IRST-BP: Preposition Disambiguation based on Chain Clarifying Relationships Contexts

Octavian Popescu
FBK-IRST, Trento (Italy)
popescu@itc.it

Sara Tonelli
FBK-IRST, Trento (Italy)
satonelli@itc.it

Emanuele Pianta
FBK-IRST, Trento (Italy)
pianta@itc.it

Abstract

We are going to present a technique of preposition disambiguation based on sense discriminative patterns, which are acquired using a variant of Angluin's algorithm. They represent the essential information extracted from a particular type of local contexts we call Chain Clarifying Relationship contexts. The data set and the results we present are from the Semeval task, WSD of Preposition (Litkowski 2007).

1 Introduction

Word Sense Disambiguation (WSD) is a problem of finding the relevant clues in a surrounding context. Context is used with a wide scope in the NLP literature. However, there is a dichotomy among two types of contexts, local and topical contexts (Leacock et. all 1993), that is general enough to encompass the whole notion and at the same to represent a relevant distinction.

The local context is formed by information on word order, distance and syntactic structure and it is not restricted to open-class words. A topical context is formed by the list of those words that are likely to co-occur with a particular sense of a word. Generally, the WSD methods have a marked predilection for topical context, with the consequence that structural clues are rarely, if ever, taken into account. However, it has been suggested (Stetina&Nagao 1997, Dekang 1997) that structural words, especially prepositions and particles, play an important role in computing the lexical preferences considered to be the most important clues for disambiguation.

Closed class words, prepositions in particular, are ambiguous (Litkowski&Hargraves2006). Their disambiguation is essential for the correct processing of the meaning of a whole phrase. A wrong PP-attachment may render the sense of the whole sentence unintelligible. Consider for example:

(1) Joe heard the gossip about you and me.
(2) Bob rowed about his old car and his mother.

A probabilistic context free grammar most likely will parse both (1) and (2) wrongly[1]. It would attach "about" to "to hear" in (1) and would consider the "his old car and his mother" the object of "about" in (2).

The information needed for disambiguation of open class words is spread at all linguistics levels, from lexicon to pragmatics, and can be located within all discourse levels, from immediate collocation to paragraphs (Stevenson&Wilks 1999). Intuitively, prepositions have a different behavior. Most likely, their senses are determined within the government category of their

[1] Indeed, Charniak's parser, considered to be among the most accurate ones for English, parses wrongly both of them.

Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007), pages 191–194,
Prague, June 2007. ©2007 Association for Computational Linguistics

heads. We expect the local context to play the most important role in the disambiguation of prepositions.

We are going to present a technique of preposition disambiguation based on sense discriminative patterns, which are acquired using a variant of Angluin's algorithm. These patterns represent the essential information extracted from a particular type of local contexts we call Chain Clarifying Relationship contexts. The data set and the results we present are from the Semeval task, WSD of Preposition (Litkowski 2007).

In Section 2 we introduce the Chain Clarifying Relationships, which represent particular types of local contexts. In Section 3 we present the main ideas of the Angluin algorithm. We show in Section 4 how it can be adapted to accommodate the preposition disambiguation task. Section 5 is dedicated to further research.

2 Chain Clarifying Relationships

We think of ambiguity of natural language as a net - like relationship. Under certain circumstances, a string of words represents a unique collection of senses. If a different sense for one of these words is chosen, the result is an ungrammatical sentence. Consider (3) below:

> (3) Most people do not live in a state of high intellectual awareness about their every action.

Suppose one chooses the sense of "*to live*" to be "*to populate*". Then, its complement, "*state*", should be synonym with location. The analysis crashes when "*awareness*" is considered. There are two things we notice here: (a) the relationship between "*live*" and "*state*" – the only two acceptable sense combination out of four are (populate, location) and (experience, entity) – and (b) the chain like relationship between "*awareness*", "*state*", "*live*" where the sense of any of them determines the sense of all the others in a cascade effect, or results in ungrammaticality. A third thing, not directly observable in (3) is that the syntactic configuration is crucial in order for (a) and (b) to arise. Example (4) shows that in a different syntactic configuration the above sense relationship simply disappears:

> (4) The awareness of people about the state institutions is arguably the first condition to live in a democratic state.

