

Negation and Speculation Identification in Chinese Language

Bowei Zou Qiaoming Zhu Guodong Zhou*

Natural Language Processing Lab, School of Computer Science and Technology
Soochow University, Suzhou, 215006, China
zoubowei@gmail.com, {qmzhu, gdzhou}@suda.edu.cn

Abstract

Identifying negative or speculative narrative fragments from fact is crucial for natural language processing (NLP) applications. Previous studies on negation and speculation identification in Chinese language suffers much from two problems: corpus scarcity and the bottleneck in fundamental Chinese information processing. To resolve these problems, this paper constructs a Chinese corpus which consists of three sub-corpora from different resources. In order to detect the negative and speculative cues, a sequence labeling model is proposed. Moreover, a bilingual cue expansion method is proposed to increase the coverage in cue detection. In addition, this paper presents a new syntactic structure-based framework to identify the linguistic scope of a cue, instead of the traditional chunking-based framework. Experimental results justify the usefulness of our Chinese corpus and the appropriateness of our syntactic structure-based framework which obtained significant improvement over the state-of-the-art on negation and speculation identification in Chinese language.

1 Introduction

Negation and speculation are ubiquitous phenomena in natural language. While negation is a grammatical category which comprises various kinds of devices to reverse the truth value of a proposition, speculation is a grammatical category which expresses the attitude of a speaker towards a statement in terms of degree of certainty,

reliability, subjectivity, sources of information, and perspective (Morante and Sporleder, 2012).

Current studies on negation and speculation identification mainly focus on two tasks: 1) cue detection, which aims to detect the signal of a negative or speculative expression, and 2) scope resolution, which aims to determine the linguistic coverage of a cue in sentence, in distinguishing unreliable or uncertain information from facts. For example, (E1) and (E2) include a negative cue and a speculative cue respectively, both denoted in **boldface** with their linguistic scopes denoted in square brackets (adopted hereinafter). In sentence (E1), the negative cue “不(not)” triggers the scope of “不会追究酒店的这次管理失职(would not investigate the dereliction of hotel)”, within which the fragment “investigate the dereliction of hotel” is the part that is repudiated; While the speculative cue “有望(expected)” in sentence (E2) triggers the scope “后期仍有望反弹(is still expected to rebound in the late)”, within which the fragment “the benchmark Shanghai Composite Index will rebound in the late” is the speculative part.

(E1) 所有住客均表示[不会追究酒店的这次管理失职].

(All of guests said that they [would **not** investigate the dereliction of hotel].)

(E2) 尽管上周五沪指盘中还受创业板的下跌所拖累,但[后期仍有望反弹].

(Although dragged down by GEM last Friday, the benchmark Shanghai Composite Index [is still **expected** to rebound in the late].)

Negation and speculation identification is very relevant for almost all NLP applications involving text understanding which need to discriminate between factual and non-factual information. The treatment of negation and speculation in computational linguistics has been shown to be

* Corresponding author

useful for biomedical text processing (Morante et al., 2008; Chowdhury and Lavelli, 2013), information retrieval (Averbuch, 2004), sentiment analysis (Councill et al., 2010; Zhu et al., 2014), recognizing textual entailment (Snow et al., 2006), machine translation (Baker et al., 2010; Wetzell and Bond, 2012), and so forth.

The research on negation and speculation identification in English has received a noticeable boost. However, in contrast to the significant achievements concerning English, the research progress in Chinese language is quite limited. The main reason includes the following two aspects: First, the scarcity of linguistic resource seriously limits the advance of related research. To the best of our knowledge, there are no publicly available standard Chinese corpus of reasonable size annotated with negation and speculation. Second, this may be attributed to the limitations of Chinese information processing.

The contributions of this paper are as follows:

- To address the aforementioned first issue, this paper seeks to fill this gap by presenting the Chinese negation and speculation corpus which consists of three kind of sub-corpora annotated for negative and speculative cues, and their linguistic scopes. The corpus has been made publicly available for research purposes and it is freely downloadable from <http://nlp.suda.edu.cn/corpus>.
- For cue detection, we propose a feature-based sequence labeling model to identify cues. It is worth noting that the morpheme feature is employed to better represent the compositional semantics inside Chinese words. Moreover, for improving the low recall rate which suffers from the unknown cues, we propose a cross-lingual cue expansion strategy based on parallel corpora.
- For scope resolution, we present a new syntactic structure-based framework on dependency tree. Evaluation justifies the appropriateness and validity of this framework on Chinese scope resolution, which outperforms the chunking-based framework that widely used in mainstream scope resolution systems.

The layout of the rest paper is organized as follows. Section 2 describes related work. Section 3 provides details about annotation guidelines and also presents statistics about corpus characteristics. Section 4 describes our approach in detail. Section 5 reports and discusses our experimental results. Finally, we conclude our work and indicate some future work in Section 6.

