Using Convex Combination Kernel Function to Extract Entity Relation in Specific Field

Qi Shang[1], Jianyi Guo[1,2(✉)], Yantuan Xian[1,2], Zhengtao Yu[1], and Yonghua Wen[1]

[1] The School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China
gjade86@hotmail.com

[2] Computer Technology Application Key Laboratory of Yunnan Province, The Institute of Intelligent Information Processing, Kunming 650500, China

Abstract. Kernel method has been proven to be effective in measuring the similarity of two complex relation patterns. Aim at the optimization problem of compound kernel functions, this paper presents a method of finding the optimal convex combination kernel function, which is comprised of multiple kernel functions and needs to be optimized. After preprocessing the corpus and selecting features including lexical information, phrases syntax information and dependency information, the feature matrix was constructed by using these features. The optimal kernel function can be found in the process of mapping the feature matrix to different high-dimensional matrix, and the different classification models can be obtained. The experiments are conducted on the domain dataset from Web and the experimental results show that our approach outperforms state-of-the-art learning models such as ME or Convolution tree kernel.

Keywords: Entity relation extraction · Compound kernel functions · Optimization · Convex combination of kernel functions

1 Introduction

Entity relation extraction refers to the automatic identification the associated relation between two entities expressed with the natural language (e.g. in the sentence "Yunnan produces a lots of wild mushrooms", the relationship between entity "Yunnan" and "mushroom" is "production" relations). Entity relation extraction has made an important role in information extraction, automatic question answering systems, machine translation and knowledge-base construction [1].

The current methods for relation extraction mainly include: knowledge-based [2], the pattern matching [3] and machine learning-based [4–12]. Knowledge-based method requires experts build different knowledge-bases on different specific areas, this method is time-consuming and labor-intensive and has many limitations. If an entity does not exist in the knowledge base, the relationship extraction cannot be carried out. Some common knowledge repositories have now been built up [15–18], while specific areas knowledge-bases are less. Pattern matching method is based on the words located

S. Song and Y. Tong (Eds.): WAIM 2016 Workshops, LNCS 9998, pp. 137–148, 2016.
DOI: 10.1007/978-3-319-47121-1_12

before or after the given entity in a sentence, matching these words or the syntax format to the target pattern. But in the process of matching, due to the position of the words, which were usually disaccord with the given pattern that may result in a very low similarity. Especially for Chinese, because of the complexity of the language, accuracy and recall rate of this method is very low.

Machine learning methods for entity relation extraction can be divided into several ways: features-based method [4, 5], bootstrapping-based method [6, 7], deep learning method [8], and kernel methods [9–14]. The key of features-based method is the effectiveness of the selected features. For example, Dong (2007) [5] selected a syntax tree, lexical features, physical features and other features as well as their combination, using CRF model, performed relation extraction between inclusion and non-inclusion. However, it is difficult for this method to adapt to another relation extraction system, and its disadvantage lies in the difficultness of discovery of new plane features to improve the extraction performance. The bootstrapping is used to learn relation patterns. For example, Ye (2014) [6] took the bootstrapping to extract entity relation. However, Komachi' analysis in his paper [7] showed that semantic drift is an inherent property of iterative bootstrapping algorithms and, therefore, poses a fundamental problem. As a new machine learning algorithms, deep learning has made remarkable achievements in speech recognition and image processing. Recently it has been widely studied and applied on natural language processing field. Zeng (2014) in his paper [8] proposed to have used convolution method on DBN to predict the relationship between the two marks word. But the limitation of deep learning requires a lot of corpus for the experiment.

2 Related Work

Kernel methods [9–14] have been proven to be effective in measuring the similarity of two complex relation patterns.

