

Learning high-dimensional multimedia data

Xiaofeng Zhu¹ · Zhi Jin² · Rongrong Ji³

Published online: 25 July 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract The internet revolution has made information acquisition easy and cheap and has been producing massive high-dimensional multimedia data, including text, audio, images, animation, video, etc. High-dimensional multimedia data bring new opportunities to modern society and challenges to researchers of the multimedia domain as well. The goal of this special issue is bridging the gap between machine learning methods and the real requirement of high-dimensional multimedia domain, aiming at gaining insight into the relationship between the current multimedia and the past ones, and also accurately predicting the future trends of multimedia data. Specifically, this special issue targeted the most recent technical progresses on learning techniques for high-dimensional multimedia data

Keywords High-dimensional data · Multimedia · Machine learning · Editorial

The internet revolution has made information acquisition easy and cheap and has been producing massive high-dimensional multimedia data, including text, audio,

images, animation, video, etc. Both massive sample size and high feature dimensionality characterize contemporary multimedia analysis problems, such as tens of thousands of features are extracted from multimedia and hundreds of millions of descriptors are obtained from massive multimedia datasets. Moreover, once interactions have been considered, both the sample size and the feature dimensionality grows even more quickly. Thanks to that multimedia data are often collected over different platforms or locations, massive data naturally generate issues with heterogeneity, measurement errors, and experimental variations, while the high feature dimensionality often encounters issues with computational cost, algorithmic stability, spurious correlations, and so on. Hence, high-dimensional multimedia data bring new opportunities to modern society and challenges to researchers of the multimedia domain as well.

The goal of this special issue is bridging the gap between machine learning methods and the real requirement of high-dimensional multimedia domain, aiming at gaining insight into the relationship between the current multimedia and the past ones, and also accurately predicting the future trends of multimedia data. Specifically, this special issue targeted the most recent technical progresses on learning techniques for high-dimensional multimedia data, including classification [1], segmentation [2, 4], feature selection [1, 5, 8], deep learning [5], image saliency detection [7], and many others [3, 6], in many kinds of learning-based applications, including image processing sequences [1, 2, 7, 8], text processing [1, 4, 6], system applications [6]. The topics of the special issue are interesting, so in total, this special issue have received 24 submissions from at least 20 different research departments over the world. After at least two rounds of reviews, we finally accepted eight papers for publication. We summarize the introduction of accepted papers as follows:

✉ Xiaofeng Zhu
seanzhuxf@gmail.com

Zhi Jin
zhi.jin.deng@gmail.com

Rongrong Ji
rrji@xmu.edu.cn

¹ Guangxi Normal University, Guilin, China

² Peking University, Beijing, China

³ Xiamen University, Xiamen, China

The paper by Cheng et al. [1] proposed a sparse feature selection algorithm for high-dimensional data classification by integrating subspace learning with feature selection in a unified framework. The authors selected the relevant features with the least square loss function and $\ell_{2,1}$ -norm regularization term to yield the minimum representation error between the prediction and the labels. The authors also proposed an efficient optimization algorithm to solve the resulting objective function. The experiments on benchmark datasets showed that the proposed approach was more effective and robust than existing feature selection algorithms.

The paper by Du et al. [2] proposed an accurate glasses detection algorithm for in-plane rotated faces. Distinct from the previous face detection techniques, this paper extended upright face detector to different rotated detectors with a new set of harr-like features so that face detection with in-plane rotation was achieved accurately. Based on the normalized upright face, the authors first determined the eye position which made the candidate region of the glasses much simple, and then the nosepiece of the glasses was selected as the target leading to the accurate results. Different from the previous methods, their proposed method did not detect the glasses directly, but located the eye positions to reduce the candidate region and selected the fixed nosepiece part as the target to improve detection accuracy. The experimental results on the CMU-MIT, BioID, and CAS-PEAL-R1 face databases demonstrated the algorithms for face, eye and glasses detection yielded accurate results and maintained the speed advantage.

Under the background of many paradigms exist and are widely used in the context of machine learning, most of them suffer from the ‘curse of dimensionality’. The paper by Gao et al. [3] proposed a survey over feature transformation, feature selection and feature encoding, which were widely proposed to fight the consequences of the curse of dimensionality. The authors first considered the main challenges for the high-dimensional data and then surveyed some existing techniques dealing with the problem. Furthermore, the authors briefly introduced some recent progresses of effective learning algorithms.

