World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:17, No:8, 2023

Chinese Event Detection Technique Based on Dependency Parsing and Rule Matching

Weitao Lin

Open Science Index, Computer and Information Engineering Vol:17, No:8, 2023 publications.waset.org/10013234.pdf

***Abstract*—**To quickly extract adequate information from large-scale unstructured text data, this paper studies the representation of events in Chinese scenarios and performs the regularized abstraction. It proposes a Chinese event detection technique based on dependency parsing and rule matching. The method first performs dependency parsing on the original utterance, then performs pattern matching at the word or phrase granularity based on the results of dependent syntactic analysis, filters out the utterances with prominent non-event characteristics, and obtains the final results. The experimental results show the effectiveness of the method.

***Keywords*—**Natural Language Processing, Chinese event detection, rules matching, dependency parsing.

I. Introduction

WITH the advent of the information age, various media are filled with large-scale unstructured text data. For example, a Google search on the keyword "Russia-Ukraine situation" can search 9,260,000 pieces of relevant information. How to quickly extract adequate information from it and filter out repetitive and meaningless descriptive information can significantly reduce the workload of data analysts and improve their work efficiency, which can help public opinion analysis and policy making. It is natural for researchers to focus on this area. Although the tasks are the same, there are some differences in the "information sources" for this task in China and abroad, and the language expressions in Chinese are more variable than those in foreign languages. This paper introduces a Chinese event detection technique based on dependency parsing and rule matching and gives experimental results.

II. Related Work

The research on Chinese event detection methods has made significant progress compared to previous ones. However, there is still a lack of research on establishing a unified event framework and using basic grammatical expressions as a breakthrough. Many studies are still limited to specific domains and are more powerless for generalized corpora. Wu et al. [1] established a unified event framework in advance. They then built sentence templates to extract information from three types of news events: fire, mine, and air disaster, and their recall and precision reached 60.82% and 94.84%, respectively. Yang [2] summarized a set of information extraction models for breaking news events, clustered them based on word and sentence analysis, automatically obtained the information structure of events, and finally adopted the method of word string with information for breaking news event information extraction. Jiang [3] performed the normalization process for the extracted data by matching the text with predefined rules, and finally, the results were merged. Zhong and Chen [4] achieved an overall accuracy of 83.13% by classifying catastrophic events into three layers with a total of seven categories and determining information about various aspects of devastating events through a layer-based finite-state free machine.

Although establishing specific regular rules and word rules for domain-specific corpus can achieve particular results on that domain corpus, its limitations are also more apparent. Take the information extraction of catastrophic events as an example, it is mainly to extract the event itself by extracting the more focused contents such as the name of the event, the time of occurrence, the place of occurrence, and the number of casualties [5], but the occurrence rate of such information is low in military and political events. Therefore, this paper proposes a Chinese event detection technique based on dependency parsing and rule matching by analyzing the correlation in Chinese syntax and establishing a more unified event framework as the cornerstone.

III. Chinese Event Detection Technology Based on Dependency Parsing and Rule Matching

A. Dependency Parsing

Dependency parsing (DP) is a sub-task of the natural language processing task, which in general terms represents the relationship between words in a sentence through a syntactic dependency tree, assuming that the result is a directed graph $G = (V, A)$, where V represents each word in the sentence, called a node. A means a directed edge, indicating the existence of dependency relationships between phrases. In general, G satisfies the following conditions.

1) Has a unique root node that does not have an entry degree.
2) Has one and only one entry degree for other nodes.
3) For any leaf node, has one and only one path from the root node to it.

In this paper, we use the Dependent Syntactic Analysis Suite provided by Language Technology Platform [6] for the task of DP.

