A Novel Multiclass Text Classification Algorithm Based on Multiconlitron

Yuping Qin[1, a *], Fengfeng Qiu[2, b], Qiangkui Leng[3, c] and Aihua Zhang[1, d]

[1]College of Engineering, Bohai University, Jinzhou 121013, China

[2]College of Mathematics and Physics, Bohai University, Jinzhou 121013, China

[3]College of Information Science and Technology, Bohai University, Jinzhou 121013, China

[a]qlq888888@sina.com

Keywords: Multiconlitron; Multiclass classification; One-against-one; Support vector machines

Abstract. A novel multiclass text classification algorithm based on multiconlitron is proposed. The multiconlitron is constructed for each possible pair of classes in sample space, each of which is used to separate two classes. For the sample to be classified, every multiconlitron is used to judge its classman vote for the corresponding class. The final class of the sample is determined by the number of votes. The classification experiments are done on Reuters 21578, and the experimental results show that the proposed algorithm has better performance. Compared with multiclass support vector machines 1-a-1, 1-a-r and DAGSVM, while ensuring the classification accuracy, it improves the classification speed and training speed greatly.

Introduction

Multiclass text classification, which involves assigning one predefined categories to a free text according to its content, plays a key role in organizing the massive sources of unstructured text information into an organized state. With the rapid development of technology of computer and communication, there is a growing need for multiclass text categorization, because the sources of information to be managed, such as webpages, news sources, and database management, are increasing rapidly. Filtering of spam mail, quick search of interesting topics from large databases, and retrieving information based on user's preferences from information sources, are some examples, where multiclass text categorization can play an important role. Therefore, automated text classification technology has attracted much attention and has become an important research area. Many multiclass text classification algorithms based on machine learning have been proposed and received better results [1-2]. Among supervised learning algorithms, SVM has been well explored, and has proven to be a promising technique for multiclass text categorization.

Currently, main research results of multiclass SVM include one-against-rest [4], one-against-one [5], directed acyclic graph SVM (DAGSVM) [6] and so on. One-against-rest method uses less classifier with faster decision speed, while one-against-one method has higher classification accuracy, and DAGSVM performs more efficiency since it needn't traverse all classifiers under the DAG combination strategy. Meanwhile, kernel function plays a key role in SVMs, and it can implicitly solve linear separable problem in high dimensional space. According to pattern recognition theory, linearly inseparable data in low dimensional space may achieve linearly separable in high dimensional feature space through the nonlinear mapping. But if the direct mapping is assigned, then there exists some difficulties such as determining the form and parameters of nonlinear mapping function, leading to serious "curse of dimensionality". The kernel function can solve the above problems very tactfully. It transforms the inner product in high dimensional feature space into function in low dimensional input space. However, selection of kernel function often lack specific guidance [7-11], and the demand for computing resources on large scale data is very big. In addition for kernel function, it is hard to interpret the change of metric space [12, 13].

Piecewise linear classifiers (PLCs) is a special non-linear classifiers. Its decision surface is composed of a number of hyperplanes. Therefore, compared with the general hypersurface, it is still simple. The parameters of PLCs is independent of the dataset, without assumptions statistical distribution of the sample. Meanwhile, it can approach hypersurfaces with various shapes, being a strong adaptive ability [14]. However, construction of the PLCs remains some problems. In general, determining the piecewise linear boundary is a complex global optimization problem [15]. In most cases, the search for such boundaries is transformed to minimize the error function about classification problem.

Another goal for piecewise linear classifier is to select the appropriate number of hyperplanes. For the purpose, some training algorithms are designed and implemented. But they are usually very complicated and have lone running time, what limits the application of PLCs to a certain extent. To reduce training time and avoid solving the optimization problem, many heuristic methods are adopted, but they are often with more empirical operations, and not conducive to the promotion.

Multiconlitron [16] absorbs the advantages of SVMs. It realizes classification with maximum margin but without kernel function and space mapping. Furthermore, it can minimize classification error on the training set, and dynamically obtain the number of hyperplanes. On the basis of multionlitron, a multiclass text classification algorithm(MCLP) is proposed. The algorithm will construct classifiers in input space without kernel function, and realize the approximation to true class boundary by piecewise linear functions .

Multiconlitron

The main component of multiconlitron is conlitron. It is a set of linear functions. For $X, Y \subset R^n$, if $CH(X) \cap Y = \Phi$ ($CH(X)$ stands for the convex hull of X), the conlitron containing L linear functions can be defined as

$$CLP = \left\{ f_l(x) = \omega_l \cdot x + b_l, 1 \le l \le L \right\} \tag{1}$$

Accordingly, the decision function can be defined as

$$CLP(x) = \begin{cases} +1, \forall 1 \le l \le L, f_l(x) > 0 \\ -1, \exists 1 \le l \le L, f_l(x) < 0 \end{cases} \tag{2}$$

