Prepositional Phrase Attachment Through A Hybrid Disambiguation Model

Haodong Wu and **Teiji Furugori**
Department of Computer Science
University of Electro-Communications
1-5-1, Chofugaoka,Chofu, Tokyo 182, JAPAN
{wu,furugori}@phaeton.cs.uec.ac.jp

Abstract

Prepositional phrase attachment is a major cause of structural ambiguity in natural language. Recent work has been dependent on corpus-based approaches to deal with this problem. However, corpus-based approaches suffer from the sparse-data problem. To cope with this problem, we introduce a hybrid method of integrating corpus-based approach with knowledge-based techniques, using a wide-variety of information that comes from annotated corpora and a machine-readable dictionary. When the occurrence frequency on the corpora is low, we use preference rules to determine PP attachment based on clues from conceptual information. An experiment has proven that our hybrid method is both effective and applicable in practice.

1 Introduction

The resolution of prepositional phrase attachment ambiguity is a difficult problem in NLP. There have been many proposals to attack this problem. Traditional proposals are mainly based on knowledge-based techniques which heavily depend on empirical knowledge encoded in handcrafted rules and domain knowledge in knowledge base: they are therefore not scalable. Recent work has turned to corpus-based or statistical approaches (e.g. Hindle and Rooth 1993; Ratnaparkhi, Reynar and Roukos 1994, Brill and Resnik 1994, Collins and Brooks 1995). Unlike traditional proposals, corpus-based approaches need not to prepare a large amount of handcrafted rules, they have therefore the merit of being scalable or easy to transfer to new domains. However, corpus-based approaches suffer from the notorious sparse-data problem: estimations based on low occurrence frequencies are very unreliable and often result in bad performances in disambiguation. To cope with this problem, Brill and Resnik (1994) use word classes from Word-Net noun hierarchy to cluster words into semantic classes. Collins and Brooks (1995) on the other hand use morphological analysis both on test and training data. Unfortunately, all these smoothing methods are not efficient enough to make a significant improvement on performance.

Instead of using pure statistical approaches stated above, we propose a hybrid approach to attack PP attachment problem. We employ corpus-based likelihood analysis to choose most-likely attachment. Where the occurrence frequency is too low to make a reliable choice, we turn to use conceptual information from a machine-readable dictionary to to make decision on PP attachments. We use this disambiguation method to build a disambiguation module in PFTE system.[1]

In what follows we first outline the idea of using hybrid information to supply preferences for resolving ambiguous PP attachment. We then describe how this information is used in disambiguating PP attachment. We put the hybrid approach in an disambiguation algorithm. Finally, we show an experiment and its result.

2 Using Multiple Information in Disambiguation

Like other work, we use four head words to make decision on PP attachment: the main verb v, the head noun (n1) ahead of the preposition (p), and the head noun (n2) of the object of the preposition. In the later discussion, the four head words are referred to as a quadruple (v n1 p n2).

Analyzing the strategies human beings employ in PP attachment disambiguation, we found that a wide-variety of information supplies important clues for disambiguation. It includes presuppositions, syntactic and lexical cues, collocations, syntactic and semantic restrictions, features of head words, conceptual relationships, and world knowledge. We use clues that are general and reliable

[1] PFTE stands for Parser for Free Text of English. PFTE system is a versatile parsing system in development which covers a wide range of phenomena in lexical, syntactic, semantic dimensions. It is designed as a linguistic tool for applications in text understanding, database generation from text and computer-based language learning.

so that they make the computation efficient and extensible. The information or clues we use are the following:

1. *Syntactic or lexical cues.* If n1 is same as n2, for example, often n1+PP is a fixed phrase such as *step by step.*
2. *Co-occurrences.* The co-occurrences of triples and pairs in (v n1 p n2) come from annotated corpora (Section 4).
3. *Syntactic and semantic features.* Features of v or n1 n2 sometimes indicate the "correct" attachment. For example,if v is a movement, p is *to* and n2 is a place or direction, the PP tends to be attached to the verb.
4. *Conceptual relationships* between v and n2, or between n1 and n2. These relationships, which reflect the role-expections of the preposition, supply important clues for disambiguation. For example, in the sentence *Peter broke the window by a stone*, we are sure that the PP *by a stone* is attached to *broke*/v by knowing that *stone*/n2 is an instrument for *broke*/v.

We use co-occurrence information in corpus-based disambiguation and other information in rule-based disambiguation. Later, we will discuss how to acquire above information and use it in disambiguation.

