Applied Mathematics & Information Sciences
An International Journal

Domain Knowledge Blended Affinity Propagation

Wei Chen[1,2*], *Qichong Tian*[3], *Xiaorong Jiang*[1], *Zhibo Tang*[1], *Caihua Guo*[1], *Xinzheng Xu*[1], *Hong Zhu*[1,4] *and Shifei Ding*[1]

[1]School of Computer Science and Technology, China University of Mining and Technology, Xuzhou, Jiangsu, 221116, China
[2]State Key Laboratory of Coal Resources and Safe Mining, China University of Mining and Technology, Xuzhou, Jiangsu, 221116, China
[3]Department of Electronics and Information Engineering, Huazhong University of Science and Technology, Wuhan, 430074, China
[4]School of Medical Information, Xuzhou Medical College, Xuzhou, China, 221000

Received: 8 Jul. 2012; Revised 4 Oct. 2012; Accepted 6 Dec. 2012
Published online: 1 Mar. 2013

Abstract: As an important clustering algorithm, Affinity Propagation (AP) algorithm can quickly find the reasonable clustering center. But the AP algorithm is difficult to make correct clustering, when the sample in the weaker separability feature space. In this paper, the Domain Knowledge Blended Affinity Propagation (DKB-AP) algorithm is proposed. Combining the domain knowledge function and the similarity measure of the AP algorithm, the algorithm makes iterating to obtain the clustering result. The experimental data are three random sample sets, including two sample sets whose subclass aggregation degree are good, one sample set whose subclass aggregation degree is weak. The clustering results, Fowlkes-Mallows Validity Index and Error Ratefor for the Sets are analysed. The results show that the clustering result in the weaker separability feature space by DKB-AP algorithm is almost consistent with the clustering result in the separability feature space by AP algorithm.

Keywords: Affinity Propagation, Domain Knowledge, Feature Space, Similarity Measure

1. Introduction

The clustering method is an important way to find the data regularity, according to the distribution characteristics of the data in the feature space. Dynamic clustering algorithm is an important kind of clustering method, which has some key problems, including the similarity measure, the initial classification method, the selecting initial cluster center method, and the criterion function of evaluating the quality of clustering results. In the clustering process, the algorithm adjusts the sample class until the criterion function has reached a certain extreme value [1,2].

If the initial classification method and the initial cluster center are selected unsuitably, the criterion function is easy to fall into local extremum, making the data regularity of dynamic clustering algorithm largely differ from the real data regularity. To avoid this problem, Frey et al proposed the Affinity Propagation (AP) clustering algorithm [3,4]. The algorithm assumes that all data points are the initial cluster centers, to avoid that the clustering result is limited by the choice of initial cluster centers. The data regularity obtained from the AP clustering algorithm is objective in some degree. Taking into account that the AP clustering is hard to determine a good value of parameter to avoid a suboptimal clustering result, the multilevel fast AP clustering is outperforms than the original AP clustering [5]. The hierarchical strategy exploits self-similarity property to locate the position of cluster center, making the AP algorithm can analyse the large-scale data [6]. In recent years, the AP algorithm has a wide range of applications, including text clustering [7], image target analysis [8], hyperspectral band selection [9], analysis of the fMRI data [10], and so on.

Due to the feature extraction methods are not able to accord with the real characteristics of samples, the sample distribution in the feature space is not necessarily consistent with the sample distribution in the real world, which leads to the AP clustering algorithm may not be able to make the truly identical subclass cluster together. Therefore, the Domain Knowledge Blended Affinity Propagation (DKB-AP) clustering algorithm is proposed in this paper. Under the premise of not changing the feature extraction method, the proposed algorithm combines the AP

[*] Corresponding author: e-mail: davior.chen@gmail.com

clustering algorithm and the priori knowledge of the sample domain to make a correct classification.

The rest of the paper is organized as follows. Section 2 reviews the basic principle and iterative methods of the AP clustering algorithm, and analyzes the existing problems of this algorithm. Section 3 introduces the basic principle and iterative methods of the DKB-AP clustering algorithm. In Sections 4, randomly generated three sample sets are tested with the DKB-AP clustering algorithm. The experimental analysis is also given. Section 5 concludes this paper.

2. AP Clustering Algorithm

2.1. principle of the AP clustering algorithm [3]

The purpose of AP clustering algorithm is to find the optimal class representative point set, making the similarity sum of all sample points and the nearest subclass cluster centers is the largest. For the data sets X=$\{x_i | i = 1,2,\cdots,N\}$, $\forall x_i, x_i \in X$, the similarity based on squared error is $s(i,k) = -||x_i - x_k||^2$.