For extracting entity relationship in the specific field, different kernel functions for different relation extraction have differences, the usage of a single kernel cannot solve the universal problem but the compound kernel methods, which may reduce the risk caused by the loss of important features by combining different single kernel function to be a combinatorial function. However, the composite kernel functions are not optimalizing. This paper proposes a way of optimizing the compound kernel to improve extracting performance by finding the optimal kernel function. The target kernel function is formed by convex combination of each single kernel function [13, 14], and the lexical and sentence information such as context words, part of speech, phrase syntactic and dependency syntax information extracted from the pretreated corpus are all used to form the feature matrix, and then the feature matrix is mapped to the different convex functions which are made of the combination among radial basis function (RBF), linear kernel function (LKF) or the polynomial kernel. Function (PKF) The classification model can be obtained by machine-learning method which supporting the kernel function, enumerating the classified result by making use of the testing corpus to find the optimal model. At last the entity relation in specific areas can be identified with the optimal model of convex combination kernel.

The task of entity relation extraction in this paper is divided into four parts: pre-processing the corpus, extracting feature to form feature matrix, mapping kernel function to high dimensional matrix, learning different classification model to find the optimal classification model (e.g. in tourism field). As shown in Fig. 1.

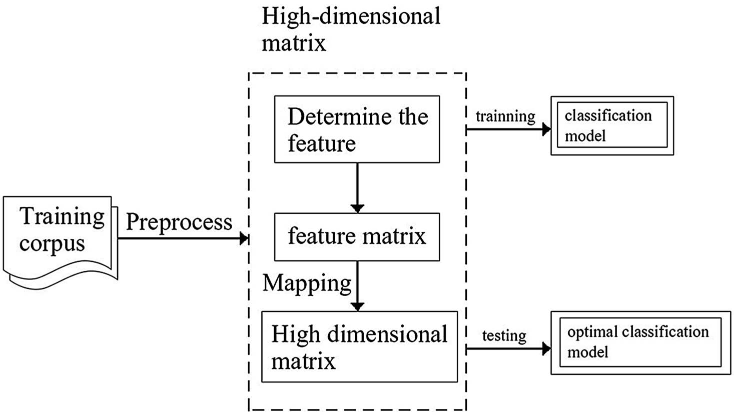

Fig. 1. General framework for entity relation extraction in specific fields

2.1 Preprocess

Preprocess includes segmentation, speech tagging, named entity recognition, sentence segmentation, relationship candidate generation and label candidate instances.

First, this paper employs ICTCLAS [19] tool from Chinese Academy of Sciences for the segmentation and speech tagging of the input text; Secondly, artificial tagging of named entities, which will be conducted on the corpus that has completed segmentation and speech tagging. In compliance with rules of conditional random fields (CRFs), the training of named entity recognition model should be performed, so as to achieve automatic recognition of named entities. Meanwhile, it requires pre-defined categories of entities, specifically: sights, location, snacks, specialties, hotels, numeric expressions, dates and festivals. Then, based on punctuation and contextual features, tagged corpus should be cut into separate sentences. It should be noted that types of entity relationships also need to be pre-defined, (for example, located, apart, adjacent, these words belong to location relationship, or ticket prices, altitudes, customs, these words belong to attribute relationship). Take the sentence “In 2004, Huangshan was first selected as a World geological park, making itself a tourist destination which won honors of World Cultural Heritage and Natural Heritage and World Geological Park at the same time.” Its segmentation result is: In 2004/m, /w Huangshan /ns was first /m selected /v as a World Geological Park /n, /w making /v itself a tourist destination /n which /w won /v honors of /n World Cultural /n and /c Natural Heritage /n and/c World Geological Park /n at the same time/c. Among them, entities are “Huangshan”, “World Geological Park”, “World Cultural Heritage”, “Natural Heritage”, “tourist destination”, whose relationship is shown in Table 1.

Table 1. Entity relationship

Entity 1	Entity 2	Relationship
Huangshan	World Geological Park	Listed (in 2004)/whole-part relation
Huangshan	tourist destination	whole-part relation
Huangshan	World Cultural Heritage	whole-part relation
Huangshan	Natural Heritage	whole-part relation

After enumerating and finding all possible combinations of entities in each sentence, each combination has become a candidate example, whose artificial tagging will be carried out according to whether relationships exist. With 500 pieces of data as artificial tagging examples, there will be examples of relations being marked out. Employing CRFs for model training, along with a test on 1000 pieces of data in the model, instances of relations will be extracted and form a desired corpus. By way of previous steps, pre-processing is completed.

