Zhou X, Yin H. Special section on recommender systems with big data. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 33(4): 621–624 July 2018. DOI 10.1007/s11390-018-1844-1

Preface

It is our great pleasure to announce the publication of this special section in JCST: "Recommender Systems with Big Data".

As guest editors of this special section, we would like to introduce some special points. Although recommender system research has made significant advances over the past decades, traditional recommendation techniques are not powerful enough to address new challenges arising from the "4Vs" (volume, variety, velocity, veracity) in the Big Data era. Firstly, large volumes of user behaviour data and media data are generated at unprecedented and ever-increasing scales. Existing recommendation techniques are designed for the conventional scale datasets, struggling to meet the requirements of scalability and storage. Secondly, media or product content data involves a great variety of data formats in different modalities: texts, images, videos and arbitrary combinations of them. Thirdly, user behaviour and product data are generated in real time and continually arrive in the form of streams. Data streams are temporally ordered, continuous, and come in high velocity. Fourthly, large amounts of biases, noises and abnormality exist in user generated behaviour and content data (e.g., user comments). What dimensions of data quality are particularly important for recommenders and what methods can address them? The data veracity is another big challenge for recommender systems. This special section aims to address these new challenges to enable both accurate and scalable recommendations in the era of Big Data. As we can see, recommender systems in this section are general, but their focus on Big Data challenges is special.

We launched call for papers and sought original and high quality research papers from all over the world. 21 submissions were received. First, quick reviews were performed on all submissions and those of insufficiently high quality were immediately rejected. Then, for the remaining submissions, at least two reviewers thoroughly reviewed each of them. Two rounds of review were carried out for all those submissions. The authors were asked to address all the major and minor issues in their submissions during the review process. Eventually we were able to accept 10 high quality submissions in terms of clarity, novelty, significance, and relevance. The accept rate was 47%. Although the topics of the accepted papers in this issue are broad, they can be roughly classified into 4 topics: active learning techniques, content-aware recommendation systems, context-aware recommendation systems, and recommender system focusing on new challenges.

Active Learning Techniques

Active learning techniques in recommender systems are used to acquire high-quality and informative user feedback data.

The paper "A Survey on Expert Recommendation in Community Question Answering" conducts a survey on expert recommendation in community question answering. Expert recommendation is effective in sending a high-quality answer to a user in community question answering platforms in a short time. The inputs of an expert recommendation problem include users (i.e., requesters and answerers), user-generated content (i.e., the questions raised by requesters and the answers provided by answerers), user profiles (such as badges, reputation scores, and links to external resources such as Web pages), user feedback on questions and answers (e.g., textual comments and voting), and more detailed information about questions (e.g., categories to which the questions belong, or the closely-related or duplicated questions). Instead of passively waiting for users to browse and find their questions of interest, an expert recommendation system actively raises the attention of users to the appropriate questions and acquires their feedback in a timely manner. This survey presents an overview of the recent research efforts

and state-of-the-art techniques for the expert recommendation in CQA, and summarizes and compares the existing methods with respect to their advantages and shortcomings.

The paper "GRIP: A Group Recommender Based on Interactive Preference Model" presents a group recommender based on interactive preference model. A group recommender reflects the analysis of multiple users' behavior, and aims to provide each user of the group with the things they prefer. Currently, most of the existing group recommenders ignore the interaction among the users. However, in the course of group activities, the interactive preferences will dramatically affect the success of a recommender. To solve this problem, an interaction-based method named GRIP (Group Recommender Based on Interactive Preference) is presented which can use group activity history information and post-rating feedback mechanism to generate interactive preference parameters. The paper will be published later in Vol.33, No.5.

Content-Aware Recommendation Systems

The papers in this category adopt most advanced algorithms and models, especially deep learning techniques, to exploit heterogeneous and multi-modal content information to provide more accurate recommendations. In this way, they are able to address data sparsity and cold start issues in recommender systems.

The paper "Jointly Recommending Library Books and Predicting Academic Performance: A Mutual Reinforcement Perspective" proposes a supervised content-aware matrix factorization for mutual reinforcement of academic performance prediction and library book recommendation in the paper. This model not only addresses the sparsity challenge by explainable dimension reduction techniques, but also quantifies the importance of library books in predicting academic performance.

