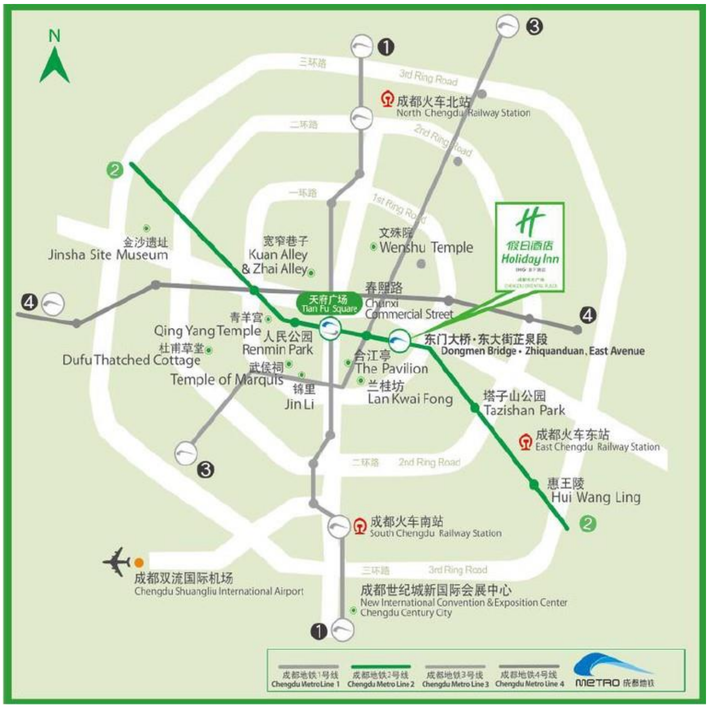
PROGRAM GUIDE

The Third APWeb-WAIM Joint Conference
on Web and Big Data
(APWeb-WAIM 2019)
August 1-3, 2019
Chengdu, China

Conference Hotel

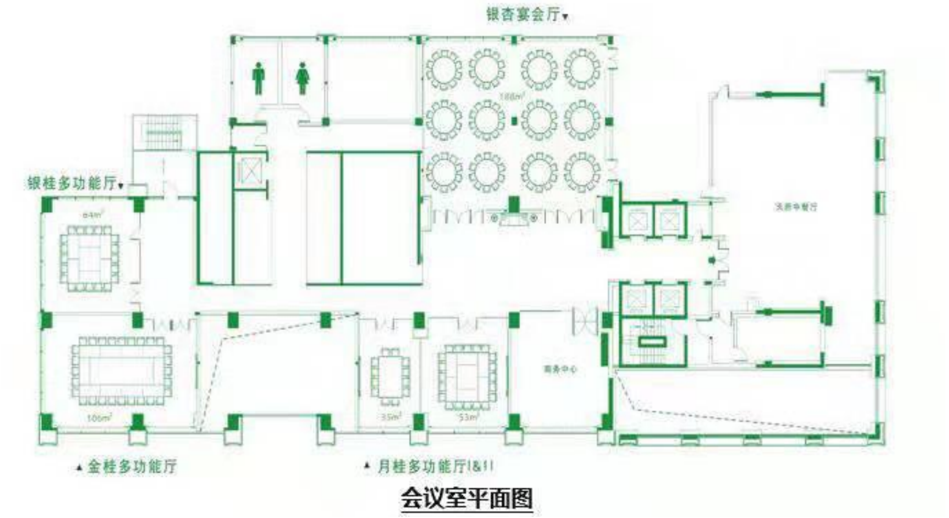
Holiday Inn Chengdu Oriental Plaza

231 Zhiquanduan, East Avenue, JinjiangDistrict, Chengdu, Sichuan Province 610061


N
三环路 3rd Ring Road
二环路 2nd Ring Road
一环路 1st Ring Road
成都火车北站 North Chengdu Railway Station
金沙遗址 Jinsha Site Museum
宽窄巷子 Kuan Alley & Zhai Alley
文殊院 Wenshu Temple
假日酒店 Holiday Inn
天府广场 Tian Fu Square
春熙路 Chunxi Commercial Street
青羊宫 Qing Yang Temple
杜甫草堂 Dufu Thatched Cottage
人民公园 Renmin Park
武侯祠 Temple of Marquis
锦里 Jin Li
合江亭 The Pavilion
兰桂坊 Lan Kwai Fong
东门大桥·东大街芷泉段 Dongmen Bridge · Zhiquanduan, East Avenue
塔子山公园 Tazishan Park
成都火车东站 East Chengdu Railway Station
惠王陵 Hui Wang Ling
成都火车南站 South Chengdu Railway Station
成都双流国际机场 Chengdu Shuangliu International Airport
成都世纪城新国际会展中心 New International Convention & Exposition Center Chengdu Century City
成都地铁1号线 Chengdu Metro Line 1
成都地铁2号线 Chengdu Metro Line 2
成都地铁3号线 Chengdu Metro Line 3
成都地铁4号线 Chengdu Metro Line 4
METRO 成都地铁

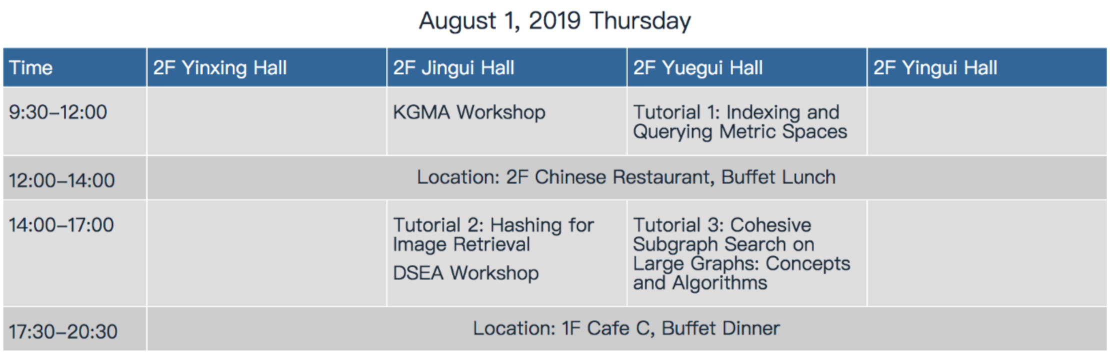
- **Seconds to metro Line 2 Dongmen Bridge Station (Exit C)**
- 5 minutes' drive to Chunxi Lu-Tai Koo Li shopping area
- 5 minutes' walking to Lan Kwai Fong and the JiuYanQiao bar street

Site Map



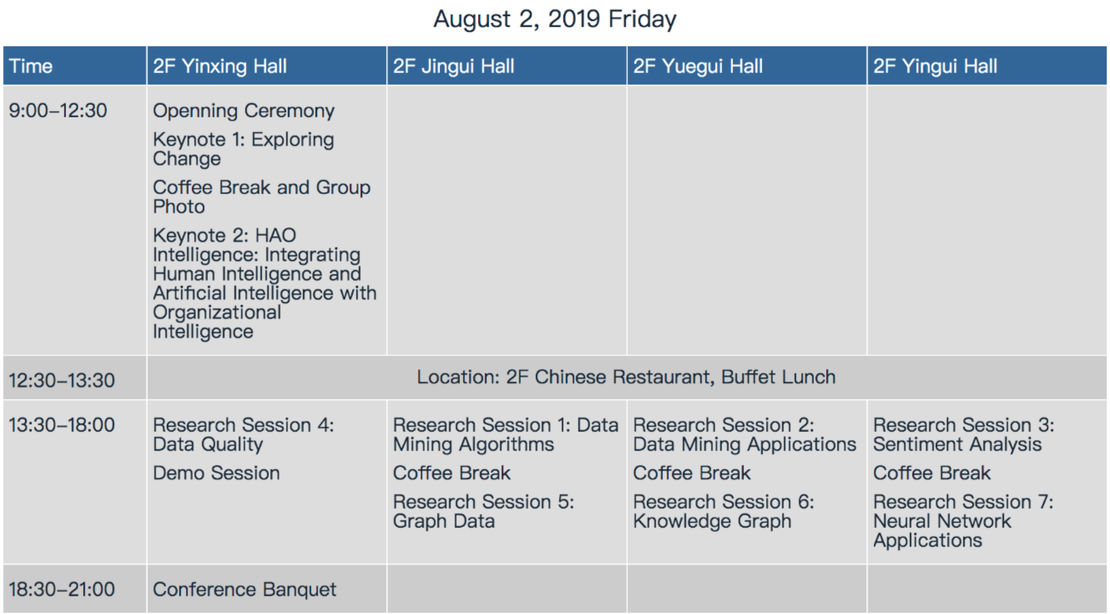
Wi-Fi connection will be provided in the meeting rooms by the hotel.

Schedule at a Glance



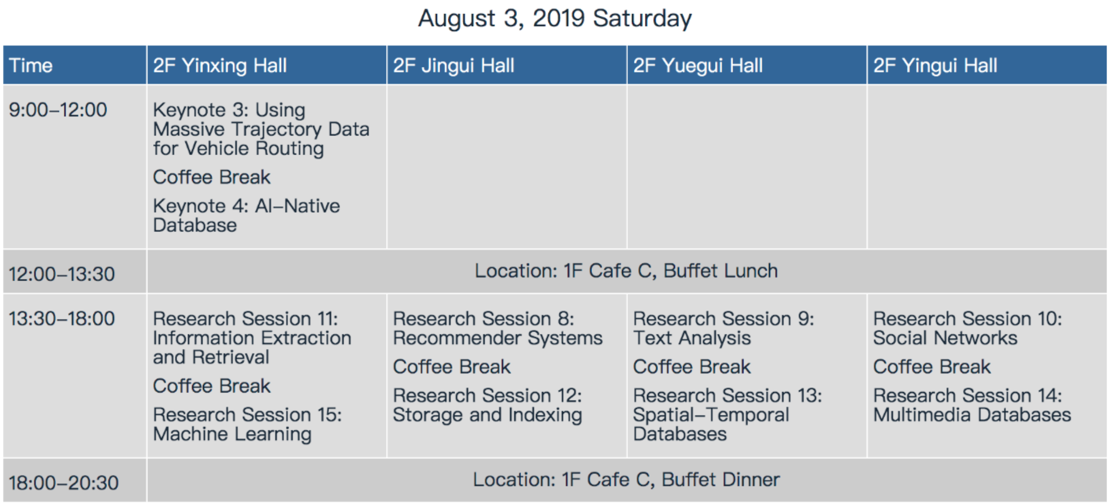
August 1, 2019 Thursday

Time	2F Yinxing Hall	2F Jingui Hall	2F Yuegui Hall	2F Yingui Hall
9:30–12:00		KGMA Workshop	Tutorial 1: Indexing and Querying Metric Spaces	
12:00–14:00	Location: 2F Chinese Restaurant, Buffet Lunch			
14:00–17:00		Tutorial 2: Hashing for Image Retrieval DSEA Workshop	Tutorial 3: Cohesive Subgraph Search on Large Graphs: Concepts and Algorithms	
17:30–20:30	Location: 1F Cafe C, Buffet Dinner			



August 2, 2019 Friday

Time	2F Yinxing Hall	2F Jingui Hall	2F Yuegui Hall	2F Yingui Hall
9:00–12:30	Openning Ceremony Keynote 1: Exploring Change Coffee Break and Group Photo Keynote 2: HAO Intelligence: Integrating Human Intelligence and Artificial Intelligence with Organizational Intelligence			
12:30–13:30	Location: 2F Chinese Restaurant, Buffet Lunch			
13:30–18:00	Research Session 4: Data Quality Demo Session	Research Session 1: Data Mining Algorithms Coffee Break Research Session 5: Graph Data	Research Session 2: Data Mining Applications Coffee Break Research Session 6: Knowledge Graph	Research Session 3: Sentiment Analysis Coffee Break Research Session 7: Neural Network Applications
18:30–21:00	Conference Banquet			



August 3, 2019 Saturday

Time	2F Yinxing Hall	2F Jingui Hall	2F Yuegui Hall	2F Yingui Hall
9:00–12:00	Keynote 3: Using Massive Trajectory Data for Vehicle Routing Coffee Break Keynote 4: AI–Native Database			
12:00–13:30	Location: 1F Cafe C, Buffet Lunch			
13:30–18:00	Research Session 11: Information Extraction and Retrieval Coffee Break Research Session 15: Machine Learning	Research Session 8: Recommender Systems Coffee Break Research Session 12: Storage and Indexing	Research Session 9: Text Analysis Coffee Break Research Session 13: Spatial–Temporal Databases	Research Session 10: Social Networks Coffee Break Research Session 14: Multimedia Databases
18:00–20:30	Location: 1F Cafe C, Buffet Dinner			

Welcome Message from the General Chairs

On behalf of the Organizing Committee, it is our great pleasure to welcome you to The Third Asia Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint Conference on Web and Big Data (APWeb-WAIM 2019) and the beautiful city of Chengdu. Chengdu is the capital of Sichuan Province and it is one of the famous historical and cultural cities in southwestern China. It is also home to the giant panda and the capital of spicy Sichuan cuisine.

APWeb and WAIM are two separate leading international conferences on research, development, and applications of Web technologies and database systems. Previous APWeb conferences were held in Beijing (1998), Hong Kong (1999), Xi’an (2000), Changsha (2001), Xi’an (2003), Hangzhou (2004), Shanghai (2005), Harbin (2006), Huangshan (2007), Shenyang (2008), Suzhou (2009), Busan (2010), Beijing (2011), Kunming (2012), Sydney (2013), Changsha (2014), Guangzhou (2015), and Suzhou (2016). Previous WAIM conferences were held in Shanghai (2000), Xi’an (2001), Beijing (2002), Chengdu (2003), Dalian (2004), Hangzhou (2005), Hong Kong (2006), Huangshan (2007), Zhangjiajie (2008), Suzhou (2009), Jiuzhaigou (2010), Wuhan (2011), Harbin (2012), Beidaihe (2013), Macau (2014), Qingdao (2015), and Nanchang (2016). Starting in 2017, the two conference committees agreed to launch a joint conference. The First APWeb-WAIM conference was held in Bejing (2017) and The Second APWeb-WAIM conference was held in Macau (2018). With the increased focus on big data, the new joint conference is expected to attract more professionals from different industrial and academic communities, not only from the Asia Pacific countries but also from other continents.

