International Journal of Research in Advent Technology, Vol.6, No.11, November 2018
E-ISSN: 2321-9637
Available online at www.ijrat.org

Review on Feature Weighted Fuzzy Particle Swarm Optimization

P.Lavanya
Department of Information Technology
Kg College of Arts and Science
Coimbatore.

Abstract: Data mining is the process of discovering required patterns in a huge amount of data set. It is an extracting knowledge form the unknown pattern. It is applied successfully in the medicine, healthcare, insurance, business background, transportation and weather forecast. The clustering is an unsupervised learning for discovering hidden patterns in data and it is a process of partition the data set into several groups based on the similarity which is done by their distance. In the fields of machine learning, pattern recognition, image process, information retrieval, bioinformatics, data compression, and computer graphics are used the clustering process. Euclidean distance is obtained by differentiating between centroid and data item. The data sets have many features but all those features are not important to solve the clustering problems. To handle this circumstance, assign weights to their features for improving the performance of clustering accuracy and reducing its computational time. Feature Weighted Fuzzy Particle Swarm Optimization (FW-PSO) algorithm is used to solve unsupervised classification and produced best result than Fuzzy Particle Swarm Optimization.

1. INTRODUCTION

Data Mining:

Data Mining is process of extracting the unknown pattern from the large set of data for business as well as real t applications. It's task requires fast as well as accu partitioning of bulk data sets, which may come with a va of attributes or features.

Knowledge Discovery in Databases (KDD) is the nontrivial selection of implicit, previously unknown and potentially required knowledge from data in given databases**.** It is actually the process of finding the hidden information or pattern from the given dataset.

Applications of data mining:

- Financial Data Analysis
- Retail Industry
- Telecommunication Industry
- Biological Data Analysis
- Other Scientific Applications
- Intrusion Detection

Techniques in data mining:

Many procedures are used in data mining for explaining the type of mining as well as the data improvement action.

1. ***Association:***

Association or relationship is to identify the pattern from the same type. To construct a simple correlation between two or more objects, can be achieved simply with different association tools.

2. ***Classification:***

Classification is one of the forms of data analysis that can be used for extracting models that describes the important classes or to predict future data trends. A model or classifier is constructed to predict the categorical labels. The Data Classification process includes two steps such as Building the Classifier or Model and Using Classifier for Classification. Classification assigns categories to a collection of data in order to aid in more accurate predictions and analysis.

3. ***Clustering:***

Clustering is the processing of grouping a set of data that are more similar is in same groups; other dissimilar is put into other groups. Clustering is one of the method in unsupervised classification, it consists the data without any predefined classes. Every good clustering method will produce with the high quality clusters in which that the intra-class similarity must be high and the inter-class similarity must be low.

Clustering methods:

i) ***Hierarchical methods:*** These methods construct the clusters by recursively partitioning the instances in either a top-down or bottom-up fashion. These methods can be sub-divided as following:

- ***Agglomerative hierarchical clustering****:* Each object initially represents a cluster of its own. After that the clusters are successively merged until when we get the desired cluster structure.

- ***Divisive hierarchical clustering**:* All the objects at the beginning belongs to only one cluster. Then the cluster is divided into sub-clusters, which are successively divided into their own sub-clusters. This process continues until the desired cluster structure is obtained.

ii) ***Partitioning methods:*** Partitioning methods is especially well-known methods that change place instances by moving them from one cluster to another, starting from an initial partitioning. These methods require that the number of clusters will be previously set by the user. To achieve global optimality in partitioned-based clustering, an exhaustive enumeration process of all possible partitions is required.

Because this is not feasible, certain greedy heuristics are used in the form of iterative optimization. Namely, a relocation method iteratively relocates points between the *k* clusters. Various types of partitioning methods are presented below.

iii) ***Density-based methods:*** Density-based methods assume that the points that belong to each cluster are drawn from a specific probability distribution

iv) ***Grid-based methods:*** These methods partition the space into a finite number of cells that form a grid structure on which all of the operations for clustering are performed.

