Word Sense Disambiguation Using Semantic Sets based on WordNet

Ganesh Ramakrishnan

Computer Sc. & Engg.
Indian Institue of Technology
Mumbai - 400076

Pushpak Bhattacharyya

Computer Sc. & Engg.
Indian Institue of Technology
Mumbai - 400076

Abstract

This paper presents an automatic method for resolving the lexical ambiguity of nouns in any free-flowing text. The method exploits the noun taxonomy present in the WordNet and also the relative position of nouns in the given text, to construct semantic sets from the text. The semantic set has been de

ned as a collection of senses of words in given text that are related through the WordNet. Two different concepts of semantic distance between words have been explored and used for disambiguation. Hand-tagging of text and training are not required by the method presented in this paper. The method has been tested against SemCor, the tagged version of the Brown corpus and compared with previous unsupervised WSD algorithms. The method is supported by good empirical results.

1 Introduction

Any language uses words with multiple meanings. Before Information Retrieval or Semantic analysis of texts, it is essential to determine the true senses of those words. The problem of determining the right sense of words, in a context, is called Word Sense Disambiguation (WSD).

The typical approaches to the problem of WSD can be classified into 3 types: *(1)Supervised, (2)Unsupervised and (3)Cross-Lingual*.

Supervised Methods require resources like semantically annotated corpora to train the WSD system, and lexical resource like WordNet which provides the sense numbers using which the annotations are made. These algorithms, like the ones considered in [1], [2] and [3] use the corpora like Grolier's encyclopedia [1] or private sense-tagged data-sets [2]. However, the semantically annotated corpora are Laborious to construct and expensive, since tagging is done manually or at most semi-automatically.

Unsupervised Methods consider the statistically relevant co-occurance of individual keywords as classes and generate a class based model to predict which will be the most likely class to follow a particular keyword. The class is treated as an equivalent of sense. Unsupervised WSD methods can be further classifed into two types, viz. WSD that makes use of the information provided by machine readable dictionaries: this is the case with the work reported by [10], [14], [4], [12] and [11]. And WSD that uses information gathered from raw corpora (unsupervised training methods); [13] presentes an unsupervised WSD method using raw corpora. From a multilingual point of view, word sense disambiguation is nothing more than determining the appropriate translation of a word or lexical item. Thus, translation presupposes word sense disambiguation. Word translation only requires only that the words should be expressing the same meaning. However, it is not necessary to know the exact meaning of the words. See [7] for further details.

2 WordNet

WordNet[9] is an online lexical reference system whose design is inspired by current psycholinguistic theories of human lexical memory. English nouns, verbs, adjectives and adverbs are organized into synonym sets, each representing one underlying lexical concept. Different relations link the synonym sets. WordNet was developed by the Cognitive Science Laboratory at Princeton University. The WordNet consists of synsets arranged in semantic relationships with one another, through hypernymy, hyponymy, holonymy, meronymy, synonymy and antonymy relationships. In our discussion, we use WordNet as the only lexical resource and all the senses are with respect to the WordNet.

3 Semantic Set

Below is a sample text of 100 words, from the Brown Corpus, with some nouns underlined. In the WordNet sub-graph in figure 1, the relationship between these nouns is shown. The words marked in ellipses are words that actually occur in the text.

