A New Simplified Gravitational Clustering Method for Multi-prototype Learning Based on Minimum Classification Error Training

Teng Long and Lian-Wen Jin

Department of Electronic and Communication Engineering,
South China University of Technology,
510640 Guangzhou, China
{tenglong, eelwjin}@scut.edu.cn

Abstract. In this paper[1], we propose a new simplified gravitational clustering method for multi-prototype learning based on minimum classification error (MCE) training. It simulates the process of the attraction and merging of objects due to their gravity force. The procedure is simplified by not considering velocity and multi-force attraction. The proposed hierarchical method does not depend on random initialization and the results can be used as better initial centers for K-means to achieve higher performance under the SSE (sum-squared-error) criterion. The experimental results on the recognition of handwritten Chinese characters show that the proposed approach can generate better prototypes than K-means and the results obtained by MCE training can be further improved when the proposed method is employed.

1 Introduction

Many real world problems of pattern classification are non-linear separable. Multi-prototype learning and classification can solve many such problems by forming complex boundary for each class of patterns. It is especially suitable for recognition of handwritten characters because characters of one category are usually written in different styles by different people [2]. The recognition accuracy can be improved significantly when multiple prototypes are well designed and a multi-prototype minimum distance classifier is employed [1][2]. Prototype selection by hand is not guaranteed to build an optimal prototype set and it is not practical for recognition of Chinese characters of which the number of categories is more than 3,000. So it is necessary to make effort on automatic prototype learning to build optimal prototype sets for classifiers. As a well known statistical clustering technique, K-means [3] is usually used to build the multiple prototypes [2][9]. The prototypes obtained by K-means can be further fine tuned to achieve much higher recognition accuracy by some prototype learning methods such as learning vector quantization (LVQ) [4] and minimum classification

[1] The paper is sponsored by New Century Excellent Talent Program of MOE (No.NCET-05-0736), NSFGD (No.04105938).

N. Zheng, X. Jiang, and X. Lan (Eds.): IWICPAS 2006, LNCS 4153, pp. 168–175, 2006.

error (MCE) training [5][6]. The empirical rules of LVQ have not enough of a mathematical basis to guarantee design optimality and the convergence mechanism has not been mathematically elaborated [7]. The MCE training which has quite similar rules as an improved version of LVQ aims at minimizing a smooth approximation function of the error rate. It can be used to adjust the initial prototype set iteratively under the minimum classification error criterion to generate high quality prototypes [6]. However, does the selection of the initial prototype set affect the local minimum finally converged by MCE training?

In this paper, we introduce a new simplified gravitational clustering method for multi-prototype learning based on MCE training. It simulates the process of the attraction and merging of objects due to their gravity force. The procedure is simplified by not considering velocity and multi-force attraction. The proposed hierarchical clustering method does not depend on thresholds which are usually required by agglomerative hierarchical clustering and density-based clustering [11]. And it does not have random initialization problems which may lead to incorrect results in K-means algorithm. The method gives not only a reasonable clustering result but also better initial centers for K-means to achieve higher performance under SSE (sum-squared-error) criterion. It can be used to generate initial prototypes for the multi-prototype learning based on MCE. Experiments were carried out on the recognition of handwritten Chinese characters and proved the efficiency of our method. The recognition performance of the 4-prototypes template generated by the proposed hybrid method is even higher than the 8-prototypes template generated by traditional K-means method. The results also indicate that when the initial prototype set is improved by our method the fine tuned prototype set obtained by MCE training achieves better performance as well.

2 A New Simplified Gravitational Clustering

The gravitational clustering algorithm was first proposed in [10], and has been discussed in a recent paper [8]. It iteratively simulates the movement of each object due to the gravity force during a time interval and check for possible merge. As it simulates the whole physical process, the velocity of each object needs to be recalculated after each time interval based on a co-efficient of the air resistance and the vector sum of the gravity forces, which the object experiences from all other objects remaining in the physical system [8].

We simplify the process by making an assumption that if each time only one pair of objects which are likely to meet and merge first are freed to move and merged, at the same time other objects are fixed and not affected by the movement and mergence, the final clustering result can still well describe the characteristics of the spatial distribution of the objects.