4. *Prediction*

Prediction is a part of a times series analysis methods that is a broad subject and runs from predicting the not a success of components or machinery, to identifying deceit and even the prediction of company earnings. Used in combination with the other data mining techniques, prediction involves analyzing trends, classification, pattern matching, and relation. By analyzing past events or instances, you can make a prediction about an event.

Sequential patterns:

Often used larger than longer-term data, in order patterns are a useful technique for identifying trends, or regular occurrences of analogous actions.

5. *Decision trees:*

An overall structure is to support the use and selection of specific data. Start with a simple question that has two answers within the decision tree. Every answer tends to a further question for helps toclassify the data, Then it can be categorized, or a prediction can also be made based on each answer.

At hand work, solved clustering technique in data mining that extremely fashionable technique in genuine globe application that technique proposed many researcher by various algorithm to address clustering techniques, give the following topic present more details about clustering technique.

2. RELATED WORK

Feature selection and weighting are normally ways to improve KNN classification algorithm. Qinghua Cao, Yu Liu were use the reverse cloud algorithm to map the training samples into clouds. Each attribute is mapped to a cloud vector. Reverse cloud algorithm is not sensitive to the noise on data sets and it can eliminate the impact of noise on classification effectively. By comparing the similarity of clouds in the cloud vector, we can find out a fitness function to measure the feature weighting results. The weighting process is a typical optimizing problem.

KNN algorithm based on particle swarm optimization (PSO) [1] feature weight learning is comparatively produce better classification accuracy with other classic KNN algorithms and other well-known improved KNN algorithms.

Min Chen and Simone A. Ludwig were developed fuzzy clustering, which is a popular unsupervised learning method that is used in cluster analysis. Fuzzy clustering allows a data point to belong to two or more clusters. Fuzzy C-Means is the most well-known method that is applied to cluster analysis, however, the shortcoming that the number of clusters needs to be predefined. This Particle Swarm Optimization approach determines the optimal number of clusters automatically with the help of a threshold vector. The algorithm first randomly partitions the data set within a preset number of clusters, and then uses a reconstruction criterion to evaluate the performance of the clustering results. The experiments conducted demonstrate that the proposed algorithm automatically finds the optimal number of clusters.

Fuzzy C-Means (FCM) algorithm is a one of the most popular fuzzy clustering techniques. However, it is easily fell in the local optima. Particle swarm optimization (PSO) is a speculative global optimization modeland it is used in many optimization problems. Shouwen Chen, Zhuoming Xu and Yan Tang [2] were developed a hybrid clustering algorithm with the help of an improved PSO and FCM is proposed, which makes use of the merits of both algorithms.

Fuzzy clustering has become an important research field with many applications to real world problems. Among fuzzy clustering methods, fuzzy c-means (FCM) is one of the best known for its simplicity and efficiency, although it shows some weaknesses, particularly its tendency to fall into local minima. To tackle this shortcoming, many optimization-based fuzzy clustering methods have been proposed in the literature. Some of these methods are based on a meta heuristic optimization, such as particle swarm optimization (PSO), others are hybrid methods such as FCM. Hybridize PSO and FCM for clustering have an improved accuracy over that the traditional partitioned clustering methods. On the other hand, PSO-based clustering methods have poor execution time in comparison to partitioned clustering techniques. Another problem with PSO-based clustering is that the current PSO algorithms require tuning a range of parameters before they are able to find good solutions. There are two hybrid methods for fuzzy clustering that deals with these shortcomings. The following methods adjusts PSO parameters dynamically during execution, that aims to provide better balance between exploration and exploitation and also avoiding the falling into local minima quickly as well as obtaining better solutions, such as FCM–IDPSO and FCM2–IDPSO and combine FCM with the latest version of PSO, the IDPSO. Ferdinando Di Martino. Salvatore Sessa was developed a new Extended Fuzzy Particle Swarm Optimization (EFPSO) [3] algorithm is presented and helped for the determination of hotspot events presented in spatial analysis. Extended Fuzzy C-Means (EFCM) can be used in the approximation of hotspot areas. EFCM gives better results than the classical Fuzzy C-Means.

