Mu Tong

1004452

Research Plan of Chinese Event Detection in Social Media

1. Motivation

Analyzing the data on online social networks draws attention in the last few years according to the greatly increasing development of social media [5], such as Twitter, Facebook, and YouTube. People use these social medias not only to socialize, but also do other things [5], which is of great value [1]. For example, News media could use social media like Twitter to spread news, because people are willing to spread something important [5]; some companies can analyze the opinions from their customers to get ideas about their products or services [2], which is called "sentiment analysis" technique; Markets can analyze their selling data and customer profiles to identify the market trend [3] or provide personal advertisements [4], which is called "personalized recommendation" [1]; there are still a lot of applications and the list goes on.

Event detection is one of these applications, and it is highly noticeable due to its difficulties and social impact [1]. As for the definition of event detection, briefly speaking, "event detection is the problem of automatically identifying significant incidents by analyzing social media data" [1], such as a baseball game or a big traffic crash.

There are five reasons why event detection in social media is more challenging [1]: volume and velocity, real-time event detection, noise and veracity, feature engineering, and evaluation. Briefly speaking, there are plenty of short and noisy information in the content, rapidly changing topics, and great volume data [5].

As for Chinese event detection, it is even harder to handle than English event detection. There are two reasons why this is more challenging. The first reason is the problem of Chinese word segmentation [8], which would not happen to English text. In English, words in a sentence can be easily separated because of the blank space, but in Chinese, words in a sentence are linked together. Therefore, word segmentation is necessary to divide a sentence into a sequence of words in Chinese. The mistakes of word segmentation may result in that the event words called event trigger (which will be explained in detail in next section) consist of multiple words or only one word. The second reason is that Chinese has lots of polysemy, which means the grammar is more complex than English. For example, the same word usually has different meanings in different contexts. Therefore, the semantic feature in Chinese event detection is increasingly important.

Chinese event detection has developed well in recent years. Several papers stated that word embedding plays a significant role in Chinese event detection, such as [9]. By the way, word embedding is a technique in natural language processing (NLP), which means that words are mapped into vectors of real numbers in computers, therefore, computers could easily deal with these words based on vectors. In addition, compared with normal machine learning algorithms, deep learning algorithms, such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), make better performances, such as [10].

Most of the studies only considered the advantages of word embedding, but not semantic features. Therefore, we propose a research question: Can a word embedding technique combined with different semantic features be used to improve the F1-measure value of Automatic Context Extraction (ACE) Chinese event detection task using Bidirectional Long Short-Term Memory (Bi-LSTM) network in social media?

The next section will explain more details of the identification of research question.

2. *Identification*

We are focusing on the Chinese event detection task on Automatic Context Extraction (ACE) [11] [12], and ACE has defined some basic concepts:

Event is "something that happens and can frequently be described as a change of state."

Event trigger: "the word that most clearly expresses the event occurrence."

Event mention: a sentence including at least one trigger, which means an event happens.

There are 8 types and 33 subtypes for event types, such as life, movement and conflict. The more details are showed in Table I.

Event Types	Event Subtypes
Life	be-born, marry, divorce, injure, die
Movement	transport
Transaction	transfer-ownership, transfer-money
Business	start-org, merge-org, declare-bankruptcy, end-org
Conflict	attack, demonstrate
Contact	meeting, 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

Table I: Details of Event Types in ACE

The task is described as follows.

Let S = $W_1W_2W_3$, …, Wn be a sentence with n words after word segmentation, where n is the length of the word list. In addition, Let L = $E_1E_2E_3$, …, E_n be labels for each word in a sentence. For every word W_i, where i means the i-th token, we should predict its label E_i whether it is a trigger, and which event subtype it should belong to.

Therefore, event detection task is divided into two parts: recognition of event trigger and its classification. The detection of trigger means that we should determine every word in a sentence whether it is a trigger for event subtype, and trigger classification means that we should classify each trigger into a correct event subtype.

3. *Methods*

To solve this task, three main methods will be used. One is the word embedding, the second is feature representation, and the final one is Bidirectional Long Short-Term Memory (Bi-LSTM) network [13] [14]. The flowchart for this model is showed as Figure I.

