A Unified Framework for Scope Learning via Simplified Shallow Semantic Parsing

Qiaoming Zhu   Junhui Li   Hongling Wang   Guodong Zhou[*]
School of Computer Science and Technology
Soochow University, Suzhou, China 215006
{qmzhu, lijunhui, hlwang, gdzhou}@suda.edu.cn

Abstract

This paper approaches the scope learning problem via simplified shallow semantic parsing. This is done by regarding the cue as the predicate and mapping its scope into several constituents as the arguments of the cue. Evaluation on the BioScope corpus shows that the structural information plays a critical role in capturing the relationship between a cue and its dominated arguments. It also shows that our parsing approach significantly outperforms the state-of-the-art chunking ones. Although our parsing approach is only evaluated on negation and speculation scope learning here, it is portable to other kinds of scope learning.

1 Introduction

Recent years have witnessed an increasing interest in the analysis of linguistic scope in natural language. The task of scope learning deals with the syntactic analysis of what part of a given sentence is under user's special interest. For example, of negation assertion concerned, a negation cue (e.g., *not*, *no*) usually dominates a fragment of the given sentence, rather than the whole sentence, especially when the sentence is long. Generally, scope learning involves two subtasks: cue recognition and its scope identification. The former decides whether a word or phrase in a sentence is a cue of a special interest, where the semantic information of the word or phrase, rather than the syntactic information, plays a critical role. The latter determines the sequences of words in the sentence which are dominated by the given cue.

Recognizing the scope of a special interest (e.g., negative assertion and speculative assertion) is essential in information extraction (IE), whose aim is to derive factual knowledge from free text. For example, Vincze et al. (2008) pointed out that the extracted information within the scope of a negation or speculation cue should either be discarded or presented separately from factual information. This is especially important in the biomedical and scientific domains, where various linguistic forms are used extensively to express impressions, hypothesized explanations of experimental results or negative findings. Besides, Vincez et al. (2008) reported that 13.45% and 17.70% of the sentences in the abstracts subcorpus of the BioScope corpus contain negative and speculative assertions, respectively, while 12.70% and 19.44% of the sentences in the full papers subcorpus contain negative and speculative assertions, respectively. In addition to the IE tasks in the biomedical domain, negation scope learning has attracted increasing attention in some natural language processing (NLP) tasks, such as sentiment classification (Turney, 2002). For example, in the sentence "*The chair is not comfortable but cheap*", although both the polarities of the words "*comfortable*" and "*cheap*" are positive, the polarity of "the chair" regarding the attribute "*cheap*" keeps positive while the polarity of "the chair" regarding the attribute "*comfortable*" is reversed due to the negation cue "*not*". Similarly, seeing the increasing interest in speculation scope learning, the CoNLL'2010 shared task (Farkas et al., 2010) aims to detect uncertain information in resolving the scopes of speculation cues.

Most of the initial research in this literature focused on either recognizing negated terms or identifying speculative sentences, using some heuristic

---

[*] Corresponding author

rules (Chapman et al., 2001; Light et al., 2004), and machine learning methods (Goldin and Chapman, 2003; Medlock and Briscoe, 2007). However, scope learning has been largely ignored until the recent release of the BioScope corpus (Szarvas et al., 2008), where negation/speculation cues and their scopes are annotated explicitly. Morante et al. (2008) and Morante and Daelemans (2009a & 2009b) pioneered the research on scope learning by formulating it as a chunking problem, which classifies the words of a sentence as being inside or outside the scope of a cue. Alternatively, Özgür and Radev (2009) and Øvrelid et al. (2010) defined heuristic rules for speculation scope learning from constituency and dependency parse tree perspectives, respectively.

Although the chunking approach has been evaluated on negation and speculation scope learning and can be easily ported to other scope learning tasks, it ignores syntactic information and suffers from low performance. Alternatively, even if the rule-based methods may be effective for a special scope learning task (e.g., speculation scope learning), it is not readily adoptable to other scope learning tasks (e.g., negation scope learning). Instead, this paper explores scope learning from parse tree perspective and formulates it as a simplified shallow semantic parsing problem, which has been extensively studied in the past few years (Carreras and Màrquez, 2005). In particular, the cue is recast as the predicate and the scope is recast as the arguments of the cue. The motivation behind is that the structured syntactic information plays a critical role in scope learning and should be paid much more attention, as indicated by previous studies in shallow semantic parsing (Gildea and Palmer, 2002; Punyakanok et al., 2005). Although our approach is evaluated only on negation and speculation scope learning here, it is portable to other kinds of scope learning.

The rest of this paper is organized as follows. Section 2 reviews related work. Section 3 introduces the Bioscope corpus on which our approach is evaluated. Section 4 describes our parsing approach by formulating scope learning as a simplified shallow semantic parsing problem. Section 5 presents the experimental results. Finally, Section 6 concludes the work.

