International Journal of Hybrid Information Technology
Vol. 9, No.6 (2016), pp. 169-178
http://dx.doi.org/10.14257/ijhit.2016.9.6.15

Affinity Propagation Algorithm Based on Locality Preserving Projections and Particle Swarm Optimization

Lele Liu, Zhiping Zhou and Jiefeng Wang

Engineering Research Center of Internet of Things Technology Applications Ministry of Education, Wuxi, 214122, China.
liulele303189997@163.com

Abstract

Affinity propagation algorithm is a new powerful and effective clustering method. One of the major problems in clustering is the determination of the optimal number of clusters. In this paper, the particle swarm optimization algorithm is utilized to cope with this problem by using the parameter p as each particle and Silhouette index as the fitness, which can search for the optimal value of p and determine the optimal number of clusters automatically. Moreover, the problem of information overlap is the main drawback of affinity propagation algorithm in dealing with complex structure or high dimensional data for clustering. Hence the enhanced Locality preserving projections method is proposed to integrate with affinity propagation algorithm to reduce the dimension of the data as a processing step. As the result of experiment shows, the proposed method can simultaneously obtain the optimal number of clusters accurately and improve the clustering accuracy by eliminating the redundant information of the data without losing the internal nonlinear structure.

Keywords*: Affinity propagation clustering (AP); Particle swarm optimization (PSO); Dimension reduction; Optimal clustering number; Silhouette*

1. Introduction

Cluster analysis is a method of creating groups of objects based on similarity degrees of relevant features, which has been used in many areas such as data mining, pattern recognition and machine learning. K-Means [1] is arguably the most popular clustering algorithm. It is intuitive and fast. However, the performance of K-means is highly dependent on the initial state of centroids, which may result in unstable clustering performance. To solve this problem, Frey *et al*. [2] proposed the Affinity propagation clustering (AP) algorithm in 2007. The algorithm does not need to define the number of clusters, instead, constant search is performed to find out suitable center of clusters during the process of iteration, which can automatically identify the location and the number of cluster centers from the data points. All the data points are represented as a candidate in the AP algorithm, which avoids the clustering results being limited by the choice of initial clustering representative points. The algorithm does not need the symmetry of the similarity matrix generated in the dataset with high operation speed in dealing with large-scale multi class data. However, the AP algorithm based on the central clustering method is not suitable for the complex shape clustering problem though it has a good performance in the data sets with compact hyper sphere distribution.

Recently many researchers have proposed different types of improved algorithms based on AP algorithm, aiming at enhancing its clustering performance. To improve the distance measurement, J. Dong *et al.* [3] proposed a variable spread neighbor clustering method of similarity measure. The work in [4] used the multi-distance measure function instead of a single Euclidean distance function to avoid the noise caused by using a single distance measure function. High dimensional data sets have a large number of

ISSN: 1738-9968 IJHIT

independent attributes, thus it is difficult to find an appropriate cluster structure in high dimensional data. H. Jia *et al*. [5] used the spectral graph theory to perform the dimension reduction operation. In [6], B. Chen *et al*. integrated PCA and AP to propose a novel hybrid algorithm, which can delete redundant information and achieve more accurate results. However, PCA cannot overcome the drawback of liner dimension reduction methods, and the low dimensional space obtained cannot retain the local structure of original data. G. Gan *et al*. [7] proposed a subspace clustering algorithm by introducing attribute weights in AP algorithm. In order to generate a given number of optimal set of exemplars through AP, a method called K-AP clustering algorithm in which a constraint function was used to limit the number of clusters to be a requested one is proposed [8]. A novel method was proposed by K. Wang [9] in 2007, which used adaptive preference scanning to search space of the number of clusters. In other aspects, F. Shang *et al*. [10] proposed a novel Fast Affinity Propagation clustering approach (FAP). The work in [11] extended the single-exemplar model to a multi-exemplar one and proposed a MEAP algorithm which has been found to be more effective than AP in the applications of multi subclasses clustering.