2.2 Extracting Features to Form Characteristic Matrix

In this paper, the selection of features will follow methods introduced in the References [4, 16]. In the field of Chinese tourism, it focuses on features of syntax and dependency information of phrases, with an expectation to optimize the performance of relation extraction. After completing the selection of features, according to which features matrix can be formed, such as the formula (1), where $X_i(i = 1, 2, \ldots m)$ is the vector of each example after characteristic extraction.

$$K = (X_1, X_2, \ldots X_m)^T \tag{1}$$

2.2.1 Lexical Information

(1) Entity Information. Entity information is the basic vocabulary information, including the first entity category, the first entity subcategory, the first entity's syntactic functions, the second entity's category, the second entity's subcategory and the second entity syntactic functions.
(2) Local Contextual Information of Words. Referred literature [16] has verified that the window of lexical features should not be too large, in order to prevent excessive noise. Usually, windows of 2-2 are selected. This paper chooses 2-3-2 mode, that is, select two words on entity one's left, two words on entity two's right and three words between the entities as features.
(3) Inclusion Information. Inclusion information mainly reflects the lexical information and inclusion relations between entities. This paper select the number of inter-entity vocabulary, number of entities and whether the entity is one with inclusion relationship as nested information.

2.2.2 Phrases Syntax Information

The syntax tree of phrases reflects the grammatical structure of sentences and expresses the semantic information over long distances. In the sentence "In 2004, Huangshan was first selected as a World geological park, making itself a tourist destination which won honors of World Cultural Heritage and Natural Heritage and World Geological Park at the same time.", its minimal complete syntax tree is shown in Fig. 2.

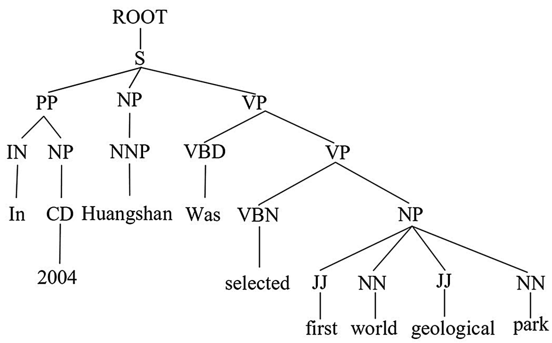

Fig. 2. Minimal complete syntax tree of an entity

Minimal complete syntax tree refers to one whose root node is the nearest public root node between two entities. Since the minimal complete syntax tree contains certain contextual information and reduces noise to some degree, this paper uses the minimal complete syntax tree to extract features. On account of the fact that paths between two entities are too specific in the syntax tree, the problem of sparse data is likely to occur. In order to avoid it, this paper regards the number of nodes in two entities' paths and the types of two entities' root nodes as features. Due to rather specific structural information contained in the syntax tree, low recall rate are to follow in addition to data sparseness.

2.2.3 Dependency Information

Dependency tree reveals long-distance dependencies of sentences, avoids the noise occurring in unstructured features, and provides more useful information for relationship extraction. For the sentence "In 2004, Huangshan was first selected as a World geological park, making itself a tourist destination which won honors of World Cultural Heritage and Natural Heritage and World Geological Park at the same time.", the dependency tree is shown in Fig. 3. Likewise, the distribution of structural information in the dependency tree is very specific, and it may generate the problem of low recall rate.

2.3 Improved Radial Base of Training Matrixes

RBF kernel function is a kernel function invariant in translation, and its concrete expression is shown as the formula (2)

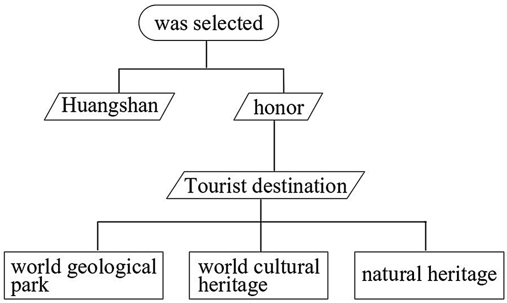

Fig. 3. Example of a dependency tree

$$rbf(x) = \exp(-a \cdot x + b) \tag{2}$$

When the coefficient a = 1, b = 0, the radial basis function is shown in Fig. 4. As can be seen from Fig. 4, when x increases, the function value quickly reaches extremely close to zero.