2 Related Work

Most of the previous research on scope learning falls into negation scope learning and speculation scope learning.

Negation Scope Learning

Morante et al. (2008) pioneered the research on negation scope learning, largely due to the availability of a large-scale annotated corpus, the Bioscope corpus. They approached negation cue recognition as a classification problem and formulated negation scope identification as a chunking problem which predicts whether a word in the sentence is inside or outside of the negation scope, with proper post-processing to ensure consecutiveness of the negation scope. Morante and Daelemans (2009a) further improved the performance by combing several classifiers and achieved the accuracy of ~98% for negation cue recognition and the PCS (Percentage of Correct Scope) of ~74% for negation scope identification on the abstracts subcorpus. However, this chunking approach suffers from low performance, in particular on long sentences. For example, given golden negation cues on the Bioscope corpus, Morante and Daelemans (2009a) only got the performance of 50.26% in PCS on the full papers subcorpus (22.8 words per sentence on average), compared to 87.27% in PCS on the clinical reports subcorpus (6.6 words per sentence on average).

Speculation Scope Learning

Similar to Morante and Daelemans (2009a), Morante and Daelemans (2009b) formulated speculation scope identification as a chunking problem which predicts whether a word in the sentence is inside or outside of the speculation scope, with proper post-processing to ensure consecutiveness of the speculation scope. They concluded that their method for negation scope identification is portable to speculation scope identification. However, of speculation scope identification concerned, it also suffers from low performance, with only 60.59% in PCS for the clinical reports subcorpus of short sentences.

Alternatively, Özgür and Radev (2009) employed some heuristic rules from constituency parse tree perspective on speculation scope identification. Given golden speculation cues, their rule-based method achieves the accuracies of 79.89%

and 61.13% on the abstracts and the full papers subcorpora, respectively. The more recent CoNLL'2010 shared task was dedicated to the detection of speculation cues and their linguistic scope in natural language processing (Farkas et al., 2010). As a representative, Øvrelid et al. (2010) adopted some heuristic rules from dependency parse tree perspective to identify their speculation scopes.

3 Cues and Scopes in the BioScope Corpus

This paper employs the BioScope corpus (Szarvas et al., 2008; Vincze et al., 2008)[1], a freely downloadable resource from the biomedical domain, as the benchmark corpus. In this corpus, every sentence is annotated with negation cues and speculation cues (if it has), as well as their linguistic scopes. Figure 1 shows a self-explainable example. It is possible that a negation/speculation cue consists of multiple words, i.e., "can not"/"*indicate that*" in Figure 1.

<sentence id="S26.8">These findings <xcope id="X26.8.2"><cue type="**speculation**" ref="X26.8.2">indicate that</cue> <xcope id="X26.8.1">corticosteroid resistance in bronchial asthma <cue type="**negation**" ref="X26.8.1">can not</cue> be explained by abnormalities in corticosteroid receptor characteristics</xcope></xcope>.</sentence>

Figure 1: An annotated sentence in the BioScope corpus

The Bioscope corpus consists of three subcorpora: biological full papers from FlyBase and from BMC Bioinformatics, biological paper abstracts from the GENIA corpus (Collier et al., 1999), and clinical (radiology) reports. Among them, the full papers subcorpus and the abstracts subcorpus come from the same genre, and thus share some common characteristics in statistics, such as the number of words in the negation/speculation scope to the right (or left) of the negation/speculation cue and the average scope length. In comparison, the clinical reports subcorpus consists of clinical radiology reports with short sentences. For detailed statistics and annotation

guidelines about the three subcorpora, please see Morante and Daelemans (2009a & 2009b).

For preprocessing, all the sentences in the Bioscope corpus are tokenized and then parsed using the Berkeley parser (Petrov and Klein, 2007)[2] trained on the GENIA TreeBank (GTB) 1.0 (Tateisi et al., 2005)[3], which is a bracketed corpus in (almost) PTB style. 10-fold cross-validation on GTB1.0 shows that the parser achieves the performance of 86.57 in F1-measure. It is worth noting that the GTB1.0 corpus includes all the sentences in the abstracts subcorpus of the Bioscope corpus.

4 Scope Learning via Simplified Shallow Semantic Parsing

In this section, we first formulate the scope learning task as a simplified shallow semantic parsing problem. Then, we deal with it using a simplified shallow semantic parsing framework.

4.1 Formulating Scope Learning as a Simplified Shallow Semantic Parsing Problem

Given a parse tree and a predicate in it, shallow semantic parsing recognizes and maps all the constituents in the sentence into their corresponding semantic arguments (roles) of the predicate or not. As far as scope learning considered, the cue can be regarded as the predicate[4], while its scope can be mapped into several constituents dominated by the cue and thus can be regarded as the arguments of the cue. In particular, given a cue and its scope which covers $word_m, ..., word_n$, we adopt the following two heuristic rules to map the scope of the cue into several constituents which can be deemed as its arguments in the given parse tree.