2 Related Work

Currently, both cue detection task and scope resolution task are always modeled as a classification problem with the purpose of predicting whether a token is inside or outside the cue and its scope. Among them, feature-based and kernel-based approaches are most popular.

In the feature-based framework, Agarwal and Yu (2010) employed a conditional random fields (CRFs) model to detect speculative cues and their scopes on the BioScope corpus. The CRFs-based model achieved an F1-measure of 88% in detecting speculative cues. We train this model on our corpus as the baseline system for cue detection. Our work is different from theirs in that we employ a new feature (morpheme feature) which is particularly appropriate for Chinese.

Besides, kernel-based approaches exploit the structure of the tree that connects cue and its corresponding scope. Zou et al. (2013) developed a tree kernel-based system to resolve the scope of negation and speculation, which captures the structured information in syntactic parsing trees. To the best of our knowledge, this system is the best English scope resolution system. For this reason, we train this system on our corpus as the baseline system for scope resolution.

Compared with a fair amount of works on English negation and speculation identification, unfortunately, few works has been published on Chinese. Ji et al. (2010) developed a system to detect speculation in Chinese news texts. However, only the speculative sentences have been found out, with no more fine-grained information such as scope. The insufficient study on Chinese negation and speculation identification drives us to construct a high-quality corpus and investigate how to find an approach that is particularly appropriate for Chinese language.

3 Corpus Construction

In this section, we elaborate on the overall characteristics of the Chinese Negation and Speculation (abbr., CNeSp) corpus we constructed, including a brief description of the sources that constitute our corpus, general guidelines which illustrated with lots of examples and some special cases, and statistics on the overall results of our corpus.

3.1 Sources

To capture the heterogeneity of language use in texts, the corpus consists of three different

sources and types, including scientific literature, product reviews, and financial articles.

Vincze et al. (2008) described that it is necessary to separate negative and speculative information from factual especially in science articles, because conclusions of science experiment are always described by using diversity of expressions and include hypothetical asserts or viewpoints. For this reason, we adopt the 19 articles from Chinese Journal of Computers (Vol.35(11)), an authoritative academic journal in Chinese, to construct the Scientific Literature sub-corpus.

Another part of the corpus consists of 311 articles from “股市及时雨(timely rain for stock market)” column from Sina.com in April, 2013. There are 22.3% and 40.2% sentences in the Financial Article sub-corpus containing negation and speculation respectively.

Many researches have investigated the role of negation in sentiment analysis task, as an important linguistic qualifier which leads to a change in polarity. For example, Councill et al. (2010) investigated the problem of determining the polarity of sentiment in movie reviews when negation words occur in the sentences. On the other hand, speculation is a linguistic expression that tends to correlate with subjectivity which is also crucial for sentiment analysis. Pang and Lee (2004) showed that subjectivity detection in the review domain helps to improve polarity classification. Therefore, the Product Review sub-corpus consists of 821 comments of hotel service from the website Ctrip.com.

3.2 Annotation Guidelines

The guidelines of our CNeSp corpus have partly referred to the existing Bioscope corpus guidelines (BioScope, 2008) in order to fit the needs of the Chinese language. In annotation process, negative or speculative cues and their linguistic scopes in sentence are annotated. There are several general principles below:

- (G1) Cue is contained in its scope.
- (G2) The minimal unit that expresses negation or speculation is annotated as a cue.
- (E3) 该股极有可能再度出现涨停。
(The stock is very likely to hit limit up.)

To G2, the modifiers such as prepositions, determiners, or adverbs are not annotated as parts of the cue. For example, in Sentence (E3), “极(very)” is only a modifier of the speculative cue “可能(likely)”, but not a constituent of the cue.

For the drawbacks of the Bioscope corpus guidelines either on itself or for Chinese lan-

guage, we introduced some modifications. These main changes are summarized below:

(G3) A cue is annotated only relying on its actual semantic in context.

(E4) 大盘不可能再次出现高开低走。

(It is **not possible** that the broader market opens high but slips later again.)

To G3, “不可能(not possible)” means that the author denies the possibility of the situation that “the broader market opens high but slips later again”, which contains negative meanings than speculative. Thus, the phrase “不可能(not possible)” should be labeled as a negative cue.

(G4) A scope should contain the subject which contributes to the meaning of the content being negated or speculated if possible.

(E5) *Once again, the Disorder module does [not contribute positively to the prediction].

The BioScope corpus suggests that the scope of negative adverbs usually starts with the cue and ends at the end of the phrase, clause or sentence (E5). However, in our view, the scope should contain the subject for the integrity of meaning. Following is an exceptional case.