The AP algorithm makes all $x_i (i = 1,2,\cdots,N)$ as the candidate cluster centers, namely, $s(k,k) = p$, $\forall(x_k) \in X$, the value of p can affect the number of final clusters. $r(i,k)$ is the responsibility accumulation of x_k as the subclass cluster centers of x_i. $a(i,k)$ is the availability accumulation of x_k as the subclass cluster centers of x_i.

In order to avoid shocks, the original AP algorithm introduces the damping factor in the information update process. The default value of the damping factor is 0.5 [3]. We set it to 0.7 in our experiment. Set the current number of iterations is t, the information update process is as follows.

$$\begin{aligned} r^{(t)}(i,k) &= (1-\lambda) \times (s(i,k) - \max_{k' \neq k}\{a^{(t-1)}(i,k) + s(i,k')\}) \\ &= \lambda \times r^{(t-1)}(i,k) \qquad (1) \end{aligned}$$

$$\begin{aligned} a^{(t)}(i,k) &= (1-\lambda) \times (min\{0, r^{(t-1)}(k,k) \\ &+ \sum_{i' \neq i,k} max\{0, r^{(t)}(i',k)\}\}) \\ &+ \lambda \times a^{(a-1)}(i,k) \quad \text{while} \quad \mathrm{i} \neq \mathrm{k} \qquad (2) \end{aligned}$$

$$\begin{aligned} a^{(t)}(k,k) &= (1-\lambda) \times \sum_{i' \neq \{i,k\}} max\{0, r^{(t)}(i',k)\} \\ &+ \lambda \times a^{(t-1)}(i,k) \qquad (3) \end{aligned}$$

The algorithm can find all class centers in each iteration, according to the above formula. The algorithm is terminated while one of the following conditions is met. The first condition is that the iteration times over a set number. The second condition is that the information update value is lower than a fixed threshold. The third condition is that the class center remains stable in the consecutive iterative steps.

2.2. disadvantages of the AP algorithm

The AP algorithm can quickly find the reasonable cluster centers, not requiring the symmetry of the similarity matrix. As a center-based clustering method, it has a better clustering performance for the data sets with compact, ultra-spherical distribution characteristics.

For a sample of the real world, it usually obtains the characteristics of the sample through some feature extraction methods. Then the separability analysis of the sample distribution is made, in the feature space. Various feature extraction methods lead to a variety of feature spaces of the same sample. Some feature spaces can be a good demonstration of the sample distribution characteristics in the real world, but some other feature spaces are difficult to correctly reflect the sample distribution characteristics in the real world. Due to the feature extraction methods are not able to accord with the sample characteristics in the real world, the sample distribution in the feature space is not necessarily consistent with the sample distribution in the real world. The center-based clustering method uses the Euclidean distance as similarity measures, making the clustering result largely differ from the real situation, as shown in Figure 1.

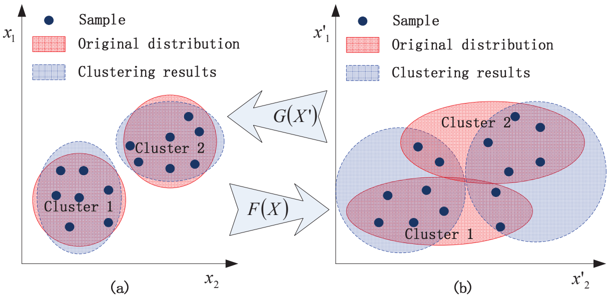

Figure 1 Clustering results in different feature spaces. (a) the distribution and clustering result of samples in the class separability feature space, (b) the distribution and clustering result of samples in the feature extraction space.

In Figure 1-(a), due to the class separability feature space can objectively reflect the distribution of the sample, it can get the correct clustering result. While in Figure 1-(b), the feature space is difficult to objectively reflect the distribution of the sample and the center-based clustering method can't adjust on this feature space, so it often obtains the incorrect clustering result.

The reason for obtaining incorrect clustering is that the similarity measure can not truly reflect the similarity of the sample in the feature space. As a center-based clustering algorithm, the AP algorithm also has such a nature. There are a variety of methods to improve the clustering performance of the AP algorithm in this case, such as establishing a soft limit strategy [11], improving the bias parameter

[9], and so on. The semi-supervised AP algorithms use the marked sample as the clustering orientation [12], making the AP algorithm can get a good clustering result for the distribution in Figure 1-(b).

