefore, it is most critical to reduce the cost of the tensor decomposition task by manipulating the data processing workflow in a way that reduces the cost of the tensor decomposition step. Therefore, in this paper, we consider data processing workflows involving tensor decomposition and union operations and we propose a novel scheme for pushing down the tensor decompositions over the union operations to reduce the overall data processing times and to promote reuse of materialized tensor decomposition results. Experimental results confirm the efficiency and effectiveness of the proposed scheme.

16:30 – 16:50

Hierarchical Latent Tree Analysis for Topic Detection

Tengfei LIU, Nevin Zhang, Peixian Chen

In the LDA approach to topic detection, a topic is determined by identifying the words that are used with high frequency when writing about the topic. However, high frequency words in one topic may be also used with high frequency in other topics. Thus they may not be the best words to characterize the topic. In this paper, we propose a new method for topic detection, where a topic is determined by identifying words that appear with high frequency in the topic and low frequency in other topics. We model patterns of word co-occurrence and co-occurrences of those patterns using a hierarchy of discrete latent variables. The states of the latent variables represent clusters of documents and they are interpreted as topics. The words that best distinguish a cluster from other clusters are selected to characterize the topic. Empirical results show that the new method yields topics with clearer thematic characterizations than the alternative approaches.

16:50 – 17:10

Causal Clustering for 2-Factor Measurement Models

Peter Spirtes, Erich Kummerfeld, Joseph Ramsey, Renjie Yang, Richard Scheines

Many social scientists are interested in inferring causal relations between "latent" variables that they cannot directly measure. One strategy commonly used to make such inferences is to use the values of variables that can be measured directly (often answers to questions in surveys or test "items'") that are thought to be measures or "indicators" of the latent variables of interest, together with a hypothesized causal graph relating the latent variables to their indicators. To use the data on the indicators to draw inferences about the causal relations between the latent variables (known as the structural model), it is necessary to hypothesize causal relations between the indicators and the latents that they are intended to indirectly measure, (known as the measurement model). The problem addressed in this paper is how to reliably infer the measurement model given measurements of the indicators, without knowing anything about the structural model, which is ultimately the question of interest. In this paper, we develop the FindTwoFactorClusters (FTFC) algorithm, a search algorithm that, in addition to being faster than existing algorithms based on vanishing tetrad constraints, also works for a more complex class of measurement models, and does not assume that the structural model is linear.

17:10 – 17:30

On learning matrices with orthogonal columns or disjoint supports

Kevin Vervier, Pierre Mahé, Alexandre d'Aspremont, Jean-Baptiste Veyrieras, Jean-Philippe Vert

We investigate new matrix penalties to jointly learn linear models with orthogonality constraints, generalizing the work of Xiao et al. (2011) who proposed a strictly convex matrix norm for orthogonal transfer. We show that this norm converges to a particular atomic norm when its convexity parameter decreases, leading to new algorithmic solutions to minimize it. We also investigate concave formulations of this norm, corresponding to more aggressive strategies to induce orthogonality, and show how these penalties can also be used to learn sparse models with disjoint supports.

17:30 – 17:50

Scalable Moment-based Inference for Latent Dirichlet Allocation

Chi Wang, Xueqing Liu, Yanglei Song, Jiawei Han

Topic models such as Latent Dirichlet Allocation have been useful text analysis methods of wide interest. Recently, moment-based inference with provable performance has been proposed for topic models. Compared with inference algorithms that approximate the maximum likelihood objective, moment-based inference has theoretical guarantee in recovering model parameters. One such inference method is tensor orthogonal decomposition, which requires only mild assumptions for exact recovery of topics. However, it suffers from scalability issue due to creation of dense, high-dimensional tensors. In this work, we propose a speedup technique by leveraging the special structure of the tensors. It is efficient in both time and space, and only requires passing the corpus twice. It improves over the state-of-the-art inference algorithm by one to three orders of magnitude, while preserving equal inference ability.

Session 9: Multi-task and multi-label learning

Room: 102 **Chair: Stefan Kramer**

16:10 – 16:30

Distinct Chains for Different Instances: an Effective Strategy for Multi-Label Classifier Chains

Pablo Silva, Eduardo Gonçalves, Alex Freitas, Alexandre Plastino

Multi-label classification (MLC) is a predictive problem in which an object may be associated with multiple labels. One of the most prominent MLC methods is the classifier chains (CC). This method induces q binary classifiers, where q represents the number of labels. Each one is responsible for predicting a specific label. These q classifiers are linked in a chain, such that at classification time each classifier considers the labels predicted by the previous ones as additional information. Although the performance of CC is largely influenced by the chain ordering, the original method uses a random ordering. To cope with this problem, in this paper we propose a novel method which is capable of finding a specific and more effective chain for each new instance to be classified. Experiments have shown that the proposed method obtained, overall, higher predictive accuracies than the well-established binary relevance, CC and CC ensemble methods.

16:30 – 16:50

Bi-Directional Representation Learning for Multi-label Classification

Xin Li, Yuhong Guo

Multi-label classification is a central problem in many application domains. In this paper, we present a novel supervised bi-directional model that learns a low-dimensional mid-level representation for multi-label classification. Unlike traditional multi-label learning methods which identify intermediate representations from either the input space or the output space but not both, the mid-level representation in our model has two complementary parts that capture intrinsic information of the input data and the output labels respectively under the auto encoder principle while augmenting each other for the target output label prediction. The resulting optimization problem can be solved efficiently using an iterative procedure with alternating steps, while closed-form solutions exist for one major step. Our experiments conducted on a variety of multi-label data sets demonstrate the efficacy of the proposed bi-directional representation learning model for multi-label classification.

16:50 – 17:10

Gaussian Process Multi-task Learning Using Joint Feature Selection

Srijith P. K., Shirish Shevade

Multi-task learning involves solving multiple related learning problems by sharing some common structure for improved generalization performance. A promising idea to multi-task learning is joint feature selection where a sparsity pattern is shared across task specific feature representations. In this paper, we propose a novel Gaussian Process (GP) approach to multi-task learning based on joint feature selection. The novelty of the proposed approach is that it captures the task similarity by sharing a sparsity pattern over the kernel hyper-parameters associated with each task. This is achieved by considering a hierarchical model which imposes a multi-Laplacian prior over the kernel hyper-parameters. This leads to a flexible GP model which can handle a wide range of multi-task learning problems and can identify features relevant across all the tasks. The hyper-parameter estimation results in an optimization problem which is solved using a block co-ordinate descent algorithm. Experimental results on synthetic and real world multi-task learning data sets demonstrate that the flexibility of the proposed model is useful in getting better generalization performance.

17:10 – 17:30

Conic Multi-Task Classification

Cong Li, Michael Georgiopoulos, Georgios Anagnostopoulos

Traditionally, Multi-task Learning (MTL) models optimize the average of task-related objective functions, which is an intuitive approach and which we will be referring to as Average MTL. However, a more general framework, referred to as Conic MTL, can be formulated by considering conic combinations of the objective functions instead; in this framework, Average MTL arises as a special case, when all combination coefficients equal 1. Although the advantage of Conic MTL over Average MTL has been shown experimentally in previous works, no theoretical justification has been provided to date. In this paper, we derive a generalization bound for the Conic MTL method, and demonstrate that the tightest bound is not necessarily achieved, when all combination coefficients equal 1; hence, Average MTL may not always be the optimal choice, and it is important to consider Conic MTL. As a byproduct of the generalization bound, it also theoretically explains the good experimental results of previous relevant works. Finally, we propose a new Conic MTL model, whose conic combination coefficients minimize the generalization bound, instead of choosing them heuristically as has been done in previous methods. The rationale and advantage of our model is demonstrated and verified via a series of experiments by comparing with several other methods.

17:30 – 17:50

Random forests with random projections of the output space for high dimensional multi-label classification

Arnaud Joly, Pierre Geurts, Louis Wehenkel

We adapt the idea of random projections applied to the output space, so as to enhance tree-based ensemble methods in the context of multi-label classification. We show how learning time complexity can be reduced without affecting computational complexity and accuracy of predictions. We also show that random output space projections may be used in order to reach different bias-variance tradeoffs, over a broad panel of benchmark problems, and that this may lead to improved accuracy while reducing significantly the computational burden of the learning stage.

Session 10: Applications and social data mining

Room: 105 **Chair: Geoffrey Holmes**

16:10 – 16:30

Learning about meetings

Been Kim, Cynthia Rudin

Most people participate in meetings almost every day, multiple times a day. The study of meetings is important, but also challenging, as it requires an understanding of social signals and complex interpersonal dynamics. Our aim this work is to use a data-driven approach to the science of meetings. We provide tentative evidence that: i) it is possible to automatically detect when during the meeting a key decision is taking place, from analyzing only the local dialogue acts, ii) there are common patterns in the way social dialogue acts are interspersed throughout a meeting, iii) at the time key decisions are made, the amount of time left in the meeting can be predicted from the amount of time that has passed, iv) it is often possible to predict whether a proposal during a meeting will be accepted or rejected based entirely on the language (the set of persuasive words) used by the speaker.

16:30 – 16:50

Approximating the Crowd

Seyda Ertekin, Cynthia Rudin, Haym Hirsh

The problem of "approximating the crowd" is that of estimating the crowd's majority opinion by querying only a subset of it. Algorithms that approximate the crowd can intelligently stretch a limited budget for a crowdsourcing task.We present an algorithm, "CrowdSense," that works in an online fashion where items come one at a time. CrowdSense dynamically samples subsets of the crowd based on an exploration/exploitation criterion. The algorithm produces a weighted combination of the subset's votes that approximates the crowd's opinion. We then introduce two variations of CrowdSense that make various distributional approximations to handle distinct crowd characteristics. In particular, the first algorithm makes a statistical independence approximation of the labelers for large crowds, whereas the second algorithm finds a lower bound on how often the current sub-crowd agrees with the crowd's majority vote. Our experiments on CrowdSense and several baselines demonstrate that we can reliably approximate the entire crowd's vote by collecting opinions from a representative subset of the crowd.

16:50 – 17:10

Revisit Behavior in Social Media: The Phoenix-R Model and Discoveries

Flavio Figueiredo, Jussara Almeida, Christos Faloutsos, Bruno Ribeiro, Yasuko Matsubara

How many listens will an artist receive on a online radio? How about plays on a YouTube video? How many of these visits are new or returning users? Modeling and mining popularity dynamics of social activity has important implications for researchers, content creators and providers. We here investigate the effect of revisits (successive visits from a single user) on content popularity. Using four datasets of social activity, with up to tens of millions media objects (e.g., YouTube videos, Twitter hashtags or LastFM artists), we show the effect of revisits in the popularity evolution of such objects. Secondly, we propose the Phoenix-R model which captures the popularity dynamics of individual objects. Phoenix-R has the desired properties of being: (1) parsimonious, being based on the minimum description length principle, and achieving lower root mean squared error than state-of-the-art baselines; (2) applicable, the model is effective for predicting future popularity values of objects.

17:10 – 17:30

Conditional Log-linear Models for Mobile Application Usage Prediction

Jingu Kim, Taneli Mielikäinen

Over the last decade, mobile device usage has evolved rapidly from basic calling and texting to primarily using applications. On average, smartphone users have tens of applications installed in their devices. As the number of installed applications grows, finding a right application at a particular moment is becoming more challenging. To alleviate the problem, we study the task of predicting applications that a user is most likely going to use at a given situation. We formulate the prediction task with a conditional log-linear model and present an online learning scheme suitable for resource-constrained mobile devices. Using real-world mobile application usage data, we evaluate the performance and the behavior of the proposed solution against other prediction methods. Based on our experimental evaluation, the proposed approach offers competitive prediction performance with moderate resource needs.

17:30 – 17:50

Students, Teachers, Exams and MOOCs: Predicting and Optimizing Attainment In Web-Based Education Using A Probabilistic Graphical Model

Yoram Bachrach, Bar Shalem, John Guiver, Chris Bishop

We propose a probabilistic graphical model for predicting student attainment in web-based education. We empirically evaluate our model on a crowdsourced dataset with students and teachers; Teachers prepared lessons on various topics. Students read lessons by various teachers and then solved a multiple choice exam. Our model gets input data regarding past interactions between students and teachers and past student attainment. It then estimates abilities of students, competence of teachers and difficulty of questions, and predicts future student outcomes. We show that our model's predictions are more accurate than heuristic approaches. We also show how demographic profiles and personality traits correlate with student performance in this task. Finally, given a limited pool of teachers, we propose an approach for using information from our model to maximize the number of students passing an exam of a given difficulty, by optimally assigning teachers to students. We evaluate the potential impact of our optimization approach using a simulation based on our dataset, showing an improvement in the overall performance.

Session 11: Pattern mining

Room: 106 **Chair: Arno Siebes**

16:10 – 16:30

Fast estimation of the pattern frequency spectrum

Matthijs van Leeuwen, Antti Ukkonen

Both exact and approximate counting of the number of frequent patterns for a given frequency threshold are hard problems. Still, having even coarse prior estimates of the number of patterns is useful, as these can be used to appropriately set the threshold and avoid waiting endlessly for an unmanageable number of patterns. Moreover, we argue that the number of patterns for different thresholds is an interesting summary statistic of the data: the pattern frequency spectrum. To enable fast estimation of the number of frequent patterns, we adapt the classical algorithm by Knuth for estimating the size of a search tree. Although the method is known to be theoretically suboptimal, we demonstrate that in practice it not only produces very accurate estimates, but is also very efficient. Moreover, we introduce a small variation that can be used to estimate the number of patterns under constraints for which the Apriori property does not hold. The empirical evaluation shows that this approach obtains good estimates for closed itemsets. Finally, we show how the method, together with isotonic regression, can be used to quickly and accurately estimate the frequency pattern spectrum: the curve that shows the number of patterns for every possible value of the frequency threshold. Comparing such a spectrum to one that was constructed using a random data model immediately reveals whether the dataset contains any structure of interest.

16:30 – 16:50

A Fast Method of Statistical Assessment for Combinatorial Hypotheses Based on Frequent Itemset Enumeration

Shin-ichi Minato, Takeaki Uno, Koji Tsuda, Aika Terada, Jun Sese

In many scientific communities using experiment databases, one of the crucial problem is how to assess the statistical significance (P-value) of a discovered hypothesis. Especially, combinatorial hypothesis assessment is a hard problem because it requires a multiple-testing procedure with a very large factor of the P-value correction. Recently, Terada et al. proposed a novel method of the P-value correction, called ``Limitless Arity Multiple-testing Procedure'' (LAMP), which is based on frequent itemset enumeration to exclude meaninglessly infrequent itemsets which never be significant. The LAMP makes much more accurate P-value correction than previous method, and it empowers the scientific discovery. However, the original LAMP implementation is sometimes too time-consuming for practical databases. We propose a new LAMP algorithm that essentially executes itemset mining algorithm once, while the previous one executes many times. Our experimental results show that the proposed method is much (10 to 100 times) faster than the original LAMP. This algorithm enables us to discover significant P-value patterns in quite short time even for very large-scale databases.

16:50 – 17:10

Ranked Tiling

Thanh Le Van, Matthijs van Leeuwen, Siegfried Nijssen, Ana Carolina Fierro, Kathleen Marchal, Luc De Raedt

Tiling is a well-known pattern mining technique. Traditionally, it discovers large areas of ones in binary databases or matrices, where an area is defined by a set of rows and a set of columns. In this paper, we introduce the novel problem of ranked tiling, which is concerned with finding interesting areas in ranked data. In this data, each transaction defines a complete ranking of the columns. Ranked data occurs naturally in applications like sports or other competitions. It is also a useful abstraction when dealing with numeric data in which the rows are incomparable. We introduce a scoring function for ranked tiling, as well as an algorithm using constraint programming and optimization principles. We empirically evaluate the approach on both synthetic and real-life datasets, and demonstrate the applicability of the framework in several case studies. One case study involves a heterogeneous dataset concerning the discovery of biomarkers for different subtypes of breast cancer patients. An analysis of the tiles by a domain expert shows that our approach can lead to the discovery of novel insights.

17:10 – 17:30

A Lossless Data Reduction For Mining Constrained Patterns in n-ary Relations

Gabriel Poesia, Loic Cerf

Given a binary relation, listing the itemsets takes exponential time. The problem grows worse when searching for analog patterns defined in n-ary relations. However, real-life relations are sparse and, with a greater number n of dimensions, they tend to be even sparser. Moreover, not all itemsets are searched. Only those satisfying some user-defined constraints, such as minimal size constraints. This article proposes to exploit together the sparsity of the relation and the presence of constraints satisfying a common property, the monotonicity per dimension. It details a pre-processing step to identify and erase n-tuples whose removal does not change the collection of patterns to be discovered. That reduction of the relation is achieved in a time and a space that is linear in the number of n-tuples. Experiments on two real-life datasets show that, whatever the algorithm used afterward to actually list the patterns, the pre-process allows to lower the overall running time by a factor typically ranging from 10 to 100.

17:30 – 17:50

Uncovering the Plot: Detecting Surprising Coalitions of Entities in Multi-Relational Schemas

Hao Wu, Jilles Vreeken, Nikolaj Tatti, Naren Ramakrishnan

Many application domains such as intelligence analysis and cybersecu- 2 rity require tools for the unsupervised identification of suspicious entities in multi- 3 relational/network data. In particular, there is a need for automated semi-automated 4 approaches to 'uncover the plot', i.e., to detect non-obvious coalitions of entities bridg- 5 ing many types of relations. We cast the problem of detecting such suspicious coalitions 6 and their connections as one of mining surprisingly dense and well-connected chains of biclusters over multi-relational data. With this as our goal, we model data by the Maxi-8 mum Entropy principle, such that in a statistically well-founded way we can gauge the 9 surprisingness of a discovered bicluster chain with respect to what we already know. 10 We design an algorithm for approximating the most informative multi-relational pat- 11 terns, and provide strategies to incrementally organize discovered patterns into the 12 background model. We illustrate how our method is adept at discovering the hidden 13 plot in multiple synthetic and real-world intelligence analysis datasets. Our approach 14 naturally generalizes traditional attribute-based maximum entropy models for single 15 relations, and further supports iterative, human-in-the-loop, knowledge discovery.

Session: Demo Track

Room: Grand Foyer (spotlight in 103-104)

BestTime: Finding Representatives in Time Series Datasets

Stephan Spiegel, David Schultz, Sahin Albayrak

Given a set of time series, we aim at finding representatives which best comprehend the recurring temporal patterns contained in the data. We demonstrate BestTime, a Matlab application that uses recurrence quantification analysis to find time series representatives.

GrammarViz 2.0: a tool for grammar-based pattern discovery in time series

Pavel Senin, Jessica Lin, Xing Wang, Tim Oates, Manfred Lerner, Arnold Boedihardjo, Crystal Chen, Susan Frankenstein, Sunil Gandhi

The problem of frequent and anomalous patterns discovery in time series has received a lot of attention in the past decade. Addressing the common limitation of existing techniques, which require a pattern length to be known in advance, we recently proposed grammar-based algorithms for efficient discovery of variable length frequent and rare patterns. In this paper we present GrammarViz 2.0, an interactive tool that, based on our previous work, implements algorithms for grammar-driven mining and visualization of variable length time series patterns.

KnowNow: a Serendipity-based Educational Tool for Learning Time-Linked Knowledge

Luigi Di Caro, Livio Robaldo, Nicoletta Bersia

In this paper we present the system KnowNow, a tool whose aim is to let the users navigate into text corpora through dynamic semantic information networks, created in real-time according to delimited time ranges. In educational scenarios, students are often asked to write short essays on different topics linked by temporal information. This usually involves a combination of several aspects to be avaluated, such as knowledge, imagination, structure and presentation. In the light of this, the introduction of Natural Language Understanding techniques together with cross-topic navigation and visualization tools and can considerably help students to retrieve, link, and create well-structured and original contributions, as we demonstrate by using KnowNow.

Spa: a web-based viewer for text mining in Evidence Based Medicine

Joël Kuiper, Iain Marshall, Byron Wallace, Morris Swertz

Summarizing the evidence about medical interventions is an immense undertaking, in part because unstructured PDF documents remain the main vehicle for disseminating scientific findings. Clinicians and researchers must therefore manually extract and synthesise information from these documents. We introduce Spá, a web-based viewer that enables automated annotation and summarisation of PDFs via machine learning. To illustrate its functionality, we use Spá to semi-automate the assessment of bias in clinical trials. Spá has a modular architecture, therefore the tool may be widely useful in other domains with a PDF-based literature, including law, physics, and biology.

Khiops CoViz: a tool for visual exploratory analysis of k-coclustering results

Bruno Guerraz, Marc Boullé, Dominique Gay, Fabrice Clérot

Identifying and visually analyzing interesting interactions between variables in large-scale data sets through k-coclustering is of high importance. We present Khiops CoViz, a tool for visual analysis of interesting relationships between two or more variables (categorical and/or numerical). The visualization of k variables coclustering takes the form of a grid/matrix whose dimensions are partitioned: categorical variables are grouped into clusters and numerical variables are discretized. The tool allows several kinds of visualization at various scales for grid representation of coclustering results by means of several criteria each of which providing different insights into the data.

