Neurocomputing 000 (2017) 1–5

Contents lists available at ScienceDirect

Neurocomputing

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

Editorial

Editorial learning for multimodal data

Multimodal data are often used to describe objectives more detail than the representation of single modal data. For example, in real applications, we usually extract all kinds of representations of the samples from its different sources, including audio, video, text, and so on [2,6]. Different sources can be extracted different kinds of features and thus contain various semantic. How to connect the gap among multiple modals or how to fuse the multimodal representations is a very popular and challenging topic in machine learning, data mining and statistics [1,3]. The previous study focused on fusing multimodal representations into a single modal representation and then building models by existing single modal learning methods. This often encounters for the issue of the "curse of dimensionality" and the issue of heterogeneity [4,5]. Hence, it will be interesting and challengeable to learn multimodal data for addressing these issues.

The main aim of this special issue is bridging the gap between multimodal learning theory and single model learning models. Specifically, this special issue targeted the most recent technical progresses on learning techniques for multimodal data, including Biclustering [28], deep learning [19], feature selection [20,22,25], convolution neural network [9,10], multi-task learning [10], cross-modal learning [7], Hop-field neural network [7], deep learning [14], probabilistic model [12], hierarchical clustering [14], hierarchical sparse classification [14], feature fusion [15], multiple kernel learning [15], Hashing [17,18], and many others, in many kinds of learning-based applications, including text mining [7,8], 3D shape retrieval [9], multimodal trajectory data [13], privacy protection [13], auroral event classification[11], social media retrieval [16], text mining [16], emotion recognition [19] 3D reconstruction [21], 3D spatial geometric model [29], phrase recognition [29], web-aided data repairing system [31], semantic segmentation [30], Chronic Gepatitis B virus assessment [27], landmark detection on human body [26], distributed multi-modal multimedia analysis [23], multimodal joint sentiment topic model [24], and so on.

The topics of the special issue are interesting, so in total, this special issue have received 61 submissions from at least 30 different research departments over the world. After at least two rounds of reviews, we finally accepted 25 papers for publication. We summarize the introduction of accepted papers as follows:

The paper by Yue et al. [7] proposed a new approach for sentence ordering with multimodal content extracted from multiple modalities. The proposed method consists of unifying the existing factors by considering the most frequent orders temporal information, and topical relevance between local themes during overall ordering process, to solve sentence ordering problem. Moreover, the cross-document sentence ordering was treated as a combinatorial optimization problem and solved with Continuous Hop-field neural network (CHNN). The experimental results demonstrated that the proposed model was effective, compared to the state-of-the-art methods.

The paper by Ju and Tao [8] proposed a general framework to measure the strengths of the relationship among different users in Wechat Friends Circle, which considers not only the similarity of users' profile information but also the interaction among users. The motivation of the proposed method is that the previous research did not consider strengths of the relationship among different users in the Wechat Friends Circle as the users can only see a common friend's comments or point praise. They conducted a set of experiments on Friends Circle dataset, from which they learned that the proposed framework was efficient and promising on the improvement of the performances of relationship strength calculation.

The paper by Tabia and Laga [9] proposed a new framework for 3D shape retrieval using queries of different modalities, which include 3D models, images and sketches. To address the problem of that different modalities had different representations and thus lie in different spaces, this paper employed a Convolutional Neural Networks (CNN) to embed all these entities into a common space. Specially, this paper first used a kernel function computed from 3D shape similarity, then built a target space in which wild images and sketches could be projected via two different CNNs. Finally, matching could be performed in the common target space between same entities (sketch-sketch, image-image and 3D shape-3D shape) and more importantly across different entities (sketch-image, sketch-3D shape and image-3D shape). The experimental results demonstrated that the performance of proposed method outperforms the state-of-the-art.

The paper by Tian et al. [10] presented a multi-task learning approach on convolution neural network for object localization. The model consisted of 3 modules, respectively extracting shared features, generating low-level features, and fusing different levels information. First, shared features were extracted from a well-designed network structure to give candidate object classification and location, and then edge map extraction module was trained to extract image edge information, and lastly different level features were combined and evaluated to give accurate object location. Experiments showed that proposed algorithm effectively and efficiently improved performance.

The paper by Zhang et al. [11] studied the issue of different types of auroral events, especially for aurora classification.

http://dx.doi.org/10.1016/j.neucom.2017.02.080
0925-2312/

To do this, the authors proposed an auroral event representation approach. Specifically, 3D steerable filters were used to generate space-time oriented energies, which could simultaneously capture texture and motion of the auroral event. Then, n-ary fusion operator was used to combine oriented energies together to generate a unified representation, which also made the representation had the property of gray-scale invariance. In order to capture roughly global shape information, a block partition scheme was used to further represent the auroral event. Finally, the performance of the representation was evaluated by the auroral event classification. The superior experimental results demonstrated the effective of proposed auroral event representation method.

The paper by Zhu et al. [12] defined a set of probabilistic spatial metrics and propose a travel-time threshold and a transfer-cost threshold convenient route planning queries to overcome two challenges in the study: (1) use multi-source human tracking data to model probabilistic transfer cost between different bus/subway lines practically, and (2) compute convenient bus/subway routes efficiently. The paper integrated multi-source human-mobility tracking data and location based social media data, including spatial, temporal, and textual data, to make human-mobility prediction and over-crowded station detection. And the authors developed a series of optimization techniques to enhance the query efficiency. They conducted extensive experiments to verify the performance of the proposed algorithms and found out that the proposed algorithms achieved better results.