1) The cue itself and all of its ancestral constituents are non-arguments.
2) If constituent X is an argument of the given cue, then X should be the highest constituent dominated by the scope of $word_m, ..., word_n$. That is to say, X's parent constituent must cross-bracket or include the scope of $word_m, ..., word_n$.
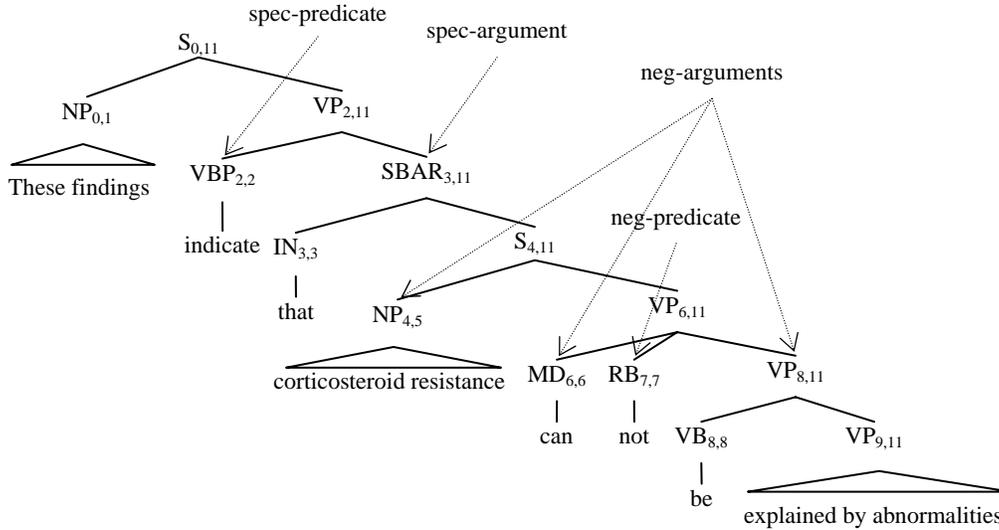

Figure 2: Examples of a negation/speculation cue and its arguments in a parse tree

The first rule ensures that no argument covers the cue while the second rule ensures no overlap between any two arguments. These two constraints between a cue and its arguments are consistent with shallow semantic parsing (Carreras and Màrquez, 2005). For example, in the sentence "*These findings indicate that corticosteroid resistance can not be explained by abnormalities*", the negation cue "*can not*" has the negation scope "*corticosteroid resistance can not be explained by abnormalities*" while the speculation cue "*indicate that*" has the speculation scope "*indicate that corticosteroid resistance can not be explained by abnormalities*". As shown in Figure 2, the node "$RB_{7,7}$" (i.e., *not*) represents the negation cue "*can not*" while its arguments include three constituents {$NP_{4,5}$, $MD_{6,6}$, and $VP_{8,11}$}. Similarly, the node "$VBP_{2,2}$" (i.e., *indicate*) represents the speculation cue "*indicate that*" while its arguments include one constituent $SBAR_{3,11}$. It is worth noting that according to the above rules, scope learning via shallow semantic parsing, i.e. determining the arguments of a given cue, is robust to some variations in the parse trees. This is also empirically justified by our later experiments. For example, if the $VP_{6,11}$ in Figure 2 is incorrectly expanded by the rule $VP_{6,11} \rightarrow MD_{6,6}+RB_{7,7}+VB_{8,8}+VP_{9,11}$, the negation scope of the negation cue "*can not*" can still be correctly detected as long as {$NP_{4,5}$, $MD_{6,6}$, $VB_{8,8}$, and $VP_{9,11}$} are predicated as the arguments of the negation cue "*can not*".

Compared with common shallow semantic parsing which needs to assign an argument with a semantic label, scope identification does not involve semantic label classification and thus could be divided into three consequent phases: argument pruning, argument identification and post-processing.

4.2 Argument Pruning

Similar to the predicate-argument structures in common shallow semantic parsing, the cue-scope structures in scope learning can be also classified into several certain types and argument pruning can be done by employing several heuristic rules accordingly to filter out constituents, which are most likely non-arguments of a given cue. Similar to the heuristic algorithm proposed in Xue and Palmer (2004) for argument pruning in common shallow semantic parsing, the argument pruning algorithm adopted here starts from designating the cue node as the current node and collects its siblings. It then iteratively moves one level up to the parent of the current node and collects its siblings. The algorithm ends when it reaches the root of the parse tree. To sum up, except the cue node itself and its ancestral constituents, any constituent in the parse tree whose parent covers the given cue will be collected as argument candidates. Taking the negation cue node "$RB_{7,7}$" in Figure 2 as an example, constituents {$MD_{6,6}$, $VP_{8,11}$, $NP_{4,5}$, $IN_{3,3}$,

Feature	Remarks
B1	Cue itself: the word of the cue, e.g., not, rather_than. (*can_not*)
B2	Phrase Type: the syntactic category of the argument candidate. (*NP*)
B3	Path: the syntactic path from the argument candidate to the cue. (*NP<S>VP>RB*)
B4	Position: the positional relationship of the argument candidate with the cue. "left" or "right". (*left*)

Table 1: Basic features and their instantiations for argument identification in scope learning, with NP4,5 as the focus constituent (i.e., the argument candidate) and "can not" as the given cue, regarding Figure 2.