The simplified gravitational clustering (SGC) is performed as follows:

Step 1. Let $\{X_1, X_2, \ldots, X_N\}$ be a set S of N objects on D dimensions. Set all objects' initial mass as:

$$m_i = 1, i = 1, 2, \ldots, N. \tag{1}$$

Step 2. Find the pair of objects which are most likely to meet and merge first by the following equation:

$$\{X_i, X_j\} \ \textit{if} \ \{i, j\} = \arg\min_{i,j}(\|X_i - X_j\| \times \frac{m_i + m_j}{2}) \tag{2}$$

Step 3. Merge the pair of objects to generate a new object. The mass of the new object is given by:

$$m_t = m_i + m_j \tag{3}$$

And the position of the new object is the centroid of the pair of objects:

$$X_t = \frac{m_i X_i + m_j X_j}{m_i + m_j} \tag{4}$$

Step 4. Add the new object X_t to the set S and delete the objects X_i and X_j from it.

Step 5. Terminate if the number of objects in the set S reaches k which is the desired number of clusters, otherwise, go to step 2.

The final remaining objects can represent the clusters by their positions. It can be easily proved that their positions are the centroids of all the objects merged to them no matter what the merging sequence is. Thus if the cluster number k is 1, the result is the centroid of all objects which is the same as the K-means algorithm.

The eq. (2) is important which determines the merging sequence. It is defined by the assumption that heavier objects move together more slowly. This makes the cluster centers distributed more equally and avoid to get centers at the outliers because of some lonely points, which is also a drawback of the K-means algorithm. It is not the same as the traditional gravitational clustering which is based on gravity theory [8], in which heavier objects have stronger gravity forces and move together earlier than light objects at the same distance. Experiments convinced us the measurement employed in eq. (2) is a good and robust choice for clustering and multi-prototype learning as well.

Because the new object generated by the mergence is located at the centroid of all the objects merged to it, the results should be the same as the K-means algorithm if the final partition of the objects given by the SGC is the same as the partition which the K-means gives. However, the SGC method does not guarantee the SSE criterion. So it can be combined with the K-means algorithm. When the centers obtained from the SGC are used as the initial centers for the K-means, the clustering results get better under the SSE criterion. The Fig. 1 shows an example, in which the cluster centers obtained by three methods, K-means, SGC and the combined are given. From the results, it can be seen that the traditional K-means algorithm with random initialization can lead to incorrect clustering results. It is also shown that the results given by the SGC are slightly different from the K-means after combined. This leads to fewer iterations of K-means in the hybrid method.

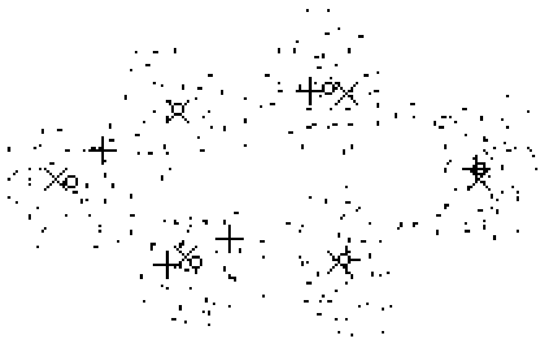

Fig. 1. An example of the clustering results ("+" represents the centers obtained by the K-means, "×" for the SGC method and "∘" for the combined)

In Fig. 2, a real world example is shown. The data points are the projection of the training samples' LDA-based features of the handwritten Chinese character "王" on the first two dimension plain. The training samples are used in the experiments in section 4. As the number of clusters is fixed, the clustering method for the prototype generation should be able to generate descriptive prototypes for the spatial distribution of the sample points while the number of clusters is not an optimal one. The example shows that with some random initialization, the K-means generate much less descriptive prototypes than the proposed method.

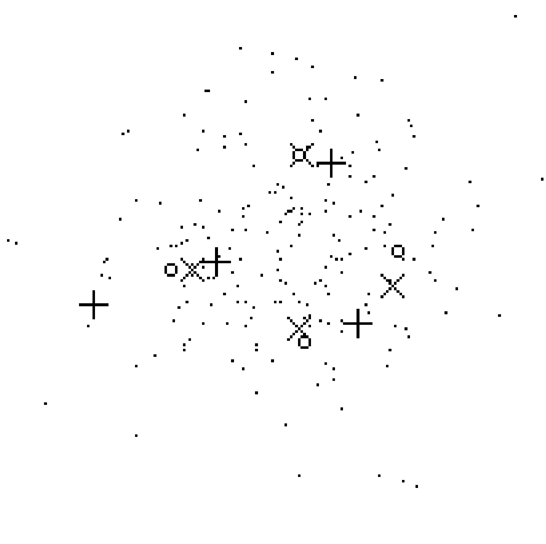

Fig. 2. A real world example of clustering results ("+" represents the centers obtained by the K-means, "×" for the SGC method and "∘" for the combined)

3 Multi-prototype Learning Based on Minimum Classification Error Training

We employed the same LDA-based Gabor feature extraction as which is discussed in [6] for the recognition of handwritten Chinese characters. All the prototypes are represented by 256-dimensional LDA-based feature vectors. The clustering methods are used to generate the multi-prototype template from the training samples. In this paper, the three clustering methods, K-means, SGC and the combined method are used to generate the prototypes respectively and the performances are compared. All of them are unsupervised clustering methods.