3 Estimation based on Corpora

In this section, we consider two kinds of PP attachment in our corpus-based approach, namely, attachment to verb phrase (VP attachment) and to noun phrase (NP attachment). Here, we use two annotated corpora: EDR English Corpus [2] and Susanne Corpus[3] to supply training data. Both of them contain tagged syntactic structure for each sentence in them. That is, each PP in the corpora has been attached to an unique phrase.

RA(v,n1,p,n2), a score from 0 to 1, is defined as a value of counts of VP attachments divided by the total of occurrences of (v,n1,p,n2) in the training data. [4]

$$\mathrm{RA(v,n1,p,n2)} = \frac{f(vp|v,n1,p,n2)}{f(v,n1,p,n2)} \simeq \frac{f(vp|v,n1,p,n2)}{f(vp|v,n1,p,n2)+f(np|v,n1,p,n2)} \quad (1)$$

In (1), the symbol f denotes frequency of a particular tuple in the training data. For example, f(vp | share,apartment,with,friend) is the number of times the quadruple (share, apartment, with, friend) is seen with a VP attachment. Thus, we could choose a attachment according to RA score: if RA>0.5 choose VP attachment, otherwise choose NP attachment.

Most of quadruples in test data are not in the training data, however. We thus turn to collect triples of (v,p,n1),(n1,p,n2),(v,n1,p) and pairs of (v,p),(n1,p),(p,n2) like Collins and Brooks (1995) did, and compute RA score by (2) and (3).

$$\mathrm{RA(v,n1,p,n2)} = \frac{f(vp|v,p,n2)+f(vp|n1,p,n2)+f(vp|v,n1,p)}{f(v,p,n2)+f(n1,p,n2)+f(v,n,p)} \quad (2)$$

or,

$$\mathrm{RA(v,n1,p,n2)} = \frac{f(vp|v,p)+f(vp|n1,p)+f(vp|p,n2)}{f(v,p)+f(n1,p)+f(p,n2)} \quad (3)$$

To avoid using very low frequencies, we set two thresholds for each one above. For triple-combination, the condition is:

$f_{\mathrm{triple}}(\mathrm{v,n1,p,n2}) \geq 2$, and
$|2*\mathrm{RA(v,n1,p,n2)}-1| * \log(f_{\mathrm{triple}}(\mathrm{v,n1,p,n2})) < 0.5$

here,

$f_{\mathrm{triple}}(\mathrm{v,n1,p,n2}) = \mathrm{f(v,p,n2)+f(n1,p,n2)+f(v,n1,p)}$

For pairs-combination, the condition is:

$f_{\mathrm{pair}}(\mathrm{v,n1,p,n2}) \geq 4$, and
$|2*\mathrm{RA(v,n1,p,n2)}-1| * \log(f_{\mathrm{pair}}(\mathrm{v,n1,p,n2})) < 0.5$

here,

$f_{\mathrm{pair}}(\mathrm{v,n1,p,n2}) = \mathrm{f(v,p)+f(n1,p)+f(p,n2)}$

With the first threshold in each case, we can avoid using low frequency tuples; with the second one in each case, we throw away the RA score which is close to 0.5 as this value is rather unstable.

4 Conceptual Information and Preference Rules

As we use only "reliable" data from corpora to make decision on PP attachment based on RA score, many PPs' attachments may be left undetermined due to sparse data. We deal these undetermined PPs with a rule-based approach. Here we use preference rules to determine PP attachments by judging features of head words and conceptual relationships among them. This information comes from a machine-readable dictionary - EDR dictionary.[5]

[2] EDR English Corpus, compiled by Japan Electronic Dictionary Research Institute, Ltd, contains 160,000 sentences with annotated morphologic, syntactic and semantic information.

[3] Susanne Corpus,compiled by Geoffrey Sampson, is an annotated corpus comprising about 130,000 words of written American English text.

[4] We assume that only two kinds of PP attachments: VP or NP attachment in the training data.

[5] EDR electronic dictionary consists of a set of machine-readable dictionaries which includes Japanese and English word dictionary, Japanese and English co-occurrence dictionary, concept dictionary, and Japanese < —— > English bilingual dictionary(EDR 1993).

4.1 Features and Concept Classes

We cluster words (verbs or nouns) which have same feature or syntactical function into a concept class. For example, we classify verbs into *active* and *passive*, and ontological classes of *mental*, *movement*, etc. Similarly, we group nouns into *place*, *time*, *state*, *direction*, etc.