Feature	Remarks
Argument Candidate (AC) related	
AC1	The headword (AC1H) and its POS (AC1P). (*resistance*, *NN*)
AC2	The left word (AC2W) and its POS (AC2P). (*that*, *IN*)
AC3	The right word (AC3W) and its POS (AC3P). (*can*, *MD*)
AC4	The phrase type of its left sibling (AC4L) and its right sibling (AC4R). (*NULL*, *VP*)
AC5	The phrase type of its parent node. (*S*)
AC6	The subcategory. (*S:NP+VP*)
Cue/Predicate (CP) related	
CP1	Its POS. (*RB*)
CP2	Its left word (CP2L) and right word (CP2R). (*can*, *be*)
CP3	The subcategory. (*VP:MD+RB+VP*)
CP4	The phrase type of its parent node. (*VP*)
Combined Features related with the Argument Candidate (CFAC1-CFAC2)	
b2&AC1H, b2&AC1P	
Combined Features related with the given Cue/Predicate (CFCP1-CFCP2)	
B1&CP2L, B1&CP2R	
Combined Features related with both the Argument Candidate and the given Cue/Predicate (CFACCP1-CFACCP7)	
B1&B2, B1&B3, B1&CP1, B3&CFCP1, B3&CFCP2, B4&CFCP1, B4&CFCP2	

Table 2: Additional features and their instantiations for argument identification in scope identification, with NP4,5 as the focus constituent (i.e., the argument candidate) and "can not" as the given cue, regarding Figure 2.

$VBP_{2,2}$, and $NP_{0,1}$} are collected as its argument candidates consequently.

4.3 Argument Identification

Here, a binary classifier is applied to determine the argument candidates as either valid arguments or non-arguments. Similar to argument identification in common shallow semantic parsing, the structured syntactic information plays a critical role in scope learning.

Basic Features

Table 1 lists the basic features for argument identification. These features are also widely used in common shallow semantic parsing for both verbal and nominal predicates (Xue, 2008; Li et al., 2009).

Additional Features

To capture more useful information in the cue-scope structures, we also explore various kinds of additional features. Table 2 shows the features in better capturing the details regarding the argument candidate and the cue. In particular, we categorize the additional features into three groups according to their relationship with the argument candidate (AC, in short) and the given cue/predicate (CP, in short).

Some features proposed above may not be effective in argument identification. Therefore, we adopt the greedy feature selection algorithm as described in Jiang and Ng (2006) to pick up positive features incrementally according to their contributions on the development data. The algorithm repeatedly selects one feature each time, which contributes most, and stops when adding any of the remaining features fails to improve the performance.

4.4 Post-Processing

Although a cue in the BioScope corpus always has only one continuous block as its scope (including the cue itself), the scope identifier may result in discontinuous scope due to independent predication in the argument identification phase. Given the golden negation/speculation cues, we observe that 6.2%/9.1% of the negation/speculation scopes predicted by our scope identifier are discontinuous.
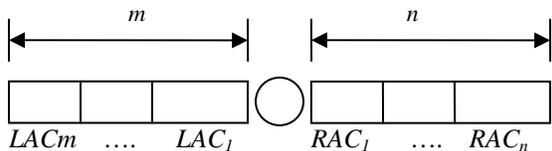
Figure 3: Projecting the left and the right argument candidates into the word level.

Figure 3 demonstrates the projection of all the argument candidates into the word level. According to our argument pruning algorithm in Section 4.2, except the words presented by the cue, the projection covers the whole sentence and each constituent (LAC_i or RAC_j in Figure 3) receives a probability distribution of being an argument of the given cue in the argument identification phase.

Since a cue is deemed inside its scope in the BioScope corpus, our post-processing algorithm first includes the cue in its scope and then starts to identify the left and the right scope boundaries, respectively.

As shown in Figure 3, the left boundary has $m+1$ possibilities, namely the cue itself, the leftmost word of constituent LAC_i ($1 <= i <= m$). Supposing LAC_i receives probability of P_i being an argument, we use the following formula to determine LAC_{k*} whose leftmost word represents the boundary of the left scope. If $k^*=0$, then the cue itself represents its left boundary.

$$k^* = \arg\max_k \prod_{i=1}^{k} P_i * \prod_{i=k+1}^{m} (1 - P_i)$$

Similarly, the right boundary of the given cue can be decided.