As the prototype learning is a supervised process, the supervised clustering methods such as LVQ and MCE can achieve better performance than the unsupervised clustering methods for prototype generation. We employed the MCE training technique as the multi-prototype learning method to fine tune the prototype sets obtained from the unsupervised clustering. It is the same as [6] but the learning strategy is modified. In our MCE training, all the training samples are used to update the prototypes no matter the samples are correctly recognized or not. In our experiment, this strategy performed better than the one used in [6].

4 Experiments

Several experiments were performed on the recognition of handwritten Chinese characters. All the experiments were based on the recognition of 3755 categories of level 1 Chinese characters in GB2312-80 which is a national Chinese character set standard. We randomly selected 250 samples for each category from the China 863 National Handwriting Database HCL2000. 200 samples among them were used as training set and other 50 samples formed the testing set, i.e. 751,000 samples for train-ing and 187,750 samples for testing in total.

In the first experiment, we tested the recognition performance on the templates of different number of prototypes generated by different clustering methods. The processing time was about 40, 175 and 200 seconds for the K-means, SGC and the combined method respectively when building the 4-prototypes template using the C++ programming language compiled program. The computation environment was on a PC with an Intel P4 3.0G CPU and 512M memory. The curves of the results are shown in Fig. 3. The results show it clearly that the multi-prototype classifier is much better than the single-prototype one. The recognition accuracy can be improved more than 2 percent by using an 8-prototypes template generated by the proposed method. They also indicate that by using the combined clustering method, the performance of 4-prototypes template is even better than which of 8-prototypes template generated by the traditional K-means method.

In the second experiment, we tried to find out how the performance is affected by the different tries of random initial points for K-means algorithm. The recognition results are listed in Table 1. By choosing the best result of 8-prototypes template obtained by K-means, the recognition rate of 90.21% is still lower than the rate of 4-prototypes template generated by the combined method, which is 90.22%. This indicates that not only the storage of the templates but also the recognition time can be saved 50% to achieve the same performance by using the proposed hybrid method.

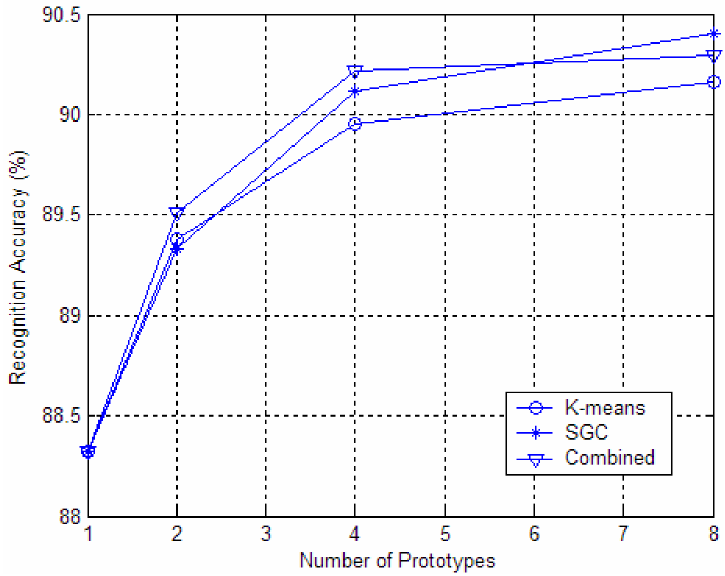

Fig. 3. Recognition performance of the different number of prototypes templates generated by the different clustering methods

Table 1. Recognition accuracy (%) of using K-Means with 10 different random initializations

Number of prototypes	Recognition accuracy by 10 different random initializations	Mean
4	89.94, 90.00, 89.97, 90.00, 89.91, 90.04, 89.96, 89.96, 89.96, 89.95	89.97
8	90.18, 90.18, 90.15, 90.19, 90.20, 90.10, 90.21, 90.20, 90.14, 90.17	90.17

In the third experiment, the MCE training was performed to fine tune the 4-prototypes template. Fig. 4 shows the learning curves for MCE training on open-test. It indicates that the recognition rate can be improved another 2 percent by using the MCE training. It also shows that the local minimum converged by MCE training can be improved by choosing a better initial prototype set. From the figure, it seems that the prototype sets obtained by the SGC and the combined method converged to the same local minimum even the performances of them are different at the beginning.