The AP algorithm makes clustering based on the similarity matrix, so it is feasible to improve the algorithm in the aspect of similarity measure of the clustering. We propose the Domain Knowledge Blended the Affinity Propagation (DKB-AP) clustering algorithm, which combines the similarity measure of the AP algorithm and the priori knowledge of sample domain knowledge.

3. The Proposed Algorithm

The domain knowledge [13] refers to the problem or concept of a specialized field, as well as the relationship between them [14]. As a highly summarization of the data source and with a strongly pertinence and objectivity, the domain knowledge can be a good description of knowledge hidden within the data source, and reflect the real data mining results. The domain knowledge can be obtained from domain experts [15] or data mining methods [16,17]. It has been successfully applied in the rapid analysis of web contents [18], text classification [19], and so on. The domain knowledge of clustering data set is not utterly ignorant, so it can correct the sample distribution distortion from the feature extraction method and improve the quality of clustering [20,21], with introducing the domain knowledge into the unsupervised learning process.

In this paper, the DKB-AP clustering algorithm is proposed. The algorithm estimates the function from the feature extraction space to a separability feature space, to obtain the quantitative domain knowledge function. Then it introduces the function into the similarity measure to improve the effect of clustering.

3.1. principle of DKB-AP

For the sample $\forall x_i \in X$ (i=1,2,$\cdots$,N), there are a number of separability feature space which can correctly classify the sample set X into several subclasses, including a d-dimensional class separability feature space $\Re^d$ and a d-dimensional feature extraction space $\aleph^d$. The mapping relationship between $\Re^d$ and $\aleph^d$ is shown in Figure 1. $\Re^d$ is an implicit space, so we do not know the exact form of $\Re^d$. The goal of sample clustering is to found the regularity of samples in the space $\Re^d$. $\aleph^d$ is the dominant space, which is obtained by the feature extraction method. The DKB-AP clustering process is in the space $\aleph^d$.

If $\aleph^d$ is approximate to $\Re^d$, then the mapping $F(X)$ is a linear or nearly linear relationship. The clustering result in the space $\aleph^d$ is also good. If $\aleph^d$ is largely different from $\Re^d$, then the clustering result in the space $\aleph^d$ is not good. So it is necessary to find the approximated inverse mapping $G(X^{'})$ of mapping $F(X)$. The inverse mapping $G(X^{'})$ can approximately map the sample in the space $\aleph^d$ to the sample in the space $\Re^d$.

$\Re^d$ is an implicit space, so the mappings $F(X)$ and $G(X^{'})$ are both implicit. It needs the quantified domain knowledge of the sample space to get the mapping $G(X^{'})$.

There are two important parameters s and p in the AP algorithm. As the independent information of the data points, the bias parameter p reflects the probability of each data point was selected to the representive point. The sample little impacts on the parameter p in the different spaces $\Re^d$ and $\aleph^d$. The similarity measure s is a certain distance in the corresponding space. There are two factors influencing the parameter s. One is the sample distribution in different spaces $\Re^d$ and $\aleph^d$. The other is the computing method of the distance.

According to the equations (1), (2) and (3) in the iterative process of the AP algorithm, the similarity matrix contains similarity information between the sample pairs. So the similarity s is very important in the algorithm, which directly impacts on the ultimate iterative results. The domain knowledge $G(X^{'})$ adjusts the similarity matrix in the space $\Re^d$, making the clustering result in the space approximates to that in the space $\aleph^d$.

Taking the Euclidean distance as an example, the sample similarity in the space $\Re^d$ is described by $s(i,k) = -\|x_i - x_k\|^2$, while the adjusted similarity in the space $\aleph^d$ is described by $s^{\aleph}(i,k) = -\|G(x_i) - G(x_k)\|^2$. The current iteration times is t. In the information update process, $r^{(t)}(i,k)$ in the formula (1) is described as following.

$$r^{(t)}(i,k) = (1-\lambda) \times (s^{\aleph}(i,k) - \max_{k^{'} \neq k}\{a^{(t-1)}(i,k) + s^{\aleph}(i,k^{'})\}) \\ + \lambda \times r^{(t-1)}(i,k) \qquad (4)$$

The formula (2) and (3) is got based on the formula (1), so their forms are unchanged when $s(i,k)$ is an implicit function.