We extract concept class from *concept classification* in EDR Concept Dictionary.[6]

4.2 Conceptual Relationship

Conceptual relationships between v and n2, or between n1 and n2 predict PP attachment quite well in many cases. We use EDR concept dictionary to acquire the conceptual relationship between two concepts. For example, given the two concepts of *open* and *key*, the dictionary will tell us that there may be a *implement* relationship between them, means that *key* may be act as an instrument for the action *open*.

4.3 Preference Rules

We introduce preference rules to encode syntactic and lexical clues, as well as clues from conceptual information to determine PP attachments. We divide these rules into two categories: a rule which can be applied to most of prepositions is called *global rule*; a rule tying to a particular preposition, on the other hand, is called *local rule*. Four global rules used in our disambiguation module are listed in Table 1.

1. lexical(passivized(v) + PP) AND prep $\neq$ 'by' $->$ vp_attach(PP)
2. n1 = n2 $->$ vp_attach(n1 + PP)
3. (prep $\neq$ 'of' AND prep $\neq$ 'for') AND (time(n2) OR date(n2)) $->$ vp_attach(PP)
4. lexical(Adjective + PP) $->$ adjp_attach(PP)

Table 1: Global rules

Local rules use conceptual information to determine PP attachment. In Table 2, we show sample local rules for preposition *with*.

with-rules:

implement(v, n2) $->$ vp_attach(PP)

(a-object(n1, n2) OR possessor(n1, n2)) AND NOT(implement(v, n2)) $->$ np_attach(PP)

Default $->$ vp_attach(PP)

Table 2: Sample local rules

On the left hand of each rule, a one-atom predicate on the left hand presents a subclass of concept in the concept hierarchy (e.g. time(n2)), and a two-atom predicate describes the concept relation between two atoms (e.g. implement(v,n2)).

Since local rules employ the senses of head words (termed as *concepts*), we should project each of v, n1 and n2 used by rules into one or several concepts which denote(s) "correct" word senses before applying local rules. The process is described in (Wu and Furugori 1995).

5 Disambiguation Module

For each sentence with ambiguous PP (both in syntactic and semantical level), PETE system will produce a structure with unattached PP(s), and call the disambiguation module to resolve ambiguous PP(s). The algorithm used in the module is shown below :

[ALGORITHM]

Phase 1. (disambiguation using global rules):
Try global rules one by one. If a rule succeeds, use it to decide the attachment, and exit.

Phase 2. (statistics-based disambiguation):

RA(v,n1,p,n2) = -1 (initial value)
f_{triple}(v,n1,p,n2) = f(v,p,n2)+f(n1,p,n2)+f(v,n1,p)
f_{pair}(v,n1,p,n2) = f(v,p)+f(n1,p)+f(p,n2)

if $f_{triple}(v,n1,p,n2) \geq 2$, then
$RA(v,n1,p,n2) = \frac{f(vp|v,p,n2)+f(vp|n1,p,n2)+f(vp|v,n1,p)}{f(v,p,n2)+f(n1,p,n2)+f(v,n,p)}$

if $|2*RA(v,n1,p,n2)-1| * \log(f_{triple}(v,n1,p,n2)) < 0.5$
then RA(v,n1,p,n2) = -1

if $RA(v,n1,p,n2) < 0$ and $f_{pair}(v,n1,p,n2) \geq 4$, then
$RA(v,n1,p,n2) = \frac{f(vp|v,p)+f(vp|n1,p)+f(vp|p,n2)}{f(v,p)+f(n1,p)+f(p,n2)}$

if $|2*RA(v,n1,p,n2)-1| * \log(f_{pair}(v,n1,p,n2)) < 0.5$
then RA(v,n1,p,n2) = -1

if $RA(v,n1,p,n2) \geq 0$, then {
if $RA(v,n1,p,n2) < 0.5$, then choose NP attachment
otherwise choose VP attachment
exit.}

Phase 3. (concept-based disambiguation):

1) Project each of v, n1, n2 into its concept sets.

2) Try the rules related to the preposition, if only one rule is applicable, use it to decide the attachment, and then exit.

Phase 4. (attachment by default):

if f(p) > 0, then {
if $\frac{f(vp|p)}{f(p)} < 0.5$, then choose NP attachment;
otherwise choose VP attachment}
otherwise choose NP attachment.