MinUS: Mining User Similarity with Trajectory Patterns

Jun Pang, Xihui Chen, Piotr Kordy, Ruipeng Lu

The development of positioning systems and wireless connectivity has made it possible to collect users' fine-grained movement data. This availability of movement data can be applied in a broad range of services. In this paper, we present a novel tool for calculating users' similarity based on their movements. This tool, MinUS, integrates the technologies of trajectory pattern mining with the state-of-the-art research on discovering user similarity with trajectory patterns. Specifically, with MinUS, we provide a platform to manage movement datasets, and construct and compare users' trajectory patterns. Tool users can compare results given by a series of user similarity metrics, which allows them to learn the importance and limitations of different similarity metrics and promotes studies in related areas, e.g., location privacy. Additionally, MinUS can also be used by researchers as a tool for preliminary process of movement data and parameter tuning in trajectory pattern mining.

WebDR: A Web Workbench for Data Reduction

Stefanos Ougiaroglou, Georgios Evangelidis

Data reduction is a common preprocessing task in the context of the k nearest neighbour classification. This paper presents WebDR, a web-based application where several data reduction techniques have been integrated and can be executed on-line. WebDR allows the performance evaluation of the classification process through a web interface. Therefore, it can be used by the academia for educational and experimental purposes.

BMaD – A Boolean Matrix Decomposition Framework

Andrey Tyukin, Stefan Kramer, Jörg Wicker

Boolean matrix decomposition is a method to obtain a compressed representation of a matrix with Boolean entries. We present a modular framework that unifies several Boolean matrix decomposition algorithms, and provide methods to evaluate their performance. The main advantages of the framework are its modular approach and hence the flexible combination of the steps of a Boolean matrix decomposition and the capability of handling missing values. The framework is licensed under the GPLv3 and can be downloaded freely at http://projects.informatik.uni-mainz.de/bmad .

Branty: a social media ranking tool for brands

Alexandros Arvanitidis, Anna Serafi, Athina Vakali, Grigorios Tsoumakas

In the competitive world of popular brands, strong presence in social media is of major importance for customer engagement and products advertising. Up to now, many such tools and applications en- able end-users to observe and monitor their company's web pro_le, their statistics, as well as their market outreach and competition status. This work goes beyond the individual brands statistics since it automates a brand ranking process based on opinions emerging in social media users' posts. Twitter streaming API is exploited to track micro-blogging activ- ity for a number of famous brands with emphasis on users' opinions and interactions. The social impact is captured from 3 di_erent perspectives (objective counts, opinion reckoning, inuence analysis), which estimate a score assigned to each brand via a multi-criteria algorithm. The results are then exposed in a Web application as a list of the most social brands on Twitter. But, are conventional metrics, such as followers, enough in order to measure the social impact of a brand? Di_erent usage scenar- ios of our application reveal that the social presence of a brand is more complex than current social impact frameworks care to admit.

Propositionalization Online

Nada Lavrac, Matic Perovšek, Anže Vavpetič

Inductive Logic Programming and Relational Data Mining address the task of inducing models or patterns from multi-relational data. An established relational data mining approach is propositionalization, characterized by transforming a relational database into a single-table representation. The paper presents a propositionalization toolkit implemented in the web-based data mining platform ClowdFlows. As a contemporary integration platform it enables workflow construction and execution, provides open access to Aleph, RSD, RelF and Wordification feature construction engines, and enables RDM performance comparison through cross-validation and ViperCharts results visualization.

PYTHIA: Employing Lexical and Semantic Features for Sentiment Analysis

Ioannis Katakis, Iraklis Varlamis, George Tsatsaronis

Sentiment analysis methods aim at identifying the polarity of a piece of text, e.g., passage, review, snippet, by analyzing lexical features at the level of the terms or the sentences. However, many of the previous works do not utilize features that can offer a deeper understanding of the text, e.g., negation phrases. In this work we demonstrate a novel piece of software, namely PYTHIA3, which combines semantic and lexical features at the term and sentence level and integrates them into machine learning models in order to predict the polarity of the input text. Experimental evaluation of PYTHIA in a benchmark movie reviews dataset shows that the suggested combination performs favorably against previous related methods. An online demo is publicly available: http://omiotis.hua.gr/pythia .

Interactive Medical Miner: Interactively exploring subpopulations in epidemiological datasets

Uli Niemann, Myra Spiliopoulou, Henry Völzke, Jens-Peter Kühn

We present our Interactive Medical Miner, a tool for classification and model drill-down, designed to study epidemiological data. Our tool encompasses supervised learning (with decision trees and classification rules), utilities for data selection, and a rich panel with options for inspecting individual classification rules, and for studying the distribution of variables in each of the target classes. Since some of the epidemiological data available to the medical researcher may be still unlabeled (e.g. because the medical recordings for some part of the cohort are still in progress), our Interactive Medical Miner also supports the juxtaposition of labeled and unlabeled data. The set of methods and scientific workflow supported with our tool have been published in [1].

Insight4News: Connecting News to Relevant Social Conversations

Georgiana Ifrim, Bichen Shi, Neil Hurley

We present the Insight4News system that connects news articles to social conversations, as echoed in microblogs such as Twitter. Insight4News tracks feeds from mainstream media, e.g., BBC, Irish Times, and extracts relevant topics that summarize the tweet activity around each article, recommends relevant hashtags, and presents complementary views and statistics on the tweet activity, related news articles, and timeline of the story with regard to Twitter reaction. The user can track their own news article or a topic-focused Twitter stream. While many systems tap on the social knowledge of Twitter to help users stay on top of the information wave, none is available for connecting news to relevant Twitter content on a large scale, in real time, with high precision and recall. Insight4News builds on our award winning Twitter topic detection approach and several machine learning components, to deliver news in a social context. Keywords: news tracking, social media, Twitter, summarization.

WEDNESDAY 17 SEPTEMBER 2014

WEDNESDAY INVITED TALK

Beyond stochastic gradient descent for large-scale machine learning

Speaker: Francis Bach
Time: 09:00 – 10:00
Room: Auditorium 850

Abstract
Many machine learning and signal processing problems are traditionally cast as convex optimization problems. A common difficulty in solving these problems is the size of the data, where there are many observations ("large n") and each of these is large ("large p"). In this setting, online algorithms such as stochastic gradient descent which pass over the data only once, are usually preferred over batch algorithms, which require multiple passes over the data. In this talk, I will show how the smoothness of loss functions may be used to design novel algorithms with improved behavior, both in theory and practice: in the ideal infinite-data setting, an efficient novel Newton-based stochastic approximation algorithm leads to a convergence rate of O(1/n) without strong convexity assumptions, while in the practical finite-data setting, an appropriate combination of batch and online algorithms leads to unexpected behaviors, such as a linear convergence rate for strongly convex problems, with an iteration cost similar to stochastic gradient descent. (joint work with Nicolas Le Roux, Eric Moulines and Mark Schmidt).

Bio
Francis Bach is a researcher at INRIA, leading since 2011 the SIERRA project-team, which is part of the Computer Science Laboratory at Ecole Normale Superieure. He completed his Ph.D. in Computer Science at U.C. Berkeley, working with Professor Michael Jordan, and spent two years in the Mathematical Morphology group at Ecole des Mines de Paris, then he joined the WILLOW project-team at INRIA/Ecole Normale Superieure from 2007 to 2010. Francis Bach is interested in statistical machine learning, and especially in graphical models, sparse methods, kernel-based learning, convex optimization vision and signal processing.

WEDNESDAY INDUSTRY INVITED TALK

Machine Learning and Data Mining in Call of Duty

Speaker: Arthur von Eschen
Time: 14:20 – 15:05
Room: Auditorium 850

Abstract
Data science is relatively new to the video game industry, but it has quickly emerged as one of the main resources for ensuring game quality. At Activision, we leverage data science to analyze the behavior of our games and our players to improve in-game algorithms and the player experience. We use machine learning and data mining techniques to influence creative decisions and help inform the game design process. We also build analytic services that support the game in real-time; one example is a cheating detection system which is very similar to fraud detection systems used for credit cards and insurance. This talk will focus on our data science work for Call of Duty, one of the bestselling video games in the world.

Bio
Arthur Von Eschen is Senior Director of Game Analytics at Activision. He and his team are responsible for analytics work that supports video game design on franchises such as Call of Duty and Skylanders. In addition to holding a PhD in Operations Research, Arthur has over 15 years of experience in analytics consulting and R&D with the U.S. Fortune 500. His work has spanned across industries such as banking, financial services, insurance, retail, CPG and now interactive entertainment (video games). Prior to Activision he worked at Fair Isaac Corporation (FICO). Before FICO he ran his own analytics consulting firm for six years.

Algorithms, Evolution and Network-Based Approaches in Molecular Discovery

Speaker: Mike Bodkin
Time: 15:05 – 15:50
Room: Auditorium 850

Abstract
Drug research generates huge quantities of data around targets, compounds and their effects. Network modelling can be used to describe such relationships with the aim to couple our understanding of disease networks with the changes in small molecule properties. This talk will build off of the data that is routinely captured in drug discovery and describe the methods and tools that we have developed for compound design using predictive modelling, evolutionary algorithms and network-based mining.

Bio
Mike did his PhD in protein de-novo design for Nobel laureate sir James Black before taking up a fellowship in computational drug design at Cambridge University. He moved to AstraZeneca as a computational chemist before joining Eli Lilly in 2000. Since 2003 he was head of the computational drug discovery group at Lilly but recently jumped ship to Evotec to work as the VP for computational chemistry and cheminformatics. His research aims are to continue to develop new algorithms and software in the fields of drug discovery and systems informatics and to deliver and apply current and novel methods as tools for use in drug research.

WEDNESDAY SESSIONS AT A GLANCE

Nectar Session 1:

Room: 102 **Chair: Pierre Geurts**

10:10 – 10:40 **Sampling-based Data Mining Algorithms: Modern Techniques and Case Studies**
Matteo Riondato

10:40 – 11:10 **Machine Learning Approaches for Metagenomics**
Huzefa Rangwala, Anveshi Charuvaka, Zeehasham Rasheed

Session 12: Optimization and approximation

Room: 105 **Chair: Ricard Gavaldà**

10:10 – 10:30 **Convergence of Min-Sum-Min Message-Passing for Quadratic Optimization**
Guoqiang Zhang, Richard Heusdens

10:30 – 10:50 **Error-bounded Approximations for Infinite-horizon Discounted Decentralized POMDPs**
Jilles Dibangoye, Olivier Buffet, Francois Charpillet

10:50 – 11:10 **Fast LSTD using stochastic approximation: Finite time analysis and application to traffic control**
Prashanth L.A, Nathaniel Korda, Remi Munos

Session 13: Kernel-based learning

Room: 106 **Chair: Thomas Gaertner**

10:10 – 10:30 **Kernel Alignment Inspired Linear Discriminant Analysis**
Shuai Zheng, Chris Ding

10:30 – 10:50 **Deconstructing Kernel Machines**
Mohsen Ali

10:50 – 11:10 **Preventing Over-Fitting of Cross-Validation with Kernel Stability**
Yong Liu, Shali Jiang, Shizhong Liao

Session 14: Community detection

Room: 103-104 **Chair: Giuseppe Manco**

10:10 – 10:30 **Overlapping community detection in labeled graphs**
Esther Galbrun, Aristides Gionis, Nikolaj Tatti

10:30 – 10:50 **Beyond Blocks: Hyperbolic Community Detection**
Miguel Araujo, Stephan Günnemann, Gonzalo Mateos, Christos Faloutsos

10:50 – 11:10 **Discovering dynamic communities in interaction networks**
Polina Rozenshtein, Nikolaj Tatti, Aristides Gionis

Session 15: Data factorization

Room: 102 **Chair: Jan Ramon**

11:40 – 12:00 **Bayesian Multi-View Tensor Factorization**
Suleiman Khan, Samuel Kaski

12:00 – 12:20 **Invariant Time-Series Factorization**
(Josif Grabocka, Lars Schmidt-Thieme)

12:20 – 12:40 **SAGA: Sparse And Geometry-Aware non-negative matrix factorization through non-linear local embedding**
Nicolas Courty, Xing Gong, Jimmy Vandel, Thomas Burger

12:40 – 13:00 **Scalable Nonnegative Matrix Factorization with Block-wise Updates**
Jiangtao Yin, Lixin Gao, Zhongfei Zhang

Session 16: Bandits

Room: 103-104 **Chair: Eyke Hüllermeier**

11:40 – 12:00 **Sub-sampling for multi-armed bandits**
Akram Baransi, Odalric-Ambrym Maillard, Shie Mannor

12:00 – 12:20 **Concurrent bandits and cognitive radio networks**
Orly Avner, Shie Mannor

12:20 – 12:40 **Experimental design in dynamical system identification: a bandit-based active learning approach**
Artemis Llamosi, Adel Mezine, Florence D'Alché-Buc, Veronique Letort, Michele Sebag

12:40 – 13:00 **Infinitely Many-Armed Bandits with Unknown Value Distribution**
Yahel David, Nahum Shimkin

Session 17: Spatial and temporal data

Room: 105 **Chair: Stefan Wrobel**

11:40 – 12:00 **L everaging the power of local spatial autocorrelation in geophysical interpolative clustering**
Annalisa Appice, Donato Malerba

12:00 – 12:20 **Detecting Localized Homogeneous Anomalies over Spatio-Temporal Data**
Aditya Telang, Deepak P, Salil Joshi, Prasad Deshpande, Ranjana Rajendran

12:20 – 12:40 **Confidence bands for time series data**
Jussi Korpela, Kai Puolamäki, Aristides Gionis

12:40 – 13:00 **Nowcasting with numerous candidate predictors**
Brendan Duncan, Charles Elkan

Session 18: Text Mining

Room: 106 **Chair: Toon Calders**

11:40 – 12:00 **Joint Prediction of Topics in a URL Hierarchy**
Michael Großhans, Christoph Sawade, Tobias Scheffer, Niels Landwehr

12:00 – 12:20 **Clustering Image Search Results by Entity Disambiguation**
Kaiqi Zhao, Zhiyuan Cai, Qingyu Sui, Enxun Wei, Kenny Zhu

12:20 – 12:40 **How Many Topics? Stability Analysis for Topic Models**
Derek Greene, Derek O'Callaghan, Pádraig Cunningham

12:40 – 13:00 **Open-domain Question Answering with Weakly Supervised Embedding Models**
Antoine Bordes, Jason Weston, Nicolas Usunier

Nectar Session 2:

Room: 103-104 **Chair: Pierre Geurts**

14:20 – 14:50 **Active Learning is Planning: Nonmyopic ε-Bayes-Optimal Active Learning of Gaussian Processes**
Trong Nghia Hoang, Bryan Kian Hsiang Low, Patrick Jaillet, Mohan Kankanhalli

14:50 – 15:20 **Heterogeneous Stream Processing and Crowdsourcing for Traffic Monitoring: Highlights**
Francois Schnitzler, Alexander Artikis, Matthias Weidlich, Ioannis Boutsis, Thomas Liebig, Nico Piatkowski, Christian Bockermann, Katharina Morik, Vana Kalogeraki, Jakub Marecek, Avigdor Gal, Shie Mannor, Dermot Kinane, Dimitrios Gunopulos

15:20 – 15:50 **Distributional Clauses Particle Filter**
Davide Nitti, Tinne De Laet, Luc De Raedt

Session 19: Neural structures and deep learning

Room: 105-106 **Chair: Patrick Gallinari**

16:20 – 16:40 **Self-Organizing Maps by Difference of Convex functions optimization**
Hoai An Le Thi, Manh Cuong Nguyen

16:40 – 17:00 **On the Equivalence Between Deep NADE and Generative Stochastic Networks**
Li Yao, Sherjil Ozair, Kyunghyun Cho, Yoshua Bengio

17:00 – 17:20 **Knowledge-Powered Deep Learning for Word Embedding**
Jiang Bian, Bin Gao, Tie-Yan Liu

17:20 – 17:40 **Learned-Norm Pooling for Deep Feedforward and Recurrent Neural Networks**
Caglar Gulcehre, Kyunghyun Cho, Razvan Pascanu, Yoshua Bengio

Session 20: Reinforcement learning

Room: 102 **Chair: Michele Sebag**

16:20 – 16:40 **Policy Search for Path Integral Control**
Vicenç Gomez, Jan Peters, Hilbert Kappen, Gerhard Neumann

16:40 – 17:00 **An Online Policy Gradient Algorithm for Markov Decision Processes with Continuous States and Actions**
Yao Ma, Tingting Zhao, Kohei Hatano, Masashi Sugiyama

17:00 – 17:20 **Boosted Bellman Residual Minimization Handling Expert Demonstrations**
Bilal Piot, Matthieu Geist, Olivier Pietquin

17:20 – 17:40 **Local Policy Search in a Convex Space and Conservative Policy Iteration as Boosted Policy Search**
Bruno Scherrer, Matthieu Geist

Session 21: Recommendation systems and dyadic data

Room: 103-104 **Chair: Martin Atzmueller**

16:20 – 16:40 **A Bayesian Generative Model for Item and User Recommendation in Social Rating Networks with Trust Relationships**
Gianni Costa, Giuseppe Manco, Riccardo Ortale

16:40 – 17:00 **Collaborative Filtering with Information-Rich and Information-Sparse Entities**
Kai Zhu, Rui Wu, Lei Ying, R. Srikant

17:00 – 17:20 **A Constrained Matrix-Variate Gaussian Process for Transposable Data**
Oluwasanmi Koyejo, Cheng Lee, Joydeep Ghosh

17:20 – 17:40 **A two-step learning approach for solving full and almost full cold start problems in dyadic prediction**
Tapio Pahikkala, Michiel Stock, Antti Airola, Tero Aittokallio, Bernard De Baets, Willem Waegeman

Session 22: Data stream mining

Room: 101 **Chair: Myra Spiliopoulou**

16:20 – 16:40 **Classy: Fast Clustering Streams of Call-Graphs**
Orestis Kostakis

16:40 – 17:00 **Classifying A Stream of Infinite Concepts: A Bayesian Non-Parametric Approach**
Abbas Hosseini, Hamid R. Rabiee, Hassan Hafez, Ali Soltani-Farani

17:00 – 17:20 **Speeding Up Recovery From Concept Drifts**
Silas Santos, Paulo Gonçalves, Geyson Silva, Roberto Barros

17:20 – 17:40 **Mining Top-K Largest Tiles in a Data Stream**
Hoang Thanh Lam, Wenjie Pei, Adriana Prado, Baptiste Jeudy, Elisa Fromont

WEDNESDAY SESSIONS, WITH ABSTRACTS

Nectar Session 1:

Room: 102 **Chair: Pierre Geurts**

10:10 – 10:40

Sampling-based Data Mining Algorithms: Modern Techniques and Case Studies

Matteo Riondato

Sampling a dataset for faster analysis and looking at it as a sample from an unknown distribution are two faces of the same coin. We discuss the use of modern techniques involving the Vapnik-Chervonenkis (VC) dimension to study the trade-off between sample size and accuracy of data mining results that can be obtained from a sample. We report two case studies where we and collaborators employed these techniques to develop efficient sampling-based algorithms for the problems of betweenness centrality computation in large graphs and extracting statistically significant Frequent Itemsets from transactional datasets.

10:40 – 11:10

Machine Learning Approaches for Metagenomics

Huzefa Rangwala, Anveshi Charuvaka, Zeehasham Rasheed

Microbes exists everywhere. Current generation of genomic technologies have allowed researchers to determine the collective DNA sequence of all microorganisms co-existing together. In this paper, we present some of the challenges related to the analysis of data obtained from the community genomics experiment (commonly referred by metagenomics), advocate the need of machine learning techniques and highlight our contributions related to development of supervised and unsupervised techniques for solving this complex, real world problem.

Session 12: Optimization and approximation

Room: 105 **Chair: Ricard Gavaldà**

10:10 – 10:30

Convergence of Min-Sum-Min Message-Passing for Quadratic Optimization

Guoqiang Zhang, Richard Heusdens

We propose a new message-passing algorithm for the quadratic optimization problem. As opposed to the min-sum algorithm, the new algorithm involves two minimizations and one summation at each iteration. The new min-sum-min algorithm exploits feedback from last iteration in generating new messages, resembling the Jacobi-relaxation algorithm. We show that if the feedback signal is large enough, the min-sum-min algorithm is guaranteed to converge to the optimal solution. Experimental results show that the min-sum-min algorithm outperforms two reference methods w.r.t. the convergence speed.

10:30 – 10:50

Error-bounded Approximations for Infinite-horizon Discounted Decentralized POMDPs

Jilles Dibangoye, Olivier Buffet, Francois Charpillet

We address decentralized stochastic control problems represented as decentralized partially observable Markov decision processes (Dec-POMDPs). This formalism provides a general model for decision-making under uncertainty in cooperative, decentralized settings, but the worst-case complexity makes it difficult to solve optimally (NEXP-complete). Recent advances suggest recasting Dec-POMDPs into continuous-state and deterministic MDPs. In this form, however, states and actions are embedded into high-dimensional spaces, making accurate estimate of states and greedy selection of actions intractable for all but trivial-sized problems. The primary contribution of this paper is the first framework for error-monitoring during approximate estimation of states and selection of actions. Such a framework permits us to convert state-of-the-art exact methods into error-bounded algorithms, which results in a scalability increase as demonstrated by experiments over problems of unprecedented sizes.

