ATLANTIS PRESS
Advances in Engineering Research (AER), volume 131
3rd Annual International Conference on Electronics, Electrical Engineering and Information Science (EEEIS 2017)

Relevance Identification of Chinese News in the New Media Environment —Taking "Shandong Vaccine Event" as an Example

Le SONG[1,a] , Ming-Chun ZHENG[2,b,*]

[1]Shandong Normal University, Jinan, China

[2]Shandong Normal University, Jinan, China

[a]songle15@163.com, [b]zhmc163@163.com

*Corresponding author

Keywords: Maximum entropy, Chinese news relevance, Micro-blog, New media environment.

Abstract. We construct the maximum entropy model, which improved the feature function and training set, to calculate the relevance degree of Chinese news and selected topics, and apply them to Chinese news event relevance recognition. According to the example of "Shandong vaccine event", the relevant data are obtained on the micro-blog platform. By selecting different numbers of features, this paper compare the maximum entropy model with the support vector machine, BP neural network, Bayes and K-means algorithm, which are four kinds of common text classification method of micro-average accuracy, by empirical analyzing. Experiments have found that although the method is not always superior to the support vector machine method, it is superior to the other three methods.

1 Introduction

Event correlation recognition is a shallow event relationship recognition task, by parsing the text of the semantics and structural characteristics to given a relevant or unrelated decision by describing a different event of the text fragment [1].

At present, the existing event relevance identification method, according to its principles can be divided into pattern matching method, elemental analysis and machine learning methods [2]. This paper hopes to use the machine learning method to establish the maximum entropy model.

We pay attention to the original maximum entropy model and it's improving the model for text classification. Xuetian Chen, etc. campared maximum entropy model with Bayes, KNN and SVM, and used different text feature generation method, the number of features and smoothing technology to test, and achieved a good classification effect [3-5]. Masaki Murata, etc. are more emphasis on the extraction of important features to improve the accuracy rate [6]. Zhang Xinhua establishes the maximum entropy model text classification method of semi-supervised learning [7]. Chunyong Yin , etc. applied the maximum entropy model to the mobile text classification in cloud computing, and added the information gain algorithm to the model to improve the classification accuracy [8]. Jun Li, etc. model of a variety of emotional tags, and they use different emotional characteristics of the text to classify the short text of micro-blog, etc. [9]. Wenming Huang , etc. use the finite Newton smoothing algorithm to smooth the emotional analysis model to optimize [10].

In this paper, we compare the average accuracy of our maximum entropy model with other methods by empirical analysis. It is important to note that our work is mainly based on the calculation of the degree of association between the event and the subject.

2 Identification of Chinese News Based on Maximum Entropy Model

2.1 Relevance of Chinese News

The words in the Chinese news event information (hereinafter referred to as "information") can be regarded as some of the characteristics of information, and a message contains a number of characteristics. The probability that the information of containing word f belongs to a certain type

of information a can be obtained by training the information set. Given a training set, $A=\{a_1,a_2\}$ is the set of categories to which the information belongs, where a_1 indicates that the information is associated, and a_2 indicates that the information is not associated. $d_i=\{b_1,b_2,...,b_m\}$ is the set of feature words of information d_i, and $D=\{d_1,d_2,...,d_n\}$ is the set of information. $num(a_i,b_j)$ is the number of occurrences of the two-tuples (a_i,b_j) in the training set. Then there will be $\sum_{i=1}^{2}\sum_{j=1}^{m}num(a_i,b_j)$ pairs in the training set, and the probability of occurrence of the two-tuples (a_i,b_j) appears to be expressed as:

$$\tilde{p}(a_i,b_j)=\frac{num(a_i,b_j)}{\sum_{i=1}^{2}\sum_{j=1}^{m}num(a_i,b_j)} \tag{1}$$

For the information containing the word b_j, select the larger of $\tilde{p}(a_1,b_j)$ and $\tilde{p}(a_2,b_j)$ as the classification result of the information.