Thomas A. Runkler and Christina Katz was introduced two new methods for minimizing the two reformulated versions of the FCM objective function by particle swarm optimization (PSO) [4]. In PSO–V method, all particles represent the component of the cluster centre. In PSO–U method, each particle should be represented as an unscaled and unnormalized membership value. PSO–V and PSO–U are compared with alternating optimization (AO) and with ant colony optimization (ACO) on two benchmark data sets: the single outlier and the lung cancer data sets. The stochastic methods ACO, PSO–V, and PSO–U are slower than AO, but in each experiment one of the two PSO variants significantly outperforms the other algorithms.

Feature-weight assignment can be based on a generalization of feature selection. That is, if all values of feature weights [5] are either 1 or 0, feature-weight assignment degenerates to the special case of feature selection. Generally speaking, a number in 0; 1 can be assigned to a feature for indicating the importance of the feature. The assignment of feature-weight can improve the performance of fuzzy c-means clustering. The weight assignment is given by learning according to the gradient descent technique.

Fuzzy C-Means (FCM) [6] is a powerful clustering algorithm and has been introduced to overcome the crisp definition of similarity and clusters. FCM ignores the importance of features in the clustering process. This affects its authenticity and accuracy.

Mousa Nazari, Jamshid Shanbehzadeh, and Abdolhossein Sarrafzadeh were overcome this problem by assigning weights to appropriate features with respect of the clustering importance.

Christian Borgelt was studies the problem of weighting and selecting attributes and principal axes in fuzzy clustering [7].The selection method that is not based on simply applying a threshold to computed feature weights, but directly assigns zero weights to features that are not informative enough. This has the important advantage that the clustering result that can be obtained on the selected subspace coincides with the projection (to the selected subspace) of the clustering result as it is obtained on the full data space.

Yue Yafan, Zeng Dayou and Hong Lei were discussed about Feature-weight [8] assignment can be regarded as a generalization of feature selection. Such as, if all the values of feature weights must be 1 or 0, then this feature-weight assignment degenerates to the special case of feature selection. Normally, a number in [0 1] can be assigned to a feature for denoting the feature's importance. It shows that an appropriate assignment of feature-weight can improve the performance of fuzzy c-means clustering. The gradient descent technique is used for the weight assignment that is given by the learning.

The guiding principle of similarity-based clustering [9] is "similar objects are within the same cluster and dissimilar objects are in different clusters," and a similarity measure must be defined to compute the degree of similarity between two objects. Two objects are considered to be similar or dissimilar based on this degree. Obviously the boundary between the two terms, similar and dissimilar, is not crisp. Thus, similarity-based clustering is a type of fuzzy clustering though the generated partitions are considered crisp.

The Fuzzy C-Means [10] program is applicable to a wide variety of geostatistical data analysis problems, it generates fuzzy partitions and prototypes for any set of given numerical data. These partitions are useful for authenticate known substructures or suggesting substructure in unexplored data. The clustering criterion helped to aggregate subsets is a generalized least-squares objective function. Features of this program include a choice of three norms (Euclidean, Diagonal, or Mahalonobis), an adjustable weighting factor controls sensitivity to noise as well as the acceptance of variable numbers of clusters, and outputs which include some measures of cluster validity.

Fasheng Xu and Wei Chen were discussed about the realistic portfolio selection problem is studied and an algorithm named improved particle swarm optimization (IPSO) [11] is presented to solve this problem. Initially, a realistic portfolio selection model, as an improved model to the standard Markowitz model, is formulated for the selecting portfolios with transaction costs and quantity constraints. Due to these complex constraints that the traditional optimization algorithms fail to work efficiently and the heuristic algorithms can be the best method, so the improved particle swarm optimization is used to solve the problem.

