An Integrated Approach for Prepositional Phrase Attachment Resolution

KOTOH GO
Dokkyo University

We in this paper propose an integrated approach for resolving prepositional phrase attachment using preference rules from syntactic cues, likelihood estimation based on corpus, and similar words acquired from thesaurus. This approach provides a higher performance and wide coverage for disposing various PP attachment ambiguities.

1. Introduction

Prepositional Phrase (PP) attachment is a paradigm case of syntactic ambiguity to be resolved. We in this paper propose a hybrid method that employs cues from corpora, thesaurus, and linguistic observations to predict the most-likely PP attachment.

In what follows we firstly outline the idea of using various types of information to supply preferences for resolving ambiguous PP attachment. We then describe how the information is employed in disambiguating PP attachment. After building a back-off model, we finally show a disambiguation experiment and discuss its result.

A typical context for disambiguating PP attachment is subject+verb+np1+prep+np2. However, as the subject has less impact on PP attachment, many proposals (e.g., Collins and Brooks 1995) use verb+np1+prep+np2 as the context for disambiguation. Np1 and np2 can be arbitrarily complicated (e.g., a beautiful, lovely and smart girl). This makes it very difficult or sometimes impossible to collect statistics or find usable cues from np's. As other many work, we assume here that the only thing important to a noun phrase is the head noun. We then use a quadruple of (v, n1, p, n2) as the short hand for representing the context. Thus, in sentence

Tom broke the boxes with a hammer.

The quadruple is (break, box, with, hammer). Here we use root forms of words.

2. Cues for Prepositional Phrase Attachment

2.1 Preference Rules

We use preference rules to encode syntactic cues to determine PP attachments. These rules are applicable to any preposition as below.

Rule 1. n 1 = n2 → vp_attachment(n1 + PP)
If n2 repeats n1 (e.g, step by step), n1 + PP is a fixed phrase to modify the verb.

Rule 2. Lexical(passiviaed(v) + PP) AND prep ≠ 'by' → vp_attachment(PP)
The PP is attached to the VP if the verb is intransitive.

Rule 3. (prep ≠ 'of' AND prep ≠ 'for') AND (time(n2) OR date(n2)) → vp_attachment(PP)
The PP is attached to the VP is n2 is a time or a date (e.g, yesterday) and the Preposition is **not** of or **for**.

Rule 4. Lexical(Adjective + PP) → adjp_attachment
If the PP comes after an adjective (including the participal adjective), it is attached to the adjective as its complement.

2.2 Likelihood Estimation Based on Corpora

We employ the RA (Ratio of Association) score to estimated the likelihood for attachment for a certain (v, n1, P, n2). The RA score is defined in (1.1) as the value of counts of VP attachments divided by the total occurrence of (v, n1, p, n2) in the training data.

$$RA(v, n1, p, n2) = \frac{f(vp|v,n1,p,n2)}{f(v,n1,p,v2)} \qquad (2.1)$$

The quadruples in test data are seldom found in training data due to the data sparseness. We then use the synonyms or strictly similar words (e.g. breakfast, lunch, dinner) of n1 and n2 from WordNet and Roget's thesaurus to improve the performance and replace formula (2.1) by (2.2).

$$RA(v, n1, p, n2) = \frac{f(vp|v,sim(n1),p,sim(n2))}{f(v,sim(n1),p,sim(n2))} \qquad (2.2)$$

This method can improve the bottleneck of data sparseness. However, many cases are also not found in (1.2). we there turn to collecting triplets of (v,p,n2),(n1,p,n2),(v,n1,p) and pairs of (v, p),(n1),(p, n2), and compute the RA score using either (2.3) or (2.4).

$$RA(v, n1, p, n2) \fallingdotseq \frac{f(vp|v,p,sim(n2))+f(vp|sim(n1),p,sim(n2))+f(vp|v,sim(n1),p)}{f(v,p,sim(n2))+f(sim(n1),p,sim(n2))+f(v,sim(n1),p)} \qquad (2.3)$$

$$RA(v, n1, p, n2) \fallingdotseq \frac{f(vp|v,p)+f(vp|sim(n1),p)+f(vp(p,sim(n2))}{f(v,p)+f(sim(n1),p)+f(p,sim(n2))} \qquad (2.4)$$

We find that the RA score makes an undesirable estimation in two cases. It is when the co-occurrences are too low (e.g. only one or two case found in the training data) or the score is close to the boundary value of 0.5. We then introduce two thresholds to deal with these problems.