(G5) Scope should be a continuous fragment in sentence.

(E6) 酒店有高档的配套设施,然而却不能多给我们提供一个枕头。

(The hotel are furnished with upscale facilities, but [cannot offer us one more pillow].)

Some rhetoric in Chinese language, such as parallelism or ellipsis, often gives rise to separation of some sentence constituents from others. For example, in Sentence (E6), the subject of the second clause should be “酒店(the hotel)”, which is omitted. In this situation, we only need to identify the negative or speculative part in sentence than all semantic constituents which can be completed through other NLP technologies, such as zero subject anaphora resolution or semantic role labeling.

(G6) A negative or speculative character or word may not be a cue.

(E7) 早茶的种类之多不得不赞。

(We are difficult not to give credit to the variety of morning snack.)

We have come across several cases where the presence of a negative or speculative character or word does not denote negative or speculative meaning. For example, there are lots of double negatives in Chinese language only for emphasizing than negative meanings. In Sentence (E7), obviously, the author wants to emphasis the praise of the variety of breakfast buffet by using

the phrase “不得不(be difficult not to)” which does not imply a negative meaning.

The CNeSp corpus is annotated by two independent annotators who are not allowed to communicate with each other. A linguist expert resolves the differences between the two annotators and modified the guidelines when they are confronted with problematic issues, yielding the gold standard labeling of the corpus.

3.3 Statistics and Agreement Analysis

Table 1 summarizes the chief characteristics of the three sub-corpora, including Scientific Literature (Sci., for short), Financial Article (Fin.), and Product Review (Prod.). As shown in Table 1, out of the total amount of 16,841 sentences more than 20% contained negation or speculation, confirming the availability for corpus.

Item	Sci.	Fin.	Prod.
#Documents	19	311	821
#Sentences	4,630	7,213	4,998
Avg. Length of Sentences	30.4	30.7	24.1
Negation			
%Sentence	13.2	17.5	52.9
Avg. Length of Scopes	9.1	7.2	5.1
Speculation			
%Sentence	21.6	30.5	22.6
Avg. Length of Scopes	12.3	15.0	6.9

(Avg. Length: The average number of Chinese characters.)

Table 1. Statistics of corpus.

Type		Sci.	Fin.	Prod.
Negation	Cue	0.96	0.96	0.93
	Cue & Scope	0.90	0.91	0.88
Speculation	Cue	0.94	0.90	0.93
	Cue & Scope	0.93	0.85	0.89

Table 2. Inter-annotator agreement.

We measured the inter-annotator agreement of annotating cues and their linguistic scope for all of three sub-corpora between the two independent annotators in terms of Kappa (Cohen, 1960). The results are shown in Table 2. The 2nd and 4th rows of the table show the kappa value of only cue annotation for negation and speculation, respectively. The 3rd and 5th rows show the agreement rate for both cue and its full scope. The most obvious conclusions here are that the identification of speculation is more complicated than negation even for humans because of the higher ambiguity of cues and the longer average length of scopes in speculation.

4 Chinese Negation and Speculation Identification

As a pipeline task, negation and speculation identification generally consists of two basic

stages, cue detection and scope resolution. The former detects whether a word or phrase implies negative or speculative meanings, while the latter determines the sequences of terms which are dominated by the corresponding cue in sentence.

In this section, we improve our cue detection system by using the morpheme features of Chinese characters and expanding the cue clusters based on bilingual parallel corpora. Then, we present a new syntactic structure-based framework for Chinese language, which regards the sub-structures of dependency tree selected by a heuristic rule as scope candidates.

4.1 Cue Detection

Most of the existing cue detection approaches are proposed from feature engineering perspective. They formulate cue detection as a classification issue, which is to classify each token in sentence as being the element of cue or not.

Feature-based sequence labeling model

At the beginning, we explore the performance of an English cue detection system, as described in Agarwal and Yu (2010), which employs a conditional random fields (abbr., CRFs) model with lexical and syntactic features. Unfortunately, the performance is very low on Chinese texts (Section 5.1). This may be attributed to the different characteristic of Chinese language, for example, no word boundaries and lack of morphologic variations. Such low performance drives us to investigate new effective features which are particularly appropriate for Chinese. We employed three kinds of features for cue detection:

1) N-gram features

For each character c_i , assuming its 5-windows characters are $c_{i-2} c_{i-1} c_i c_{i+1} c_{i+2}$, we adopt following features: $c_{i-2}, c_{i-1}, c_i, c_{i+1}, c_{i+2}, c_{i-1}c_i, c_i c_{i+1}, c_i c_{i-1} c_i, c_{i-1} c_i c_{i+1}, c_i c_{i+1} c_{i+2}$.