10:50 – 11:10

Fast LSTD using stochastic approximation: Finite time analysis and application to traffic control

Prashanth L.A, Nathaniel Korda, Remi Munos

We propose a stochastic approximation based method with randomization of samples for policy evaluation using the least squares temporal difference (LSTD) algorithm. Our method results in an O(d) improvement in complexity in comparison to vanilla LSTD, where d is the dimension of the data. We provide convergence rate results for our proposed method, both in high probability and in expectation. Moreover, we also establish that using our scheme in place of LSTD does not impact the rate of convergence of the approximate value function to the true value function. This result coupled with the low complexity of our method makes it attractive for implementation in big data settings, where d is large. Further, we also analyse a similar low-complexity alternative for least squares regression and provide finite-time bounds there. We demonstrate the practicality of our method for LSTD empirically by combining it with the LSPI algorithm in a traffic signal control application.

Session 13: Kernel-based learning

Room: 106 **Chair: Thomas Gaertner**

10:10 – 10:30

Kernel Alignment Inspired Linear Discriminant Analysis

Shuai Zheng, Chris Ding

Kernel alignment measures the degree of similarity between two kernels. In this paper, inspired from kernel alignment, we propose a new Linear Discriminant Analysis (LDA) formulation, kernel alignment LDA (kaLDA). We first define two kernels, data kernel and class indicator kernel. The problem is to find a subspace to maximize the alignment between subspace-transformed data kernel and class indicator kernel. Surprisingly, the kernel alignment induced kaLDA objective function is very similar to classical LDA and can be expressed using between-class and total scatter matrices. This can be extended to multi-label data. We use a Stiefel-manifold gradient descent algorithm to solve this problem. We perform experiments on 8 single-label and 6 multi-label data sets. Results show that kaLDA has very good performance on many single-label and multi-label problems.

10:30 – 10:50

Deconstructing Kernel Machines

Mohsen Ali

This paper studies the following problem: Given an SVM (kernel)-based binary classifier $\bC$ as a black-box oracle, how much can we learn of its internal working by querying it? Specifically, we assume the feature space $\RR^d$ is known and the kernel machine has m support vectors such that $d > m$ (or $d >> m$), and in addition, the classifier $\bC$ is laconic in the sense that for a feature vector, it only provides a predicted label (± 1) without divulging other information such as margin or confidence level. We formulate the the problem of understanding the inner working of $\bC$ as characterizing the decision boundary of the classifier, and we introduce the simple notion of bracketing to sample points on the decision boundary within a prescribed accuracy. For the five most common types of kernel function, linear, quadratic and cubic polynomial kernels, hyperbolic tangent kernel and Gaussian kernel, we show that with O(*dm*) number of queries, the type of kernel function and the (kernel) subspace spanned by the support vectors can be determined. In particular, for polynomial kernels, additional O(m^3) queries are sufficient to reconstruct the entire decision boundary, providing a set of quasi-support vectors that can be used to efficiently evaluate the deconstructed classifier. We speculate briefly on the future application potential of deconstructing kernel machines and we present experimental results validating the proposed method. A simple user-friendly MATLAB implementation has been submitted as supplemental material for review.

10:50 – 11:10

Preventing Over-Fitting of Cross-Validation with Kernel Stability

Yong Liu, Shali Jiang, Shizhong Liao

Kernel selection is critical to kernel methods. Cross-validation (CV) is a widely accepted kernel selection method. However, the CV based estimates generally exhibit a relatively high variance and are therefore prone to over-fitting. In order to prevent the high variance, we first propose a novel version of stability, called kernel stability. This stability quantifies the perturbation of the kernel matrix with respect to the changes in the training set. Then we establish the connection between the kernel stability and variance of CV. By restricting the derived upper bound of the variance, we present a kernel selection criterion, which can prevent the high variance of CV and hence guarantee good generalization performance. Furthermore, we derive a closed form for the estimate of the kernel stability, making the criterion based on the kernel stability computationally efficient. Theoretical analysis and experimental results demonstrate that our criterion is sound and effective.

Session 14: Community detection

Room: 103-104 **Chair: Giuseppe Manco**

10:10 – 10:30

Overlapping community detection in labeled graphs

Esther Galbrun, Aristides Gionis, Nikolaj Tatti

We present a new approach for the problem of finding overlapping communities in graphs and social networks. Our approach consists of a novel problem definition and three accompanying algorithms. We are particularly interested in graphs that have labels on their vertices, although our methods are also applicable to graphs with no labels. Our goal is to find k communities so that the total edge density over all k communities is maximized. In the case of labeled graphs, we require that each community is succinctly described by a set of labels. This requirement provides a better understanding for the discovered communities. The proposed problem formulation leads to the discovery of vertex-overlapping and dense communities that cover as many graph edges as possible. We capture these properties with a simple objective function, which we solve by adapting efficient approximation algorithms for the generalized maximum-coverage problem and the densest-subgraph problem. Our proposed algorithm is a generic greedy scheme. We experiment with three variants of the scheme, obtained by varying the greedy step of finding a dense subgraph. We validate our algorithms by comparing with other state-of-the-art community- detection methods on a variety of performance measures. Our experiments confirm that our algorithms achieve results of high quality in terms of the reported measures, and are practical in terms of performance.

10:30 – 10:50

Beyond Blocks: Hyperbolic Community Detection

Miguel Araujo, Stephan Günnemann, Gonzalo Mateos, Christos Faloutsos

What do real communities in social networks look like? Community detection plays a key role in understanding the structure of real-life graphs with impact on recommendation systems, load balancing and routing. Previous community detection methods look for uniform blocks in adjacency matrices. However, after studying 4 real networks with ground-truth communities, we provide empirical evidence that communities are best represented as having an hyperbolic structure. We detail HyCoM - the Hyperbolic Community Model - as a better representation of communities and the relationships between their members, and show improvements in compression compared to standard methods. We also introduce HyCoM-FIT, a fast, parameter free algorithm to detect communities with hyperbolic structure. We show that our method is effective in finding communities with a similar structure to self-declared ones. We report findings in real social networks, including a community in a blogging platform with over 34 million edges in which more than 1000 users established over 300000 relations.

10:50 – 11:10

Discovering dynamic communities in interaction networks

Polina Rozenshtein, Nikolaj Tatti, Aristides Gionis

Very often online social networks are defined by aggregating information regarding the interaction between the nodes of the network. For example, a call graph is defined by considering an edge for each pair of individuals who have called each other at least once; or at least k times. Similarly, an implicit social network in a social-media site is defined by considering an edge for each pair of users who have interacted in some way, e.g., have made a conversation, commented to each other's content, etc. Despite the fact that this type of definitions have been used to obtain a lot of insights regarding the structure of social networks, it is obvious that they suffer from a severe limitation: they neglect the precise time that the interaction between network nodes occurs. In this paper we propose to study interaction networks, where one considers not only the underlying topology of the social network, but also the exact time instances that nodes interact. In an interaction network an edge is associated with a time stamp, and multiple edges may occur for the same pair of nodes. Consequently, interaction networks offer a rich fine-grained representation that can be used to reveal dynamic phenomena in the network. In the context of interaction networks, we study the problem of discovering communities, which are dense, and whose edges occur in short time intervals. Such communities represent groups of individuals who interact with each other in some specific time instances, for example, a group of employees working on the same project and whose interaction intensifies before certain milestones of the project. We prove that the problem we define is NP-hard, and we provide effective algorithms by adapting techniques used to find dense subgraphs. We perform extensive evaluation of the proposed methods on synthetic and real datasets, which demonstrates the validity of our concepts and the good performance of our algorithms.

Session 15: Data factorization

Room: 102 **Chair: Jan Ramon**

11:40 – 12:00

Bayesian Multi-View Tensor Factorization

Suleiman Khan, Samuel Kaski

We introduce a Bayesian extension of the tensor factorization problem to multiple coupled tensors. For a single tensor it reduces to standard PARAFAC-type Bayesian factorization, and for two tensors it is the first Bayesian Tensor Canonical Correlation Analysis method. It can also be seen to solve a tensorial extension of the recent Group Factor Analysis problem. The method decomposes the set of tensors to factors shared by subsets of the tensors, and factors private to individual tensors, and does not assume orthogonality. For a single tensor, the method empirically outperforms existing methods, and we demonstrate its performance on multiple tensor factorization tasks in toxicogenomics and functional neuroimaging.

12:00 – 12:20

Invariant Time-Series Factorization

Josif Grabocka, Lars Schmidt-Thieme

Time-series analysis is an important domain of machine learning and a plethora of methods have been developed for the task. This paper proposes a new representation of time series, which in contrast to existing approaches, decomposes a time-series dataset into latent patterns and membership weights of local segments to those patterns. The process is formalized as a constrained objective function and a tailored stochastic coordinate descent optimization is applied. The time-series are projected to a new feature representation consisting of the sums of the membership weights, which captures frequencies of local patterns. Features from various sliding window sizes are concatenated in order to encapsulate the interaction of patterns from different sizes. The derived representation offers a set of features that boosts classification accuracy. Finally, a large-scale experimental comparison against 11 baselines over 43 real life datasets, indicates that the proposed method achieves state-of-the-art prediction accuracy results.

12:20 – 12:40

SAGA: Sparse And Geometry-Aware non-negative matrix factorization through non-linear local embedding

Nicolas Courty, Xing Gong, Jimmy Vandel, Thomas Burger

This paper presents a new non-negative matrix factorization technique which (1) allows the decomposition of the original data on multiple latent factors accounting for the geometrical structure of the manifold embedding the data; (2) provides an optimal representation with a controllable level of sparsity; (3) has an overall linear complexity allowing handling in tractable time large and high dimensional datasets. It operates by coding the data with respect to local neighbors with non-linear weights. This locality is obtained as a consequence of the simultaneous sparsity and convexity constraints. Our method is demonstrated over several experiments, including a feature extraction and classification task, where it achieves better performances than the state-of-the-art factorization methods, with a shorter computational time.

12:40 – 13:00

Scalable Nonnegative Matrix Factorization with Block-wise Updates

Jiangtao Yin, Lixin Gao, Zhongfei Zhang

Nonnegative Matrix Factorization (NMF) has been applied with great success to many applications. As NMF is applied to massive datasets such as web-scale dyadic data, it is desirable to leverage a cluster of machines to speed up the factorization. However, it is challenging to efficiently implement NMF in a distributed environment. In this paper, we show that by leveraging a new form of update functions, we can perform local aggregation and fully explore parallelism. Moreover, under the new form of update functions, we can perform frequent updates, which aim to use the most recently updated data whenever possible. As a result, frequent updates are more efficient than their traditional concurrent counterparts. Through a series of experiments on a local cluster as well as the Amazon EC2 cloud, we demonstrate that our implementation with frequent updates is up to two orders of magnitude faster than the existing implementation with the traditional form of update functions.

Session 16: Bandits

Room: 103-104 **Chair: Eyke Hüllermeier**

11:40 – 12:00

Sub-sampling for multi-armed bandits

Akram Baransi, Odalric-Ambrym Maillard, Shie Mannor

The stochastic multi-armed bandit problem is a popular model of the exploration/exploitation trade-off in sequential decision problems. We introduce a novel algorithm that is based on sub-sampling. Despite its simplicity, we show that the algorithm demonstrates excellent empirical performances against state-of-the-art algorithms, including Thompson sampling and KL-UCB. The algorithm is very flexible, it does need to know a set of reward distributions in advance nor the range of the rewards. It is not restricted to Bernoulli distributions and is also invariant under rescaling of the rewards. We provide a detailed experimental study comparing the algorithm to the state of the art, the main intuition that explains the striking results, and conclude with a finite-time regret analysis for this algorithm in the simplified two-arm bandit setting.

12:00 – 12:20

Concurrent bandits and cognitive radio networks

Orly Avner, Shie Mannor

We consider the problem of multiple users targeting the arms of a single multi-armed stochastic bandit. The motivation for this problem comes from cognitive radio networks, where selfish users need to coexist without any side communication between them, implicit cooperation or common control. Even the number of users may be unknown and can vary as users join or leave the network. We propose an algorithm that combines an ε-greedy learning rule with a collision avoidance mechanism. We analyze its regret with respect to the system-wide optimum and show that sub-linear regret can be obtained in this setting. Experiments show dramatic improvement comparing to other algorithms for this setting.

12:20 – 12:40

Experimental design in dynamical system identification: a bandit-based active learning approach

Artemis Llamosi, Adel Mezine, Florence D'Alché-Buc, Veronique Letort, Michele Sebag

This paper is interested in dynamical system identification, with the reverse modeling of a gene regulatory network as motivating application. An active learning approach is used to iteratively select the most informative experiments needed to improve the parameters and hidden variables estimates in a dynamical model given a budget of experiments. The design of experiments under these budgeted resources is formalized in terms of sequential optimization. A local optimization criterion (reward) is designed to assess each experiment in the sequence, and the global optimization of the sequence is tackled in a game-inspired setting, within the Upper Confidence Tree framework combining Monte-Carlo tree-search and multi-armed bandits. The approach, called EDEN for Experimental Design for parameter Estimation in a Network, shows very good performances on several realistic simulated problems of gene regulatory network reverse-modeling, inspired from the international challenge DREAM7.

12:40 – 13:00

Infinitely Many-Armed Bandits with Unknown Value Distribution

Yahel David, Nahum Shimkin

We consider a version of the classical stochastic Multi-Armed bandit problem in which the number of arms is large compared to the time horizon, with the goal of minimizing the cumulative regret. Here, the mean-reward (or value) of newly chosen arms is assumed to be i.i.d. We further make the simplifying assumption that the value of an arm is revealed once this arm is chosen. We present a general lower bound on the regret, and learning algorithms that achieve this bound up to a logarithmic factor. Contrary to previous work, we do not assume that the functional form of the tail of the value distribution is known. Furthermore, we also consider a variant of our model where sampled arms are non-retainable, namely are lost if not used continuously, with similar near-optimality results.

Session 17: Spatial and temporal data

Room: 105 **Chair: Stefan Wrobel**

11:40 – 12:00

Leveraging the power of local spatial autocorrelation in geophysical interpolative clustering

Annalisa Appice, Donato Malerba

Nowadays ubiquitous sensor stations are deployed worldwide, in order to measure several geophysical variables (e.g. temperature, humidity, light) for a growing number of ecological and industrial processes. Although these variables are, in general, measured over large zones and long (potentially unbounded) periods of time, stations cannot cover any space location. On the other hand, due to their huge volume, data produced cannot be entirely recorded for future analysis. In this scenario, summarization, i.e. the computation of aggregates of data, can be used to reduce the amount of produced data stored on the disk, while interpolation, i.e. the estimation of unknown data in each location of interest, can be used to supplement station records. We illustrate a novel data mining solution, named interpolative clustering, that has the merit of addressing both these tasks in time-evolving, multivariate geophysical applications. It yields a time-evolving clustering model, in order to summarize geophysical data and computes a weighted linear combination of cluster prototypes, in order to predict data. Clustering is done by accounting for the local presence of the spatial autocorrelation property in the geophysical data. Weights of the linear combination are defined, in order to reflect the inverse distance of the unseen data to each cluster geometry. The cluster geometry is represented through shape-dependent sampling of geographic coordinates of clustered stations. Experiments performed with several data collections investigate the trade-off between the summarization capability and predictive accuracy of the presented interpolative clustering algorithm.

12:00 – 12:20

Detecting Localized Homogeneous Anomalies over Spatio-Temporal Data

Aditya Telang, Deepak P, Salil Joshi, Prasad Deshpande, Ranjana Rajendran

The last decade has witnessed an unprecedented growth in availability of data having spatio-temporal characteristics. Given the scale and richness of such data, finding spatio-temporal patterns that demonstrate significantly different behavior from their neighbors could be of interest for various application scenarios such as—weather modeling, analyzing spread of disease outbreaks, monitoring traffic congestions, and so on. In this paper, we propose an automated approach of exploring and discovering such anomalous patterns irrespective of the underlying domain from which the data is recovered. Our approach differs significantly from traditional methods of spatial outlier detection, and employs two phases—(i) discovering homogeneous regions, and (ii) evalating these regions as anomalies based on their statistical difference from a generalized neighborhood. We evaluate the quality of our approach and distinguish it from existing techniques via an extensive experimental evaluation.

12:20 – 12:40

Confidence bands for time series data

Jussi Korpela, Kai Puolamäki, Aristides Gionis

Simultaneous confidence intervals, or confidence bands, provide an intuitive description of the variability of a time series. Given a set of N time series of length M, we consider the problem of finding a confidence band that contains a (1 − α)-fraction of the observations. We construct such confidence bands by finding the set of N −K time series whose envelope is minimized. We refer to this problem as the minimum width envelope problem. We show that the minimum width envelope problem is NP-hard, and we develop a greedy heuristic algorithm, which we compare to quantile- and distance-based confidence band methods. We also describe a method to find an effective confidence level αeff and an effective number of observations to remove Keff, such that the resulting confidence bands will keep the family-wise error rate below α. We evaluate our methods on synthetic and real datasets. We demonstrate that our method can be used to construct confidence bands with guaranteed familywise error rate control, also when there is too little data for the quantile-based methods to work.

12:40 – 13:00

Nowcasting with numerous candidate predictors

Brendan Duncan, Charles Elkan

The goal of nowcasting, or "predicting the present," is to estimate up-to-date values for a time series whose actual observations are available only with a delay. Methods for this task leverage observations of correlated time series to estimate values of the target series. This paper introduces a nowcasting technique called FDR (false discovery reduction) that combines tractable variable selection with a time series model trained using a Kalman filter. The FDR method guarantees that all variables selected have statistically significant predictive power. We apply the method to sales figures provided by the United States census bureau, and to a consumer sentiment index. As side data, the experiments use time series from Google Trends of the volumes of search queries. In total, there are 39,059 potential correlated time series. We compare results from the FDR method to those from several baseline methods. The new method outperforms the baselines and achieves comparable performance to a state-of-the-art nowcasting technique on the consumer sentiment time series, while allowing variable selection from over 250 times as many side data series.

Session 18: Text Mining

Room: 106 **Chair: Toon Calders**

11:40 – 12:00

Joint Prediction of Topics in a URL Hierarchy

Michael Großhans, Christoph Sawade, Tobias Scheffer, Niels Landwehr

We study the problem of jointly predicting topics for all web pages within URL hierarchies. We employ a graphical model in which latent variables represent the predominant topic within a subtree of the URL hierarchy. The model is built around a generative process that infers how web site administrators hierarchically structure web site according to topic, and how web page content is generated depending on the page topic. The resulting predictive model is linear in a joint feature map of content, topic labels, and the latent variables. Inference reduces to message passing in a tree-structured graph; parameter estimation is carried out using concave-convex optimization. We present a case study on web page classification for a targeted advertising application.

12:00 – 12:20

Clustering Image Search Results by Entity Disambiguation

Kaiqi Zhao, Zhiyuan Cai, Qingyu Sui, Enxun Wei, Kenny Zhu

Existing key-word based image search engines return images whose title or immediate surrounding text contains the search term as a keyword. When the search term is ambiguous and means different things, the results often come in a mixed bag of different entities. This paper proposes a novel framework that understands the context and thus infers the most likely entity in the given image by disambiguating the terms in the context into the corresponding concepts from external knowledge in a process called conceptualization. The images can subsequently be clustered by the most likely associated entities. This approach outperforms the best competing image clustering techniques by 29.2% in NMI score. In addition, the framework automatically annotates each cluster of images by its key entities which allows users to quickly identify the images they want.

12:20 – 12:40

How Many Topics? Stability Analysis for Topic Models

Derek Greene, Derek O'Callaghan, Pádraig Cunningham

Topic modeling refers to the task of discovering the underlying thematic structure in a text corpus, where the output is commonly presented as a report of the top terms appearing in each topic. Despite the diversity of topic modeling algorithms that have been proposed, a common challenge in successfully applying these techniques is the selection of an appropriate number of topics for a given corpus. Choosing too few topics will produce results that are overly broad, while choosing too many will result in the "over-clustering" of a corpus into many small, highly-similar topics. In this paper, we propose a term-centric stability analysis strategy to address this issue, the idea being that a model with an appropriate number of topics will be more robust to perturbations in the data. Using a topic modeling approach based on matrix factorization, evaluations performed on a range of corpora show that this strategy can successfully guide the model selection process.

12:40 – 13:00

Open-domain Question Answering with Weakly Supervised Embedding Models

Antoine Bordes, Jason Weston, Nicolas Usunier

Building computers able to answer questions on any subject is a long standing goal of artificial intelligence. Promising progress has recently been achieved by methods that learn to map questions to logical forms or database queries. Such approaches can be effective but at the cost of either large amounts of human-labeled data or by defining lexicons and grammars tailored by practitioners. In this paper, we instead take the radical approach of learning to map questions to vectorial feature representations. By mapping answers into the same space one can query any knowledge base independent of its schema, without requiring any grammar or lexicon. Our method is trained with a new optimization procedure combining stochastic gradient descent followed by a fine-tuning step using the weak supervision provided by blending automatically and collaboratively generated resources. We empirically demonstrate that our model can capture meaningful signals from its noisy supervision leading to major improvements over PARALEX, the only existing method able to be trained on similar weakly labeled data.

Nectar Session 2:

Room: 103-104 **Chair: Pierre Geurts**

14:20 – 14:50

Active Learning is Planning: Nonmyopic ε-Bayes-Optimal Active Learning of Gaussian Processes

Trong Nghia Hoang, Bryan Kian Hsiang Low, Patrick Jaillet, Mohan Kankanhalli

A fundamental issue in active learning of Gaussian processes is that of the exploration-exploitation trade-off. This paper presents a novel nonmyopic ε-Bayes-optimal active learning (ε-BAL) approach that jointly optimizes the trade-off. In contrast, existing works have primarily developed greedy algorithms or performed exploration and exploitation separately. To perform active learning in real time, we then propose an anytime algorithm based on ε-BAL with performance guarantee and empirically demonstrate using a real-world dataset that, with limited budget, it outperforms the state-of-the-art algorithms.