In this paper, we use the following formula to calculate the information entropy of the text based on the principle of the maximum entropy and the uniformity of the conditional distribution $p(a\mid b)$:

$$H(p)=-\sum_{a,b}\tilde{p}(b)p(a\mid b)\log_2 p(a\mid b) \tag{2}$$

$\tilde{p}(b)$ is the empirical distribution of b in the training sample, $p(a\mid b)$ is the probability that the text belongs to category a_i in the case where the feature word b_i appears. We need to calculate the maximum information entropy of the text, and solve the probability distribution formula under the maximum entropy principle:

$$p^*=\arg\max_{p\in P}H(p) \tag{3}$$

That is, solving the p value that maximizes the information entropy $H(p)$.

In the absence of any prior knowledge, by the nature of the entropy we can see that when the distribution is the most uniform, the entropy is the largest, the formula (2) to take the maximum condition is $p(a\mid b)=\frac{1}{|A|}$.

The sum of the probabilities that b_i belongs to each category is 1, which is $\sum_{a\in A}p(a\mid b_i)=1$.That is to say $p(a_1\mid b_i)+p(a_2\mid b_i)=1$ and $p(a\mid b_i)\geq 0$.

By training the text, we can get the probability values of some of the two-tuples (a_i,b_j), or the conditions that certain probabilities need to be met. Therefore, the classification of Chinese news turns into solving the maximum entropy under the condition of partial information, or finding the optimal solution satisfying certain constraints.

In order to express the known information, we introduce the feature function. In general, the characteristic function is a binary function, $f(a,b)\rightarrow\{0,1\}$. For the Chinese news event classification problem, the characteristic function can be defined as:

$$f(a,b)=\begin{cases}1, a=a_i\wedge b=b_j, i=1,2. j=1,2,...\\ 0, \quad otherwise\end{cases} \tag{4}$$

For the characteristic function f_i, the expected value of the characteristic function f_i for model $p(a|b)$ is (5). We are limited to the training set, the two expectations are the same, that is (6).

$$E_p f_i = \sum_{a,b} \tilde{p}(b) p(a|b) f_i(a,b) \tag{5}$$

$$E_{\tilde{p}} f_i = E_p f_i, i = 1,2,...,k \tag{6}$$

We call (6) a constraint. We can define multiple such independent characteristic functions. Many characteristic functions can describe the problem from different perspectives, and combine the scattered knowledge together to complete the classification task. Given k characteristic functions $f_1, f_2, ..., f_k$, we can get the k constraints of the required probability distribution.

Thus, we turn the problem into an optimal solution for satisfying a set of constraints, that is:

$$P = \{p \mid E_{\tilde{p}} f_i = E_p f_i, i = 1,2,...,k\} \tag{7}$$

$$p^* = \arg\max_{p \in P} H(p) \tag{8}$$

Use the Lagrangian multiplier method to find the optimal solution to get:

$$p^*(a|b) = \frac{1}{\sum_a \exp(\sum_{i=1}^{k} f_i(a,b))} \exp(\sum_{i=1}^{k} f_i(a,b)) \tag{9}$$

The formula (9) is recorded as (10). Where $\pi(b)$ is the normalization factor:

$$p^*(a|f) = \frac{1}{\pi(b)} \exp(\sum_{i=1}^{k} f_i(a,b)) \tag{10}$$

$$\pi(b) = \sum_a \exp(\sum_{i=1}^{k} f_i(a,b)) \tag{11}$$

We can use formula (10) to find probability that any information d_i belongs to the class a_1 and a_2, and we only need to select the larger probability of a_i as the category of document d_i. According to the classification result, If two or more events are categorized with the selected topic, select $p(a_1, d_i)$. As a result, the events are considered to be associated. The degree of association of two or more event information can be calculated using the following formula:

$$\eta = \frac{\sum_{i=1}^{n} p(a_1, d_i)}{n} \tag{12}$$

2.2 Improve the characteristic equation

The feature that is more in the association event and fewer in the non-associated event has a reference value for event relevance identification. If the number of occurrences of a feature is small, or the number of occurrences in the text of the association and non-associated events is greater, then this feature has a lower reference value in event relevance recognition.

Given the feature b, we can classify the training set into four categories, which show in table 1.