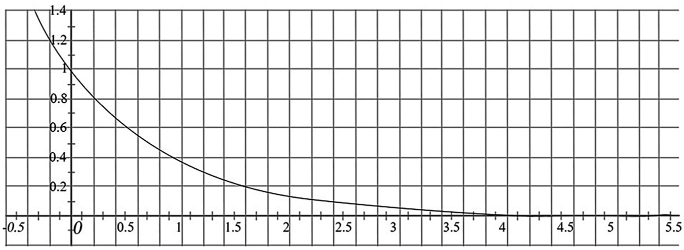

Fig. 4. Curve of radial basis functions

For two random feature vectors X_i, X_j, i, j•1,2,..., m, their maps form the j-th feature of i-th vectors in the new map space. Due to the property of radial basis functions, each character approaches very close to 0, making it very detrimental to classification. For example, there are three two-dimensional vectors $a, b, c,$ respectively: $a = (2, 3)$; $b = (4, 10)$; $c = (3, 5)$; the training matrixes of their corresponding space that they mapped to the radial basis functions are:

$$K_{a,b,c} = \begin{bmatrix} 0 & 9.602 \times 10^{-24} & 6.738 \times 10^{-3} \\ 9.602 \times 10^{-24} & 0 & 5.109 \times 10^{-12} \\ 6.738 \times 10^{-3} & 5.109 \times 10^{-12} & 0 \end{bmatrix} \tag{3}$$

As it can be seen, in the new feature space, three new vectors are very close to the origin, but data like these are quite unfavorable for classification. In order to solve problems mentioned above, this paper employs a method that limits each feature of the training matrix within an appropriate range to facilitate classification. For example, the range is limited to between 0.2 and 1. The standardized method of training matrix is:

(1) Calculate $Feature_{ij} = (Xi - Xj) \cdot (Xi - Xj)^T$, among them, $i,j \in 1,2,\ldots, m$;
(2) Enumerate each $Feature_{ij}$, find the maximum value, denoted as $Feature_{\max}$;
(3) The matrix is multiplied by a constant $\delta = (-ln_{0.2})/Feature_{\max}$;
(4) The training matrix is standardized as:

$$K_{new} = [k(Featureij \times \delta)]_{m \times m} \tag{4}$$

Among them, $i,j \in 1,2,\ldots,m$。

If three vectors mentioned previously, a, b, c, have been standardized, the training matrix will be:

$$K_{new} = \begin{bmatrix} 0 & 0.2000 & 0.8593 \\ 0.2000 & 0 & 0.4541 \\ 0.8593 & 0.4541 & 0 \end{bmatrix} \tag{5}$$

It can be perceived that three standardized vectors are more distant than that without standardizing in spacing.

2.4 Getting the Optimal Convex Combination Kernel Function

In terms of varied Chinese relation extraction, there are differences in effects for different kernel functions. In order to make kernel functions perform good adaptability on characteristic represented by different pieces of information, this paper will make convex combination of different single-core functions, expecting that the optimal convex combination of kernel functions also possess good adaptability. According to different expressions of kernel functions, they can be divided into translational invariant kernel functions and inner product kernel functions, and their expressions are: $\mathrm{k}(x, y) = f(x - z)$ and $\mathrm{k}(x, y) = f(<x, z>)$, respectively.