The paper "Multiple Auxiliary Information Based Deep Model for Collaborative Filtering" integrates plot information as auxiliary information into the denoising autoencoder (DAE), called SemRe-DCF, which aims at learning semantic representations of item descriptions and succeeds in capturing fine-grained semantic regularities by using vector arithmetic to get better rating prediction in the paper. The results manifest that the proposed method can effectively improve the accuracy of prediction, and solve the cold start problem.

The paper "Exploiting Pre-Trained Network Embeddings for Recommendations in Social Networks" takes the advantage of network embedding techniques and proposes an embedding-based recommendation method, which is composed of the embedding model and the collaborative filtering model. Specifically, to exploit the deep structure hidden in social networks and rating patterns, a neural network based embedding model is first pre-trained and then these extracted factors are incorporated into collaborative filtering model by fusing them with latent factors linearly.

Context-Aware Recommendation Systems

Context-aware recommender systems exploit and integrate various contextual information, such as spatial, temporal and social information, to improve recommendation performance.

The paper "Discovering Functional Organized Point of Interest Groups for Spatial Keyword Recommendation" designs clustering algorithms to obtain organized POI (i.e., point of interest) groups (OPGs) and utilizes OPGs-LDA (Latent Dirichlet Allocation) model to reveal functions of OPGs for spatial keyword recommendation. To the best of our knowledge, this is the first to study functional organized POI groups which have important applications in urban planning and social marketing.

Hashtag recommendation for microblogs is a very hot research topic that is useful to many applications involving microblogs. However, since short text in microblogs and low utilization rate of hashtags will lead to the data sparsity problem, it is difficult for typical hashtag recommendation methods to achieve accurate recommendation. The paper

"Hashtag Recommendation Based on Multi-Features of Microblogs" proposes HRMF, a hashtag recommendation method based on multi-features of microblogs. HRMF expands short text into long text, and then it simultaneously models multi-features (i.e., user, hashtag, text) of microblogs by designing a new topic model. To further alleviate the data sparsity problem, HRMF exploits hashtags of both similar users and similar microblogs as the candidate hashtags. In particular, to find similar users, HRMF combines the designed topic model with typical user-based collaborative filtering method.

The paper "A Generative Model Approach for Geo-Social Group Recommendation" focuses on meeting the needs of a group of users in social networks by addressing two major challenging problems. The first challenge is that the choice of a member in a group is influenced by various factors, i.e., personal preference, group topic, and social relationship. The second challenge lies in that users have different influences when in different groups.

Recommender System Focusing on New Challenges

This category of papers presents recommender systems that focus on at least one of the new challenges or new problems.

The paper "Illuminating Recommendation by Understanding the Explicit Item Relations" presents a survey of recent advances on recommender systems from the perspective of explicit item relation understanding. For quite a long time, most researchers have been pursing recommendation performances with predefined metrics, e.g., accuracy. However, in real-world applications, users select items from a huge item list by considering their internal personalized demand and external constraints. Thus, Liu *et al.* argue that explicitly modeling the complex relations among items under domain-specific applications is an indispensable part for enhancing the recommendations.

The paper "PTM: A Topic Model for the Inferring of the Penalty" focuses on exploiting artificial intelligence and machine learning in the problem of recommending penalty to assist the judicial study. It proposes a latent-class probabilistic generative model to infer the topic of law cases, and the temporal and spatial patterns of topics embedded in the case judgement. Then, the learnt knowledge is utilized to automatically cluster all cases accordingly in a unified way.

Acknowledgment

Finally, we would like to thank all the authors who submitted papers to this special section. We would also like to express our sincere gratitude to the reviewers of this special section who diligently assisted us in reviewing the submissions. Without the reviewers' significant contributions, this special section would not have been published. We also thank the Editorial Director Ms Fengdi Shu for her encouragement, guidance, and help.

We hope that you enjoy reading this special section as much as we did.