4.5 Cue Recognition

Automatic recognition of cues of a special interest is the prerequisite for a scope learning system. The approaches to recognizing cues of a special interest usually fall into two categories: 1) substring matching approaches, which require a set of cue words or phrases in advance (e.g., Light et al., 2004); 2) machine learning approaches, which train a classifier with either supervised or semi-supervised learning methods (e.g., Özgür and Radev, 2009; Szarvas, 2008). Without loss of generality, we adopt a machine learning approach and train a classifier with supervised learning. In particular, we make an independent classification for each word with a BIO label to indicate whether it

is the first word of a cue, inside a cue, or outside of it, respectively.

Inspired by previous studies on similar tasks such as WSD and nominal predicate recognition in shallow semantic parsing (Lee and Ng, 2002; Li et al., 2009), where various features on the word itself, surrounding words and syntactic information have been successfully used, we believe that such information is also valuable to automatic recognition of cues. Table 3 shows the features employed for cue recognition. In particular, we categorize these features into three groups: 1) features about the cue candidate itself (CC in short); 2) features about surrounding words (SW in short); and 3) structural features derived from the syntactic parse tree (SF in short).

Feature	Remarks
Cue Candidate (CC) related	
CC1	The cue candidate itself. (*indicate*)
CC2	The stem of the cue candidate. (*indicate*)
CC3	The POS tag of the cue candidate. (*VBP*)
Surrounding Words (SW) related	
SW1	The left surrounding words with the window size of 3. (*these*, *findings*)
SW2	The right surrounding words with the window size of 3. (*that*, *corticosteroid*, *resistance*)
Structural Features (SF)	
SF1	The subcategory of the candidate node. (*VP-->VBP+SBAR*)
SF2	The subcategory of the candidate node's parent. (*S-->NP+VP*)
SF3	POS tag of the candidate node + Phrase type of its parent node + Phrase type of its grandpa node. (*VBP + VP + S*)

Table 3: Features and their instantiations for cue recognition, with $VBP_{2,2}$ as the cue candidate, regarding Figure 2.

5 Experimentation

We have evaluated our simplified shallow semantic parsing approach to negation and speculation scope learning on the BioScope corpus.

5.1 Experimental Settings

Following the experimental setting in Morante et al. (2008) and Morante and Daelemans (2009a & 2009b), the abstracts subcorpus is randomly divided into 10 folds so as to perform 10-fold cross-validation, while the performance on both the pa-

pers and clinical reports subcorpora is evaluated using the system trained on the whole abstracts subcorpus. In addition, SVMLight[5] is selected as our classifier.

For cue recognition, we report its performance using precision/recall/F1-measure. For scope identification, we report the accuracy in PCS (Percentage of Correct Scopes) when the golden cues are given, and report precision/recall/F1-measure when the cues are automatically recognized.

5.2 Experimental Results on Golden Parse Trees and Golden Cues

In order to select beneficial features from the additional features proposed in Section 4.3, we randomly split the abstracts subcorpus into the training data and the development data with proportion of 4:1. After performing the greedy feature selection algorithm on the development data, 7 features {CFACCP5, CP2R, CFCP1, AC1P, CP3, CFACCP7, AC4R} are selected consecutively for negation scope identification while 11 features {CFACCP5, AC2W, CFACCP2, CFACCP4, AC5, CFCP1, CFACCP7, CFACCP1, CP4, AC3P, CFAC2} are selected for speculation scope identification. Table 4 gives the contribution of additional features on the development data. It shows that the additional features significantly improve the performance by 11.66% in accuracy from 74.93% to 86.59% ($\chi^2; p < 0.01$) for negation scope identification and improve the performance by 11.07% in accuracy from 77.29% to 88.36% ($\chi^2; p < 0.01$) for speculation scope identification. The feature selection experiments suggest that the features (e.g., CFACCP5, AC2W, CFCP1) related to neighboring words of the cue play a critical role for both negation and speculation scope identification. This may be due to the fact that neighboring words usually imply important sentential information. For example, "*can not be*" indicates a passive clause while "*did not*" indicates an active clause.

Since the additional selected features significantly improve the performance for both negation and speculation scope identification, we will include those additional selected features in all the remaining experiments.

