



Intelligent computational techniques for multimodal data

Shishir Kumar¹ · Prabhat Mahanti² · Su-Jing Wang³

Published online: 11 July 2019

© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

This guest editorial introduces the special issue on “Intelligent Computational Techniques for Multimodal Data”. The goal of this special issue was to solve a variety of real-life problems which have uncertainty, imprecision, vagueness, resulting in high performance applications or prototypes for real time system. The special issue touched different hot topics related to Computer Vision, Computational Biology, Multimedia data mining, High-dimensional multimedia data, Deep convolution network, Deep semantic preserving hashing, High-dimensional multimedia classification, Deep CNN and extended residual units, particle swarm optimization, Cyberbullying detection on social multimedia, Multimedia detection algorithm of malicious nodes, Multimedia based fast face recognition, Diversifying personalized mobile multimedia application, Multimedia image edge extraction algorithm, Rumour veracity detection.

Keywords Multimedia data mining · Deep convolution network · Deep semantic preserving hashing · Particle swarm optimization · Cyberbullying detection · Mobile multimedia application · Rumour veracity detection

✉ Shishir Kumar
dr.shishir@yahoo.com

Prabhat Mahanti
pmahanti@unb.ca

Su-Jing Wang
wangpujing@psych.ac.cn

¹ Jaypee University of Engineering & Technology, Guna, India

² University of New Brunswick, Saint John, New Brunswick, Canada

³ Institute of Psychology, Visual Information Processing and Learning Laboratory, Chinese Academy of Sciences, Beijing, China

1 Introduction

We are in the technology age where most of the data is stored in digital format. The internet revolution make the explosive growth of multimedia data possible and easily available to the public, thus producing an immense amount of data with multiple modalities such as text, images, audio, video etc. [1]. Growing amount of research conducted in this field, combined with advances in Artificial Intelligence, Computer Vision and Machine learning, Deep learning has led to the development of advanced intelligent systems that aim to detect and process affective information contained in multimedia and multi-modal sources. The massive size and pretty high dimension features in multimedia data characterize the existing analysis problems in this subject. Extraction of meaningful inferences from such a large-scale, multimodal and noisy data is a challenging and interesting research topic. Moreover, along with quantity and dimensionality, the complexity and diversity of such datasets are also increasing day by day and thus, the existing learning tasks are computationally inapplicable and incompetent for analysing and modelling the multimedia data [3].

This special issue intends to bring together for theoreticians and practitioners from academic fields and industries worldwide working in the broad range of topics relevant to machine learning techniques and deep learning techniques in the field of multimedia data classification and analysis, including multi-modal data modelling, face recognition, audio feature extraction, multimedia text classification and multimedia application recommendation.

To give an overall picture of the special issue contents and coverage, the papers part of it are briefly presented in the following:

In “Deep Convolution Network for Surveillance Records Super-Resolution” Shamsolmoali et al. proposed a deep learning based super-resolution model for surveillance records. The deep Convolutional Neural Network (CNN) with 9 layers has been used to recover the low resolution objects and solve the image boundary problems with the data padded with zeros. The CNN model has the local interconnection between neighbouring layers which extracts the details of the low resolution images efficiently. The proposed model has been trained and tested on the two surveillance dataset SCface dataset and Chokepoint Dataset. The proposed model outperforms the existing approaches in these datasets with good amount of error rate and accuracy [12].

In “Deep semantic preserving hashing for large scale image retrieval” Zareapoor et al. proposed a deep hashing method to extract the abstract features and learning the hash function from the high dimensional images. The proposed method contains the encoder-decoder and a supervisory sub-network which generates a low dimensional binary codes from the hash codes. The performance of the proposed approach has been evaluated on the several large scale image dataset in terms of mAP and precision-recall. The proposed model proves that hashing based autoencoder network is an optimal solution for analysing the large scale visual data [16].

In “Discriminant maximum margin projections for face recognition” Yang et al. presented a novel dimensionality reduction algorithm called discriminant maximum margin projections for the task of face recognition. The proposed algorithm retains the local structure of the data and maximize the between class margin at all local areas. The experiments have been conducted on face recognition datasets such as the ORL, Yale and FERET. The proposed algorithm outperforms the most other advanced approaches in terms of optimal recognition accuracy [15].