2) Lexical features

To achieve high performance as much as possible, we also use some useful basic features which are widely used in other NLP tasks on Chinese. The basic feature set consists of POS tag, the left/right character and its PoS tag. It is worth noting that the cue candidates in our model are characters. Thus, in order to get these features, we substitute them with corresponding features of the words which contain the characters.

3) Morpheme features

The word-formation of Chinese implies that almost all of the meanings of a word are made up by the morphemes, a minimal meaningful unit in Chinese language contained in words. This more

fine-grained semantics are the compositional semantics inside Chinese words namely. We assume that the morphemes in a given cue are also likely to be contained in other cues. For example, “猜测(guess)” is a given speculative cue which consists of “猜(guess)” and “测(speculate)”, while the morpheme “猜(guess)” could be appeared in “猜想(suppose)”. In consideration of the Chinese characteristics, we use every potential character in cues to get the morpheme feature.

A Boolean feature is taken to represent the morpheme information. Specifically, the characters which appear more than once within different cues in training corpus were selected as the features. The morpheme feature is set to 1, if the character is a negative or speculative morpheme.

For the ability of capturing the local information around a cue, we choose CRFs, a conditional sequence model which represents the probability of a hidden state sequence given some observations, as classifier to label each character with a tag indicating whether it is out of a cue (O), the beginning of the cue (B) or a part of the cue except the beginning one (I). In this way, our CRFs-based cue identifier performs sequential labeling by assigning each character one of the three tags and a character assigned with tag B is concatenated with following characters with tag I to form a cue.

Cross-lingual Cue Expansion Strategy

The feature-based cue detection approach mentioned above shows that a bottleneck lies in low recall (see Table 4). This is probably due to the absence of about 12% negation cues and 17% speculation cues from the training data. It is a challenging task to identify unknown cues with the limited amount of training data. Hence, we propose a cross-lingual cue expansion strategy.

In the approach, we take use of the top 5 Chinese cues in training corpus as our “anchor set”. For each cue, we search its automatically aligned English words from a Chinese-English parallel corpus to construct an English word cluster. The parallel corpus consisting of 100,000 sentence pairs is built by using Liu’s approach (Liu et al., 2014), which combines translation model with language model to select high-quality translation pairs from 16 million sentence pairs. The word alignment was obtained by running Giza++ (Och and Ney, 2003). In each cluster, we record the frequency of each unique English word. Considering the word alignment errors in cross-lingual

clusters, we filter the clusters by word alignment probability which is formulated as below:

$$\begin{aligned} P_A &= \alpha P(w_E | w_C) + (1 - \alpha) P(w_C | w_E) \\ &= \alpha \frac{P(w_E, w_C)}{P(w_C)} + (1 - \alpha) \frac{P(w_E, w_C)}{P(w_E)} \\ &= \alpha \frac{\sum_i align(w_{Ei}, w_C)}{\sum_i align(w_{Ei}, w_C)} + (1 - \alpha) \frac{\sum_i align(w_E, w_{Ci})}{\sum_i align(w_E, w_{Ci})} \quad (1) \end{aligned}$$

where $P(w_E | w_C)$ is the translation probability of English word w_E conditioned on Chinese word w_C , reversely, while $P(w_C | w_E)$ is the translation probability of Chinese word w_C conditioned on English word w_E . $align(w_m, w_n)$ is the number of alignments of word w_m and word w_n in parallel corpus. $\sum_i align(w_{mi}, w_n)$ is the sum of the number of alignments which contain word w_n . The parameter $\alpha \in [0, 1]$ is the coefficient controlling the relative contributions from the two directions of translation probability.

Then we conduct the same procedure in the other direction to construct Chinese word clusters anchored by English cues, until no new word comes about. For example, applying the above approach from the cue “可能(may)”, we obtain 59 Chinese speculative cues. All of words in the final expansion cluster are identified as cues.

4.2 Scope Resolution

Currently, mainstream approaches formulated the scope resolution as a chunking problem, which classifies every word of a sentence as being inside or outside the scope of a cue. However, unlike in English, we found that plenty of errors occurred in Chinese scope resolution by using words as the basic identifying candidate.

In this paper we propose a new framework using the sub-structures of dependency tree as scope candidates. Specifically, given a cue, we adopt the following heuristic rule to get the scope candidates in the dependency tree.

Setting constituent X and its siblings as the root nodes of candidate structure of scope, X should be the ancestor node of cue or cue itself.

