Pattern Recognition 50 (2016) 155–177

Contents lists available at ScienceDirect

Pattern Recognition

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

Cross-domain, soft-partition clustering with diversity measure and knowledge reference

Pengjiang Qian [a,b,c,*], Shouwei Sun [a], Yizhang Jiang [a], Kuan-Hao Su [b,c], Tongguang Ni [d,,a], Shitong Wang [a], Raymond F. Muzic Jr. [b,c]

[a] School of Digital Media, Jiangnan University, Wuxi, Jiangsu 214122, China
[b] Case Center for Imaging Research, Case Western Reserve University, Cleveland, OH 44106, USA
[c] Department of Radiology, University Hospitals Case Medical Center, Case Western Reserve University, Cleveland, OH 44106, USA
[d] School of Information Science and Engineering, Changzhou University, Changzhou, Jiangsu 213164, China

ARTICLE INFO

Article history:
Received 22 December 2014
Received in revised form
6 August 2015
Accepted 11 August 2015
Available online 22 August 2015

Keywords:
Soft-partition clustering
Fuzzy *c*-means
Maximum entropy
Diversity index
Transfer learning
Cross-domain clustering

ABSTRACT

Conventional, soft-partition clustering approaches, such as fuzzy *c*-means (FCM), maximum entropy clustering (MEC) and fuzzy clustering by quadratic regularization (FC-QR), are usually incompetent in those situations where the data are quite insufficient or much polluted by underlying noise or outliers. In order to address this challenge, the quadratic weights and Gini-Simpson diversity based fuzzy clustering model (QWGSD-FC), is first proposed as a basis of our work. Based on QWGSD-FC and inspired by transfer learning, two types of cross-domain, soft-partition clustering frameworks and their corresponding algorithms, referred to as type-I/type-II knowledge-transfer-oriented *c*-means (TI-KT-CM and TII-KT-CM), are subsequently presented, respectively. The primary contributions of our work are four-fold: (1) The delicate QWGSD-FC model inherits the most merits of FCM, MEC and FC-QR. With the weight factors in the form of quadratic memberships, similar to FCM, it can more effectively calculate the total intra-cluster deviation than the linear form recruited in MEC and FC-QR. Meanwhile, via Gini-Simpson diversity index, like Shannon entropy in MEC, and equivalent to the quadratic regularization in FC-QR, QWGSD-FC is prone to achieving the unbiased probability assignments, (2) owing to the reference knowledge from the source domain, both TI-KT-CM and TII-KT-CM demonstrate high clustering effectiveness as well as strong parameter robustness in the target domain, (3) TI-KT-CM refers merely to the historical cluster centroids, whereas TII-KT-CM simultaneously uses the historical cluster centroids and their associated fuzzy memberships as the reference. This indicates that TII-KT-CM features more comprehensive knowledge learning capability than TI-KT-CM and TII-KT-CM consequently exhibits more perfect cross-domain clustering performance and (4) neither the historical cluster centroids nor the historical cluster centroid based fuzzy memberships involved in TI-KT-CM or TII-KT-CM can be inversely mapped into the raw data. This means that both TI-KT-CM and TII-KT-CM can work without disclosing the original data in the source domain, i.e. they are of good privacy protection for the source domain. In addition, the convergence analyses regarding both TI-KT-CM and TII-KT-CM are conducted in our research. The experimental studies thoroughly evaluated and demonstrated our contributions on both synthetic and real-life data scenarios.