APWeb-WAIM 2019 will enable you to enjoy an outstanding program, exchange your ideas with leading researchers in various disciplines, and make new friends in the international science community. Some highlights include four keynote talks on the latest exciting topics of Web and big data, ranging from the fundamental topic of core database systems to the fast-growing artificial intelligence applications; a diverse range of tutorials and workshops; technical sessions with exciting talks and demonstrations, and social events.

We are grateful to the strong support of the Steering Committee of APWeb and WAIM, and we are honored to serve as General Chairs for such a unique joint conference. The conference would not have been possible without the dedication and the hard work of all members of the Organizing Committee. The Program Committee Chairs, Jie Shao

(University of Electronic Science and Technology of China, China), Man Lung Yiu (Hong Kong Polytechnic University, Hong Kong SAR), and Masashi Toyoda (The University of Tokyo, Japan) put tremendous effort into the creation of an exciting program. Many other individuals and organizations contributed to the success of this conference. We would like to acknowledge the efforts of Workshop Chairs (Jingkuan Song and Xiaofeng Zhu), Tutorial Chairs (Shaojie Qiao and Jiajun Liu), Demo Chairs (Wei Lu and Jizhou Luo), Industry Chairs (Jianjun Chen and Jia Zhu), Publication Chairs (Dongxiang Zhang, Wei Wang and Bin Cui) and Publicity Chairs (Lei Duan, Yoshiharu Ishikawa, Jianxin Li, and Weining Qian).

In addition to members of the Organization Committee, many volunteers have contributed to the success of the conference. Volunteers helped in editing this conference booklet, and helped with local arrangements and on-site setups, and many other important tasks. While it is difficult to list all their names here, we would like to take this opportunity to sincerely thank them all.

Last but not least, we would like to extend our most sincere congratulations to all authors and speakers for a job well done. We look forward to welcoming you in person, and we hope that you will enjoy APWeb-WAIM 2019 and the beautiful summer of Chengdu!

General Chairs

Heng Tao Shen
University of Electronic Science and Technology of China, China
Kotagiri Ramamohanarao
University of Melbourne, Australia
Jiliu Zhou
Chengdu University of Information Technology, China

Welcome Message from the Program Committee Chairs

On behalf of the APWeb-WAIM 2019 Program Committee, we are delighted to welcome you to Chengdu! For more than 20 years in the past, Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) have attracted professionals of different communities related to Web and big data who have common interests in interdisciplinary research to share and exchange ideas, experiences, and the underlying techniques and applications, including Web technologies, database systems, information management, software engineering, and big data.

The technical program APWeb-WAIM 2019 features four keynotes by Dr. Divesh Srivastava (AT&T Labs-Research, USA), Dr. Xindong Wu (Mininglamp Technology, China), Prof. Christian S. Jensen (Aalborg University, Denmark), and Prof. Guoliang Li (Tsinghua University, China), as well as three tutorials by Prof. Yunjun Gao (Zhejiang University, China), Dr. Lu Chen (Aalborg University, Denmark) and Mr. Keyu Yang (Zhejiang University, China), Dr. Zi Huang (The University of Queensland, Australia), and Dr. Rong-Hua Li (Beijing Institute of Technology, China) and Mr. Hongchao Qin (Northeastern University, China). We are grateful to these distinguished scientists for their invaluable contributions to the conference program.

Our gratitude goes to Program Committee members and external reviewers whose technical expertise and dedication were not only thorough and crucial for the technical assessment of the selection of papers, but also inspirational in making the whole process even more pleasurable. During the double-blind review process, each paper submitted to APWeb-WAIM 2019 received at least three high quality review reports. Based on the obtained reviews, our Senior Program Committee members provided recommendations for each paper so that the difficult task of making decisions for acceptance could be performed. Finally, out of 180 submissions in total, the conference accepted 42 regular (23.33%), 17 short research papers, and 6 demonstrations. The contributed papers address a wide range of topics, such as big data analytics, data and information quality, data mining and application, graph data and social networks, information extraction and retrieval, knowledge graph, machine learning, recommender systems, storage, indexing and physical database design, and text analysis and mining. Amongst a number of highly rated manuscripts, several candidates for best papers have been shortlisted for awards, where the final selection will be decided during the conference. In particular, we would like to thank Springer for its cash sponsorship of the APWeb-WAIM 2019 Best Paper Award, which will be announced at the conference banquet.

In addition to the main conference program, we would also like to thank Xin Wang (Tianjin University, China) and Yuan-Fang Li (Monash University, Australia) for organizing The Second International Workshop on Knowledge Graph Management and Applications (KGMA 2019), and Lianli Gao (University of Electronic Science and Technology of China, China), Han Su (University of Electronic Science and Technology of China, China), Jiajie Xu (Soochow University, China) and Zhixu Li (Soochow University, China) for organizing The First International Workshop on Data Science for Emerging Applications (DSEA 2019), which are in conjunction with APWEB-WAIM 2019.

We thank the General Chairs Heng Tao Shen, Kotagiri Ramamohanarao, and Jiliu Zhou for their patience and support, and Yanchun Zhang representing the Steering Committee of APWeb and WAIM for the guidance. Many thanks also to all the members of the Organizing Committee for their full support in preparation of the conference, especially with respect to Website, publications, registration and local arrangements, without which the conference would not be possible to be put together.

Finally, the high-quality program would not have been possible without the authors who chose APWeb-WAIM for disseminating their findings. We would like to thank our authors whose valuable and novel contributions are essential for both the continued success of APWeb-WAIM and the advancement of technology for humanity.

Program Committee Chairs

Jie Shao
University of Electronic Science and Technology of China, China
Man Lung Yiu
Hong Kong Polytechnic University, Hong Kong SAR
Masashi Toyoda
The University of Tokyo, Japan

Keynotes

Keynote Speech I: Exploring Change

Time: 9:20-10:40, August 2, 2019
Location: 2F Yinxing Hall
Chair: Heng Tao Shen

Abstract: Data and schema in datasets experience many different kinds of change. Values are inserted, deleted or updated; rows appear and disappear; columns are added or repurposed, and so on. In such a dynamic situation, users might wonder: How many changes have there been in the recent minutes, days or years? What kind of changes were made at which points of time? How dirty is the data? The fact that data changed can hint at different hidden processes: a frequently crowd-updated city name may be controversial; a person whose name has been recently changed may be the target of vandalism; and so on. To interactively explore such changes, we present our DBChEx (Database Change Explorer) prototype system. Using two real-world datasets, IMDB and Wikipedia infoboxes, we illustrate how users can gain valuable insights into data generation processes and data or schema evolution over time by a mix of serendipity and guided investigation using DBChEx. Finally, we identify a range of technical challenges that need to be addressed to fully realize our vision of change exploration.

This is joint work with T. Bleifuß, L. Bornemann, T. Johnson, D. Kalashnikov and F. Naumann.

Divesh Srivastava
Head of Database Research, AT&T Labs-Research
Speaker Bio: Divesh Srivastava is the Head of Database Research at AT&T Labs-Research. He is a Fellow of the Association for Computing Machinery (ACM), the Vice President of the VLDB Endowment, on the ACM Publications Board and an associate editor of the ACM Transactions on Data Science (TDS). He has served as the managing editor of the Proceedings of the VLDB Endowment (PVLDB), as associate editor of the ACM Transactions on Database Systems (TODS), and as associate Editor-in-Chief of the IEEE Transactions on Knowledge and Data Engineering (TKDE). He has presented keynote talks at several international conferences, and his research interests and publications span a variety of topics in data management. He received his Ph.D. from the University of Wisconsin, Madison, USA, and his Bachelor of Technology from the Indian Institute of Technology, Bombay, India.

Keynote Speech II: HAO Intelligence: Integrating Human Intelligence and Artificial Intelligence with Organizational Intelligence

Time: 11:10-12:30, August 2, 2019
Location: 2F Yinxing Hall
Chair: Jie Shao

Abstract: We present a HAO Intelligence framework, which integrates human intelligence (HI), artificial intelligence (AI) and organizational intelligence, for domain-specific industrial applications. HAO Intelligence starts with Bigdata, discovers Big Knowledge, and facilitates human and machine synergism for complex problem solving. This talk discusses Bigdata, Big Knowledge and Big Wisdom, and instantiates HAO Intelligence with a Big Wisdom case study for intelligent catering services.

Xindong Wu
President of Mininglamp Academy of Sciences, Mininglamp Technology. Director of the Key Laboratory of Knowledge Engineering with Big Data (Hefei University of Technology), Ministry of Education, China

Speaker Bio: Dr. Xindong Wu is President of Mininglamp Academy of Sciences, Mininglamp Technology, Beijing, and Director of the Key Laboratory of Knowledge Engineering with Big Data (Hefei University of Technology), Ministry of Education, China. He is a Fellow of the IEEE and the AAAS. He holds a PhD in Artificial Intelligence from the University of Edinburgh and Bachelor's and Master's degrees in Computer Science from the Hefei University of Technology, China. Dr. Wu's research interests include data mining, Bigdata analytics, knowledge engineering, and Web systems.

Dr. Wu is the Steering Committee Chair of the IEEE International Conference on Data Mining (ICDM), the Editor-in-Chief of Knowledge and Information Systems (KAIS, by Springer), the Founding Chair (2002-2006) of the IEEE Computer Society Technical Committee on Intelligent Informatics (TCII), and an Editor-in-Chief of the Springer Book Series on Advanced Information and Knowledge Processing (AI&KP). He was the Editor-in-Chief of the IEEE Transactions on Knowledge and Data Engineering (TKDE, by the IEEE Computer Society) between January 1, 2005 and December 31, 2008, and is currently a Co-Editor-in-Chief of the ACM Transactions on Knowledge Discovery from Data (TKDD, by ACM). He has served as Program Committee Chair/Co-Chair for ICDM '03 (the 3rd IEEE International Conference on Data Mining), KDD-07 (the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining), CIKM 2010 (the 19th ACM Conference on Information and Knowledge Management), ASONAM 2014 (the 2014 IEEE/ACM International

Conference on Advances in Social Network Analysis and Mining), and ICBK 2017 (the 8th IEEE International Conference on Big Knowledge).

Professor Wu is the 2004 ACM SIGKDD Service Award winner and the 2006 IEEE ICDM Outstanding Service Award winner. He received the 2012 IEEE Computer Society Technical Achievement Award "for pioneering contributions to data mining and applications", and the 2014 IEEE ICDM 10-Year Highest-Impact Paper Award. He won the Best Paper Awards from the 2005 and 2011 IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2005 & 2011) and the 2012 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2012), and also the IEEE ICDM-2007 Best Theory/Algorithms Paper Runner-Up Award.

Keynote Speech III: Using Massive Trajectory Data for Vehicle Routing

Time: 9:00-10:20, August 3, 2019
Location: 2F Yinxing Hall
Chair: Man Lung Yiu

Abstract: Massive vehicle trajectory data is becoming available that captures detailed information about vehicular transportation and holds the potential to transform vehicle routing profoundly. The availability of trajectory data renders the traditional routing paradigm obsolete. Instead, new and data-intensive paradigms that thrive on massive trajectory data are called for. The talk will cover three such paradigms, including so-called path-based routing, where costs are associated with paths and not just edges; on-the-fly routing, where all weights are not pre-computed, but are computed as needed during routing; and cost-oblivious routing, where no costs are associated with routes, but where historical trajectories are used directly for routing. These paradigms aim to yield better and more efficient routing.

Christian S. Jensen
Professor, Aalborg University
Speaker Bio: Christian S. Jensen is Professor of Computer Science at Aalborg University, Denmark. His research concerns data management and data-intensive systems, and its focus is on temporal and spatio-temporal analytics. Christian is an ACM and an IEEE Fellow, and he is a member of Academia Europaea, the Royal Danish Academy of Sciences and Letters, and the Danish Academy of Technical Sciences. He has received several national and international awards for his research, most recently the 2019 IEEE TCDE Impact Award. He serves on the board of Villum Fonden, a major funder of technical and natural science research in Denmark; he is President of the steering committee of the Swiss National Research Program on Big Data; and in Germany, he serves on the scientific advisory board the Max Planck Institute for Informatics. He is Editor-in-Chief of ACM Transactions on Database Systems.

Keynote Speech IV: AI-Native Database

Time:10:40-12:00, August 3, 2019
Location: 2F Yinxing Hall
Chair: Shaojie Qiao

Abstract: In big data era, database systems face three challenges. Firstly, the traditional heuristics-based optimization techniques (e.g., cost estimation, join order selection, knob tuning) cannot meet the high-performance requirement for large-scale data, various applications and diversified data. We can design learning-based techniques to make database more intelligent. Secondly, many database applications require to use AI algorithms, e.g., image search in database. We can embed AI algorithms into database, utilize database techniques to accelerate AI algorithms, and provide AI capability inside databases. Thirdly, traditional databases focus on using general hardware (e.g., CPU), but cannot fully utilize new hardware (e.g., ARM, AI chips). Moreover, besides relational model, we can utilize tensor model to accelerate AI operations. Thus, we need to design new techniques to make full use of new hardware.

To address these challenges, we design an AI-native database. On one hand, we integrate AI techniques into databases to provide self-configuring, self-optimizing, self-healing, self-protecting and self-inspecting capabilities for databases. On the other hand, we can enable databases to provide AI capabilities using declarative languages, in order to lower the barrier of using AI.

In this talk, I will introduce the five levels of AI-native databases and provide the open challenges of designing an AI-native database. I will also take automatic database knob tuning, deep reinforcement learning based optimizer, machine-learning based cardinality estimation, automatic index/view advisor as examples to showcase the superiority of AI-native databases.