For example, in the sentence “所有住客均表示不会追究酒店的这次管理失职(All of guests said that they would not investigate the dereliction of hotel)”, the negative cue “不(not)” has four constituent Xs and seven scope candidates, as shown in Figure 1. According to the above rule, three ancestor nodes {Xa: “表示(said)”, Xb: “追究(investigate)”, and Xc: “会(would)”,} correspond to three scope candidates (a, b1, and c),

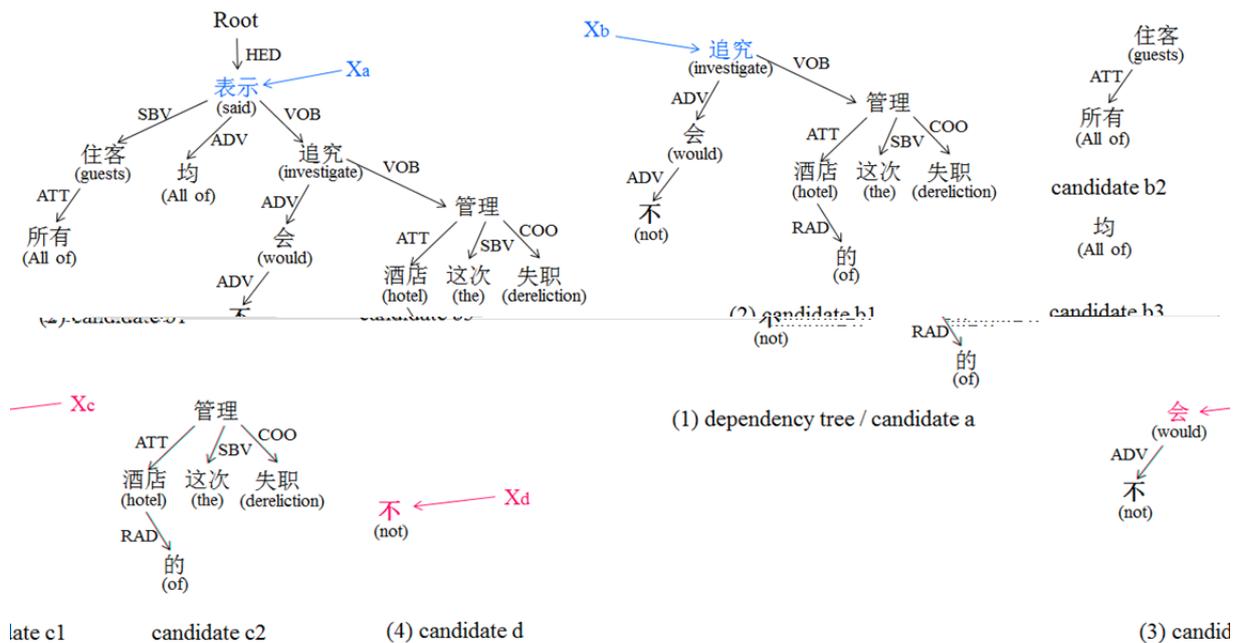


Figure 1. Examples of a negative cue and its seven scope candidates in dependency tree.

Feature	Description	Instantiation
Cue:		
C1: Itself	Tokens of cue	不(not)
C2: PoS	PoS of cue	d(adverb)
Scope candidate:		
S1: Itself	Tokens of headword	追究(investigate)
S2: PoS	PoS of headword	v(verb)
S3: Dependency type	Dependency type of headword	VOB
S4: Dependency type of child nodes	Dependency type of child nodes of headword	ADV+VOB
S5: Distance<candidate, left word>	Number of dependency arcs between the first word of candidate and its left word	3
S6: Distance<candidate, right word>	Number of dependency arcs between the last word of candidate and its right word	0
Relationship between cue and scope candidate:		
R1: Path	Dependency relation path from cue to headword	ADV-ADV
R2: Distance<cue, headword>	Number of dependency arcs between cue and headword	2
R3: Compression path	Compression version of path	ADV
R4: Position	Positional relationship of cue with scope candidate	L_N(Left-nested)

Table 3. Features and their instantiations for scope resolution.

and the cue itself is certainly a scope candidate (d). In addition, the Xb node has two siblings in dependency tree {“住客(guests)” and “均(all of)”}. Therefore, the two scope candidates corresponding to them are b2 and b3, respectively. Similarly, the sibling of the Xc node is labeled as candidate c2.

A binary classifier is applied to determine each candidate as either part of scope or not. In this paper, we employ some lexical and syntactic features about cue and candidate. Table 3 lists all of the features for scope resolution classification (with candidate b1 as the focus constituent (i.e., the scope candidate) and “不(not)” as the given cue, regarding candidate b1 in Figure 1(2)).

For clarity, we categorize the features into three groups according to their relevance with the given cue (C, in short), scope candidate (S, in short), and the relationship between cue and candidate (R, in short). Figure 2 shows four kinds of positional features between cue and scope candidate we defined (R4).