1. Introduction

As we know well, partition clustering is one of the conventional clustering methods in pattern recognition which attempts to determine the optimal partition with minimum intra-cluster deviations as well as maximum inter-cluster separations according to the given cluster number and a distance measure criterion. The studies began with hard-partition clustering in this field, such as *k*-means [1–3] (also known as crisp *c*-means [3]), i.e., the ownership of one pattern to one cluster is definite, without any ambiguity. Then, benefiting from Zadeh's fuzzy-set theory [4,5], soft-partition clustering [6–24,26–43] emerged, such as classic fuzzy *c*-means (FCM) [3,6], where the memberships regarding one data instance to all underlying clusters are in the form of uncertainties (generally measured by probabilities [6,17,18] or possibilities [7–9]), i.e. fuzzy memberships. So far soft-partition clustering has triggered extensive research and the representative work can be reviewed from the following four aspects:

* Corresponding author at: School of Digital Media, Jiangnan University, Wuxi, Jiangsu, China. Tel.: +86 137 71510961.
E-mail address: qpengjiang@gmail.com (P. Qian).

http://dx.doi.org/10.1016/j.patcog.2015.08.009
0031-3203/

(1) FCM's derivatives [6–14]. For improving the robustness against noise and outliers, two major families of derivatives of FCM, i.e., possibilistic *c*-means (PCM) [3,7–9] and evidential *c*-means (ECM) [10–13], were presented by relaxing the normalization constraint defined on the memberships of one pattern to all classes, and based on the concepts of possibilistic partition and credal partition, respectively. In addition, Pal and Sarkar [14] analyzed the conditions in which we can or should not use the kernel version of FCM; and the convergence analyses regarding FCM were studied in [15,16], (2) maximum entropy clustering (MEC) [3,17–23]. Karayiannis [17] and Li and Mukaidono [18] initially developed the MEC models by incorporating the Shannon entropy term into the total intra-cluster distortion measure. After that, Li and Mukaidono [19] further designed a complete Gaussian membership function for MEC; Wang et al. [20] incorporated the concepts of Vapnik's ε-insensitive loss function as well as weight factor into the original MEC framework in order to improve the identification ability of outliers; Zhi et al. [21] presented a meaningful joint framework by combining the fuzzy linear discriminant analysis with the original MEC objective function; and the convergence of MEC was studied in [22,23], (3) hybrid rough-fuzzy clustering approaches [13,24–30]. Dubois and Prade [24] fundamentally addressed the rough-fuzzy and fuzzy-rough hybridization as early as 25 years ago. Then quite quantities of fuzzy and rough hybridization clustering approaches have been developed. For example, Mitra et al. [25] introduced a hybrid rough-fuzzy clustering algorithm with fuzzy lower approximations and fuzzy boundaries; Maji and Pal [26] varied Mitra's et al. method [25] into the rough-fuzzy *c*-means with crisp lower approximations and fuzzy boundaries for heightening the impact of the lower approximation on clustering; Mitra et al. [27] suggested the shadowed *c*-means algorithm as an integration of fuzzy and rough clustering; and Zhou et al. [28] discussed shadowed sets in the characterization of rough-fuzzy clustering, (4) other fuzzy clustering models as well as applications. Aside from the above mentioned three aspects of literature, there exists a plenty of other work regarding soft-partition clustering. For example, Miyamoto and Umayahara [3,29] regarded FCM as a regularization of crisp *c*-means, and then via the quadratic regularization function of memberships they designed another regularization method named fuzzy clustering by quadratic regularization (FC-QR); Yu [30] devised the general *c*-means model by extending the definition of the mean from a statistical point of view; Gan and Wu [31] proposed a classic fuzzy subspace clustering model and further analyzed its convergence; Wang et al. [32] proposed another fuzzy subspace clustering method for handling high-dimensional, sparse data; and in addition, some application studies with respect to soft-partition clustering were also conducted, such as image compression [33,34], image segmentation [35–37], real-time target tracking [38,39], and gene expression data analysis [40].