3.2. algorithm steps

DKB-AP is an improved AP algorithm based on adjusting the the similarity measure. The steps of this algorithm are as follows.

Step 1: It estimates the domain knowledge function $G(X^{'})$, according to the sample distribution in the feature extraction space $\aleph^d$.

Step 2: It calculates the adjusted similarity matrix $S^{\aleph} = [s^{\aleph}(i,k)]_{n\times n}$ in the feature extraction space $\aleph^d$. Which n is the number of sample points, the diagonal elements of $S^{\aleph}$ are $S^{\aleph}(k,k) = p,\ p < 0$. The initialized values $a^{(0)}(i,k) = 0$, $r^{(0)}(i,k) = 0$.

Step 3: Information updated. It updates $r(i,k)$ and $a(i,k)$ according to the formulas (2), (3) and (4). It calculates the information of all data points to find the class center point of each point. The conditions of the algorithm termination are the same as the AP algorithm.

Step 4: If the number of cluster centers does not meet the requirements, then the vaule of p changes. It repeats the iteration process until the number of clusters to meet the requirements. Then, the final clustering result is output.

4. Experiments

Taking the two-dimensional separability feature space $\Re^2$ as an example, we make the randomly generated normally distributed samples as the sample sets. The sample set Ran_set1 is generated based on the centers (-6, -6), (-6,6), (0, 0), (5,5), (5, -5). The sample set Ran_set2 is generated based on the centers (-3, -3), (0,0), (3,3). The sample set Ran_set3 is generated based on the centers (-5,5), (0,0), (5,5). They are as shown in Table 1.

Table 1 Experimental sample sets for DKB-AP

Sample sets	Dimensions	The number of samples	The number of classes
Ran_set1	2	150	5
Ran_set2	2	150	3
Ran_set3	2	300	3

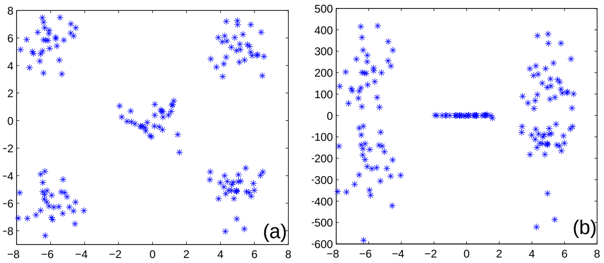

Figure 2 The distribution of the sample set Ran_set1 in different feature spaces. (a) In the space $\Re^d$, (b) In the space $\aleph^d$.

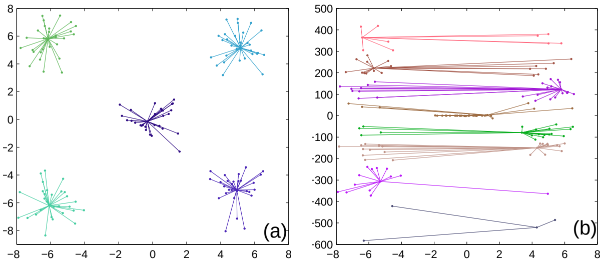

Figure 3 The clustering results of the sample set Ran_set1 in the spaces $\Re^2$ and $\aleph^2$ with the AP algorithm. (a) In the space $\Re^2$, it gets the correct clustering result, with 5 clusters. (b) In the space $\aleph^2$, the clustering result is 8 clusters.

In the feature extraction space $\aleph^2$, it assumes that the mapping between samples in the space $\Re^2$ and samples in the space $\aleph^2$ is $F(X)$: $x_1^{'} = x_1$, $x_2^{'} = ax_2^c + b$. Which $a, b \in R$, $a \neq 0$, $c = 2k-1$, $k \in N$. Taking the sample set Ran_set1 as an example, the distribution of the sample set in the space $\Re^2$ and $\aleph^2$ are as shown in Figure 2 when a=1, b=0, c=3.

In Figure 2-(a), the sample set Ran_set1 has obvious clustering regularity in the space $\Re^2$. The samples of different subclasses cluster well, with 5 subclasses. There are a variety of clustering methods to cluster correctly. In Figure 2-(b), due to the two different dimension scales in the space $\aleph^2$, the sample distribution distorts, with 3 subclasses. It can't reflect the real distribution of sample sets in new feature spaces. In the real clustering process, the reasonable number of subclasses may become larger, so it is difficult to find the right clustering regularity.