Guoliang Li
Professor, Tsinghua University
Speaker Bio: Guoliang Li is a tenured full Professor of Department of Computer Science, Tsinghua University, Beijing, China. His research interests include AI-native database, big data analytics and mining, crowdsourced data management, big spatio-temporal data analytics, large-scale data cleaning and integration. He has published more than 100 papers in premier conferences and journals, such as SIGMOD, VLDB, ICDE, SIGKDD, SIGIR, TODS, VLDB Journal, and TKDE. He is a PC co-chair of DASFAA 2019, WAIM 2014, WebDB 2014, and NDBC 2016. He servers as associate editor for IEEE Transactions and Data Engineering, VLDB Journal, ACM Transaction on Data Science, IEEE Data Engineering Bulletin. He has regularly served as the (senior) PC members of many premier conferences, such as SIGMOD, VLDB, KDD, ICDE, WWW, IJCAI, and AAAI. His papers have been cited

more than 6000 times. He got several best paper awards in top conferences, such as CIKM 2017 best paper award, ICDE 2018 best paper candidate, KDD 2018 best paper candidate, DASFAA 2014 best paper runner-up, APWeb 2014 best paper award, etc. He received VLDB Early Research Contribution Award 2017, IEEE TCDE Early Career Award 2014, The National Youth Talent Support Program 2017, ChangJiang Young Scholar 2016, NSFC Excellent Young Scholars Award 2014, CCF Young Scientist 2014.

Tutorials

Tutorial I: Indexing and Querying Metric Spaces

Time: 9:30-12:00, August 1, 2019
Location: 2F Yuegui Hall
Chair: Han Su

Abstract: With the rapid advances in Internet, wireless, and other techniques, there is an exploration of big data with three typical characteristics, i.e., volume, velocity, and variety. Volume denotes that the amount of data is extremely large; velocity represents that the speed of data input and output is extremely high; and variety indicates that the range of data types and sources is extremely wide. A lot of studies have been done on volume and velocity, but not as much has been reported on variety. To handle the variety of data, metric space can be used. Metric space is a general model that can represent any type of data as long as its distance metric satisfies the triangle-inequality. Thus, based on metric space model, we can develop a unified solution to process all various data types. In this tutorial, the speakers systematically present indexing and query processing technologies in metric spaces, including metric index structures, metric query processing, and metric space mining.

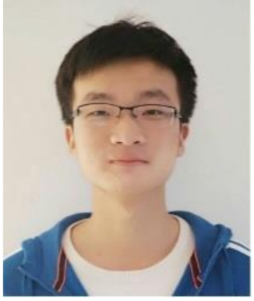
Yunjun Gao

Speaker Bio: Yunjun Gao is now a full professor at the College of Computer Science, Zhejiang University (ZJU), Hangzhou, China. He received the Ph.D. degree in computer science from ZJU in 2008. Prior to joining ZJU in 2010, he was a research assistant or postdoctoral research fellow (scientist) or visiting professor/scholar in the City University of Hong Kong (CityU, China), Singapore Management University (SMU, Singapore), Simon Fraser University (SFU, Canada), and Nanyang Technological University (NTU, Singapore), respectively. His primary research areas are Database, Big Data Management and Analytics, and AI Interaction with DB Technology. He has published more than 130 papers on several premium/leading journals including TODS, VLDBJ, TKDE, TOIS, TMC, TFS, TITS, and DKE, and various prestigious international conferences such as SIGMOD, VLDB, ICDE, SIGIR, AAAI, EDBT, and DASFAA. He is a senior member of the CCF; a member of the ACM and the IEEE; a/an (young) editorial board member or associate editor of JCST, DAPD, and FCS; and a guest editor of WWWJ, IJDSN, and DSE. He is/was a referee/reviewer of several top/important journals such as TODS, VLDBJ, TKDE, TMC, TKDD, Information Sciences, GeoInformatica, etc.; and he is/has serving/served as an organization committee (e.g., PC co-chairs, workshop co-chairs, publication chair, publicity co-chair, etc.) or a program committee member for various conferences such as SIGMOD, VLDB, ICDE, CIKM, SIGSPATIAL GIS, DASFAA, etc. He won the Best Paper Award of APWeb-WAIM 2018, 2017 CCF Outstanding Doctoral Dissertation

Award (Supervisor), the First Prize of the Ministry of Education Science and Technology Progress Award (2016), the Winner of National Outstanding Young Scientist Fund Project (2015), the Nomination of the Best Paper Award of SIGMOD 2015, One of the Best Papers of ICDE 2015, and The First Prize of Zhejiang Province Science and Technology Award First (2011).

Lu Chen

Speaker Bio: Lu Chen is an Assistant Professor in Aalborg University, Denmark. She received the Ph.D. degree in computer science from Zhejiang University, China, in 2016, and then worked as a research fellow in Nanyang Technological University from Oct. 2016 to Sep. 2017. Her research concerns data management and data-intensive systems, and its focus is on metric data management. She has published more than 30 papers on several top/important database conferences (e.g., SIGMOD, VLDB, ICDE, SIGIR) and journals (e.g., VLDBJ, TKDE, Information Sciences). Her paper was selected as one of best papers in ICDE 2015, her thesis is selected as one of the excellent PHD theses by CCF, and her paper won APWeb-WAIM 2018 best paper award. She was also a publication chair of WISE 2017, a PHD colloquium co-chair of MDM 2019, a publicity co-chair of IEEE ICBK 2019, and a guest editor of WWWJ and DSE.



Keyu Yang

Speaker Bio: Keyu Yang received the BS degree in computer science from Zhejiang University of Technology, China, in 2016. He is currently working toward the Ph.D. degree in the College of Computer Science, Zhejiang University, Hangzhou, China. His research interests include metric data management and machine learning interaction with data management technologies.

Tutorial II: Hashing for Image Retrieval

Time: 14:00-15:00, August 1, 2019
Location: 2F Jingui Hall
Chair: Yang Yang

Abstract: Hashing has shown its efficiency and effectiveness in facilitating large-scale multimedia applications. In this tutorial, we will introduce the motivation and advantages of applying hashing in the task of image retrieval. In the first part, a number of classic hashing methods will be discussed, including PCA Hashing, Spectral Hashing, Supervised Hashing with Kernels, and Supervised Discrete Hashing. In this second part, we will present our recent work of zero-shot hashing, robust hashing, and scalable hashing, which are designed for different retrieval scenarios.

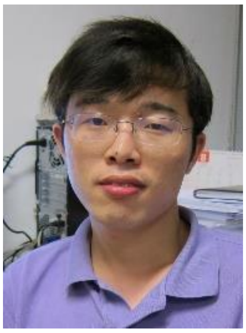
Zi Huang

Speaker Bio: Zi Huang is an Associate Professor (Reader) and ARC Future Fellow in School of ITEE, The University of Queensland. She received her BSc degree from Department of Computer Science, Tsinghua University, China, and her PhD in Computer Science from School of ITEE, The University of Queensland in 2001 and 2007 respectively. Dr. Huang's research interests mainly include multimedia indexing and search, social data analysis and knowledge discovery. She has published 100+ papers in prestigious venues, and is currently an Associate Editor of The VLDB Journal. Dr. Huang has received 2016 Chris Wallace Award from Computing Research and Education (CORE) Australasia for a notable breakthrough or a contribution of particular significance in Computer Science, and Women in Technology (WiT) Infotech Research Award 2014, Queensland. She was also a recipient of the Excellence in Higher Degree by Research Supervision Award, University of Queensland, 2018.

Tutorial III: Cohesive Subgraph Search on Large Graphs: Concepts and Algorithms

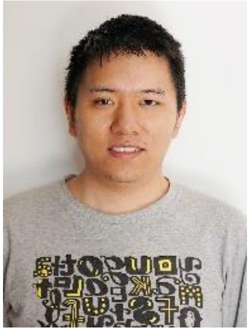
Time: 14:00-17:00, August 1, 2019
Location: 2F Yuegui Hall
Chair: Peng Peng

Abstract: The problem of finding cohesive subgraphs from a large graph is a fundamental problem in graph data mining which has attracted much attention in recent years due to a large number of practical applications. In this tutorial, I will first introduce two widely-used cohesive subgraph models: k-core and k-truss. Then, I will present a peeling algorithm and an h-index iteration algorithm to efficiently compute the k-core and k-truss decomposition on large graphs. Finally, I will introduce some generalized k-core concepts and algorithms on attributed and temporal graphs.



Rong-Hua Li

Speaker Bio: Rong-Hua Li received the Ph.D. degree from the Chinese University of Hong Kong in 2013. He is currently an associate Professor at Beijing Institute of Technology (BIT), Beijing, China. Before joining BIT in 2018, he was an assistant professor at Shenzhen University. His research interests include graph data management and mining, social network analysis, graph computation systems, and graph-based machine learning.



Hongchao Qin

Speaker Bio: Hongchao Qin is currently a Ph.D. Candidate in Northeastern University, China. He received the B.S. degree in mathematics and M.E. degree in computer science from Northeastern University in 2013 and 2015, respectively. His current research interests include social network analysis and data-driven graph mining.

Conference Sessions

Research Session 1: Data Mining Algorithms

Time: 13:30-15:30, August 2, 2019
Location: 2F Jingui Hall
Chair: Lianli Gao

Coupled Semi-supervised Clustering: Exploring Attribute Correlations in Heterogeneous Information Networks

Jianan Zhao, Ding Xiao, Linmei Hu, Chuan Shi
Beijing University of Posts and Telecommunications, China

Abstract. Heterogeneous Information Network (HIN) has been widely adopted in various tasks due to its excellence in modeling complex network data. To handle the additional attributes of nodes in HIN, the Attributed Heterogeneous Information Network (AHIN) was brought forward. Recently, clustering on HIN becomes a hot topic, since it is useful in many applications. Although existing semi-supervised clustering methods in HIN have achieved performance improvements to some extent, these models seldom consider the correlations among attributes which typically exist in real applications. To tackle this issue, we propose a novel model SCAN for semi-supervised clustering in AHIN. Our model captures the coupling relations between mixed types of node attributes and therefore obtains better attribute similarity. Moreover, we propose a flexible constraint method to leverage supervised information and network information for flexible adaption of different datasets and clustering objectives. Extensive experiments have shown that our model outperforms state-of-the-art algorithms.

Exploring Nonnegative and Low-Rank Correlation for Noise-Resistant Spectral Clustering

Zheng Wang[1], Cai Na[2], Zeyu Ma[1], Si Chen[3], Lingyun Song[4], Yang Yang[1]
[1]University of Electronic Science and Technology of China, China, [2]Beijing Institute of Computer Technology and Application, China, [3]China Electronics Technology Group Corporation, China, [4]Xi'an Jiaotong University, China

Abstract. Clustering has been extensively explored in pattern recognition and data mining in order to facilitate various applications. Due to the presence of data noise, traditional clustering approaches may become vulnerable and unreliable, thereby degrading clustering performance. In this paper, we propose a robust spectral clustering approach, termed Non-negative Low-rank Self-reconstruction (NLS), which simultaneously a) explores the nonnegative low-rank properties of data correlation as well as b) adaptively models the structural sparsity of data noise. Specifically, in order to discover the intrinsic correlation among data, we devise a self-reconstruction approach to jointly consider the nonnegativity and low-rank property of data correlation matrix. Meanwhile, we propose to model data noise via a structural norm, i.e., ℓp,2-norm, which not only naturally conforms to genuine patterns of data noise in real-world situations, but also provides more adaptivity and flexibility to different noise levels. Extensive experiments on various real-world datasets illustrate the advantage of the proposed robust spectral clustering approach compared to existing clustering methods.

PrivBUD-Wise: Differentially Private Frequent Itemsets Mining in High-dimensional Databases

Jingxin Xu[1], Kai Han[1], Pingping Song[2], Chaoting Xu[1], Fei Gui[1]

[1]University of Science and Technology of China, China, [2]Anhui University, China

Abstract. In this paper, we study the problem of mining frequent itemsets in high-dimensional databases with differential privacy, and propose a novel algorithm, PrivBUD-Wise, which achieves high result utility as well as a high privacy level. Instead of limiting the cardinality of transactions by truncating or splitting approaches, which causes extra information loss and result in unsatisfactory performance in utility, PrivBUDWise doesn't make any preprocessing on original database and guarantees high result utility by reducing extra privacy budget consumption on irrelevant itemsets as much as possible. To achieve that, we first propose a Report Noisy mechanism with optional number of reported itemsets: SRNM, and what is more important is that we give a strict proof for SRNM in the appendix. Moreover, PrivBUD-Wise first proposes a biased privacy budget allocation strategy and no assumption or estimation on the maximal cardinality needs to be made. The good performance in utility and efficiency of PrivBUD-Wise is shown by experiments on three real-world datasets.

Who is the abnormal user: Anomaly Detection Framework based on the Graph Convolutional Networks

Zetao Zheng[1], Jia Zhu[1], Yong Tang[1], Jiabing Du[2]

[1]South China Normal University, China, [2]Guangdong Grid Co, China

Abstract. Anomaly detection is the identification of items, events or observations which do not conform to an expected pattern in a dataset. It is applicable in a variety of domains, such as intrusion detection, fault detection, medical and public health anomaly monitoring. Existing model usually detects the anomaly according to the data's feature. However, two drawbacks exist if the model only detects anomaly by using the feature. On the one hand, model could not make use of the relationship between users, which contains a large amount of potential information that can strengthen the prediction ability of the model. On the other hand, existing model could not adjust their learning ability automatically with the increasing of the data. To address the issues referred above, we focus on proposing an anomaly detection system based on the Graph Convolutional Networks (GCN). The framework consists of four mechanisms. It can detect the anomalies by using the user features as well as the relationship between users. Experiment result shows that our framework has outstanding performance compared with other state-of-the-art detection models.