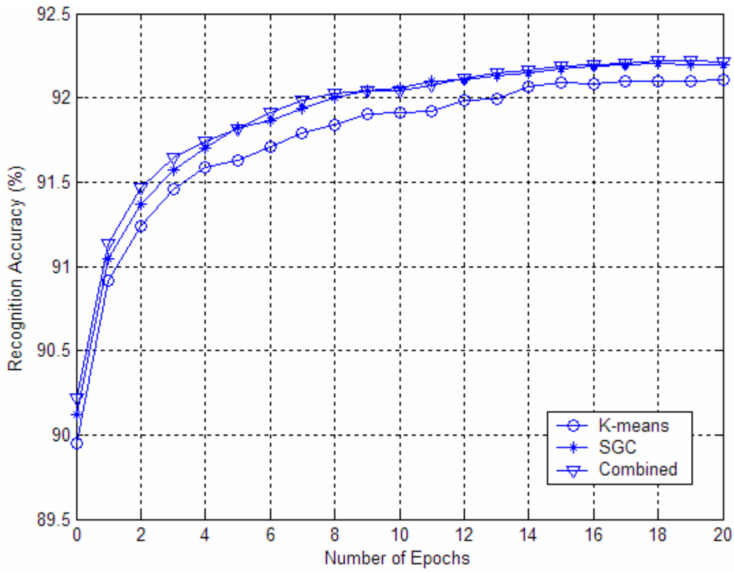

Fig. 4. Learning curves for MCE training on open-test: recognition accuracies (%) as a function of epoch number by the different initial templates obtained by the different clustering methods

5 Conclusion

In this paper, we have introduced a new simplified gravitational clustering method for multi-prototype learning based on MCE training. It's a hierarchical agglomerative clustering method which does not depend on thresholds or random initialization. The main idea is to find equally distributed centers to better describe the spatial distribution of the sample data by using a modified distance. The results of the proposed method can be used as the initial centers for K-means to achieve higher performance under the SSE criterion. Experiments on the recognition of handwritten Chinese characters proved the efficiency of the proposed method. As the number of the prototypes can be reduced to achieve the same performance, the proposed technique saves not only the storage of the templates but also the computation time of the recognition. The experiments also showed us an interesting result that the prototypes obtained by MCE training can be further improved by choosing a better initial prototype set.

References

1. Liu, C.-L., Jaeger, S., and Nakagawa, M.: Online Recognition of Chinese Characters: The State-of-the-Art. IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 26, No. 2, (2004) 198–213

2. Rahman, A.F.R., Fairhurst, M.C.: Multi-prototype Classification: Improved Modeling of the Variability of Handwritten Data using Statistical Clustering Algorithms. Electronics Letters, Vol. 33, No. 14, (1997) 1208–1210
3. Kanungo, T., et al: An Efficient k-Means Clustering Algorithm: Analysis and Implementation. IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 24, No. 7, (2002) 881–892
4. Kohonen, T.: The Self-Organizing Map. IEEE Proceedings, Vol. 78, No. 9, (1990) 1464–1480
5. Juang, B.-H. and Katagiri, S.: Discriminative Learning for Minimum Error Classification. IEEE Trans. on Signal Processing, Vol. 40, (1992) 3043–3054
6. Huo, Q., Ge, Y., Feng, Z.-D.: High Performance Chinese OCR Based on Gabor Features, Discriminative Feature Extraction and Model Training. Proc. ICASSP 2001, Vol. 3, (2001) 1517–1520
7. Katagiri, S., Juang, B.-H., Lee, C.-H.: Pattern Recognition Using a Family of Design Algorithms Based Upon the Generalized Probabilistic Descent Method. IEEE Proceedings, Vol. 86, No. 11, (1998) 2345–2373
8. Chena, C.-Y., Hwanga, S.-C., Oyanga, Y.-J.: A statistics-based approach to control the quality of subclusters in incremental gravitational clustering. Pattern Recognition, Vol. 38, (2005) 2256–2269
9. Wang, Q., et al: Match between Normalization Schemes and Feature Sets for Handwritten Chinese Character Recognition. Proc. ICDAR 2001, (2001) 551–555
10. Wright, W.E.: Gravitational Clustering. Pattern Recognition, Vol. 9 (1997) 1149–1160
11. Berkhin, P.: Survey of Clustering Data Mining Techniques. Technical Report, Accrue. Software Inc., San Jose, CA, USA, 2002