We call the relationship between "*live*", "*state*", "*awareness*" a Chain Clarifying Relationship (CCR). In that specific syntactic configuration their senses are interdependent and independent of the rest of the sentence. To each CCR corresponds a sense discriminative pattern. Our goal is to learn which local contexts are CCRs. Each CCR is a pattern of words on a syntactic configuration. Each slot can be filled only by words defined by certain lexical features. To learn a CCR means to discover the syntactic configuration and the respective features. For example consider (5) and (6) with their CCRs in (CCR5) and (CCR6) respectively:

> (5) Some people lived in the same state of disappointment/ optimism/ happiness.
> (CCR5) (vb=live_sense_2, prep1=in_1, prep1_obj=state_sense_1,prep2=of_sense_1a,prep2_obj=[State_of_Spirit])
> (6) Some people lived in the same state of Africa/ Latin America/ Asia.
> (CCR6) (vb=live_sense_1, prep1=in_1, prep1_obj=state_sense_1,prep2=of_1b,prep2_obj = [Location])

The lexical features of the open class words in a specific syntactic configuration trigger the senses of each word, if the context is a CCR. In (CCR5) any word that has the same lexical trait as the one required by prep2_obj slot will determine a unique sense for all the other words, including the preposition. The same holds for (CCR6). The difference between (CCR5) and (CCR6) is part of the linguistic knowledge (which can be clearly shown: "*how*" (5) vs. "*where*" (6)).

The CCR approach proposes a deterministic approach to WSD. There are two features of CCRs which are interesting from a strictly practical point of view. Firstly, CCR proposal is a way to determine the size of the window where the disambiguation clues are searched for (many WSD algorithms arbitrarily set it apriori). Secondly, within a CCR, by construction, the sense of one word determines the senses of all the others.

3 Angluin Learning Algorithm

Our working hypothesis is that we can learn the CCRs contexts by inferring differences via a regular language learning algorithm. What we want to learn is which features fulfil each syntactic slot. First we introduce the original Angluin's algorithm and then we mention a variant of it admitting unspecified values.

Angluin proved that a regular set can be learned in polynomial time by assuming the existence of an oracle which can gives "yes/no" answers and counterexamples to two types of queries: membership queries and conjecture queries (queries about the form of the regular language) (Angluin 1998).

The algorithm employs an observation table built on prefix /suffix closed classes. To each word a {1, 0} value is associated, "1" meaning that the word belongs to the target regular language. Initially the table is empty and is filled incrementally. The table is closed if all prefixes of the already seen examples are in the table and is consistent if two rows dominated by the same prefix have the same value, "0" or "1".

If the table is not consistent or closed then a set of membership queries is made. If the table is consistent and closed then a conjecture query is made. If the oracle responds "no", it has to provide a counterexample and the previous steps are cycled till "yes" is obtained.

The role of the oracle for conjecture questions can be substituted by a stochastic process. If strict equality is not requested, then a probably approximately correct identification of language can be obtained (PAC identification), which guarantees that the two languages (the identified one, L_i, and the target one, L_t) are equal up to a certain extent. The approximation is constrained by two parameters ε – accuracy and δ – confidence, and the constraint is $P(d(L_i, L_t) \leq \varepsilon) \geq \delta)$, where the distance between two languages is the probability to see a word in just one of them.

The algorithm can be further generalized to work with unspecified values. The examples may have three values ("yes", "no", "?"), as in many domains one has to deal with partial knowledge The main result is that a variant of the above algorithm successfully halts if the number of counterexamples provided by the oracle have O(log *n*) missing attributes, where *n* is the number of attributes (Goldmann et all 2003).

4 Preposition Disambiguation Task

The CCR extraction algorithm is supervised. Consider that you have a sense annotated corpora. Extract the dependency paths and filter out the ones which are not sense discriminative. Try to generalize each slot and retain the minimal ones. What is left are CCRs.

Unfortunately, for the preposition disambiguation task the training set is sense annotated only for prepositions. We have undertaken a different strategy. The training corpus can be used as an oracle. The main idea is to start with a set of few examples for each sense from the training set which are considered to be the most representative ones. We try to generalize each of them independently and to tackle down the border cases (the cases that may correspond to two different senses) which are considered unspecified examples. The process stops when the oracle does not bring any new information (the training cases have been learned). Below we explain this process step by step.