B. General Design of the Algorithm

In traditional event recognition tasks, action features are often used as event trigger words to mark the occurrence of an

Weitao Lin is with the University of Electronic Science and Technology of China, Chengdu, Sichuang 611731 China (phone: +86-15928806209; e-mail: clyderlin@gmail.com).

event and thus determine whether a text sentence is an event sentence. However, this approach is too broad, e.g., "A cruise ship is located off the coast of Los Angeles." However, such sentences, which indicate a state of affairs, do not change form and cannot be classified as event sentences. Let us think in another direction: Is it possible to find the type of sentences that are none event sentences, and then the event sentences will be solved? We studied a large amount of news corpus and concluded five categories of sentences with prominent non-event characteristics. We define them as "speech sentences," "declarative sentences," "conditional sentences," "able-wish sentences," and "negative sentences", which have more focused trigger word features or sentence features. Therefore, we propose identifying non-event-oriented trigger word features for initial screening, first removing the non-event sentences that are not focused. Naturally, the result is left with event sentences. We use the root node in the dependency syntax analysis tree as the core word that triggers all dependencies. The dictionary of non-event trigger words rules is shown in Table I (Note: due to translation differences between Chinese and English, words may have unknown meanings or be repeated).

TABLE I
DICTIONARY OF NON-EVENT TRIGGERING WORD RULES

Type	Rules
Non-event sentences	1 category: have, wish, want, figure, is, period, does, belong, may 2 categories: not, not yet, should, will, to, want, can, figure, is, like, according to, must, must, if, as, predicted, may (These words are divided into two categories, one is the verb itself contains the semantic meaning of emotion, tense, prediction, speech, statement, such words in the core word match; 2 is in the immediate vicinity of the core word before the core word makes the behavior of the core word becomes uncertain or becomes a description of the state of behavior, for such words should be expanded in front of the core word a word, and if it is independent)
Words of statements	point out, declare, set forth, affirm, reiterate, emphasize, claim, report, think, respond, declare, announce, proclaim, comment, imply, signify, indicate, express, mention, present, send, address, interpret, introduce, address, report, mention, describe, criticize, analyze, release, publish, announce, reveal, learn, according to ... news, will ... verb, observe, question, ask, news, show, take a stand, see, of is, words, claim, say, answer, ask, speak
Emotional words	feel, must, hope, vow, worry (2), count (2), worry, hope, wish, hope, hope, willing, deeply, hope, expect, feel, feel (2), with a view to, should, should, can, can, need, will, will be, expected, in advance, soon, plan (2), attempt, delusion, try, may, can, destined, believe, can be expected, so, therefore, gladly, gladly, willingly, must, dare, will, willingly
Speech words	is, so, for example, aim, record (2), include, know,, into, prove, illustrate (2), also, also have, lie, exist, based on, depending on, show (2), reflect (2), destined, exemplify, enumerate, have, have, have (2), witness, focus, concern, contain, include, as follows, important, confess, recount, become, divided (2), composed (2), called, summarize, summarize discuss, list, recount, name, remember, divide (2), set out, present (2), contain, sign (2), include, in fact, regard, see, insist (2), resolute (2), firm (2), keep, constitute, quote, narrate, for example, such as, recognize, conform, ignore, always, relative to, equivalent to, no different from, see, contrast, admonish, urge, decide (2), may be said, think, see, see, look, claim , yet, meaning, consisting of, in
Negative words	can't, no, except, won't, not, shall not, can't, can't, don't, unless, must not, difficult, should have, forbidden, not, can't, mustn't, can't, can't commit, difficult
Conditional Words	if, seem, only, as long as, but, even if, even though, whether, seem, would have, appear, maybe, perhaps, feel, might, seem, speculate, consider, prepare, once, seem

The general logic diagram of the algorithm is shown in Fig. 1.

