Guest Editor's Introduction

Big Cross-Modal Social Media Data Analytics With Deep Intelligence

Yang Wang
Hefei University of Technology

Tingting Mu
The University of Manchester

Meng Fang and Joey Tianyi Zhou
A-star

Dacheng Tao
The University of Sydney

Big cross-modal social Media Data Analytics with Deep Intelligence aims to handle data sampling from multimodal deep spaces, so as to well characterize the big data. This article reports some recent advancements in this area. The special issue accepted 13 papers after a careful peer review process. The addressed topic span from the range of human action recognition to affective computing, disaster detection, classification, retrieval, clustering, vehicle reidentification, and data security. An overview of accepted papers is presented as follows.

With the aim of mining the connotation of human action from skeleton data, Fu *et al.*[1] proposed directed graph convolution neural networks with a novel residual split structure to better mine the spatiotemporal relationship between nodes and bones in skeleton data. It extracted human behavior by two kinds of graph models, afterwards, employing the novel residual split structure to divide the high-dimensional features into several shallow ones. Finally, randomly sampling data in the training can reduce the resource consumption and enhance the robustness. The experiments on NTU-RGBD dataset validate its effectiveness.

Mental health has caught much attention especially under global outbreak of COVID-19, Sun *et al.*[2] proposed a naturalistic paradigm database related to mental health in the universal scene, which consists of data from multimodal. Based on that, it developed a long-term ubiquitous interpretable psychological computing model based on prior knowledge and multimodal information fusion, which effectively enhances the accuracy of mental health status recognition and prediction. Future work will implement with more knowledge from psychology, psychiatry, brain science, and other fields.

The social media can help the government make better decision during the natural disasters. However, the existing methods can only extract specific features that cannot be shared among different disaster events from textual messages and images. In addition, the use of images are not as frequently as textual messages either. Gao *et al.*[3] proposed a novel multimodal adversarial neural network (MANN) to detect the potential new disasters. The MANN exploits the extracted features to identify disaster-related tweets and estimates the similarity between different disasters that can be used to

Digital Object Identifier 10.1109/MMUL.2020.3034589
Date of current version 20 November 2020.

1070-986X © 2020 IEEE
Published by the IEEE Computer Society

capture transferable features that generalize well for unseen disaster events. The MANN outperforms other baselines on several evaluation metrics.

Guan *et al.*[4] tackled the problem of the incomplete labels in multilabel text classification to exploit label correlations with two label constraints. On one hand, the label manifold regularization can help handle the incomplete ones. On the other hand, the label confidence constraint can avoid overestimating negative labels caused by regularizing labels, leading to a safer inference. Based on that, Bayesian Model with Label manifold Regularization and Label Confidence Constraint (BM-LRC) is proposed to estimate the missing label weights. Empirical studies validate its advantages.

The Android malware detection technology develops with the rapid growth of Android applications. To enhance the ability of existing classification models, Wu *et al.*[5] presented an Android malicious application detection framework termed multimodal information integration technology (MVIIDroid) which integrates a variety of information from an Android app' permissions, APIs and opcodes to determine whether the app is malicious or not. The experiments demonstrate its effectiveness.

Searching for products with images for item retrieval becomes more popular. While directly utilizing detection or parsing algorithms may be deleterious due to distortion and others. So the clothing translation framework named WarpClothingOut, proposed by Zhang *et al.*[6], learned the mapping from the human body clothing images to tiled clothing images, and fuses the two kinds of coarse images generated from generative adversarial networks and thin plate spline transformation. The generated tiled clothing images can preserve clothing characteristic and have the similar distribution with real data which can directly used for clothing retrieval.

A paper relates to the work discussed above is the one authored by Cao *et al.*[7] that addressed the issues of lacking the capability of discovering global statistical distribution of feature representations at multiscales among the existing methods in image retrieval. Cao *et al.*[7] proposed a gated multiscale NetVLAD network to capture content semantic and structural information at each scale of images. This architecture combines the gated FPN with NetVLAD, the FPN captured multiscale information implied in the images and the NetVLAD learned the features' statistical distribution over the global feature space. The gating mechanism on every levels of the feature pyramid network can further balance the information flowing.

The studies of cross-modal retrieval seldom combines correlation learning and adversarial learning to bridge the intermodal semantic gap and diminish cross-modal heterogeneity. Based on that, Zhu *et al.*[8] proposed a novel cross-modal retrieval method, named adversarial learning based semantic correlation representation, this is an end-to-end framework to jointly leverage cross-modal representation, correlation, and adversarial learning. The integration of VisNet, TxtNet, and CCA can help learn the cross-modal nonlinear correlation. Besides, cross-modal classifier and modality classifier are used to learn intramodal discrimination and minimize the cross-modal heterogeneity.