The Fulton_County_Grand_Jury said Friday an investigation of Atlanta 's recent primary_election produced no *evidence* that any irregularities took_place. The jury further said in *term* end presentments that the City_Executive_Committee which had over-all charge of the election deserves the praise and thanks of the City_of_Atlanta for the manner in which the election was conducted The September-October *term* jury had been charged by Fulton Superior_Court_Judge_Durwood_Pye to investigate reports of possible irregularities in the hard-fought primary which was won by Mayor-nominate_Ivan_Allen_Jr ..... It recommended that Fulton legislators act to have these laws studied and revised to the end of modernizing and improving them. The grand_jury commented on a number of other *topics* among them the Atlanta and Fulton_County purchasing_departments which it said are well operated and follow generally accepted practices which inure to the best interest of both governments However the jury said it believes these two offices should be combined to achieve greater efficiency and reduce the cost of administration ... Implementation of Georgia 's automobile title *law* was also recommended by the outgoing jury It urged that the next Legislature provide enabling funds and re-set the effective date so_that an orderly implementation of the *law* may be effected. ... This is one of the major items in the Fulton_County general assistance *program* the jury said but the State_Welfare_Department has seen_fit to distribute these funds through the welfare departments of all the counties in the state with the exception of Fulton_County which receives none of_this money The jurors said they realize a proportionate distribution of these funds might disable this *program* in our less populous counties. The jurors said Failure to do this will continue to place a disproportionate *burden* on Fulton taxpayers.

Figure 1: Sample text from SemCor, br-a01 with the word *program* (word number 93) in consideration

The number in the brackets, by the side of the word, is its WordNet sense number. The numbers mentioned in the square brackets are the textual positions. For example, the word law appears in textual positions 66 and 72. The arrows going up-down show the hyponymy relations. Thus, 2 hyponyms of sense number 1 of cognition are shown.

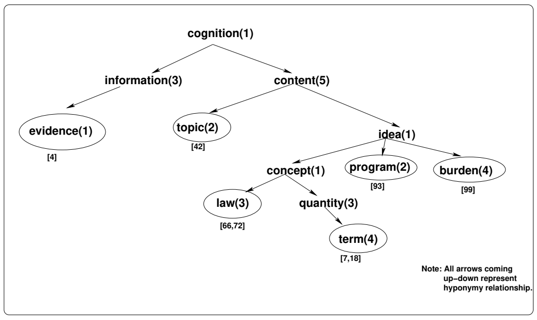

Figure 2: An extract of the WordNet graph, corresponding to the nouns underlined in figure 1

In the same way, one can consider the *holonymy*, *meronymy*, *synonymy* and *antonymy* relationships from the WordNet to capture all the nouns in a given piece of text. Consider the resultant WordNet subgraph. Also, suppose that distances are measured over edges, with every edge of unit distance and the distances are additive. Consider all the words that occur in the graph, within a distance of 4 from the 2^{nd} sense of the word *program*. We call the set of word-senses, within a fixed distance from the chosen synset as the *semantic set* corresponding to that synset. Fig. 3 is an example. The notations and the definitions are given in section 4.

In the same way, one can consider the holonymy, meronymy, synonymy and antonymy relations from the WordNet to capture all the nouns in a givenpiece of text. Consider the resultant WordNet subgraph. Also, suppose that distances are measured over edges, with every edge of unit distance and the distances are additive. Consider all the words that occur in the graph, within a distance of 4 from the 2^{nd} sense of the word program. We call the set of word-senses, within a fixed distance from the chosen synset as the *semantic set* corresponding to that synset.

Fig. 3 is an example. The notations and the definitions are given in section 4.

program<80,2,<0,0,0,0,0,0>>, portion<95,1,<0,3,1,0,1,0>>, policy<58,2,<0,2,0,0,0,0>>, term<7,4,<0,3,1,0,0,0>>, topic<42,2,<0,1,2,0,0,0>>, law<66,1,<0,3,1,0,0,0>>, end<8,3,<0,4,0,1,0,0>>, term<18,4,<0,3,1,0,0,0>>, end<8,4,<0,1,0,1,0,1>>, practice<45,5,<0,1,3,0,0,0>>, manner<14,3,<0,4,1,0,0,0>>, law<66,3,<0,2,1,0,0,0>>, end<39,3,<0,4,0,1,0,0>>, end<39,4,<0,1,0,1,0,1>>, law<72,1,<0,3,1,0,0,0>>, burden<99,4,<0,1,1,0,0,0>>, city<59,1,<0,2,3,0,0,0>>, law<72,3,<0,2,1,0,0,0>>, city<31,1,<0,2,3,0,0,0>>, city<56,1,<0,2,3,0,0,0>>, evidence<4,1,<0,2,3,0,0,0>>

Figure 3: Semantic set corresponding to sense number 2 of the word *program*

4 Terminology

We want to find the correct senses of the words in a text T. Let W be a window in T having n nouns, $(w_1,\ w_2,\ w_3,\ .....w_n)$. For every w_i its s_i senses are $\sigma_{i_1}, ...\sigma_{i_{s_i}}$. Let P_i be the position of w_i in the text.