Task	Features	Acc (%)
Negation scope identification	Baseline	74.93
	+selected features	86.59
Speculation scope identification	Baseline	77.29
	+selected features	88.36

Table 4: Contribution of additional selected features on the development dataset of the abstracts subcorpus

Since all the sentences in the abstracts subcorpus are included in the GTB1.0 corpus while we do not have golden parse trees for the sentences in the full papers and the clinical reports subcorpora, we only evaluate the performance of scope identification on the abstracts subcorpus with golden parse trees. Table 5 presents the performance on the abstracts subcorpus by performing 10-fold cross-validation. It shows that given golden parse trees and golden cues, speculation scope identification achieves higher performance (e.g., ~3.3% higher in accuracy) than negation scope identification. This is mainly due to the observation on the BioScope corpus that the scope of a speculation cue can be usually characterized by its POS and the syntactic structures of the sentence where it occurs. For example, the scope of a verb in active voice usually starts at the cue itself and ends at its object (e.g., the speculation cue "*indicate that*" in Figure 2 scopes the fragment of "*indicate that corticosteroid resistance can not be explained by abnormalities*"). Moreover, the statistics on the abstracts subcorpus shows that the number of arguments per speculation cue is smaller than that of arguments per negation cue (e.g., 1.5 vs. 1.8).

Task	Acc (%)
Negation scope identification	83.10
Speculation scope identification	86.41

Table 5: Accuracy (%) of scope identification with golden parse trees and golden cues on the abstracts subcorpus using 10-fold cross-validation

It is worth nothing that we adopted the post-processing algorithm proposed in Section 4.4 to ensure the continuousness of identified scope. As to examine the effectiveness of the algorithm, we abandon the proposed algorithm by simply taking the left and right-most boundaries of any nodes in the tree which are classified as in scope. Experiments on the abstracts subcorpus using 10-fold cross-validation shows that the simple post-processing rule gets the performance of 80.59 and 86.08 in accuracy for negation and speculation

scope identification, respectively, which is lower than the performance in Table 5 achieved by our post-processing algorithm.

5.3 Experimental Results on Automatic Parse Trees and Golden Cues

The GTB1.0 corpus contains 18,541 sentences in which 11,850 of them (63.91%) overlap with the sentences in the abstracts subcorpus[6]. In order to get automatic parse trees, we train the Berkeley parser with the remaining 6,691 sentences in GTB1.0, which achieves the performance of 85.22 in F1-measure on the remaining 11,850 sentences in GTB1.0. Table 6 shows the performance of scope identification on automatic parse trees and golden cues. In addition, we also report an oracle performance to explore the best possible performance of our system by assuming that our scope finder can always correctly determine whether a candidate is an argument or not. That is, if an argument candidate falls within the golden scope, then it is a argument. This is to measure the impact of automatic syntactic parsing itself. Table 6 shows that:

1) For both negation and speculaiton scope identification, automatic syntactic parsing lowers the performance on the abstracts subcorpus (e.g., from 83.10% to 81.84% in accuracy for negation scope identification and from 86.41% to 83.74% in accuracy for speculaiton scope identification). However, the performance drop shows that both negation and speculation scope identification are not as senstive to automatic syntactic parsing as common shallow semantic parsing, whose performance might decrease by about ~10 in F1-measure (Toutanova et al., 2005). This indicates that scope identification via simplified shallow semantic parsing is robust to some variations in the parse trees.
2) Although speculation scope identification consistently achieves higher performance than negaiton scope identification when golden parse trees are availabe, speculation scope identification achieves comparable performance with negation scope identification on the abstracts subcorpus and the full papers

subcorpus while speculation scope identification even performs ~20% lower in accuracy than negation scope identification on the clinical report subcorpus. This is largely due to that specuaiton scope identification is more sensitive to syntactic parsing errors than negation scope identification due to the wider scope of a speculation cue while the sentences of the clinical reports come from a different genre, which indicates low performance in syntactic parsing.
3) Given the performance gap between the performance of our scope finder and the oracle performance, there is still much room for further performance improvement.

Task	Method	Abstracts	Papers	Clinical
Negation scope identification	auto	81.84	62.70	85.21
	oracle	94.37	83.33	98.39
Speculation scope identification	auto	83.74	61.29	67.90
	oracle	95.69	83.72	83.29

Table 6: Accuracy (%) of scope identification on the three subcorpora using automatic parser trained on 6,691 sentences in GTB1.0

Task	Method	Abstracts	Papers	Clinical
Negation scope identification	M et al. (2008)	57.33	n/a	n/a
	M & D (2009a)	73.36	50.26	87.27
	Our baseline	73.42	53.70	88.42
	Our final	81.84	64.02	89.79
Speculation scope identification	M & D (2009b)	77.13	47.94	60.59
	Ö & R (2009)	79.89	61.13	n/a
	Our baseline	77.39	54.55	61.92
	Our final	83.74	63.49	68.78

Table 7: Performance comparison of our system with the state-of-the-art ones in accuracy (%). Note that all the performances achieved on the full papers subcorpus and the clinical subcorpus are achieved using the whole GTB1.0 corpus of 18,541 sentences while all the performances achieved on the abstract subcorpus are achieved using 6,691 sentences from GTB1.0 due to overlap of the abstract subcorpus with GTB1.0.

