National Journal of Multidisciplinary Research and Development
ISSN: 2455-9040; Impact Factor: RJIF 5.22
Received: 25-04-2019; Accepted: 31-05-2019
www.nationaljournals.com
Volume 4; Issue 3; July 2019; Page No. 25-27

Multi view clustering with fuzzy and view weighting

Antim Yadav[1], Shadab Ali[2]
[1] Institute of Engineering and Technology, Computer Science, Rajasthan University Jaipur, Alwar, Rajasthan, India
[2] Professor, Institute of Engineering and Technology, Computer Science, Rajasthan University Jaipur, Alwar, Rajasthan, India

Abstract
Internet and its related applications like social networking, online shopping, virtual classrooms etc are growing and access to more and more people around the world at a fast pace. A lot of information is collected through clustering from this data. Ordinarily clustering performed over a single view of data is not sufficient to derive proper information. Recently, practice of combining different views of data through clustering has increased. Many research works prove that such method, called multi-view clustering, produces better clustering results. As the dimensionality of data increases, inclusion of all dimensions in clustering becomes time consuming. Moreover, the semantics of data. Inherently gives preference to some attributes of data objects over others. This leads to feature selection as an essential step before clustering is performed. When multiple views are involved, a ranking system among the views can also benefit by producing results oriented towards what analyst desires. Hence, this dissertation focuses on designing a multi-view clustering method which involves feature selection and view weighing. scheme.

Keywords: fuzzy, clustering, anamoly, hybridness

1. Introduction
The data as present in today's world of increased Internet usage does not have single coherent view. The interactions among users are of many varieties and lead to many views of a single dataset. Different views of webpages are a classic example. Data analysts hold that if information from multiple views is combined through clustering, it gives better results. Thus, multi view clustering has become active research topic.

1.1 Clustering
The need of analyzing data available in the market for purposes like storage, sorting, updation, searching and more brings with it the need of new tools for the same. While developing new technologies for the purpose, the data analysts have to deal with one more drastic issue - hybrid nature of data. The contents in a webpage data clearly explain the hybridness of data In a single webpage, we can have images, videos, texts, hyperlinks, links to other pages, audios and more. Data analysis in such a complicated environment is a sure tedious task. Learning/discovering patterns in data in order to analyze data and use it further comes in the area of data mining and particularly one task of its six common classes of tasks-Clustering, other being Anamoly detection, Association Rule Learning, Classification, Regression and Summarization. The aim of clustering is to sort and group similar data from a bulk into clusters while maximizing the intra-cluster similarity and minimizing the inter-cluster similarity. The formed clusters are non-empty, may possess balance, differ from each other and may have even different criterion for cluster formation like density, distance, pairwise relation and more.

1.2 Multi-view data
The contents in a webpage data, being complex to handle, show a new perspective of viewing data and later clustering it accordingly. Taking the images, audios, videos, textual content, hyperlinks and other content of a single webpage as different views of the same page, one for each type of content, each aspect of the webpage can be put into consideration. This perspective of representing a complex form of data is termed as the multi-view approach and the data seen through various views as multi-view data. In easier words, multiple views represent different representations of the same set of instances. Set of instances here refers to the single platform from where the data is extracted for making its analysis easier.

1.3 Multi-view clustering
Clustering of multi-view data is a tedious and challenging task, the major challenge being techniques to combine together multiple views or representations of the same set of instances in order to get as output the best clustering result. Since the conventional clustering is applicable on only single-view data and cannot be directly applied to multi-view data, there arises a need to develop newer algorithms for handling the multi-viewed nature of data. The next challenge is to handle the different dimensions and semantics of the various views of the same set of instances which are actually the reasons that data of multiple views cannot be compared to each other. Clustering should be such that the effect of such heterogeneity is minimized. Clustering may also require simplification tasks like feature selection or adding weights to views in the process because some views can be of high-dimensionality and therefore incur high computational complexity and low clustering accuracy.

1.4 Literature Survey
The need to learn multi-viewed data is emerging with each passing day. This chapter surveys the various approaches towards multi-view learning and the various field multi-

view learning is applicable. Also, some of the noteworthy research works headed in the direction of multi-view learning are briefly discussed. One more topic under recent research is the hybrid nature of data as shown in Fig. of the webpage example discussed above. (While the term hybrid may even refer to the incompleteness or missing data, the scope of this dissertation is to handle the different types of data to be clustered together.

Fig 1

2. Proposed Clustering Algorithm Proposed Clustering Algorithm

A fuzzy clustering method for multi-view datasets is proposed here. It involves weight learning for views and features both. The proposed algorithm is based on a hard-clustering method WMCFS of Xu *et al.* [11].

2.1 Problem formulation

Multi-view clustering refers to the problem of partitioning a set of objects into clusters based on their similarity. The semantics of similarity are different in each view of the objects and the clustering output summarizes it.

a) Data Model

A dataset of objects and views is denoted through as. Here is object from the view. If each view is considered separately, then is actually a collection of datasets which can be denoted as

$$\chi = \{\mathbf{X}^1, \mathbf{X}^2, \ldots, \mathbf{X}^V\},$$

$$\mathbf{X}^1 = \{x_1^1, x_2^1, \ldots, x_N^1\}$$

$$\mathbf{X}^V = \{x_1^V, x_2^V, \ldots, x_N^V\}$$

Thus, a generalized denotes the set of objects from view in Fig. 3.1 shows this pictorially. The width of each view depends on the number of dimensions in it. That is, different views can have different number of dimensions. A view has number of dimensions; hence any object in view is a tuple of real numbers.