For (1.2), the condition is

f(v, sim(n1),p, sim(n2))>=3, and

|2*RA(v, n1, p, n2)-1|*log(f(v, n1, p, n2)) < 0.5

For (1.3), the condition is

f(v, sim(n1),p, sim(n2))>=5, and

|2*RA(v, n1, p, n2)-1|*log(f(v, n1, p, n2)) < 0.5

For(1.4), the condition is

f(v, sim(n1),p, sim(n2))>=9, and

|2*RA(v, n1, p, n2)-1|*log(f(v, n1, p, n2)) < 0.5

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

3. The Integrated PP Attachment Algorithm

For each sentence with ambiguous PP(s), the disambiguation process of our model is put in an algorithm in the following:

Phase 1. (disambiguation using rules)
Try rule described in 2.1 one by one. If a rule succeeds, employ it to decide the attachment, and exit.

Phase 2 (statistics-based disambiguation extended with thesaurus)
Set the initial value for RA(v, n1, p, n2) = -1;
Define quadruple(v, n1, p, n2) as f(v, sim(n1), p, sim(n2)), and
ftriplt(v, n1, p, n2) as f(v, p, sim(n2)) + f(sim(n1), p, sim(n2)) + f(v, sim(n1), p), and
fpair(v, n1, p, n2) as f(v, p) + f(sim(n1),p) + f(p, sim(n2))

If fquadruple(v, n1, p, n2) >= 3, then

$$RA(v, n1, p, n2) = \frac{f(vp|v,n1,p,n2)}{f(v,n1,p,v2)}$$

and If |2*RA(v, n1, p, n2)-1|*log(f(v, n1, p, n2)) < 0.5 then

RA(v, n1, p, n2)=-1

else if RA(v, n1, p, n2) <0 and ftriplt(v, n1, p, n2) >= 5, then

$$RA(v, n1, p, n2) \fallingdotseq \frac{f(vp|v,p,sim(n2))+f(vp|sim(n1),p,sim(n2))+f(vp|v,sim(n1),p)}{f(v,p,sim(n2))+f(sim(n1),p,sim(n2))+f(v,sim(n1),p)}$$

and if |2*RA(v, n1, p, n2)-1|*log(f(v, n1, p, n2)) < 0.5 then

RA(v, n1, p, n2)=-1

else if RA(v, n1, p, n2) <0 and fpair(v, n1, p, n2) >= 9, then

$$RA(v, n1, p, n2) \fallingdotseq \frac{f(vp|v,p)+f(vp|sim(n1),p)+f(vp(p,sim(n2))}{f(v,p)+f(sim(n1),p)+f(p,sim(n2))}$$

else if |2*RA(v, n1, p, n2)-1|*log(f(v, n1, p, n2)) < 0.5

then RA(v, n1, p, n2) = -1

if RA(v, n1, p, n2) >= 0 then {

if RA(v, n1, p, n2)< 0.5,

then choose NP attachment;

else choose VP attachment.

}

Phrase 3. (attachment by default)

if f(vp|p)/f(p) < 0.5, then choose NP attachment;

else choose VP attachment.

4. Results

A total of 10694 sentences with PP ambiguities were extracted from the Wall Street Journal corpus of the Penn Treebank. The sentences are converted to quadruples. In the first test, all four-tuples with the preposition *of* are removed. This leaves 7810 sentences. Of these, the first 2000 are retained for testing, with the remaining 5810 are used for training.

4.1 Lower and Upper Bounds on Performance

Finding the lower and upper bounds on the performance of a method is helpful in evaluating its effectiveness. The table 1 shows some reference performances. Here, 'Most likely for each preposition' means an attachment of the PP to either a noun or a verb depending on which case is predominant in the training data. 'Average human (4 head words only)' and are taken from the experiment done by Hindle et al. (1990).

