Information Technology Journal, 2010
ISSN 1812-5638

Chinese Event Extraction Based on Feature Weighting

Jianfeng Fu, Zongtian Liu, ZhaoMan Zhong and Jianfang Shan
School of Computer Engineering and Science, Shanghai University, Shanghai, China

Abstract: In this study, we focus on the two subtasks of Chinese event extraction: (1) Chinese event detection and identification; (2) Chinese event argument extraction. Some features for the two subtasks are provided. Considering the particular contributions of different features on classification analysis in the subtasks, we weight features by introducing ReliefF algorithm. Experimental results show that comparing with normal K-Nearest Neighbor algorithm, feature weighting obviously improves the F-Measures in Chinese event detection and identification and Chinese event argument extraction.

Key words: Chinese event extraction, feature weighting, ReliefF algorithm, ACE

INTRODUCTION

The Automatic Content Extraction (ACE) program is funded by National Institute of Standards and Technology (NIST). The objective of the ACE program is to develop automatic content extraction technology to support automatic processing of source language data. Event extraction has been defined as a fundamental task in the ACE program. The task mainly involves: (1) detecting specified types of events and (2) extracting information about these events. The information includes event attributes, event arguments and event mentions. Event attributes are event type, subtype, modality, polarity, genericity and tense. An event argument is an entity, or a temporal expression or a value that plays a certain role in the event. An event mention is a sentence or phrase within which the event is described. The event extraction of ACE program is supported for two languages (Chinese and English). In this study, we focus on the two subtasks of Chinese event extraction: (1) Chinese Event Detection and Identification (CEDI); (2) Chinese Event Argument Extraction (CEAE).

In ACE Chinese corpus (Consortium, 2005), 8 types and 33 subtypes of events are annotated. The event types and subtypes are life (Be-born, Marry, Divorce, Injure and Die), movement (Transport), transaction (Transfer-ownership, Transfer-money), business (Start-org, Merge-org, Declare-bankruptcy, End-org), conflict (Attack, demonstrate), contact (Meet, Phone-write), personnel (Start-position, End-position, Nominate, Elect), justice (Arrest-jail, Release-parole, Trial-hearing, Charge-indict, Sue, Convict, Sentence, Fine, Execute, Extradite, Acquit, Appeal, Pardon). Each subtype of an event has its own arguments. An example of Be-born event in the ACE Chinese corpus is shown as follows:

Table 1: Results for event extraction

Type		Arguments		
Event type	Event Subtype	Person-Arg	Time-Arg	Place-Arg
Life	Be-born	赖斯女士	1954年	亚拉巴马 州的伯明 翰

赖斯女士 1954 年出生于亚拉巴马州的伯明翰

Ms. Rice was born in 1954 in Birmingham, Alabama.

For the example, the task of CEDI is to identify the event type (life) and event subtype (Be-born). The task of CEAE is to extract the three arguments (赖斯女士, 1954年, 亚拉巴马州的伯明翰), which involved in the event. The results of event extraction are shown in Table 1.

Recent years, some event extraction systems have been reported. Ahn (2006) presented a simple, modular approach to event extraction. The K-Nearest Neighbor (KNN) and Maximum Entropy (ME) classifiers were introduced to his systems for extracting ACE English event. Tan *et al.* (2008) presented strategies of feature selection and pattern matching to extract ACE Chinese event. Chen and Ji (2009) presented an ACE Chinese event extraction system based on feature selection and detailed analyzed the selected features. Most of their study focused on how to select appropriate features to improve the event extraction performance, however, ignored the particular contributions of the different features on classification analysis. In this study, we present a novel method of Chinese event extraction. ReliefF algorithm (Robnik-Sikonja and Kononenko, 2003), a well known feature weighting algorithm, is employed to weight the features in present method. By using feature

Corresponding Author: Jianfeng Fu, School of Computer Engineering and Science, Shanghai University, Shanghai, China

weighting, different features are assigned with different weights according to their particular contributions on classification analysis. Experimental results show that the feature weighting dramatically improves the system performance.