2.2 Proposed Fuzzy Multi-view Clustering

An iterative algorithm is proposed here to perform clustering over multiple views of a dataset simultaneously. The information in different views is combined through a combined global objective function. The process involves fuzzy clustering at view level, view weight learning and feature weight adjustment. The learning phase involves the clustering phase as iterative structure, shown in Fig 2.2.

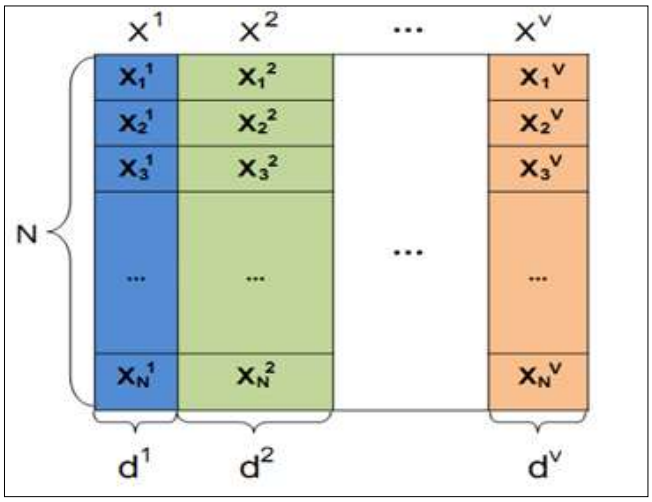

Fig 2: Illustration of multi-view data mode

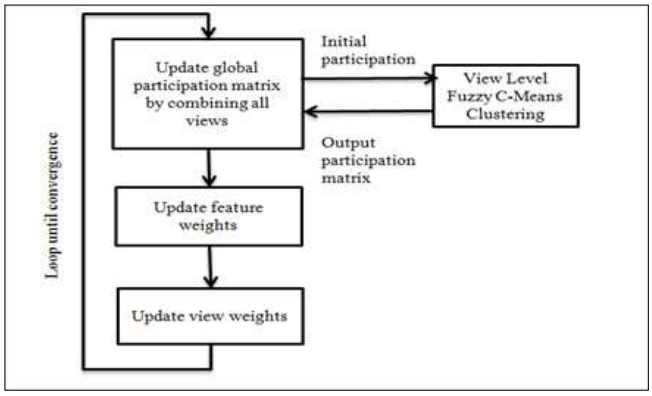

Fig 3: Iterative structure of the proposed algorithm

3. Conclusion

Modern real life applications, like face recognition, sentiment analysis, handwritten character recognition, webpage design analysis etc have an underlying data bank which though referring to same set of objects has different representations.

a) Summary of proposal: The various representations of the data are called views of data. Owing to high dimensionality of data, feature selection through a weighing scheme has been incorporated. A ranking (weighting) system for the views is also included. The cluster output produced is fuzzy in the sense that a data object may belong to more than one cluster as indicated by the output participation matrix. Thus, a global objective function has been proposed which considers fuzzy clustering of multiple views of data with feature and view weights. The clustering process involves automatic adjust-and-update computations for the feature and view weights. View level clustering is similar to Fuzzy C-Means.

b) Experimental Validation: Fuzzy Multi-view Clustering captures the underlying information from data better than the hard clustering of multi-view data. Also, the rate of convergence is high due to Fuzzy C-Means performed at view level. The global objective function which combines all views in linearly weighted manner serves better than the plain objective function of Fuzzy C-Means of single view clustering. All these observations are validated through experiments on real-life data with known ground truths. The results prove the proposal to be better both in terms of convergence speed and clustering accuracy.

4. Future Scope

With datasets having objects uniformly divided over all classes, Fuzzy C-Means does not produce significant improvement over any crisp clustering. Hence, the proposal can be adapted in future with a fuzzification process which gives better results on such balanced datasets too. Other fuzzy clustering approaches, besides Fuzzy C-Means, can be included. The proposed FMVC algorithm works entirely on numeric data. A variant to deal with mixed type of data can be useful.

5. References

1. Xu YM, Wang CD, Lai JH. Weighted Multi-view Clustering with Feature Selection, Pattern Recognition Letters. 2016; 53:25-35.
2. Kettenring J. Canonical analysis of several sets of variables, Biometrika, 58, 433-451.
3. White M, Yu Y, Zhang X, Schuurmans D. Convex multi-view subspace learning, Advances in Neural Information Processing Systems. 2012; 25:1-9.
4. Celebi ME, Kingravi HA. Deterministic Initialization of the K-Means Algorithm using Hierarchical Clustering, International Journal of Pattern Recognition and Artificial Intelligence. 2012; 26(7):1250018-1250041.
5. Duwairi R, Md. Rahmeh A. A novel approach for initializing the spherical K-means clustering algorithm in Simulation Modeling practice and Theory papers. 2015; 54:49-63.
6. Wang X, Qian B, Ye J, Davidson I. Multi-Objective Multi-View Spectral Clustering via Pareto Optimization, Proceedings of the 2013 SIAM International Conference on Data Mining, 2013.
7. Wang Y, Lin X, Wu L, Zhang W, Zhang Q, Huang X. Robust Subspace Clustering for Multi-View Data by Exploiting Correlation Consensus, IEEE Transactions on Image Processing. 2015; 24(11):3939-3949.
8. Serra A, Greco D, Tagliaferri R. Impact of different Metrics on Multi-View Clustering, Proceedings of the 2015 International Joint Conference on Neural Networks (IJCNN), 2015, pp. 1-8.
9. Tyagi G, Patel N, Sethi I. Soft-Hard Clustering for Multiview Data, Proceedings of the 2015 IEEE 16th International Conference on Information Reuse and Integration, 2015, 464-469.
10. Xu YM, Wang CD, Lai JH. Weighted Multi-view Clustering with Feature Selection, Pattern Recognition Letters. 2016; 53:25-35.