Table 1. Lower and Upper Bounds for Evaluating Performance

Method	Percentage Accuracy
Most likely for each preposition	72.2
Average human (4 head word only)	88.3
Average human (whole sentence)	93.3

For our experiment, the lower bound seems to be 72.2% scored by 'Most likely for each preposition' and the upper bound may be 88.3% by 'Average human (4 head words only)'. I say from this result that the success rate of our method, 86.8%, is closed to the performance human beings.

4.2 Comparison with Other Work

We make a comparison of the performance of my method with that of other work. Table 2 gives the results of comparison. (1) shows the result of an example-based method by Sumita et al. (1994), where both the training data (3299 handcrafted examples) and the test data are taken in a small domain (international conference reservation). (3) and (4) show the results of two statistics-based methods (Brill and Resnik 1994; Collins and Brook 1995), where both of them use the WSJ corpus and the IBM data as their test data. (5) shows our previous work that use both statistic information and conceptual information extracted from EDR dictionaries.

We attribute our sound result to the integrated model in which different cues are explored and the sound cues are applied before less sound cues. Using thesaurus and two thresholds are also of great help to overcome the sparse data problem and improve the performance for predicting the correct PP attachment.

Table 2. Comparison with Other work

Method	Number Test	Success Rate
(1) Example-based (SFI 94)	131	85.7
(2) Rule-based (BR 94)	3097	81.9
(3) Backed-off (CB 95)	3097	84.5
(4) Hybrid model (WT 96)	3093	86.9
(5) Our method	2000	86.8

Table 2 is not a direct comparison between systems because the systems have been trained and tested on different corpora. None the less, the system described in this paper performs better than almost all of the other systems. Moreover, the algorithm is a relatively simple one based on the combination of hierarchies.

5. Conclusion

Pure statistic-based models, example-based models and dictionary-based models for PP attachment resolution task suffer from the sparse data problem and/or not so robust. This experiment has proved that the integrated approach described in this paper is both effective and applicable in practice.

This method has some weakness. Although full-sentence context and full-text context are helpful for improving the performance for resolving PP attachment ambiguities, the disambiguation model do not use them at all. Further improvement may be attained by using larger corpus and other linguistic resources such as Web, and perhaps an incremental disambiguation procedure in consulting a wider context.

References

1) M. Atterer, H. Schütze, "Prepositional Phrase Attachment Without Oracles." *Computational Linguistics*. 33(4), pages 469-476, 2007.

2) E. Brill and P. Resnik. "A Rule-based Approach to Prepositional Phrase Attachment Disambiguation." In *Proceedings Of the 29th COLING*, pages 1198-1204, 1994.

3) M. Collions and J. Brooks "Prepositional Phrase Attachment Through a Backed-off Model, *acl.ldc.upenn.edu/W/W95/W95-0103.pdf.*

4) S., Jiri and M. Nagao. "Corpus based PP Attachment Ambiguity resolution with a Semantic Dictionary". In *Proceedings of the Fifth Workshop on Very Large Corpora,* 4) M. McLauchlan, "Thesauruses for Prepositional Phrase pages 66–80, Beijing/Hong Kong, 1997.

5) M. McLauchlan, "Thesauruses for Prepositional Phrase Attach-ment."*www.aclweb.org/anthology-new/.../W04-2410.pdf*

6) K. Nadh, C. Huyck. "A Neurocomputational Approach to Prepositional Phrase Attachment Ambiguity Resolution". *Neural Computation* 24(7), pages 1906-1925, Online publication date: 1-Jul-2012.

7) A. Ratnaparkhi, J. Reynar, and S. Roukos. "A Maximum Entropy Model for Prepositional Phrase Attachment." In *Proceedings of the ARPA Human Language Technology Workshop*, pages 250–252, 1994.

8) Z., Shaojun and D. Lin. 2004. "A Nearest-neighbor Method for Resolving PP-attachment Ambiguities". In *The First International Joint Conference on Natural Language Processing* (IJCNLP-04), pages 545–554, Hainan Island, China, 2004.

9) C. Schutze and E. Gibson." Argumenthood and English Prepositional Phrase Attachment." *Journal of Memory and Language*, 40:409–431, 1999.

10) H. Wu and T. Furugori, "A Hybrid Disambiguation Model for Prepositional Phrase Attachment." *Literary and Linguistic Computing*. 11(4). 1996.