SENSEVAL-3: Third International Workshop on the Evaluation of Systems
for the Semantic Analysis of Text, Barcelona, Spain, July 2004
Association for Computational Linguistics

KUNLP System in SENSEVAL-3

Hee-Cheol Seo, Hae-Chang Rim
Dept. of Computer Science
and Engineering,
Korea University
1, 5-ka, Anam-dong, Seongbuk-Gu,
Seoul, 136-701, Korea
{hcseo, rim}@nlp.korea.ac.kr

Soo-Hong Kim
Dept. of Computer Software Engineering,
College of Engineering,
Sangmyung University,
San 98-20, Anso-Dong,
Chonan, Chungnam, Korea
soohkim@smuc.ac.kr

Abstract

We have participated in both English all words task and English lexical sample task of SENSEVAL-3. Our system disambiguates senses of a target word in a context by selecting a substituent among WordNet relatives of the target word, such as synonyms, hypernyms, meronyms and so on. The decision is made based on co-occurrence frequency between candidate relatives and each of the context words. Since the co-occurrence frequency is obtainable from raw corpus, our method is considered to be an unsupervised learning algorithm that does not require a sense-tagged corpus.

1 Introduction

At SENSEVAL-3, we adopted an unsupervised approach based on WordNet and raw corpus, which does not require any sense tagged corpus. WordNet specifies relationships among the meanings of words.

Relatives of a word in WordNet are defined as words that have a relationship with it, e.g. they are synonyms, antonyms, superordinates (hypernyms), or subordinates (hyponyms). Relatives, especially those in a synonym class, usually have related meanings and tend to share similar contexts. Hence, some WordNet-based approaches extract relatives of each sense of a polysemous word from WordNet, collect example sentences of the relatives from a raw corpus, and learn the senses from the example sentences for WSD. Yarowsky (1992) first proposed this approach, but used International Roget's Thesaurus as a hierarchical lexical database instead of WordNet. However, the approach seems to suffer from examples irrelevant to the senses of a polysemous word since many of the relatives are polysemous. Leacock et al. (1998) attempted to exclude irrelevant or spurious examples by using only monosemous relatives in WordNet. However, some senses do not have short distance monosemous relatives through a relation such as synonym, child, and parent. A possible alternative of using only monosemous relatives in the long distance, however, is problematic because the longer the distance of two synsets in WordNet, the weaker the relationship between them. In other words, the monosemous relatives in the long distance may provide irrelevant examples for WSD.

Our approach is somewhat similar to the WordNet based approach of Leacock et al. (1998) in that it acquires relatives of a target word from WordNet and extracts co-occurrence frequencies of the relatives from a raw corpus, but our system uses polysemous as well as monosemous relatives. To avoid a negative effect of polysemous relatives on the co-occurrence frequency calculation, our system handles the example sentences of each relative separately instead of putting together the example sentences of all relatives into a pool. Also we devised our system to efficiently disambiguate senses of all words using only co-occurrence frequency between words.

2 KUNLP system

2.1 Word Sense Disambiguation

We disambiguate senses of a word in a context[1] by selecting a substituent word from the WordNet[2] relatives of the target word. Figure 1 represents a flowchart of the proposed approach. Given a target word and its context, a set of relatives of the target word is created by searches in WordNet. Next, the most appropriate relative that can be substituted for the word in the context is chosen. In this step, co-occurrence frequency is used. Finally, the sense of the target word that is related to the selected relative is determined.

The example in Figure 2 illustrates how the proposed approach disambiguates senses of the target word *chair* given the context. The set of relatives {*president*, *professorship*, ...} of *chair* is built by WordNet searches, and the probability,

[1] In this paper, a context indicates a target word and six words surrounding the target word in an instance.

[2] The WordNet version is 1.7.1.