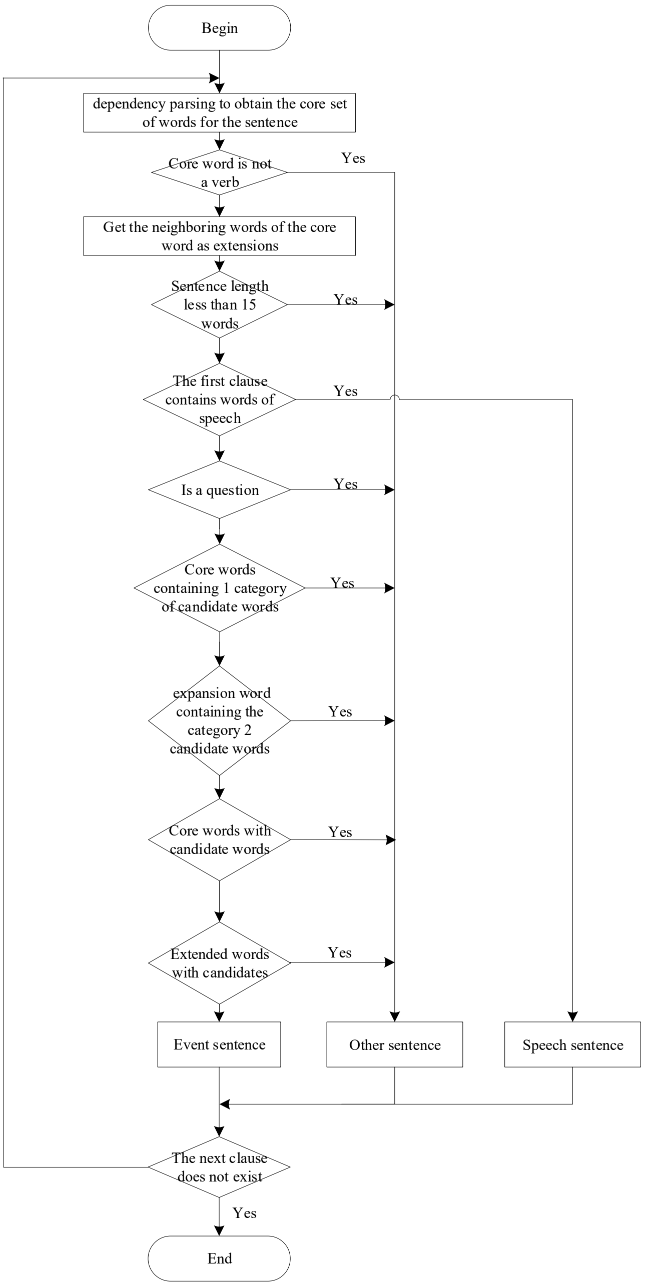
Begin
dependency parsing to obtain the core set of words for the sentence
Core word is not a verb
Yes
Get the neighboring words of the core word as extensions
Sentence length less than 15 words
Yes
The first clause contains words of speech
Yes
Is a question
Yes
Core words containing 1 category of candidate words
Yes
expansion word containing the category 2 candidate words
Yes
Core words with candidate words
Yes
Extended words with candidates
Yes
Event sentence
Other sentence
Speech sentence
The next clause does not exist
Yes
End

Fig. 1 Overall logic diagram of the algorithm

The specific steps of the algorithm are as follows:

1) First, we obtain the core words through DP. DP reveals the syntactic structure of a language unit by analyzing the dependencies between its components. Intuitively, DP identifies the "subject-verb-object" and "definite complement" grammatical features of a sentence and examines the relationships between them. Generally speaking, there is a "core" in the grammatical logic of a sentence, and this "core" determines the meaning of the sentence together with the other components of the

sentence. Therefore, we first use DP to obtain the "core words" of the input sentence.

2) The "core word" refers to the core of the sentence, in which all the relations of the sentence are developed. The event sentence must represent the occurrence of an action or a change of state, and the words that describe the circumstance and change are usually verbs. Therefore, we can filter sentences with non-verbs as the core word according to the linguistic nature of the "core word." The shorter sentences have a lower probability of describing events, so we filter them out.

3) In the news corpus, there are more speech sentences, such as "XX said" and "XX answered," which cause more interference to the target, and the speech sentences have prominent "cue words," such as "said, claimed," and these "cue words" are concentrated in the first of clauses. For example, "Taiwan is an inseparable part of China, a Foreign Ministry spokesperson said at a press conference today." But in some cases, the algorithm does not find the first clause of the sentence very well, even if it is what we consider to be a sentence, e.g., "On October 20, the Foreign Ministry spokesperson said at a press conference that Taiwan is an inseparable part of China." In this sentence, the first clause is "October 20", so we add an algorithm to find the "speech word" from the second clause if the length of the first clause is less than 10 words and define it as a "speech sentence."