The paper by Lin et al. [4] proposed to compare document images in multilingual corpus, which was composed of character segmentation, feature extraction and similarity measure. The paper applied projection and self-adaptive threshold to analyze the layout and then segment the text line by horizontal projection. Then, English, Chinese and Japanese were recognized by different methods based on the distribution and ratios of text line. Finally, character segmentation with different languages was done using different strategies. In feature extraction and similarity measure, four features were given for coarse measurement, and then a template was set up. Based on the templates, a fast

template matching method based on coarse-to-fine strategy and bit memory was presented for precise matching. The experimental results demonstrated that the proposed method could handle multilingual document images of different resolutions and font sizes with high precision and speed.

The paper by Nie et al. [5] proposed a feature extraction model based on convolutional neural networks (CNN). First, the paper extracted a set of 2D images from 3D model to represent each 3D object. Second, a single CNN layer learnt low-level features which were then given as inputs to multiple recursive neural networks (RNN) to compose higher order features. Finally, nearest neighbor was used to compute the similarity between different 3D models to handle the retrieval problem. The results demonstrated the superiority of the proposed method.

This paper by Uljarevic et al. [6] proposed a system of information transfer within Microsoft Word documents, by utilizing the JPEG images dataset to experiment then presented the steganographic tool StegApp 1.1.0. In particular, the novel tool utilized the domain transformation technique with parameters obtained by the discrete cosine transforms (DCT) that is based on the transformation of the JPEG image. Then compared with the commercial SteganPEG tool, the proposed tool can get better results.

This paper by Zhao et al. [7] proposed an image saliency detection method by means of two effective maps. They used the new distinction measure and background pixel distribution approximation method, by weighting the combination of the color contrast and spatial distance criteria in accordance with the previous works and utilized patches sampled near the image borders, respectively. And ultimately joint the distinction map and the background probability map into a hierarchical framework to generate the final saliency map result. In several recent work experiments, the presented method has achieved the best results.

The paper by Zhu et al. [8] proposed a self-representation graph feature selection method for classification. The objective function includes a self-representation loss function, a graph regularization term and an $\ell_{2,1}$ -norm regularization term. Aim at effectively selecting representative features and ensuring the robustness to outliers, the proposed self-representation loss function pushes to represent each feature with a linear combination of its relevant features, which is different from traditional least square loss function focuses on achieving the minimal regression error between the class labels and their corresponding predictions. The graph regularization terms include the relationship between samples and between features. The sample–sample ratio reflects the similarity between two samples and preserves the relation into the coefficient matrix, while the feature–feature relation reflects the similarity between two features and preserves the relation into the coefficient matrix. The

$\ell_{2,1}$ -norm regularization term is used to conduct feature selection, which satisfying the characteristics above mentioned. The experimental results showed that this method has a better performance comparing with state-of-the-art methods.

References

1. Cheng, D., Zhang, S., Liu, X., Sun, K., Zong, M.: Feature selection by combining subspace learning with sparse representation. doi:[10.1007/s00530-015-0487-0](https://doi.org/10.1007/s00530-015-0487-0)
2. Du, S., Liu, J., Liu, Y., Zhang, X., Xue, J.: Precise glasses detection algorithm for face with in-plane rotation. doi:[10.1007/s00530-015-0483-4](https://doi.org/10.1007/s00530-015-0483-4)
3. Gao, L., Song, J., Liu, X., Shao, J., Liu, J., Shao, J.: Learning in high-dimensional multimedia data: the state of the art. doi:[10.1007/s00530-015-0494-1](https://doi.org/10.1007/s00530-015-0494-1)
4. Lin, Y., Li, Y., Song, Y., Wang, F.: Fast document image comparison in multilingual corpus without OCR. doi:[10.1007/s00530-015-0484-3](https://doi.org/10.1007/s00530-015-0484-3)
5. Nie, W., Cao, Q., Liu, A., Su, Y.: Convolutional deep learning for 3D object retrieval. doi:[10.1007/s00530-015-0485-2](https://doi.org/10.1007/s00530-015-0485-2)
6. Uljarevic, D., Veinovic, M., Kunjadic, G., Tepsic, D.: A new way of covert communication by steganography via JPEG images within a Microsoft Word document. doi:[10.1007/s00530-015-0492-3](https://doi.org/10.1007/s00530-015-0492-3)
7. Zhao, S., Shen, J., Li, F.: A hierarchical visual saliency detection method by combining distinction and background probability maps. doi:[10.1007/s00530-015-0490-5](https://doi.org/10.1007/s00530-015-0490-5)
8. Zhu, Y., Liang, Z., Liu, X., Sun, K.: Self-representation graph feature selection method for classification. doi:[10.1007/s00530-015-0486-1](https://doi.org/10.1007/s00530-015-0486-1)