Perspective

iMedPub Journals
www.imedpub.com

Journal of Obesity & Eating Disorders
ISSN 2471-8203

2022
Vol.8 No.6:128

Common Clustering Assignment Regularization is Adopted to Explore the Consistency among Multiple Views

Liping Dong*

Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, CA, USA

***Corresponding author:** Liping Dong, Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, CA, USA. E-mail: donggping@gmail.com

Received date: October 17, 2022, Manuscript No. IPJOED-22-15192; **Editor assigned date**: October 19, 2022, PreQC No. IPJOED-22-15192 (PQ); **Reviewed date**: October 28, 2022, QC No. IPJOED-22-15192; **Revised date** November 07, 2022, Manuscript No. IPJOED-22-15192 (R); **Published date**: November 16, 2022.DOI: 10.36648/2471-8203.8.6.128

Citation: Dong L (2022) Common Clustering Assignment Regularization is Adopted to Explore the Consistency among Multiple Views. J Obes Eat Disord Vol.8 No.6: 128

Description

Multi-view clustering integrates multiple feature sets, which usually have a complementary relationship and can reveal distinct insights of data from different angles, to improve clustering performance. It remains challenging to productively utilize complementary information across multiple views since there is always noise in real data, and their features are highly redundant. Moreover, most existing multi-view clustering approaches only aimed at exploring the consistency of all views, but overlooked the local structure of each view. However, it is necessary to take the local structure of each view into consideration, because individual views generally present different geometric structures while admitting the same cluster structure. To ease the above issues, in this paper, a novel multi-view subspace clustering method is established by concurrently assigning weights for different features and capturing local information of data in view-specific self-representation feature spaces. In particular, common clustering assignment regularization is adopted to explore the consistency among multiple views.

An alternating iteration algorithm based on the augmented Lagrangian multiplier is also developed for optimizing the associated objective. Clustering aims to organize reasonably unlabeled data to discover meaningful patterns. Clustering has an important role in many fields, including machine learning data mining and pattern recognition. Clustering methods can be mainly categorized into two groups: partitioning clustering and hierarchical clustering.

In particular, spectral clustering is a graph-based algorithm for partitioning arbitrarily shaped data structure into disjoint clusters. Numerous spectral clustering methods and their variants have been proposed, such as Ratio Cut, K-way Ratio Cut, Min Cut, Normalized Cut and Spectral Embedded Clustering. The clustering performance of all of these methods is largely dependent on the quality of the so-called similarity graph, which is learned according to the similarities between the corresponding data points.

Sparse Subspace Clustering

Recent works on spectral clustering-based subspace clustering have attracted considerable attention due to the promising performance in data clustering. Subspace clustering works on the assumption that data are drawn from a union of low-dimensional subspaces, *i.e.,* every subspace is equivalent to a cluster. According to the self-expression of data and by imposing a properly chosen constraint on the representation coefficients, a representation matrix uncovering the intrinsic subspace structure of data can be obtained. For example, Elhamifar and Vida proposed sparse subspace clustering, which learns a graph by adaptively and flexibly selecting data points. Liu et al. imposed a low-rank constraint on the representation matrix to capture the global structure of data. Dornaika and Weng incorporated a manifold regularization term into SSC to capture the manifold structure. More recently, Peng et al. proposed a new deep model—Structured auto Encoder by preserving the global and local structure of subspace for clustering. Nevertheless, these approaches are mostly applied to group single-view data. In the era of Big Data, various data sources are represented by multiple distinct feature sets. Traditional clustering methods can be used to group multi-view data by simply concatenating all views into a monolithic one. However, compatible and complementary information across all views can be typically under-utilized. Recently, numerous multi-view clustering approaches have become available. Bickel and Scheffer generalized the conventional K-means and expectation–maximization clustering methods to the multi-view case. However, these methods are directly conducted in the original feature space, and then they fail to discover the geometric structures existing in each feature space. Canonical correlation analysis is a prominent statistical approach for learning from multiple views, but is restricted to linear transformation for each view. Liu et al. proposed a multi-view non-negative matrix factorization algorithm for clustering that learns a common representation of all views and then feeds it into K-means clustering. Such a common coefficient matrix does not consider the flexible structures of different feature spaces because of the heterogeneity of different views. Assuming that each sample should share the same cluster in all views, co-

regularized spectral clustering proposed two co-regularization terms to force the consistency of clustering labels criss-crossing different views. The Laplacian matrix used in CRMSC is learned from the original data, and then the obtained low-dimensional representation of the data is used to perform clustering. The quality of the Laplacian matrix may not be good enough because of the influence of noise. As a result, the clustering performance may be seriously affected. To construct a reliable similarity graph from different views, a low-rank and sparse decomposition technique is adopted to improve the robustness of multi-view spectral clustering. However, the above methods ignore the importance of different views, *i.e.,* treating all of the features equally.

Feature Selection

Recently, several multiple-graph learning-based clusteringmethods have been established to identify clustering ability of different views. Auto-weighted multiple-graph learning simultaneously learns an optimal weight for each Laplacian matrix and conducts spectral clustering. Multi-view learning with adaptive neighbors can automatically learn the weights of all views during each iteration. Simultaneously, the local manifold structure of multi-view data is captured. The above methods achieve better performances compared to those methods that do not consider the weights of views. To learn a better graph from multiple views, Liang et al. proposed to simultaneously explore the consistency and inconsistency across different views. However, the similarity graph is usually learned in the original feature space. Such a graph might be severely damaged, and the true similarities among samples cannot be guaranteed due to the influence of noise and redundancy. The above-mentioned work assumes that data lie in a single subspace. This assumption contradicts the observation, *i.e.,* in many problems, data in a cluster often lie in a low-dimensional subspace of the high-dimensional ambient space. Recent studies have focused on developing multi-view subspace clustering methods. For example, Gao et al. unified classic spectral clustering and self-representation learning into a framework. Brbić and Kopriva introduced simultaneous sparsity and low-rankness constraints on the coefficient matrix to extract underlying structure in multi-view data. Luo et al. jointly exploited consistency and specificity for subspace representation learning by using a set of specific representations and a shared consistent representation. The above methods mainly treat all of the features of each view as a whole to learn a single representation or common representation of all views. However, features of real data may exhibit a high degree of redundancy, and hence unrelated information can be introduced to degrade clustering performance. Although various feature-selection methods have been proposed, conducting feature selection and clustering in two separate steps may not obtain the optimal clustering results. Despite the recent progress within multi-view clustering, most existing methods ignore the feature-level relationship or the local structure of multi-view data. Taking Fig. 1 as an example, five independent subspaces are constructed and then 20 vectors from each subspace are sampled. The affinity matrix in which all of the vectors of the same subspace are connected, while those vectors belonging to other subspaces are not connected. The affinity matrices learned by LSR and SSC respectively. Compared to LSR and SSC, which capture the global and local structure of data, respectively, and ignore the feature-level relationship of data, the affinity matrix provided by the proposed method much better approximates the ideal affinity matrix. To ease the above limitations, a novel multi-view subspace clustering method is proposed based on data self-representation, which simultaneously weights features and learns the local structure of each view. By weighting the original features, the effective and robust features can be extracted, thus alleviating the influence of redundancy. Moreover, data self-representation and local structure learning are integrated into a unified framework such that the inherent difference in each view can be captured *via* simultaneously exploiting the global and local information of each view. In particular, common cluster structure regularization is used to capture consistency across different views.