Semi-supervised Kernel-Based Fuzzy C-Means

Daoqiang Zhang, Keren Tan, and Songcan Chen

Department of Computer Science and Engineering
Nanjing University of Aeronautics and Astronautics
Nanjing 210016, P.R. China
{dqzhang,s.chen}@nuaa.edu.cn

Abstract. This paper presents a semi-supervised kernel-based fuzzy c-means algorithm called S^{2}KFCM by introducing semi-supervised learning technique and the kernel method simultaneously into conventional fuzzy clustering algorithm. Through using labeled and unlabeled data together, S^{2}KFCM can be applied to both clustering and classification tasks. However, only the latter is concerned in this paper. Experimental results show that S^{2}KFCM can improve classification accuracy significantly, compared with conventional classifiers trained with a small number of labeled data only. Also, it outperforms a similar approach S^{2}FCM.

1 Introduction

Recently, semi-supervised learning has attracted much attention in machine learning community. One reason is that in many learning tasks, there is a large supply of unlabeled data but insufficient labeled data because the latter is much more expensive to obtain than the former. In other words, labeled data is accurate but the number is few, and unlabeled data is not accurate whereas their amount is huge. To break away from that dilemma, semi-supervised learning combines labeled and unlabeled data together during training to improve performance. Typically, semi-supervised learning is applicable to both clustering and classification. In semi-supervised clustering, some labeled data is used along with the unlabeled data to obtain a better clustering. However, in semi-supervised classification, additional unlabelled data are exploited with labeled data to obtain a good classification function. A lot of semi-supervised learning algorithms have been proposed to date [1]-[4]. Among them, most semi-supervised clustering algorithms originate from classical clustering algorithms and are for clustering tasks, whereas most semi-supervised classification algorithms originate from classical classification algorithms and are for classification tasks.

In this paper, we present a semi-supervised kernel-based fuzzy c-means algorithm called S^{2}KFCM, which is based on our previously proposed kernel-based fuzzy c-means clustering algorithm (KFCM) [5][6]. S^{2}KFCM is the semi-supervised KFCM, and here our goal is to use S^{2}KFCM not for clustering but for classification tasks. We made comparisons between S^{2}KFCM and classical classifiers trained with a small number of labeled data, e.g. *k*-nearest neighbor classifier (*k*-NN) [7] and support vector machines (SVM) [7]. Comparisons of classification performances are also made between S^{2}KFCM and another similar algorithm S^{2}FCM which originated from

N.R. Pal et al. (Eds.): ICONIP 2004, LNCS 3316, pp. 1229–1234, 2004.

fuzzy c-means algorithm (FCM) and was for clustering. Experimental results demonstrate the advantages of the proposed approach over other algorithms. In Section 2, we first review the KFCM algorithm. In Section 3 the detailed S^2KFCM algorithm is proposed. Section 4 presents the experimental results. Conclusions are made in Section 5.

2 KFCM

Given $X=\{x_1,\ldots,x_n\}$ where x_k in R^s, the original FCM algorithm partitions X into c fuzzy subsets by minimizing the following objective function [8]

$$J_m(U,V)=\sum_{i=1}^{c}\sum_{k=1}^{n}u_{ik}^m \parallel x_k-v_i \parallel^2 . \tag{1}$$

Here c is the number of clusters, n is the number of data points, and u_{ik} is the membership of x_k in class i takes value in the interval [0,1] such that $\sum_i u_{ik}=1$ for all k. By optimizing the objective function of FCM, one can obtain the alternate iterative equations.

By using the popular 'kernel method', we constructed a kernel version of FCM in our early works, where the original Euclidian distance in FCM is replaced with the following kernel-induced distance measures [5]

$$d(x,y)=\parallel \Phi(x)-\Phi(y) \parallel=\sqrt{K(x,x)-2K(x,y)+K(y,y)} . \tag{2}$$

Here Φ is a nonlinear function mapping x_k from the input space X to a new space F with higher or even infinite dimensions. $K(x,y)$ is the kernel function which is defined as the inner product in the new space F with: $K(x,y) = \Phi(x)^{\mathrm{T}}\Phi(y)$, for x, y in input space X.