Semantic Graph

Let G be that minimal sub-graph of the WordNet, which includes all the noun-senses σ_{i_k}, $1 \leq k \leq s_i$ and $1 \leq i \leq n$, from T. We call G, the *Semantic Graph* for the text T.

Let σ_{i_p} and σ_{j_q} be two noun-senses in the subgraph G. Consider the shortest path from σ_{i_p} to σ_{j_q}. Let $\eta_1(\sigma_{i_p}, \sigma_{j_q}$, $\eta_2(\sigma_{i_p}, \sigma_{j_q})$, $\eta_3(\sigma_{i_p}, \sigma_{j_q})$, $\eta_4(\sigma_{i_p}, \sigma_{j_q})$, $\eta_5(\sigma_{i_p}, \sigma_{j_q})$ and $\eta_6(\sigma_{i_p}, \sigma_{j_{q6}})$ respectively be the number of *hyponymy*, *hypernymy*, *meronymy*, *holonymy*, *synonymy* and *antonymy* arcs on this path.

Semantic vector

The *semantic vector* between two noun senses σ_{i_p} and σ_{j_q} in the graph G is the sequence

$$(\eta_1(\sigma_{i_p}, \sigma_{j_q})), ...(\eta_6(\sigma_{i_p}, \sigma_{j_q})) \tag{1}$$

where the $\eta_i(\sigma_i, \sigma_{j_q})$'s are as given in previous definition. We denote the semantic vector by $N(\sigma_{i_p}, \sigma_{j_q})$.

Semantic distance

The concept of *semantic distance* has been explored in [15]. Broadly, two concepts of *semantic distance* have been mentioned there. They are *semantic similarity* and *semantic relatedness*. In this paper, we talk of semantic relatedness, as explored in [16]. But the measures of *semantic distance* that we adopt are little variants of what has been proposed by [16]. The first measure of the *semantic distance* of a noun-sense σ_{j_q} from σ_{i_p} in G corresponds to the minimum number of arcs that must be traversed in order to reach σ_{j_q} from σ_{i_p}.

From the fact that *hypernymy*, *hyponymy* and *meronymy*, *holonymy* are complementary, and that *synonymy* and *antonymy* are *symmetric*, it follows that the semantic distance is *commutative*.

The second measure of *semantic distance* will be given in section 5.2.

Semantic form

Recall the definition that P_i is the position of w_i in the text. The expression, $w_j < P_j, q, < N(\sigma_{i_p}, \sigma_{j_q}) >>$ is called the *semantic form* for σ_{j_q} *with respect to* σ_{i_p}. We will denote it by $F(\sigma_{i_p}, \sigma_{j_q})$.

Semantic set

Consider every noun-sense σ_{j_q} in G, within a maximum semantic distance of R from σ_{i_p}. The collection of all the semantic forms $F(\sigma_{i_p}, \sigma_{j_q})$ is called the *semantic set* S_{i_p} for σ_{i_p}, with radius R. σ_{i_p} is called the *reference noun-sense* for S_{i_p}.