This algorithm differs from the previous one described in (Wu and Furugori 1995) in which preference rules were applied before statistical computing. We have changed the order for the following reasons: an experiment has proven that using the

[6] Concept Dictionary consists of about 400,000 concepts, where, for concept classification, related concepts are organized in hierarchical architecture and a concept in lower level inherits the features from its upper level concepts.

data of quadruples and triples, as well as tuples with high occurrences is good enough in success rate (See Table 3). and statistic models have a ground mathematical basis.

6 Experiment and Evaluation

We did an experiment to test our method. First, we prepared test data of 3043 ambiguous PPs in texts randomly taken from a computer manual, a grammar book and *Japan Times*.

Phase	Total Number	Number Correct	Success rate
global rules	507	487	96.1%
triples	564	518	91.8%
pairs	1093	931	85.3%
local rules	662	557	84.1%
others	217	151	69.6%
Total	3043	2644	86.9%

Table 3: Results of the test in PP attachment

The results are shown in Table 3. We successfully disambiguated 86.9% of the test data. To reduce sparse data problem and deal with undefined words in the dictionary, we use a procedure similar to that of Collins and Brook (1995) to process head words both in training data and in test data. The procedure is shown as follows:

- All 4-digit numbers are replaced with 'date'.
- All verbs are replaced with their stems in lower cases.
- Nouns starting with a capital letter are replaced with 'name'.
- Personal pronouns in the n2 field are replaced with 'person'.

As the result, we acquired an accurate rate of 87.5% (Table 4), an improvement of 0.6% on the previous one.

Phase	Total Number	Number Correct	Success rate
global rules	507	487	96.1%
triples	659	601	90.9%
pairs	1134	965	84.9%
local rules	628	527	83.9%
others	115	81	70.4%
Total	3043	2661	87.5%

Table 4: Results with processing head words

The result is rather good, comparable to the performance of an "average" human looking at (v,n1,p,n2) alone (about 85% to 90% according to Hindle and Rooth 1993, Collins and Brooks 1995). We attribute this result to the hybrid approach we used, in which preferences with higher reliabilities are used prior to other ones in the disambiguation process. We found that two thresholds are very helpful in improving the result. If we set the first threshold as 0 and throw away the second threshold, then the success rates in triple-combination will become 89.1% (-1.8%), and 81.2% (-3.7%) in pair-combination. Moreover, using local rules to tackle unattached PPs by statistical model is also helpful in improving the overall success rate since local rules in Phase 3 work much better than default decision in Phase 4.

7 Conclusion

Pure statistical models for disambiguation tasks suffer from sparse-data problem. We noted that even when applying smooth techniques such as semantic similarity or clustering, it is hard to avoid making poor estimations on low occurrences in corpora. On-line dictionaries which contain rich semantic or conceptual information may be of help in improving the performance. Our experiment shows that the hybrid approach we taken is both effective and applicable in practice.

References

Brill, E. and Resnik, P. 1994. A rule-based approach to prepositional phrase attachment disambiguation. In *Proc. of the 15th Coling*, 1198 1204.

Collins, M. and Brooks, J. 1995. Prepositional phrase attachment through a backed-off model. http://xxx.lanl.gov/cmp-lg/9506021.

Dahlgren, K. and McDowell, J. 1986. Using commonsense knowledge to disambiguate prepositional phrase modifiers. In *Proc. of the 5th AAAI*, 589-593.

Japan Electronic Dictionary Research Institute, Ltd. 1993. EDR electronic dictionary specifications guide.

Jensen, K. and Binot, J. 1987. Disambiguating prepositional phrase attachments by using online dictionary definition. In *Computational Linguistics*. 13(3-4) : 251-260.

Hindle, D. and Rooth, M. 1993. Structural ambiguity and lexical relations. In *Computational Linguistics*, 19(1) : 103-120.

Luk, A. K. 1995. Statistical sense disambiguation with relatively small corpora using dictionary definitions. In *Proc. of the 33rd ACL Meeting*, 181-188.

Whittemore, G.; Ferrara, K.; and Brunner, H. 1990. Empirical study of predictive powers of simple attachment schemes for post-modifiers prepositional phrases. In *Proc. of the 28th ACL Meeting*, 23-30.

Wu, H., Takeshi, I. and Furugori, T. 1995. A preferential approach for disambiguating prepositional phrase modifiers. In *Proc. of the 3th Natural Language Processing Pacific Rim Symposium*, 745-751.