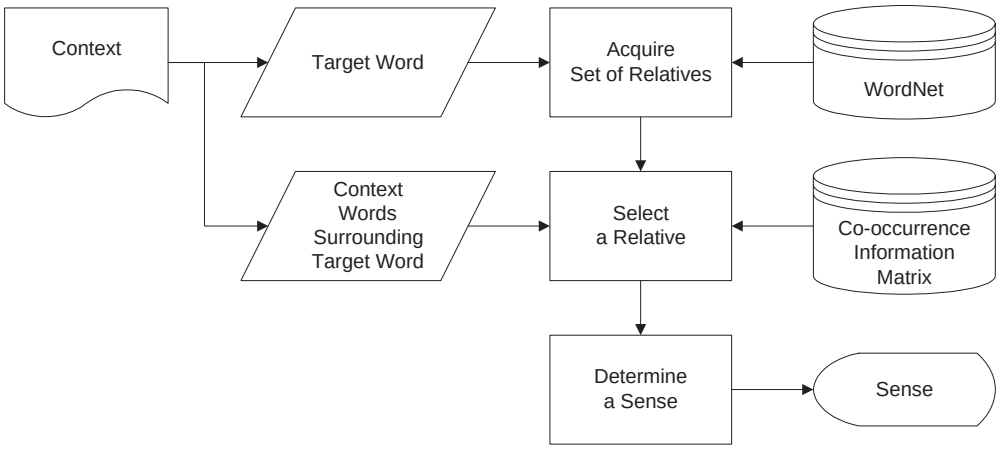

Figure 1: Flowchart of KUNLP System

"$Pr(professorship|Context)$," that a relative can be substituted for the target word in the given context is estimated by the co-occurrence frequency between the relative and each of the context words. In this example, the relative, *seat*, is selected with the highest probability and the proper sense, "*a seat for one person, with a support for the back*," is chosen.

Thus, the second step of our system (i.e. selecting a relative) has to be carefully implemented to select the proper relative that can substitute for the target word in the context, while the first step (i.e. acquiring the set of relatives) and the third step (i.e. determining a sense) are done simply through searches in WordNet.

The substituent word of the i-th target word tw_i in a context C is defined to be the relative of tw_i which has the largest co-occurrence probability with the words in the context:

$$SW(tw_i, C) \stackrel{\text{def}}{=} \arg\max_{r_{ij}} P(r_{ij}^{\alpha}|C) \qquad (1)$$

where SW is the substituent word, r_{ij} is the j-th relative of tw_i, and r_{ij}^{α} is the α-th sense related to tw_i[3]. If α is 2, the 2-nd sense of r_{ij} is related to tw_i. The right hand side of Equation 1 is calculated with logarithm as follows:

$$\begin{aligned}
&\arg\max_{r_{ij}} P(r_{ij}^{\alpha}|C) \\
&= \arg\max_{r_{ij}} \frac{P(C|r_{ij}^{\alpha})P(r_{ij}^{\alpha})}{P(C)} \\
&= \arg\max_{r_{ij}} P(C|r_{ij}^{\alpha})P(r_{ij}^{\alpha}) \\
&= \arg\max_{r_{ij}} \{logP(C|r_{ij}^{\alpha}) + logP(r_{ij}^{\alpha})\} \qquad (2)
\end{aligned}$$

[3] α is a function with two parameters tw_i and r_{ij}, but it can be written in brief without parameters.

Instance :
He should sit in the chair beside the desk.

Target Word :
'chair'

Context :
sit in the chair beside the desk

Set of Relatives :
{professorship, president, chairman,
electronic chair, death chair, seat,
office, presiding officer, ...}

Probability of Relative given the Context :
P(professorship | Context)
P(president | Context)
...
P(seat | Context)
...

Selected Relative :
'seat' - it is the most likely word occurred
from the above context among
the relatives of 'chair'

Determined Sense :
chair%1:06:00 - "a seat for one person,
with a support for the back."
'seat' - the hypernym of chair%1:06:00.

Figure 2: Example of sense disambiguation procedure for *chair*

Then Equation 2 may be calculated under the assumption that words in C occur independently:

$$\begin{aligned}&\arg\max_{r_{ij}}\{logP(C|r_{ij}^{\alpha}) + logP(r_{ij}^{\alpha})\}\\ &\approx \arg\max_{r_{ij}}\{\sum_{k=1}^{n} logP(w_k|r_{ij}^{\alpha}) + logP(r_{ij}^{\alpha})\} \quad (3)\end{aligned}$$

where w_k is the k-th word in C and n is the number of words in C. In Equation 3, we assume independence among words in C.