A semantic set S_{i_p} is of the form given below:

$$S_{i_p} = F(\sigma_{i_p}, \sigma_{i_p}), F(\sigma_{i_p}, \sigma_{i_{1_{p_1}}})....F(\sigma_{i_p}, \sigma_{i_{k_{p_k}}}) \quad (2)$$

where k is the length of the semantic set. For word w_i we have s_i semantic sets S_{i_1}, S_{i_2} , ... $S_{i_{s_i}}$. Also, for the word sense σ_{i_p} we define the position vector $\overline{P_{i_p}}$ and $\overline{N_{i_p}}$ as in equations 3 and 4.

$$\overline{P_{i_p}} = < P_{i_1}...P_{i_k} > \quad (3)$$

$$\overline{N_{i_p}} = < N(\sigma_{i_p}, \sigma_{i_{1_{p_1}}}), ...N(\sigma_{i_p}, \sigma_{i_{k_{p_k}}}) > \quad (4)$$

An Example

Consider again the figure 1 which shows a sample from the text br-a01 of SemCor. The wordsenses of the underlined nouns in the text, form a semantic graph, part of which has been depicted in figure 2. For instance consider the word *program* which has position number 93 in br-a01 and *burden* which has position number 93 in br-a01. In figure 2, it is shown that sense number 4 of *burden* and sense number 2 of *program* have the same hypernym - the sense number 1 of *idea*. Thus, the semantic distance between *program(2)* and *burden(4)* is 2. The semantic vector from *program(2)* to *burden(4)*, keeping *program(2)* as the reference word is $(\eta_1(\sigma_{93_2}, \sigma_{99_4}))$,.. $(\eta_6(\sigma_{93_2}, \sigma_{99_4}))$ = (0), (1), (1), (0), (0), (0)

The distances traversed along the different relation arcs, in the figure 2 from *program(2)* to *burden(4)* are as given in the table 1.

The semantic form $F(\sigma_{93_2}, \sigma_{99_4})$ is given as *burden(99,4,0,1,1,0,0,0)*. σ_{99_4} is within a semantic distance of 4 from σ_{93_2}. The collection of all $F(\sigma_{93_2}, \sigma_{j_q}), 1 \leq q \leq s_j \forall$ words $\sigma_j, j \neq i$ in the text T such that, σ_{j_q} is within a semantic distance 4 from σ_{93_2} is called the semantic set for σ_{93_2}, $S(\sigma_{93_2})$. This semantic set is given in figure 3.

Table 1: The distance along the different relation arcs, between *program(2)* and *burden(4)* as depicted in 2

Relation	Notation for dist.	Distance
hyponymy	$\eta_1(\sigma_{93_2}, \sigma_{99_4})$	0
hypenymy	$\eta_2(\sigma_{93_2}, \sigma_{99_4})$	1
meronymy	$\eta_3(\sigma_{93_2}, \sigma_{99_4})$	1
holonymy	$\eta_4(\sigma_{93_2}, \sigma_{99_4})$	0
synonymy	$\eta_5(\sigma_{93_2}, \sigma_{99_4})$	0
antonymy	$\eta_6(\sigma_{93_2}, \sigma_{99_4})$	0

5 The Approach

The problem of finding the appropriate sense for w_i can be transformed to the problem of choosing the corresponding appropriate semantic set for w_i. This means we intend to find a measure function $M(S_{i_p}) = m_{i_p}$ such that $\text{argmax}_{1 \leq p \leq s_i} M(S_{i_p})$ gives the correct sense for the word w_i.

The idea is that, a word-sense in the text indicates the presence of other word-senses in the piece of text in such a way that **semantically close word senses should also appear textually close**. Therefore, a word-sense in the text is affected by another word-sense in the text in two ways. First is that, *as the semantic distance between them increases, the influence should decrease*. Secondly, *as the textual distance between them increases, the influence should decrease.*

Intuitively, the first factor plays a predominant role in determining the sense of the word under consideration. This follows from the fact that **slight variation in textual position of a word-sense should not influence the sense of the passage as such. But a slight variation in semantic distance should considerably alter the sense of the passage.**

Based on these two observations, we state the hypothesis in section 5.1

5.1 Simple Manhattan measure

Hypothesis

The measure $M(S_{i_p})$ is of the form $M(\overline{P_{i_p}}, \overline{N_{i_p}})$. The contribution of each word σ_{i_j} in the semantic set to the score $M(S_{i_p})$ decreases exponentially its the semantic distance from w_i and decreases inversely with its textual distance from w_i.