FEATURES FOR CHINESE EVENT EXTRACTION

Extracting ACE Chinese event is a complex task. For simplicity, we break down the task into two subtasks, CEDI and CEAE. The goal of each subtask is defined as:

- **CEDI:** Finding event trigger (the main word which most clearly expresses an event occurrence) in text and assigning it an appropriate event type
- **CEAE:** Determining which entity mentions, temporal expressions and values are arguments of each event mention

In present experiments, the K-Nearest Neighbor (KNN) classifier is employed in the two subtasks to extract Chinese event. For the classification learner, we need a range of information to build feature vectors. In CEDI, the features which we selected are as follows:

- **Lexical features:** The current word w_i, the POS tag of w_i
- **Semantic feature:** The semantic tag of w_i
- **Context features:** The left word w_{i-1}, the POS tag of w_{i-1}, the semantic tag of w_{i-1}, the right word w_{i+1}, the POS tag of w_{i+1}, the semantic tag of w_{i+1}
- **Dependency features:** The dependency relations of w_i, w_{i-1} and w_{i+1} in the sentence. The dependency relation involves (1) the tag of the dependency relation and (2) the depth of the word in the dependency tree

In CEAE, all the features which mentioned above are covered. Besides, we add the following features:

- **Related trigger features:** The trigger t, the event type of t, the POS tag of t, the semantic tag of t, the dependency relation of t in the sentence
- **Position feature:** The word w_i before or after the trigger t

WEIGHTING FEATURES

The problem of weighting the quality of features is an important issue in the machine learning and has received much attention in the literature (Wettschereck *et al*., 1997; Modha and Spangler, 2003; Sun, 2007). Not only can feature weighting reduce system complexity and processing time, but it can also enhance system performance in many cases.

In traditional KNN algorithm, a hypothesis is implied that all the features have the same contributions to the classification while computing the similarity between two objects. However, intuitively, different features have different importance, ignoring their contributions may affect classification performance. An appropriate approach is to compute weights for different features. Each feature has a weight associating with it.

Given a data set $X = \{x_1, x_2, ..,x_n\}$, where n is the number of the objects, x_i is the ith object in X. Feature vector is $F = \{f_1, f_2,...,f_m\}$, where f_j is the jth feature. Function (f_j, x_i) gives the weight of x_i on feature f_j. Feature weight vector is $W = \{w_1, w_2,...,w_m\}$, where w_j is a weight which assigned to feature f_j. For two objects x_a and x_b, their similarity $Sim(x_a, x_b)$ is defined as:

$$\mathrm{Sim}(x_a, x_b) = \frac{\sum_{j=1}^{m} w_j E\left(\mathrm{value}(f_j, x_a), \mathrm{value}(f_j, x_b)\right)}{\sum_{j=1}^{m} w_j} \tag{1}$$

where, function:

$$E\left(\mathrm{value}(f_j, x_a), \mathrm{value}(f_j, x_b)\right) = \begin{cases} 1, \text{ if } \mathrm{value}(f_j, x_a) = (f_j, x_b) \\ 0, \text{ otherwise} \end{cases} \tag{2}$$

We normalize the similarity by dividing it with $\sum_{j=1}^{m} w_j$ to ensure $Sim(x_a, x_b) \in [0, 1]$.

For computing feature weights, ReliefF algorithm, a well known weighting algorithm, is employed in our method.

Suppose we randomly select an object x in data set X, ReliefF searches l nearest neighbors p_r, r = 1,...,l from the same class and also l nearest neighbors q_r, r = 1,,,.l from each of the different classes. For each f in feature vector F, its weight w is defined as:

$$\begin{aligned} w = w &- \sum_{r=1}^{l} \mathrm{diff}(f, x, p_r) \\ &+ \sum_{q_r \notin \mathrm{class}(x)} \left[\frac{P(\mathrm{class}(q_r))}{1 - P(\mathrm{class}(x))} \sum_{r=1}^{l} \mathrm{diff}(f, x, q_r) \right] \end{aligned} \tag{3}$$

where, P(class(x)) is the probability of class which x is contained, it is defined as:

$$P(\mathrm{class}(x)) = \frac{\text{No. of objects in class(x)}}{\text{No. of objects in data set X}} \tag{4}$$