Step 1. Get the seed examples. For each preposition and sense get the seed examples. This operation is performed by a human expert. It may be the case that the glosses or the dictionary definition are a good starting point (with the advantage that the intervention of a human is no more required). However, we preferred do to it manually for better precision.

Besides the most frequent sense, we have considered, in average, another two senses. There is a practical reason for this limitation: the number of examples for the rest of the senses is insufficient. In total we have considered 149 senses out of the 241 senses present in the training set. For each an average of three examples has been chosen.

Step 2. Get the CCRs. For each example we read the lex units associated with its frame from FrameNet. Our goal is to identify the relevant syntactic and lexical features associated with each slot. We have undertaken two simplifying assumptions. Firstly, only the government category of the head of the PP is considered (which can be a verb, a noun or an adjective). Secondly,

the lexical features are identified with synsets from WordNet.

We have used the Charniak's parser to extract the structure of the PP-phrases and further we have used Collin's algorithm to implement a head recogniser.

A head can have many synsets. In order to understand which sense the word has in the respective construction we look for the synset common to the elements extracted from lex. If the proposed synset uniquely identifies just one sense then it is considered a CCR. If not, we are looking for the next synset. This step corresponds to membership queries in Angluin's algorithm.

Step 3. Generalize the CCRs. At the end of step 2 we have a set of CCRs for each sense. We obtained 395 initial CCRs. We tried to extend the coverage by taking into account the hyperonyms of each synsets. Only approximately 10% of these new patterns have received an answer from the oracle. Consequently, for our approach ,a part of the training corpus has not been used. It serves only 15 examples in average to get a correct CCR. All the instances of the same CCR do not bring any new information to our approach.

Posteriori, we have noticed that the initial patterns have an almost 50% (48.57%) coverage in the test data. The generalized patterns obtained after the third step have 82% test corpus coverage. For the rest 18%, which are totally unknown cases, we have chosen the most frequent sense.

In table 1 we present the performances of our system. It achieves 0.65 (FF-score), which compares favourably against baseline – the most frequent -of 0.53. On the first column of Table 1 we write the FF score interval - more than 0.75, between 0.75 and 0.5, and less than 0.5 respectively, - on the second column we present the number of cases within that interval the system solved and on the third column we include the corresponding number for baseline.

Table 1

Interval	System	Baseline
1.00 - 0.75	18	8
0.75 - 0.50	15	6
0.00 – 0.50	2	20

5 Conclusion and Further Research

Our system did not perform very well (third position out of three). Analyzing the errors, we have noticed that our system systematically confound two senses in some cases (for example "*by*" 5(2) vs. 15(3), for "*on*" 4(1c) vs. 1(1) etc.). We would like to see whether these errors are due to a misclassification in training.

References

Angluin, D. (1987): "Learning Regular Sets from Queries and Counterexamples", Information and Computation Volume 75 , Issue 2

Goldman, S., Kwek, S., Scott, S. (2003): "Learning from examples with unspecified attribute values", Information and Computation, Volume 180

Leacock, C., Towell, G., Voorhes, E. (1993): "Towards Building Contextual Representations of Word Senses Using Statistical Models", In Proceedings, SIGLEX workshop: Acquisition of Lexical Knowledge from Text

Lin, D. (1997): "Using syntactic dependency as local context to resolve word sense ambiguity".ACL/EACL-97, Madrid

Litkowski, K. C. (2007):"Word Sense Disambiguation of Prepositions" , The Semeval 2007 WePS Track. In Proceedings of Semeval 2007, ACL

Litkowski, K. C., Hargraves O. (2006): "Coverage and Inheritance in the Preposition Project", Proceedings of the Third ACL-SIGSEM Workshop on Prepositions, Trento,

Stetina J, Nagao M (1997): "Corpus based PP attachment ambiguity resolution with a semantic dictionary.", Proc. of the 5th Workshop on very large corpora, Beijing and Hongkong, pp 66-80

Stevenson K., Wilks, Y.,(2001): "The interaction of knowledge sources in word sense disambiguation", Computational Linguistics, 27(3):321–349.