According to the above analysis, a novel method called Silhouette clustering validity index is utilized to determine bias parameter of AP algorithm without prior knowledge. Then, AP is integrated into the enhanced Locality preserving projections (LPP) [14] and experiment results indicate that the proposed algorithm can obtain the optimal number of clusters, and also delete the redundant information in the data space under the conditions of effectively keeping the data inner nonlinear structure.

The remainder of this paper is organized as follows: Section 2 gives a brief review of the original AP algorithm and PSO. In section 3, we introduce the main idea and details of our algorithm. Firstly, the PSO algorithm is utilized to obtain the optimal parameter and then the enhanced LPP is utilized to deal with high dimensional data. The optimal parameter *p* obtained is used to determine the final dimension. Finally, the AP algorithm is utilized to analyze the low dimensional data obtained by LPP. Section 4 reviews the results of our experiments. Section 5 provides our conclusions.

2. Preliminary

2.1. Affinity Propagation Algorithm

AP algorithm is a new unsupervised clustering algorithm, which firstly sets all data points as potential cluster centers then establishes attraction information passing among data points to find clusters. The similarities $s(i,k)$ between any two points are the negative of Euclidean distance square. For instance, point x_i and point x_k :

$$s(i,k) = -\left\| x_i - x_k \right\|^2 \tag{1}$$

Where, the similarity $s(i,k)$ indicates how well the data point x_k is suited to be the cluster center for data point x_i. Before clustering, the AP clustering algorithm sets the bias parameter $p = s(i,i)$ for each data point x_i. The bigger value of the bias parameter *p* is, the more likely the point x_i is selected as the representative cluster. The original algorithm [2] sets *p* to the median value of the similarity matrix in the absence of a priori knowledge, as shown in formula (2) :

$$p_m = \frac{\sum_{i,j,i\neq j}^{N} s(i,j)}{N \times (N-1)} \tag{2}$$

There are two kinds of message exchanged between data points, the responsibility $r(i,k)$ and the availability $a(i,k)$, which take different kinds of competition into account respectively. For the point x_i, when $a(i,k)+r(i,k)$ reach the maximum, it is more likely that the point x_k is the final cluster center. When $a(i,k)+r(i,k)$ remains invariant, the message-passing procedure will stop. By iteration, the sample points are competing to obtain the final clustering center.

2.2. Particle Swarm Optimization Algorithm

In swarm intelligence optimization algorithm, particle swarm optimization (PSO) [12] algorithm is an effective global optimization algorithm which has many advantages such as simple principle, fast convergence speed. In PSO, each of solutions can be considered as a particle in the search space, and the fitness value is determined by the objective function. The particle will perform regulation of its velocity V_i and position X_i to search new solutions in the solution space. At the same time, the optimal solution P_{id} and the global optimal solution P_{gd} are recorded. Each particle moves according to its velocity. At each iteration, the particle movement is computed as follows:

$$v_{k+1} = \omega v_k + c_1 r_1 (P_{id} - x_k) + c_2 r_2 (P_{gd} - x_k) \quad (3)$$

$$x_{k+1} = x_k + v_{k+1} \quad (4)$$

Where, in Eq.(3) and (4), x_k is the position of particle i at time k, v_k is the velocity of particle i at time k, P_{id} is the personal best for particle i in the d th dimension and P_{gd} is the global best in d th dimension. ω is an inertia weight scaling the previous time step velocity. c_1 and c_2 represents the acceleration coefficients and r_1, r_2 is the random numbers distributed uniformly within the range $[0,1]$.