14:50 – 15:20

Heterogeneous Stream Processing and Crowdsourcing for Traffic Monitoring: Highlights

Francois Schnitzler, Alexander Artikis, Matthias Weidlich, Ioannis Boutsis, Thomas Liebig, Nico Piatkowski, Christian Bockermann, Katharina Morik, Vana Kalogeraki, Jakub Marecek, Avigdor Gal, Shie Mannor, Dermot Kinane, Dimitrios Gunopulos

We give an overview of an intelligent urban traffic management system. Complex events related to congestions are detected from heterogeneous sources involving fixed sensors mounted on intersections and mobile sensors mounted on public transport vehicles. To deal with data veracity, sensor disagreements are resolved by crowdsourcing. To deal with data sparsity, a traffic model offers information in areas with low sensor coverage. We apply the system to a real-world use-case.

15:20 – 15:50

Distributional Clauses Particle Filter

Davide Nitti, Tinne De Laet, Luc De Raedt

We review the Distributional Clauses Particle Filter (DCPF), a statistical relational framework for inference in hybrid domains over time such as vision and robotics. Applications in these domains are challenging for statistical relational learning as they require dealing with continuous distributions and dynamics in real-time. The framework addresses these issues, it supports the online learning of parameters and it was tested in several tracking scenarios with good results.

Session 19: Neural structures and deep learning

Room: 105-106 **Chair: Patrick Gallinari**

16:20 – 16:40

Self-Organizing Maps by Difference of Convex functions optimization

Hoai An Le Thi, Manh Cuong Nguyen

We offer an efficient approach based on Difference of Convex functions (DC) optimization for Self-OrganizingMaps (SOM). We consider SOM as an optimization problem with a nonsmooth, nonconvex energy function and investigated DC Programming and DC Algorithm (DCA), an innovative approach in nonconvex optimization framework to effectively solve this problem. Furthermore an appropriate training version of this algorithm is proposed. The numerical results on many real-world datasets show the efficiency of the proposed DCA based algorithms on both quality of solutions and topographicmaps.

16:40 – 17:00

On the Equivalence Between Deep NADE and Generative Stochastic Networks

Li Yao, Sherjil Ozair, Kyunghyun Cho, Yoshua Bengio

Neural Autoregressive Distribution Estimators (NADEs) have recently been shown as successful alternatives for modeling high dimen- sional multimodal distributions. One issue associated with NADEs is that they rely on a particular order of factorization for P(x). This issue has been recently addressed by a variant of NADE called Orderless NADEs and its deeper version, Deep Orderless NADE. Orderless NADEs are trained based on a criterion that stochastically maximizes P (x) with all possible orders of factorizations. Unfortunately, ancestral sampling from deep NADE is very expensive, corresponding to running through a neural net separately predicting each of the visible variables given some others. This work makes a connection between this criterion and the training criterion for Generative Stochastic Networks (GSNs). It shows that training NADEs in this way also trains a GSN, which defines a Markov chain associated with the NADE model. Based on this connection, we show an alternative way to sample from a trained Orderless NADE that allows to trade-off computing time and quality of the samples: a 3 to 10-fold speedup (taking into account the waste due to correla- tions between consecutive samples of the chain) can be obtained without noticeably reducing the quality of the samples. This is achieved using a novel sampling procedure for GSNs called annealed GSN sampling, sim- ilar to tempering methods that combines fast mixing (obtained thanks to steps at high noise levels) with accurate samples (obtained thanks to steps at low noise levels).

17:00 – 17:20

Knowledge-Powered Deep Learning for Word Embedding

Jiang Bian, Bin Gao, Tie-Yan Liu

Recent years have witnessed the increasing efforts that apply deep learning techniques to solve text mining and natural language processing tasks. The basis of these tasks is to obtain high-quality distributed representations of words, i.e., word embeddings, from large amounts of text data. However, text itself usually contains limited information, which makes necessity to leverage extra knowledge to understand it. Fortunately, since text is generated by human, it already contains well-defined morphological and syntactic knowledge; moreover, the large amount of human-generated texts on the Web enable the extraction of plenty of semantic knowledge. Thus, novel deep learning algorithms and systems are needed in order to leverage the above knowledge to compute more effective word embedding. In this paper, we conduct an empirical study on the capacity of leveraging morphologic, syntactic, and semantic knowledge to achieve high-quality word embeddings. Our study explores these types of knowledge to define new basis for word representation, provide additional input information, and serve as auxiliary supervision in deep learning, respectively. Experiments on a popular analogical reasoning task, a word similarity task, and a word completion task have all demonstrated that knowledge-powered deep learning can enhance the effectiveness of word embedding.

17:20 – 17:40

Learned-Norm Pooling for Deep Feedforward and Recurrent Neural Networks

Caglar Gulcehre, Kyunghyun Cho, Razvan Pascanu, Yoshua Bengio

In this paper we propose and investigate a novel nonlinear unit, called Lp unit, for deep neural networks. The proposed Lp unit receives signals from several projections of a subset of units in the layer below and computes a normalized Lp norm. We notice two interesting interpretations of the Lp unit. First, the proposed unit can be understood as a generalization of a number of conventional pooling operators such as average, root-mean-square and max pooling widely used in, for instance, convolutional neural networks (CNN), HMAX models and neocognitrons. Furthermore, the Lp unit is, to a certain degree, similar to the recently proposed maxout unit which achieved the state-of-the-art object recognition results on a number of benchmark datasets. Secondly, we provide a geometrical interpretation of the activation function based on which we argue that the Lp unit is more efficient at representing complex, nonlinear separating boundaries. Each Lp unit defines a superelliptic boundary, with its exact shape defined by the order p. We claim that this makes it possible to model arbitrarily shaped, curved boundaries more efficiently by combining a few Lp units of different orders. This insight justifies the need for learning different orders for each unit in the model. We empirically evaluate the proposed Lp units on a number of datasets and show that multilayer perceptrons (MLP) consisting of the Lp units achieve the state-of-the-art results on a number of benchmark datasets. Furthermore, we evaluate the proposed Lp unit on the recently proposed deep recurrent neural networks (RNN).

Session 20: Reinforcement learning

Room: 102 **Chair: Michele Sebag**

16:20 – 16:40

Policy Search for Path Integral Control

Vicenç Gomez, Jan Peters, Hilbert Kappen, Gerhard Neumann

Path integral (PI) control defines a general class of control problems for which the optimal control computation is equivalent to an inference problem that can be solved by evaluation of a path integral over state trajectories. However, this potential is mostly unused in real-world problems because of two main limitations: first, current approaches can typically only be applied to learn open-loop controllers and second, current sampling procedures are inefficient and not scalable to high dimensional systems. We introduce the efficient Path Integral Relative-Entropy Policy Search (PI-REPS) algorithm for learning feedback policies with PI control. Our algorithm is inspired by information theoretic policy updates that are often used in policy search. We use these updates to approximate the state trajectory distribution that is known to be optimal from the PI control theory. Our approach allows for a principled treatment of different sampling distributions and can be used to estimate many types of parametric or non-parametric feedback controllers. We show that PI-REPS significantly outperforms current methods and is able to solve tasks that are out of reach for current methods.

16:40 – 17:00

An Online Policy Gradient Algorithm for Markov Decision Processes with Continuous States and Actions

Yao Ma, Tingting Zhao, Kohei Hatano, Masashi Sugiyama

We consider the learning problem under an online Markov decision process (MDP), which is aimed at learning the time-dependent decision-making policy of an agent that minimizes the regret — the difference from the best fixed policy. The difficulty of online MDP learning is that the reward function changes over time. In this paper, we show that a simple online policy gradient algorithm could achieves regret bound in order of squre root T for T steps with concavity assumption. To the best of our knowledge, this is the first work to give an online MDP algorithm that can handle continuous state, action and parameter spaces. We demonstrate the performance of the online policy gradient method through experiments.

17:00 – 17:20

Boosted Bellman Residual Minimization Handling Expert Demonstrations

Bilal Piot, Matthieu Geist, Olivier Pietquin

This paper addresses the problem of batch Reinforcement Learning with Expert Demonstrations (RLED). In RLED, the goal is to find an optimal policy of a Markov Decision Process (MDP), using a data set of fixed sampled transitions of the MDP as well as a data set of fixed expert demonstrations. This is slightly different from the batch Reinforcement Learning (RL) framework where only fixed sampled transitions of the MDP are available. Thus, the aim of this article is to propose algorithms that leverage those expert data. The idea proposed here differs from the Approximate Dynamic Programming methods in the sense that we minimize the Optimal Bellman Residual (OBR), where the minimization is guided by constraints defined by the expert demonstrations. This choice is motivated by the the fact that controlling the OBR implies controlling the distance between the estimated and optimal quality functions. However, this method presents some difficulties as the criterion to minimize is non-convex, non-differentiable and biased. Those difficulties are overcome via the embedding of distributions in a Reproducing Kernel Hilbert Space (RKHS) and a boosting technique which allows obtaining non-parametric algorithms. Finally, our algorithms are compared to the only state of the art algorithm, Approximate Policy Iteration with Demonstrations (APID) algorithm, in different experimental settings.

17:20 – 17:40

Local Policy Search in a Convex Space and Conservative Policy Iteration as Boosted Policy Search

Bruno Scherrer, Matthieu Geist

Local Policy Search is a popular reinforcement learning approach for handling large state spaces. Formally, it searches locally in a parameterized policy space in order to maximize the associated value function averaged over some predefined distribution. The best one can hope in general from such an approach is to get a local optimum of this criterion. The first contribution of this article is the following surprising result: if the policy space is convex, any (approximate) local optimum enjoys a global performance guarantee. Unfortunately, the convexity assumption is strong: it is not satisfied by commonly used parameterizations and designing a parameterization that induces this property seems hard. A natural solution to alleviate this issue consists in deriving an algorithm that solves the local policy search problem using a boosting approach (constrained to the convex hull of the policy space). The resulting algorithm turns out to be a slight generalization of conservative policy iteration; thus, our second contribution is to highlight an original connection between local policy search and approximate dynamic programming.

Session 21: Recommendation systems and dyadic data

Room: 103-104 **Chair: Martin Atzmueller**

16:20 – 16:40

A Bayesian Generative Model for Item and User Recommendation in Social Rating Networks with Trust Relationships

Gianni Costa, Giuseppe Manco, Riccardo Ortale

A Bayesian generative model is presented for recommending interesting items and trustworthy users to the targeted users in social rating networks with asymmetric and directed trust relationships. The proposed model is the first unified approach to the combination of the two recommendation tasks. Within the devised model, each user is associated with two latent-factor vectors, i.e., her susceptibility and expertise. Items are also associated with corresponding latent-factor vector representations. The probabilistic factorization of the rating data and trust relationships is exploited to infer user susceptibility and expertise. Statistical social-network modeling is instead used to constrain the trust relationships from a user to another to be governed by their respective susceptibility and expertise. The inherently ambiguous meaning of unobserved trust relationships between users is suitably disambiguated. An intensive comparative experimentation on real-world social rating networks with trust relationships demonstrates the superior predictive performance of the presented model in terms of RMSE and AUC.

16:40 – 17:00

Collaborative Filtering with Information-Rich and Information-Sparse Entities

Kai Zhu, Rui Wu, Lei Ying, R. Srikant

In this paper, we consider a popular model for collaborative filtering in recommender systems. In particular, we consider both the clustering model, where only users (or items) are clustered, and the co-clustering model, where both users and items are clustered, and further, we assume that some users rate many items (information-rich users) and some users rate only a few items (information-sparse users). When users (or items) are clustered, our algorithm can recover the rating matrix with ω(M K log M) noisy entries while M K entries are necessary, where K is the number of clusters and M is the number of items. In the case of co-clustering, we prove that K^2 entries are necessary for recovering the rating matrix, and our algorithm achieves this lower bound within a logarithmic factor when K is sufficiently large. Extensive simulations on Netflix and MovieLens data show that our algorithm outperforms the alternating minimization (AM) and the popularity-among-friends (PAF) algorithm. The performance difference increases even more when noise is added to the datasets.

17:00 – 17:20

A Constrained Matrix-Variate Gaussian Process for Transposable Data

Oluwasanmi Koyejo, Cheng Lee, Joydeep Ghosh

Transposable data represents interactions among two sets of entities, and are typically represented as a matrix containing the known interaction values. Additional side information may consist of feature vectors specific to entities corresponding to the rows and/or columns of such a matrix. Further information may also be available in the form of interactions or hierarchies among entities along the same mode (axis). We propose a novel approach for modeling transposable data with missing interactions given additional side information. The interactions are modeled as noisy observations from a latent noise free matrix generated from a matrix-variate Gaussian process. The construction of row and column covariances using side information provides a flexible mechanism for specifying a-priori knowledge of the row and column correlations in the data. Further, the use of such a prior combined with the side information enables predictions for new rows and columns not observed in the training data. In this work, we combine the matrix-variate Gaussian process model with low rank constraints. The constrained Gaussian process approach is applied to the prediction of hidden associations between genes and diseases using a small set of observed associations as well as prior covariances induced by gene-gene interaction networks and disease ontologies. The proposed approach is also applied to recommender systems data which involves predicting the item ratings of users using known associations as well as prior covariances induced by social networks. We present experimental results that highlight the performance of constrained matrix-variate Gaussian process as compared to state of the art approaches in each domain.

17:20 – 17:40

A two-step learning approach for solving full and almost full cold start problems in dyadic prediction

Tapio Pahikkala, Michiel Stock, Antti Airola, Tero Aittokallio, Bernard De Baets, Willem Waegeman

Dyadic prediction methods operate on pairs of objects (dyads), aiming to infer labels for out-of-sample dyads. We consider the full and almost full cold start problem in dyadic prediction, a setting that occurs when both objects in an out-of-sample dyad have not been observed during training, or if one of them has been observed, but very few times. A popular approach for addressing this problem is to train a model that makes predictions based on a pairwise feature representation of the dyads, or, in case of kernel methods, based on a tensor product pairwise kernel. As an alternative to such a kernel approach, we introduce a novel two-step learning algorithm that borrows ideas from the fields of pairwise learning and spectral filtering. We show theoretically that the two-step method is very closely related to the tensor product kernel approach, and experimentally that it yields a slightly better predictive performance. Moreover, unlike existing tensor product kernel methods, the two-step method allows closed-form solutions for training and parameter selection via cross-validation estimates both in the full and almost full cold start settings, making the approach much more efficient and straightforward to implement.

Session 22: Data stream mining

Room: 101 **Chair: Myra Spiliopoulou**

16:20 – 16:40

Classy: Fast Clustering Streams of Call-Graphs

Orestis Kostakis

An abstraction resilient to common malware obfuscation techniques is the call-graph. A call-graph is the representation of an executable file as a directed graph with labeled vertices, where the vertices correspond to functions and the edges to function calls. Unfortunately, most of the interesting graph comparison problems, including full-graph comparison and computing the largest common subgraph, belong to the NP-hard class. This makes the study and use of graphs in large scale systems difficult. Existing work has focused only on offline clustering and has not addressed the issue of clustering streams of graphs. In this paper we present Classy, a scalable distributed system that clusters streams of large call-graphs for purposes including automated malware classification and facilitating malware analysts. Since algorithms aimed at clustering sets are not suitable for clustering streams of objects, we propose the use of a clustering algorithm that relies on the notion of candidate clusters and reference samples therein. We demonstrate via thorough experimentation that this approach yields results very close to the off-line optimal. Graph similarity is determined by computing a Graph Edit distance (GED) of pairs of graphs using an adapted version of Simulated Annealing. Furthermore, we present a novel lower bound for the GED. We also study the problem of approximating statistics of clusters of graphs when the distances of only a fraction of all possible pairs have been computed. Finally, we present results and statistics from a real production-side system that has clustered and contains more than 0.8 million graphs.

16:40 – 17:00

Classifying A Stream of Infinite Concepts: A Bayesian Non-Parametric Approach

Abbas Hosseini, Hamid R. Rabiee, Hassan Hafez, Ali Soltani-Farani

Classifying streams of data, for instance financial transactions or emails, is an essential element in applications such as online advertising and spam or fraud detection. The data stream is often large or even unbounded, and in many instances non-stationary. Therefore, an adaptive approach that can manage concept drift in an online fashion, is often required. This paper presents a probabilistic, non-parametric, and generative model for stream classification that can handle concept drift efficiently, and adjust its complexity over time. Unlike recent methods, the proposed model handles concept drift by adapting data-concept association without the unnecessary i.i.d. assumption among the data, in a batch. This allows the model to efficiently classify data by using fewer and simpler base classifiers. Moreover, an online algorithm for making inference on the proposed non-conjugate, time-dependent, and non-parametric model is proposed. Extensive experimental results on several stream datasets demonstrate the effectiveness of the proposed model.

17:00 – 17:20

Speeding Up Recovery From Concept Drifts

Silas Santos, Paulo Gonçalves, Geyson Silva, Roberto Barros

The extraction of knowledge from data streams is an activity that has progressively been receiving an increased demand. However, in this type of environment, changes in data distribution, or concept drift, can occur constantly and is a challenge. This paper proposes the Adaptable Diversity-based Online Boosting (ADOB), a modified version of the online boosting, as proposed by Oza and Russell, which is aimed at speeding up the experts recovery after concept drifts. We performed experiments to compare the accuracy as well as the execution time and memory use of ADOB against a number of other methods using several artificial and real world datasets, chosen from the most used ones in the area. Results suggest that, in many different situations, the proposed approach maintains a high accuracy, outperforming the other tested methods in regularity, with no significant change in the execution time and memory use. In particular, ADOB was specially efficient in situations where frequent and abrupt concept drifts occur.

17:20 – 17:40

Mining Top-K Largest Tiles in a Data Stream

Hoang Thanh Lam, Wenjie Pei, Adriana Prado, Baptiste Jeudy, Elisa Fromont

Large tiles in a database are itemsets with the largest area which is defined as the itemset frequency in the database multiplied by its size. Mining these large tiles is an important pattern mining problem since tiles with a large area describe a large part of the database. In this paper, we introduce the problem of mining top-k largest tiles in a data stream under the sliding window model. We propose a candidate-based approach which summarizes the data stream and produces the top-k largest tiles efficiently for moderate window size. We also propose an approximation algorithm with theoretical bounds on the error rate to cope with large size windows. In the experiments with two real-life datasets, the approximation algorithm is up to hundred times faster than the candidate-based solution and the baseline algorithms based on the state-of-the-art solutions. We also investigate an application of large tile mining in computer vision and in emerging search topics monitoring.

THURSDAY 18 SEPTEMBER 2014

THURSDAY INVITED TALK

Machine Learning for Search Ranking and Ad Auction

Speaker: Tie-Yan Liu
Time: 09:00 – 10:00
Room: Auditorium 850

Abstract
In the era of information explosion, search has become an important tool for people to retrieve useful information. Every day, billions of search queries are submitted to commercial search engines. In response to a query, search engines return a list of relevant documents according to a ranking model. In addition, they also return some ads to users, and extract revenue by running an auction among advertisers if users click on these ads. This "search + ads" paradigm has become a key business model in today's Internet industry, and has incubated a few hundred-billion-dollar companies. Recently, machine learning has been widely adopted in search and advertising, mainly due to the availability of huge amount of interaction data between users, advertisers, and search engines. In this talk, we discuss how to use machine learning to build effective ranking models (which we call learning to rank) and to optimize auction mechanisms. (i) The difficulty of learning to rank lies in the interdependency between documents in the ranked list. To tackle it, we propose the so-called listwise ranking algorithms, whose loss functions are defined on the permutations of documents, instead of individual documents or document pairs. We prove the effectiveness of these algorithms by analyzing their generalization ability and statistical consistency, based on the assumption of a two-layer probabilistic sampling procedure for queries and documents, and the characterization of the relationship between their loss functions and the evaluation measures used by search engines (e.g., NDCG and MAP). (ii) The difficulty of learning the optimal auction mechanism lies in that advertisers' behavior data are strategically generated in response to the auction mechanism, but not randomly sampled in an i.i.d. manner. To tackle this challenge, we propose a game-theoretic learning method, which first models the strategic behaviors of advertisers, and then optimizes the auction mechanism by assuming the advertisers to respond to new auction mechanisms according to the learned behavior model. We prove the effectiveness of the proposed method by analyzing the generalization bounds for both behavior learning and auction mechanism learning based on a novel Markov framework.

Bio
Tie-Yan Liu is a senior researcher and research manager at Microsoft Research. His research interests include machine learning (learning to rank, online learning, statistical learning theory, and deep learning), algorithmic game theory, and computational economics. He is well known for his work on learning to rank for information retrieval. He has authored the first book in this area, and published tens of highly-cited papers on both algorithms and theorems of learning to rank. He has also published extensively on other related topics. In particular, his paper won the best student paper award of SIGIR (2008), and the most cited paper award of the Journal of Visual Communication and Image Representation (2004-2006); his group won the research break-through award of Microsoft Research Asia (2012). Tie-Yan is very active in serving the research community. He is a program committee co-chair of ACML (2015), WINE (2014), AIRS (2013), and RIAO (2010), a local co-chair of ICML 2014, a tutorial co-chair of WWW 2014, a demo/exhibit co-chair of KDD (2012), and an area/track chair of many conferences including ACML (2014), SIGIR (2008-2011), AIRS (2009-2011), and WWW (2011). He is an associate editor of ACM Transactions on Information System (TOIS), an editorial board member of Information Retrieval Journal and Foundations and Trends in Information Retrieval. He has given keynote speeches at CCML (2013), CCIR (2011), and PCM (2010), and tutorials at SIGIR (2008, 2010, 2012), WWW (2008, 2009, 2011), and KDD (2012). He is a senior member of the IEEE and the ACM.