The first probability in Equation 3 is calculated as follows:

$$\begin{aligned}&P(w_k|r_{ij}^{\alpha})\\ &\approx P(w_k|r_{ij})\\ &= \frac{P(r_{ij}|w_k)P(w_k)}{P(r_{ij})} \quad (4)\end{aligned}$$

The second probability in Equation 3 is computed as follows:

$$P(r_{ij}^{\alpha}) = \beta(r_{ij}^{\alpha})P(r_{ij}) \quad (5)$$

where $\beta(r_{ij}^{\alpha})$ is the ratio of the frequency of r_{ij}^{α} to that of r_{ij}:

$$\beta(r_{ij}^{\alpha}) = \frac{WNf(r_{ij}^{\alpha}) + 0.5}{n * 0.5 + WNf(r_{ij})}$$

where $WNf(r_{ij}^{\alpha})$ is the frequency of r_{ij}^{α} in WordNet, $WNf(r_{ij})$ is the frequency of r_{ij} in WordNet, 0.5 is a smoothing factor, and n is the number of senses of r_{ij}.

Applying Equations 4 and 5 to Equation 3, we have the following equation for acquiring the relative with the largest co-occurrence probability:

$$\begin{aligned}&\arg\max_{r_{ij}} P(r_{ij}^{\alpha}|C)\\ &\approx \arg\max_{r_{ij}} \sum_{k=1}^{n} log\frac{P(r_{ij}|w_k)P(w_k)}{P(r_{ij})}\\ &\quad + log\beta(r_{ij}^{\alpha})P(r_{ij})\\ &= \arg\max_{r_{ij}} \sum_{k=1}^{n} log\frac{P(r_{ij}|w_k)}{P(r_{ij})} + log\beta(r_{ij}^{\alpha})P(r_{ij})\end{aligned}$$

In the case that several relatives have the largest co-occurrence probability, all senses related to the relatives are determined as proper senses.

2.2 Co-occurrence Frequency Matrix

In order to select a substituent word for a target word in a given context, we must calculate the probabilities of finding relatives, given the context. These probabilities can be estimated based on the co-occurrence frequency between a relative and context words as follows:

$$P(r_{ij}) = \frac{freq(r_{ij})}{CS} \quad (6)$$

$$\begin{aligned}P(r_{ij}|w_k) &= \frac{P(r_{ij}, w_k)}{P(w_k)}\\ &= \frac{freq(r_{ij}, w_k)}{freq(w_k)} \quad (7)\end{aligned}$$

where $freq(r_{ij})$ is the frequency of r_{ij}, CS is the corpus size, $P(r_{ij}, w_k)$ is the probability that r_{ij} and w_k co-occur, and $freq(r_{ij}, w_k)$ is the frequency that r_{ij} and w_k co-occur.

In order to calculate these probabilities, frequencies of words and word pairs are required. For this, we build a co-occurrence frequency matrix that contains co-occurrence frequencies of words pairs. In this matrix, an element m_{ij} represents the frequency that the i-th word and j-th word in the vocabulary co-occur in a corpus[4]. The frequency of a word can be calculated by counting all frequencies in the same row or column. The vocabulary is composed of all content words in the corpus. Now, the equations 6 and 7 can be calculated with the matrix.

The matrix is easily built by counting each word pair in a given corpus. It is not necessary to make an individual matrix for each polysemous word, since the matrix contains co-occurrence frequencies of all word pairs. Hence, it is possible to disambiguate all words with only one matrix. In other words, the proposed method disambiguates the senses of all words efficiently with only one matrix.

2.3 WordNet Relatives

Our system used most of relationship types in WordNet, except sister and attribute types, to acquire the relatives of target words. For a nominal word, we included all hypernyms and hyponyms in distance 3 from a sense, which indicate parents, grandparents and great-grand parents for hypernymy and children, grandchildren and great-children for hyponymy[5].