Group technology (GT) is a useful way to increase productivity with high quality in cellular manufacturing systems (CMSs), in which cell formation (CF) is a key step in the GT philosophy. Fuzzy clustering has been successfully used to solve the CF problem; But, it may result uneven distribution of parts/machines where as the problem becomes larger. In this case, particle swarm optimization (PSO) can be used to tackle such a hard problem. The hybrid algorithm based on the fuzzy clustering and particle swarm optimization (FPSO) [12] to solve the CF problem.

Chih-Cheng Hung *and* Bor-Chen Kuo were proposed a new weighted fuzzy C-Means (NW-FCM) algorithm [13] is proposed to improve the performance of both FCM and FWCM models for high-dimensional multiclass pattern recognition problems. Nonparametric weighted feature extraction (NWFE) and Discriminated analysis feature extraction (DAFE) concepts are combined in NW-FCM for unsupervised clustering. The main features of NW-FCM, when compared to FCM, are the inclusion of the weighted mean to increase the accuracy, and, when compared to FWCM, the centroid of each cluster is included to increase the stability. NW-FCM algorithm gives greater classification accuracy and stability than that of FCM and FWCM.

In cluster analysis, some of the features of a given data set may fall in higher relevance than others. For avoid this issue, Feature-Weighted Fuzzy C-Means (FWFCM) approaches are helpful in recent years. An Improved FWFCM (IFWFCM) [14] is overcome the certain deficiencies in the existing FWFCMs, e.g., the elements in a feature-weight vector cannot be adaptively adjusted during the training phase and also the update formulas of a feature-weight vector cannot be derived analytically.

3. UNSUPERVISED CLASSIFICATION

Clustering is an unsupervised learning [15] method which is introduced by Jar dine and Sibson in 1968. The aim of the clustering is to discovering groups of similar instances within the dataset. Unsupervised classification is done by without class label in dataset.

The clustering is a partition set of the data into subset of clusters which is done by minimizing the objective function of the clustering models.

The k-means algorithm is vital role for solving clustering algorithm. The k-means is the simplest and most commonly used clustering algorithm. The simplicity is due to the use of squared error as the stopping criteria, which tends to work well with isolated and compact clusters. Its time complexity depends on the number of data points to be clustered and the number of iteration.

Fuzzy C-Means clustering [10] is an effective algorithm, even though the random selection in center points makes that the iterative process falling into the local optimal solution easily. For solving this problem, recently evolutionary algorithms such as Genetic Algorithm (GA), Simulated Annealing (SA), Ant Colony Optimization (ACO), and Particle Swarm Optimization (PSO) [23]have been successfully applied.

4. PARTICLE SWARM OPTIMIZATION

Particle swarm optimization (PSO) is a population based stochastic optimization technique inspired by bird flocking and fish schooling [16] that was originally designed and introduced by Kennedy and Eberhart in 1994. The particle swarm optimization is uses iterative practice to find several better solution fitness functions in a search space.

In PSO, the potential solutions are called particles and they fly through the problem space by following the in progress optimum. The algorithm flow in PSO starts with a population of particles whose positions corresponds to the potential solutions for the studied problem, and velocities are randomly initialized in the search space.

In each iteration, the search for optimal position is personal best position and global best position. The personal best position, pbest, is the best position the particle has visited and gbest is the best position the swarm has visited since the first time step.

5. DRAWBACK OF FUZZY PARTICLE SWARM OPTIMIZATION

- It take more computing time for processing application or methods
- Accuracy of clustering algorithms is very low
- Concentrate on important features only in dataset

6. FUZZY PRTICLE SWARM OPTIMIZATION BASED FEATURE LEARNING VECTOR:

This algorithm is automatically assign the weights to the attributes or even selects a proper subset for working out. The feature weights assigning to dataset is done in the form of the feature selection. Filter Approach, Wrapper Approach[17] are used in feature selection for clustering.

Assign the weights to each feature for given dataset to increase the accuracy and reducing the computational time by means of feature learning vector combined with particle swarm optimization[10, 26] is called FW-PSO (Feature Weight-Particle Swarm Optimization).