Research Session 2: Data Mining Applications

Time: 13:30-15:30, August 2, 2019

Location: 2F Yuegui Hall
Chair: Xingquan Zhu

A Survival Certification Model Based on Active Learning over Medical Insurance Data

Yongjian Ren[1], Kun Zhang[1,2], Yuliang Shi[1,2]
[1]Shandong University, China, [2]Dareway Software Co., Ltd., China

Abstract. In China, Survival Certification (SC) is a work carried out for the implementation of Social Insurance (SI) policies, mainly for retirees. If a retiree is dead but his family has not notified the SI institution, then the SI institution will continue to issue pensions to the retiree. This will lead to the loss of pensions. The purpose of SC is to block the “black hole” of pension loss. However, currently, SC work mainly relies on manual services, which leads to two problems. First, due to the large number of retirees, the implementation of SC usually occupies a large amount of manpower. Secondly, at present, SC work requires all retirees to cooperate with the work of local SI institutions, while some retirees have problems with inconvenient movement or distant distances. These phenomena will lead to an increase of social costs and a waste of social resources. Thus, in this paper, a SC model based on active learning is proposed, which helps staff to narrow the scope of attention. First, we extract features from medical insurance data and analyze their effectiveness. Then, we study the effects of kinds of feature selection functions and classifiers on the SC model. The experimental results show that the model can effectively predict death and can greatly reduce the range of high-risk populations.

Predictive Role Discovery of Research Teams Using Ordinal Factorization Machines

Tong Liu[1], Weijian Ni[1], Qingtian Zeng[1], Nengfu Xie[2]
[1]Shandong University of Science and Technology, China, [2]Chinese Academy of Agricultural Sciences, China

Abstract. In this paper, we address the problem of research role discovery, especially for large research institutes where similar yet separated teams co-exist. The roles that researchers play in a research team, i.e., principal investigator, sub-investigator and research staff, typically exhibit an ordinal relationship. In order to better incorporate the ordinal relationship into a role discovery model, we approach research role discovery as an ordinal regression problem. In the proposed approach, we represent a research team as a heterogeneous teamwork network and propose OrdinalFM, short for Ordinal Factorization Machines, to learn the role prediction function. OrdinalFM extends the traditional Factorization Machines (FM) in an effort to handle the ordinal relationship among learning targets. Experiments with a real-world research team dataset verify the advantages of OrdinalFM over state-of-the-art ordinal regression methods.

Deep Learning for Online Display Advertising User Clicks and Interests Prediction

Zhabiz Gharibshah[1], Xingquan Zhu[1], Arthur Hainline[2], Michael Conway[2]
[1]Florida Atlantic University, USA, [2]Bidtellect Inc., USA

Abstract. In this paper, we propose a deep learning based framework for user interest modeling and click prediction. Our goal is to accurately predict (1) the probability that a user clicks on an ad, and (2) the probability that a user clicks a specify type of campaign ad. To achieve the goal, we collect page information displayed to users as a temporal sequence, and use long-term-short-term memory (LSTM) network to learn latent features representing user interests. Experiments and comparisons on real-world data shows that, compared to existing static set based approaches, considering sequences and temporal variance of user requests results in an improvement in performance ad click prediction and campaign specific ad click prediction.

Data Driven Charging Station Placement

Yudi Guo, Junjie Yao, Jiaxiang Huang, Yijun Chen
East China Normal University, China

Abstract. With the rapid increasing availability of EV (electric vehicle) users, the demand for charging stations has also become vast. In the meanwhile, where to place the stations and what factors have major influence, remains unclear. These problems are bothering when EV companies tries to decide the locations for charging stations. Therefore, we tried to find an effective and interpretable approach to place them in more efficient locations. In common sense, a better location to place a station should relatively has a higher usage rate. Intuitively, we decided to predict usage rates of the candidate locations and tried to explain the result in the meantime, i.e. to find out how much important each feature is or what kind of influence they have. In this paper, we implement 2 models for the usage rate prediction. We also conducted experiments on real datasets, which contains the real charging records of anyo charging company in Shanghai. Further analysis is conducted as well for interpretation of the experiment result, including feature importance.

Research Session 3: Sentiment Analysis

Time: 13:30-15:30, August 2, 2019
Location: 2F Yingui Hall
Chair: Lei Duan

Improved Review Sentiment Analysis with A Syntax-aware Encoder

Jiangfeng Zeng[1], Ming Yang[2], Ke Zhou[1], Xiao Ma[3], Yangtao Wang[1], Xiaodong Xu[1], Zhili Xiao[4]
[1]Huazhong University of Science and Technology, China, [2]Wuhan cciisoft Co Ltd, China, [3]Zhongnan University of Economics and Law, China

Abstract. Review sentiment analysis has drawn a lot of active research interest because of the explosive growth in the amount of available reviews in our day-to-day activities. The current review sentiment classification work often models each sentence as a sequence of words, thus simply training sequence-structured recurrent neural networks

(RNNs) end-to-end and optimizing via stochastic gradient descent (SGD). However, such sequence-structured architectures overlook the syntactic hierarchy among the words in a sentence. As a result, they fail to capture the syntactic properties that would naturally combine words to phrases. In this paper, we propose to model each sentence of a review with an attention-based dependency tree-LSTM, where a sentence embedding is obtained relying on the dependency tree of the sentence as well as the attention mechanism in the tree structure. Then, we feed all the sentence representations into a sequence-structured long short-term memory network (LSTM) and exploit attention mechanism to generate the review embedding for final sentiment classification. We evaluate our attention-based tree-LSTM model on three public datasets, and experimental results turn out that it outperforms the state-of-the-art baselines.

PowerMonitor: Aspect Mining and Sentiment Analysis on Online Reviews

Zhibin Zhao, Lan Yao, Siyuan Wang, Ge Yu
Northeastern University, China

Abstract. Customer reviews on a product regard multi-aspect with emotional tendencies. Aspects in a review show what properties customers concern about and sentiment towards an aspect reveals how a customer evaluates it. The aspect mining and sentiment analysis provides a lot of valuable references and market feedback information to online commercial platforms. Due to the unpredictability of aspects appearing in a review, the method proposed in this paper is supposed to be dynamic and intelligent and to define the sentiment related to an aspect negative or positive polarity in semantic analysis. Based on the improved aspect dictionary and sentiment dictionary, this paper presents a framework for aspect mining and sentiment analysis for online customer reviews PowerMonitor. The experimental results show that the framework performs well in aspect extraction and aspect emotion judgment. We evaluate the model using small, widely used sentiment and subjectivity corpora from JD.com and find it out-performs several previously introduced methods for sentiment classification. We also introduce future works to serve as a reference for efforts in this area.

Transformer and Multi-scale Convolution for Target-oriented Sentiment Analysis

Yinxu Pan[1], Binheng Song[1], Ningqi Luo[1], Xiaojun Chen[2], Hengbin Cui[2]
[1]Tsinghua University, China, [2]Ant Financial Services Group, China

Abstract. Target-oriented sentiment analysis aims to extract the sentiment polarity of a specific target in a sentence. In this paper, we propose a model based on transformers and multi-scale convolutions. The transformer which is based solely on attention mechanisms generalizes well in many natural language processing tasks. Convolution layers with multiple filters can efficiently extract n-gram features at many granularities on each receptive field. We conduct extensive experiments on three datasets: SemEval ABSA challenge Restaurant and Laptop dataset, Twitter dataset. Our framework achieves state-of-the-art results, including improving the accuracy of Restaurant dataset to 84.20% (5.81% absolute improvement), improving the accuracy of the Laptop dataset to 78.21% (4.23% absolute improvement), and improving the accuracy of the

Twitter dataset to 72.98% (0.87 % absolute improvement).

MBMN: Multivariate Bernoulli Mixture Network for News Emotion Analysis

Xue Zhao, Ying Zhang, Wenya Guo, Xiaojie Yuan
Nankai University, China

Abstract. In the text classification task, besides the text features, labels are also crucial to the final classification performance, which have not been considered in most existing works. In the context of emotions, labels are correlated and some of them can coexist. Such label features and label dependencies as auxiliary text information can be helpful for text classification. In this paper, we propose a Multivariate Bernoulli Mixture Network (MBMN) to learn a text representation as well as a label representation. Specifically, it generates the text representation with a simple convolutional neural network, and learns a mixture of multivariate Bernoulli distribution which can model the label distribution as well as label dependencies. The labels can be sampled from the distribution and further used to generate a label representation. With both text representation and label representation, MBMN can achieve better classification performance. Experiments show the effectiveness of the proposed method against competitive alternatives on two public datasets.

Research Session 4: Data Quality

Time: 13:30-15:00, August 2, 2019
Location: 2F Yinxing Hall
Chair: Xiang Zhao

DeepAM: Deep Semantic Address Representation for Address Matching

Shuangli Shan[1,2], Zhixu Li[1,3], Yang Qiang[4], An Liu[1], Jiajie Xu[1], Zhigang Chen[3]
[1]Soochow University, China, [2]Neusoft Corporation, China, [3]IFLYTEK, China, [4]King Abdullah University of Science and Technology, SA

Abstract. Address matching is a crucial task in various location-based businesses like take-out services and express delivery, which aims at identifying addresses referring to the same location in address databases. It is a challenging one due to various possible ways to express the address of a location, especially in Chinese. Traditional address matching approaches relying on string similarities and learning matching rules to identify addresses referring to the same location, could hardly solve the cases with redundant, incomplete or unusual expression of addresses. In this paper, we propose to map every address into a fixed-size vector in the same vector space using state-of-the-art deep sentence representation techniques and then measure the semantic similarity between addresses in this vector space. The attention mechanism is also applied to the model to highlight important features of addresses in their semantic representations. Last but not least, we novelly propose to get rich contexts for addresses from the web through web search engines, which could strongly enrich the semantic meaning of addresses that could be learned. Our empirical study conducted on two real-world address datasets demonstrates that our approach greatly improves both precision (up to

5%) and recall (up to 8%) of the state-of-the-art existing methods.

Drawing CoCo Core-sets from Incomplete Relational Data

Yongnan Liu[1,2], Jianzhong Li[1]
[1]Harbin Institute of Technology, China, [2]Heilongjiang University, China

Abstract. Incompleteness is a pervasive issue and brings challenges to answer queries with high-quality tuples. Since not all missing values can be repaired by complete values, it is crucial to provide completeness of a query answer for further decisions. To estimate such completeness results fast and objectively, CoCo core-sets are proposed in this paper. A CoCo core-set is a subset of an incomplete relational dataset, which contains tuples providing enough complete values on attributes of interest and whose ratio of complete values is close to that of the entire dataset. Based on CoCo core-sets reliable mechanisms can be designed to estimate query completeness on incomplete datasets. This paper investigates the problem of drawing CoCo core-sets on incomplete relational data. To the best of our knowledge, there is no such a proposal in the past. (1) We formalize the problem of drawing CoCo core-sets, and prove that the problem is NP-Complete. (2) An efficient approximate algorithm to draw an approximate CoCo core-set is proposed, where uniform sampling technique is employed to efficiently select tuples for coverage and completeness. (3) Analysis of the proposed approximate algorithm shows both coverage of attributes of interest and the relative error of ratio of complete attribute values between drawn tuples and the entire data can be within a given relative error bound. (4) Experiments on both real-world and synthetic datasets demonstrate that the algorithm can effectively and efficiently draw tuples preserving properties of entire datasets for query completeness estimation, and have a well scalability.

Reducing Wrong Labels for Distant Supervision Relation Extraction with Selective Capsule Network

Zihao Wang, Yong Zhang, Chunxiao Xing
Tsinghua University, China

Abstract. Distant Supervision is a common technique for relation extraction from large amounts of free texts, but introduces wrong labeled sentences at the same time. Existing deep learning approaches mainly rely on CNN-based models. However, they fail to capture spatial patterns due to the inherent drawback of pooling operations and thus lead to suboptimal performance. In this paper, we propose a novel framework based on Selective Capsule Network for distant supervision relation extraction. Compared with traditional CNN-based models, the involvement of capsule layers in the sentence encoder makes it more powerful in encoding spatial patterns, which is very important in determining the relation expressed in a sentence. To address the wrong labeling problem, we introduce a high-dimensional selection mechanism over multiple instances. It is one generalization of traditional selective attention mechanism and can be seamlessly integrated with the capsule network based encoder. Experimental results on a widely used dataset (NYT) show that our model significantly outperform all the state-of-the-art methods.

Research Session 5: Graph Data

Time: 16:00-18:00, August 2, 2019
Location: 2F Jingui Hall
Chair: Bolong Zheng

Distributed Landmark Selection for Lower Bound Estimation of Distances in Large Graphs

Mingdao Li, Peng Peng, Yang Xu, Hao Xia, Zheng Qin
Hunan University, China

Abstract. Given two vertices in a graph, computing their distance is a fundamental operation over graphs. However, classical exact methods for this problem often cannot scale up to the rapidly evolving graphs in recent applications. Many approximate methods have been proposed, including some landmark-based methods that have been shown to have good scalability and estimate the upper bound of the distance in acceptable accuracy. In this paper, we propose a new landmark-based framework based a new measure called coverage to more accurately estimate the lower bound of the distance. Although we can prove that selecting the optimal set of landmarks is NP-hard, we propose a heuristic algorithm that can guarantee the approximation ratio. Furthermore, we implement our method through the distributed graph processing systems while considering the characteristic of the distributed graph processing systems. Experiments on large real graphs confirm the superiority of our methods.

Cider: Highly Efficient Processing of Densely Overlapped Communities in Big Graphs

Yadi Chen, Wen Bai, Runyuan Chen, Di Wu, Guoqiao Ye, Zhichuan Huang
Sun Yat-sen University, China

Abstract. As one of the most fundamental operations in graph analytics, community detection is to find groups of vertices that are more densely connected internally than with the rest of the graph. However, the detection of densely overlapped communities in big graphs is extremely challenging due to high time complexity. In this paper, we propose an effective and efficient graph algorithm called Cider to detect densely overlapped communities in big graphs. The intuition behind our algorithm is to exploit inherent properties of densely overlapped communities, and expand the community by minimizing its conductance. To make Cider more efficient, we extend the algorithm to expand the community more quickly by merging vertices in batches. We explicitly derive the time complexity of our algorithm and conclude that it can be implemented in near-linear time. Besides, we also implement a parallelized version of Cider to further improve its performance. Experimental results on real datasets show that our algorithms outperform existing approaches in terms of Flake Out Degree Fraction (FODF) and F1 Score.