In order to identify part-of-speech (POS) of words including target words in instances, our system uses TreeTagger (Schmid, 1994). After POS

[4]The co-occurrence frequency matrix is a symmetric matrix, thus m_{ij} is the same as m_{ji}.

[5]We implemented WordNet APIs with index files and data files in WordNet package, which is downloadable from http://www.cogsci.princeton.edu/ wn/.

	fine grained		coarse grained	
	recall	prec.	recall	prec.
noun	0.451	0.451	0.556	0.556
verb(R)	0.354	0.354	0.496	0.496
adjective	0.497	0.497	0.610	0.610
overall	0.404	0.404	0.528	0.528

Table 1: Official Results : English Lexical Sample

	with_U		without_U	
	recall	prec.	recall	prec.
overall	0.500	0.500	0.496	0.510

Table 2: Official Results (fine grained) : English All Words

of the target word is determined, relationship types related to the POS are considered to acquire the candidate relatives of the target word. For instance, if a target word is adverb, the following relationships of the word are considered: synonymy, antonymy, and derived.

2.4 WordNet Multiword Expression

Our system recognizes multiword expressions of WordNet in an instance by a simple string match before disambiguating senses of a target word. If the instance has a multiword expression including the target word, our system does not disambiguate the senses of the multiword expression but just assigns all senses of the multiword expression to the instance.

3 Official Results

We have participated in both English lexical sample task and English all words task. Table 1 and 2 show the official results of our system for two tasks. Our system disambiguates all instances, thus the coverage of our system is 100% and precision of our system is the same as the recall.

Our system assigns WordNet sense key to each instance, but verbs in English lexical sample task are annotated based on Wordsmyth definitions. In official submission, we did not map the WordNet sense keys of verbs to Wordsmyth senses, thus the recall of our system for verbs is 0%. Table 1 shows the results after a mapping between Wordsmyth and WordNet verb senses using the file EnglishLS.dictionary.mapping.xml.

In English all word task, there are two additional scoring measures in addition to fine- and coarse-grained scoring: *with_U* and *without_U*[6]. In *with_U*, any instance without a WN sensekey is assumed to be tagged with a 'U' and thus is tagged as correct if the answer file (i.e. answer.key) has a 'U', incorrect otherwise. In *without_U*, any instance without a WN sensekey is assumed to have been skipped, thus precision will not be affected, but recall will be lowered.

4 Conclusions

In SENSEVAL-3, we participated in both English all words task and English lexical sample task with an unsupervised system based on WordNet and a raw corpus, which did not use any sense tagged corpus. Our system disambiguated the senses of a target word by selecting a substituent among WordNet relatives of the target word, which frequently co-occurs with each word surrounding the target word in a context. Since each relative is usually related to only one sense of the target word, our system identifies the proper sense with the selected relative. The substituent word is selected based on the co-occurrence frequency between the relative and the words surrounding the target word in a given context. We collected the co-occurrence frequency from a raw corpus, not a sense-tagged one that is often required by other approaches. In short, our system disambiguates senses of words only through the set of WordNet relatives of the target words and a raw corpus. The system was simple but seemed to achieve a good performance when considered the performance of systems in last SENSEVAL-2 English tasks.

For future research, we will investigate the dependency between the types of relatives and the characteristics of words or senses in order to devise an improved method that better utilizes various types of relatives for WSD.

References

Claudia Leacock, Martin Chodorow, and George A. Miller. 1998. Using corpus statistics and WordNet relations for sense identification. *Computational Linguistics*, 24(1):147–165.

Helmut Schmid. 1994. Probabilistic part-of-speech tagging using decision trees. In *Proceedings of International Conference on New Methods in Language Processing*, Manchester,U.K.

David Yarowsky. 1992. Word-sense disambiguation using statistical models of Roget's categories trained on large corpora. In *Proceedings of COLING-92*, pages 454–460, Nantes, France, July.

[6]These measures are described in Benjamin Synder's mail about English all words task results