Where the r,s respectively represent the number of associated and non-associated events that contain feature b. The t,u respectively represent the number of associated and non-associated events that don't contain feature b. If the feature b appears more in the association event category and less in the non-associated event category, or if a feature appears less in association events and more in non-associated events, then such a feature should be chosen as an effective feature and given a higher weight.

Table 1 Features-Class Tables

	Association event	Non-associated event	Total
Contains the feature b	r	s	r+s
Does not contain feature b	t	u	t+u
Total	r+t	s+u	r+s+ t+u

Using $m_1,m_2,...,m_r$ respectively represent that the number of feature b which is included in the first incident to the r-related events. And using $n_1,n_2,...,n_s$ respectively represent that the number of feature b which is included in the first non-associated events to the s-non-associated events. The value of Difference Counting for feature b is:

$$DC(b)=\alpha*\frac{\sum_r^1 m_i}{\sum_1^r m_i+\sum_1^s n_i}+(1-\alpha)\frac{u}{t+u} \tag{13}$$

Obviously, the greater the $DC(b)$, the greater the reference value for event relevance identification. After we calculate the $DC(b)$ of all the features, we can sort the features according to $DC(b)$, and arrange the feature functions according to the characteristics of the first n. After the difference calculation method is established, the feature function is further improved:

$$f(a,b)=\begin{cases} DC(b)+\min(DC(b)), a=a_i \\ \min(DC(b)), otherwise \end{cases} \tag{14}$$

3 Experimental Design and Analysis of Results

In this article, we selected "Shandong vaccine case" that belongs to food and drug safety incident field as an example. We chose one of the representatives of the new media in China which is "micro-blog". In the micro-blog platform, we collect the official micro-blog of People's Daily, Xinhua News, The Paper, Headlines, which released 111 pieces of association information, but also collect the "Yao Chen" and other 200 more active micro-blog users 681 pieces of related information in the subject. And in the field of this, we selected 800 pieces of irrelevant information, and finally in the micro-blog popular information, "big V" published news and other information randomly selected 2000 pieces of real information. Where the length of the selected information is 80 to 140 words, which avoids the lack of reference value because the information is too short to contain effective information.

The selection of the training sets need to be explained as follows: (1) Training set 1: Select 792 pieces of related event information and 2000 pieces of random real information as training set 1 to train the maximum entropy model. (2) Training set 2: 800 pieces of random information in training set 1 are replaced by 800 pieces of unrelated information in this field, and others are the same as training set 1.

In this paper, we use the NLPIR word segmentation system to segment the word and use the artificial way to re-train the training set of information. The micro-average accuracy rate is used as

the evaluation index, and we trained the classifiers with two training sets, and tested the classification using the different number of features under the original feature function and the improved feature function.

Table 2 Micro-Average Accuracy Comparison Among Different Training Sets

Number of Features	Training Set 1		Training Set 2	
	Characteristic Function	Improved Characteristic Function	Characteristic Function	Improved Characteristic Function
10	68.17	70.73	69.39	71.34
15	71.43	72.70	72.59	74.80
20	72.69	74.94	75.01	79.22
25	74.74	75.62	76.74	82.17
30	72.65	74.37	75.53	76.20
35	71.10	72.99	73.31	75.84

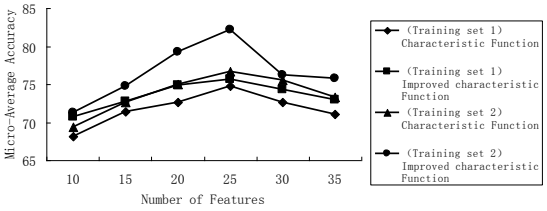

Fig.1 Comparison of micro-average accuracy for different training sets and feature functions.