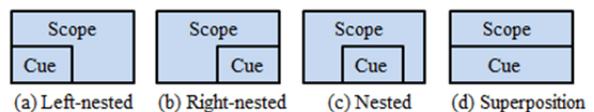


Figure 2. Positional features.

Some features proposed above may not be effective in classification. Therefore, we adopt a greedy feature selection algorithm as described in (Jiang and Ng, 2006) to pick up positive features incrementally according to their contribu-

tions on the development data. Additionally, a cue should have one continuous block as its scope, but the scope identifier may result in discontinuous scope due to independent candidate in classification. For this reason, we employ a post-processing algorithm as described in Zhu et al. (2010) to identify the boundaries.

5 Experimentation

In this section, we evaluate our feature-based sequence labeling model and cross-lingual cue expansion strategy on cue detection, and report the experimental results to justify the appropriateness of our syntactic structure-based framework on scope resolution in Chinese language.

The performance is measured by Precision (P), Recall (R), and F1-score (F). In addition, for scope resolution, we also report the accuracy in PCS (Percentage of Correct Scopes), within which a scope is fully correct if the output of scope resolution system and the correct scope have been matched exactly.

5.1 Cue Detection

Results of the Sequence Labeling Model

Every sub-corpus is randomly divided into ten equal folds so as to perform ten-fold cross validation. Lexical features are gained by using an open-source Chinese language processing platform, LTP¹(Che et al., 2010) to perform word segmentation, POS tagging, and syntactic parsing. CRF++² toolkit is employed as our sequence labeling model for cue detection.

Table 4 lists the performances of cue detection systems using a variety of features. It shows that the morpheme features derived from the word-formation of Chinese improve the performance for both negation and speculation cue detection systems on all kinds of sub-corpora. However, the one exception occurs in negation cue detection on the Product Review sub-corpus, in which the performance is decreased about 4.55% in precision. By error analysis, we find out the main reason is due to the pseudo cues. For example, “非常(very)” is identified by the negative morpheme “非(-un)”, which is a pseudo cue.

Table 4 also shows a bottleneck of our sequence labeling model, which lies in low recall. Due to the diversity of Chinese language, many cues only appear a few times in corpus. For ex-

ample, 83% (233/280) of speculative cues appear less than ten times in Financial Article sub-corpus. This data sparse problem directly leads to the low recall of cue detection.

Sci.	Negation			Speculation		
	P	R	F1	P	R	F1
Agarwal’s	48.75	36.44	41.71	46.16	33.49	38.82
N-gram	64.07	49.64	55.94	62.15	42.87	50.74
+Lexical	76.68	57.36	65.63	70.47	48.31	57.32
+Morpheme	81.37	59.11	68.48	76.91	50.77	61.16
Fin.						
Agarwal’s	41.93	39.15	40.49	50.39	42.80	46.29
N-gram	56.05	45.48	50.21	60.37	44.16	51.01
+Lexical	71.61	50.12	58.97	68.96	48.72	57.10
+Morpheme	78.94	53.37	63.68	75.43	51.29	61.06
Prod.						
Agarwal’s	58.47	47.31	52.30	45.88	34.13	39.14
N-gram	71.33	54.69	61.91	49.38	39.31	43.77
+Lexical	86.76	65.41	74.59	64.85	44.63	52.87
+Morpheme	82.21	66.82	73.72	70.06	45.31	55.03

Table 4. Contribution of features to cue detection.

Results of the Cross-lingual Cue Expansion Strategy

Before cue expansion, we select the parameter α as defined in formula (1) by optimizing the F1-measure score of on Financial Article sub-corpus. Figure 3 shows the effect on F1-measure of varying the coefficient from 0 to 1. We can see that the best performance can be obtained by selecting parameter 0.6 for negation and 0.7 for speculation. Then we apply these parameter values directly for cue expansion.

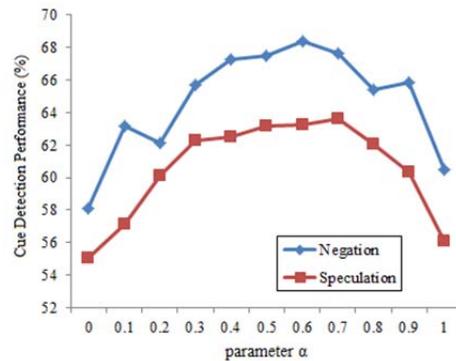


Figure 3. The effect of varying the value of parameter α on Financial Article sub-corpus.

Table 5 lists the performances of feature-based system, expansion-based system, and the combined system. A word is identified as a cue by combined system if it is identified by one of the above systems (Feat-based or Exp-based) at least.