It clusters the sample set Ran_set1 in the spaces $\Re^2$ and $\aleph^2$ with the AP algorithm, adopting the negative squared Euclidean distance $s(i,j) = -[(x_1^i - x_1^j)^2 + (x_2^i - x_2^j)^2]$ as the similarity measure between two samples (x_1^i, x_2^i) and (x_1^j, x_2^j), $i, j = 1, 2, \cdots, n$. The clustering results are as shown in Figure 3, whose evaluation index are Fowlkes-Mallows Validity Index and Error Rate for the Set.

In Figure 3-(a), the AP algorithm gets the correct clustering result, with that the number of clusters is 5, the Fowlkes-Mallows Validity Index is 1.00, and the Error Rate for the Set is 0.00%. The result indicates that the AP algorithm can get the correct clustering result in the space $\Re^2$. In Figure 3-(b), the number of clusters is 8, and the samples are banded structure, making the incorrect clustering result. The Fowlkes-Mallows Validity Index is 0.53, and the Error Rate for the Set is 73.33%. The result indicates that the AP algorithm can't get the correct clustering result in the space $\aleph^2$. The main reason is that the distribution of the sample set Ran_set1 in the space $\aleph^2$ can't reflect its real distribution regularity.

The DKB-AP algorithm clusters the sample sets in the space $\aleph^2$. It needs to estimate the domain knowledge function $F(X)$. In the experiment, the function $G(X^{'})$ is known, so the ideal domain knowledge function is described as $G(X^{'}) x_1 = x_1^{'}$, $x_2 = \sqrt[c]{(x_2^{'} - b)/a}$. The similarity measure between the samples $(x_1^{'i}, x_2^{'i})$ and $(x_1^{'j}, x_2^{'j})$ $(i, j = 1, 2, \cdots, n)$ is described as $s^{\aleph} = -[(x_1^{'i} - x_1^{'j})^2 + (\sqrt[c]{(x_2^{'i} - b)/a} - \sqrt[c]{(x_2^{'j} - b)/a})^2]$. The clustering result of the sample sets by DKB-AP algorithm is as shown in Figure 4.

In Figure 4, the DKB-AP algorithm gets the clustering result of sample set Ran_set1 in the space $\aleph^2$, with that the number of clusters is 5, the Fowlkes-Mallows Validity Index is 1.00, and the Error Rate for the Set is 0.00%. The result indicates that the DKB-AP algorithm can get the correct clustering result in the space $\aleph^2$. The result of

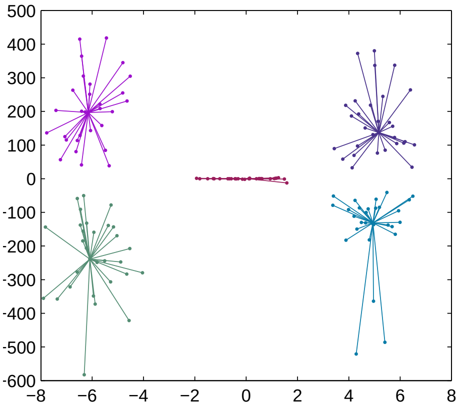

Figure 4 The clustering result of sample set Ran_set1 in the space $\aleph^2$ by DKB-AP algorithm.

sample set Ran_set1 in the space $\aleph^2$ by the DKB-AP algorithm is almost consistent with the result of sample set Ran_set1 in the space $\Re^2$ by the AP algorithm. This indicates that the DKB-AP algorithm can reflect the real distribution regularity of the sample set Ran_set1 in the space $\aleph^2$.

The experimental results of sample sets Ran_set2 and Ran_set3 are as shown in Figure 5 and Figure 6.

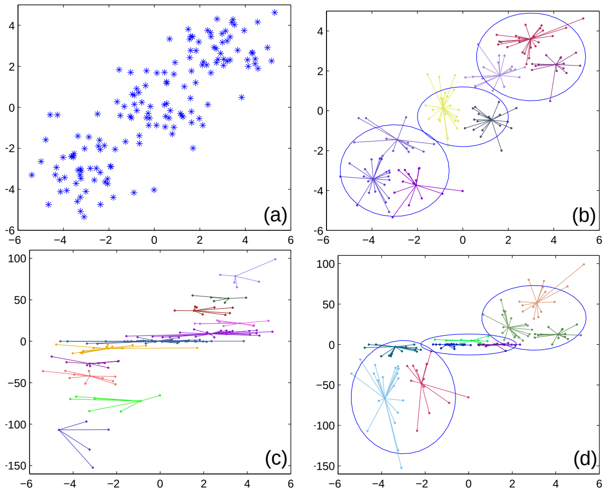

Figure 5 The clustering results of sample set Ran_set2. (a) the distribution of the sample set in the space $\Re^2$, whose subclass clustering regularity is weaker than that of Ran_set1. (b) the clustering result of sample set in the space $\Re^2$ by the AP algorithm, with 8 clusters. (c) the clustering result of sample set in the space $\aleph^2$ by the AP algorithm, with 11 clusters. (d) the clustering result of sample set in the space $\aleph^2$ by the DKB-AP algorithm, with 8 clusters.