Table 3 Comparison of micro-average accuracy ratios for different classification methods

Number of Features	Maximum Entropy	SVM	BP Neural Network	Bayes	K-means
10	71.34	72.38	69.07	67.73	69.00
15	74.80	74.33	73.58	70.87	71.81
20	79.22	78.60	76.44	74.34	73.25
25	82.17	81.89	78.73	77.71	78.64
30	76.20	81.10	75.71	75.36	74.40
35	75.84	76.54	74.29	72.37	73.63

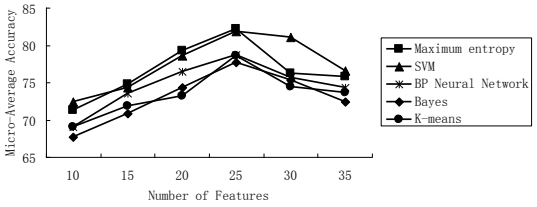

Fig.2 Comparing the micro-average accuracy of different classification methods.

According to the statistics of Table 2 and Figure 1, we can get the following conclusions:

(1) With the increase of the number of features, the classification accuracy is gradually improved. However, when the number of features increases to a certain number, classification accuracy will decline, which due to the length of the text limit, resulting in part of the text can not obtain the required number of features. With the selection of 25 features, the classification gets the best classifier.

(2) The classification accuracy of the classifier using the improved training set 2 is improved significantly.

(3) When using improved feature functions, the classification accuracy is higher.

On the basis of the above research, we use training set 2 and improved feature function to compare the classification effect of the maximum entropy model with the other text classifiers who commonly used in text recognition.

Through the analysis of Table 3 and Figure 2 can be the following conclusions:

(1) The micro-average accuracy of different methods varies with the number of features. Initially, the accuracy rate increases with the increase of the number of features. When the number of features exceeds 30, the recognition accuracy of the five methods has decreased in different degrees.

(2) When the number of features is less than or equal to 25, the accuracy rate of Chinese news text recognition based on maximum entropy model is about the same as that of support vector machine, which is higher than that of BP neural network, K-means method and Bayesian method.

4 Conclusions

In this paper, the maximum entropy model is constructed to calculate the correlation degree between Chinese news and selected topics, and we apply them to Chinese news event relevance identification. According to the empirical analysis, it can be seen that the Chinese news event relevance recognition method based on the maximum entropy model is feasible, but the model needs to be improved to make the model more suitable for the identification of Chinese news, so as to improve the accuracy of recognition rate.

5 Acknowledgement

We acknowledge the National Social Science Foundation of China (Grant No.14BTQ049), and the Soft Science Project Foundation of Shandong Province (Grant No.2016RZB01029).

References

[1] B. Ma, Y. Hong, X.R. Yang, J.M. Yao, Q.M. Zhu, Using Event Dependency Cue Inference to Recognize Event Relation, J. Acta Scientiarum Naturalium Universitatis Pekinensis, 49 (2013) 109-116.

[2] P.P. Liu, Research on the Relevance Identification of Chinese News Subject Events, D. Kunming University of Science and Technology (2016).

[3] X.T. Chen, R.L. Li, Text classification using the maximum entropy model, J. Computer Engineering and Applications. 40 (2004) 78-79.

[4] R.L. Li, J.H. Wang, X.Y. Chen, X.P. Tao, Y.F. Hu, Using the maximum entropy model for Chinese text classification, J. Computer Research and Development. 42 (2005) 94-101.

[5] R.L. Li, X.P. Tao, L. Tang, Y.F. Hu, Using Maximum Entropy Model for Chinese Text Categorization, j. LNCS-Advanced Web Technologies and Applications, 2005, pp. 578-587.

[6] M. Murata, K. Uchimoto, M. Utiyama, et al., Using the maximum entropy method for natural language processing: Category estimation, feature extraction, and error correction, J. Cognitive computation, 2(2010): 272-279.

[7] X.H. Zhang, Building Maximum Entropy Text Classifier Using Semi-supervised Learning, D. National University of Singapore(2011).

[8] C.Y. Yin, J.W. Xi, Maximum entropy model for mobile text classification in cloud computing using improved information gain algorithm, J. Multimedia Tools and Applications, 2016, pp. 1-17.

[9] J. Li, Y.H. Rao, F.M. Jin, et al., Multi-label maximum entropy model for social emotion classification over short text, J. Neurocomputing, 210(2016): 247-256.

[10] W.M. Huang, Y.Q. Sun, Emotional Analysis of Chinese Short Text Based on Maximum Entropy, J.Computer Engineering and Design. 38(2017): 138-143.