In “High-dimensional multimedia classification using deep CNN and extended residual units” Shamsolmoali et al. the hybrid convolutional neural network and residual network is presented to reduce the dimensionality in the high-dimensional multimedia classification task.

The proposed model extracts the deep multimedia features with hundreds of layers and fully connected layers. The presented model considers the features from the fully connected layers because it is more discriminative than the features extracted from convolution layer. The results of the proposed model generates the low dimensional deep features and surpass state-of-the-art image classification approaches [11].

In “Exploiting aggregate channel features for urine sediment detection” Sun et al. proposed a aggregate channel feature plus detector based on aggregate channel features for urine sediment detection. The urine sediment examination refers to the use of microscope to examine various components like red blood cell, white blood cell, tube and crystal etc. The proposed method provides an automatic recognition for the components from the microscopic image for the task of urine sediment analysis. The dataset has been collected from the professional medical institutions which consists of 240 high resolution urine sediment images. The model has been compared with HOG- SVM, ACF, ACDS and evaluated using precision, recall, f1 score and miss rate. As a result the method is more outstanding both in effectiveness and efficiency as compared to other models [14].

In “Multimodal data modelling for efficiency assessment of social priority based urban bus route transportation system using GIS and data envelopment analysis” Singh et al. used multi modal data to access and design a socially efficient public transport bus route plan for the Allahabad city of Uttar Pradesh state, India. The aim of the study is to suggest a scientific approach to evaluate the performance of a particular route from the social perspective. They used a data endvelopment analysis (DEA) and geographical information system (GIS) for the efficiency assessment of existing 24 public transport bus routes. DEA measures the efficiency of public bus transportation route whereas the GIS is applied for social priority based route planning. As a result the study helps the bus routes with low efficiency to efficiently change the routes of buses [13].

In “Modified particle swarm optimization for multimodal functions and its application” Kushwaha et al. proposed a modified variant of particle swarm optimization (PSO) algorithm to improve the performance of the original PSO algorithm. In the proposed algorithm the particles not only learn from their personal best position but also learn from their neighbours. Here each neighbour is weighted according to score which is calculated by PageRank algorithm. The scale-free network is also proposed for the interaction among particles in the population. The proposed algorithm is compared with nine PSO variants on 17 benchmark functions. It has been concluded that the proposed modifications in PSO enhances the performance of PSO in terms of convergence and solution quantity [8].

In “An effective analysis of deep learning based approaches for audio based feature extraction and its visualization” Biswas et al. conducted a critical analysis of the effectiveness of various state-of-the-art deep neural networks in visualizing music. The aim of the study was to capture the types of features such as mood of a song, lyrical aspect and distant temporal relations from audio tracks. The features were extracted by implementing auto encoders and genre classifiers then mapped to the parameters that drive audio visualization. The methodology enables the visualization to be responsive to the music and also provide unique visual experiences across different songs [2].

In “Cyberbullying Detection on Social Multimedia using Soft Computing Techniques: A Meta-Analysis” Kumar et al. studied the significance of using soft computing techniques for cyberbullying detection on social multimedia. Study shows that conventional methods incapable to deal with the mounting velocity, volume and variety of data generated by the social media. Therefore A systematic literature review was conducted on the use of soft computing

techniques to detect cyberbullying activities across various social media domains to understand the theory, research and practice trends within the domain. After conducting the review it was observed that the use of soft computing techniques for cyberbullying detection provided the intelligent analytic paradigm essential for predicting bullying behaviours and activities on both textual and non-textual social media [6].

In “Multimedia detection algorithm of malicious nodes in intelligent grid based on fuzzy logic” Gao et al. proposed a fuzzy logic trust model (FLTM) to detect the malicious nodes in the smart grid. FLTM model makes full use of fuzzy logic system to deal with uncertainties, estimates the overall trust value of nodes, and then detects malicious nodes. The proposed model takes the direct trust, indirect trust and past trust as the input, take its output as the trust value of the node. The experimentation shows that the FLTM model improves the detection ratio of the malicious nodes [4].