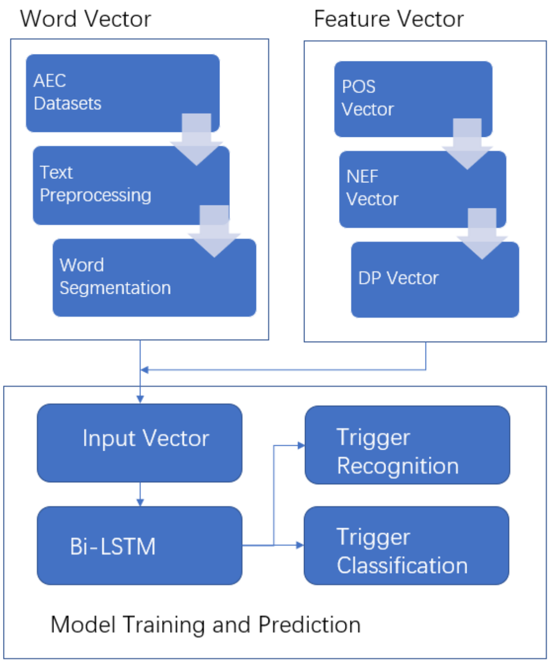

Figure I: Event Trigger Detection Flowchart

3.1 Word Embedding

Word embedding is necessary for computers to convert words into numerical symbol in order to easily process and calculate. In traditional, One-hot representation is used a lot in word embedding. In one-hot representation, words are considered as a long binary vector, and the length of this vector is the length of the dictionary. Only one dimension is 1, and others will be 0. The example is showed in Figure II.

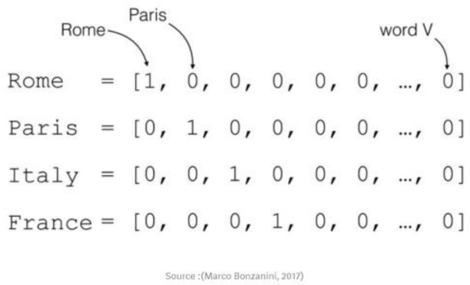

Figure II: Example for One-Hot Representation, from:

https://medium.com/@athif.shaffy/one-hot-encoding-of-text-b69124bef0a7

There are two problems for this presentation. One is that the similarity between the two words can't be calculated, and the other is the dimension for each vector is extremely large, and lots of dimension is useless in practice.

Therefore, we used a distributed representation for word embedding. In a low-dimensional word vector, a fixed length vector is used through training usually 50 or 100 in practice. Then, all the word vectors will be formed a vector space. This method has much smaller dimension and the distance between word vectors, which means the similarity between words, than one-hot representation. In this paper, we will use word2vec model, which was proposed by Mikolov [15]. Word2vec is famous for distributed representation, including two methods: skip-gram model and continuous bags of words (CBOW). Skip-gram model is to predict the context word from word vectors, and CBOW is to predict the target vector from context words. Therefore, we will use skip-gram model word2vec to form the basic vector space.

3.2 Feature Representation

We will use three types of feature representation in order to add more semantic features into word vector space, including part-of-speech (POS) vector, Named Entity Feature (NEF) vector and dependency parsing (DP) [16][17]. They are all binary representation model, which means only one dimension is one, and other dimensions are zeros.

3.2.1 *POS vector*

The aim of this method is to assign a word type (such as verb, noun, adjective or other types) to words in a sentence. POS vector plays an important role in Chinese event detection [17]. According to the analysis in [17], they found event trigger words are often verb or noun. Therefore, using this method could greatly improve the accuracy of finding event trigger. There are some details of POS with 25 dimensions in Figure III from [16].

ID	Label	Dsecription	ID	Label	Dsecription
0	nt	time noun	13	c	conjunction
1	m	number	14	nl	place noun
2	wp	punctuation	15	nh	person noun
3	n	common nouns	16	j	abbreviations
4	ns	geographic nouns	17	ws	foreign word
5	nd	direction noun	18	b	noun modifier
6	q	quantifier	19	r	pronoun
7	v	verb	20	i	idiom
8	p	preposition	21	z	descriptive words
9	u	auxiliary verb	22	k	suffix
10	d	adverb	23	ni	organization name
11	nz	proper nouns	24	o	onomatopoeia
12	a	adjective			

Figure III: Details of POS vector

3.2.2 *NEF vector*

The target for this method is to determine the special meaning for words in a sentence, such as space name, institution name or human name, which really helps the recognition of trigger word. The details of NEF vector with 13 dimensions are showed in Figure IV from [16].