3. Proposed Method

3.1. Obtain the Optimal Parameter By PSO

When the AP algorithm is carried out, *p* is the parameter to adjust the number of clusters, which can affect the optimal clustering results. In traditional AP algorithm, the value of input preference *p* is usually set to p_m (the median value of the similarity matrix). However, the value cannot lead to a suboptimal clustering solution in many cases since the underlying manifold structure of the dataset is not considered, consequently the number of clusters for some data sets is different from the real number. In this paper, a new method is introduced to determine the optimal clustering number, in which the bias parameter *p* is utilized as the position coordinates of each particle, different values of *p* can be used to initialize the position of all particles. And then, the positions and directions of all particles are updated continuously, namely the new bias parameter *p* for AP can be achieved to conduct the clustering process, and the fitness of each particle can also be calculated by using the validity index Silhouette.

In the number of clustering validity index, Silhouette index can be used to estimate the optimal number of clustering. It reflects the intracluster compactness and intercluster separation of the clustering structure. Therefore, it can be utilized as the fitness evaluation index in this paper.

Assuming that there is a data set with n points and k clusters $C_i (i = 1,2,...,k)$ in which the Silhouette index validity of a point t can be calculated as following:

$$Sil(t) = \frac{b(t) - a(t)}{\max\{a(t), b(t)\}} \tag{5}$$

where $b(t)$ is the min value of $d(t, C_j)$, which is the mean dissimilarity or distance between the point t in cluster C_i and all points within another cluster C_j, and $a(t)$ is the mean dissimilarity or distance between the point t and other points within cluster C_i. S_{il-av} is the average value of the Silhouette index which can effectively reflect the quality of the whole cluster, the higher the value is, the higher the clustering quality is. So the fitness function of the algorithm can be defined as follows:

$$F = S_{il-av}(clu) \tag{6}$$

Among it, clu is the label of the data set obtained by using AP algorithm.

Inertia weight plays a vital role in controlling the process of exploration and exploitation by maintaining a balance in their capabilities. The large inertia weight is advantageous to the process of exploration, but lead to a low efficiency. The smaller inertia weight can accelerate the convergence of the algorithm, but may easily lead to a local optimum. The key to avoid falling into a local optimum and achieving an efficient search is to set up the reasonable inertia weight. In this paper, we have adopted a method for the best particle [13]. At any given moment a particle that has achieved the global best value will be self adjustment by its inertia weight to accelerate its search towards the global optimum. The particle which achieves the global best position will accelerate the speed in the direction in search of global optimum. And the rest of the particles follow the standard procedure of linearly decreasing the inertia weight. The self-regulating inertia weight method is defined as follows:

$$\omega_i = \begin{cases} \omega_i(t) + \eta\Delta\omega, & \text{best particle} \\ \omega_i(t) - \Delta\omega, & \text{otherwise} \end{cases} \tag{7}$$

Where, $\omega_i(t)$ is the current inertia weight. $\Delta\omega = {\omega_I - \omega_F}/{N_{Iter}}$ (N_{Iter} is the number of iterations, ω_I and ω_F are the initial and final values of inertia weight respectively) and η is the constant to control the rate of acceleration. In our simulation, the rate η is set as 1. Finally, we also need to design the interval of p to improve the efficiency. When the bias parameter space is $[-\infty, 0]$, its corresponding clustering number is $[1, N]$. The upper bound of the optimal number of cluster number of N data points is $\sqrt{N}$, which needs setting a bigger initial deviation as the starting point of the $-\infty$ direction. Literature [9] shows that when the preference value is set to half of the median of p, the number of exemplars identified by AP will always be equal to or larger than $\sqrt{N}$, so the starting value can be set to half of p_m.

The steps of our method can be summarized as follows:

Step1 Initialize the position and the speed of the particle.

Step2 Evaluate the fitness values of particles by using the Silhouette index as fitness function.

Step3 Look for individual extremum and group extremum adapted to AP algorithm.

Step4 Update the speed and the position of the particle and perform recursive search for the optimal individual extremum and group extremum.

Step5 If the convergence condition is satisfied or the maximum number of iterations is achieved, stop and output the clustering results, else go to step 2.

