Neurocomputing ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Editorial

Learning for visual semantic understanding in big data

ARTICLE INFO

Keywords:
Visual semantic understanding
Big data
Editorial

In recent years, we have witnessed the popularity of big multimedia data in both online websites, such as Flickr, Twitter, YouTube and Facebook, and local data centers. Under these circumstances, the visual semantic understanding meets its new challenges when dealing with such big data. Extensive research efforts have been dedicated to learning the semantic of big visual data. However effective tools to manipulate these data are still at their infancy. This special issue targets the most recent technical progresses on learning techniques for visual semantic understanding in big data, such as the content/concept-based retrieval/recognition, action recognition, event detection and semantic analysis. This special issue also targets on novel learning-based applications, such as recommendation, summarization and evaluation. The primary objective of this special issue fosters focused attention on the latest research progress in this interesting area.

The topics of the special issue are interesting. In total, 37 submissions have been received. After two rounds of reviews, 20 papers were accepted for publication. The introductions of the accepted papers are summarized as follows:

The paper by Gu et al. [1] proposed a new spatial–spectral joint sparsity algorithm for target detection in hyperspectral imagery (HSI). The proposed algorithm embeds the sparse representation (SR) into the conventionally subspace target detector in hyperspectral images. This algorithm is based on such an idea that a pixel in HSI lies in a low-dimensional subspace and can be represented as a sparse linear combination of the training samples. Substituting SR for the conventional subspace method, a sparse matched subspace detector (SMSD) is developed. Moreover, 3D discrete wavelet transform (DWT) and independent component analysis (ICA) are exploited to extract the spatial and spectral distribution in the hyperspectral imagery and capture the joint spatial–spectral sparsity structure. By integrating the structured sparsity and the SMSD, the proposed algorithm is able to carry out target detection task in the hyperspectral images. Experiments are conducted on real hyperspectral data. The experimental results show that the proposed algorithm outperforms both the conventional matched subspace detector (MSD) and the state-of-the-arts sparse detection algorithm.

The paper by Feng et al. [2] deals with the issues of software reliability allocation, which have been discussed from many aspects, such as mathematical models and solutions to maximize the reliability. However, most of this research has concentrated on single software. The goal of this work is to investigate the possibility of solving multi-software reliability allocation in multimedia systems with budget constraints. For this purpose, they first developed an architecture-based multi-software Budget-Constrained Reliability-maximization model. In addition, they introduced Dempster–Shafer theory to identify the relative reliability weights of each element in the proposed model and present a searching algorithm based on differential evolution and encoding repair. Finally, contrast experiments are illustrated to demonstrate the proposed methods.

The paper by Ji et al. [3] proposed to combine multiple visual features together in 3D model retrieval. An efficient Semisupervised Multiple Feature Fusion (SMFF) method was proposed for view-based 3D model retrieval. Specifically, they first extracted multiple visual features to describe both the local and global appearance characteristic of multiple 2D projected images that are generated from 3D models. Then, SMFF was adopted to learn a more compact and discriminative low-dimensional feature representation via multiple feature fusion using both the labeled and unlabeled 3D models. Once the low-dimensional features have been learned, many existing methods such as SVM and KNN can be used in the subsequent retrieval phase. Moreover, an out-of-sample extension of SMFF was provided to calculate the low-dimensional features for the newly added 3D models in linear time. Experiments on two public 3D model datasets demonstrate that using such a learned feature representation can significantly improve the performance of 3D model retrieval and the proposed method outperforms the other competitors.

The paper by Li et al. [4] proposed a cloud image detection method based on SVM vector machine to remove thick cloud data to reduce the amount of data to improve the efficiency of the data. Firstly, the satellite remote sensing image was divide into small blocks, and the brightness characteristics of the sub-block image was extracted to accomplish the preliminary detection. Then the average gradient and the angle of the gray level co-occurrence matrix second-order moment for sub-block image based on the texture features of the sub-block image was calculated as the basic of SVM victor machine. The sub-block cloud image is used as learning samples of the SVM classifier that has brightness characteristics,

http://dx.doi.org/10.1016/j.neucom.2015.05.023

and the classification model is obtained from the training of the SVM classifier to realize a detail classification of the cloud image detect based on the SVM victor machine. Finally, they conducted experiments on cloud image detection method based on SVM vector machine. Experiment results demonstrate detection accuracy of the method proposed could reach above 90%.

The paper by Zhu et al. [5] proposed a new Multi-view Multi-sparsity Kernel Reconstruction (MMKR in short) model for conducting multi-class image classification. Specifically, the proposed MMKR first represents each test image by part representative training images on a Reproducing Kernel Hilbert Space (RKHS), and then outputs the coefficient between each test image and training images, which is then used for devising a classification rule to conduct the classification task of test images. Experimental results on two real datasets indicates that the new model helps enhance the performances of multi-class image classification, outperforming the state-of-the-art methods.

The paper by Shang et al. [6] proposed a dimension reduction method called Meta object-Group Component (MGC) to tackle this problem. MGC aims at discovering the hidden relations of objects by examining the correlations between dimensions in the BoW features and maximizing the relations of the members in a meta object-group. By exchanging message passing between object-groups, meta object-groups are identified for a dataset. A meta objectgroup does not only contain visually similar objects, but also includes objects that are likely to co-occur with each other. As the meta object groups are obtained, group-specific dimension reduction is performed to obtain denser representations for efficient retrieval. The framework was evaluated on the NUS-Wide image dataset with approximately 270,000 images represented by BoW features, and demonstrate its advantage over existing method.