An important fact about kernel function is that it can be directly constructed in original input space without knowing the concrete form of Φ. That is, a kernel function implicitly defines a nonlinear mapping function. There are several typical kernel functions, e.g. the Gaussian kernel: $K(x,y)=exp(-\|x-y\|^2/\sigma^2)$, and the polynomial kernel: $K(x,y)=(x^{\mathrm{T}}y + 1)^d$. Especially for Gaussian kernel, we have $K(x,x)=1$ for all x. For simplicity, we only consider the Gaussian kernel in this paper. By replacing the Euclidean distance in Eq. (1) with the Gaussian kernel-induced distance, we the objective function of KFCM as follows

$$J_m(U,V)=2\sum_{i=1}^{c}\sum_{k=1}^{n}u_{ik}^m\left(1-K(x_k,v_i)\right) . \tag{3}$$

By optimizing Eq. (3), we have the following alternative iterative equations for the Gaussian kernel

$$u_{ik}=\frac{\left(1/\left(1-K(x_k,v_i)\right)\right)^{1/(m-1)}}{\sum_{j=1}^{c}\left(1/\left(1-K(x_k,v_j)\right)\right)^{1/(m-1)}} , \tag{4}$$

and

$$v_i = \frac{\sum_{k=1}^{n} u_{ik}^m K(x_k, v_i) x_k}{\sum_{k=1}^{n} u_{ik}^m K(x_k, v_i)}. \tag{5}$$

3 S²KFCM

Now we are position to present the semi-supervised KFCM. Assume there are a mall number of labeled data and a large amount of unlabeled data. Each data can be represented as a vector in R^s. Thus all the labeled and unlabeled data can be denoted in a whole matrix form as follows

$$X = \left\{ \underbrace{x_1^l, \ldots, x_{n_l}^l}_{labeled} \middle| \underbrace{x_1^u, \ldots, x_{n_u}^u}_{unlabeled} \right\} = X^l \bigcup X^u. \tag{6}$$

Here the superscript l and u indicate the labeled or unlabeled data respectively, and n_l and n_u denote the number of labeled and unlabeled data respectively. The total number of data is represent with $n = n_l + n_u$. In conventional approach to classifier design, e.g. k-nearest neighbor classifier, only X^l is used to train the classification function, and then use that function to label X^u.

Similarly, a matrix representation of the fuzzy c-partition of X in Eq. (6) has the following form

$$U = \left\{ \underbrace{U^l = \left\{u_{ik}^l\right\}}_{labeled} \middle| \underbrace{U^u = \left\{u_{ik}^u\right\}}_{unlabeled} \right\}. \tag{7}$$

Here the value of the component u^l_{ik} in U^l is known beforehand and typically is set to 1 if the data x_k is labeled with class i, and 0 otherwise. From U^l, we can obtain an initial set of cluster centers or prototypes as follows

$$v_i^0 = \frac{\sum_{k=1}^{n_l} (u_{ik}^l)^m x_k^l}{\sum_{k=1}^{n_l} (u_{ik}^l)^m}, 1 \le i \le c. \tag{8}$$

Consequently, the membership u^u_{ik} in U^u is updated as follows

$$u_{ik}^u = \frac{\left(1/\left(1 - K(x_k^u, v_i)\right)\right)^{1/(m-1)}}{\sum_{j=1}^{c} \left(1/\left(1 - K(x_k^u, v_j)\right)\right)^{1/(m-1)}}, 1 \le i \le c, 1 \le k \le n_u. \tag{9}$$

Finally, the cluster centers are updated by calculating

$$v_i = \frac{\sum_{k=1}^{n_l} (u_{ik}^l)^m K(x_k^l, v_i) x_k^l + \sum_{k=1}^{n_u} (u_{ik}^u)^m K(x_k^u, v_i) x_k^u}{\sum_{k=1}^{n_l} (u_{ik}^l)^m K(x_k^l, v_i) + \sum_{k=1}^{n_u} (u_{ik}^u)^m K(x_k^u, v_i)}. \tag{10}$$

We summarize the above discussion by formalizing the developed algorithm.

The proposed S²KFCM algorithm

Step 1 Fix c, and select parameters, t_{max}, m >1 and ε >0 for some positive constant.

Step 2: Initilize $U_0=[U^d|U^u{}_0]$.

Step 3: Compute initial prototypes using Eq. (8).

Step 4: For t = 1, 2, …, t_{max}.

(a) Compute the membership $u^u{}_{ik}$ in U^u using Eq. (9).

(b) Compute $E_t = \|U^u{}_t - U^u{}_{t-1}\|$.

(c) If $E_t \leqslant \varepsilon$, then stop; else compute the prototypes using Eq. (10), next t.

4 Experiments

In this section, we make numerical comparison between the proposed S²KFCM and other algorithms including S²FCM, nearest neighbor (1-NN) and unsupervised FCM and KFCM on some benchmark data sets. We use the Gaussian kernel for S²KFCM and KFCM and the parameter σ is computed as follows

$$\sigma = \frac{1}{c}\left(\sqrt{\frac{\sum_{j=1}^{n}\left\|x_j - m\right\|^2}{n}}\right). \tag{11}$$

Here c is number of clusters, x_j is the labeled or unlabeled data, n is the total number of labeled and unlabeled data and m is the centroid of the n data. In all the experiments, we set the parameters m=2, ε =0.001 and t_{max}=50.