ID	Tag	Meaning
0	O	The word is not a named entity
1	B-Ns	Begin of place name
2	E-Ns	End of place name
3	S-Ns	Place name
4	I-Ns	Middle of place name
5	B-Ni	Begin of institution name
6	I-Ni	Middle of institution name
7	E-Ni	End of institution name
8	S-Ni	Institution name
9	B-Nh	Begin of human name
10	E-Nh	Middle of human name
11	I-Nh	End of human name
12	S-Nh	Human name

Figure IV: Details of NEF vector

3.2.3 *DP vector*

The purpose of this method is to describe the dependency relationship between the center word and dependent word, which shows the semantic dependence. The details of DP vector with 14 dimensions are showed in Figure V from [16].

ID	Tag	Meaning
0	ATT	attribute
1	ADV	adverbial
2	RAD	right adjunct
3	IS	independent structure
4	CMP	complement
5	POB	preposition-object
6	COO	coordinate
7	VOB	verb-object
8	SBV	subject-verb
9	HED	head
10	FOB	fronting-object
11	LAD	left adjunct
12	DBL	double
13	IOB	indirect-object

Figure V: Details of DP vector

3.2.4 Examples for Feature Representation

Word	Type of Feature	Vector Value
Study	POS	00000000000000000100000000
	NEF	00000100000000
	DP	00000001000000
Phone	POS	00000000000010000000000000
	NEF	00000000000100
	DP	00001000000000
paper	POS	00000100000000000000000000
	NEF	00010000000000
	DP	00000000000100

Table II: Examples for Feature Representation

The final vector space combines the feature vector space with the word vector space, which is used to become an input in a neural network called Bidirectional Long Short-Term Memory (Bi-LSTM).

3.3 Bi-LSTM Network

Bidirectional Long Short-Term Memory network (Bi-LSTM) is kind of variation in RNN

(Recurrent Neural Network). The input of this method is a sequence of tokens and predict a sequence of labels corresponding to each token as output.

As for this task, the concatenation of word vector and feature vector as input will be put into two LSTM components in two directions: forward and backward. They both will capture the feature information and following information. Then, the result will be mapped into n dimension, where n is the dimension of word vector, resulting in a probability distribution of each word. And the specified type of the event trigger will be given due to the probability distribution.

4. *Analysis Method*

We will use AEC corpus [12] as our experiment datasets to train and predict the labels, and our evaluation method is using the precision, the recall, and the F1-measure, which is widely used in the field of text analyzing. The formula for these three criteria is showed as follows in Figure VI.

$$
\begin{aligned}
precision &= \frac{TP}{TP + FP} \\
recall &= \frac{TP}{TP + FN} \\
F1 &= \frac{2 \times precision \times recall}{precision + recall}
\end{aligned}
$$

Figure VI: Formula for Precision, Recall, and F-measure

Some identification should be clearly proposed. TP means that the number of event triggers correctly recognized, and FP means the number of event triggers falsely recognized, and FN means the number of missing triggers.

We will use these evaluation criteria to compare the results of different features vectors. The details of sample analysis table are showed in Table III.

Method	Trigger Recognition			Trigger Classification		
	Precision	Recall	F1-Measure	Precision	Recall	F1-Measure
BI-LSTM + word embedding						
BI-LSTM + word embedding + POS						
BI-LSTM + word embedding + NEF						
BI-LSTM + word embedding + DP						
BI-LSTM + word embedding + POS + NEF + DP						

Table III: Sample of Analysis Table

The baseline is the combination of Bi-LSTM networks and word embedding without any feature representation. From this table, we could get the information about the contribution of improving F1-measure for different feature representations, and so that we can answer the research question.

In addition, we could compare the results between different event types, so that we can get the information about which event type we could get higher F1-measure value, which means that we could predict the specified event type more accurately. The sample table is showed in Table IV.