As is well known, however, the effectiveness of usual soft-partition clustering methods in complex data situations still faces challenges. Specifically, their clustering performance depends to a great extent on the data quantity and quality in the target dataset. They can achieve desirable clustering performance only in relatively ideal situations where the data are comparatively sufficient and have not been distorted by lots of noise and outliers. Nevertheless, these conditions are usually difficult to be satisfied in reality. Particularly, new things frequently appear in modern high-technology society, e.g., load balancing in distributed systems [41] and attenuation correction in medical imaging [42], and it is difficult to accumulate abundant, reliable data in the beginning phase in these new applications. Therefore, this issue strictly restricts the practicability of partition clustering, in both cases of hard-partition and soft-partition. In our view, there exist two countermeasures to this challenge. That is, on one hand, we try our best to go on refining the self-formulations of partition clustering, like the trials from crisp *c*-means to FCM, PCM, MEC, and the others (e.g., [10,27,29]); on the other hand, the collaboration between partition clustering and fashionable techniques in pattern recognition should also be feasible, including semi-supervised learning [43–45], transfer learning [46–59], multi-task learning [60–62], multi-view learning [63,64], co-clustering [65–67], etc. Semi-supervised learning utilizes partial data labels or must-link/cannot-link constraints as the reference in order to improve the learning effectiveness on the target dataset. Transfer learning aims to enhance the processing performance on the target domain by migrating some auxiliary information from other correlative domains into the target domain. Multi-task learning concurrently performs multiple tasks with interactivities among them so that they can achieve better performance than that of each separate one. Multi-view learning regards as well as processing the data from multiple perspectives, and then eventually combines the result of each individual view according to a certain strategy. Co-clustering attempts to perform clustering on both the samples and the attributes of a dataset, i.e. it simultaneously processes the dataset from the perspectives of both row and column. As far as these techniques are concerned, however, we prefer transfer learning due to its specific mechanism. Transfer learning works in at least two, correlative data domains, i.e. one source domain and one target domain, and the case of more than one source domain is also allowed if necessary. Transfer learning first identifies useful information in the source domain, in the form of either raw data or knowledge, and then it handles the data in the target domain with such information acting as the reference and supplements. This usually enhances the learning quality of intelligent algorithms in the target domain. When current data are insufficient or impure (namely, polluted by noise or outliers), but some helpful information from other, related fields or previous studies is available, transfer learning is definitely the appropriate choice. Currently, many methodologies regarding transfer learning have also been deployed. For example, Pan and Yang [46] made an outstanding survey on transfer learning. The transfer learning based classification methods were investigated in [47–50], and the classification problem could currently be the most extensive research field on transfer learning. Several transfer regression models were proposed in [51–53]. Two dimension reduction approaches via transfer learning were presented in [54,55]. In addition, the trials connecting clustering problems with transfer learning were studied in [56–59], and several transfer clustering approaches were consequently put forward.

In this literature, we focus on the combination of the new soft-partition clustering model with transfer learning, due to the following two aspects of facts. First, conventional soft-partition clustering approaches, such as FCM and MEC, are prone to being confused by the apparent data distribution when the data in the target dataset are too sparse or distorted by noise or outliers. This usually causes their inefficient and even invalid results. Second, transfer learning offers us additional, supplemental information from other correlative domains in addition to these existing data in the target domain. With such auxiliary information acting as the reference, it is possible to approach the underlying, unknown data structure in the target domain. To this end, we conduct our work in two ways, i.e., refining the soft-partition clustering formulation as well as incorporating the transfer learning mechanism. In the first point, in light of the separate advantages in different, existing soft-partition models, e.g., FCM, MEC, and FC-QR, we first propose a new, concise, but meaningful fuzzy clustering model, referred to as quadratic weights and Gini-Simpson diversity based fuzzy clustering (QWGSD-FC), which aims at simultaneously inheriting the most merits of these existing methods. Then, based on this new model, by means of transfer learning, two types of cross-domain, soft-partition clustering frameworks and their corresponding algorithms, called Type-I/Type-II knowledge-transfer-oriented *c*-means (TI-KT-CM/TII-KT-CM), are separately developed. The primary contributions of our studies in this manuscript can be concluded as follows.

(1) As a basis of our work, the delicate QWGSD-FC model concurrently has the advantages of FCM, MEC and FC-QR. That