For discrete features, function diff(f, x, p_r) is defined as:

$$\mathrm{diff}(f,x,p_r)=\begin{cases}0, \text{if } \mathrm{value}(f,x)=\mathrm{value}(f,p_r)\\ 1, \text{otherwise}\end{cases} \quad (5)$$

For each feature f, we calculated its weight w by using Eq. 3 to get weight vector W. The whole process is repeated for t times, where t is a user-defined parameter. In the first step of the repeating process, the w on the right side of the equal sign is initialized to 0. The calculated w on the left side will be assigned to the w on the right side in the next step until the repeating process is terminated.

RESULTS AND DISCUSSION

Experimental data: We used 2005 ACE training corpus for our experiments. The corpus contains 633 Chinese documents. We split this corpus into training and test sets at the document level, with 573 training documents and 60 test documents. Each document is split into sentences. These sentences are parsed by the LTP system developed by IR-Lab in HIT (http://ir.hit.edu.cn/) to obtain the feature information of the words.

Results of feature weighting: We repeated the process of calculating feature weights for 20 times in ReliefF algorithm. In Table 2, we present the results in details. The results contain the feature weights for CEDI and CEAE.

In Table 2, feature 1 represents the word information of the $w_i/w_{i-1}/w_{i+1}$/trigger; feature 2 represents the POS tag information of the $w_i/w_{i-1}/w_{i+1}$/trigger; feature 3 represents the semantic information of the $w_i/w_{i-1}/w_{i+1}$/trigger; feature 4 and feature 5 represent the dependency information of the $w_i/w_{i-1}/w_{i+1}$/trigger (feature 4 is the tag of the dependency relation and feature 5 is the depth of the word in the dependency tree); feature 6 represents the event type of the trigger; feature 7 represents the position information of the trigger with w_i.

Table 2: Feature weights for CEDI and CEAE

Word	Feature number	Feature weights	
		CEDI	CEAE
w_i	1	10	2
	2	5	6
	3	3	1
	4	6	6
	5	3	4
w_{i-1}	1	3	2
	2	2	2
	3	1	1
	4	2	2
	5	1	1
w_{i+1}	1	3	2
	2	3	2
	3	1	1
	4	2	2
	5	1	1
Trigger	1	/	1
	2	/	1
	3	/	2
	4	/	2
	5	/	1
	6	/	2
	7	/	1

From the Table 2, we see most feature weights of current word w_i are higher then the others. Thus, we can come to the conclusion that the features of the current word w_i are more important and discriminative than the others for CEDI and CEAE. However, the weights of the feature 1 of the current word w_i for CEDI and CEAE are very different (10 vs. 2). It can be explained that, for example, a Chinese Be-Born event might be described as a pattern of [Entity] [Time]出生于(born in) [Place], the arguments for Entity, Time and Place can be assigned with various allowable values to describe different Be-Born event instances, conversely, triggers for the Be-Born event are numbered, furthermore, if the current word is 出生 (born) or other words, for example, 诞生 (naissance), in a sentence, it is probably described a Be-Born event. For this reason, the word information of w_i is more discriminative in CEDI and the weight is correspondingly higher than in CEAE.

In addition, we also observe that the feature weights of the left word w_{i-1} and right word w_{i+1} are almost equal. It can be understood that the features of w_{i-1} and w_{i+1} are effective as fairly as possible in CEDI and CEAE.

Comparison of systems performance: We compare present system with Baseline (KNN only, no feature weighting) on CEDI and CEAE. The system performance is evaluated with Precision (P), Recall (R) and F-Measure (F). For the KNN classifier, we set the parameter k = 5.