Moreover, multiconlitron is a set of conditions. For $X, Y \subset R^n$, if $X \cap Y = \Phi$, the multiconlitron containing K conlitrons can be defined as

$$MCLP = \left\{ CLP_k, 1 \le k \le K \right\} \tag{3}$$

Its decision function is defined as:

$$MCLP(x) = \begin{cases} +1, \forall 1 \le k \le K, CLP_k(x) = +1 \\ -1, \exists 1 \le k \le K, CLP_k(x) = -1 \end{cases} \tag{4}$$

Fig. 1 shows the schematic of a multiconlitron. Given the training direction from X to Y, it can be obtained three conlitrons, i.e., CLP_1, CLP_2, and CLP_3. Each of them consists of different linear decision functions, e.g., CLP_1 consists of five linear decision functions, CLP_2 consists of three linear decision functions, and CLP_3 consists of one linear decision function. The final multiconlitron is expressed as $MLCP = \left\{ CLP_k, 1 \le k \le 3 \right\}$. The corresponding decision rule can be described as following: if a prediction sample locates one of CLP_s, then its label is marked +1, otherwise -1.
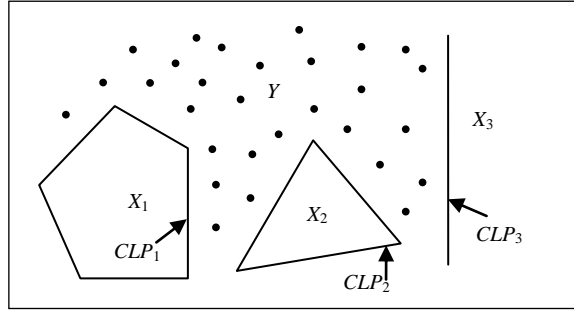
Figure 1. The schematic of a multiconiltron

Multiclass Text Classification Algorithm based on Multiconlitron

For multiclass classification by using multiconlitron, the basic idea is that constructing multiconlitron for each possible pair of classes in sample space, and each of which is used to separate two classes. The classification decision employs the voting method, i.e., the final class is decided by the number of votes. While the two classes have the same number of votes [17, 18], the smaller label is determined as the class.

Assume that the given multiclass sample set $A = \{x_i, y_i\}_{i=1}^{l}$, where $x_i \in R^n$, $y_i \in \{1,2,\cdots,N\}$, l is the number of samples, and N is the number of classes. The multiclass text classification algorithm based on multiconlitron is described in detail as follows:

Step 1: Constructing $N(N-1)/2$ multiconlitrons $MCLP_{ij}$ ($1 \le i \le N-1, i < j \le N$) by Eq. (1)-Eq. (4), where $MCLP_{ij}$ is the multiconlitron separating the i-th and the j-th class.

Step 2: For the sample x to be classified, judging its class by $MCLP_{ij}$ ($1 \le i \le N-1, i < j \le N$). If $MCLP_{ij}(x) = +1$, then x belongs to the i-th class, and the number of votes of the class will increase by one.

Step 3: Computing the final number of votes that the sample x belong to the i-th class by Eq. (5)

$$D_i(x) = \sum_{j \ne i, j=1}^{N} sgn(MCLP_{ij}(x))$$

(5)

$$sgn(x) = \begin{cases} 1 & x > 0 \\ 0 & x \le 0 \end{cases}$$

(6)

Step 4: Sorting the sequences $D_1, D_2, D_3, \cdots, D_N$ in ascending order. If $D_i = D_j (i<j)$, then D_i is placed in front of D_j.

Step 5: Determining the class of the sample x by Eq. (7)

$$class(x) = \arg \max_{i=1,2,\cdots,N} (D_i(x))$$

(7)

Experimental Results and Analysis

Experiments are made on Reuters 21578, in which five categories, 896 texts are used. 596 texts are used as training set, and the rest as testing set (see Tab. 1). Information Gain is used to reduce feature dimension and the weight of each word is computed according to TF-IDF.

Table 1 Training set and testing set

class	wheat	corn	coffee	soybean	cocoa
training size	204	168	97	79	50
testing size	101	84	48	40	25

The computational experiments were done on a CPU Pentium 2 G with 512M memory, Windows Xp operation system. In SVMs, RBF kernel function is selected, $K(x,y) = e^{-\gamma \|x-y\|^2}$ with $\gamma = 0.1$. The penalty parameter C is set to 10.

The standard precision, recall and F_1 value are used to evaluate the classification performance of the algorithm. The function of Recall, Precision and F_1 value are as follows:

$$R_i = \frac{N_{CP_i}}{N_{C_i}} \tag{8}$$

$$P_i = \frac{N_{CP_i}}{N_{P_i}} \tag{9}$$

$$F_1^i = \frac{P_i \times R_i \times 2}{P_i + R_i} \tag{10}$$

N_{C_i} the number of texts belonging to class C_i; N_{P_i} is the number of texts predicted to belong to class C_i; N_{CP_i} is number of texts that have been correctly classified.

Micro average and Macro average are two methods that are applied to evaluate the system performance. The Precision, Recall, and F_1 value of one class are called Micro average. The Precision, Recall, and F_1 value of all classes are called Macro average.