Iterative Hypergraph Computation based on Hyperedge-Connected Graphs

Kaiqiang Yu, Yu Gu, Shuo Yao, Zhen Song, Ge Yu

Northeastern University, China

Abstract. A hypergraph allows a hyperedge to connect arbitrary number of vertices, which can be used to capture the complex and high-order relationships. By analyzing the iterative processing on bipartite graphs, a method of converting the original hypergraph into a hyperedgeconnected graph and corresponding iterative processing method are proposed. Then, the iterative processing solution based on hyperedge-connected graphs is combined with Push-based and Pull-based message acquisition mechanisms. On top of the distributed graph processing system HybridGraph, a hypergraph iterative processing framework HyraphD is implemented. Finally, extensive experiments are conducted on several real-world datasets and hypergraph learning algorithms. Experimental results confirm the efficiency and the scalability of HyraphD.

How to Reach: Discovering Multi-Resolution Paths on Large Scale Networks

Zhaokun Zhang[1], Ning Yang[1], Philip S. Yu[2]
[1]Sichuan University, China, [2]University of Illinois at Chicago, USA

Abstract. Reachability query is a fundamental problem which has been studied extensively due to its important role in various application domains, such as social networks, communication networks, biological networks, etc. However, few existing works pay attention to the problem of how two vertices are connected. In this paper, we investigate the problem of discovering paths of multiple resolutions connecting two vertices in large scale networks. We propose a new structure, called Muti-Resolution Path (MRP), to describe how two vertices are connected at different resolution levels. To facilitate the building of MRPs on a network of large scale, we propose a new search structure, called Hierarchical Compressed Network (HCN), which can represent a network at multiple resolution levels and can be built offline. At last, extensive experiments conducted on real-world datasets verify the effectiveness and efficiency of the proposed approach.

Research Session 6: Knowledge Graph

Time: 16:00-18:00, August 2, 2019
Location: 2F Yuegui Hall
Chair: Jia Zhu

Leveraging Domain Context for Question Answering over Knowledge Graph

Peihao Tong, Junjie Yao, Linzi He, Liang Xu
East China Normal University, China

Abstract. This paper focuses on the problem of question answering over knowledge graph (KG-QA). With the increasing availability of different knowledge graphs in a variety of domains, KG-QA becomes a prevalent information interaction approach. Current KG-QA methods usually resort to semantic parsing, retrieval or neural matching based models. However, current methods generally ignore the rich domain context, i.e., category and surrounding descriptions within the knowledge graphs.

Experiments shows that they can not well tackle the complex questions and information needs. In this work, we propose a new KG-QA approach, leveraging the domain context. The new method designs a neural cross-attention QA framework. We incorporate the new approach with question and answer domain contexts. Specifically, for questions, we enrich them with users' access log, and for the answers, we equip them with meta-paths within the target knowledge graph. Experimental study on real datasets verifies its improvement. The new approach is especially beneficial for domain knowledge graphs.

Leveraging Lexical Semantic Information for Learning Concept-Based Multiple Embedding Representations for Knowledge Graph Completion

Yashen Wan[1], Yifeng Liu[1], Huanhuan Zhang[1], Haiyong Xie[1,2]

[1]China Academy of Electronics and Information Technology, China, [2]University of Science and Technology of China, China

Abstract. Knowledge graphs (KGs) are important resources for a variety of natural language processing tasks but suffer from incompleteness. To address this challenge, a number of knowledge graph completion (KGC) methods have been developed using low-dimensional graph embeddings. Most existing methods focus on the structured information of triples in encyclopaedia KG and maximize the likelihood of them. However, they neglect semantic information contained in lexical KG. To overcome this drawback, we propose a novel KGC method (named as TransC), that integrates the structured information in encyclopaedia KG and the entity concepts in lexical KG, which describe the categories of entities. Since all entities appearing in the head (or tail) position with the same relation have some common concepts, we introduce a novel semantic similarity to measure the distinction of entity semantics with the concept information. And then TransC utilizes concept-based semantic similarity of the related entities and relations to capture prior distributions of entities and relations. With the concept based prior distributions, TransC generates multiple embedding representations of each entity in different contexts and estimates the posterior probability of entity and relation prediction. Experimental results demonstrate the efficiency of the proposed method on two benchmark datasets.

Efficient Distributed Knowledge Representation Learning for Large Knowledge Graphs

Lele Chai, Xin Wang, Baozhu Liu, Yajun Yang

Tianjin University, China

Abstract. Knowledge Representation Learning (KRL) has been playing an essential role in many AI applications and achieved desirable results for some downstream tasks. However, two main issues of existing KRL embedding techniques have not been well addressed yet. One is that the size of input datasets processed by these embedding models is typically not large enough to accommodate large-scale real-world knowledge graphs; the other issue is that lacking a unified framework to integrate current KRL models to facilitate the realization of embeddings for various applications. We propose DKRL, which is a distributed KRL training framework that can incorporate different

KRL models in the translational category using a unified algorithm template. In DKRL, a set of primitive interface functions is defined to be implemented by various knowledge embedding models to form a unified algorithm template for distributed KRL. The effectiveness and efficiency of our framework have been verified by extensive experiments on both benchmark and real world knowledge graphs, which show that our approach can outperform the existing ones by a large margin.

Coherence and Salience-Based Multi-Document Relationship Mining

Yongpan Sheng, Zenglin Xu

University of Electronic Science and Technology of China, China

Abstract. In today’s interconnected world, there is an endless 24/7 stream of new articles appearing online. Faced with these overwhelming amounts of data, it is often helpful to consider only the key entities and concepts and their relationships. This is challenging, as relevant connections may be spread across a number of disparate articles and sources. In this paper, we propose a unified framework to aid users in quickly discerning salient connections and facts from a set of related documents, and presents the resulting information in a graph-based visualization. Specifically, given a set of relevant documents as input, we firstly extract candidate facts from above sources by exploiting Open Information Extraction (Open IE) approaches. Then, we design a Two-Stage Candidate Triple Filtering (TCTF) approach based on a self-training framework to maintain only coherent facts associated with the specified document topic from the candidates and connect them in the form of an initial graph. We further construct this graph by a heuristic to ensure the final conceptual graph only consist of facts likely to represent meaningful and salient relationships, which users may explore graphically. The experiments on two real-world datasets illustrate that our extraction approach achieves 2.4% higher on the average of F-score over several OpenIE baselines. We also further present an empirical evaluation of the quality of the final generated conceptual graph towards different topics on its coverage rate of topic entities and concepts, confidence score, and the compatibility of involved facts. Experimental results show the effectiveness of our proposed approach.

Research Session 7: Neural Network Applications

Time: 16:00-18:00, August 2, 2019
Location: 2F Yingui Hall
Chair: Jingkuan Song

DeepDial: Passage Completion on Dialogs

Nan Hu, Jianyun Zhou, Xiaojun Wan

Peking University, China

Abstract. Many neural models have been built to carry out reading comprehension tasks. However, these models mainly focus on formal passages like news and book stories. Although human dialog is the most important part of daily life, machine reading comprehension on dialogs (i.e., passage completion on dialogs) has not been sufficiently explored. Existing models show some weaknesses when comprehending dialogs and they are unable to capture global information over a distance and local detailed information at the same time. This paper introduces a neural network model DeepDial that aims at addressing the problems mentioned above. The model explores both word-level and utterance-level information in a dialog, and achieves the state-of-the-art performance on the benchmark dataset constructed from a TV series Friends.

Medical Treatment Migration Prediction in Healthcare via Attention-based Bidirectional GRU

Lin Cheng[1], Yongjian Ren[1], Kun Zhang[1,2], Yuliang Shi[1,2]

[1]Shandong University, China, [2]Dareway Software Co., Ltd, China

Abstract. With the rapid expansion of the number of floating populations in China, a large number of people are gradually migrating to different hospitals to seek medical treatment. How to accurately predict the future medical treatment behaviors of patients has become an important research issue in healthcare. In this paper, an Attention-based Bidirectional Gated Recurrent Unit (AB-GRU) medical treatment migration prediction model is proposed to predict which hospital patients will go to in the future. The model groups patients who are prone to medical treatment migration, and achieves disease prediction and medical treatment migration prediction for each group. In terms of disease prediction, considering the predictive performance problem of a single prediction algorithm, a standard deviation weight recombination method is used to achieve disease prediction. When disease prediction has been completed, considering the impact of medical visit on the future medical behavior, on the basis of bidirectional gated recurrent unit (GRU) framework, we introduce an attention mechanism to determine the strength of hidden state at different moments, which can improve the predictive performance of the model. The experiment demonstrates that the predictive model proposed in this paper is more accurate than the traditional predictive models.

A Novel Approach for Air Quality Inference and Prediction based on DBU-LSTM

Liang Ge, Aoli Zhou, Junling Liu, Hang Li

Chongqing University, China

Abstract. The inference and prediction of fine-grained air quality are two important directions in urban air computing. Solving these two problems can provide useful information for urban environmental governance and residents' health improvement. In this paper, we propose a general approach to solve these two problems with one model, while most other existing works use different models to solve them. Our model is based on deep bidirectional and unidirectional long short-term memory (DBULSTM) neural network, which can capture bidirectional temporal dependencies and spatial correlation from time series data containing spatial information. To infer and predict the air quality of the target region, we use the historical meteorological data of the target region and the historical air quality data of regions which are similar to the target. Urban heterogeneous data such as point of interest (POI) and road network are used to evaluate the similarities between urban regions. We also use a tensor decomposition method to complete the missing historical air quality data onto monitoring stations, which reduces the error of our model. We evaluated our approach on real data sources obtained in Beijing, and the results show its advantages over recent literature.

WRL: A Combined Model for Short-Term Load Forecasting

Yuecan Liu[1], Kun Zhang[1,2], Shuai Zhen[2], Yongming Guan[2], Yuliang Shi[1,2]
[1]Shandong University, China, [2]Dareway Software Co., Ltd, China

Abstract. Load forecasting plays a vital role in economic construction and national security. The accuracy of short-term load forecasting will directly affect the quality of power supply and user experience, and will indirectly affect the stability and safety of the power system operation. In this paper, we present a novel short-term load forecasting model, which combines influencing factors analysis, Wavelet Decomposition feature extraction, Radial Basis Function (RBF) neural networks and Bidirectional Long Short-Term Memory (Bi-LSTM) networks (WRL below). The model uses wavelet decomposition to extract the main features of load data, analyzes its correlation with influencing factors, and then constructs corresponding adjustment factors. The RBF neural networks are used to forecast the feature subsequence related to external factors. Other subsequences are input into Bidirectional LSTM networks to forecast future values. Finally, the forecasting results are obtained by wavelet inverse transform. Experiments show that the proposed short-term load forecasting method is effective and feasible.

Demo Session

Time: 15:00-17:00, August 2, 2019
Location: 2F Yinxing Hall
Chair: Jizhou Luo

FMQO: A Federated RDF System Supporting Multi-Query Optimization

Qi Ge[1], Peng Peng[1], Zhiwei Xu[1], Lei Zou[2], Zheng Qin[1]
[1]Hunan University, China, [2]Peking University, China

Abstract. This demo designs and implements a system called FMQO that can support multiple query optimization in federated RDF systems. Given a set of queries posed simultaneously, we propose a heuristic query rewriting-based approach to share the common computation during evaluation of multiple queries. Furthermore, we propose an efficient method to use the interconnection topology between SPARQL endpoints to filter out irrelevant sources and join intermediate results during multiple query evaluation. The experimental studies over both real federated RDF datasets show that the demo is effective, efficient and scalable.

DataServiceHatch: Generating and Composing Continuous Data Services

Guiling Wang[1,2], Tongtong Cui[1], Xiaojiang Zuo[1], Yao Xu[2], Yanbo Han[1]
[1]North China University of Technology, China, [2]China Electronics Technology Group, China

Abstract. In this paper, we present DataServiceHatch, a Web-based system that semi-

automatically converts relational databases and stream data sources into Web Services and answers continuous queries on both traditional database tables and data streams by composing Web Services, rather than accesses databases or big data stream infrastructures directly. This can help organizations to unify the access entrance of all their data sources with just a few simple configurations and avoid exposing their data directly. DataServiceHatch also provides mechanisms to remove the need for manually writing a complex SQL-like query expression or service composition plan to answer a continuous query on data streams.

A Mobile Phone Data Visualization Tool for People Flow Analysis

Liangjian Chen[1], Siyu Chen[1], Shengnan Guo[1], Yue Yang[2], Jianqiu Xu[1]
[1]Nanjing University of Aeronautics and Astronautics, China, [2]Jiangsu Academy of Architectural Sciences, China

Abstract. Mobile phone data contain the information of each interaction between mobile phones and telecommunication infrastructures. These data provide a wealth of information about urban dynamics and human activities since each mobile phone can be seen as a sensor that senses the geographic position of the subscriber holder in real time. In this paper, we introduce an open-source and web-based data visualization tool for analyzing and displaying people flow information using mobile phone data. The developed tool provides users a user-friendly interface for data visualization. We demonstrate how to install and display this tool by using real mobile phone data.

NativeHelper: A Bilingual Sentence Search & Recommendation Engine for Academic Writing

Weijian Ni[1], Yujian Sun[1], Tong Liu[1], Qingtian Zeng[1], Nengfu Xie[2]
[1]Shandong University of Science and Technology, China, [2]Chinese Academy of Agricultural Sciences, China

Abstract. This demo presents NativeHelpler, a bilingual Chinese-English sentence search engine that aims to provide assistance for non-English academic writers. As opposed to most existing bilingual sentence search engines that rely heavily on parallel corpora, our system is built on monolingual sentence corpus and bilingual dictionaries which are more readily available. The system is implemented based on a large-scale English sentence database and a simple yet practically efficient bilingual language model. A screen cast is available at https://www.youtube.com/watch?v=oNYOYPDeTyM.