The paper by Cui et al. [7] deals with the problem of “curse of dimensionality“ in the hyper dimensionality of BOF vectors, which usually leads to huge computation and storage complexity. To create a compact and discriminative BOF representations, in this paper they proposed a novel unsupervised dimensionality reduction framework for the histogram vectors. First, they built the dissimilarity matrix between every histogram pairs, and then performed Multidimensional Scaling technique to obtain a low-dimensional Euclidean embedding of the original BOF while simultaneously preserving the inherent neighborhood structure. The widely used metrics for measuring dissimilarity, including distance and kernel, were investigated to build the dissimilarity matrix as the input of our dimensionality reduction model. Extensive experiment results show that a very low dimension is sufficient for the learning tasks using BOF or SPM without losing the classification accuracy. Comparatively, the state-of-the-art methods can hardly achieve high accuracy on the very low dimension. Furthermore, a compact representation of BOFs can improve the classification accuracy compared with the original BOF. Finally, they also showed that the compact representation is effective for image retrieval tasks.

The paper by Feng et al. [8] deals with the problem that the standard NMF is sensitive to the noise and outliers. To overcome this problem, they proposed a noise robust NMF method named as Locally Weighted Sparse Graph regularized Nonnegative Matrix Factorization (LWSG_NMF). Since many real-world noises can be broadly decomposed into the dense Gaussian random noise and the sparse block noise, they proposed a sparse noise assumption. Based on this assumption, they reformulated the empirical likelihood term of the standard NMF by explicitly imposing a sparse noise term. Meanwhile, a locally weighted sparse graph regularization term is also incorporated in the proposed model to exploit the local geometric structure information of data. Different from the other existing graph-based methods, they took effect of noise into account in learning the graph regularization term. An iterative optimization method was also proposed to solve the objective function of LWSG_NMF. Extensive experiments on three public benchmark datasets demonstrated the robustness and the effectiveness of the proposed method for human face recognition and handwritten digital recognition in the presence of noise.

The paper by Li et al. [9] proposed a new traffic sign detection method by integrating color invariants based image segmentation and pyramid histogram of oriented gradients (PHOG) features based shape matching. Given the target image, they first extracted its color invariants in Gaussian color model, and then segmented the image into different regions to get the candidate regions of interests (ROIs) by clustering on the color invariants. Next, PHOG was adopted to represent the shape features of ROIs and support vector machine is used to identify the traffic signs. The traditional PHOG is sensitive to the cluttered background of traffic sign when extracting the object contour. To boost the discriminative power of PHOG, they proposed introducing Chromatic-edge to enhance object contour while suppress the noises. Extensive experiments demonstrate that the method can robustly detect traffic signs under varying weather, shadow, occlusion and complex background conditions.

The paper by Hebboul et al. [10] proposed an Incremental Neural Network for Classification and Clustering (INNCC). The main advantages of this neural network are the linkage between data topology preservation and classes representation by using the cluster posterior probabilities of classes. It is a constructive model without prior conditions such as a suitable number of nodes. A new neuron is inserted when new data are not represented by existing neurons. In training step, both supervised and unsupervised learning were used. The training dataset contains few samples with class labels and several unlabeled ones. The Support Vector Machines (SVM) operates in the training step to assess the INNCC classification result. The proposed approach has been tested on synthetic and real datasets. Obtained results are very promising.

The paper by Lai et al. [11] tried to explore the behaviors of DNS lookup by mining DNS logs from three primary DNS servers in a large university campus network in China. The used dataset is made up of two parts, namely DNS query logs and messages received or sent by DNS servers. Firstly, through analyzing these DNS query logs, the overall trend of users' surfing was understood. For dealing with huge DNS dataset, they introduced an algorithm, termed as *DNSReduce*, which can be used to dig out top 10 client IP addresses and top 10 destination domain names efficiently. Moreover, they made comparative analysis of lookup behavior between wired and wireless users. Secondly, with messages received or send by DNS servers they can find these DNS servers behaviors, i.e., TTLs, equivalent answers and are able to accurately identify domain names with dynamic IP addresses. They provided different and specific visualization techniques for presenting these analysis results and show these techniques are very useful for understanding user behaviors, analyzing security events and characterizing overall tendency in campus network management.

The paper by Zhang et al. [12] presents a novel learning-free image classification algorithm under the framework of Naive-Bayes Nearest-Neighbor (NBNN) and collaborative representation, where nonnegative sparse coding, low-rank matrix recovery and collaborative representation were employed to obtain more robust and discriminative representation. Firstly, instead of general sparse coding, nonnegative sparse coding combined with max pooling was introduced to further reduce information loss. Secondly, they used the low-rank matrix recovery technique to decompose the training data of the same class into a discriminative low rank matrix, which preserves more structurally correlated information. As for test images, a low-rank projection matrix was also learned to remove possible image corruptions. Finally, the classification process was implemented by simply comparing the responses over the different bases. Experimental results on several