3.2. The enhanced Locality Preserving Projections

When high dimensional space data have redundancy information, it is too difficult to accurately cluster by barely relying on the similarity measure method of AP clustering. In order to improve the clustering quality, we need to reduce the dimension of the original data to effectively remove redundant information of space data. The LPP algorithm was proposed for manifold learning algorithm in 2002, which was mainly used for nonlinear manifold study and analysis. The LPP method can extract the most distinguishing feature for the purpose of dimension reduction. It is a local information preserved dimension reduction method, which overcomes the shortcomings of linear dimension reduction algorithm. Besides it skillfully combined the idea of Laplace feature mapping algorithm, so it can effectively keep the nonlinear structure of the data within the dimension reduction mapping of high dimensional data. In this paper, the LPP method is utilized to reduce the data dimension.

The steps of LPP algorithm can be summarized as follows: the first step is to create an adjacency graph that is constructed using the *k* nearest neighbor (KNN) method, by which a point is connected to the *k* nearest neighbor with a side link. The second step is to determine the weight and the third step is to obtain the eigenvalues, finally we can complete the mapping. But most of the existing LPP algorithms use the Euclidean distance to choose a global optimal *k* or conduct the adaptive neighborhood selection. In many cases, the KNN method based on Euclidean distance does not reflect the neighborhood structure of the real data. For instances, x_i and x_j of *n* dimensions, the Euclidean distance between them, is defined as following:

$$d_{ij} = \sqrt{\sum_{l=1}^{n} (x_{il} - x_{jl})^2} \tag{8}$$

It is known that the Euclidean distance perform well when the data sets are subject to the globular cluster structure or the global linear structure. But most of the real data are nonlinear and corresponding to the random manifold distribution of low dimensions to some degree, hence the Euclidean distance based on neighborhood selection method cannot obtain the real neighborhood structure of the data.

The neighborhood selection based on the manifold distance is defined by the redefinition of the distance between the sample points in the manifold space to determine the nearest *k* nearest neighbor. In this paper, an improved manifold space distance is proposed to choose the neighborhood, and the distance between the two points x_i and x_j in the manifold space is formulated as equation (9):

$$L(x_i, x_j) = e^{\frac{D(x_i, x_j)}{\sigma}} - 1 \tag{9}$$

Where, $D(x_i, x_j) = \dfrac{d_{ij} - \min(d_{ij})}{\max(d_{ij}) - \min(d_{ij})}$ is Normalized Euclidean distance, σ is an adjustable parameter. The data points are the vertices of graph $G(V, E)$ where V is the vertex set and E is the edge set. If P_{ij} represents the set of all paths connecting data x_i and x_j in the graph, then the manifold distance between them can be formulated as equation (10):

$$MD(x_i, x_j) = \min_{p \in P_{ij}} \sum_{k=1}^{|p|-1} L(P_k, P_{k+1}) \tag{10}$$

As we can see from the above formula, the manifold distance between two points is the minimum total length of the line among all paths connecting the two points in the graph

$G(V,E)$. Where, p represents a path connecting x_i and x_j and $|p|$ is the number of nodes on the path.

The distance between any two points can be calculated by using the manifold distance formula and all the edges associated with each vertex are sorted in ascending order. Then the top *k* vertexes are selected as the *k* neighborhood vertexes in the manifold space. After dimensionality reduction, the final dimension is determined and the low dimensional data is obtained.

Accordingly, our proposed algorithm can be summarized as follows:

1. The improved manifold space distance is used to determine the neighborhood in LPP.
2. The final dimension is determined by the optimal parameter *p* according to subsection 3.1, hence the low dimensional data can be obtained.
3. The AP with the optimal parameter *p* is executed as the clustering operation.