Table 7 compares our performance with related ones. It shows that even our baseline system with the four basic features presented in Table 1 achieves comparable performance with Morante et al. (2008) and Morante and Daelemans (2009a & 2009b). This further indicates the appropriateness of our simplified shallow semantic parsing approach and the effectiveness of structured syntactic information on scope identification. It also shows that our final system significantly outperforms the

---

[6] There are a few cases where two sentences in the abstracts subcorpus map into one sentence in GTB1.0.

state-of-the-art ones using a chunking approach, especially on the abstracts and full papers subcorpora. However, the improvement on the clinical reports subcorpora for negation scope identification is much less apparent, partly due to the fact that the sentences in this subcorpus are much simpler (with average length of 6.6 words per sentence) and thus a chunking approach can achieve high performance. Table 7 also shows that our parsing approach to speculation scope identification outperforms the rule-based method in Özgür and Radev (2009), where 10-fold cross-validation is performed on both the abstracts and the full papers subcorpora.

5.4 Experimental Results with Automatic Parse Trees and Automatic Cues

So far negation/speculation cues are assumed to be manually annotated and available. Here we turn to a more realistic scenario in which cues are automatically recognized. In the following, we first report the results of cue recognition and then the results of scope identification with automatic cues.

Cue Recognition

Task	Features	R (%)	P (%)	F1
Negation cue recognition	CC + SW	93.80	94.39	94.09
	CC+SW+SF	95.50	95.72	95.61
Speculation cue recognition	CC + SW	83.77	92.04	87.71
	CC+SW+SF	84.33	93.07	88.49

Table 8: Performance of automatic cue recognition with gold parse trees on the abstracts subcorpus using 10-fold cross-validation

Table 8 lists the performance of cue recognition on the abstracts subcorpus, assuming all words in the sentences as candidates. It shows that as a complement to features derived from word/pos information (CC+SW features), structural features (SF features) derived from the syntactic parse tree significantly improve the performance of cue recognition by about 1.52 and 0.78 in F1-measure for negation and speculation cue recognition, respectively, and thus included thereafter. In addition, we have also experimented on only these words, which happen to be a cue or inside a cue in the training data as cue candidates. However, this experimental setting achieves a lower performance than that when all words are considered.

Task	Corpus	R (%)	P (%)	F1
Negation cue recognition	Abstracts	94.99	94.35	94.67
	Papers	90.48	87.47	88.95
	Clinical	86.81	88.54	87.67
Speculation cue recognition	Abstracts	83.74	93.14	88.19
	Papers	73.02	82.31	77.39
	Clinical	33.33	91.77	48.90

Table 9: Performance of automatic cue recognition with automatic parse trees on the three subcorpora

Table 9 presents the performance of cue recognition achieved with automatic parse trees on the three subcorpora. It shows that:
1) The performance gap of cue recognition between golden parse trees and automatic parse trees on the abstracts subcorpus is not salient (e.g., 95.61 vs. 94.67 in F1-measure for negation cues and 88.49 vs. 88.19 for speculation cues), largely due to the features defined for cue recognition are local and insenstive to syntactic variations.
2) The performance of negation cue recognition is higher than that of speculation cue recognition on all the three subcorpora. This is prabably due to the fact that the collection of negation cue words or phrases is limitted while speculation cue words or phrases are more open. This is illustrated by our statistics that about only 1% and 1% of negation cues in the full papers and the clinical reports subcorpora are absent from the abstracts subcorpus, compared to about 6% and 20% for speculation cues.
3) Unexpected, the recall of speculation cue recognition on the clinical reports subcorpus is very low (i.e., 33.33% in recall measure). This is probably due to the absence of about 20% speculation cues from the training data of the abstracts subcorpus. Moreover, the speculation cue "*or*", which accounts for about 24% of specuaiton cues in the clinical reports subcorpus, only acheives about 2% in recall largely due to the errors caused by the classifier trained on the abstracts subcorpus, where only about 11% of words "*or*" are annotated as speculation cues.

Scope Identification with Automatic Cue Recognition

Table 10 lists the performance of both negation and speculation scope identification with automatic cues and automatic parse trees. It shows that automatic cue recognition lowers the performance by

722

3.34, 6.80, and 8.38 in F1-measure for negation scope identification on the abstracts, the full papers and the clinical reports subcorpora, respectively, while it lowers the performance by 6.50, 13.14 and 31.23 in F1-measures for speculation scope identification on the three subcorpora, respectively, suggesting the big challenge of cue recognition in the two scope learning tasks.