THURSDAY INDUSTRY INVITED TALK

Making smart metering smarter by applying data analytics

Speaker: Georges Hébrail
Time: 14:20 –15:05
Room: Auditorium 850

Abstract
New data is being collected from electric smart meters which are deployed in many countries. Electric power meters measure and transmit to a central information system electric power consumption from every individual household or enterprise. The sampling rate may vary from 10 minutes to 24 hours and the latency to reach the central information system may vary from a few minutes to 24h. This generates a large amount of – possibly streaming – data if we consider customers from an entire country (ex. 35 millions in France). This data is collected firstly for billing purposes but can be processed with data analytics tools with several other goals. The first part of the talk will recall the structure of electric power smart metering data and review the different applications which are considered today for applying data analytics to such data. In a second part of the talk, we will focus on a specific problem: spatio-temporal estimation of aggregated electric power consumption from incomplete metering data.

Bio
Georges Hébrail is a senior researcher at EDF Lab, the research centre of Electricité de France, one of the world's leading electric utility. His background is in Business Intelligence covering many aspects from data storage and querying to data analytics. From 2002 to 2010, he was a professor of computer science at Telecom ParisTech, teaching and doing research in the field of information systems and business intelligence, with a focus on time series management, stream processing and mining. His current research interest is on distributed and privacy-preserving data mining on electric power related data.

Ads that matter

Speaker: Alexandre Cotarmanac'h
Time: 15:05 – 15:50
Room: Auditorium 850

Abstract
The advent of realtime bidding and online ad-exchanges has created a new and fast-growing competitive marketplace. In this new setting, media-buyers can make fine-grained decisions for each of the impressions being auctioned taking into account information from the context, the user and his/her past behavior. This new landscape is particularly interesting for online e-commerce players where user actions can also be measured online and thus allow for a complete measure of return on ad-spend.
Despite those benefits, new challenges need to be addressed such as:

- the design of a realtime bidding architecture handling high volumes of queries at low latencies,
- the exploration of a sparse and volatile high-dimensional space
- as well as several statistical modeling problems (e.g. pricing, offer and creative selection).

In this talk, I will present an approach to realtime media buying for online e-commerce from our experience working in the field. I will review the aforementioned challenges and discuss open problems for serving ads that matter.

Bio
Alexandre Cotarmanac'h is Vice-President Distribution & Platform for Twenga.
Twenga is a services and solutions provider generating high value-added leads to online merchants that was founded in 2006.
Originally hired to help launch Twenga's second generation search engine and to manage the optimization of revenue, he launched in 2011 the affinitAD line of business and Twenga's publisher network. Thanks to the advanced contextual analysis which allows for targeting the right audience according to their desire to buy e-commerce goods whilst keeping in line with the content offered, affinitAD brings Twenga's e-commerce expertise to web publishers. Alexandre also oversees Twenga's merchant program and strives to offer Twenga's merchants new services and solutions to improve their acquisition of customers.
With over 14 years of experience, Alexandre has held a succession of increasingly responsible positions focusing on advertising and web development. Prior to joining Twenga, he was responsible for the development of Search and Advertising at Orange.
Alexandre graduated from Ecole polytechnique.

THURSDAY SESSIONS AT A GLANCE

Nectar Session 3:

Room: 103-104 **Chair: Loïc Cerf**

10:10 – 10:40 **Be certain of how-to before mining uncertain data**
Francesco Gullo, Giovanni Ponti, Andrea Tagarelli

10:40 – 11:10 **Generalized Online Sparse Gaussian Processes with Application to Persistent Mobile Robot Localization**
Bryan Kian Hsiang Low, Nuo Xu, Jie Chen, Keng Kiat Lim, Etkin Ozgul

Session 23: Classifier evaluation

Room: 102 **Chair: Pierre Dupont**

10:10 – 10:30 **Leave-one-out cross-validation is risk consistent for lasso**
Darren Homrighausen, Daniel McDonald

10:30 – 10:50 **Reliability maps: A tool to enhance probability estimates and improve classification accuracy**
Meelis Kull, Peter Flach

10:50 – 11:10 **A Peek into the Black Box: Exploring Classifiers by Randomization**
Andreas Henelius, Kai Puolamäki, Henrik Boström, Lars Asker, Panagiotis Papapetrou

Session 24: Data Mining tools and frameworks

Room: 105 **Chair: Nada Lavrac**

10:10 – 10:30 **Ontology of Core Data Mining Entities**
Pance Panov, Larisa Soldatova, Saso Dzeroski

10:30 – 10:50 **Code you are Happy to Paste: an Algorithmic Dictionary of Exponential Families**
Olivier Schwander

10:50 – 11:10 **On Combining Machine Learning with Decision Making**
Theja Tulabandhula, Cynthia Rudin

Session 25: Spectral learning

Room: 106 **Chair: Floriana Esposito**

10:10 – 10:30 **GMRF Estimation under Topological and Spectral constraints**
Victorin Martin, Cyril Furtlehner, Yufei Han, Jean-Marc Lasgouttes

10:30 – 10:50 **Discriminative Subnetworks with Regularized Spectral Learning for Global-state Network Data**
Xuan-Hong Dang, Ambuj K. Singh, Petko Bogdanov, Hongyuan You, Bayyuan Hsu

10:50 – 11:10 **Hypernode Graphs for Spectral Learning on Binary Relations over Sets**
Thomas Ricatte, Rémi Gilleron, Marc Tommasi

Session 26: Neural networks and deep learning

Room: 103-104 **Chair: Cho Kyunghyun**

11:40 - 12:00 **Restricted Boltzmann Machines with Overlapping Partitions**
Hasari Tosun, John Sheppard

12:00 - 12:20 **Recurrent Greedy Parsing with Neural Networks**
Joël Legrand, Ronan Collobert

12:20 - 12:40 **Large-scale Multi-label Text Classification — Revisiting Neural Networks**
Jinseok Nam, Jungi Kim, Eneldo Loza Mencia, Iryna Gurevych, Johannes Fuernkranz

12:40 - 13:00 **Neural Gaussian Conditional Random Fields**
Vladan Radosavljevic, Slobodan Vucetic, Zoran Obradovic

Session 27: Partially and semi-supervised learning

Room: 102 **Chair: Hendrik Blockeel**

11:40 - 12:00 **Statistical Hypothesis Testing in Positive Unlabelled Data**
Konstantinos Sechidis, Borja Calvo, Gavin Brown

12:00 - 12:20 **Hetero-Labeled LDA: A partially supervised topic model with heterogeneous labels**
Dongyeop Kang, Youngja Park, Suresh Chari

12:20 - 12:40 **Semi-Supervised Learning Using an Unsupervised Atlas**
Chris Russell, Lourdes Agapito , Nikos Pitelis

12:40 - 13:00 **Consistency of losses for learning from weak labels**
Jesus Cid Sueiro, Raul Santos-Rodriguez, Dario Garcia-Garcia

Session 28: Reliable prediction

Room: 105 **Chair: Indrė Žliobaitė**

11:40 - 12:00 **Transductive Minimax Probability Machine**
Gao Huang, Shiji Song, Zhixiang Xu, Kilian Weinberger

12:00 - 12:20 **Regression Conformal Prediction with Random Forests**
Ulf Johansson, Henrik Boström, Tuve Löfström, Henrik Linusson

12:20 - 12:40 **Combination of one-class support vector machines for classification with reject option**
Blaise Hanczar, Michele Sebag

12:40 - 13:00 **Cautious Ordinal Classification by Binary Decomposition**
Sebastien Destercke, Gen Yang

Session 29: Multi-target and transfer learning

Room: 106 **Chair: Michelangelo Ceci**

11:40 - 12:00 **Multi-Target Regression via Random Linear Target Combinations**
Grigorios Tsoumakas, Eleftherios Spyromitros-Xioufis, Aikaterini Vrekou, Ioannis Vlahavas

12:00 - 12:20 **Transfer Learning with Multiple Sources via Consensus Regularized Autoencoders**
Fuzhen Zhuang, Xiaohu Cheng, Sinno Jialin Pan, Qing He, zhongzhi Shi

12:20 - 12:40 **Importance Weighted Inductive Transfer Learning for Regression**
Thomas Vanck, Jochen Garcke

12:40 - 13:00 **Domain adaptation with regularized optimal transport**
Nicolas Courty, Devis Tuia, Rémi Flamary

Nectar Session 4:

Room: 103-104 **Chair: Marc Plantevit**

14:20 - 14:50 **Network reconstruction for the identification of miRNA:mRNA interaction networks**
Gianvito Pio, Michelangelo Ceci, Domenica D'Elia, Donato Malerba

14:50 - 15:20 **Analyzing and Grounding Social Interaction in Online and Offline Networks**
Martin Atzmueller

15:20 - 15:50 **Agents Teaching Agents in Reinforcement Learning (Nectar Abstract)**
Matthew Taylor, Lisa Torrey

Session 30: Support vector machines

Room: 102 **Chair: Ulf Brefeld**

16:20 - 16:40 **Active Learning for Support Vector Machines with Maximal Model Change**
Wenbin Cai, Ya Zhang, Siyuan Zhou, Wenquan Wang, Chris Ding, Xiao Gu

16:40 - 17:00 **Support Vector Machines for Differential Prediction**
Finn Kuusisto, Vitor Santos Costa, Houssam Nassif, Elizabeth Burnside, David Page, Jude Shavlik

17:00 - 17:20 **Accelerating Model Selection with Safe Screening for L1-Regularized L2-SVM**
Zheng Zhao, Jun Liu, James Cox

17:20 - 17:40 **A Theoretical and Empirical Analysis of Support Vector Machine Methods for Multiple-Instance Classification**
Gary Doran, Soumya Ray

Session 31: Privacy and anti-discrimination in data mining

Room: 105 **Chair: Bart Goethals**

16:20 - 16:40 **Anti-discrimination Analysis Using Privacy Attack Strategies**
Salvatore Ruggieri, Sara Hajian, Faisal Kamiran, Xiangliang Zhang

16:40 - 17:00 **Neutralized Empirical Risk Minimization with Generalization Neutrality Bound**
Kazuto Fukuchi, Jun Sakuma

17:00 - 17:20 **Generalization-based Privacy Preservation and Discrimination Prevention in Data Publishing and Mining**
Sara Hajian, Josep Domingo-Ferrer, Oriol Farras

17:20 - 17:40 **Preserving Worker Privacy in Crowdsourcing**
Hiroshi Kajino, Hiromi Arai, Hisashi Kashima

Session 32: Probabilistic and Bayesian methods

Room: 106 **Chair: Szymon Jaroszewicz**

16:20 - 16:40 **Linear State-Space Model with Time-Varying Dynamics**
Jaakko Luttinen, Tapani Raiko, Alexander Ilin

16:40 - 17:00 **Nonparametric Markovian Learning of Triggering Kernels for Mutually Exciting and Mutually Inhibiting Multivariate Hawkes Processes**
Remi Lemonnier, Nicolas Vayatis

17:00 - 17:20 **Bayesian Models for Structured Sparse Estimation via Set Cover Prior**
Xianghang Liu, Xinhua Zhang, Tiberio Caetano

17:20 - 17:40 **Cutset Networks: A Simple, Tractable, and Scalable Approach for Improving the Accuracy of Chow-Liu Trees**
Tahrima Rahman, Prasanna Kothalkar, Vibhav Gogate

Session 33: Time-evolving graphs and Dynamic Networks

Room: 103-104 **Chair: Pauli Miettinen**

16:20 - 16:40
Fast Nearest Neighbor Search on Large Time-Evolving Graphs
Leman Akoglu, Rohit Khandekar, Vibhore Kumar, Srinivasan Parthasarathy, Deepak Rajan, Kun-Lung Wu, Christos Faloutsos

16:40 - 17:00
Scalable Information Flow Mining in Networks
Karthik Subbian, Chidananda Sridhar, Charu Aggarwal, Jaideep Srivastava

17:00 - 17:20
Discovering Bands from Graphs
Nikolaj Tatti

17:20 - 17:40
Communication-Efficient Distributed Online Prediction by Decentralized Variance Monitoring
Michael Kamp, Mario Boley, Daniel Keren, Assaf Schuster, Izchak Sharfman

THURSDAY SESSIONS, WITH ABSTRACTS

Nectar Session 3:

Room: 103-104 **Chair: Loïc Cerf**

10:10 – 10:40

Be certain of how-to before mining uncertain data

Francesco Gullo, Giovanni Ponti, Andrea Tagarelli

The purpose of this technical note is to introduce the problems of similarity detection and summarization in uncertain data. We provide the essential arguments that make the problems relevant to the data-mining and machine-learning community, stating major issues and summarizing our contributions in the field. Further challenges and directions of research are also issued.

10:40 – 11:10

Generalized Online Sparse Gaussian Processes with Application to Persistent Mobile Robot Localization

Bryan Kian Hsiang Low, Nuo Xu, Jie Chen, Keng Kiat Lim, Etkin Ozgul

This paper presents a novel online sparse Gaussian process (GP) approximation method that is capable of achieving constant time and memory (i.e., independent of the size of the data) per time step. We theoretically guarantee its predictive performance to be equivalent to that of a sophisticated offline sparse GP approximation method. We empirically demonstrate the practical feasibility of using our online sparse GP approximation method through a real-world persistent mobile robot localization experiment.

Session 23: Classifier evaluation

Room: 102 **Chair: Pierre Dupont**

10:10 – 10:30

Leave-one-out cross-validation is risk consistent for lasso

Darren Homrighausen, Daniel McDonald

The lasso procedure pervades the statistical and signal processing literature, and as such, is the target of substantial theoretical and applied research. While much of this research focuses on the desirable properties that lasso possesses—predictive risk consistency, sign consistency, correct model selection— these results assume that the tuning parameter is chosen in an oracle fashion. Yet, this is impossible in practice. Instead, data analysts must use the data twice, once to choose the tuning parameter and again to estimate the model. But only heuristics have ever justified such a procedure. To this end, we give the first definitive answer about the risk consistency of lasso when the smoothing parameter is chosen via cross-validation. We show that under some restrictions on the design matrix, the lasso estimator is still risk consistent with an empirically chosen tuning parameter.

10:30 – 10:50

Reliability maps: A tool to enhance probability estimates and improve classification accuracy

Meelis Kull, Peter Flach

We propose a general method to assess the reliability of two-class probabilities in an instance-wise manner. This is relevant, for instance, for obtaining calibrated multi-class probabilities from two-class probability scores. The LS-ECOC method approaches this by performing least-squares fitting over a suitable error-correcting output code matrix, where the optimisation resolves potential conflicts in the input probabilities. While this gives all input probabilities equal weight, we would like to spend less effort fitting unreliable probability estimates. We introduce the concept of a reliability map to accompany the more conventional notion of calibration map; and LS-ECOC-R which modifies LS-ECOC to take reliability into account. We demonstrate on synthetic data that this gets us closer to the Bayes-optimal classifier, even if the base classifiers are linear and hence have high bias. Results on UCI data sets demonstrate that multi-class accuracy also improves.

10:50 – 11:10

A Peek into the Black Box: Exploring Classifiers by Randomization

Andreas Henelius, Kai Puolamäki, Henrik Boström, Lars Asker, Panagiotis Papapetrou

Classifiers are often opaque and cannot easily be inspected to gain understanding of which factors are of importance. We propose an efficient iterative algorithm to find the attributes and dependencies used by any classifier when making predictions. The performance and utility of the algorithm is demonstrated on two synthetic and 26 real-world datasets, using 15 commonly used learning algorithms to generate the classifiers. The empirical investigation shows that the novel algorithm is indeed able to find groupings of interacting attributes exploited by the different classifiers. These groupings allow for finding similarities among classifiers for a single dataset as well as for determining the extent to which different classifiers exploit such interactions in general.

Session 24: Data Mining tools and frameworks

Room: 105 **Chair: Nada Lavrac**

10:10 – 10:30

Ontology of Core Data Mining Entities

Pance Panov, Larisa Soldatova, Saso Dzeroski

In this article, we present OntoDM-core, an ontology of core data mining entities. OntoDM-core defines the most essential data mining entities in a three-layered ontological structure comprising of a specification, an implementation and an application layer. It provides a representational framework for the description of mining structured data, and in addition provides taxonomies of datasets, data mining tasks, generalizations, data mining algorithms and constraints, based on the type of data. OntoDM-core is designed to support a wide range of applications/use cases, such as semantic annotation of data mining algorithms, datasets and results; annotation of QSAR studies in the context of drug discovery investigations; and disambiguation of terms in text mining. The ontology has been thoroughly assessed following the practices in ontology engineering, is fully interoperable with many domain resources and is easy to extend. OntoDM-core is available at http://www.ontodm.com .

10:30 – 10:50

Code you are Happy to Paste: an Algorithmic Dictionary of Exponential Families

Olivier Schwander

We describe a library and a companion website designed to ease the usage of exponential families in various programming languages. Implementation of mathematical formulas in computer programs is often error-prone, difficult to debug and difficult to read afterwards. Moreover, this implementation is heavily dependent of the programming language used and often needs an important knowledge of the idioms of the language. In our system, formulas are described in a high-level language and mechanically exported to the chosen target language and a LaTeX export allows to quickly review correctness of formulas. Although our system is not limited by design to exponential families, we focus on this kind of formulas since they are of great interest for machine learning and statistical modeling applications. Besides, exponential families are a good usecase of our dictionary: among other usages, they may be used with generic algorithms for mixture models such as Bregman Soft Clustering, in which case lots of formulas from the canonical decomposition of the family need to be implemented. We thus illustrate our library by generating code which can be plugged into generic Expectation-Maximization schemes written in multiple languages.

10:50 – 11:10

On Combining Machine Learning with Decision Making

Theja Tulabandhula, Cynthia Rudin

We present a new application and covering number bound for the framework of "Machine Learning with Operational Costs (MLOC)," which is an exploratory form of decision theory. The MLOC framework incorporates knowledge about how a predictive model will be used for a subsequent task, thus combining machine learning with the decision that is made afterwards. In this work, we use the MLOC framework to study a problem that has implications for power grid reliability and maintenance, called the Machine Learning and Traveling Repair-man Problem (ML&TRP). The goal of the ML&TRP is to determine a route for a "repair crew," which repairs nodes on a graph. The repair crew aims to minimize the cost of failures at the nodes, but as in many real situations, the failure probabilities are not known and must be estimated. The MLOC framework allows us to understand how this uncertainty influences the repair route. We also present new covering number generalization bounds for the MLOC framework.

Session 25: Spectral learning

Room: 106 **Chair: Floriana Esposito**

10:10 – 10:30

GMRF Estimation under Topological and Spectral constraints

Victorin Martin, Cyril Furtlehner, Yufei Han, Jean-Marc Lasgouttes

We investigate the problem of Gaussian Markov random field selection under a non-analytic constraint: the estimated models must be compatible with a fast inference algorithm, namely the Gaussian belief propagation algorithm. To address this question, we introduce the *-IPS framework, based on iterative proportional scaling, which incrementally selects candidate links in a greedy manner. Besides its intrinsic sparsity-inducing ability, this algorithm is flexible enough to incorporate various spectral constraints, like e.g. walk summability, and topological constraints, like short loops avoidance. Experimental tests on various datasets, including traffic data from San Francisco Bay Area, indicate that this approach can deliver, with reasonable computational cost, a broad range of efficient inference models, which are not accessible through penalization with traditional sparsity-inducing norms.

10:30 – 10:50

Discriminative Subnetworks with Regularized Spectral Learning for Global-state Network Data

Xuan-Hong Dang, Ambuj K. Singh, Petko Bogdanov, Hongyuan You, Bayyuan Hsu

Data mining practitioners are facing challenges from data with network structure. In this paper, we address a specific class of global-state networks which comprises of a set of network instances sharing a similar structure yet having different values at local nodes. Each instance is associated with a global state which indicates the occurrence of an event. The objective is to uncover a small set of discriminative subnetworks that can optimally classify global network values. Unlike most existing studies which explore an exponential subnetwork space, we address this difficult problem by adopting a space transformation approach. Specifically, we present an algorithm that optimizes a constrained dual-objective function to learn a low-dimensional subspace that is capable of discriminating networks labelled by different global states, while reconciling with common network topology sharing across instances. Our algorithm takes an appealing approach from spectral graph learning and we show that the globally optimum solution can be achieved via matrix eigen-decomposition.