4. Experiment and Analysis

In order to verify the effectiveness of the proposed algorithm, the clustering performance of AP algorithm, adAP algorithm [9], PCA-AP [6] and our method are compared in this section in terms of the internal evaluation index Silhouette and the external index F-Measure. All algorithms use the same damping factor $\lambda = 0.5$, and the bias parameter *p* of AP is set as p_m. The parameter settings of PSO are configured: the maximum iteration is 50 times, the number of particles is 20, $c_1 = c_2 = 1.496$. $\omega_I = 1.05$, $\omega_F = 0.5$, which are the same as that in literature [13]. In this paper, 10 common data sets from UCI are selected for experiment, whose characteristics are shown in Table 1. The number of clusters (NC) and performance index values of each algorithm are respectively shown in Table 2.

Table 1. UCI Datasets

Datasets	Classes	Size	Dimensions
air	3	359	65
glass	6	214	10
wine	3	178	13
hearts	2	270	13
iris	3	150	4
zoo	7	101	17
sonar	2	208	61
WBC	2	683	10
vote	2	435	17
X8D5K	5	1000	9

Table 2. The Results of Three Algorithms

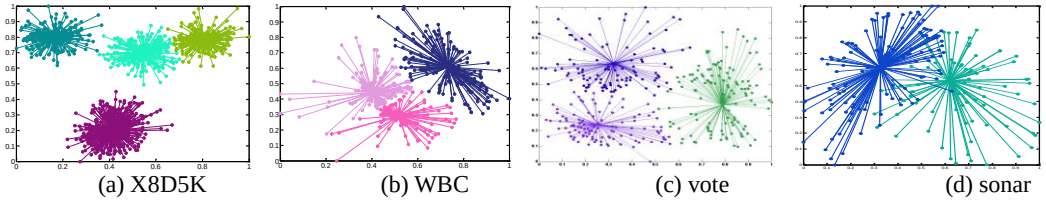
Datasets	True class	AP			adAP			Proposed method		
		NC	FM	$S_{il\text{-}av}$	NC	FM	$S_{il\text{-}av}$	NC	FM	$S_{il\text{-}av}$
air	3	59	0.197	0.229	2	0.546	0.792	3	0.668	0.801
glass	6	23	0.205	0.283	3	0.386	0.562	6	0.452	0.573
wine	3	11	0.465	0.311	3	0.801	0.725	3	0.816	0.759
hearts	2	21	0.252	0.276	2	0.590	0.560	2	0.672	0.564
iris	3	6	0.599	0.402	2	0.764	0.703	3	0.882	0.722
zoo	7	9	0.619	0.406	4	0.718	0.448	4	0.739	0.454
sonar	2	16	0.372	0.199	2	0.628	0.352	2	0.796	0.454
WBC	2	33	0.330	0.305	2	0.861	0.579	2	0.928	0.676
vote	2	30	0.269	0.126	2	0.867	0.332	2	0.930	0.584
X8D5K	5	7	0.562	0.480	5	0.934	0.586	5	0.996	0.619

As can be seen from Table 2, the number of clusters obtained by AP algorithm is relatively large compared with the real number of classes and that obtained by adAP algorithm for only 6 types of data sets are consistent with the real number of classes, whereas 9 types of data sets can obtain the correct number of clusters using our proposed algorithm. For the data set zoo, all algorithms fail to get the correct number of clusters, but the Silhouette value and F-Measure value of the proposed algorithm is higher than the other two algorithms, which indicates that the performance of our method is the best. Therefore, it can be concluded that the optimal *p* value of AP can be obtained by using PSO and Silhouette index effectively, and then correct number of clusters and better clustering results can be obtained.

In this paper, the standard dimension reduction is conducted for each data set, and the final dimension can be determined by using the Silhouette index and the optimal *p* value determined in advance, which is shown in Table 3. Four data sets X8D5K, WBC, vote and sonar are chosen to intuitively analyze the result of dimensionality reduction, and the low dimensional clustering results of AP using PCA and LPP are drawn respectively in Figure 1 and Figure 2.