PKRS: A Product Knowledge Retrieve System

Taoyi Huang, Yuming Lin, Haibo Tang, You Li, Huibing Zhang
Guilin University of Electronic Technology, China

Abstract. In this demo paper, we present the Product Knowledge Retrieve System (PKRS), which can retrieve the large-scale product knowledge efficiently. The PKRS has three features. Firstly, PKRS can retrieve not only the objective knowledge (e.g. categories) but also the subjective knowledge (e.g. users' opinion). Secondly, a learned mapping dictionary (LMD) is devised to accelerate the query parsing. Thirdly, PKRS adopts optimized join strategy to improve the retrieval effectiveness. For demonstration, we compare the performance of our PKRS with a state-of-the-art knowledge management system. The experimental results show that the PKRS can process the queries on product knowledge more effectively.

DataSESec: Security Monitoring for Data Share and Exchange Platform

Guowei Shen[1,2], Lu Liu[1], Qin Wei[1], Chun Guo[1]
[1]GuiZhou University, China, [2]CETC Big Data Research Institute Co. Ltd., China

Abstract. Data share and exchange platform is an infrastructure of data open and share. How to ensure the security of government data in the exchange and sharing platform is a key problem. To solve this problem, we developed a security monitoring system for data share and exchange platform - DataSESec. A multi-layer graph model is provided to realize multi-source heterogeneous security monitoring metadata organization, data tracking and forensics, and multi-dimensional security monitoring data analysis. The system extracts network traffic data without authorization, which can achieve early security warning. The deployment of the security monitoring system is very flexible, the interface that interacts with the existing platform is very flexible, and the impact on the existing data share and exchange platform is very small.

Research Session 8: Recommender Systems

Time: 13:30-15:30, August 3, 2019
Location: 2F Jingui Hall
Chair: Yong Zhang

Streaming Recommendation Algorithm with User Interest Drift Analysis

Jianzong Chen, Hanlu Li, Qing Xie, Lin Li, Yongjian Liu
Wuhan University of Technology, China

Abstract. Recommender system is an effective way to solve the problem of information overload, and remarkable progress has been achieved along with the research and applications in both academic and industrial communities. However, the scalability of the conventional recommendation algorithms has been challenged by the exponential growth of the resource data size, and the increasing time span of the data also raises new requirements on the time-awareness of the algorithm. Therefore, a dynamic recommendation model monitoring the user interest drift has become an important task for streaming recommender system. In this paper, an incremental matrix factorization model named streamGBMF is proposed which utilizes the genre information as the resource feature. The proposed model can be updated in real-time according to the streaming data. To achieve the online updating, two kinds of forgetting mechanism are embedded to analyze the users' current interest and preference accurately and timely. To evaluate the performance of our proposed model, the experiments are designed on the popular dataset MovieLens, and different algorithms are compared in streaming environment. The results show that our approach can effectively accelerate the model training process, and the recommendation performance can be improved by real-time

user interest drift detection with proposed forgetting mechanisms.

Unified Group Recommendation towards Multiple Criteria

Yi Wu, Ning Yang, Huanrui Luo
Sichuan University, China

Abstract. In online social networks, a growing number of people are willing to share their activities with ones who have common interests. This motivates the research on group recommendation, which focuses on the issue of recommending items to a group of users. The existing methods on addressing the problem of grouping users and making recommendations for the formed groups simultaneously, however, often suffer from two defects. The first one is that they separate group partition and group recommendation, which often reduce the overall group satisfaction. The second one is that they tend to pursue a single objective optimum instead of making a balance between multiple objectives. In this paper, we strive to tackle the key problem of grouping users and making recommendations for the formed groups simultaneously. It is a challenging problem due to the differences between user preferences over items, and how to make a trade-off among their preferences for the recommended items is still the main research point. To address these challenges, we present a Unified Group Recommendation (UGR) model, which intertwines the user grouping and group recommendation in a unified multi-objective optimization process that makes a balance between multiple criteria, including maximizing overall group satisfaction, social relationship density, and overall group fairness. Extensive experiments on two real-world datasets verify the effectiveness of our method.

Latent Path Connected Space Model for Recommendation

Lang Mei[1], Jun He[1], Hongyan Liu[2], Xiaoyong Du[1]
[1] Renmin University of China, China, [2] Tsinghua University, China

Abstract. Matrix Factorization (MF) is a latent factor model, which has been one of the most popular techniques for recommendation systems. Performance of MF-based recommender models degrades as the sparseness of user-item rating data increases. MF-based models map each user and each item into a low dimensional space, where either of them is represented by a point in the space. While a point is a concise and simple representation of a user’s preference or an item’s characteristics, it is hard to learn the precise position of the point, especially when the data is very sparse. In this paper we propose an alternative latent space model, Latent Path Connected Space model (LSpace), to address this issue. In this model, users and items are both represented by path connected space described by different latent dimensions and spatial intersection between user space and item space reflects their matching degree. Extensive evaluations on four real-world datasets show that our approach outperforms the Matrix Factorization model on rating prediction task especially when the rating data is extremely sparse.

A Novel Ensemble Approach for Click-Through Rate Prediction based on Factorization Machines and Gradient Boosting Decision Trees

Xiaochen Wang, Gang Hu, Haoyang Lin, Jiayu Sun
University of Electronic Science and Technology of China, China

Abstract. Click-Through Rate (CTR) prediction is a significant technique in the field of computational advertising, its accuracy directly affects companies profits and user experience. Achieving great ability of generalization by learning complicated feature interactions behind user behaviors is critical in improving CTR for recommender systems. Factorization Machines (FM) is a hot recommender method for efficiently modeling features' second-order interactions. Nevertheless, FM cannot capture the nonlinear and complex modes implied in the real-world data while it models feature in a linear way and just uses the second-order feature interactions. In this paper, we propose a model named GFM, which is an ensemble learning of FM and Gradient Boosting Decision Trees (GBDT) for recommendations. We use FM to model linear features and second-order feature interactions and use GBDT to model the side information for transforming the raw features to cross-combined features. In addition, we import the attention mechanism to calculate users' latent attention on different features. To illustrate the performance of GFM, we conduct experiments on two real-world datasets, including a movie dataset and a music dataset, the results show that our model is effective in providing accurate recommendations.

Research Session 9: Text Analysis

Time: 13:30-15:30, August 3, 2019
Location: 2F Yuegui Hall
Chair: Xujian Zhao

Multi-label Text Classification: Select Distinct Semantic Understanding for Different Labels

Wei Sun, Xiangying Ran, Xiangyang Luo, Yunlai Xu, Chongjun Wang
Nanjing University, China

Abstract. Multi-label classification is a challenging task in natural language processing. Most of existing methods tend to ignore the semantic information of the text. Besides, different parts of the text contribute differently to each label, which is not considered by most of existing methods. In this paper, we propose a novel model for multi-label text classification. This model generates high-level semantic understanding representations with a multi-level dilated convolution. The multi-level dilated convolution effectively reduces dimension and expands the receptive fields without loss of information. Moreover, a hybrid attention mechanism is designed to capture most relevant information of the text based on trainable label embeddings and semantic understanding. Experimental results on the dataset AAPD and RCV1-V2 show that our model has significant advantages over baseline methods.

A New Feature Selection Algorithm based on Category Difference for Text Categorization

Wang Zhang[1], Chanjuan Chen[2], Lei Jiang[1], Xu Bai[1]

[1]Chinese Academy of Sciences, China, [2]China National Machinery Industry Corporation, China

Abstract. The feature selection is an important step which can reduce the dimensionality and improve the performance of the classifiers in text categorization. Many popular feature selection methods do not consider the difference in the distribution of different categories on a feature. In this paper, we propose a new filter based feature selection algorithm, namely fused distance feature selection (FDFS), which evaluates the significance of a feature by taking account of the difference in the distribution of different categories and selects more discriminative features with the minimal number. The proposed algorithm is investigated both inside and outside perspectives on four benchmark document datasets, 20- Newsgroups, WebKB, CSDMC2010 and Ohsumed, using Linear Support Vector Machine (LSVM) and Multinomial Naive Bayes (MNB) classifiers. The experimental results indicate that our proposed method provides a competitive result, where its average ranking is 1.25 on LSVM and 1 on MNB.

Opinion-aware Knowledge Embedding for Stance Detection

Zhenhui Xu[1], Qiang Li[2], Wei Chen[1], Yingbao Cui[2], Zhen Qiu[2], Tengjiao Wang[1]
[1]Peking University, China, [2]State Grid Information and Telecommunication Group, China

Abstract. As an emerging text classification task, stance detection is much helpful in reviewing subjective text and mining expressed attitudes of a person or organization towards an object. Due to the similarity with other text classification tasks, stance detection is always tackled by conventional classification methods. However, there is a big difference between stance detection and others, since stance detection depends much on human background knowledge while others do not. Therefore, to address such a unique problem, we propose a novel method, which leverages knowledge graph and incorporates text-mentioned knowledge with a deep classifier, by a key component named Opinion-aware Knowledge Embedding (OKE). The proposed OKE can integrate the objective knowledge facts and subjective text opinion well by a customized and effective attention mechanism. Our experiments also show that the proposed method comprehensively outperforms all the baselines on real datasets.

History-driven Entity Categorization

Yijun Duan, Adam Jatowt, Katsumi Tanaka
Kyoto University, Japan

Abstract. Knowledge of entity histories is often necessary for comprehensive understanding and characterization of entities. In this paper we introduce a novel task of history-based entity categorization. Taking a set of entity-related documents as an input we detect latent entity categories whose members share similar histories, effectively, grouping entities based on the similarities of their historical developments. Next, we generate comparative timelines for each generated group allowing users to spot similarities and differences in entity histories. We evaluate our approach on several datasets of different entity types demonstrating its effectiveness against competitive

baselines.

Research Session 10: Social Networks

Time: 13:30-15:30, August 3, 2019
Location: 2F Yingui Hall
Chair: Bin Liu

A Learning Approach for Topic-aware Influence Maximization

Shan Tian[1], Ping Zhang[2], Songsong Mo[1], Liwei Wang[1], Zhiyong Peng[1]
[1]Wuhan University, China, [2]Huawei, China

Abstract. Motivated by the application of viral marketing, the topic-aware influence maximization (TIM) problem has been proposed to identify the most influential users under given topics. In particular, it aims to find k seeds (users) in social network G, such that the seeds can maximize the influence on users under the specific query topics. This problem has been proved to be NP-hard and most of the proposed techniques suffer from the efficiency issue due to the lack of generalization. Even worse, the design of these algorithms requires significant specialized knowledge which is hard to be understood and implemented. To overcome these issues, this paper aims to learn a generalized heuristic framework to solve TIM problems by meta-learning. To this end, we encode the feature of each node by a vector and introduce a deep learning model, called deep-influence-evaluation-model (DIEM), to evaluate users' influence under different circumstances. Based on this model, we can construct the solution according to the influence evaluations efficiently, rather than spending a high cost to compute the exact influence by considering the complex graph structure. We conducted experiments on generated graph instances and real-world social networks. The results show the superiority in performance and comparable quality of our framework.

In Pursuit of Social Capital: Upgrading Social Circle Through Edge Rewiring

Qian Chen[1], Hongyi Su[1], Jiamou Liu[2], Bo Yan[1], Hong Zheng[1], He Zhao[1]
[1]Beijing Institute of Technology, China, [2]The University of Auckland, New Zealand

Abstract. The paper investigates the dynamics of the social circle of an individual in a social network. Updating social circle affects two kinds social assets: Bonding social capital which affects trusts and social support, and bridging social capital that determines information access. We address a rewiring process that enables the individual to upgrade her social circle. The questions are (1) what strategies would guide the individual to iteratively rewire social ties to gain bridging while maintain bonding social capital, and (2) what structural properties will arise as a result of applying the strategies. For the first problem, we put forward three greedy rewiring strategies based on scoping the network access for the individual. We conduct experiments over four random graph models and five real-world datasets to evaluate these strategies. The results reveal a striking difference between bonding and bridging social capitals, while the community-based strategy is able to achieve a balance between bridging with bonding social capital. For the second problem, we correlate

social capital with structural features such as centrality and embeddedness. In this respect, the paper advances understanding to social capital and its connections to network structures.

AERIAL: An Efficient Randomized Incentive-based Influence Maximization Algorithm

Yongyue Sun, Qingyun Wang, Hongyan Li
Peking University, China

Abstract. In social networks, once a user is more willing to influence her neighbors, a larger influence spread will be boosted. Inspired by the economic principle that people respond rationally to incentives, properly incentivizing users will lift their tendencies to influence their neighbors, resulting in a larger influence spread. However, this phenomenon is ignored in traditional IM studies. This paper presents a new diffusion model, IB-IC Model (Incentive-based Independent Cascade Model), to describe this phenomenon, and considers maximizing the influence spread under this model. However, this work faces great challenge under high solution quality and time efficiency. To tackle the problem, we propose AERIAL algorithm with solutions not worse than existing methods in high probability and $O(n^2)$ average running time. We conduct experiments on several real-world networks and demonstrate that our algorithms are effective for solving IM Problem under IB-IC Model.

Time Optimal Profit Maximization in a Social Network

Yong Liu, Zitu Liu, Shengnan Xie, Xiaokun Li
HeiLongJiang University, China

Abstract. Influence maximization aims to seek k nodes from a social network such that the expected number of activated nodes by these k nodes is maximized. However, influence maximization is different from profit maximization for a real marketing campaign. We observe that when promotion time increases, the number of activated nodes tends to be stable eventually. In this paper, we first use real action log to propose a novel influence power allocation model with time span called IPA-T, and then present time optimal profit maximization problem called TOPM based on IPA-T. To address this problem, we propose an effective approximation algorithm called Profit-Max. Experimental results on real datasets verify the effectiveness and efficiency of Profit-Max.

Research Session 11: Information Extraction and Retrieval

Time: 13:30-15:30, August 3, 2019
Location: 2F Yinxing Hall
Chair: Yongpan Sheng

Two-Encoder Pointer-Generator Network for Summarizing Segments of Long Articles

Junhao Li, Mizuho Iwaihara

Waseda University, Japan

Abstract. Usually long documents contain many sections and segments. In Wikipedia, one article can usually be divided into sections and one section can be divided into segments. But although one article is already divided into smaller segments, one segment can still be too long to read. So, we consider that segments should have a short summary for readers to grasp a quick view of the segment. This paper discusses applying neural summarization models including Seq2Seq model and pointer generator network model to segment summarization. These models for summarization can take target segments as the only input to the model. However, in our case, it is very likely that the remaining segments in the same article contain descriptions related to the target segment. Therefore, we propose several ways to extract an additional sequence from the whole article and then combine with the target segment, to be supplied as the input for summarization. We compare the results against the original models without additional sequences. Furthermore, we propose a new model that uses two encoders to process the target segment and additional sequence separately. Our results show our two-encoder model outperforms the original models in terms of ROGUE and METEOR scores.