10:50 – 11:10

Hypernode Graphs for Spectral Learning on Binary Relations over Sets

Thomas Ricatte, Rémi Gilleron, Marc Tommasi

We introduce hypernode graphs as weighted binary relations between sets of nodes: a hypernode is a set of nodes, a hyperedge is a pair of hypernodes, and each node in a hypernode of a hyperedge is given a non negative weight that represents the node contribution to the relation. Hypernode graphs model binary relations between sets of individuals while allowing to reason at the level of individuals. We present a spectral theory for hypernode graphs that allows us to introduce an unnormalized Laplacian and a smoothness semi-norm. In this framework, we are able to extend spectral graph learning algorithms to the case of hypernode graphs. We show that hypernode graphs are a proper extension of graphs from the expressive power point of view and from the spectral analysis point of view. Therefore hypernode graphs allow to model higher order relations whereas it is not true for hypergraphs as shown in Higher Order Learning with Graphs (http://vision.ucsd.edu/~kbranson/HigherOrderLearningWithGraphs.pdf). In order to prove the potential of the model, we represent multiple players games with hypernode graphs and introduce a novel method to infer skill ratings from game outcomes. We show that spectral learning algorithms over hypernode graphs obtain competitive results with skill ratings specialized algorithms such as Elo duelling and TrueSkill.

Session 26: Neural networks and deep learning

Room: 103-104 **Chair: Cho Kyunghyun**

11:40 - 12:00

Restricted Boltzmann Machines with Overlapping Partitions

Hasari Tosun, John Sheppard

Restricted Boltzmann Machines (RBM) are energy-based models that are successfully used as generative learning models as well as crucial components of Deep Belief Networks (DBN). The most successful training method to date for RBMs is the Contrastive Divergence method. However, Contrastive Divergence is inefficient when the number of features is very high and the mixing rate of the Gibbs chain is slow. We propose a new training method that partitions a single RBM into multiple overlapping small RBMs. The final RBM is learned by layers of partitions. We show that this method is not only fast, it is also more accurate in terms of its generative power.

12:00 - 12:20

Recurrent Greedy Parsing with Neural Networks

Joël Legrand, Ronan Collobert

In this paper, we propose a bottom-up greedy and purely discriminative syntactic parsing approach that relies only on a few simple features. The core of the architecture is a simple neural network architecture, trained with an objective function similar to a Conditional Random Field. This parser leverages continuous word vector representations to model the conditional distributions of context-aware syntactic rules. The learned distribution rules are naturally smoothed, thanks to the continuous nature of the input features and the model. Generalization accuracy compares very well with the existing generative or discriminative (non-reranking) parsers (despite the greedy nature of our approach), and prediction speed is very fast.

12:20 - 12:40

Large-scale Multi-label Text Classification — Revisiting Neural Networks

Jinseok Nam, Jungi Kim, Eneldo Loza Mencia, Iryna Gurevych, Johannes Fuernkranz

Recent works have proposed the use of neural networks for multi-label classification because they are able to capture and model label dependencies in the output layer. In this work, we investigate limitations of BP-MLL, a neural network (NN) architecture that aims at minimizing pairwise ranking error. Instead, we propose to use a comparably simple NN approach with recently proposed learning techniques for large-scale multi-label text classification tasks. In particular, we show that BP-MLL's ranking loss minimization can be efficiently and effectively replaced with the commonly used cross entropy error function, and demonstrate that several advances in neural network training that have been recently developed in the realm of deep learning can be effectively employed in this setting. Our experimental results show that simple NN models equipped with recent advanced techniques such as rectified linear units, dropout, and adagrad perform as well as or even outperform state-of-the-art approaches on six large-scale textual datasets with diverse characteristics.

12:40 - 13:00

Neural Gaussian Conditional Random Fields

Vladan Radosavljevic, Slobodan Vucetic, Zoran Obradovic

We propose a Conditional Random Field (CRF) model for structured regression. By constraining the feature functions as quadratic functions of outputs, the model can be conveniently represented in a Gaussian canonical form. We improved the representational power of the resulting Gaussian CRF (GCRF) model by (1) introducing an adaptive feature function that can learn nonlinear relationships between inputs and outputs and (2) allowing the weights of feature functions to be dependent on inputs. Since both the adaptive feature functions and weights can be constructed using feedforward neural networks, we call the resulting model Neural GCRF. The appeal of Neural GCRF is in conceptual simplicity and computational efficiency of learning and inference through use of sparse matrix computations. Experimental evaluation on the remote sensing problem of aerosol estimation from satellite measurements and on the problem of document retrieval showed that Neural GCRF is more accurate than the benchmark predictors.

Session 27: Partially and semi-supervised learning

Room: 102 **Chair: Hendrik Blockeel**

11:40 - 12:00

Statistical Hypothesis Testing in Positive Unlabelled Data

Konstantinos Sechidis, Borja Calvo, Gavin Brown

We propose a set of novel methodologies which enable valid statistical hypothesis testing when we have only positive and unlabelled (PU) examples. This type of problem, a special case of semi-supervised data, is common in text mining, bioinformatics, and computer vision. Focusing on a generalised likelihood ratio test, we have 3 key contributions: (1) a proof that assuming all unlabelled examples are negative cases is sufficient for independence testing, but not for power analysis activities; (2) a new methodology that compensates this and enables power analysis, allowing sample size determination for observing an effect with a desired power; and finally, (3) a new capability, supervision determination, which can determine a-priori the number of labelled examples the user must collect before being able to observe a desired statistical effect. Beyond general hypothesis testing, we suggest the tools will additionally be useful for information theoretic feature selection, and Bayesian Network structure learning.

12:00 - 12:20

Hetero-Labeled LDA: A partially supervised topic model with heterogeneous labels

Dongyeop Kang, Youngja Park, Suresh Chari

We propose, Hetero-Labeled LDA (hLLDA), a novel semi-supervised topic model, which can learn from multiple types of labels such as document labels and feature labels, and also accommodate labels for only a subset of classes (i.e., partial labels). This addresses two major limitations in existing semi-supervised learning methods: they can incorporate only one type of domain knowledge (e.g. document labels or feature labels), and they assume that provided labels cover all the classes in the problem space. This limits their applicability in real-life situations where domain knowledge for labeling comes in different forms from different groups of domain experts and some classes may not have labels. hLLDA resolves both the label heterogeneity and label partialness problems in a unified generative process. hLLDA can leverage different forms of supervision and discover semantically coherent topics by exploiting domain knowledge mutually reinforced by different types of labels. Experiments with three document collections–Reuters, 20Newsgroup and Delicious– validate that our model generates a better set of topics and efficiently discover additional latent topics not covered by the labels resulting in better classification and clustering accuracy than existing supervised or semi-supervised topic models. The empirical results demonstrate that learning from multiple forms of domain knowledge in a unified process creates an enhanced combined effect that is greater than a sum of multiple models learned separately with one type of supervision.

12:20 - 12:40

Semi-Supervised Learning Using an Unsupervised Atlas

Chris Russell, Lourdes Agapito , Nikos Pitelis

In many machine learning problems, high-dimensional datasets often lie on or near manifolds of locally low-rank. This knowledge can be exploited to avoid the ""curse of dimensionality"" when learning a classifier. Explicit manifold learning formulations such as LLE are rarely used for this purpose, and instead classifiers may make use of methods such as local co-ordinate coding or auto-encoders to implicitly characterise the manifold. We propose novel manifold-based kernels for semi-supervised and supervised learning. We show how smooth classifiers can be learnt from existing descriptions of manifolds that characterise the manifold as a set of piecewise affine charts, or an atlas. We experimentally validate the importance of this smoothness vs. the more natural piecewise smooth classifiers, and we show a significant improvement over competing methods on standard datasets. In the semi-supervised learning setting our experiments show how using unlabelled data to learn the detailed shape of the underlying manifold substantially improves the accuracy of a classifier trained on limited labelled data.

12:40 - 13:00

Consistency of losses for learning from weak labels

Jesus Cid Sueiro, Raul Santos-Rodriguez, Dario Garcia-Garcia

In this paper we analyze the consistency of loss functions for learning from weakly labelled data, and its relation to properness. We show that the consistency of a given loss depends on the mixing matrix, which is the transition matrix relating the weak labels and the true class. A linear transformation can be used to convert a conventional classification-calibrated (CC) loss into a weak CC loss. By comparing the maximal dimension of the set of mixing matrices that are admissible for a given CC loss with that for proper losses, we show that classification calibration is a much less restrictive condition than properness. Moreover, we show that while the transformation of conventional proper losses into a weak proper losses does not preserve convexity in general, conventional convex CC losses can be easily transformed into weak and convex CC losses. Our analysis provides a general procedure to construct convex CC losses, and to identify the set of mixing matrices admissible for a given transformation. Several examples are provided to illustrate our approach.

Session 28: Reliable prediction

Room: 105 **Chair: Indrė Žliobaitė**

11:40 - 12:00

Transductive Minimax Probability Machine

Gao Huang, Shiji Song, Zhixiang Xu, Kilian Weinberger

The Minimax Probability Machine (MPM) is an elegant machine learning algorithm for inductive learning. It learns a classifier that minimizes an upper bound on its own generalization error. In this paper, we extend its celebrated inductive formulation to an equally elegant transductive learning algorithm. In the transductive setting, the label assignment of a test set is already optimized during training. This optimization problem is an intractable mixed-integer programming. Thus, we provide an efficient label-switching approach to solve it approximately. The resulting method scales naturally to large data sets and is very efficient to run. In comparison with nine competitive algorithms on eleven data sets, we show that the proposed Transductive MPM (TMPM) almost outperforms all the other algorithms in both accuracy and speed.

12:00 - 12:20

Regression Conformal Prediction with Random Forests

Ulf Johansson, Henrik Boström, Tuve Löfström, Henrik Linusson

Regression conformal prediction produces prediction intervals that are valid, i.e., the probability of excluding the correct target value is bounded by a predefined confidence level. The most important criterion when comparing conformal regressors is efficiency; the prediction intervals should be as tight (informative) as possible. In this study, the use of random forests as the underlying model for regression conformal prediction is investigated and compared to existing state-of-the-art techniques, which are based on neural networks and k-nearest neighbors. In addition to their robust predictive performance, random forests allow for determining the size of the prediction intervals by using out-of-bag estimates instead of requiring a separate calibration set. An extensive empirical investigation, using 33 publicly available data sets, was undertaken to compare the use of random forests to existing state-of-the-art conformal predictors. The results show that the suggested approach, on almost all confidence levels and using both standard and normalized nonconformity functions, produced significantly more efficient conformal predictors than the existing alternatives.

12:20 - 12:40

Combination of one-class support vector machines for classification with reject option

Blaise Hanczar, Michele Sebag

This paper focuses on binary classification with reject option, enabling the classifier to detect and abstain hazardous decisions. While reject classification produces in more reliable decisions, there is a tradeoff between accuracy and rejection rate. Two type of rejection are considered: ambiguity and outlier rejection. The state of the art mostly handles ambiguity rejection and ignored outlier rejection. The proposed approach, referred as CONSUM, handles both ambiguity and outliers detection. Our method is based on a quadratic constrained optimization formulation, combining one-class support vector machines. An adaptation of the sequential minimal optimization algorithm is proposed to solve the minimization problem. The experimental study on both artificial and real world datasets exams the sensitivity of the CONSUM with respect to the hyper-parameters and demonstrates the superiority of our approach.

12:40 - 13:00

Cautious Ordinal Classification by Binary Decomposition

Sebastien Destercke, Gen Yang

We study the problem of performing cautious inferences for an ordinal classification (a.k.a. ordinal regression) task, that is when the possible classes are totally ordered. By cautious inference, we mean that we may produce partial predictions when available information is insufficient to provide reliable precise ones. We do so by estimating probabilistic bounds instead of precise ones. These bounds induce a (convex) set of possible probabilistic models, from which we perform inferences. As the estimates or predictions for such models are usually computationally harder to obtain than for precise ones, we study the extension of two binary decomposition strategies that remain easy to obtain and computationally efficient to manipulate when shifting from precise to bounded estimates. We demonstrate the possible usefulness on such a cautious attitude on tests performed on benchmark data sets.

Session 29: Multi-target and transfer learning

Room: 106 **Chair: Michelangelo Ceci**

11:40 - 12:00

Multi-Target Regression via Random Linear Target Combinations

Grigorios Tsoumakas, Eleftherios Spyromitros-Xioufis, Aikaterini Vrekou, Ioannis Vlahavas

Multi-target regression is concerned with the simultaneous prediction of multiple continuous target variables based on the same set of input variables. It arises in several interesting industrial and environmental application domains, such as ecological modelling and energy forecasting. This paper presents an ensemble method for multi-target regression that constructs new target variables via random linear combinations of existing targets. We discuss the connection of our approach with multi-label classification algorithms, in particular RAkEL, which originally inspired this work, and a family of recent multi-label classification algorithms that involve output coding. Experimental results on 12 multi-target datasets show that it performs significantly better than a strong baseline that learns a single model for each target using gradient boosting and compares favourably to the state-of-the-art multiobjective random forest approach. The experiments further show that our approach improves more when stronger unconditional dependencies exist among the targets.

12:00 - 12:20

Transfer Learning with Multiple Sources via Consensus Regularized Autoencoders

Fuzhen Zhuang, Xiaohu Cheng, Sinno Jialin Pan, Qing He, zhongzhi Shi

Knowledge transfer from multiple source domains to a target domain is crucial in transfer learning. Most existing methods are focused on learning weights for different domains based on the similarities between each source domain and the target domain or learning more precise classifiers from the source domain data jointly by maximizing their consensus of predictions on the target domain data. However, these methods only consider measuring similarities or building classifiers on the original data space, and fail to discover a more powerful feature representation of the data when transferring knowledge from multiple source domains to the target domain. In this paper, we propose a new framework for transfer learning with multiple source domains. Specifically, in the proposed framework, we adopt autoencoders to construct a feature mapping from an original instance to a hidden representation, and train multiple classifiers from the source domain data jointly by performing an entropy-based consensus regularizer on the predictions on the target domain. Based on the framework, a particular solution is proposed to learn the hidden representation and classifiers simultaneously. Experimental results on image and text real-world datasets demonstrate the effectiveness of our proposed method compared with state-of-the-art methods.

12:20 - 12:40

Importance Weighted Inductive Transfer Learning for Regression

Thomas Vanck, Jochen Garcke

We consider inductive transfer learning for dataset shift, a situation in which the distributions of two sampled, but closely related, datasets differ. When the target data to be predicted is scarce, one would like to improve its prediction by employing data from the other, secondary, dataset. Transfer learning tries to address this task by suitably compensating such a dataset shift. In this work we assume that the distributions of the covariates and the dependent variables can differ arbitrarily between the datasets. We propose two methods for regression based on importance weighting. Here to each instance of the secondary data a weight is assigned such that the data contributes positively to the prediction of the target data. Experiments show that our method yields good results on benchmark and real world datasets.

12:40 - 13:00

Domain adaptation with regularized optimal transport

Nicolas Courty, Devis Tuia, Rémi Flamary

We present a new and original method to solve the domain adaptation problem using optimal transport. By searching for the best transportation plan between the probability distribution functions of a source and a target domain, a non-linear and invertible transformation of the learning set samples can be estimated. Any standard machine learning method can then be applied on the transformed set, which makes our method very generic. We propose an new optimal transport algorithm that incorporates the information contained in the labels available in the learning set in the optimization: this is achieved by combining an efficient matrix scaling technique together with a majoration of a non-convex regularization term. By using the proposed optimal transport with label regularization, we obtain significant increases of the performances compared with the original transport solution. The proposed algorithm is computationally efficient and effective, as illustrated by its evaluation on a toy example and a challenging real life vision dataset, against which it achieves competitive results with respect to state-of-the-art methods.

Nectar Session 4:

Room: 103-104 **Chair: Marc Plantevit**

14:20 - 14:50

Network reconstruction for the identification of miRNA:mRNA interaction networks

Gianvito Pio, Michelangelo Ceci, Domenica D'Elia, Donato Malerba

Network reconstruction from data is a data mining task which is receiving a significant attention due to its applicability in several domains. For example, it can be applied in social network analysis, where the goal is to identify connections among users and, thus, sub-communities. Another example can be found in computational biology, where the goal is to identify previously unknown relationships among biological entities and, thus, relevant interaction networks. Such task is usually solved by adopting methods for link prediction and for the identification of relevant sub-networks. Focusing on the biological domain we proposed two methods for learning to combine the output of several link prediction algorithms and for the identification of biological significant interaction networks involving two important types of RNA molecules, i.e. microRNAs (miRNAs) and messenger RNAs (mRNAs). The relevance of this application comes from the importance of identifying (previously unknown) regulatory and cooperation activities for the understanding of the biological roles of miRNAs and mRNAs. In this paper, we review the contribution given by the combination of the proposed methods for network reconstruction and the solutions we adopt in order to meet specific challenges coming from the specific domain we consider.

14:50 - 15:20

Analyzing and Grounding Social Interaction in Online and Offline Networks

Martin Atzmueller

In social network analysis, there are a variety of options for investigating social interactions. This paper reviews our recent work on analyzing and grounding social interactions in online and offline networks considering distributional semantics, structural network correlation and network inter-dependencies. Specifically, we focus on the analysis of user relatedness, community structure, and relations on online and offline networks. We discuss findings and results that justify the use of even implicitly accruing social interaction networks for the analysis of user-relatedness, community structure, etc. Furthermore, we provide insights into recent work on analyzing and grounding offline social networks.

15:20 - 15:50

Agents Teaching Agents in Reinforcement Learning (Nectar Abstract)

Matthew Taylor, Lisa Torrey

Using reinforcement learning (RL), agents can autonomously learn a control policy to master sequential-decision tasks. Rather than always learning tabula rasa, our recent work considers how an experienced RL agent, the teacher, can help another RL agent, the student, to learn. As a motivating example, consider a household robot that has learned to perform tasks in a household. When the consumer purchases a new robot, she would like the student robot to quickly learn to perform the same tasks as the teacher robot, even if the new robot has different state representation, learning method, or manufacturer. Our goals are to: 1) Allow the student to learn faster with the teacher than without it, 2) Allow the student and teacher to have different learning methods and knowledge representations, 3) Not limit the student's performance when the teacher is sub-optimal, 4) Not require a complex, shared language, and 5) Limit the amount of communication required between the agents.

Session 30: Support vector machines

Room: 102 **Chair: Ulf Brefeld**

16:20 - 16:40

Active Learning for Support Vector Machines with Maximal Model Change

Wenbin Cai, Ya Zhang, Siyuan Zhou, Wenquan Wang, Chris Ding, Xiao Gu

Margin-based strategies and model change based strategies represent two important types of strategies for active learning. While margin-based strategies have been dominant for Support Vector Machines (SVMs), most methods are based on heuristics and lack a solid theoretical support. In this paper, we propose an active learning strategy for SVMs based on Maximum Model Change (MMC). The model change is defined as the difference between the current model parameters and the updated parameters obtained with the enlarged training set. Inspired by Stochastic Gradient Descent (SGD) update rule, we measure the change as the gradient of the loss at a candidate point. We analyze the convergence property of the proposed method, and show that the upper bound of label requests made by MMC is smaller than passive learning. Moreover, we connect the proposed MMC algorithm with the widely used simple margin method in order to provide a theoretical justification for margin-based strategies. Extensive experimental results on various benchmark data sets from UCI machine learning repository have demonstrated the effectiveness and efficiency of the proposed method.

16:40 - 17:00

Support Vector Machines for Differential Prediction

Finn Kuusisto, Vitor Santos Costa, Houssam Nassif, Elizabeth Burnside, David Page, Jude Shavlik

Machine learning is continually being applied to a growing set of fields, including the social sciences, business, and medicine. Some fields present problems that are not easily addressed using standard machine learning approaches and, in particular, there is growing interest in differential prediction. In this type of task we are interested in producing a classifier that specifically characterizes a subgroup of interest by maximizing the difference in predictive performance for some outcome between subgroups in a population. We discuss adapting maximum margin classifiers for differential prediction. We first introduce multiple approaches that do not affect the key properties of maximum margin classifiers, but which also do not directly attempt to optimize a standard measure of differential prediction. We next propose a model that directly optimizes a standard measure in this field, the uplift measure. We evaluate our models on real data from two medical applications and show excellent results.

17:00 - 17:20

Accelerating Model Selection with Safe Screening for L1-Regularized L2-SVM

Zheng Zhao, Jun Liu, James Cox

The L1-regularized support vector machine (SVM) is a powerful predictive learning model that can generate sparse solutions. Compared to a dense solution, a sparse solution is usually more interoperable and more effective for removing noise and preserving signals. The L1-regularized SVM has been successfully applied in numerous applications to solve problems from text mining, bioinformatics, and image processing. The regularization parameter has a significant impact on the performance of an L1-regularized SVM model. Therefore, model selection needs to be performed to choose a good regularization parameter. In model selection, one needs to learn a solution path using a set of predefined parameter values. Therefore, many L1-regularized SVM models need to be fitted, which is usually very time consuming. This paper proposes a novel safe screening technique to accelerate model selection for the L1-regularized L2-SVM, which can lead to much better efficiency in many scenarios. The technique can successfully identify most inactive features in an optimal solution of the L1-regularized L2-SVM model and remove them before training. To achieve safe screening, the technique solves a minimization problem for each feature on a convex set that is formed by the intersection of a tight n-dimensional hyperball and the upper half-space. An efficient algorithm is designed to solve the problem based on zero-finding. Every feature that is removed by the proposed technique is guaranteed to have zero weight in the optimal solution. Therefore, an L1-regularized L2-SVM solver achieves exactly the same result by using only the selected features as when it uses the full feature set. Empirical study on high-dimensional benchmark data sets produced promising results and demonstrated the effectiveness of the proposed technique.