For both negation and speculation, the cross-lingual cue expansion approach provides significant improvement over the feature-based sequence labeling model, achieving about 15-20%

¹ <http://www.ltp-cloud.com>

² <https://crfpp.googlecode.com/svn/trunk/doc/index.html>

better recall with little loss in precision. More importantly, the combined system obtains the best performance.

Sci.	Negation			Speculation		
	P	R	F1	P	R	F1
Feat-based	81.37	59.11	68.48	76.91	50.77	61.16
Exp-based	68.29	76.24	72.05	62.74	68.07	65.30
Combined	75.17	78.91	76.99	70.98	75.71	73.27
Fin.						
Feat-based	78.94	53.37	63.68	75.43	51.29	61.06
Exp-based	70.31	64.49	67.27	67.46	68.78	68.11
Combined	72.77	67.02	69.78	71.60	69.03	70.29
Prod.						
Feat-based	82.21	66.82	73.72	70.06	45.31	55.03
Exp-based	78.30	86.47	82.18	62.18	63.47	62.82
Combined	81.94	89.23	85.43	67.56	69.61	68.57

Table 5. Performance of cue detection.

5.2 Syntactic Structure-based Scope Resolution

Considering the effectiveness of different features, we divide the Financial Article sub-corpus into 5 equal parts, within which 2 parts are used for feature selection. Then, the feature selection data are divided into 5 equal parts, within which 4 parts for training and the rest for developing. On this data set, a greedy feature selection algorithm (Jiang and Ng, 2006) is adopted to pick up positive features proposed in Table 3. In addition, SVM^{Light}³ with the default parameter is selected as our classifier.

Table 6 lists the performance of selected features. 7 features {C1, C2, S4, S5, S6, R1, R4} are selected consecutively for negation scope resolution, while 9 features {C2, S1, S3, S4, S5, R1, R2, R3, R4} are selected for speculation scope resolution. We will include those selected features in all the remaining experiments.

Type	Feature set	Sci.	Fin.	Prod.
Negation	Selected features	62.16	56.07	60.93
	All features	59.74	54.20	55.42
Speculation	Selected features	54.16	49.64	52.89
	All features	52.33	46.27	48.07

Table 6. Feature selection for scope resolution on golden cues (PCS %).

The feature selection experiments suggest that the feature C2 (POS of cue) plays a critical role for both negation and speculation scope resolution. It may be due to the fact that cues of different POS usually undertake different syntactic roles. Thus, there are different characteristics in triggering linguistic scopes. For example, an adjective cue may treat a modificatory structure as

its scope, while a conjunction cue may take the two connected components as its scope.

As a pipeline task, the negation and speculation identification could be regarded as a combination of two sequential tasks: first, cue detection, and then scope resolution. Hence, we turn to a more realistic scenario in which cues are automatically recognized.

Type	Corpus	P	R	F1	PCS
Negation	Sci.	55.32	53.06	54.17	59.08
	Fin.	42.14	46.37	44.15	49.24
	Prod.	50.57	48.55	49.54	52.17
Speculation	Sci.	45.68	47.15	46.40	48.36
	Fin.	34.21	31.80	32.96	41.33
	Prod.	32.64	33.59	33.11	39.78

Table 7. Performance of scope resolution with automatic cue detection.

Table 7 lists the performance of scope resolution by using automatic cues. It shows that automatic cue detection lowers the performance by 3.08, 6.83, and 8.76 in PCS for the three sub-corpora, respectively; while it lowers the performance by 5.80, 8.31 and 13.11 in PCS for speculation scope resolution on the three sub-corpora, respectively (refer to Table 6). The main reason of performance lost is the error propagation from the automatic cue detection.

We employ a start-of-the-art chunking-based scope resolution system (described in Zou et al., (2013)) as a baseline, in which every word in sentence has been labelled as being the element of the scope or not. Table 8 compares our syntactic structure-based framework with the chunking-based framework on scope resolution. Note that all the performances are achieved on Financial Article sub-corpus by using golden cues. The results in Table 8 shows that our scope resolution system outperforms the chunking ones both on negation and speculation, improving 8.75 and 7.44 in PCS, respectively.

Type	System	PCS
Negation	Chunking-based	47.32
	Ours	56.07
Speculation	Chunking-based	42.20
	Ours	49.64

Table 8. Comparison with the chunking-based system on Financial Article sub-corpus.

6 Conclusion

In this paper we construct a Chinese corpus for negation and speculation identification, which annotates cues and their linguistic scopes. For cue detection, we present a feature-based sequence labeling model, in which the morpheme

³ <http://svmlight.joachims.org>

feature is employed to better catch the composition semantics inside the Chinese words. Complementally, a cross-lingual cue expansion strategy is proposed to increase the coverage in cue detection. For scope resolution, we present a new syntactic structure-based framework to identify the linguistic scope of a cue. Evaluation justifies the usefulness of our Chinese corpus and the appropriateness of the syntactic structure-based framework. It also shows that our approach outperforms the state-of-the-art chunking ones on negation and speculation identification in Chinese language.