4) Some words precede the core word or are the core word itself, causing the behavior to become uncertain or making it a description of a state of behavior, e.g., "During Obama's term, the United States will restart its Asia-Pacific rebalancing strategy." The word "will" in the example indicates that the event did not happen but may happen at some point in the future, so it is not an event sentence. Therefore, based on long-term iterative adjustment, we summarized four significant categories of candidate words (emotive, declarative, negative, conditional), as well as specific individual fields, for further screening, and finally left the sentence, naturally, as an event sentence.

IV. Experimental Results and Analysis

A. Dataset

The experiment uses the ACE2005 Chinese event dataset, which contains 633 articles, and the dataset defines eight event types and 33 event subtypes [7].

B. Baseline

The experiments are compared with some typical deep learning event detection approaches, and the baselines used in the experiments are described below.

(1) DMCNN [8]: Event detection approach based on trigger words for argumentative elements using multilayer dynamic words to retain critical information.

(2) HNN [9]: A convolutional neural network and bi-directional long and short-term memory network to capture sequential information in context.

(3) Bert+CRF+Bi-LSTM [10]: A classical trigger word-based event detection approach using a bidirectional long and short-term memory network and a random state field to capture contextual information and constrain trigger word boundaries thoroughly.

C. Evaluation Indicators

Although precision and recall are the most intuitive representation of model evaluation metrics, they are not reasonable in the state of data imbalance, so this paper uses precision (P), recall (R), and F1-score (F1) as experimental evaluation metrics.

D. Experimental Results Analysis

The specific experimental results are shown in Table II. The Chinese detection technique based on DP and rule matching proposed in this paper does not involve event classification-related experiments, so the three experimental metrics of the baseline model also only involve the trigger word detection task.

TABLE II
Comparison of Different Event Detection Methods

Model	P	R	F1
DMCNN	66.60	63.60	65.07
HNN	74.20	63.10	68.20
Bert+CRF+Bi-LSTM	74.50	64.30	69.00
Our Model	67.70	65.80	66.50

The proposed method does not suffer prediction bias due to biased training datasets. It is not affected by unregistered words and multiple meanings of words in Chinese contexts, so its experimental metrics achieve results that exceed those of some deep learning models.

V. Conclusion

This paper proposes a Chinese event detection technique based on DP and rule matching, rethinking the task from the opposite side of event detection by summarizing the non-event features in the corpus and proposing a method to filter non-event sentences using DP and rule matching, to achieve the goal of Chinese event detection. Experiments show that the proposed method achieves an effect comparable to that of deep learning models in terms of metrics. Although it is still far from the most widely used deep learning models, it has certain advantages regarding training time and prediction efficiency. It is highly portable and can be quickly modified in engineering practice as an initial baseline model to speed up project implementation and has broader application scenarios.

References

[1] Wu, Pingbo, Q. X. Chen, and Liang Ma. "Research on extraction and integration of developing event based on analysis of space-time information." *Journal of Chinese Information Processing* 20.1 (2006): 21-28.

[2] Yang, Erhong. "On the Information Extraction of the Sudden Events." *Doctor, Beijing Language and Culture University Beijing* (2005).

[3] Jiang, De-Liang. "Research on extraction of emergency event information based on rules matching." *Computer engineering and Design* 31.14 (2010): 3294-3297.

[4] Zhong Tao, Chen Qunxiu. "One Event Information Extraction System Based on the Cascade Finite State Automata." Proceedings of the 3rd National Conference on Information Retrieval and Content Security 2007:33-39

[5] Guoyin Lv."Research on Sudden Event Information Extraction of Tracking Reports Based on Rules." Computer Development & Applications 25.06(2012):7-9+13

[6] Harbin Institute of Technology. http://ltp.ai/

[7] Xiang, Wei, and Bang Wang. "A survey of event extraction from text." *IEEE Access* 7 (2019): 173111-173137.

[8] Chen, Yubo, et al. "Event extraction via dynamic multi-pooling convolutional neural networks." *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)*. 2015.

[9] Feng, Xiaocheng, Bing Qin, and Ting Liu. "A language-independent neural network for event detection." *Science China Information Sciences* 61.9 (2018): 1-12.

[10] Tian, Z., and X. Li. "Research on Chinese event detection method based on BERT-CRF model." *Computer Engineering and Applications* 57.ll (2021): 135-139.