The paper by Sui and Li [13] proposed a km-anonymity-rho-uncertainty privacy protection model to protect the privacy information in integrating transaction data with trajectory data in a tree-structured data model. The authors presented a tree-structure data model to describe integration data and then utilized the km-anonymity-rho-uncertainty model to defend identity disclosure and sensitive item disclosure. And that partition and copy could not only eliminate identity disclosure but also preserve other sequential patterns. The paper first demonstrated a privacy leakage model caused by integration of multimodal datasets, where integrated data are modeled as a tree. To address the identity disclosure of trajectories, the authors partitioned location sequences to meet privacy demands, and copied locations to offset information loss caused by partition; then, to deal with the sensitive item disclosure of transactions, they used suppression technique to eliminate sensitive association rules. Finally, the experimental results showed that the proposed algorithm achieved better effectiveness and efficiency, compared to the the state-of-the-art methods.

The paper by Wang et al. [14] proposed a hierarchical sparse representation based classification method by augmenting the single-layer sparse representation into the hierarchical representation with a deep dictionary to improve the classification accuracy and robustness. Specifically, the authors first divided the features from all training samples that were first divided into several groups according to their training labels. Then they employed hierarchical clustering in each group and combined them to form a deep dictionary such that the root layer includes only a certain amount of the most representative exemplars and while the subsequent layers focus on characterizing the remaining individual information across different groups. Furthermore, they used the layer-after-layer residuals to encode the variation patterns across individuals in different scales. Given the deep dictionary, the paper then presented a hierarchical sparse representation based classification method to determine the label for each of the new test sample by iteratively representing its primary part with the exemplars in different clinical groups and the remaining parts by the variation patterns encoded in different layers. Experiments on multiple features datasets showed better results compared with the state-of-the-art classification methods.

The paper by He et al. [15] proposed a method named Score-Distribution Multiple Kernel Learning (SD-MKL) for image classification to address the feature fusion problem. The proposed algorithm mainly contained two stages in off-line part: (1) independent data was used to construct reference curves according to classes and feature type; (2) samples and corresponding score-distribution weights were put into multikernel support vector machine (MKSVM) to learn feature weights. The experimental results on two datasets showed the proposed SD-MKL perform better than MKL, and thus the score distribution could be considered as a useful information in classification.

The paper by Gao et al. [16] proposed an interval-at-a-time (IAAT) framework to provide a one-sizefits-all solution to social media retrieval with spatial, temporal and social constraints. The main advantages of this framework are its versatility and effectiveness. The authors had also proposed another algorithmic frameworks based on inverted index to support various search scenarios in social media retrieval which was based on top-k aggregation that conducted dimension reduction for the spatial and social attributes. The experimental results on Twitter dataset with up to 100 million geo-tweets and 20 million users showed that IAAT was a more efficient framework than the top-k aggregation when multiple retrieval constraints were considered.

The paper by Li et al. [17] proposed a method named Kernel based Latent Semantic Sparse Hashing (KLSSH) to consider the semantic correlation between multi-modal representations. The proposed KLSSH method first employed Sparse Coding for obtaining primary latent features of image and Matrix Factorization for generate semantic features of text concepts to learn latent semantic features in a high level abstraction space. Next, it mapped the latent semantic feature to compact binary codes using kernel method. Meanwhile it reduced the quantization loss obviously. Extensive experiments conducted on three real-world large-scale datasets demonstrated the effectiveness and efficiency of the proposed model compared to state-of-the-art hashing algorithms.

The paper by Chen et al. [18] proposed an adaptive discrete cyclic coordinate descent (ACC) method to effectively solve discrete optimization problem. The proposed method could single out which discrete binary code was worth to be updated. In updating process, an indicating function was brought in to determine to stop in a proper time. Therefore, binary codes learning process could be more efficient and effective. The experimental results on image and multimedia databases: CIFAR-10, NUS-WIDE and MIRFLickr-25k showed that the proposed method achieved speed-up over compared to the state-of-the-art methods.

The paper by Barros et al. [19] proposed a neurocomputational model that learned to emotional expressions and modulate emotion recognition. The authors used a deep neural architecture to learn emotional attention. Their learning architecture was based on Convolutional Neural Networks, using the filtering capability of the convolution layers to learn the location of emotional expressions conveyed by two visual cues: face expressions and body movements. They utilized probability distributions that indicated the focus on attention to train the model. Experimental results showed that the proposed model learned how to filter multi-cue information with neurons that react to face expressions and others that react to body movements emerge without been explicitly defined.

The paper by Cheng et al. [20] proposed a feature selection method that simultaneously embeds the low-rank constraint, sparse representation, global and local structure learning into a unified framework. The proposed method first utilized the conventional regression function (i.e., the least square loss function) to form a novel regression framework by introducing a low-rank constraint and a relaxation term, then employed an $\ell_{2,1}$-norm regularization term to filter out the redundant and irrelative features.