To verify the performance of the proposed algorithm, SVM multiclass classification algorithm 1-a-1, 1-a-r and DAGSVM are all employed here for comparison. The Macro average precision, Macro average recall and Macro average $F1$ of four algorithms are given in Table. 2. The training time and testing time of four algorithms are given in Table. 3.

Table 2 Comparison of macro average precision, recall and F_1 Value

algorithm	Macro Average Precision (%)	Macro Average Recall (%)	Micro Average F_1 value (%)
1-a-1	68.27	62.91	64.71
1-a-r	68.51	64.11	65.61
DAGSVM	68.22	62.75	64.23
MCLP	68.16	62.38	64.18

Table 3 Comparison of training time and testing time

algorithm	Training time [ms]	Testing time[ms]
1-a-1	1038	469
1-a-r	2109	422
DAGSVM	1038	357
MCLP	1053	346

The experimental results show that the accuracy of MCLP is basically equal to 1-a-1, 1-a-r and DAGSVM methods. The training speed of MCLP is basically equal to 1-a-1 method, but higher than 1-a-r and DAGSVM methods,The classification speed of MCLP is faster than 1-a-1,1-a-r and DAGSVM methods. The key reason is that the proposed method does not used kernel function, and reduces the computational complexity. If the dada is linear separable in the feature space, the superiority of MCLP is more obvious.

Conclusion

Multiclass text classification is of great value in many applications. On the basis of in-depth research on multiclass SVMs, a novel multiclass text classification algorithm based on multiconlitron is proposed. The experimental results show, Compared with the support vector machine multiclass classification algorithms, the proposed algorithm not only improves the training speed and classification speed significantly, but also ensures the classification accuracy. In the future work, we will focus on improving the accuracy of the algorithms and testing the algorithm on the large data set with more classes.And data driven idea[18-20] can be considered also in this field.

Acknowledgements

References

[1] Y. M. Yang and X. Liu: In Proceeding of ACM SIGIR Conference on Research and Development in Information Retrieval, p.42.

[2] S. Fabrizio: *Machine Learning in Automated Text Categorization* (Statistical Learning Theory, New York: Wiley 1998), p.1.

[3] S. Yin, X. Li, H. GAO, . Kaynak: Data-based techniques focused on modern industry: An overview, IEEE Transactions on Industrial Electronics, Vol.62 (2015) No.1, P. 657.

[4] U.G. Krebel: *Pairwise classification and support vector machines* (MA: MIT Press, Cambridge 1999), p.255.

[5] Platt J, Cristianini N and Shawe-Taylor J: *Large margin DAGs for multiclass classification*. (MA: MIT Press, Cambridge 2000), p. 547.

[6] K. Muller, S. Mika, G. Ratsch et al: IEEE Transactions on Neural Networks,Vol.12 (2001) No.2, p.181.

[7] Shawe-Taylor J and Cristianini N: *Kernel methods for pattern analysis* (Cambridge university press, Cambridge 2004).

[8] S. Sonnenburg, G. Rätsch and C.A.: *Proceedings of 2005 Annual Conference on Neural Information Processing Systems* (Vancouver, Canada, 2005), P. 1273.

[9] S. Sonnenburg, G.Rätsch, C.Schäfer et al: Journal of Machine Learning Research, Vol.7 (2006) , p.1531.

[10]Z. Wang, S. Chen and T. Sun: IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.30 (2008) No.2, p.348.

[11]M. Doumpos, C. Zopounidis and V. Golfinopoulou: IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, Vol.37 (2007) No.3, p.540.

[12]O. Pujol and D.: IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.31 (2009) No.6, p.1140.

[13]D. Webb: *Efficient piecewise linear classifiers and applications* (Ballarat: University of Ballarat).

[14] J. Sklansky and G.N. Wessel: *Pattern classifiers and trainable machines* (Springer, Berlin 1981) .

[15] Y. Li, B. Liu, X. Yang, Y. Fu et al: IEEE Transactions on Neural Networks, Vol.22 (2011) No.2, p.276.

[16] C.W. Hsu and C.J. Lin: Neural Networks, IEEE Transactions on Neural Networks, Vol.13 (2002) No.2, p.415.

[17] Y. Shan, H. Wang and S. Dong: Computer Engineering and Design, Vol.33 (2012) No.5, p.1838.

[18] S. Yin, Z. Huang: Performance monitoring for vehicle suspension system via fuzzy positivistic C-means clustering based on accelerometer measurements, IEEE/ASME Transactions on Mechatronics, Vol.20 (2015), No.5, p.2613.

[19] S. Yin, X. Zhu, and O. Kaynak: Improved PLS focused on key performance indictor related fault diagnosis, IEEE Transactions on Industrial Electronics, Vol.62(2015), No.3, p.1651.

[20] S. Yin, X. Zhu: Intelligent particle filter and its application on fault detection of nonlinear system, IEEE Transactions on Industrial Electronics, Vol.62 (2015), No.6, p.3852.