In “Multimedia document image retrieval based on regional correlation fusion texture feature FDPC” Zeng et al. presented a new clustering algorithm called Fast density peak clustering (FDPC) in order to realize the retrieval efficiency and detection precision of digital library collection resources. This work aimed at image information retrieval of digital library to organically integrate colour features and correlation for different areas of image information. This procedure was followed to obtain image feature extraction strategy of regional correlation integration. The image is classified with the help of density peak clustering (DPC) and its performance is improved by using a dynamic truncation distance mode. The experimental results obtained shows that the proposed algorithm has higher retrieval efficiency and retrieval accuracy when validated on standard test library (Corel) [17].

In “Multimedia based fast face recognition algorithm of speed up robust features” Zhang et al. improved the original speed up robust feature (SURF), the face recognition algorithm with the K-mean clustering technology. Authors found the defects of the original SURF like the direction distribution, descriptor vector generation and interest point matching which are improved by introducing the K-mean clustering idea. The results of the experiment on FERET and Yale face database show that the proposed algorithm has higher recognition rate and efficiency than other face recognition techniques [18].

In “Diversifying personalized mobile multimedia application recommendations through the Latent Dirichlet Allocation and clustering optimization” Raja et al. presented the Diversifying Personalized Mobile Multimedia Application Recommendation (DIPMMAR) for the app stores to recommend the desired applications to the users. The model analyses the fusion of user ratings, review texts, application description and application popularity. The hidden relationships among the different user reviews and application descriptions are analysed by applying the Latent Dirichlet Allocation (LDA) based topic model. Later the DIPMMAR is used to compute the user preferences of the local popularity score for each user by applying the clustering optimization of K-means clustering and PSO algorithm. After ranking the applications based on the preferred inherent features of applications, the DIPMMAR approach applies the reranking procedure to ensure the relevance as well as diversity in the recommendation list related to the preferred sub-categories for each user. The extensive evaluation results reveal that the proposed mobile application recommender system significantly outperforms the existing recommender system by 10.8% of higher recall and 0.8% of higher diversity [10].

In “A Multimedia Image Edge Extraction Algorithm Based on Flexible Representation of Quantum” Lu et al. proposed an algorithm to solve the real time problem of edge extraction and improve image edge continuity, based on quantum flexible representation (flexible representation of Quantum, FRQ) in combination with features of Sobel operator. At first

the authors used FRQ to quantify digital image as quantum image and stored in the quantum register. Secondly by the translation transformation of the X and Y direction of the FRQ image, the relative quanta of the neighbouring pixels of the whole image is obtained. For the image edge extraction different types of pixels are judged according to Sobel gradient. It is proved through experimental result that proposed algorithm has quicker edge extraction speed than current edge extraction algorithm [9].

In “Rumour veracity detection on twitter using particle swarm optimized shallow classifiers” Kumar et al. proposed a particle swarm algorithm (PSO) using the optimal feature selection method for rumour veracity classification problem. A total of nearly 14 k tweets pertaining to the recent mob lynching fuelled by rumours on suspected child-lifters in the Indian sub-continent were analysed based on accuracy as the performance metric of the classifier. At First the authors implemented five classification algorithms for rumour veracity detection: SVM, DT, K-NN, NB, NN and PSO for the optimal feature selection. The empirical analysis validates that the use of PSO for feature selection in rumour veracity classification task selects the features with the highest importance/influence on the target variable, from the existing set of features. It is proved the this automatic prediction of rumour veracity will debunk false rumours and mitigate their spread and impact [7].

In “Sentiment analysis of multimodal twitter data” Kumar et al. proposed a multimodal sentiment analysis model to determine the sentiment polarity and score for any incoming tweet i.e. textual, image or info-graphic. The author conducted the image sentiment scoring using SentiBank and SentiStrength scoring for Regions with convolutional neural network(R-CNN). The text scoring is conducted with a novel context aware hybrid lexicon and machine learning technique and the multimodal sentiment scoring is done by separating text from image using an optical character recognizer. At last the resultant scores from text and image module are combined to produce aggregate sentiment score for the multimodal tweet. The authors found that the performance results are motivating and improves the generic sentiment analysis task [5].

2 Conclusion

We hope these contributions will be of interest and value to readers from a wide range of subject areas and form a reference for future development. The experience of serving as Guest Editors for this special issue on “Multimedia Tools and Applications” is excellent in form of analysing the topics relevant to the theory and practice of intelligent computational techniques, models and empirical analysis in the multimedia development area.