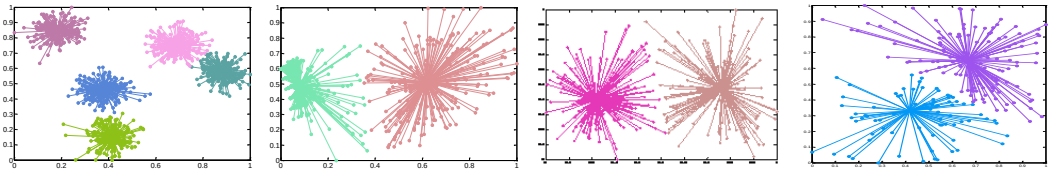
Table 3. Dimension Determination based on Silhouette

Datasets	Silhouette	Dimensions(Original dimension)
air	0.882	4(65)
glass	0.652	2(10)
wine	0.891	2(13)
hearts	0.573	2(13)
iris	0.842	2 (4)
zoo	0.645	2(17)
sonar	0.561	6(61)
WBC	0.688	2(10)
vote	0.584	3(17)
X8D5K	0.688	2(9)



(a) X8D5K (b) WBC (c) vote (d) sonar

Figure 1. Optimal Clustering Result Diagram of PCA



(a) X8D5K (b) WBC (c) vote (d) sonar

Figure 2. Optimal Clustering Result Diagram of LPP

According to Table 3, if the original dimension of the data is high, the dimension obtained is relatively high. For the data set with 10 dimensions, usually 2 or 3 dimensions are chosen. As we can see from Figure 1 and Figure 2, for the four data sets using LPP, the result between the clusters is relatively disperse and within the clusters is relatively compact, which can reflect the internal structure information of the data well. Also, the optimal number of clusters obtained is consistent with the real numbers. However, the numbers obtained by PCA for data sets X8D5K, WBC and vote are different from the real numbers and it cannot clearly distinguish the data within and between clusters, which indicates that the local structure of original data is not maintained. Although the correct number of clusters for data set sonar is obtained, there are some overlaps between the clusters.

Table 4. Clustering Results

	AP		PCA-AP		Proposed method	
Datasets	FM	$S_{il\text{-}av}$	FM	$S_{il\text{-}av}$	FM	$S_{il\text{-}av}$
air	0.197	0.229	0.725	0.821	0.856	0.882
glass	0.205	0.283	0.421	0.589	0.543	0.652
wine	0.465	0.311	0.721	0.725	0.823	0.891
hearts	0.252	0.276	0.526	0.560	0.682	0.573
iris	0.599	0.402	0.632	0.733	0.774	0.842
zoo	0.619	0.406	0.732	0.586	0.756	0.645
sonar	0.372	0.199	0.673	0.475	0.698	0.561
WBC	0.330	0.305	0.861	0.579	0.963	0.688
vote	0.269	0.126	0.834	0.537	0.868	0.584
X8D5K	0.562	0.480	0.621	0.636	0.998	0.688

In addition, the clustering performances of AP, PCA-AP and the proposed algorithm are analyzed in terms of F-Measure and Silhouette, which are shown in Table 4. It is demonstrated that the $S_{il\text{-}av}$ value and F-measure value is larger than that of PCA-AP for all data sets and also the $S_{il\text{-}av}$ value is higher than 0.5, which indicates that the proposed algorithm can effectively partition all data sets. Hence the reduced attribute space obtained by our method can effectively maintain the local structure of the original high dimensional space, meanwhile the compactness within clusters and separation between clusters can be achieved.

5. Conclusion

In this paper, a high-performance method based on the Locality Preserving Projections method is proposed to deal with the main drawback of affinity propagation algorithm in complex structure or high dimensional data for clustering. The method can delete the redundant information in the data space under the conditions of effectively keeping the data inner nonlinear structure. Meanwhile, in order to well determine bias parameter of affinity propagation algorithm without prior knowledge, the PSO algorithm is utilized to cope with this problem by using the parameter *p* as each particle and Silhouette index as the fitness. The experiment results verify the accuracy and the effectiveness of the proposed algorithm in this paper.