In order to fuse these two kernel functions' features in the relationship extraction, this paper focuses on three single-core functions, which include radical kernel functions, polynomial kernel functions and linear kernel functions, as shown in formulas (6), (7) and (8), where x, y are two vectors with arbitrary dimension.

$$K_{RBF}(x, y) = \exp(-|x - y|^2) \tag{6}$$

$$K_{polynomial}(x, y) = (x \cdot y^T)^3 \tag{7}$$

$$K_{linear}(x, y) = x \cdot y^T \tag{8}$$

The compound of various kernel functions offers a solution to the problem that the single-core functions do not have universal properties. For convex combination kernel functions, determining parameters of the combination ratio is of great importance. Suppose there are n kinds of kernel functions k1,…, kn, parameters options of combination ratios may have m kinds; then enumerate all the computational complexity of the kernel functions as O(m^n). And the n kinds of kernel functions of convex combination kernel functions are shown in formula (9), where a_i is called convex combination parameter, $\sum_{i=1}^{n} a_i = 1$.

$$CCK = \sum_{i=1}^{n} a_i k_i \tag{9}$$

This paper presents a strategy that by constantly looking for the optimal convex combination kernel function of these two kernel functions based on some principles to substitute these two kernel functions, so as to find out the optimal convex combination kernel function of various kernel functions. This strategy can reduce computational complexity to O (m*(n−1)). The algorithm to get the optimal convex combination of kernel functions is shown in Fig. 5.

Generation Algorithm of Optimal Convex Combined Kernel Functions

Algorithm Input: n kinds kernel functions PPI extracted performance collection F

Algorithm Output: Optimal Convex Combined Kernel Functions (OCCK)

Algorithm Steps:

```
While(number(F)>1)
candidate1←min(F)
F=F-min(F)
candidate2←min(F)
F=F-min(F)
F=F+Optimal(candidate1,candidate2)
if number(F)=1
then
Return Optimal(candidate1,candidate2)
```

∆number(F) is the number of F_i in collection F, min(F) is the minimum in collection F;

∆Optimal (k1,k2) is the optimal kernel function of kernel functions k1,k2

Fig. 5. Generation algorithm of optimal convex combination of kernel functions

After the kernel function is determined, feature matrix can be mapped to high-dimensional matrix by formula (9).

2.5 Finding the Optimal Classification Model

Via Sect. 2.4, a high-dimensional matrix which is mapped by different convex combinations of the kernel functions can be acquired. It is not probable to tell classification model obtained from high dimensional matrix training possess better extraction performance in entity relations, that is, which convex combination of kernel functions has better adaptability, by direct observation of high dimensional matrixes. In order to get the optimal convex combination of kernel function, it is a top priority to train each high-dimensional matrix and to obtain the corresponding training model. By running tests on the corpus, a subsequent step is to test out a training model with the best performance in extraction of entity relations. The corresponding kernel function of this training model is the optimal convex combination of kernel function.

3 Experimental Results and Discussion

3.1 The Experimental Dataset

The corpus used in this paper has been manually acquired by us from the Web and relevant literature, more than 1000 Chinese tourism texts. Among various machine learning algorithms applicable to kernel functions in the field of entity relation extraction, support vector machines (SVMs) have the best performance. So this paper chooses SVM as the machine learning algorithms, taking LIBSVM developed by Lin Zhiren from Taiwan University as SVM tools [20]. In the analysis of phrases, Stanford Parser [21] is used to conduct phrases parsing and dependency study, selecting probability context-free grammars. CCprocessed dependency expression is used to form the dependency information. As for training, a method of 10-fold cross-validation is employed to maximize the use of data. The performance is measured by precision, recall, and F-measure.

$$F = \frac{2 \times P \times R}{P + R} \times 100\% \tag{10}$$

3.2 Experiments

Experiment 1, on the lexical information basis of word, speech, etc. in contexts of entities, the syntactic information and dependency syntactic information of phrases are added as features, to find the impacts on the extraction performance of entity relation in Chinese. In the test, different features are added, with a purpose to find the optimal feature combination. Parameters in formula (9), $\alpha_1 = 1$, $\alpha_2 = 0$, $\alpha_3 = 0$ are chosen as the weighting parameter of convex combined kernel functions.