References

1. Bhatt CA, Kankanhalli MS (2011) Multimedia data mining: state of the art and challenges. *Multimed Tools Appl* 51(1):35–76
2. Biswas R, Ghattamaraju N (2018) An effective analysis of deep learning based approaches for audio based feature extraction and its visualization. *Multimed Tools Appl*. <https://doi.org/10.1007/s11042-018-6706-x>
3. Gao L, Song J, Liu X, Shao J, Liu J, Shao J (2017) Learning in high-dimensional multimedia data: the state of the art. *Multimedia Systems* 23(3):303–313
4. Gao M, Wu Y, Nan J, Cui S (2019) Multimedia detection algorithm of malicious nodes in intelligent grid based on fuzzy logic. *Multimed Tools Appl*. <https://doi.org/10.1007/s11042-019-7191-6>

5. Kumar A, Garg G (2019) Sentiment analysis of multimodal twitter data. *Multimed Tools Appl.* <https://doi.org/10.1007/s11042-019-7390-1>
6. Kumar A, Sachdeva N (2019) Cyberbullying detection on social multimedia using soft computing techniques: a meta-analysis. *Multimed Tools Appl.* <https://doi.org/10.1007/s11042-019-7234-z>
7. Kumar A, Sangwan SR, Nayyar A (2019) Rumour veracity detection on twitter using particle swarm optimized shallow classifiers. *Multimed Tools Appl.* <https://doi.org/10.1007/s11042-019-7398-6>
8. Kushwaha N, Pant M (2018) Modified particle swarm optimization for multimodal functions and its application. *Multimed Tools Appl.* <https://doi.org/10.1007/s11042-018-6324-7>
9. Lu Z, Wang X, Shang J, Luo Z, Sun C, Wu G (2019) A multimedia image edge extraction algorithm based on flexible representation of quantum. *Multimed Tools Appl.* <https://doi.org/10.1007/s11042-019-7173-8>
10. Raja DK, Pushpa S (2019) Diversifying personalized mobile multimedia application recommendations through the Latent Dirichlet Allocation and clustering optimization. *Multimed Tools Appl.* <https://doi.org/10.1007/s11042-019-7190-7>
11. Shamsolmoali P, Jain DK, Zareapoor M, Yang J, Alam MA (2018) High-dimensional multimedia classification using deep CNN and extended residual units. *Multimed Tools Appl.* <https://doi.org/10.1007/s11042-018-6146-7>
12. Shamsolmoali P, Zareapoor M, Jain DK, Jain VK, Yang J (2018) Deep convolution network for surveillance records super-resolution. *Multimed Tools Appl.* <https://doi.org/10.1007/s11042-018-5915-7>
13. Singh P, Singh AK, Singh P, Kumari S, Sangaiah AK (2018) Multimodal data modeling for efficiency assessment of social priority based urban bus route transportation system using GIS and data envelopment analysis. *Multimed Tools Appl.* <https://doi.org/10.1007/s11042-018-6147-6>
14. Sun Q, Yang S, Sun C, Yang W (2018) Exploiting aggregate channel features for urine sediment detection. *Multimed Tools Appl.* <https://doi.org/10.1007/s11042-018-6241-9>
15. Yang Z, Huang P, Wan M, Zhan T, Zhang F, Luo L (2018) Discriminant maximum margin projections for face recognition. *Multimed Tools Appl.* <https://doi.org/10.1007/s11042-018-6242-8>
16. Zareapoor M, Yang J, Jain DK, Shamsolmoali P, Jain N, Kant S (2018) Deep semantic preserving hashing for large scale image retrieval. *Multimed Tools Appl.* <https://doi.org/10.1007/s11042-018-5970-0>
17. Zeng F, Xu J (2019) Multimedia document image retrieval based on regional correlation fusion texture feature FDPC. *Multimed Tools Appl.* <https://doi.org/10.1007/s11042-019-7161-z>
18. Zhang Y, Fang H (2019) Multimedia based fast face recognition algorithm of speed up robust features. *Multimed Tools Appl.* <https://doi.org/10.1007/s11042-019-7209-0>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.