Acknowledgements

This work was supported by the Cooperative Industry-Academy-Research Innovation Foundation of Jiangsu Province (BY2013015-33) and the Fundamental Research Funds for the Central Universities (JUSRP51510).

Reference

[1] A. K. Jain, "Data clustering: 50 years beyond K-means", Journal Pattern recognition letters, **(2010)**, pp.651-666.
[2] B. J. Frey and D. Dueck, "Clustering by Passing Messages between Data Points", Journal Science, **(2007)**, pp.972-976.
[3] J. Dong, S. Wang and F. Xiong, "Affinity Propagation Clustering Based on Variable-Similarity Measure", J. Journal of Electronics & Information Technology, **(2010)**, pp.509- 514.
[4] D. Wang, L. Li and Z. Yu, "AP2CE: Double affinity propagation based cluster ensemble", 2013 International Conference on. IEEE, **(2013)**, pp.16-23.
[5] H. Jia, S. Ding and L. Meng, "A density-adaptive affinity propagation clustering algorithm based on spectral dimension reduction", Neural Computing and Applications, **(2014)**, pp.1557-1567.
[6] B. Chen and K. Song, "Affinity propagation clustering algorithm based on PCA", the 2014 International Conference on Applied Mechanics and Mechanical Automation. Macao, China, **(2014)**, pp.688-692.
[7] G. Gan and M. K. P. Ng, "Subspace clustering using affinity propagation", Journal Pattern Recognition **(2015)**, pp.1455-1464.
[8] X. Zhang, W. Wang and K. Nrvag, "K-AP: Generating specified K clusters by efficient Affinity Propagation", the 10th IEEE International Conference on Data Mining (ICDM), Sydney, NSW, Australia, **(2010)**, pp.1187-1192.
[9] K. Wang, J. Zhang and D. Li, "Adaptive Affinity Propagation Clustering", Journal ACTA AUTOMATICA SINICA **(2007)**, pp.1242- 1246.
[10] F. Shang, L. Jiao and J. Shi, "Fast affinity propagation clustering: A multilevel approach", Journal Pattern recognition **(2012)**, pp. 474-486.
[11] C. Wang, J. Lai and C. Suen, "Multi-exemplar affinity propagation", Journal Pattern Analysis and Machine Intelligence, IEEE Transactions on, **(2013)**, pp.2223-2237.
[12] R. C. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory", Proceedings of the sixth international symposium on micro machine and human science, **(1995)**, pp.39-43.
[13] M. R. Tanweer, S. Suresh and N. Sundararajan, "Self regulating particle swarm optimization algorithm", Journal Information Sciences **(2015)**, pp.182-202.
[14] X. yogi, "Locality preserving projections", Neural information processing systems, MIT, **(2004)**, pp.16: 153

Authors

Lele Liu, Master Degree Candidate; Institution: Engineering Research Center of Internet of Things Technology Applications Ministry of Education, Jiangnan University; Address: Binhu District Road No.1800, Wuxi Jiangsu, China; Tel: 15961709635; Email: liulele303189997@163.com; Subject: data mining.

Zhiping Zhou, PhD, professor in Jiangnan University. His main research interests include detection technology and automation devices, information security, *etc*. Tel: 13861867502; Email: zzp@jiangnan.edu.cn.

Jiefeng Wang, Master Degree Candidate; Institution: Engineering Research Center of Internet of Things Technology Applications Ministry of Education, Jiangnan University; Address: Binhu District Road No.1800, Wuxi Jiangsu, China; Tel: 18896724565; Email: 18352513420@163.com; Subject: data mining, pattern recognition.

.

