Word Sense Disambiguation Using Pairwise Alignment

Koichi Yamashita **Keiichi Yoshida** **Yukihiro Itoh**

Faculty of Administration and Informatics
University of Hamamatsu
1230 Miyakoda-cho, Hamamatsu, Shizuoka, Japan
yamasita@hamamatsu-u.ac.jp

Abstract

In this paper, we proposed a new supervised word sense disambiguation (WSD) method based on a pairwise alignment technique, which is used generally to measure a similarity between DNA sequences. The new method obtained 2.8%-14.2% improvements of the accuracy in our experiment for WSD.

1 Introduction

WSD has been recognized as one of the most important subjects in natural language processing, especially in machine translation, information retrieval, and so on (Ide and Véronis, 1998). Most of previous supervised methods can be classified into two major ones; approach based on association, and approach based on selectional restriction. The former uses some words around a target word, represented by *n*-word window. The latter uses some syntactic relations, say, verb-object, including necessarily a target word.

However, there are some words that one approach gets good result for them while another gets worse, and vice versa. For example, suppose that we want to distinguish between "go off or discharge" and "terminate the employment" as a sense of "fire". Consider the sentence in Brown Corpus[1]:

> My Cousin Simmons carried a musket, but he had loaded it with bird shot, and as the officer came opposite him, he rose up behind the wall and fired.

[1]In this case, we consider only one sentential context for the simplicity.

The words such as "musket", "loaded" and "bird shot" would seem useful in deciding the sense of "fire", and serve as clue to leading the sense to "go off or discharge". It seems that there is no clue to another sense. For this case, an approach based on association is useful for WSD. However, an approach based on selectional restriction would not be appropriate, because these clues do not have the direct syntactic dependencies on "fire". On the other hand, consider the sentence in EDR Corpus:

> Police said Haga was immediately fired from the force.

The most significant fact is that "Haga" (a person's name) appears as the direct object of "fire". A selectional restriction approach would use this clue appropriately, because there is the direct dependency between "fire" and "Haga". However, an association approach would make an error in deciding the sense, because "Police" and "force" tend to be a noise, from the point of view of an unordered set of words. Generally, an association does not use a syntactic dependency, and a selectional restriction uses only a part of words appeared in a sentence.

In this paper, we present a new method for WSD, which uses syntactic dependencies for a whole sentence as a clue. They contain both of all words included in a sentence and all syntactic dependencies in it. Our method is based on a technique of pairwise alignment, and described in the following two sections. Using our method, we have gotten appropriate sense for various cases including above examples. In section 4, we describe our experimental result for WSD on some verbs in SENSEVAL-1 (Kilgarriff, 1998).

2 Our Method

Our method has the features on an association and a selectional restriction approach both. It can be applied with the various sentence types because our method can treat a local (direct) and a whole sentence dependency. Our method is based on the following steps;

Step 1. Parse the input sentence with syntactic parser[2], and find all paths from root to leaves in the resulting dependency tree.

Step 2. Compare the paths from **Step 1.** with prototype paths prepared for each sense of the target word.

Step 3. Find a summation of similarity between each prototype and input path for each sense.

Step 4. Select the sense with the maximum value of the summation.

We describe our method in detail in the followings.

In our method, we consider paths from root to leaves in a dependency tree. For example, consider the sentence "we consider a path in a graph". This sentence has three leaves in the dependency structure, and consequently has three paths from root to leaves; (consider, SUB, we), (consider, OBJ, path, a) and (consider, OBJ, path, in, graph, a). "SUB" and "OBJ" in the paths are the elements added automatically using some rules in order to make a remarkable difference between verb-subject and verb-object. We think this sequence structure of word would serve as a clue to WSD very well, and we regard a set of the sequences obtained from an input sentence as the context of a target word.

The general intuition for WSD is that words with similar context have the same sense (Charniak, 1993; Lin, 1997). That is, once we prepare the prototype sequences for each sense, we can determine the sense of the target word as one with the most similar prototype set. We measure a similarity between a set of prototype sequences T and a set of sequences from input sentence T'. Let T and T' have a set of sequences, $P_T = \{p_1, p_2, \cdots, p_n\}$ and

[2]We assume that we can get the correct syntactic structure here. (See section 4)

fire: go off or discharge
fire, SUB, person
fire, OBJ, [weapon, rocket]
fire, [on, upon, at], physical_object
fire, *, load, [into, with], weapon
fire, *, set_up, OBJ, weapon

fire: terminate the employment
fire, SUB, company
fire, OBJ, [person, people, staff]
fire, from, organization
fire, *, hire
fire, *, job

Figure 1: Prototype sequence for verb "fire"