Event Type	Precision	Recall	F1-Measure
Life			
Movement			
Transaction			
Business			
Conflict			
Contact			
Personnel			
Justice			

Table IV: Sample of Analysis Table

5. *Contribution*

This paper proposes a practical model that can recognize event trigger and classify them into specified event type in Chinese. According to the polysemy and complex grammar in Chinese, we will combine more semantic information into word vector, in order to make it more accurate. In addition, we will compare the results between different combinations of feature representation and get the information about which combination could get the best F1-measure value for event trigger detection in Chinese. During this model, we would prove that we could get a better performance for event trigger in Chinese, if we add a better combination of feature representation into word embedding.

References

[1] Panagiotou, N., Katakis, I., & Gunopulos, D. (2016). Detecting events in online social networks: Definitions, trends and challenges. In *Solving Large Scale Learning Tasks. Challenges and Algorithms* (pp. 42-84). Springer, Cham.

[2] Bifet, A., & Frank, E. (2010, October). Sentiment knowledge discovery in twitter streaming data. In International conference on discovery science (pp. 1-15). Springer, Berlin, Heidelberg.

[3] Mathioudakis, M., & Koudas, N. (2010, June). Twittermonitor: trend detection over the twitter stream. In Proceedings of the 2010 ACM SIGMOD International Conference on Management of data (pp. 1155-1158). ACM.

[4] Chen, K., Chen, T., Zheng, G., Jin, O., Yao, E., & Yu, Y. (2012, August). Collaborative personalized tweet recommendation. In Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval (pp. 661-670). ACM.

[5] Nurwidyantoro, A., & Winarko, E. (2013, June). Event detection in social media: A survey. In International Conference on ICT for Smart Society (pp. 1-5). IEEE.

[6] Dou, W., Wang, X., Ribarsky, W., & Zhou, M. (2012, October). Event detection in social media data. In IEEE VisWeek Workshop on Interactive Visual Text Analytics-Task Driven Analytics of Social Media Content (pp. 971-980).

[7] Petrović, S., Osborne, M., & Lavrenko, V. (2010, June). Streaming first story detection with application to twitter. In Human language technologies: The 2010 annual conference of the north american chapter of the association for computational linguistics (pp. 181-189). Association for Computational Linguistics.

[8] Chen, Z., & Ji, H. (2009, June). Language specific issue and feature exploration in Chinese event

extraction. In Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers (pp. 209-212).

[9] Nie, Y., Rong, W., Zhang, Y., Ouyang, Y., & Xiong, Z. (2015). Embedding assisted prediction architecture for event trigger identification. Journal of bioinformatics and computational biology, 13(03), 1541001.

[10] Chen, Y., Xu, L., Liu, K., Zeng, D., & Zhao, J. (2015, July). Event extraction via dynamic multi-pooling convolutional neural networks. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers) (pp. 167-176).

[11] Doddington, G. R., Mitchell, A., Przybocki, M. A., Ramshaw, L. A., Strassel, S. M., & Weischedel, R. M. (2004, May). The Automatic Content Extraction (ACE) Program-Tasks, Data, and Evaluation. In Lrec (Vol. 2, p. 1).

[12] Linguistic Data Consortium. (2009). ACE (Automatic Content Extraction) Chinese Annotation Guidelines for Events. 2009-09-08]. Version 5.5.1. http://www.ldc.upenn.edu/Projects/ACE.

[13] Zeng, Y., Yang, H., Feng, Y., Wang, Z., & Zhao, D. (2016). A convolution BiLSTM neural network model for Chinese event extraction. In Natural Language Understanding and Intelligent Applications (pp. 275-287). Springer, Cham.

[14] Feng, X., Qin, B., & Liu, T. (2018). A language-independent neural network for event detection. Science China Information Sciences, 61(9), 092106.

[15] Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

[16] Yin, M., Chen, X., Wang, H., Wang, Q., Ma, C., & Qin, X. (2019). An Automatic Event Detection Method for Massive Wireless Access Prediction. IEEE Access, 7, 113404-113416.\

[17] Wu, Y., & Zhang, J. (2018). Chinese Event Extraction Based on Attention and Semantic Features: A Bidirectional Circular Neural Network. Future Internet, 10(10), 95.