In Table 3, we represent the comparison results of CEDI between Baseline and Feature Weighting. It can be seen that the Feature Weighting improves the F-measure in each event type and enhances the F-Measure of Macro-average (Form 70.1 to 78.4%). Table 4 shows the

Table 3: Comparison results for CEDI

Event type	Baseline (%)			Feature weighting (%)		
	P	R	F	P	R	F
Life	81.7	73.3	77.3	88.9	82.7	85.70
Movement	69.1	62.8	65.8	76.8	70.5	73.50
Transaction	50.3	40.9	45.1	63.5	53.9	58.30
Business	79.5	72.4	75.8	87.4	80.1	83.63
Conflict	76.4	71.0	73.6	84.9	77.8	81.20
Contact	72.8	69.1	70.9	82.6	76.4	79.40
Personnel	76.6	70.6	73.5	82.9	79.0	80.90
Justice	81.9	75.4	78.5	87.1	82.8	84.90
Macro-average	73.5	66.9	70.1	81.8	75.4	78.40

Table 4: Comparison results for CEAE

	P	R	F
		(%)	
Baseline	50.2	45.4	47.7
Feature weighting	55.8	51.8	53.7

comparison results of CEAE between baseline and feature weighting. It also indicates that the feature weighting performs better than the baseline (6% improvement in F-Measure). By introducing feature weighting, present method achieves significant improvement on CEDI and CEAE.

From Table 3 and 4, we also observe that the feature weighting based CEDI performs better than CEAE (78.4 vs. 53.7% in F-Measure). Compared with CEDI, CEAE is more complex and difficult because, as mentioned above, the Chinese event arguments are variable in human language.

In present experiments, it is found that feature weighting plays an important role in Chinese event extraction. By employing ReliefF algorithm, the particular contributions of the features which selected in Chinese event extraction are nicely quantified. An important feature is assigned with a higher weight, or vice versa. Compared with baseline, feature weighting obviously enhance the F-Measure in CEDI and CEAE. Therefore, Feature weighting is helpful to improve the performance of Chinese event extraction.

CONCLUSIONS AND FUTURE WORK

Event extraction is a complex and challenging task in ACE program. In this study, we have presented a novel method for Chinese event extraction. Considering the particular contributions of different features on classification analysis, ReliefF algorithm, a feature weighting algorithm has been introduced to our method. By weighting features, different weights were assigned to different features according to their importance of contribution on classification analysis. Then the feature weights were applied to KNN to extract Chinese event. The comprehensive experimental results demonstrated that the feature weighting dramatically improved the F-Measure in CEDI and CEAE.

As we have seen, the performance of CEAE was not so satisfied. For future research, we intend to explore more effective machine learning methods to improve our system.

ACKNOWLEDGMENTS

This study is supported by the projects of National Science Foundation of China (NSFC No. 60575035, 60975033), Shanghai Leading Academic Discipline Project (J50103) and Postgraduate Innovation Fund of Shanghai University (SHUCX091041).

REFERENCES

Ahn, D., 2006. The stages of event extraction. Proceedings of the COLING-ACL 2006 Workshop on Annotating and Reasoning about Time and Events, Jul. 23, Sydney, Australia, pp: 1-8.

Chen, Z. and H. Ji, 2009. Language specific issue and feature exploration in chinese event extraction. Proceedings of the North American Chapter of the Association for Computational Linguistics-Human Language Technologies, Jun. 1-3, Boulder, Colorado, pp: 209-212.

Consortium, L.D., 2005. ACE (Automatic Content Extraction) chinese annotation guidelines for events. http://projects.ldc.upenn.edu/ace/docs/Chinese-Entities-Guidelines_v5.5.pdf.

Modha, D. and W. Spangler, 2003. Feature weighting in k-means clustering. Machine Learn., 52: 217-237.

Robnik-Sikonja, M. and I. Kononenko, 2003. Theoretical and empirical analysis of ReliefF and RReliefF. Machine Learn., 53: 23-69.

Sun, Y., 2007. Iterative RELIEF for feature weighting: Algorithms, theories and applications. IEEE Trans. Pattern Analysis Machine Intell., 29: 1035-1051.

Tan, H., T. Zhao and J. Zheng, 2008. Identification of chinese event and their argument roles. Proceedings of the IEEE 8th International Conference on Computer and Information Technology Workshops, Jul. 08-11, IEEE Computer Society, Sydney, Australia, pp: 14-19.

Wettschereck, D., D. Aha and T. Mohri, 1997. A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms. Artificial Intell. Rev., 11: 273-314.