In the future we will explore more effective features to improve the negation and speculation identification in Chinese language, and focus on joint learning of the two subtasks.

Acknowledgments

This research is supported by the National Natural Science Foundation of China, No.61272260, No.61331011, No.61273320, No.61373097, and the Major Project of College Natural Science Foundation of Jiangsu Province, No.11KJA520003. The authors would like to thank the anonymous reviewers for their insightful comments and suggestions.

Reference

- Shashank Agarwal and Hong Yu. 2010. Detecting hedge cues and their scope in biomedical text with conditional random fields. *Journal of Biomedical Informatics*, 43, 953-961.
- Mordechai Averbuch, Tom H. Karson, Benjamin Ben-Ami, Oded Maimon, and Lior Rokach. 2004. Context-sensitive medical information retrieval. In *Proceedings of the 11th World Congress on Medical Informatics (MEDINFO'04)*, 1-8.
- Kathrin Baker, Michael Bloodgood, Bonnie Dorr, Nathaniel W. Filardo, Lori Levin, and Christine Pitko. 2010. A modality lexicon and its use in automatic tagging. In *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10)*, 1402-1407.
- BioScope. 2008. Annotation guidelines. [http://www.inf.u-szeged.hu/rgai/project/nlp/bioscope/Annotation guidelines2.1.pdf](http://www.inf.u-szeged.hu/rgai/project/nlp/bioscope/Annotation%20guidelines2.1.pdf)
- Wanxiang Che, Zhenghua Li, Ting Liu. 2010. LTP: A Chinese language technology platform. In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING'10): Demonstrations*, 13-16.
- Md. Faisal Mahbub Chowdhury and Alberto Lavelli. 2013. Exploiting the scope of negations and heterogeneous features for relation extraction: A case study for drug-drug interaction extraction. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT'13)*, 765-771.
- Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.
- Isaac Council, Ryan McDonald, and Leonid Velikovich. 2010. What's great and what's not: Learning to classify the scope of negation for improved sentiment analysis. In *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, 51-59.
- Zhengping Jiang and Hwee T. Ng. 2006. Semantic role labeling of NomBank: A maximum entropy approach. In *Proceedings of the Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (EMNLP'06)*, 138-145.
- Le Liu, Yu Hong, Hao Liu, Xing Wang, and Jianmin Yao. 2014. Effective selection of translation model training data. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL'14)*, Short Papers, 569-573.
- Feng Ji, Xipeng Qiu, Xuanjing Huang. 2010. Exploring uncertainty sentences in Chinese. In *Proceedings of the 16th China Conference on Information Retrieval*, 594-601.
- Roser Morante, Anthony Liekens, and Walter Daelemans. 2008. Learning the scope of negation in biomedical texts. In *Proceedings of the Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (EMNLP'08)*, 715-724.
- Roser Morante and Caroline Sporleder. 2012. Modality and negation: an introduction to the special issue. *Comput. Linguist.* 38, 2, 223-260.
- Franz Josef Och and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Comput. Linguist.* 29, 1, 19-51.
- Bo Pang and Lillian Lee. 2004. A sentimental education: sentiment analysis using subjectivity. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL'04)*, 271-278.
- Rion Snow, Lucy Vanderwende, and Arul Menezes. 2006. Effectively using syntax for recognizing false entailment. In *Proceedings of the Main Conference on Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT'06)*, 103-111.

ation of Computational Linguistics (HLT-NAACL'06), 33-40.

- Veronika Vincze, György Szarvas, Richárd Farkas, György Móra and János Csirik. 2008. The BioScope corpus: biomedical texts annotated for uncertainty, negation and their scopes. *BMC Bioinformatics*, 9(Suppl 11):S9.
- Dominikus Wetzel, and Francis Bond. 2012. Enriching parallel corpora for statistical machine translation with semantic negation rephrasing. In *Proceedings of the 6th Workshop on Syntax, Semantics and Structure in Statistical Translation*, 20-29.
- Qiaoming Zhu, Junhui Li, Hongling Wang, and Guodong Zhou. 2010. A Unified Framework for Scope Learning via Simplified Shallow Semantic Parsing. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing (EMNLP'10)*, 714-724.
- Xiaodan Zhu, Hongyu Guo, Saif Mohammad, and Svetlana Kiritchenko. 2014. An empirical study on the effect of negation words on sentiment. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL'14)*, 304-313.
- Bowei Zou, Guodong Zhou, and Qiaoming Zhu. 2013. Tree kernel-based negation and speculation scope detection with structured syntactic parse features. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP'13)*, 968-976.