The first benchmark data is the well-known Iris data set [9]. It contains 3 clusters with 50 samples each. We choose from the total 150 data one portion as the labeled data set and the other as the unlabeled data set. 1-NN uses only labeled data set for training; FCM and KFCM use only unlabeled data set for clustering, while S²FCM and S²KFCM both use labeled and unlabeled data set for better performance. Table 1 shows the numbers of misclassified data of the five algorithms when different sizes of labeled data set are used.

Table 1. Number of misclassified data on Iris data set (#: number of labeled data)

#	1-NN	FCM	KFCM	S²FCM	S²KFCM
45	5	11	11	8	6
60	6	9	8	7	5
75	4	6	5	5	4
90	2	3	3	1	1

Table 2. Number of misclassified data on Wine data set (#: number of labeled data)

#	1-NN	FCM	KFCM	S²FCM	S²KFCM
45	43	38	36	38	37
60	41	34	34	33	32
75	29	25	26	24	24
90	22	23	22	21	18

From Table 1, we see that in nearly all cases S^{2}KFCM achieves the best performance. Also we know that the semi-supervised algorithms are superior to the corresponding unsupervised one for classification. Finally, as the labeled data increases, the numbers of misclassified data of all algorithms decrease. It was reported in [10] that when 60 labeled data is used for training, the classification accuracy of SVM is 94.0%, which is very close to that of S^{2}KFCM.

The second example is the Wine data set [9]. It contains 3 clusters with 59, 71 and 48 samples respectively. We choose from the total 178 data one portion as the labeled data set and the other as the unlabeled data set, and Table 2 gives the results. Clearly, S^{2}KFCM also shows good advantage over other algorithms.

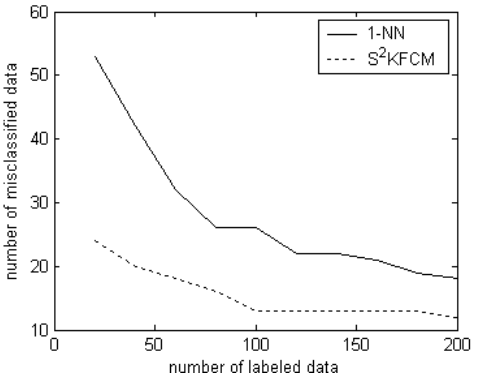

Fig. 1. Comparisons of number of misclassified data on Wisconsin Breast-Cancer data set of 1-NN and S^{2}KFCM

Finally, we made comparisons between 1-NN and S^{2}KFCM on Wisconsin Breast-Cancer data set [9], which contains 2 clusters with 444 and 239 data respectively. Fig. 1 shows the results. From Fig. 1, we see that there is a significant increase in the performance of S^{2}KFCM by using the unlabeled data for classification. When 200 labeled data is used, S^{2}KFCM has 12 misclassified data, which is comparative to SVM [10]. As number of the labeled data increases, the difference between 1-NN and S^{2}KFCM decrease, which support our claim that when a small number of labeled data is available, S^{2}KFCM can obtain better performance compared with classical classifiers.

5 Conclusions

A semi-supervised fuzzy kernel clustering algorithm called S^{2}KFCM is proposed, which can utilize unlabeled data together with labeled data for classification. Experiments are carried out on benchmark data sets and the results show that by combing both labeled and unlabeled data in the learning process, S^{2}KFCM achieves good classification results.

References

1. Saint-Jean, C., Frelicot, C.: A Robust Semi-Supervised EM Clustering Algorithm with a Reject Option. In: Proceedings of the International Conference on Pattern Recogntion, 2002
2. Bensaid, A.M., Hall, L.O., Bezdek, J.C., Clarke, L.P.: Partial Supervised Clustering for Image Segmentation. Pattern Recognition 29 (1996) 859-871
3. Pedrycz, W., Waletzky, J.: Fuzzy Clustering with partial supervision. IEEE Trans. on Systems, Man, and Cybernetics, Part B-Cybernetics 27 (1997) 787-795
4. Bennett, K., Demiriz, A.: Semi-Supervised Support Vector Machines. Advances in Neural Information Processing Systems 11 (1999) 368-374
5. Zhang, D.Q., Chen, S.C.: Kernel-based Fuzzy and Possibilistic C-Means Clustering. In: Proceedings of the International Conference on Artificial Neural Networks, Istanbul, Turkey, 2003, 122-125
6. Zhang, D.Q., Chen, S.C.: A Novel Kernelised Fuzzy C-Means Algorithm with Application in Medical Image Segmentation. Artificial Intelligence in Medicine, in press (2004)
7. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification. John Wiley (2001)
8. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York (1981)
9. UCI Repository of Machine Learning Databases, University of California, Irvine**,** available from: http://www.ics.uci.edu/~mlearn/MLRository.html
10. Chen, S.C., Yang, X.B.: Alternative Linear Discriminant Analysis. Pattern Recognition, in press (2004)