17:20 - 17:40

A Theoretical and Empirical Analysis of Support Vector Machine Methods for Multiple-Instance Classification

Gary Doran, Soumya Ray

The standard support vector machine (SVM) formulation, widely used for supervised learning, possesses several intuitive and desirable properties. In particular, it is convex and assigns zero loss to solutions if, and only if, they correspond to consistent classifying hyperplanes with some nonzero margin. The traditional SVM formulation has been heuristically extended to multiple-instance (MI) classification in various ways. In this work, we analyze several such algorithms and observe that all MI techniques lack at least one of the desirable properties above. Further, we show that this tradeoff is fundamental, stems from the topological properties of consistent classifying hyperplanes for MI data, and is related to the computational complexity of learning MI hyperplanes. We then study the empirical consequences of this three-way tradeoff in MI classification using a large group of algorithms and datasets. We find that the experimental observations generally support our theoretical results, and properties such as the labeling task (instance versus bag labeling) influence the effects of different tradeoffs.

Session 31: Privacy and anti-discrimination in data mining

Room: 105 **Chair: Bart Goethals**

16:20 - 16:40

Anti-discrimination Analysis Using Privacy Attack Strategies

Salvatore Ruggieri, Sara Hajian, Faisal Kamiran, Xiangliang Zhang

Social discrimination discovery from data is an important task to identify illegal and unethical discriminatory patterns towards protected-by-law groups, e.g., ethnic minorities. We deploy privacy attack strategies as tools for discrimination discovery under hard assumptions which have rarely tackled in the literature: indirect discrimination discovery, privacy-aware discrimination discovery, and discrimination data recovery. The intuition comes from the intriguing parallel between the role of the anti-discrimination authority in the three scenarios above and the role of an attacker in private data publishing. We design strategies and algorithms inspired/based on Frèchet bounds attacks, attribute inference attacks, and minimality attacks to the purpose of unveiling hidden discriminatory practices. Experimental results show that they can be effective tools in the hands of anti-discrimination authorities.

16:40 - 17:00

Neutralized Empirical Risk Minimization with Generalization Neutrality Bound

Kazuto Fukuchi, Jun Sakuma

Currently, machine learning plays an important role in the lives and individual activities of numerous people. Accordingly, it has become necessary to design machine learning algorithms to ensure that discrimination, biased views, or unfair treatment do not result from decision making or predictions made via machine learning. In this work, we introduce a novel empirical risk minimization (ERM) framework for supervised learning, neutralized ERM (NERM) that ensures that any classifiers obtained can be guaranteed to be neutral with respect to a viewpoint hypothesis. More specifically, given a viewpoint hypothesis, NERM works to find a target hypothesis that minimizes the empirical risk while simultaneously identifying a target hypothesis that is neutral to the viewpoint hypothesis. Within the NERM framework, we derive a theoretical bound on empirical and generalization neutrality risks. Furthermore, as a realization of NERM with linear classification, we derive a max-margin algorithm, neutral support vector machine (SVM). Experimental results show that our neutral SVM shows improved classification performance in real datasets without sacrificing the neutrality guarantee.

17:00 - 17:20

Generalization-based Privacy Preservation and Discrimination Prevention in Data Publishing and Mining

Sara Hajian, Josep Domingo-Ferrer, Oriol Farras

Living in the information society facilitates the automatic collection of huge amounts of data on individuals, organizations, etc. Publishing such data for secondary analysis (e.g. learning models and finding patterns) may be extremely useful to policy makers, planners, marketing analysts, researchers and others. Yet, data publishing and mining do not come without dangers, namely privacy invasion and also potential discrimination of the individuals whose data are published. Discrimination may ensue from training data mining models (e.g. classifiers) on data which are biased against certain protected groups (ethnicity, gender, political preferences, etc.). The objective of this paper is to describe how to obtain data sets for publication that are: i) privacy-preserving; ii) unbiased regarding discrimination; and iii) as useful as possible for learning models and finding patterns. We present the first generalization- based approach to simultaneously offer privacy preservation and discrimination prevention. We formally define the problem, give an optimal algorithm to tackle it and evaluate the algorithm in terms of both general and specific data analysis metrics (i.e. various types of classifiers and rule induction algorithms). It turns out that the impact of our transformation on the quality of data is the same or only slightly higher than the impact of achieving just privacy preservation. In addition, we show how to extend our approach to different privacy models and anti-discrimination legal concepts.

17:20 - 17:40

Preserving Worker Privacy in Crowdsourcing

Hiroshi Kajino, Hiromi Arai, Hisashi Kashima

This paper proposes a crowdsourcing quality control method with worker-privacy preservation. Crowdsourcing allows us to outsource tasks to a number of workers. The results of tasks obtained in crowdsourcing are often low-quality due to the difference in the degree of skill. Therefore, we need quality control methods to estimate reliable results from low-quality results. In this paper, we point out privacy problems of workers in crowdsourcing. Personal information of workers can be inferred from the results provided by each worker. To formulate and to address the privacy problems, we define a worker-private quality control problem, a variation of the quality control problem that preserves privacy of workers. We propose a worker-private latent class protocol where a requester can estimate the true results with worker privacy preserved. The key ideas are decentralization of computation and introduction of secure computation. We theoretically guarantee the security of the proposed protocol and experimentally examine the computational efficiency and accuracy.

Session 32: Probabilistic and Bayesian methods

Room: 106 **Chair: Szymon Jaroszewicz**

16:20 - 16:40

Linear State-Space Model with Time-Varying Dynamics

Jaakko Luttinen, Tapani Raiko, Alexander Ilin

This paper introduces a linear state-space model with time-varying dynamics. The time dependency is obtained by forming the state dynamics matrix as a time-varying linear combination of a set of matrices. The time dependency of the weights in the linear combination is modelled by another linear Gaussian dynamical model allowing the model to learn how the dynamics of the process changes. Previous approaches have used switching models which have a small set of possible state dynamics matrices and the model selects one of those matrices at each time, thus jumping between them. Our model forms the dynamics as a linear combination and the changes can be smooth and more continuous. The model is motivated by physical processes which are described by linear partial differential equations whose parameters vary in time. An example of such a process could be a temperature field whose evolution is driven by a varying wind direction. The posterior inference is performed using variational Bayesian approximation. The experiments on stochastic advection-diffusion processes and real-world weather processes show that the model with time-varying dynamics can outperform previously introduced approaches.

16:40 - 17:00

Nonparametric Markovian Learning of Triggering Kernels for Mutually Exciting and Mutually Inhibiting Multivariate Hawkes Processes

Remi Lemonnier, Nicolas Vayatis

In this paper, we address the problem of fitting multivariate Hawkes processes to potentially large-scale data in a setting where series of events are not only mutually-exciting but can also exhibit inhibitive patterns. We focus on nonparametric learning and propose a novel algorithm called MEMIP (Markovian Estimation of Mutually Interacting Processes) that makes use of polynomial approximation theory and selfconcordant analysis in order to learn both triggering kernels and base intensities of events. Moreover, considering that N historical observations are available, the algorithm performs log-likelihood maximization in O(N) operations, while the complexity of non-Markovian methods is in O(N2). Numerical experiments on simulated data, as well as real-world data, show that our method enjoys improved prediction performance when compared to state-of-the art methods like MMEL and exponential kernels.

17:00 - 17:20

Bayesian Models for Structured Sparse Estimation via Set Cover Prior

Xianghang Liu, Xinhua Zhang, Tiberio Caetano

A number of priors have been recently developed for Bayesian estimation of sparse models. In many applications the variables are simultaneously relevant or irrelevant in groups, and appropriately modeling this correlation is important for improved sample efficiency. Although group sparse priors are also available, most of them are either limited to disjoint groups, or do not infer sparsity at group level, or fail to induce appropriate patterns of support in the posterior. In this paper we tackle this problem by proposing a new framework of prior for overlapped group sparsity. It follows a hierarchical generation from group to variable, allowing group-driven shrinkage and relevance inference. It is also connected with set cover complexity in its maximum a posterior. Analysis on shrinkage profile and conditional dependency unravels favorable statistical behavior compared with existing priors. Experimental results also demonstrate its superior performance in sparse recovery and compressive sensing.

17:20 - 17:40

Cutset Networks: A Simple, Tractable, and Scalable Approach for Improving the Accuracy of Chow-Liu Trees

Tahrima Rahman, Prasanna Kothalkar, Vibhav Gogate

In this paper, we present cutset networks, a new tractable probabilistic model for representing multi-dimensional discrete distributions. Cutset networks are rooted OR search trees, in which each OR node represents conditioning of a variable in the model, with tree Bayesian networks (Chow-Liu trees) at the leaves. From an inference point of view, cutset networks model the mechanics of Pearl's cutset conditioning algorithm, a popular exact inference method for probabilistic graphical models. We present efficient algorithms, which leverage and adopt vast amount of research on decision tree induction for learning cutset networks from data. We also present an expectation-maximization (EM) algorithm for learning mixtures of cutset networks. Our experiments on a wide variety of benchmark datasets clearly demonstrate that compared to approaches for learning other tractable models such as thin-junction trees, latent tree models, arithmetic circuits and sum-product networks, our approach is significantly more scalable, and provides similar or better accuracy.

Session 33: Time-evolving graphs and Dynamic Networks

Room: 103-104 **Chair: Pauli Miettinen**

16:20 - 16:40

Fast Nearest Neighbor Search on Large Time-Evolving Graphs

Leman Akoglu, Rohit Khandekar, Vibhore Kumar, Srinivasan Parthasarathy, Deepak Rajan, Kun-Lung Wu, Christos Faloutsos

Finding the k nearest neighbors (k-NNs) of a given vertex in a graph has many applications such as link prediction, keyword search, and image tagging. An established measure of vertex-proximity in graphs is the Personalized Page Rank (PPR) score based on random walk with restarts. Since PPR scores have long-range correlations, computing them accurately and efficiently is challenging when the graph is too large to fit in main memory, especially when it also changes over time. In this work, we propose an efficient algorithm to answer PPR-based k-NN queries in large time-evolving graphs. Our key approach is to use a divide-and-conquer framework and efficiently compute answers in a distributed fashion. We represent a given graph as a collection of dense vertex- clusters with their inter connections. Each vertex-cluster maintains certain information related to internal random walks and updates this information as the graph changes. At query time, we combine this information from a small set of relevant clusters and compute PPR scores efficiently. We validate the effectiveness of our method on large real-world graphs from diverse domains. To the best of our knowledge, this is one of the few works that simultaneously addresses answering k-NN queries in possibly disk-resident and time-evolving graphs.

16:40 - 17:00

Scalable Information Flow Mining in Networks

Karthik Subbian, Chidananda Sridhar, Charu Aggarwal, Jaideep Srivastava

The problem of understanding user activities and their patterns of communication is extremely important in social and collaboration networks. This can be achieved by tracking the dominant content flow trends and their interactions between users in the network. Our approach tracks all possible paths of information flow using its network structure, content propagated and the time of propagation. We also show that the complexity class of this problem is \#P-complete. Because most social networks have many activities and interactions, it is inevitable the proposed method will be computationally intensive. Therefore, we propose an efficient method for mining information flow patterns, especially in large networks, using distributed vertex-centric computational models. We use the Gather-Apply-Scatter (GAS) paradigm to implement our approach. We experimentally show that our approach achieves over three orders of magnitude advantage over the state-of-the-art, with an increasing advantage with a greater number of cores. We also study the effectiveness of the discovered content flow patterns by using it in the context of an influence analysis application.

17:00 - 17:20

Discovering Bands from Graphs

Nikolaj Tatti

Discovering the underlying structure of a given graph is one of the fundamental goals in graph mining. Given a graph, we can often order vertices in a way that neighboring vertices have a higher probability of being connected to each other. This implies that the edges form a band around the diagonal in the adjacency matrix. Such structure may rise for example if the graph was created over time: each vertex had an active time interval during which the vertex was connected with other active vertices. The goal of this paper is to model this phenomenon. To this end, we formulate an optimization problem: given a graph and an integer K, we want to order graph vertices and partition the ordered adjacency matrix into K bands such that bands closer to the diagonal are more dense. We measure the goodness of a segmentation using the log-likelihood of a log-linear model, a flexible family of distributions containing many standard distributions. We divide the problem into two subproblems: finding the order and finding the bands. We show that discovering bands can be done in polynomial time with isotonic regression, and we also introduce a heuristic iterative approach. For discovering the order we use Fiedler order accompanied with a simple combinatorial refinement. We demonstrate empirically that our heuristic works well in practice.

17:20 - 17:40

Communication-Efficient Distributed Online Prediction by Decentralized Variance Monitoring

Michael Kamp, Mario Boley, Daniel Keren, Assaf Schuster, Izchak Sharfman

We present the first protocol for distributed online prediction that aims to minimize online prediction loss and network communication at the same time. This protocol can be applied wherever a prediction-based service must be provided timely for each data point of a multitude of high frequency data streams, each of which is observed at a local node of some distributed system. Exemplary applications include social content recommendation and algorithmic trading. The challenge is to balance the joint predictive performance of the nodes by exchanging information between them, while not letting communication overhead deteriorate the responsiveness of the service. Technically, the proposed protocol is based on controlling the variance of the local models in a decentralized way. This approach retains the asymptotic optimal regret of previous algorithms. At the same time, it allows to substantially reduce network communication, and, in contrast to previous approaches, it remains applicable when the data is non-stationary and shows rapid concept drift. We demonstrate empirically that the protocol is able to hold up a high predictive performance using only a fraction of the communication required by benchmark methods.

FRIDAY 19 SEPTEMBER 2014

FRIDAY INVITED TALK

Network Analysis in the Big Data Age: Mining Graph and Social Streams

Speaker: Charu Aggarwal
Time: 09:00 – 10:00
Room: Auditorium 300

Abstract
The advent of large interaction-based communication and social networks has led to challenging streaming scenarios in graph and social stream analysis. The graphs that result from such interactions are large, transient, and very often cannot even be stored on disk. In such cases, even simple frequency-based aggregation operations become challenging, whereas traditional mining operations are far more complex. When the graph cannot be explicitly stored on disk, mining algorithms must work with a limited knowledge of the network structure. Social streams add yet another layer of complexity, wherein the streaming content associated with the nodes and edges needs to be incorporated into the mining process. A significant gap exists between the problems that need to be solved, and the techniques that are available for streaming graph analysis. In spite of these challenges, recent years have seen some advances in which carefully chosen synopses of the graph and social streams are leveraged for approximate analysis. This talk will focus on several recent advances in this direction.

Bio
Charu Aggarwal is a Research Scientist at the IBM T. J. Watson Research Center in Yorktown Heights, New York. He completed his B.S. from IIT Kanpur in 1993 and his Ph.D. from Massachusetts Institute of Technology in 1996. His research interest during his Ph.D. years was in combinatorial optimization (network flow algorithms), and his thesis advisor was Professor James B. Orlin. He has since worked in the field of data mining, with particular interests in data streams, privacy, uncertain data and social network analysis. He has published over 200 papers in refereed venues, and has applied for or been granted over 80 patents. Because of the commercial value of the above-mentioned patents, he has received several invention achievement awards and has thrice been designated a Master Inventor at IBM. He is a recipient of an IBM Corporate Award (2003) for his work on bio-terrorist threat detection in data streams, a recipient of the IBM Outstanding Innovation Award (2008) for his scientific contributions to privacy technology, and a recipient of an IBM Research Division Award (2008) for his scientific contributions to data stream research. He has served on the program committees of most major database/data mining conferences, and served as program vice-chairs of the SIAM Conference on Data Mining, 2007, the IEEE ICDM Conference, 2007, the WWW Conference 2009, and the IEEE ICDM Conference, 2009. He served as an associate editor of the IEEE Transactions on Knowledge and Data Engineering Journal from 2004 to 2008. He is an associate editor of the ACM TKDD Journal, an action editor of the Data Mining and Knowledge Discovery Journal, an associate editor of the ACM SIGKDD Explorations, and an associate editor of the Knowledge and Information Systems Journal. He is a fellow of the ACM (2013) and the IEEE (2010) for "contributions to knowledge discovery and data mining techniques".

FRIDAY WORKSHOPS

DARE'14: Data Analytics for Renewable Energy Integration

Wei Lee Woon, Zeyar Aung, Stuart Madnick
Room: 101

Climate change, the depletion of natural resources and rising energy costs have led to an increasing focus on renewable sources of energy. A lot of research has been devoted to the technologies used to extract energy from these sources; however, equally important is the storage and distribution of this energy in a way that is efficient and cost effective. Achieving this would generally require integration with existing energy infrastructure.

The challenge of renewable energy integration is inherently multidisciplinary and is particularly dependant on the use of techniques from the domains of data analytics, pattern recognition and machine learning. Examples of relevant research topics include the forecasting of electricity supply and demand, the detection of faults, demand response applications and many others. This workshop will provides a forum where interested researchers from the various related domains will be able to present and discuss their findings.

More information: http://dare2014.dnagroup.org/

LD4KD: Linked Data for Knowledge Discovery

Ilaria Tiddi, Mathieu d'Aquin, Nicolas Jay
Room: 205

Linked Data have attracted a lot of attention from both developers and researchers in recent years, as the underlying technologies and principles provide new ways, following the Semantic Web standards, to overcome typical data management and consumption issues such as reliability, heterogeneity, provenance or completeness. Many different areas of research, from social media analysis to biomedical research, have adopted these principles both for the management and dissemination of their own data and for the combined reuse of external data sources. However, the way in which Linked Data can be applicable and beneficial to the Knowledge Discovery (KDD) process is still not completely understood. It is therefore worth exploring the question of the benefit of Linked Data principles and technologies for knowledge discovery, together with addressing the new challenges that will emerge from joining the two fields, beyond the traditional data management and consumption issues in KDD.

While one of the most obvious scenarios for using Linked Data in a KDD process is the representation of the underlying data following Semantic Web standards, many other aspects of KDD can benefit from including some elements of Linked Data, as a way to reuse external data or to produce new information that can be easily shared or integrated. Contributions here might range from the mining of Linked Data sources to the use of Linked Data to enrich and integrate local data for the purpose of data preparation, results interpretation or visualisation.

This interdisciplinary workshop will therefore provide a forum for researchers to discuss and investigate established as well as potential avenues for interaction and cross-fertilisation between the two fields of Knowledge Discovery and Linked Data, both considered broadly. It will be an opportunity for practitioners of both fields to create communication and collaboration channels, in which they will be able to share their experience and bridge the gap between these overlapping, but mostly isolated communities.

This workshop is kindly supported by www.KDnuggets.com ： Analytics, Data Mining, & Data Science Resources

More information: http://events.kmi.open.ac.uk/ld4kd2014/

NFmcp 2014: New Frontiers in Mining Complex Patterns

Annalisa Appice, Michelangelo Ceci, Corrado Loglisci, Giuseppe Manco, Elio Masciari, Zbigniew W. Ras
Room: 203

Modern automatic systems are able to collect huge volumes of data, often with a complex structure (e.g. multi-table data, XML data, web data, time series and sequences, graphs and trees). This fact poses new challenges for current information systems with respect to storing, managing and mining these sets of complex data. The purpose of this workshop is to bring together researchers and practitioners of data mining who are interested in the advances and latest developments in the area of extracting patterns from complex data sources like blogs, event or log data, medical data, spatio-temporal data, social networks, mobility data, sensor data and streams, and so on. The workshop aims at integrating recent results from existing fields such as data mining,statistics, machine learning and relational databases to discuss and introduce new algorithmic foundations and representation formalisms in pattern discovery. We are interested in advanced techniques which preserve the informative richness of data and allow us to efficiently and efficaciously identify complex information units present in such data.

More information: http://www.di.uniba.it/~loglisci/NFmcp2014/

LMCE 2014: Generalization and reuse of machine learning models over multiple contexts

Cèsar Ferri, Peter Flach, Nicolas Lachiche
Room: 202

Adaptive reuse of learnt knowledge is of critical importance in the majority of knowledge-intensive application areas, particularly when the context in which the learnt model operates can be expected to vary from training to deployment. In machine learning this has been studied, for example, in relation to variations in class and cost skew in (binary) classification, leading to the development of tools such as ROC analysis to adjust decision thresholds to operating conditions concerning class and cost skew. More recently, considerable effort has been devoted to research on transfer learning, domain adaptation, and related approaches.

Given that the main business of predictive machine learning is to generalise from training to deployment, there is clearly scope for developing a general notion of operating context. Without such a notion, a model predicting sales in Prague for this week may perform poorly in Nancy for next Wednesday. The operating context has changed in terms of location as well as resolution. While a given predictive model may be sufficient and highly specialised for one particular operating context, it may not perform well in other contexts. If sufficient training data for the new context is available it might be feasible to retrain a new model; however, this is generally not a good use of resources, and one would expect it to be more cost-effective to learn one general, versatile model that effectively generalizes over multiple and possibly previously unseen contexts.

The aim of this workshop is to bring together people working in areas related to versatile models and model reuse over multiple contexts. Given the advances made in recent years on specific approaches such as transfer learning, an attempt to start developing an overarching theory is now feasible and timely, and can be expected to generate considerable interest from the machine learning community. Papers are solicited in all areas relating to model reuse and model generalization including the following areas: transfer learning, data shift and concept drift, domain adaptation, transductive learning, multi-task learning, ROC analysis and cost-sensitive learning, background knowledge, relational learning, context-aware applications, incomplete information, abduction, meta-learning.