Enhancing Joint Entity and Relation Extraction with Language Modeling and Hierarchical Attention

Renjun Chi, Bin Wu, Linmei Hu, Yunlei Zhang
Beijing University of Posts and Telecommunications, China

Abstract. Both entity recognition and relation extraction can benefit from being performed jointly, allowing them to enhance each other. However, existing methods suffer from the sparsity of relevant labels and strongly rely on external natural language processing tools, leading to error propagation. To tackle these problems, we propose an end-to-end joint framework for entity recognition and relation extraction with an auxiliary training objective on language modeling, i.e., learning to predict surrounding words for each word in sentences. Furthermore, we incorporate hierarchical multi-head attention mechanisms into the joint extraction model to capture vital semantic information from the available texts. Experiments show that the proposed approach consistently achieves significant improvements on joint extraction task of entities and relations as compared with strong baselines.

An Unsupervised Learning Approach for NER based on Online Encyclopedia

Maolong Li[1], Qiang Yang[3], Fuzhen He[1], Zhixu Li[1,2], Pengpeng Zhao[1], Lei Zhao[1], Zhigang Chen[2]
[1]Soochow University, China, [2]IFLYTEK, China, [3]King Abdullah University of Science and Technology, SA

Abstract. Named Entity Recognition (NER) is a core task of NLP. State-of-art supervised NER models rely heavily on a large amount of high-quality annotated data, which is quite expensive to obtain. Various existing ways have been proposed to reduce the heavy reliance on large training data, but only with limited effect. In this paper, we propose a novel way to make full use of the weakly-annotated texts in encyclopedia pages for exactly unsupervised NER learning, which is expected to provide an opportunity to train the NER model with no manually-labeled data at all. Briefly, we roughly divide the sentences of encyclopedia pages into two parts simply according to the density of inner url links contained in each sentence. While a relatively small number of sentences with dense links are used directly for training the NER model initially, the left sentences with sparse links are then smartly selected for gradually promoting the model in several self-training iterations. Given the limited number of sentences with dense links for training, a data augmentation method is proposed, which could generate a lot more training data with the help of the structured data of encyclopedia to greatly augment the training effect. Besides, in the iterative self-training step, we propose to utilize a graph model to help estimate the labeled quality of these sentences with sparse links, among which those with the highest labeled quality would be put into our training set for updating the model in the next iteration. Our empirical study shows that the NER model trained with our unsupervised learning approach could perform even better than several state-of-art models fully trained on newswires data.

Pseudo Topic Analysis for Boosting Pseudo Relevance Feedback

Rong Yan, Guanglai Gao

Inner Mongolia University, China

Abstract. Traditional Pseudo Relevance Feedback (PRF) approaches fail to mode real-world intricate user activities. They naively assume that the first-pass top-ranked search results, i.e. the pseudo relevant set, have potentially relevant aspects for the user query. It is make the major challenge in PRF lies in how to get the reliability relevant feedback contents for the user real information need. Actually, there are two problems should not be ignored: (1) the assumed relevant documents are intertwined with the relevant and the non-relevant content, which influence the reliability of the expansion resource and can not concentrate in the real relevant portion; (2) even if the assumed relevant documents are real relevant to the user query, but they are always semantic redundance with various forms because the peculiarity of natural language expression. Furthermore, it will aggravate the ‘query drift’ problem. To alleviate these problems, in this paper, we propose a novel PRF approach by diversifying feedback source, which main aim is to converge the relatively single semantic as well as diversity relevant information from the pseudo relevant set. The key idea behind our PRF approach is to construct an abstract pseudo content obtained from topical networks modeling over the set of top-ranked documents to represent the feedback documents, so as to cover as diverse aspects of the feedback set as possible in a small semantic granularity. Experimental results conducted in real datasets indicate that the proposed strategies show great promise for searching more reliable feedback source by helping to achieve query and search result diversity without giving up precision.

Research Session 12: Storage and Indexing

Time: 16:00-18:00, August 3, 2019

Location: 2F Jingui Hall
Chair: Hyuk-Yoon Kwon

FreshJoin: an Efficient and Adaptive Algorithm for Set Containment Join

Jizhou Luo, Wei Zhang, Shengfei Shi, Hong Gao, Jianzhong Li, Tao Zhang, Zening Zhou
Harbin Institute of Technology, China

Abstract. This paper revisits set containment join (SCJ), which has many fundamental applications in commercial and scientific fields. To improve the performance further, this paper proposes a new adaptive parameter-free in-memory algorithm for SCJ, named as FreshJoin. It accomplishes this by exploiting two flat indices, which record three kinds of signatures (i.e., the two least frequent elements and a hash signature). Experiments on 16 real-life datasets show that FreshJoin usually reduces more than 50% of space costs while remains as competitive as the state-of-the-art algorithms in running time.

Apara: Workload-aware Data Partition and Replication for Parallel Databases

Xiaolei Zhang, Chunxi Zhang, Yuming Li, Rong Zhang, Aoying Zhou
East China Normal University, China

Abstract. Data partition and replication mechanisms directly determine query execution patterns in parallel database systems, which have a great impact on system performance. Recently, there have been some workload-aware data storage techniques, but they suffer from problems of narrow support to complex workloads or large requirements for storage. In order to enable the support for complex analytical workloads over massive distributed database systems, we design and implement a workload-aware data partition and replication tool, called Apara. We design two heuristic algorithms and define two cost models for effective data partition calculation and efficient replication usages. We run a set of experiments to compare and demonstrate the performance between Apara and the other representative work. The results show that Apara consistently outperforms the primary solutions on TPC-H workload.

An Efficient Top-k Spatial Join Query Processing Algorithm on Big Spatial Data

Baiyou Qiao, Bing Hu, Xiyu Qiao, Laigang Yao, Junhai Zhu, Gang Wu
Northeastern University, China

Abstract. Based on Spark platform, we propose an efficient top-k spatial join query processing algorithm on big spatial data, in which, the whole data space is divided into same-sized cells by using a grid partitioning method. Then spatial objects in two data sets are projected and replicated to these cells by projection and replication operations respectively, meanwhile a filtering operation is used to speed up the processing. After that, an R-tree based local top-k spatial join algorithm is proposed to compute the top-k candidate results in each cell, which extends the traditional R-tree index and combines threshold filtering techniques to reduce the communication and computation costs, therefore speeding up the query processing. Experimental results on synthetic data sets show that the proposed algorithm is significantly better than the existing top-k spatial join query processing algorithms in performance.

Which Category Is Better: Benchmarking the RDBMSs and GDBMSs

Pengjie Ding[1], Yijian Cheng[1], Wei Lu[1], Hao Huang[2], Xiaoyong Du[1]
[1]Renmin University of China, China, [2]Wuhan University, China

Abstract. Relational database management systems (RDBMSs) have been a common option to manage structured data over the past decades. In recent years, with the prevalence of big data applications, vast unstructured and semi-structured data are generated, deeply challenging the relational model used in RDBMSs. For this reason, a wide spectrum of NoSQL databases are developed for managing unstructured, semi-structured or structured data. For example, graph database management systems (GDBMSs) are commonly used as an important category of NoSQL databases, to manage sophisticated graph data as well as relational data. Nonetheless, as claimed in existing literatures, both RDBMSs and GDBMSs are capable of managing graph data and relational data, the boundaries of them still remain unclear. In this paper, we propose a unified benchmark for RDBMSs and GDBMSs, to evaluate them under the same metrics, and report which category is better in different application scenarios. We conduct extensive experiments over the unified benchmark, and report our findings: (1) RDBMSs are significantly faster for aggregations and order by operations, (2) GDBMSs are shown to be superior for projection, multi-table join and deep recursive operations, (3) GDBMSs and RDBMSs are comparable for two-table join and shallow recursive operations.

Research Session 13: Spatial-Temporal Databases

Time: 16:00-18:00, August 2, 2019
Location: 2F Yuegui Hall
Chair: Jianqiu Xu

A Meta-Path-Based Recurrent Model for Next POI Prediction with Spatial and Temporal Contexts

Hengpeng Xu[1], Peizhi Wu[1], Jinmao Wei[1], Zhenglu Yang[1], Jun Wang[2]
[1]Nankai University, China, [2]Ludong University, China

Abstract. Predicting next point of interest (POI) of users in location-based social networks has become an increasingly significant requirement, because of its potential benefits for individuals and businesses. Recently, various recurrent neural network architectures have incorporated contextual information associated with users' sequence of check-ins to capture their dynamic preferences. However, these architectures are limited because they only take the sequential order of check-ins into account and face difficulties in remembering long-range dependencies. In this work, we resort to the heterogeneous of information network (HIN) to address these issues. Specifically, a novel attentional meta-path-based recurrent neural network is proposed, dubbed ST-HIN. ST-HIN predicts the next POI of users from their spatial-temporal incomplete historical check-in sequences, and uses the multi-modal recurrent neural network to capture the complex transition relationship. Furthermore, a meta-path attention embedding module is devised to capture the mutual influence between the users meta-path-based global information in HIN and the dynamic status of their current mobility. The results of extensive experiments performed on real-world datasets demonstrate the effectiveness of our proposed model.

Spatial Temporal Trajectory Similarity Join

Tangpeng Dan[1], Changyin Luo[1], Yanhong Li[2], Bolong Zheng[3], Guohui Li[3]
[1]Central China Normal University, China, [2]South-Central University for Nationalities, China [3]Huazhong University of Science and Technology, China

Abstract. Existing works only focus on spatial dimension without the consideration of combining spatial and temporal dimensions together when processing trajectory similarity join queries, to address this problem, this paper proposes a novel two-level grid index which takes both spatial and temporal information into account when processing spatial-temporal trajectory similarity join. A new similarity function MOGS is developed to measure the similarity in an efficient manner when our candidate trajectories have high coverage rate CR. Extensive experiments are conducted to verify the efficiency of our solution.

Multi-View based Spatial-Keyword Query Processing for Real Estate

Xi Duan, Liping Wang, Shiyu Yang
East China Normal University, China

Abstract. The real estate search web systems such as Zillow, Anjuke, and Lianjia have become very popular in daily life. Generally, the comprehensive query results combined with transportation, health care, education, POIs, etc. are expected, but those surrounding information are rarely utilized in traditional query methods, which thereby restricts the results of the query. In this paper, we address the above limitations and provide a novel multi-view based query method, named KBHR. We investigate feature extraction method and introduce multi-view to represent comprehensive real estate data. The proposed method, KBHR, is based on BHR-tree which is a hybrid indexing structure and a kernel based similarity function developed to rank the query results of multi-view data. We construct experiments and evaluate KBHR on real-world data sets. The experimental results demonstrate the efficiency and effectiveness of our method.

ST-DCN: A Spatial-Temporal Densely Connected Networks for Crowd Flow Prediction

Longlong Xu, Xiansheng Chen, Yue Xu, Wei Chen, Tengjiao Wang
Peking University, China

Abstract. The accurate prediction of crowd flow is of great significance for urban traffic management and public safety. Its key challenge lies in how to model the complex non-linear spatial-temporal dependencies and other external factors such as holidays and weather conditions. In this paper, we propose a novel deep-learning-based approach to address this problem, called Spatial-Temporal Densely Connected

Networks (STDCN), which is able to predict both inflow and outflow of crowds in every region of a city. Specifically, ST-DCN consists of three parts: spatial module, temporal module and external module. The spatial module is designed with a densely connected convolutional structure to capture the spatial dependencies at a citywide level. The temporal module is composed of ConvLSTM units to learn long-term temporal dependencies. We propose an external module consisting of fully connected layers for modeling the external factors. Then the outputs of these three modules are merged to predict the final crowd flow in each region. ST-DCN can alleviate the vanishing-gradient problem and strengthen the propagation of spatial features in very deep network. In addition, the spatial features structure can be maintained throughout the network to avoid losing implied spatial information of crowd flow. Experimental results on two real-world datasets demonstrate that ST-DCN achieves significant improvements over the state-of-the-art methods.

Research Session 14: Multimedia Databases

Time: 16:00-18:00, August 3, 2019
Location: 2F Yingui Hall
Chair: Xing Xu

A Framework for Image Dark Data Assessment

Yu Liu[1], Yangtao Wang[1], Ke Zhou[1], Yujuan Yang[1], Yifei Liu[1], Jingkuan Song[2], Zhili Xiao[3]
[1]Huazhong University of Science and Technology, China, [2]University of Electronic Science and Technology of China, China, [3]Tencent Inc., China

Abstract. Blindly applying data mining techniques on image dark data whose content and value are not clear, is highly likely to bring undesired result. Therefore, we propose an assessment framework which includes offline and online stages for image dark data. In offline stage, we first transform images into hash codes by Deep Self-taught Hashing (DSTH) algorithm, then construct a semantic graph, and finally use our designed Semantic Hash Ranking (SHR) algorithm to calculate the importance score. During online stage, we first translate the user's query into hash codes, then match the suitable data contained in the dark data, and finally return the weighted average value of these matched data to help the user cognize the dark data. The results on real-world dataset show our framework can apply to large-scale datasets, help the user conduct subsequent data mining work.

