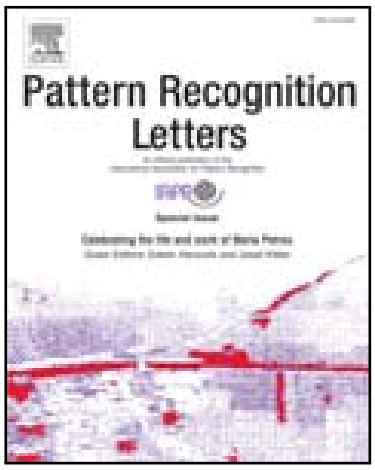

Editorial Pattern Discovery from Multi-Source Data

Xiaofeng Zhu, Jie Shao, Jilian Zhang

PII: S0167-8655(18)30089-8
DOI: 10.1016/j.patrec.2018.03.012
Reference: PATREC 7107

To appear in: *Pattern Recognition Letters*

Please cite this article as: Xiaofeng Zhu, Jie Shao, Jilian Zhang, Editorial Pattern Discovery from Multi-Source Data, *Pattern Recognition Letters* (2018), doi: 10.1016/j.patrec.2018.03.012

Editorial
Pattern Discovery from Multi-Source Data

Xiaofeng Zhu[a], Jie Shao[b], Jilian Zhang[c]

[a]*Xiaofeng Zhu is with Guangxi Normal University, Guilin, China (Email address: seanzhuxf@gmail.com).*
[b]*Jie Shao is with University of Electronic Science and Technology of China, Chengdu, China (Email address: shaojie@uestc.edu.cn).*
[c]*Jilian Zhang is with Jinan University, Guangzhou, China (Email address: Jilian.z.2007@smu.edu.sg).*

Keywords:

Advanced data acquisition technologies have been producing massive amounts of data in engineering sciences, medicine/healthcare sciences, and computer science [1, 2, 3]. In addition to volume, all kinds of data are naturally comprised of multiple representations in many real applications since only single-source data do not always meet all types of scenarios. For example, in image analysis, images are represented by local features and global features. Usually, different sources describe different characteristics of images. Thanks to the massive volume and multi-source structure of data, studies have shown that it is very difficult to deal with multi-source data using conversional analysis tools [4, 5, 6, 7]. It has been noticed that pattern recognition from multi-source data is different activity than that from sing-source data. Thus the understanding and analysis of multi-source data has been a very popular topic in machine learning and computer vision. Meanwhile the advent of multi-source data creates new challenges for current information technology.

The main aim of this special issue is bridging the gap between multi-source data and machine learning theory. Specifically, this special issue targeted the most recent technical progresses on learning techniques for data analysis, including classification [8, 15, 22], index system [9], clustering [14], regression [15], query [10], zero-short learning [11], kernelized ridge regression [11], unsupervised feature selection [14], sparse learning [15], deep learning [16], principal component analysis [17], gene expression programming and cloud computing [18], spatial clustering [19], supervised feature selection [22], metric learning [23], label propagation [24], and many others, in many kinds of learning-based applications, including genetic data analysis [8], agricultural engineering [9], image retrieval [10], objective recognition and retrieval task [11], missing value imputation [15], semantic segmentation [16], the aging principle of face sketch and photo [17], electricity load forecast [18], grid-based and density-based learning [19], boundary node algorithm [24], and so on. Moreover, Wang *et al.* reviewed the state-of-the-art techniques from the study of multi-source data [25], aim at systematically exploring the ideas behind current multiple data source mining methods and to consolidate recent research results in this field.

In [8], Chen L. *et al.* proposed a distributed framework to predict potential miRNA-disease associations. By assuming that diseases and miRNAs could be represented as feature matrices, potential associations in miRNA-disease association matrix were predicted based on estimated feature matrices. Experimental results verified that the proposed computational model achieved the best performance, compared to the state-of-the-art models.

In [9], Chen X. *et al.* devised an index system to evaluate the application of agricultural engineering technology via proposing an algorithm to select appropriate indicators.

In [10], Zhou Z. *et al.* proposed a multiple contextual clue encoding approach for partial-duplicate image retrieval. By treating each visual word of a given query or an image as a center, the authors first proposed an asymmetrical context selection strategy to select the contextual visual words for the query and database images differently, and then captured the multiple contextual clues such as the geometric relationships, the visual relationships, and the spatial configurations between the center and its contextual visual words. Experiments conducted on the large-scale partial-duplicate image datasets demonstrated that the proposed method provided the highest retrieval accuracy and comparable performances in time and space efficiency, compared to the state-of-the-art methods.

In [11], Long T. *et al.* proposed a zero-short learning method to recognize the classes whose samples did not appear during training by using two different ways such as Kernelized Linear Discriminant Analysis (KLDA) and Central-loss based Network (CLN). In particular, both KLDA and CLN could force samples to be intra-class aggregation and inter-class separation. Experimental results showed that the proposed methods outperformed the state-of-the-art approaches in term of recognition and retrieval task.

In [14], Li Y. *et al.* proposed a novel unsupervised feature selection method to 1) use the property of the data to construct self-representation coefficient matrix [12], 2) utilize sparse representation to find the sparse structure of the self-representation coefficient matrix [12], and 3) embed a hypergraph Laplacian regularization term to make up the insignificance of the ordinary graph in the representation of multiple relations. Experimental results on real datasets showed that the proposed method outperformed the state-of-the-art methods in term of clustering.