Research Collection School Of Computing and Information Systems

School of Computing and Information Systems

11-2019

Special issue on multimedia recommendation and multi-modal data analysis

Xiangnan HE

Zhenguang LIU

Hanwang ZHANG

Chong-wah NGO
Singapore Management University, cwngo@smu.edu.sg

Svebor KARAMAN

See next page for additional authors

Citation

Author

Xiangnan HE, Zhenguang LIU, Hanwang ZHANG, Chong-wah NGO, Svebor KARAMAN, and Yongfeng
ZHANG

EDITORIAL

Special issue on multimedia recommendation and multi-modal data analysis

Xiangnan He[1] · Zhenguang Liu[2] · Hanwang Zhang[3] · Chong-Wah Ngo[4] · Svebor Karaman[5] · Yongfeng Zhang[6]

Rich multimedia contents are dominating the Web. In popular social media platforms such as FaceBook, Twitter, and Instagram, there are over millions of multimedia contents being created by users on a daily basis. In the meantime, multimedia data consist of data in multiple modalities, such as text, images, audio, and so on. Users are heavily overloaded by the massive multi-modal data, and it becomes critical to explore advanced techniques for heterogeneous big data analytics and multimedia recommendation. Traditional multimedia recommendation and data analysis technologies cannot well address the problem of understanding users' preference in the feature-rich multimedia contents, and have challenges in processing massive and multi-modal data. Moreover, previous works on multimedia recommendation and multi-modal data analysis mainly used shallow features and conventional deep learning methods to process the multimedia contents. Advanced deep learning and machine learning methods and novel deep-feature extraction mechanism are yet to be explored. This special issue seeks original contributions related to either multimedia recommendation or multi-modal data analysis related to recommendation.

Submissions came from an open call for paper and with the assistance of professional referees. Six papers were finally selected after three rounds of rigorous peer review. These accepted papers cover several popular topics of recommendation and multi-modal data analysis, including music recommendation and retrieval, clothes recommendation, app usage recommender system, and multi-modal 3D model recognition. We summarize these accepted papers as follows.

In the paper entitled "Fashion Clothes Matching and Recommendation Scheme Based On Siamese Network and AutoEncoder", the authors present a new fashion clothes matching and recommendation scheme based on improving the method of neural compatibility modeling. There exist excessive complicated factors influencing the relation among clothes items, such as color or material. The paper attempts to address how to extract efficient and accurate features. An efficient clothes matching scheme with Siamese network and autoencoder is proposed, which is based on both labeled data from dataset FashionVC and unlabeled data from MicroBlog.

In the paper entitled "An App Usage Recommender System: Improving Prediction Accuracy for Both Warm and Cold Start Users", an approach is proposed to predict next app usage. It is becoming increasingly difficult to find a particular app on a smartphone due to the increasing number of apps installed. Consequently, it is important to be able to quickly and accurately predict the next app to be used. A modified incremental k-nearest neighbors algorithm is introduced to address the prediction problem, and a cold start strategy is proposed to provide app cold start prediction.

In the paper entitled "Deep learning-based automatic downbeat tracking: a brief review", the authors conduct a survey on music downbeat tracking. Automatically analyzing music is a significant step to satisfy people's need for music retrieval and music recommendation. The paper provides detailed discussions on system architecture, feature extraction, deep neural network algorithms, datasets, and evaluation strategy of existing downbeat tracking methods. The results from the annual benchmark evaluation—Music Information Retrieval Evaluation eXchange, as well as the developments in software implementations are also presented.

In the paper entitled "Toward Efficient Indexing Structure for Scalable Content based Music Retrieval", the authors

✉ Zhenguang Liu
  liuzhenguang2008@gmail.com

[1] National University of Singapore, Singapore, Singapore

[2] Zhejiang Gongshang University, Hangzhou, China

[3] Nanyang Technological University, Singapore, Singapore

[4] City University of Hong Kong, Kowloon Tong, Hong Kong

[5] Columbia University, New York City, USA

[6] Rutgers University, News Brunswick, USA

propose an indexing technique to facilitate scalable and accurate multi-modal music retrieval. The method consists of two main functionality modules: (1) music classification—a novel semantic-sensitive classification to identify an input songs category; and (2) indexing module—multiple local indexing structures, one for each semantic category to reduce query response time significantly.

In the paper entitled "PANORAMA based on Multi-channel-Attention CNN for 3D Model Recognition", to address the multi-modal 3D model recognition problem, the authors propose a panorama based on multi-channel attention CNN network for the representation of 3D models. The proposed method is composed of three parts: extracting views, transform function learning, and generating 3D model descriptor. The main idea is to first extract the 2D panoramic views for each 3D model, and then use the multi-channel attention neural network to extract the descriptor for each 3D model. Finally, the fusion feature is used to handle the 3D model classification and retrieval problem.

In the paper entitled "Multi-Guiding Long-Short Term Memory for Video Captioning", the authors propose a multi-modal long–short-term memory, which is an extension of the conventional LSTM network for video captioning. Global information (i.e., detected attributes) and local information (i.e., appearance features) extracted from the video are added as extra inputs to each cell of LSTM, with the aim of collaboratively guiding the model towards solutions that are more tightly coupled to the video content.