Experiment 2, comparing convex combined kernel functions composed of three single-core functions with single kernel functions and combinations of two kernel functions in relation extraction performance, In this test, coefficients of the convex combination in formula (9) are set, with the upper and lower limits from 0 to1, steps of 0.1, 11 options in total. By means of enumeration, the best coefficients can be determined. For n kinds of single-core functions, it requires a calculation of $11 \times (n - 1)$ times to find the best convex combination.

Experiment 3, in order to verify the effectiveness of the system proposed in this paper, a comparison with other methods has been conducted under the same corpus used in this paper.

3.3 Results and Discussion

3.3.1 Performance of Different Features in Relation Extraction

As it can be seen from Table 2, the entity relation extraction performance is unsatisfactory when only lexical information is selected as features in the field of Chinese tourism. After syntax and dependency information of phrases are added, the extraction performance improves, particularly with an increase in recall rate. It indicates that it is an effective method, which increases the performance of entity relation extraction in the field of Chinese tourism, to include syntax and dependency syntactic information of phrases to features.

Table 2. Performance under different features

Features	Accuracy	Recall Rate	F value
Lexical information	61.3 %	48.4 %	53.7 %
+ Syntactic information of phrases	62.6 %	52.7 %	57.1 %
+ Dependency of syntactic information	62.9 %	53.5 %	57.8 %

3.3.2 Comparison of Relation Extraction Performance Between Three Single-Core Functions and Convex Combined Kernel Functions

As Table 3 shows, in the entity relation extraction performance of single kernel function concerning Chinese tourism, polynomial kernel functions(PKF) have highest accuracy and F value; Radial kernel functions(RBF) have the highest recall rate; the performance of linear kernel function(LKF) is rather poor. These have also verified that single-core functions have differences in extraction performance in regard to the same feature matrix. In combinations of two convex combinations of kernel functions, the convex combination kernel functions, consisted of radial basis kernel functions and polynomial kernel functions, have the best performance, proving that the performance of convex combination bases has a positive impact on the performance related to convex combination of kernel functions. The best extraction performance is the convex combination of three single-core functions. And the performance of convex combinations of two single-core functions is superior to that of single-core functions. It can be described that convex combinations of single-core functions can solve the problem that single-core functions do not possess universality property.

3.3.3 Comparison with Other Methods

Table 4 suggests that the use of maximum entropy method in the field of Chinese tourism for relation extraction generates the highest accuracy. Using the shortest path tree in convolution tree kernel functions in relation extraction can get the best recall rate.

Table 3. Performance of single-core functions and convex combination kernel function

Categories of kernel functions	Accuracy	Recall Rate	F value
RBF	62.7 %	53.4 %	57.65 %
PKF	70.1 %	49.5 %	57.78 %
LKF	65.8 %	51.9 %	58.1 %
RBF + LKF	63.7 %	53.9 %	58.5 %
RBF + PKF	72.6 %	55.5 %	62.9 %
LKF + PKF	70.5 %	49.7 %	58.3 %
RBF + LKF + PKF	72.6 %	56.1 %	63.4 %

The method of optimal convex combination of kernel functions proposed in this paper has better recall rate and accuracy and the best F value, which fully verifies the effectiveness of the proposed optimal convex combination method.

Table 4. Comparison with other methods

Methods	Accuracy	Recall Rate	F value
Maximum entropy	74.1 %	48.9 %	58.9 %
Convolution tree kernel	61.3 %	58.9 %	60.1 %
Proposed method	72.6 %	56.1 %	63.4 %

4 Conclusions

Based on the lexical information of word, speech, etc. in entity contexts, this paper includes syntactic and dependency information of phrases as features. With different convex combinations of radial kernel functions, polynomial kernel functions and linear kernel functions, the feature matrix is mapped to a different high-dimensional matrix. To obtain different classification models, training support vector machine is used; and the optimal performance of the classification model is found by enumerating. Finally, the classification model is used in the field of Chinese entities' relation extraction. In the field of tourism, the optimal convex combination kernel extraction system proposed in this paper has achieved an F value of 63.4. In the further studies, efforts will be made to unveil other useful information as features, to find the deep relationship between kernel functions and the corpus, and to further improve the performance of relation extraction concerning entities in Chinese.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China (Grant Nos. 61262041, 61472168 and 61562052) and the key project of National Natural Science Foundation of Yunnan province (Grant No. 2013FA030).