Supervised Hashing with Recurrent Scaling

Xiyao Fu[1,2] , Yi Bin[2], Zheng Wang[2] , Qin Wei[1], Si Chen[3]
[1]GuiZhou University, China, [2]University of Electronic Science and Technology of China, China, [3]China Electronics Technology Group Corporation, China

Abstract. Learning to hash is a method that can deal with content-based information retrieval efficiently. Traditional learning to hash methods, however, lack the ability to map the generated hash codes to the high-level semantic space. Attributes, as a kind of higher level of visual data representation compared to features, have the potential ability in deep learning to boost the performance. Utilizing attributes from visual data in deep learning to hash can link every bit of the hash codes and a certain type of attributes, therefore giving the hash code an explicit explanation. This paper presents a novel framework, named Deep Recurrent Scaling Hashing (DRSH), to solve the traditional image retrieval problem. The hash codes generated from DRSH are a combination of the outputs of each step of an enhanced LSTM and features generated from convolutional neural nets and are learned through images' attributes. This RNN is reformed to adjust the decorrelation of data flowing between each cell step, which not only makes the learning phase benefit from the ability of recurrent neural nets to learn with recurrent memory but also enable the availability of each hash bit to preserve distinct information. Experiments show that this framework can achieve appreciable performance on major datasets, and also have the ability to explain the meaning of hash codes based on attributes.

Analysis and Management to Hash-Based Graph and Rank

Yangtao Wang[1], Yu Liu[1], Yifei Liu[1], Ke Zhou[1], Yujuan Yang[1], Jiangfeng Zeng[1], Xiaodong Xu[1], Zhili Xiao[2]

[1]Huazhong University of Science and Technology, China, [2]Tencent Inc., China

Abstract. We study the problem of how to calculate the importance score for each node in a graph where data are denoted as hash codes. Previous work has shown how to acquire scores in a directed graph. However, never has a scheme analyzed and managed the graph whose nodes consist of hash codes. We extend the past methods and design the undirected hash-based graph and rank algorithm. In addition, we present addition and deletion strategies on our graph and rank.

Firstly, we give a mathematical proof and ensure that our algorithm will converge for obtaining the ultimate scores. Secondly, we present our hash based rank algorithm. Moreover, the results of given examples illustrate the rationality of our proposed algorithm. Finally, we demonstrate how to manage our hash-based graph and rank so as to fast calculate new scores in the updated graph after adding and deleting nodes.

Discovering Attractive Segments in the User Generated Video Streams

Jie Zhou[1], Jiangbo Ai[2], Zheng Wang[2], Si Chen[3], Qin Wei[1]

[1]GuiZhou University, China, [2]University of Electronic Science and Technology of China, China, [3]China Electronics Technology Group Corporation, China

Abstract. With the rapid development of digital equipment and the continuous upgrading of online media, a growing number of people are willing to post videos on the web to share their daily lives [1, 2]. Generally, not all video segments are popular with audiences, some of which may be boring. In recent years, crowd-sourced time-sync video comments have emerged worldwide, supporting further research on temporal video labelling. In this paper, we propose a novel framework to achieve the following goal: Predicting which segment in a newly generated video stream will be popular among the audiences. At last, experimental results on real-world data demonstrate the effectiveness of the proposed framework and justify the idea of

predicting the popularities of segments in a video exploiting crowd-sourced time-sync comments as a bridge to analyse videos.

Research Session 15: Machine Learning

Time: 16:00-18:00, August 3, 2019
Location: 2F Yinxing Hall
Chair: Jie Shao

FeatureBand: A Feature Selection Method by Combining Early Stopping and Genetic Local Search

Huanran Xue[1], Jiawei Jiang[1,2], Yingxia Shao[3], Bin Cui[1]
[1]Peking University, China, [2]Tencent Inc, China, [3]BUPT, China

Abstract. Feature selection is an important problem in machine learning and data mining. In reality, the wrapper methods are broadly used in feature selection. It treats feature selection as a search problem using a predictor as a black-box. However, most wrapper methods are time-consuming due to the large search space. In this paper, we propose a novel wrapper method, called FeatureBand, for feature selection. We use the early stopping strategy to terminate bad candidate feature subsets and avoid wasting of training time. Further, we use a genetic local search to generate new subsets based on previous ones. These two techniques are combined under an iterative framework in which we gradually allocate more resources for more promising candidate feature subsets. The experimental result shows that FeatureBand achieves a better trade-off between search time and search accuracy. It is 1.45× to 17.6× faster than the state-of-the-art wrapper-based methods without accuracy loss.

I-mRMR: Incremental Max-Relevance, and Min-Redundancy Feature Selection

Yeliang Xiu[1], Suyun Zhao[1,2], Hong Chen[1,2], Cuiping Li[1,2]
[1]Renmin University, China, [2]Key Laboratory of Data Engineering and Knowledge Engineering, Ministry of Education, Ministry of Education, China

Abstract. An incremental method of feature selection based on mutual information, called incremental Max-Relevance, and Min-Redundancy (I-mRMR), is presented. I-mRMR is an incremental version of Max-Relevance, and Min-Redundancy feature selection (mRMR), which is used to handle streaming data or large-scale data. First, Incremental Key Instance Set is proposed which composes of the non-distinguished instances by the historical selected features. Second, an incremental feature selection algorithm is designed in which the incremental key instance set, replacing of all the seen instances so far, is used in the process of adding representative features. Since the Incremental Key Instance Set is far less than the whole instances, the incremental feature selection by using this key set avoids redundant computation and save computation time and space. Finally, the experimental results show that I-mRMR could significantly or even dramatically reduce the time of feature selection with an acceptable classification accuracy.

TRPN: Matrix Factorization Meets Recurrent Neural Network for Temporal Rating Prediction

Haozhe Zhu, Yanyan Shen, Xian Zhou
Shanghai Jiao Tong University, China

Abstract. Traditional matrix factorization techniques for recommendation have a basic assumption that user interests will not change over time, which is not consistent with the reality. To this end, temporal user-item interaction sequences are important to capture users' dynamic interests towards more accurate and timely recommendation. Previous works used to capture dynamic interests based on the basic recurrent neural networks. However, they do not distinguish the static interests which reflect user's long-term preferences from temporal interests caused by occasional incidents. They also treat all the user's past temporal interests equally when performing future rating prediction. In this paper, we leverage Probabilistic Matrix Factorization (PMF) to learn both static and temporal interests for users, and design a new filtering layer to adaptively feed the static and temporal user information to RNN at different time step. We also apply item-dependent attention mechanism to discriminate the importance of different temporal interactions. We conduct extensive experiments to evaluate the performance of our proposed temporal rating prediction method named TRPN. The results show that TRPN can achieve higher performance than several state-of-the-art methods.

Using Sentiment Representation Learning to Enhance Gender Classification for User Profiling

Yunpei Zheng[1], Lin Li[1], Jianwei Zhang[2], Qing Xie[1], Luo Zhong[1]
[1]Wuhan University of Technology, China, [2]Iwate University, Japan

Abstract. User profiling means exploiting the technology of machine learning to predict attributes of users, such as demographic attributes, hobby attributes, preference attributes, etc. It's a powerful data support of precision marketing. Existing methods mainly study network behavior, personal preferences and post texts to build user profile. Through our data analysis of micro-blog, we find that females show more positive and have richer sentiments than males in online social platform. This difference is very conducive to the distinction between genders. Therefore, we argue that sentiment context is important as well for user profiling. In this paper, we propose to predict one of the demographic labels: gender by exploring micro-blog user posts. Firstly we build a sentiment polarity classifier in advance by training a Long Short-Term Memory (LSTM) model. Next we extract sentiment representations from LSTM middle layer. Lastly we combine sentiment representations with virtual document vectors to train a basic MLP network for gender classification. We conduct experiments on a dataset provided by SMP CUP 2016 in China. Experimental results show that our approach can improve gender classification accuracy by 5.53%, compared with classical MLP gender classification without sentiment context.

Workshops

The 2nd International Workshop on Knowledge Graph Management and Applications (KGMA 2019)

Time: 9:30-12:00, August 1, 2019
Location: 2F Jingui Hall
Chair: Xin Wang

Invited Talk: Processing SPARQL Queries Over Distributed RDF Graphs - A Partial Evaluation-based Approach
Peng Peng
Hunan University

Abstract: With the increasing size of RDF data published on the Web, it is necessary for us to design a distributed database system to process SPARQL queries. In many applications, the RDF graphs are geographically or administratively distributed over the sites, and the RDF repository partitioning strategy is not controlled by the distributed RDF system itself. Thus, partitioning-tolerant SPARQL processing is desirable. We adopt a "partial evaluation and assembly" framework for processing SPARQL queries over a large RDF graph in a distributed environment. Based on properties of subgraph matching over a distributed graph, we introduce local partial match as partial answers in each fragment of RDF graph G. For assembly, we propose two methods: centralized and distributed assembly.

Classification-based Emoji Recommendation for User Social Networks
Yuan Wang, Yukun Li, Fenglian Liu
Tianjin University of Technology

Joint Entity and Relation Linking Based on Context Information
Yao Zhao[1], Zhuoming Xu[1], Wei Hu[2]
[1]Hohai University, [2]Nanjing University

Community Detection in Knowledge Graph Network with Matrix Factorization Learning
Xiaohua Shi, Yin Qian, Hongtao Lu
Shanghai Jiao Tong University

Research Review on Relationship Extraction of Knowledge Graph
Aoran Li, Xinmeng Wang, Bohan Li
Nanjing University of Aeronautics and Astronautics

The First International Workshop on Data Science for Emerging Applications (DSEA 2019)

Time: 15:20-17:00, August 1, 2019
Location: 2F Jingui Hall
Chair: Zhixu Li

PEVR: Pose Estimation for Vehicle Re-identification

Saifullah Tumrani, Zhiyi Deng, Abdullah Aman Khan, Waqar Ali
University of Electronic Science and Technology China

The Research of Chinese Ethnical Face Recognition Based on Deep Learning

Qike Zhao[1], Tangming Chen[2], Xiaosu Zhu[2], Jingkuan Song[2]
[1]GuiZhou University, [2] University of Electronic Science and Technology of China

Model of Charging Stations Construction and Electric Vehicles Development Prediction

Qilong Zhang[1], Zheyong Qiu[2], Jingkuan Song[1]
[1]University of Electronic Science and Technology of China, [2]Hangzhou Dianzi University

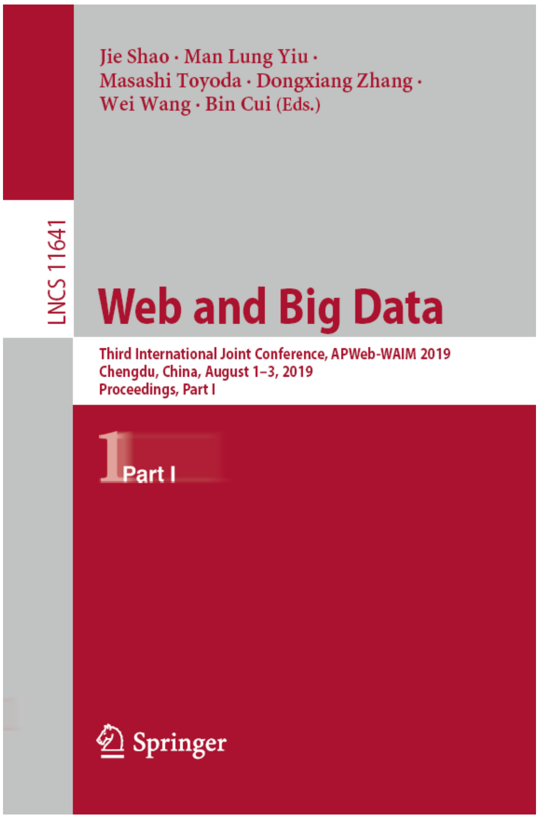
Boundary Detector Encoder and Decoder with Soft Attention for Video Captioning

Tangming Chen[1], Qike Zhao[2], Jingkuan Song[1]
[1]GuiZhou University, [2] University of Electronic Science and Technology of China

Conference Proceedings

The proceedings of APWeb-WAIM 2019 main conference are free to access for the period July 30 to August 30, 2019.

Please visit http://cfm.uestc.edu.cn/apwebwaim2019/proceedings.html or scan the following QR code for access (please use only the provided links on the conference website to made the access and its authentication working).



Jie Shao · Man Lung Yiu ·
Masashi Toyoda · Dongxiang Zhang ·
Wei Wang · Bin Cui (Eds.)

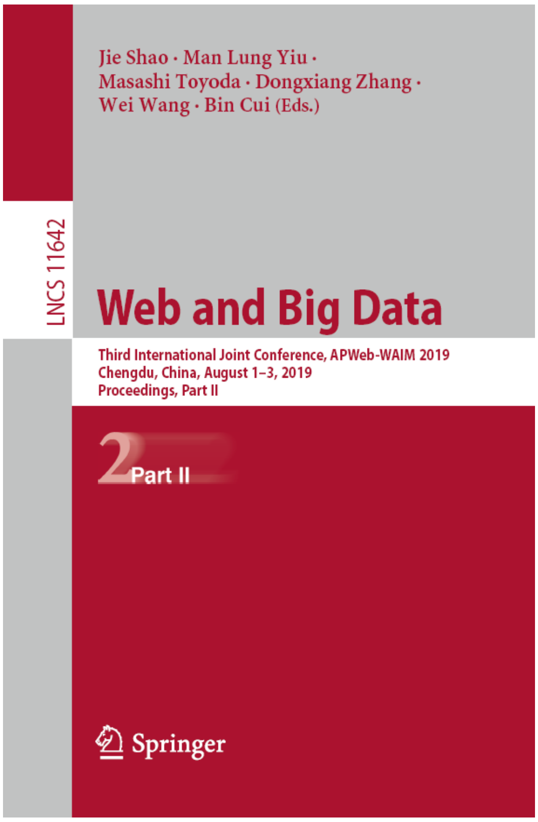
LNCS 11641

Web and Big Data

Third International Joint Conference, APWeb-WAIM 2019
Chengdu, China, August 1–3, 2019
Proceedings, Part I

1 Part I

Springer



Jie Shao · Man Lung Yiu ·
Masashi Toyoda · Dongxiang Zhang ·
Wei Wang · Bin Cui (Eds.)

LNCS 11642

Web and Big Data

Third International Joint Conference, APWeb-WAIM 2019
Chengdu, China, August 1–3, 2019
Proceedings, Part II

2 Part II

Springer

The papers presented at APWeb-WAIM 2019 workshops (KGMA and DSEA) will be published in a volume of post-proceedings by Springer LNCS after the conference.

Note

Note