More information: http://users.dsic.upv.es/~flip/LMCE2014/

LEMA 2014: Learning with Multiple Views: Applications to Computer Vision and Multimedia Mining

Stéphane Ayache, Matthieu Cord, François-Xavier Dupé, Emilie Morvant
Room: 102

Recent years have witnessed new frameworks/algorithms able to deal with multiple views, such as Multiple Kernel Learning (MKL), Boosting, Co-regularized approach. Such algorithms come from the Machine Learning community and find applications in many different areas, such as Multimedia Indexing, Computer Vision, Bio-informatics, Neuro-imaging... Multiview learning, naturally enough, emphasise the potential benefits of learning through collaboration with multiple sources of data (e.g. video document can be described through images, sound, motion, text). Depending on the context, this issue of learning from multiple descriptions of data goes under the name of multiview learning (machine learning, computer vision), multimodality fusion (multimedia), among others. This workshop is the opportunity to bring together theoretical and applicative communities around multiview learning, which could lead to significant contributions in Machine Learning, Multimedia Mining and Computer Vision.

This workshop builds upon successful previous machine learning workshops on multiview learning or connections between ML and applications, like Machine Learning techniques for processing multimedia content (ICML 2005), Learning with multiples views (ICML 2005), Learning from multiples sources (NIPS 2008), Learning from multiples sources with applications to robotics (NIPS 2009) where links between theory and applications of the multiview paradigms are made. The literature and the advances on multiview learning have grown up to a point where a broad synthesis is required.

More information: http://qarma.lif.univ-mrs.fr/lema/

FRIDAY MORNING DISCOVERY CHALLENGE

C2: Predictive Web Analytics

Carlos Castillo, Joshua Schwartz
Time: 10:00 – 13:00
Room: 106

Studying the behavior of visitors on web sites is extremely important: the number of visitors and social media reactions to a web site are indirect measures of its influence, and in the case of commercial operations, typically translate into income and profits.

Predictive tasks on the web are an active research topic. This includes, for instance, predicting the number of visits an article will obtain using signals from early visits to it. Until now, such tasks have done on proprietary data because behavioral data is delicate from the point of view of privacy, and it is also a highly sensitive business secret for most large-scale web sites. This has made it difficult to compare and evaluate different approaches.

This predictive challenge will allow for the first time to compare predictive web analytic systems on a shared dataset.

More information: https://sites.google.com/site/predictivechallenge2014/home

FRIDAY MORNING TUTORIALS

T5: The Lunch is Never Free: How Information Theory, MDL, and Statistics are Connected

Nikolaj Tatti
Room: 105

Model techniques are becoming increasingly popular in many diverse data mining subfields such as sequence mining, graph mining, and pattern mining. One particularly popular approach, due to its interpretability and practicality, is Minimum Description Length (MDL) principle which is based on information-theoretic approach. In this tutorial we present basic concepts of MDL, Information Theory, and Bayesian Statistics with the emphasis on how they are connected, and what are the consequences of these connections. These connections provide additional insights into MDL principle and information theory, provide a stronger theoretical background, and allow us to use tools from statistics, but also point out limitations that are not immediately apparent.

Outline

- Introduction
- Information Theory and Statistics
  - Definitions of Kullback-Leibler and entropy [Mac02]
  - Shannon theory on lower bound for encoding [Sha48]
  - Huffman encoding
  - Connection of entropy to the log-likelihood
- Maximum entropy models
  - Definition with the emphasis on subjectivity of the model [Csi75]
  - Maximum entropy model as a log-linear model [Csi75]
  - Algorithm for solving maximum entropy model [DR72, Csi75]
  - Comparing two log-linear models leads to a G-test
  - Mutual information as an example of G-test
- MDL and Bayesian model selection
  - Bayesian model selection and BIC [Sch78]
  - Kolmogoroff complexity [LV93]
  - Practical version based on Information Theory [Gru07]
  - MDL and connection to Bayesian model selection [Mac02]
  - Refined MDL and connection to BIC [Ris96, Gru07]
- Conclusions

T9: Deep Learning

Hugo Larochelle, Aaron Courville
Room: 103-104

Deep learning is one of the most rapidly growing areas of machine learning. It concerns the learning of multiple layers of representation that gradually transform the input into a form where a given task can be performed more effectively. Deep learning has recently been responsible for an impressive number of state-of-the-art results in a wide array of domains, including object detection and recognition, speech recognition, natural language processing tasks, bio-informatics and r einforcement learning.

Outline

In this tutorial we will cover the foundations of deep learning: neural networks, convolutional neural networks, recurrent neural networks, autoencoders and Boltzmann machines. We will discuss why models with many layers of representation can be hard to learn and present strategies that have been developed to overcome these challenges. We will also discuss more recent innovations including dropout training that has proved to be an extremely effective regularization technique for training neural networks. Finally, we will cover some concrete and successful applications of deep learning.

FRIDAY AFTERNOON TUTORIALS

T6: Information Theoretic Methods in Data Mining

Matthijs van Leeuwen, Jilles Vreeken, Arno Siebes
Room: 105

Selecting a model for a given set of data is at the heart of what data analysts do, whether they are statisticians, machine learners or data miners. However, the philosopher Hum already pointed out that the 'Problem of Induction' is unsolvable; there are infinitely many functions that touch any finite set of points. So, it is not surprising that there are many different principled approaches to guide the search for a good model. Well-known examples are Bayesian Statistics and Statistical Learning Theory.
In the last decade Information Theoretic methods to select the best model slowly but surely became popular in the data mining community. In this tutorial we present an overview of these methods. Starting from the basics — i.e., Information Theory — to how one defines and finds good sets of patterns using Information Theory to how such patterns can be used for many data mining tasks.

Outline

- Basics of Information Theory
- Patterns and Information Theory
- Information Theory for Data Mining Tasks
- Information Theory and Descriptive Data Mining

T7: Machine Learning with Analogical Proportions

Laurent Miclet, Henri Prade, Gilles Richard
Room: 106

Reasoning by analogy has been recognized as a major cognitive capability of human mind, and studied in AI, among other fields. In the last decade, there has been a renewal of interest around the notion of analogical proportion, i.e., statements of the form "a is to b as c is to d".

Formal models of analogical proportions have been proposed in various settings including sets, lattices, trees, etc. In logical terms, analogical proportion states that "a differs from b as c differs from d" and vice-versa. This shows that analogy making is both a matter of similarity and dissimilarity. Analogical proportions provide a symbolic counterpart to numerical proportions.

Instead of dealing exclusively with numbers, analogical proportions transpose the "rule of threeâ€™â€™" to symbolic items, allowing to induce a 4th item when only the 3 others are known. This is the core of analogical-based learning methods. Its interest relies on the "creative" nature of the process which looks at similar items (as in the neighborhood-based methods), but takes also advantage of dissimilar, but "parallel" cases.

The aim of this tutorial is to provide the audience with :
- an overview of computationally oriented models of analogical reasoning
- technical knowledge about the use of analogical proportions for inductive tasks

T8: Preference Learning Problems

Yann Chevaleyre, Frédéric Koriche
Room: 103-104

We will start with an overview of the various preference learning problems which have emerged these past years, including instance ranking and label ranking. We will see how these problems can be formulated as (possibly convex) optimization problems, or reduced to other well-known machine learning problems. Then, we will discuss about the main preference models and how to learn them. In particular, we will first introduce ordinal preference models, including CP-nets and lexicographic preference networks, and then discuss about utility-based models such as generalized additive independence (GAI) networks. Finally, to broaden the talk, we will mention how preference learning may be used in other setting such as Markov Decision Processes or Computational Social Choice.

Outline

- Some Preference Learning Problems
  - Instance ranking
  - Label ranking
- Learning Ordinal Preference Models
  - CP-nets
  - Lexicographic preferences
- Learning Utility-based Models
  - GAI-nets
- Beyond
  - Learning Preferences in Markov Decision Processes

PHD SESSION

Radim Belohlavek, Bruno Crémilleux
Room: 204

The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD) includes a PhD session on machine learning and knowledge discovery in databases and related application domains. The objective of the session is to provide an environment for students to exchange in an interactive atmosphere their ideas and experiences with peers and to get constructive feedback from senior researchers in machine learning, data mining and related areas.

The topics for discussion will be ideas of students and their ongoing work in preparation of their PhD dissertations and their interests in machine learning and data mining. The target audience for the session is mainly PhD students, their advisors, and interested researchers and attendees of ECML/PKDD. During the PhD session, researchers with experience in supervising and examining PhD students will be able to participate and provide feedback and advice to participants. It will be an excellent opportunity for developing person-to-person networks to the benefit of the community and PhD students in their future careers.

More information: https://phdsession-ecmlpkdd2014.greyc.fr/

ILP – Inductive Logic Programming

OVERVIEW

The 2014 edition of the annual International Conference on Inductive Logic Programming (ILP 2014) will be held in Nancy, France, on September 14-16th, 2014. Additionally, ILP 2014 will be co-located with ECML/PKDD.

Inductive Logic Programming is a subfield of machine learning that uses logic programming as a uniform representation language for examples, background knowledge and hypotheses. Due to its expressive representation formalism based on first-order logic, ILP provides suitable means for multi-relational learning and data mining.

The ILP conference is the premier international forum on logic-based and relational learning. The conference has broadened its scope and welcomes contributions to learning from multi-relational databases and non-trivially structured data, ranging from purely logic-based to alternative approaches, such as probabilistic or connectionist. Work on ILP has explored several intersections to statistical learning and other probabilistic approaches, expanding research horizons significantly.

PROGRAM

Sunday

09:00 – 09:30	Registration
09:30 – 09:45	Opening
09:45 – 10:45	**Invited talk: Helmut Simonis**
10:45 – 11:15	Coffee break
11:15 – 12:45	**Session 1: Life Science Applications**
12:45 – 14:30	Lunch
14:30 – 16:10	**Session 2: Logic & Time**
16:10 – 16:30	Coffee break
16:30 – 18:05	**Session 3: Applications**
20:00	**Gala dinner**

Monday

09:00 – 10:30	**Session 4: Probabilistic Logics & Graphs**
10:30 – 11:00	Coffee break
11:00 – 12:00	**Invited talk: Sumit Gulwani**
12:00 – 14:00	Lunch
14:00 – 15:40	**Session 5: Feature Construction & Applications**
15:40 – 16:00	Coffee break
16:00 – 17:30	**Session 6: Constraints & Predicate Invention**
18:30 – 19:30	**Invited talk: Lise Getoor** (joint with ECML/PKDD)

Tuesday

11:00 – 12:40	**Joint ECML/PKDD - ILP session**
12:40 – 14:00	Lunch
14:00 – 15:00	**Community meeting**

INVITED TALK

Helmut Simonis

Learning Constraint Models from Example Solutions

Abstract
In this talk we will give an overview of ModelSeeker, a tool to generate constraint programs for combinatorial problems from small numbers of example solutions. It is based on the Global Constraint Catalog, a detailed, executable description of over 400 global constraints, which form the basic building blocks of our models. Models are presented as a conjunction of conjunctions of global constraints over various partitions of the problem variables. The approach has been validated on more than 300 problems from different domains, ranging from puzzles to real-world problem instances in scheduling, placement and time tabling. The initial approach required that all example solutions for a problem had identical size, we now present an extension which predicts growth laws from examples of varying sizes. Finally, we will discuss an application of the approach to real-world data from EDF about production schedules for all electric power stations in France, extracting plant-specific constraints for the daily unit commitment model.

Bio
Helmut Simonis has been working on constraint programming for over 25 years. He was member of the CHIP project at ECRC in Munich, co-founder and technical director for COSYTEC in Orsay, France, and a principal research fellow at Imperial College London, as well as working for start-up companies Parc Technologies and CrossCore before joining the Cork Constraint Computation Centre (4C) at University College Cork in Ireland. His main interests are applications of constraint programming and other optimisation techniques. Recently he has been working on the automatic generation of constraint models from example solutions, and more generally on the integration of machine learning and optimisation. He currently (2013-2014) is president of the Association for Constraint Programming.

Sumit Gulwani

Applications of Program Synthesis to End-User Programming and Intelligent Tutoring Systems

Abstract
Computing devices have become widely available to billions of end users, yet a handful of experts have the needed expertise to program these devices. Automated program synthesis has the potential to revolutionize this landscape, when targeted for the right set of problems and when allowing the right interaction model. The first part of this talk discusses techniques for programming using examples and natural language. These techniques have been applied to various end-user programming domains including data manipulation and smartphone scripting. The second part of this talk presents surprising applications of program synthesis technology to automating various repetitive tasks in Education including problem, solution, and feedback generation for various subject domains including math and programming. These results advance the state-of-the-art in intelligent tutoring, and can play a significant role in enabling personalized and interactive education in both standard classrooms and MOOCs.

Bio
Sumit Gulwani is a principal researcher at Microsoft Research, and an adjunct faculty in the Computer Science Department at IIT Kanpur. He has expertise in formal methods and automated program analysis and synthesis techniques. As part of his vision to empower masses, he has recently focused on cross-disciplinary areas of automating end-user programming (for various systems like spreadsheets, smartphones, and robots), and building intelligent tutoring systems (for various subject domains including programming, logic, and math). Sumit's programming-by-example work led to the famous Flash Fill feature of Microsoft Excel 2013 that is used by hundreds of millions of people. Sumit obtained his PhD in Computer Science from UC-Berkeley in 2005, and was awarded the ACM SIGPLAN Outstanding Doctoral Dissertation Award. He obtained his BTech in Computer Science and Engineering from IIT Kanpur in 2000, and was awarded the President's Gold Medal.

Lise Getoor

Scalable Collective Reasoning using Probabilistic Soft Logic

Abstract
One of the challenges in big data analytics is to efficiently learn and reason collectively about extremely large, heterogeneous, incomplete, noisy interlinked data. Collective reasoning requires the ability to exploit both the logical and relational structure in the data and the probabilistic dependencies. In this talk I will overview our recent work on probabilistic soft logic (PSL), a framework for collective, probabilistic reasoning in relational domains. PSL is able to reason holistically about both entity attributes and relationships among the entities. The underlying mathematical framework, which we refer to as a hinge-loss Markov random field, supports extremely efficient, exact inference. This family of graphical models captures logic-like dependencies with convex hinge-loss potentials. I will survey applications of PSL to diverse problems ranging from information extraction to computational social science. Our recent results show that by building on state-of-the-art optimization methods in a distributed implementation, we can solve large-scale problems with millions of random variables orders of magnitude faster than existing approaches.

Bio
In 1995, Lise Getoor decided to return to school to get her PhD in Computer Science at Stanford University. She received a National Physical Sciences Consortium fellowship, which in addition to supporting her for six years, supported a summer internship at Xerox PARC, where she worked with Markus Fromherz and his group. Daphne Koller was her PhD advisor; in addition, she worked closely with Nir Friedman, and many other members of the DAGS group, including Avi Pfeffer, Mehran Sahami, Ben Taskar, Carlos Guestrin, Uri Lerner,Ron Parr, Eran Segal, Simon Tong. In 2001, Lise Getoor joined the Computer Science Department at the University of Maryland, College Park.

SESSIONS, WITH ABSTRACTS

All ILP sessions are in room 101.

11:15 – 12:45 **Session 1: Life Science Applications**

Short paper **Learning symbolic features for rule induction in computer aided diagnosis**
Sebastijan Dumancic, Antoine Adam and Hendrik Blockeel

Short paper **Is Medical Reasoning Relational?**
Arjen Hommersom

Short paper **Towards machine learning of predictive models from ecological data**
Alireza Tamaddoni-Nezhad, David Bohan, Alan Raybould and Stephen Muggleton

Short paper **Nonmonotonic Learning in Large Biological Networks**
Stefano Bragaglia and Oliver Ray

Short paper **Mining Double Strand Break Model for Cancer Prognostic**
Andrei Doncescu, Chloe Millan and Pierre Siegel

Short paper **Learning meets Sequencing: a Generality Framework for Read-Sets**
Filip Zelezny, Karel Jalovec and Jakub Tolar

14:30 – 16:10 **Session 2: Logic & Time**

Long paper **Logical minimisation of metarules in Meta-Interpretive Learning**
Andrew Cropper and Stephen Muggleton

Short paper **Computing Least Generalization by Anti-combination**
Mikio Yoshida and Chiaki Sakama

Short paper **A First-Order Logic Representation Based Distance Function**
Nirattaya Khamsemanan, Cholwich Nattee and Masayuki Numao

Short paper **Learning Delayed Influence of Dynamical Systems From Interpretation Transition**
Tony Ribeiro, Morgan Magnin and Katsumi Inoue

Short paper **Collaborative on line learning of an action model**
Christophe Rodrigues, Henry Soldano, Gauvain Bourgne and Celine Rouveirol

Long paper **Learning Prime Implicant Conditions From Interpretation Transition**
Tony Ribeiro and Katsumi Inoue

16:30 – 18:05 **Session 3: Applications**

Long paper **Effectively creating weakly labeled training examples via approximate domain knowledge**
Sriraam Natarajan, Jose Manuel Picado Leiva, Tushar Khot, Kristian Kersting, Christopher Re and Jude Shavlik

Short paper **Relational Learning from Ambiguous Examples**
Dominique Bouthinon and Henry Soldano

Short paper **Modeling Semantic Web Services by Learning from Users' Feedback**
Francesca Alessandra Lisi and Floriana Esposito

Short paper **Collective Document Classification over a Large Label Space from Active Learning over a Relational Graph**
Ramakrishna Bairi and Ganesh Ramakrishnan

Short paper **ILP for Mining Linked Open Data**
Gabin Personeni, Simon Daget, Céline Bonnet, Philippe Jonveaux, Marie-Dominique Devignes, Malika Smail-Tabbone and Adrien Coulet

Short paper **Visualizations of First-Order Logic Representation Based Dataset**
Nirattaya Khamsemanan, Cholwich Nattee and Masayuki Numao

09:00 – 10:30 **Session 4: Probabilistic Logics & Graphs**

Short paper **PageRank, ProPPR, and Stochastic Logic Programs**
Dries Van Daele, Angelika Kimmig and Luc De Raedt

Long paper **The Most Probable Explanation for Probabilistic Logic Programs with Annotated Disjunctions**
Dimitar Shterionov, Joris Renkens, Jonas Vlasselaer, Angelika Kimmig, Wannes Meert and Gerda Janssens

Short paper **Goal recognition from incomplete action sequences by probabilistic grammars**
Ryosuke Kojima and Taisuke Sato

Long paper **Statistical Relational Learning for Handwriting Recognition**
Arti Shivram, Tushar Khot, Sriraam Natarajan and Venu Govindaraju

Short paper **On the Complexity of Frequent Subtree Mining in Very Simple Structures**
Pascal Welke, Tamas Horvath and Stefan Wrobel

14:00 – 15:40 **Session 5: Feature Construction & Applications**

Long paper **Reframing on Relational Data**
Chowdhury Farhan Ahmed, Clément Charnay, Nicolas Lachiche and Agnès Braud

Long paper **Construction of Complex Aggregates with Random Restart Hill-Climbing**
Clément Charnay, Nicolas Lachiche and Agnès Braud

Short paper **Consensus-Based Modelling using Distributed Feature Construction**
Haimonti Dutta and Ashwin Srinivasan

Short paper **Propositionalization Online**
Nada Lavrač, Matic Perovšek and Anže Vavpetič

Short paper **Learning Characteristics and Antecedent Behaviours of Lone-Actor Terrorists**
Dalal Alrajeh, Paul Gill and Duangtida Athakravi

Short paper **Declarative Machine Learning for Energy Efficient Compiler Optimizations**
Craig Blackmore, Kerstin Eder and Oliver Ray

16:00 – 17:30 **Session 6: Constraints & Predicate Invention**

Short paper **Learning Constraint Satisfaction Problems: an ILP Perspective**
Luc De Raedt, Anton Dries, Tias Guns and Christian Bessiere

Short paper **Inductive Learning of Answer Set Programs**
Mark Law, Alessandra Russo and Krysia Broda

Short paper **Inductive Learning using Constraint-driven Bias**
Duangtida Athakravi, Dalal Alrajeh, Krysia Broda and Alessandra Russo

Short paper **Bias reformulation for one-shot function induction**
Dianhuan Lin, Eyal Dechter, Joshua Tenenbaum and Stephen Muggleton

Short paper **Incremental Graph-Based Discovery of Relational Concepts**
Ana Tenorio-Gonzalez and Eduardo Morales

Short paper **Can predicate invention in meta-interpretive learning compensate for incomplete background knowledge?**
Andrew Cropper and Stephen Muggleton

11:00 – 12:40 **Joint session with ECML/PKDD**

ECML/PKDD **Evidence-based Clustering for Scalable Inference in Markov Logic**
Deepak Venugopal, Vibhav Gogate

ECML/PKDD **Effective Blending of Two and Three-way Interactions for Modeling Multi-relational Data**
Alberto Garcia-Duran, Antoine Bordes, Nicolas Usunier

ECML/PKDD **Towards Automatic Feature Construction for Supervised Classification**
Marc Boullé

Long paper **Fast Learning of Relational Dependency Networks**
Oliver Schulte, Zhensong Qian, Arthur E. Kirkpatrick, Xiaoqian Yin, Yan Sun

Long paper **Complex aggregates over subsets of elements**
Celine Vens, Sofie Van Gassen, Tom Dhaene, Yvan Saeys

www.ecmlpkdd2014.org