Leading Editor

Xiaofang Zhou, Professor, IEEE Fellow, School of Information Technology and Electrical Engineering
The University of Queensland, Brisbane zxf@itee.uq.edu.au

Guest Editor

Hongzhi Yin, Assistant Professor, School of Information Technology and Electrical Engineering
The University of Queensland, Brisbane h.yin1@uq.edu.au

Xiaofang Zhou is a professor of computer science at the University of Queensland (UQ), Brisbane, and the leader of Data Science Research Group at UQ, which includes the Data and Knowledge Engineering (DKE) Group. Prof. Zhou received his B.Sc. and M.Sc. degrees in computer science from Nanjing University, Nanjing, and Ph.D. degree in computer science from UQ. Before joining UQ in 1999, he worked as a researcher in Commonwealth Scientific and Industrial Research Organisation (CSIRO), leading its Spatial Information Systems group. His research focus is to find effective and efficient solutions for managing, integrating and analyzing very large amount of complex data for business, scientific and personal applications. He has been working in the area of spatiotemporal and multimedia databases, data mining, data quality, high performance query processing, big data analytics and machine learning, and co-authored over 300 research papers with many published in top journals and conferences such as SIGMOD, VLDB , ICDE, ACM Multimedia, AAAI, IJCAI, The VLDB Journal, ACM and IEEE Transactions. He was the Program Committee Chair of Australasian Database Conferences (ADC 2002 and 2003), International Conference on Web Information Systems Engineering (WISE 2004), Asia Pacific Web Conferences (APWeb 2003 and 2006), International Conference on Databases Systems for Advanced Applications (DASFAA 2009, DASFAA 2012 and DASFAA 2015), International Conference on Cooperative Information Systems (CoopIS 2012), IEEE International Conference on Data Engineering (ICDE 2013), ACM International Conference on Information and Knowledge Management (CIKM 2016), International Symposium on Spatial and Temporal Databases (SSTD 2017), and International Conference on Very Large Databases (VLDB 2020). He is a General Co-Chair of ACM Multimedia Conference 2015, IEEE International Conference on Data Management (MDM 2018), and China Big Data Technology Conference 2017. He has been on the program committees of numerous international conferences, often as a Senior PC member, including SIGMOD, VLDB, ICDE, WWW, ACM Multimedia, ICDM, ICDCS and AAAI. He was the Convenor and Director of ARC Research Network in Enterprise Information Infrastructure from 2004-2011 (a major national research collaboration initiative in Australia), and the founding chair of ACM SIGSPATIAL Australian Chapter from 2010-2011. Currently he is an Associate Editor of IEEE Transactions on Cloud Computing, World Wide Web Journal, Distributed and Parallel Databases, Knowledge and Information Systems (since 2017), Springer's Encyclopedia of Database Systems, and Springer's Web Information System Engineering book series. He is the current Chair of IEEE Technical Committee on Data Engineering (TCDE, 2015 present), and the Steering Committees of ICDE, DASFAA, WISE, APWeb and Australasian Database Conferences. In the past he was an Associate Editor of VLDB Journal (2008∼2015), IEEE Data Engineering Bulletin, Information Processing Letters (2009-2015), and IEEE Transactions on Knowledge and Data Engineering (2009∼2015). Xiaofang is a specially appointed Adjunct Professor under the Chinese State Qianren Scheme (short-term) hosted by Renmin University of China (2010∼2013), and by Soochow University (since 2013) where he leads the Research Center on Advanced Data Analytics (ADA). He is a Fellow of IEEE.

Hongzhi Yin is now working as a lecturer in data science (permanent position) and an ARC DECRA Fellow with The University of Queensland, Brisbane. He received his Ph.D. degree from Peking University, Beijing, in July 2014. After graduation, he joined the School of Information Technology and Electrical Engineering, the University of Queensland. He successfully won the ARC DECRA Award in 2015 and obtained an ARC Discovery Project grant as a chief investigator in 2016. His current main research interests include social media analytic, user profiling, recommender system, especially spatial-temporal recommendation, topic discovery and event detection, deep learning, user linkage across social networks, and knowledge graph mining and construction. He has published over 80 peer-reviewed papers in prestigious journals and top international conferences including ACM TOIS, VLDBJ, IEEE TKDE, ACM TKDD, ACM TIST, ACM SIGMOD, ACM SIGKDD, VLDB, IEEE ICDE, AAAI, SIGIR, WWW, ACM Multimedia, ICDM, WSDM and CIKM. He has been actively engaged in professional services by serving as conference organizers, conference PC members for PVLDB, SIGIR, ICDE, IJCAI, AAAI, ICDM, CIKM, DASFAA, ASONAM, MDM, WISE, PAKDD and reviewer of more than 10 reputed journals such as VLDB Journal, and TKDE, TOIS, TKDD, TWeb, IEEE Transactions on Cybernetics, WWW Journal, Knowledge-Based System.