In Figure 5-(a), the clustering regularity of the sample set in the space $\Re^2$ is not very obvious. In Figure 5-(b), the number of clusters is 8, the Fowlkes-Mallows Validity Index is 0.56, and the Error Rate for the Set is 85.33%. Although the clustering result of sample set in the space $\Re^2$ by the AP algorithm is not good, each subclass clusters together. In Figure 5-(c), the number of clusters is 11, the Fowlkes-Mallows Validity Index is 0.52, and the Error Rate for the Set is 93.33%. The clustering result is so bad, and can't show the clustering regularity. In Figure 5-(d), the number of clusters is 8, the Fowlkes-Mallows Validity Index is 0.54, and the Error Rate for the Set is 86%. The clustering result is almost consistent with the result in Figure 5-(b).

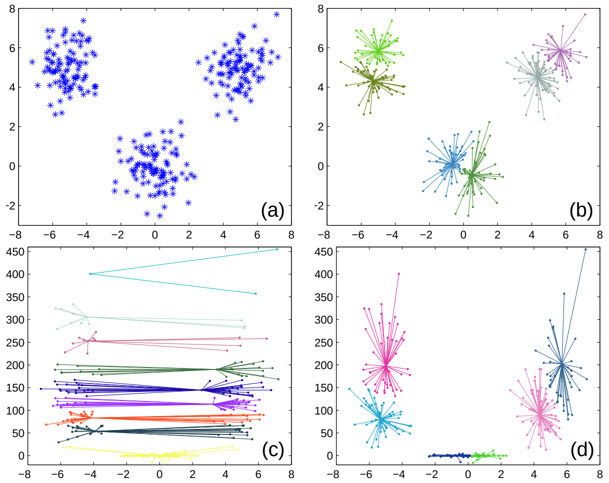

Figure 6 The clustering results of sample set Ran_set3. (a) the distribution of the sample set in the space $\Re^2$, whose subclass clustering regularity is weaker than that of Ran_set1. (b) the clustering result of sample set in the space $\Re^2$ by the AP algorithm, with 6 clusters. (c) the clustering result of sample set in the space $\aleph^2$ by the AP algorithm, with 9 clusters. (d) the clustering result of sample set in the space $\aleph^2$ by the DKB-AP algorithm, with 6 clusters.

In Figure 6-(a), the clustering regularity of the sample set in the space $\Re^2$ is very obvious. In Figure 6-(b), the number of clusters is 6, the Fowlkes-Mallows Validity Index is 0.71, and the Error Rate for the Set is 85.33%. The clustering result of sample set in the space $\Re^2$ by the AP algorithm is not good. The reason is that the variance of the sample subclass is relatively large, and the same class is divided into two subclasses. However, the different classes have not been clustered incorrectly. In Figure 6-(c), the number of clusters is 9, the Fowlkes-Mallows Validity Index is 0.58, and the Error Rate for the Set is 92.67%. The clustering result is so bad, and can't show the clustering regularity. In Figure 6-(d), the number of clusters is 6, the Fowlkes-Mallows Validity Index is 0.71, and the Error

Rate for the Set is 85.33%. The clustering result is very consistent with the result in Figure 6-(b).

For the sample sets Ran_set2 and Ran_set3, the clustering result in the space $\aleph^2$ by the DKB-AP algorithm is not the same as the clustering result in the space $\Re^2$ by the AP algorithm, but it is also better than the clustering result in the space $\aleph^2$ by the AP algorithm.

The Fowlkes-Mallows Validity Index of three sample sets by the AP algorithm and DKB-AP algorithm are as shown in Table 2. And the Error Rate for the Set of three sample sets by the AP algorithm and DKB-AP algorithm are as shown in Table 3.