Semantic distance (defined in section 4) can be restated as the *Manhattan distance*, $H(\sigma_{i_p}, \sigma_{j_q})$ in equation 5.

$$H(\sigma_{i_p}, \sigma_{j_q}) = \Sigma_{m=1}^{6} \eta_m(\sigma_i, \sigma_{j_q}) \tag{5}$$

According to the hypothesis mentioned above, the expression for the measure function is as given in equation 6

$$M(\overline{P_{i_p}}, \overline{N_{i_p}}) = \Sigma_{F(\sigma_{i_p}, \sigma_{j_q}) \in S(\sigma_{i_p})} \frac{1}{|P_{j_q} - P_i|} \times e^{-H(\sigma_{i_p}, \sigma_{j_q})} \tag{6}$$

For a word w_i, the appropriate sense number is p and the second most appropriate sense number is $\overline{p}$ iff the conditions given in equations (7) and (8) are satisfied.

$$p = \underset{0 \le j \le s_i}{\operatorname{argmax}} M(S_{i_j}) \tag{7}$$

$$\overline{p} = \underset{0 \le j \le s, j \ne k}{\operatorname{argmax}} M(S_{i_j}) \tag{8}$$

5.2 Eucledian measure

Instead of using the *Manhattan distance*, one can use the *Eucledian distance* . The intuition is given in the figure 4

We can look upon the words as being arranged in a six dimensional space, with each space corresponding to one of the 6 relations (*hypernymy etc*). The figure 4 for instance, shows the word-senses $end(4)$, $burden(4)$ and $term(4)$, with respect to the word-sense $program(2)$ in 3-D space of *hypernymy, hyponymy* and *synonymy*.

Instead of using the distance measure as in equation 5, we can use the measure $H(\sigma_{i_p}, \sigma_{j_q})$ as in equation 9

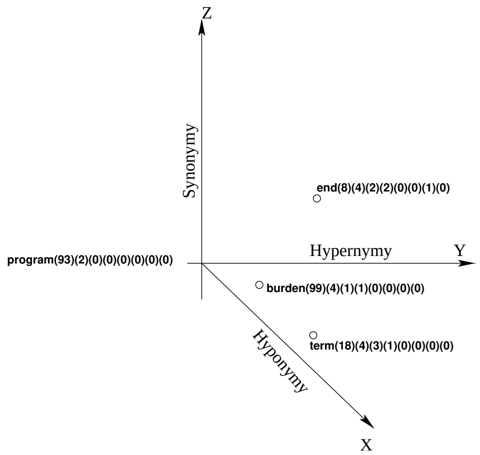

Figure 4: 3-D Graph showing the relative positions of three words with respect the the word *program*

$$H(\sigma_{i_p}, \sigma_{j_q}) = \sqrt{\Sigma_{m=1}^{6} (\eta_m(\sigma_i, \sigma_{j_q})^2)} \tag{9}$$

The appropriate sense for the word w_i can be found as before, using equation 6 and 7. In the measure in equation 9, we give uniform weight-age to all the six relations - *hypernyms* etc. One can instead, give more weight-age to the *hypernymy* and *synonymy* relations as compared to the other relations (say, by taking cubes instead of squares), since, they determine the context of a passage of text, to a greater extent. This gives us the equations 10 and 11 for $H(\sigma_{i_p}, \sigma_{j_q})$.

$$E(\sigma_{i_p}, \sigma_{j_q}) = (\eta_2(\sigma_{i_p}, \sigma_{j_q})^3) + (\eta_5(\sigma_{i_p}, \sigma_{j_q})^3) \tag{10}$$

$$H(\sigma_{i_p}, \sigma_{j_q}) = \sqrt{\Sigma_{m=1, m \ne 2,5}^{6} (\eta_m(\sigma_i, \sigma_{j_q})^2) + E(\sigma_{i_p}, \sigma_{j_q})} \tag{11}$$

Again, one can employ equations 6 and 7 to find the appropriate sense for w_i.