References

1. Zhao, J., Liu, K., Zhou, G.Y.: Open information extraction. J. Chin. Inf. Process. **25**(6), 98–110 (2011)
2. Aone, C., Ramos-Santacruz, M.: Rees: A large-scale relation and event extraction system. In: Proceedings of the 6th Applied Natural Language Processing Conference, pp. 76–83. ACM Press, New York (2000)
3. Califf, M.E., Mooney, J.: Bottom-up relational learning of pattern matching rules for information extraction. J. Mach. Learn. Res. **4**, 177–210 (2003)
4. Zhou, G., Su, J., Zhang, J.: Exploring various knowledge in relation extraction. In: ACL, June 2005, pp. 427–434 (2005)
5. Dong, J., Sun, L., Feng, Y.Y.: Chinese automatic entity relation extraction. J, Chin. Inf. Process. **21**(4), 80–85 (2007)
6. Ye, F., Shi, H., Wu, S.: Research on pattern representation method in semi-supervised semantic relation extraction based on bootstrapping. In: 2014 Seventh International Symposium on Computational Intelligence and Design (ISCID). IEEE(2014)
7. Komachi, M., Kudo, T., Shimbo, M., Matsumoto, Y.: Graph-based analysis of semantic drift in espresso-like bootstrapping algorithms. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, pp. 1011–1020 (2008)
8. Zeng, D., Liu, K., Lai, S.: Relation classification via convolutional deep neural network. In: Proceedings of COLING (2014)
9. Liu, K.B., Li, F., Liu, L., Han, Y.: Implementation of a kernel-based Chinese relation extraction system. J. Comput. Res. Dev. **44**(8), 1406–1411 (2007)
10. Zhuang, C.L., Qian, L.H., Zhou, G.D.: Research on tree kernel-based entity semantic relation extraction. J. Chin. Inf. Process. **23**(1), 3–9 (2009). ISSN: 1003-0077
11. Yang, Z., Tang, N., Zhang, X., et al.: Multiple kernel learning in protein–protein interaction extraction from biomedical literature. J. Artif. Intell. Med. **51**(3), 163–173 (2011)
12. Peng, C., Gu, J., Qian, L.: Research on tree kernel-based personal relation extraction. In: Zhou, M., Zhou, G., Zhao, D., Liu, Q., Zou, L. (eds.) NLPCC 2012. CCIS, vol. 333, pp. 225–236. Springer, Heidelberg (2012). doi:10.1007/978-3-642-34456-5_21
13. Arenas-García, J., Martínez-Ramón, M., Gómez-Verdejo, V., Figueiras-Vidal, A.R.: Multiple plant identifier via adaptive LMS convex combination. In: Proceedings of the IEEE International Symposium on Intelligent Signal Processing, Budapest, Hungary, pp. 137–142 (2003)
14. Arenas-García, J., Figueiras-Vidal, A.R., Sayed, A.H.: Mean-square performance of a convex combination of two adaptive filters. IEEE Trans. Signal Process. **54**(3), 1078–1090 (2006)
15. Knowloge- base, CYC. http://www.cyc.com/2008
16. Miller, G.: Introduction to wordnet: an on-line lexical database. Int. J. Lexicograhy **3**(4), 235–3244 (1990)
17. Dong, Z.D., Dong, Q.: National Knowledge Infrastructure (2005)
18. Suchanek, F.M., Kasneci, G., Weikum, G.: Yago: a core of semantic knowledge. In: 16th International World Wide Web Conference (WWW2007). ACM Press, New York (2007)
19. ICTCLAS tool from Chinese Academy of Sciences. http://ictclas.nlpir.org/downloads
20. LIBSVM developed by Lin, Z.R from Taiwan University. http://www.csie.ntu.edu.tw/~cjlin/libsvm
21. Stanford Parser. http://nlp.stanford.edu/software/lexparser.shtml