Table 2 The Fowlkes-Mallows Validity Index of three sample sets by AP and DKB-AP

Sample sets	AP+ $\Re^2$	AP+ $\aleph^2$	$\mathrm{DKB}-\mathrm{AP}+$ $\aleph^2$
Ran_set1	1.0	0.53	1.0
Ran_set2	0.56	0.52	0.54
Ran_set3	0.71	0.57	0.71

Table 3 The Error Rate for the Set of three sample sets by AP and DKB-AP

Sample sets	AP+ $\Re^2$	AP+ $\aleph^2$	$\mathrm{DKB}-\mathrm{AP}+$ $\aleph^2$
Ran_set1	0.00%	73.33%	0.00%
Ran_set2	85.33%	93.33%	86.00%
Ran_set3	85.33%	92.67%	85.33%

From the Table 2 and Table 3, we can draw a conclusion that the clustering result in the space $\aleph^2$ by DKB-AP algorithm is almost consistent with the clustering result in the space $\Re^2$ by AP algorithm, when the distribution of the sample set in the space $\aleph^2$ does not reflect the real distribution regularity.

5. Conclusion

The distribution of samples in the weaker separability feature space can not reflect the real distribution regularity, so the AP algorithm is difficult to obtain the satisfactory clustering result. In this paper, the DKB-AP algorithm is proposed, which is an improved AP algorithm based on adjusting the the similarity measure. According to the distribution of samples in the feature space, the proposed algorithm estimates the domain knowledge mapping from this feature space to a separability feature space. And the domain knowledge mapping is introduced to the similarity measure of the AP algorithm. Then, the algorithm gets the clustering result with an iteration process. In the experiment, the test sample sets are three randomly generated sample sets. The subclass aggregation degree of two sample sets is good, and the subclass aggregation degree of another sample set is weak. The clustering results and evaluation index are analysed. We can draw a conclusion that the clustering result in the weaker separability feature space by DKB-AP algorithm is almost consistent with the clustering result in the separability feature space by AP algorithm.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (Grant No. 51104157), the Fundamental Research Funds for the Central Universities (Grant No. 2011QNA30), the Ph.D. Programs Foundation of Ministry of Education of China (Grant No. 20110095120008), the China Postdoctoral Science Foundation (Grant No. 20100481181) and Jiangsu Overseas Research & Training Program for University Prominent Young & Middle-aged Teachers and Prsesidents.

References

[1] R. Xu and D. Wunsch, Survey of clustering algorithms, IEEE Transactions on Neural Networks, **16**, 645-678 (2005).

[2] A. K. Jain, Data clustering: 50 years beyond K-means, Pattern Recognition Letters, **31**, 651-666 (2010).

[3] B. J. Frey and D. Dueck, Clustering by passing messages between data points, Science, **315**, 972-976 (2007).

[4] I. E. Givoni and B. J. Frey, A Binary Variable Model for Affinity Propagation, Neural Computation, **21**, 1589-1600 (2009).

[5] F. H. Shang, L. C. Jiao, J. R. Shi, F. Wang, and M. G. Gong, Fast affinity propagation clustering: A multilevel approach, Pattern Recognition, **45**, 474-486 (2012).

[6] C. Furtlehner, M. Sebag, and X. L. Zhang, Scaling analysis of affinity propagation, Physical Review E, **81**, (2010).

[7] R. C. Guan, X. H. Shi, M. Marchese, C. Yang, and Y. C. Liang, Text Clustering with Seeds Affinity Propagation, IEEE Transactions on Knowledge and Data Engineering, **23**, 627-637 (2011).

[8] H. B. Yang and X. Hou, Video People Counting Using Trajectories Affinity Propagation Clustering, Journal of Nanoelectronics and Optoelectronics, **7**, 186-190 (2012).

[9] H. J. Su, Y. H. Sheng, P. J. Du, and K. Liu, Adaptive affinity propagation with spectral angle mapper for semi-supervised hyperspectral band selection, Applied Optics, **51**, 2656-2663 (2012).

[10] J. Zhang, X. G. Tuo, Z. Yuan, W. Liao, and H. F. Chen, Analysis of fMRI Data Using an Integrated Principal Component Analysis and Supervised Affinity Propagation Clustering Approach, IEEE Transactions on Biomedical Engineering, **58**, 3184-3196 (2011).

[11] M. Leone, Sumedha, and M. Weigt, Clustering by soft-constraint affinity propagation: applications to gene-expression data, Bioinformatics, **23**, 2708-2715 (2007).

[12] M. Leone, Sumedha, and M. Weigt, Unsupervised and semi-supervised clustering by message passing: soft-constraint affinity propagation, European Physical Journal B, **66**, 125-135 (2008).