Mutual Reinforcement

We may note that **a word w which has a unique sense in WordNet, helps disambiguate other words related to it**. That is, if word w_j has only one WordNet sense, we would like to give special attention to this information, in all the sets that contain σ_{j_1}. For instance, if the p^{th} semantic set for w_i, i.e S_{i_p} has the word w_j, with w_j having only one sense in the WordNet, giving more weight-age to w_j, sense number 1, will add additional emphasis on the p^{th} sense of w_i.

Moreover, we would like that this effect on σ_{i_p} be reflected on all the sets that contain σ_{i_p}.in turn. To ensure that this happens, we make the following changes to equation 6. Initially, we set the score for each semantic set to 1. Next, within the semantic sets for a word, we normalize the scores. Not that sets corresponding to unambiguous word senses (i.e word senses for the words having just one WordNet sense) will have a score of 1 initially. Then we find the new measure for each semantic set using equations 13 and 12:

$$I(\sigma_{i_p}, \sigma_{j_q}) = M(S_{j_q}) \times \frac{1}{|P_{j_q} - P_i|} \times e^{-H(\sigma_{i_p}, \sigma_{j_q})} \quad (12)$$

$$M(\overline{P_{i_p}}, \overline{N_{i_p}}) = \Sigma_{F(\sigma_{i_p}, \sigma_{j_q}) \in S(\sigma_{i_p})} I(\sigma_{i+p}, \sigma_{j_q}) \quad (13)$$

After updating all set measures for w_i using equation 13, we normalize the measures for the sets corresponding to w_i using equation 14.

$$M(S_{i_p}) = M(\overline{P_{i_p}}, \overline{N_{i_p}}) = \frac{M(S_{i_p})}{\Sigma_{r=1}^{s_i} M(S_{i_r})} \quad (14)$$

Note that in the equation 13 we have scaled the entry for each term σ_{j_q} in the set S_{i_p}, by the measure $M(S_{j_q})$ for the corresponding set S_{j_q}. This means that, if in a particular iteration, sense number q of w_j is found to be more probable than the other senses of w_j, then it's contribution to the scores of other sets is more than the other senses of w_j.

The pseudo-code is summarized in figures 5 (INITIALIZATION) and 6 (MUTUAL REINFORCEMENT).

6 Experiments and results

Experiments were performed over nouns in Brown corpus and checked against SemCor for correctness. As an example case, consider the 93^{rd} noun, *program* in the text in figure 1. It is tagged with sense number 2 in SemCor. Figure 7 shows the 8 semantic sets for the word *program*.

Using equations 5 and 6, we get the scores for the different sets as indicated by the bold number to the right of each set in the figure. *The scores stabilize after around 10 iterations.* We find highest score for the second set - thus indicating *sense number 2*. Thus, as per our expectation, the algorithm correctly disambiguated the word *program*. On the other hand, using equations 9 and 6, we get the scores as the underlined number, to the right of each set, in the figure 7. As far as the *Eucledian distance* was concerned, it did not

1. **INTIALISATION**

2. Incrementally construct semantic chains S_{ip}, $1 \leq p \leq i_s$, for each of the i_s WordNet senses of σ_i, $1 \leq i \leq n$.

3. for all $1 \leq i \leq n$ do

   (a) for all $1 \leq p \leq s_i$ do

      i. $M(S_{ip}) = \frac{1}{s_i}$ /* Note that we have combined 2 steps into 1; setting $M(S_{ip})$ to 1 and then normalising */

Figure 5: The INITIALIZATION Pseudo-code for the method

make a big difference, whether we used the measure as suggested in equation 9 or 11. The experiments were carried out on the fist 100 nouns for each of 10 documents from the *Brown corpus*. 2 tests were done - (1) comparing the top ranked sense p and (2) comparing the 2 top ranked senses, p and $\overline{p}$ derived using equation 7. The results for 5 of them are tabulated below in table 2.