$P_{T'} = \{p'_1, p'_2, \cdots, p'_m\}$ respectively. p_i and p'_j are sequences of words. We define the similarity between T and T', $sim(T, T')$, as following:

$$sim(T,T') = \sum_{p_i \in P_T} f_i \max_{p'_j \in P_{T'}} alignment(p_i, p'_j) \quad (1)$$

$sim(T,T')$ is not commutative. That is, $sim(T,T') \neq sim(T',T)$. $alignment(p_i, p'_j)$ is an alignment score between the sequences p_i and p'_j, defined in the next section. f_i is a weight function characteristic of the sequence p_i, defined as following:

$$f_i = \begin{cases} u_i & \text{if } \max\limits_{p'_j \in P_{T'}} alignment(p_i, p'_j) \geq t_i \\ v_i & \text{otherwise} \end{cases} \quad (2)$$

where u_i and v_i are arbitrary constants and t_i is arbitrary threshold.

Using equation (1), we can estimate a similarity between the context of a target word and prototype context, and can determine the sense of a target word by selecting the prototype with the maximum similarity.

An example of the prototype sequences for verb "fire" is shown in Figure 1. A prototype sequence is represented like a regular expression. For the present, we obtain the sequence by hand. The basic policy to obtain prototypes is to observe the common features on dependency trees in which target word is used in the same sense. We have some ideas about a method to obtain prototypes automatically.

3 Pairwise Alignment

We attempt to apply the method of pairwise alignment to measuring the similarity between sequences. Recently, the technique of pairwise alignment is

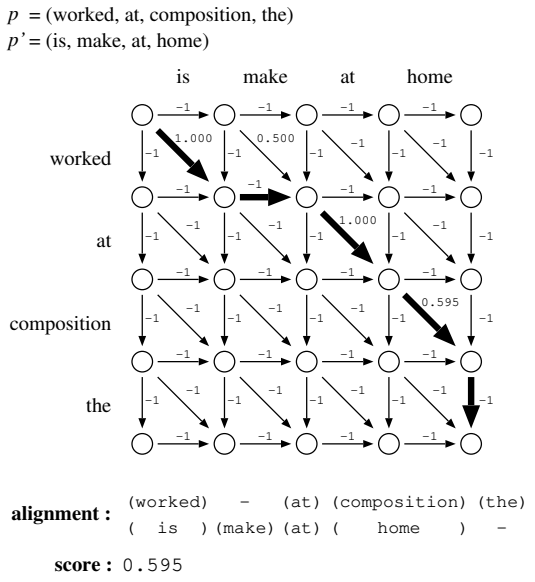

Figure 2: Pairwise alignment

used generally in molecular biology research as a basic method to measure the similarity between proteins or DNA sequences (Mitaku and Kanehisa, 1995).

There have been several ways to find the pairwise alignment, such as the method based on Dynamic Programming, one based on Finite State Automaton, and so on (Durbinet al., 1998). In our method, we apply the method using DP matrix, as in Figure 2. We have shown the pairwise alignment between sequences $p =$ (worked, at, composition, the) and $p' =$ (is, make, at, home) as an example.

In a matrix, a vertical and horizontal transition means a gap and is assigned a gap score. A diagonal transition means a substitution and is assigned a score based on the similarity between two words corresponding to that point in the matrix. Actually, the following value is calculated in each node, using values which have been calculated in its three previous nodes.

$$F_{i,j} = \max \begin{cases} F_{i-1,j} + subst(-, w'_j) \\ F_{i,j-1} + subst(w_i, -) \\ F_{i-1,j-1} + subst(w_i, w'_j) \end{cases} \quad (3)$$

where $subst(-, w'_j)$ and $subst(w_i, -)$ represent respectively to substitute w'_j and w_i with a gap $(-)$, and return the gap score. $subst(w_i, w'_j)$ represent the score of substituting w_i with w'_j or vice versa.

Now let the word w has synsets $s_1, s_2, \cdots, s_k$ and w' has $s'_1, s'_2, \cdots, s'_l$ on WordNet hierarchy (Miller et al., 1990). For simplicity, we define the $subst(w, w')$ as following, based on the semantic distance (Stetina and Nagao, 1998).

$$subst(w, w') = 2 \cdot \max_{i,j}(sd(s_i, s'_j)) - 1 \quad (4)$$

where $sd(s_i, s'_j)$ is the semantic distance between two synsets s_i and s'_j. Because $0 \leq sd(s_i, s'_j) \leq 1$, $-1 \leq subst(w, w') \leq 1$. The score of the substitution between identical words is 1, and one between two words with no common ancestor in the hierarchy is -1. We simply define the gap score as -1.