Task	Corpus	R (%)	P (%)	F1
Negation scope identification	Abstracts	78.77	78.24	78.50
	Papers	58.20	56.27	57.22
	Clinical	80.62	82.22	81.41
Speculation scope identification	Abstracts	73.34	81.58	77.24
	Papers	47.51	53.55	50.35
	Clinical	25.59	70.46	37.55

Table 10: Performance of both negation and speculation scope identification with automatic cues and automatic parse trees

6 Conclusion

In this paper we have presented a new approach to scope learning by formulating it as a simplified shallow semantic parsing problem, which has been extensively studied in the past few years. In particular, we regard the cue as the predicate and map its scope into several constituents which are deemed as arguments of the cue. Evaluation on the Bioscope corpus shows the appropriateness of our parsing approach and that structured syntactic information plays a critical role in capturing the domination relationship between a cue and its dominated arguments. It also shows that our parsing approach outperforms the state-of-the-art chunking ones. Although our approach is only evaluated on negation and speculation scope learning here, it is portable to other kinds of scope learning.

For the future work, we will explore tree kernel-based methods to further improve the performance of scope learning in better capturing the structural information, and apply our parsing approach to other kinds of scope learning.

Acknowledgments

References

Xavier Carreras and Lluís Màrquez. 2005. Introduction to the CoNLL-2005 Shared Task: Semantic Role Labeling. *CoNLL' 2005*.

Wendy W. Chapman, Will Bridewell, Paul Hanbury, Gregory F. Cooper, and Bruce G. Buchanan. 2001. A Simple Algorithm for Identifying Negated Findings and Diseases in Discharge Summaries. *Journal of Biomedical Informatics*, 34: 301-310.

Nigel Collier, Hyun Seok Park, Norihiro Ogata, et al. 1999. The GENIA Project: Corpus-Based Knowledge Acquisition and Information Extraction from Genome Research Papers. *EACL'1999*.

Richárd Farkas, Veronika Vincze, György Móra, János Csirik, and György Szarvas. 2010. The CoNLL-2010 Shared Task: Learning to Detect Hedges and their Scope in Natural Language Text. *CoNLL'2010: Shared Task*.

Daniel Gildea and Martha Palmer. 2002. The Necessity of Parsing for Predicate Argument Recognition. *ACL'2002*.

Ilya M. Goldin and Wendy W. Chapman. 2003. Learning to Detect Negation with 'Not' in Medical Texts. *SIGIR'2003*.

Zheng Ping Jiang and Hwee Tou Ng. 2006. Semantic Role Labeling of NomBank: A Maximum Entropy Approach. *EMNLP' 2006*.

Yoong Keok Lee and Hwee Tou Ng. 2002. An Empirical Evaluation of Knowledge Sources and Learning Algorithms for Word Sense Disambiguation. *EMNLP'2002*.

Junhui Li, Guodong Zhou, Hai Zhao, Qiaoming Zhu, and Peide Qian. 2009. Improving Nominal SRL in Chinese Language with Verbal SRL Information and Automatic Predicate Recognition. *EMNLP' 2009*.

Marc Light, Xin Ying Qiu, and Padmini Srinivasan. 2004. The Language of Bioscience: Facts, Speculations, and Statements in Between. *BioLink'2004*.

Ben Medlock and Ted Briscoe. 2007. Weakly Supervised Learning for Hedge Classification in Scientific Literature. *ACL'2007*.

Roser Morante, Anthony Liekens, and Walter Daelemans. 2008. Learning the Scope of Negation in Biomedical Texts. *EMNLP'2008*.

Roser Morante and Walter Daelemans. 2009a. A Metalearning Approach to Processing the Scope of Negation. *CoNLL'2009*.

Roser Morante and Walter Daelemans. 2009b. Learning the Scope of Hedge Cues in Biomedical Texts. *BioNLP'2009*.

Lilja Øvrelid, Erik Velldal, and Stephan Oepen. 2010. Syntactic Scope Resolution in Uncertainty Analysis. *COLING'2010*.

Arzucan Özgür, Dragomir R. Radev. 2009. Detecting Speculations and their Scopes in Scientific Text. *EMNLP'2009*.

Slav Petrov and Dan Klein. 2007. Improved Inference for Unlexicalized Parsing. *NAACL'2007*.

Vasin Punyakanok, Dan Roth, and Wen-tau Yih. 2005. The Necessity of Syntactic Parsing for Semantic Role Labeling. *IJCAI' 2005*.

György Szarvas. 2008. Hedge Classification in Biomedical Texts with a Weakly Supervised Selection of Keywords. *ACL'2008*.

György Szarvas, Veronika Vincze, Richárd Farkas, and János Csirik. 2008. The BioScope corpus: Annotation for Negation, Uncertainty and their Scope in Biomedical Texts. *BioNLP'2008*.

Yuka Tateisi, Akane Yakushiji, Tomoko Ohta, and Jun'ichi Tsujii. 2005. Syntax Annotation for the GENIA Corpus. *IJCNLP'2005* (Companion volume).

Peter D. Turney. 2002. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. *ACL'2002*.

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.

Nianwen Xue and Martha Palmer. 2004. Calibrating Features for Semantic Role Labeling. *EMNLP'2004*.

Nianwen Xue. 2008. Labeling Chinese Predicates with Semantic Roles. *Computational Linguistics*, 34(2):225-255.