[13] M. D'Hondt and T. D'Hondt, Is Domain Knowledge an Aspect?, Lecture notes in Computer science, 293-293 (1999).

[14] M. D'Hondt, W. De Meuter, and R. Wuyts, Using reflective logic programming to describe domain knowledge as an aspect, Proc. the Generative and Component-Based Software Engineering, (2000).

[15] D. Z. Hambrick and R. W. Engle, Effects of domain knowledge, working memory capacity, and age on cognitive performance: An investigation of the knowledge-is-power hypothesis, Cognitive Psychology, **44**, 339-387 (2002).

[16] T. Yu, S. Simoff, and T. Jan, VQSVM: A case study for incorporating prior domain knowledge into inductive machine learning, Neurocomputing, **73**, 2614-2623 (2010).

[17] J. Wang, Y. P. Wu, X. N. Liu, and X. Y. Gao, Knowledge acquisition method from domain text based on theme logic model and artificial neural network, Expert Systems with Applications, **37**, 267-275 (2010).

[18] T. Willoughby, S. A. Anderson, E. Wood, J. Mueller, and C. Ross, Fast searching for information on the Internet to use in a learning context: The impact of domain knowledge, Computers & Education, **52**, 640-648 (2009).

[19] Z. T. Yy, L. Han, J. Y. Guo, X. Y. Meng and Z. K. Zhang, Study on the construction of domain text classification model with the help of domain knowledge, Proc. 2008 International Conference on Machine Learning and Cybernetics, 1-7, (2008).

[20] A. R. Sinha and H. M. Zhao, Incorporating domain knowledge into data mining classifiers: An application in indirect lending, Decision Support Systems, **46**, 287-299 (2008).

[21] X. G. Hu, X. F. Hu, D. X. Wang, D. Y. Zhang and C. L. Hu, A classification algorithm based on multi-relation domain knowledge, Proc. 2005 International Conference on Machine Learning and Cybernetics, 1-9, (2005).

Wei Chen received the B.Eng. Degree in medical imaging and the M.S. degree in paleontology and stratigraphy from China University of Mining and Technology, Xuzhou, China, in 2001 and 2005, respectively, and the Ph.D degree in communications and information systems from China University of Mining and Technology, Beijing, China, in 2008. In 2008, he joined the School of Computer Science and Technology, China University of Mining and Technology, where he is currently an Associate Professor. His research interests include machine learning, image processing, and wireless communications.

Qichong Tian received the B.S. degree in electronic information science and technology from China University of Mining and Technology, Xuzhou, China, in 2010. Currently, he is pursuing the M.S. degree in communication and information systems at Huazhong University of Science and Technology, Wuhan, China. His research interests include image processing, machine learning, and wireless communications. He is a Student Member of IEEE and China Computer Federation.

Xiaorong Jiang a student in China University of Mining and Technology, Xuzhou, China. He is studying Coal Mining Engineering in School of Computer Science and Technology. He received her Bachelor's degree in Mining Engineering from China University of Mining and Technology in 2012.

Zhibo Tang is a student in China University of Mining and Technology, Xuzhou, China. He is studying Electronic Information Science and Technology in School of Computer Science and Technology. He received her Bachelor's degree in Electronic Information Science and Technology from China University of Mining and Technology in 2012.

Caihua Guo is a student in China University of Mining and Technology, Xuzhou, China. She is studying Electronic Information Science and Technology in School of Computer Science and Technology. She received her Bachelor's degree in Electronic Information Science and Technology from China University of Mining and Technology in 2012.

Xinzheng Xu received his Ph.D.degree in Computer Application Technology from China University of Mining and Technology in 2012, and his MS degree in Computer Application Technology from Xiamen University in 2005, and his BS degree in electrical engineering from Shandong University of Science and Technology in 2002. He is currently a lecturer at School of Computer Science and Technology, China University of Mining and Technology. He is a member of China Computer Federation, and China Association for Artificial Intelligence. His research interests include intelligent information processing, pattern recognition, machine learning, and granular computing et al.

Hong Zhu (1970-), Ph.D. candidate, associate professor, her interesting field includes granule computing, clustering, parallel computing et al.

Shifei Ding is currently a professor at School of Computer Science and Technology, China University of Mining and Technology. He is a member of China Computer Federation, and China Association for Artificial Intelligence. His research interests include granular computing, machine learning, and et al.