Ou *et al.*[9] observed that there are two major issues in the area of multiview clustering, i.e., handling large-scale data efficiently and incorporating the complementary multiple sources efficiently. This paper proposes an anchor-based multiview subspace clustering with diversity regularization (AMVSS-DR) to solve these problems. First, Ou *et al.*[9] evaluated the self-expression similarity matrix to reduce the high time complexities. Then, the HSIC regularization term is imposed to encourage independence and diversity of each views. Finally, compared with several state-of-the-art multiview clustering methods, our proposed algorithm achieves superior clustering performance with low time complexity.

Zhang *et al.*[10] showed that the previous clustering model, named FU-PCM, is difficult to obtain the satisfactory performance when the category distribution is imbalance. This paper proposed a novel clustering model, which exploits Pearson Correlation Coefficient to autobalance the optimal equation according to the category distribution. Moreover, the novel clustering model can extend to incremental learning that learns new samples at current frame and

updates the original model which can enhance the efficiency.

The vehicle reidentification (reID) targets at recognizing the same vehicle images captured by different cameras. In particular, distinguishing vehicles that belong to the same or similar models can be more challenging. The attribute of vehicles could provide rich information to learn the correlation, while the existing methods overlooked the relationship between the attributes and identity. Hence, Wang *et al.*[11] proposed an attribute-guided network (AGNet) that can select details in category feature maps and uses the ALS loss to better regular the AGNet for training more discriminative features. As demonstrated, AGNet could achieve competitive results in comparison with other existing vehicle reID methods.

Zhao *et al.*[12] showed a novel framework based on MM-Fashion to address the topic about clothing recommendation. This network includes new feature extraction and fusion modules which can merge low-level feature information extracted by Inception V3 and high-level semantic information extracted by region convolutional neural network. In addition, the proposed method of extract high-level semantic information will avoid the disadvantages of MM-Fashion like the consistency of the tag information, the validity of the semantic feature and applying on images without semantic expressions. This end-to-end clothing collocation learning framework can jointly learn the compatibility relationships among fashion items as well as visual semantic embedding, which can also be utilized for online clothes collocation.

To protect the EPR data for the healthcare system. Anand *et al.*[13] developed a CTE-based dual watermarking. Utilizing turbo code to encode the EPR data before inserted into the image watermark. The RDWT-RSVD and SPIHT-STE guarantee the robustness and security of EPR. The better performance on imperceptibility, robustness, and security show this technique will be an attractive tool.

ACKNOWLEDGMENTS

The work was supported by National Natural Science Foundation of China, under Grant no. 61806035, U1936217.

REFERENCES

1. B. Fu, S. Fu, L. Wang, Y. Dong, and Y. Ren, "Deep residual split directed graph convolutional neural networks for action recognition," *IEEE MultiMedia*, to be published.
2. X. Sun, Y. Song, and M. Wang, "Towards sensing emotion with deep visual analysis: A long-term psychological modeling approach," *IEEE MultiMedia*, to be published.
3. W. Gau, X. Zhu, Y. Wang, and L. Li, "Detecting disaster-related tweets via multi-modal adversarial neural network," *IEEE MultiMedia*, to be published.
4. Y. Guan and X. Li, "Multi-label text classification with incomplete labels: A safe generative model with label manifold regularization and confidence constraint," *IEEE MultiMedia*, to be published.
5. Q. Wu, M. Li, X. Zhu, and B. Liu, "Mviidroid: A multiple view information integration approach for android malware detection and family identification," *IEEE MultiMedia*, to be published.
6. H. Zhang, X. Wang, L. Liu, D. Zhou, and J. Z. Zhang, "Warpclothingout: A stepwise framework for clothes translation from the human body to tiled images," *IEEE MultiMedia*, to be published.
7. Y. Cao, J. Zhang, and J. Yu, "Image retrieval via gated multi-scale netvlad for social media applications," *IEEE MultiMedia*, to be published.
8. L. Zhu, J. Song, X. Zhu, C. Zhang, S. Zhang, and X. Yuan, "Adversarial learning based semantic correlation representation for cross-modal retrieval," *IEEE MultiMedia*, to be published.
9. Q. Ou, S. Wang, S. Zhou, M. Li, X. Guo, and E. Zhu, "Anchor-based multi-view subspace clustering with diversity regularization," *IEEE MultiMedia*, to be published.
10. J. Zhang, T. Chen, and Y. Zhang, "Incremental clustering with hard centers," *IEEE MultiMedia*, to be published.
11. H. Wang, J. Peng, D. Chen, G. Jiang, T. Zhao, and X. Fu, "Attribute-guided feature learning network for vehicle re-identification," *IEEE MultiMedia*, to be published.
12. M. Zhao, Y. Liu, X. Li, J. Z. Zhang, and Y. Zhang, "An end-to-end framework for clothing collocation based on semantic feature fusion," *IEEE MultiMedia*, to be published.
13. A. Anand, A. K. Singh, Z. Lv, and G. Bhatnagar, "Compression-then-encryption based secure watermarking technique for smart healthcare system," *IEEE MultiMedia*, to be published.