4 Experimental Result

Up to the present, we have obtained the experimental results on 7 verbs in SENSEVAL-1[3]. In our experiment, for all sentences including target word in the training and test corpus of SENSEVAL-1, we make a parsing using Apple Pie Parser (Sekine, 1996) and additional vertices using some rules automatically. If the resulted parsing includes some errors, we remove them by hand. Then we obtain the sequence patterns by hand from training data and attempt WSD using equation (1) for test data. Because of various length of sequence, we assign score zero to the preceding and right-end gaps in an alignment.

We show our experimental results in Table 1. In SENSEVAL-1, precisions and recalls are calculated by three scoring ways, fine-grained, mixed-grained and coarse-grained scoring. We show the results only by fine-grained scoring which is evaluated by distinguishing word sense in the strictest way. It is impossible to make simple comparison with the participants in SENSEVAL-1 because our method needs supervised learning by hand. However, 2.8%-14.2% improvements of the accuracy compared with the best system seems significant, suggesting that our method is promising for WSD.

5 Future Works

There are two major limitations in our method; one of syntactic information and of knowledge acquisi-

[3]We have experimented on verbs in SENSEVAL-1 one by one alphabetically. The word "amaze" is omitted because it has only one verbal sense.

Table 1: Experimental results for some verbs (in fine-grained scoring)

bet the numbers of test instance:117

	precision (recall)
our method	0.880 (0.880)
best system in SENSEVAL-1	0.778 (0.778)
human	0.924 (0.916)

bother the numbers of test instance:209

	precision (recall)
our method	0.900 (0.900)
best system in SENSEVAL-1	0.866 (0.866)
human	0.976 (0.976)

bury the numbers of test instance:201

	precision (recall)
our method	0.667 (0.667)
best system in SENSEVAL-1	0.572 (0.572)
human	0.928 (0.923)

calculate the numbers of test instance:218

	precision (recall)
our method	0.950 (0.950)
best system in SENSEVAL-1	0.922 (0.922)
human	0.954 (0.950)

consume the numbers of test instance:186

	precision (recall)
our method	0.645 (0.645)
best system in SENSEVAL-1	0.503 (0.500)
human	0.944 (0.939)

derive the numbers of test instance:217

	precision (recall)
our method	0.751 (0.751)
best system in SENSEVAL-1	0.664 (0.664)
human	0.965 (0.961)

float the numbers of test instance:229

	precision (recall)
our method	0.616 (0.616)
best system in SENSEVAL-1	0.555 (0.555)
human	0.927 (0.923)

tion by hand.

The former is that our method assumes we can get the correct syntactic information. In fact, the accuracy and performance of syntactic analyzer are being improved more and more, consequently this disadvantage would become a minor problem. Because a similarity between sequences derived from syntactic dependencies is calculated as a numerical value, our method would also be suitable for integration with a probabilistic syntactic analyzer.

The latter, which is more serious, is that the sequence patterns used as clue to WSD are acquired by hand at the present. In molecular biology research, several attempts to obtain sequence patterns automatically have been reported, which can be expected to motivate ours for WSD. We plan to construct an algorithm for an automatic pattern acquisition from large scale corpora based on those biological approaches.

References

Eugene Charniak. 1993. Statistical Language Learning. MIT Press, Cambridge.

Richard Durbin, Sean R. Eddy, Andrew Krogh and Graeme Mitchison. 1998. Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids. Cambridge University Press.

Marti A. Hearst. 1991. Noun Homograph Disambiguation Using Local Context in Large Text Corpora. In *Proceedings of the 7th Annual Conference of the University of Waterloo Center for the New OED and Text Research*, pp.1-22.

Nancy Ide and Jean Véronis. 1998. Introduction to the Special Issue on Word Sense Disambiguation: The State of the Art. *Computational Linguistics*, 24(1):1-40.

Adam Kilgarriff. 1998. Senseval: An exercise in evaluating word sense disambiguation programs. In *Proceedings of the 1st International Conference on Language Resources and Evaluation (LREC98)*, volume 1, pp.581-585.

Dekang Lin. 1997. Using Syntactic Dependency as Local Context to Resolve Word Sense Ambiguity. In *Proceedings of ACL/EACL-97*, pp.64-71.

Christopher D. Manning and Hinrich Schütze. 1999. Foundations of Statistical Natural Language Processing. MIT Press, Cambridge.

George A. Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross and Katherine J. Miller. 1990. Introduction to WordNet: an on-line lexical database. In *International Journal of Lexicography*, 3(4):235-244.

Shigeki Mitaku and Minoru Kanehisa (ed). 1995. Human Genom Project and Knowledge Information Processing (in Japanese). Baifukan.

Satoshi Sekine. 1996. Manual of Apple Pie Parser. URL: `http://nlp.cs.nyu.edu/app/`.

Jiri Stetina and Makoto Nagao. 1998. General Word Sense Disambiguation Method Based on a Full Sentential Context. In *Journal of Natural Language Processing*, 5(2):47-74.