Table 2: Results with top sense for each of 10 *brown corpus* documents

Text	Coverage (%)	Precision (%)	Recall (%)
a01	99	70	69.3
a02	98	69	67.6
a11	96	63	60.5
a12	95	65.0	61.8

The average *precision* obtained using the *Eucledian measure* was $3-4\%$ lower than that obtained using the *Manhattan measure*.

The comparison of our algorithm was done with [4], one of the best known *Unsupervised* WSD algorithms.The comparison was performed on the entire text of *br-a01*. The results were as mentioned in table 4

The window size $|W| = n$, for all the above tests was chosen as 100. Changing it to 150 produced im-

1. do till the scores stabilise
   (a) **MUTUAL REINFORCEMENT**
   (b) for all $1 \leq i \leq n$ do
      i.
      ii. for all $1 \leq p \leq s_i$ do
         A. $H(\sigma_{i_p}, \sigma_{j_q}) = \Sigma_{m=1}^{6} \eta m(\sigma_i, \sigma_{j_q})$ /*This could be replaced by the Eucledian measure.*/
         B. $M(\overline{P_{i_p}}, \overline{N_{i_p}}) = \Sigma_{F(\sigma_{i_p}, \sigma j_q) \in S(\sigma_{i_p})} M(S_{j_q}) \times \frac{1}{|P_{jq} - P_i|} \times e^{-H(\sigma_{i_p}, \sigma_{j_q})}$
   (c) **NORMALISATION**
   (d) for all $1 \leq i \leq n$ do
      i. for all $1 \leq p \leq s_i$ do
         A. $M(S_{i_p}) = M(\overline{P_{i_p}}, \overline{N_{i_p}}) = \frac{M(S_{ip})}{\Sigma_{r=1}^{s_i} M(S_{ir})}$

Figure 6: The MUTUAL REINFORCEMENT Pseudo-code for the method

program<93,1,<0,0,0,0,0,0>>, ... evidence<4,1,<0,2,2,1,0,0>> = 0.312, **0.023**
program<93,2,<0,0,0,0,0,0>>, ... portion<95,1,<0,3,1,0,1,0>> = 1.129, **0.088**
program<93,3,<0,0,0,0,0,0>>, ... election<35,2,<0,3,1,0,1,0>> = 0.144, **0.009**
program<93,4,<0,0,0,0,0,0>>, ... report<24,1,<0,1,1,0,0,0>> = 0.899, **0.611**
program<93,5,<0,0,0,0,0,0>>, ... title<65,1,<0,1,1,1,0,0>> = 0.186, **0.017**
program<93,6,<0,0,0,0,0,0>>, . distribution<91,3,<0,1,3,1,0,0>> = 1.107, **0.053**
program<93,7,<0,0,0,0,0,0>>, ... laws<38,1,<0,2,2,0,0,1>> = 0.629, **0.045**
program<93,8,<0,0,0,0,0,0>>, ... city<56,1,<0,1,3,0,0,1>> = 0.473, **0.022**

Figure 7: Example of 8 semantic sets for the word *program*

provement by $5 - 7\%$.

7 Conclusions

The algorithm discussed in this paper is unsupervised. Currently, it is designed only for disambiguating nouns. All it needs is WordNet, an extensively used lexical database. It can disambiguate any free running text, provided that the *part of speech tags* are provided. The idea behind the algorithm is theoretically well supported. It has many special features compared to previous unsupervised algorithms. Even though a window of words is used for disambiguation, all the nouns in the window are not considered with equal importance for disambiguating a word in the text - the importance decreases with increasing distance in the text as well as with increasing *Manhattan* or *Eucledian* distance in the WordNet. Also note that

Table 3: Results with top 2 senses for each of 10 *brown corpus* documents

Text	Coverage (%)	Precision (%)	Recall (%)
a01	99	83.8	83.0
a02	98	75.5	74.0
a11	96	79.2	76.0
a12	95	74.7	71.0

Table 4: Comparison with [4]

	Aigrre		Our algo	
	precision	recall	precision	recall
br-a01	66.4	58.8	76.9	68.2
br-a02	-	-	70.9	68.8
br-b13	-	-	77.8	75.5
br-c04	-	-	67.3	64.10

the same word, occurring in different parts of the window is disambiguated in a different way - it considers separately, the multiple occurrences of same word in the same window.