7. CONCLUSION

The classification based on unsupervised approach plays the vital role in data mining and machine learning community. There are many unsupervised classification algorithms have been introduced by various researchers for different applications. Each approach has its own advantages and disadvantages. The Feature Weighted Fuzzy Particle Swarm Optimization algorithm is overcome the demerits of the Fuzzy Particle Swarm Optimization.

REFERENCES

[1] Q. Cao and Y. Liu, "A knn classifier with pso feature weight learning ensemble," in *Intelligent Control and Information Processing (ICICIP), 2010 International Conference on*, 2010, pp. 110-114.

[2] S. Chen, Z. Xu, and Y. Tang, "A hybrid clustering algorithm based on fuzzy c-means and improved particle swarm optimization," *Arabian Journal for Science and Engineering*, vol. 39, pp. 8875-8887, 2014.

[3] F. Di Martino and S. Sessa, "A fuzzy particle swarm optimization algorithm and its application to hotspot events in spatial analysis," *Journal of Ambient Intelligence and Humanized Computing,* vol. 4, pp. 85-97, 2013.

[4] T. A. Runkler and C. Katz, "Fuzzy clustering by particle swarm optimization," in *2006 IEEE international conference on fuzzy systems*, 2006, pp. 601-608.

[5] X. Wang, Y. Wang, and L. Wang, "Improving fuzzy c-means clustering based on feature-weight learning," *Pattern recognition letters,* vol. 25, pp. 1123-1132, 2004.

[6] M. Nazari, J. Shanbehzadeh, and H. Sarrafzadeh, "Fuzzy c-means based on automated variable feature weighting," 2013.

[7] C. Borgelt, "Feature weighting and feature selection in fuzzy clustering," in *Fuzzy Systems, 2008. FUZZ-IEEE 2008.(IEEE World Congress on Computational Intelligence). IEEE International Conference on*, 2008, pp. 838-844.

[8] Y. Yafan, Z. Dayou, and H. Lei, "Improving fuzzy c-means clustering by a novel feature-weight learning," in *Computational Intelligence and Industrial Application, 2008. PACIIA'08. Pacific-Asia Workshop on*, 2008, pp. 173-177.

[9] D. S. Yeung and X. Wang, "Improving performance of similarity-based clustering by feature weight learning," *IEEE transactions on pattern analysis and machine intelligence,* vol. 24, pp. 556-561, 2002.

[10] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy c-means clustering algorithm," *Computers & Geosciences,* vol. 10, pp. 191-203, 1984.

[11] F. Xu, W. Chen, and L. Yang, "Improved particle swarm optimization for realistic portfolio selection," in *Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, 2007. SNPD 2007. Eighth ACIS International Conference on*, 2007, pp. 185-190.

[12] E. Mehdizadeh and R. Tavakkoli-Moghaddam, "A Fuzzy Particle Swarm Optimization Algorithm for a Cell Formation Problem," in *IFSA/EUSFLAT Conf.*, 2009, pp. 1768-1772.

[13] C.-C. Hung, S. Kulkarni, and B.-C. Kuo, "A new weighted fuzzy C-means clustering algorithm for remotely sensed image classification," *IEEE Journal of Selected Topics in Signal Processing,* vol. 5, pp. 543-553, 2011.

[14] H.-J. Xing and M.-H. Ha, "Further improvements in feature-weighted fuzzy c-means," *Information Sciences*, vol. 267, pp. 1-15, 2014.

[15] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*: Elsevier, 2011.

[16] E. Mehdizadeh, "A fuzzy clustering PSO algorithm for supplier base management," *International Journal of Management Science and Engineering Management,* vol. 4, pp. 311-320, 2009.

[17] KrissnaPriya.R., “A Improved Classification of Network Traffic using Adaptive Nearest cluster Based Classifier”, International Journal of Computer Trends and Technology ISSN:2231-2803,Vol.18, Issue No.1 January 2015.