With slight modification, this algorithm can be used for disambiguating *verbs*, *adjectives* in any text. The corresponding *verb* and *adjective* taxonomies in the WordNet can be used for these purposes - in a most similar way.

The algorithm can be improved by choosing a different measure function or choosing different measures of semantic distance, than the two mentioned in this chapter. Also, consideration of collocation of words and verb-noun collocations, should give additional clues for disambiguation.

References

[1] David Yarowsky, "Word-sense disambiguation using statistical models of Roget's categories trained on large corpora", in *Proceedings of the 14^{th} International Conference on Computational Linguistics (COLING-92), pages 454-460, Nantes, France, 1992.*

[2] Hwee Tou Ng and Hian Beng Lee, "Integrating Multiple Knowledge Sources to Disambiguate Word Sense: An Exemplar-Based Approach", in *Proceedings of ACL96, 1996.*

[3] Adam Kilgarriff, "Gold Standard Data-sets for Evaluating Word Sense Disambiguation Programs", in *Computer Speech and Language 12 (4), Special Issue on Evaluation, 1998.*

[4] Agirre, E. and Rigau, G, "Word sense disambiguation using conceptual density" *Proceedings of COLING'96.*

[5] Phil Resnik, "Selectional preference and sense disambiguation" in *Proceedings of ANLP-97 Workshop on Tagging Text with Lexical Semantics: Why, What, and How?, Washington D.C., USA, 1997.*

[6] Hinrich Schtze, "Automatic Word Sense Discrimination", *Computational Linguistics, Volume 24, Number 1, 1998.*

[7] Nancy Ide, "Parallel Translations as Sense Discriminators", in *Proceedings of SIGLEX99, Washington D.C, USA, 1999.*

[8] Green, S.J., "Automatically Generating Hypertext by Computing Semantic Similarity", in *Ph.D. Thesis, University of Toronto, 1997.*

[9] Fellbaum, Christiane, ed., "WordNet: An Electronic Lexical Database", *MIT Press, May 1998.*

[10] Cowie.J, Guthrie.L and Guthrie.J, "Lexical Disambiguation using simulated annealing", in *Proceedings of the 5th International Conference on Computational Linguistics. COLING-93 (1992), pp157-161.*

[11] McRoy.S, "Using multiple knowledge sources for Word Sense Disambiguation". *Computational Linguistics 18.1 (1992), pages 1-30.*

[12] Li.X, Szpakowicz.S and Matwin.M, "A WordNet based algorithm for word sense disambiguation". *Proceedings of the 14 th Joint International Conference on Artificial Intelligence.*

[13] Resnik.P, "Selectional preference and sense disambiguation", in *Proceedings of the ACL Singlex Workshop on Tagging Text with Lexical Semantics, Why, What and How?* (Washington DC, April 1997).

[14] Miller.G, Chodorow.M, Landes.S, Leacock.C and Thomas.R, "Using a semantic concordance for sense indentification", in *Proceedings of the* 37^th *Annual Meeting of the Association for Computational Linguistics (ACL-99).*

[15] Alexander Budanitsky, Graeme Hirst, "Semantic Distance in WordNet: An experimental, application-oriented evaluation of five measures", in *Proceedings of WordNet and Other Lexical Resources Workshop*, Second meeting of the North American Chapter of the Association for Computational Linguistics, Pittsburgh, June 2001.

[16] Graeme Hirst, David St-Onge, "Lexical chains as representations of context for the detection and correction of malapropisms", In *Christiane Fellbaum (editor), WordNet: An electronic lexical database*, Cambridge, MA: The MIT Press, 1998, 305-332.