On Some Optimization Heuristics
for Lesk-like WSD Algorithms[*]

Alexander Gelbukh,[1] Grigori Sidorov,[1] and Sang-Yong Han[2+]

[1] Natural Language and Text Processing Laboratory,
Center for Computing Research, National Polytechnic Institute, 07738, Mexico
{gelbukh, sidorov}@cic.ipn.mx, www.gelbukh.com

[2] Department of Computer Science and Engineering, Chung-Ang University,
221 Huksuk-Dong, DongJak-Ku, Seoul, 156-756, Korea. [+] *Corresponding author.*
hansy@cau.ac.kr

Abstract. For most English words, dictionaries give various senses: e.g., "*bank*" can stand for a financial institution, shore, set, etc. Automatic selection of the sense intended in a given text has crucial importance in many applications of text processing, such as information retrieval or machine translation: e.g., "(*my account in the*) *bank*" is to be translated into Spanish as "(*mi cuenta en el*) *banco*" whereas "(*on the*) *bank* (*of the lake*)" as "(*en la*) *orilla* (*del lago*)." To choose the optimal combination of the intended senses of all words, Lesk suggested to consider the global coherence of the text, i.e., which we mean the average relatedness between the chosen senses for all words in the text. Due to high dimensionality of the search space, heuristics are to be used to find a near-optimal configuration. In this paper, we discuss several such heuristics that differ in terms of complexity and quality of the results. In particular, we introduce a dimensionality reduction algorithm that reduces the complexity of computationally expensive approaches such as genetic algorithms.

1   Introduction

Most words we use in our everyday communication have several possible interpretations, called senses and listed in dictionaries. E.g., the word *bank* can be interpreted as a financial institution, river shore, stock of some objects, etc. For correct understanding of a text, the reader—be it a human being or a computer program—must be able to determine what sense is intended for each word in the text. Apart from message understanding, there are a number of important applications where automatically determining the correct sense of a word is crucial, e.g., information retrieval.

Given a dictionary and a specific occurrence of a word in a specific text, the problem of the choice, out of the senses listed for this word in the dictionary, of the one intended for this occurrence is called the word sense disambiguation (WSD) [2]. One
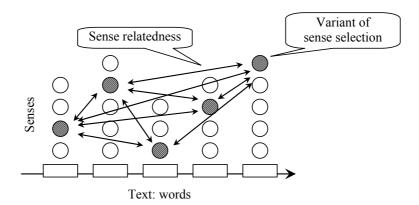

Figure 1. A variant of sense selection and the relatedness measure. The task is to find the variant with the maximum total relatedness.

of possible approaches to this problem is global optimization of text coherence, i.e., of average relatedness between the chosen senses for all words in the text [5], see Fig. 1. Due to high computational cost of this approach, evolutionary approaches were used to find a near-optimal solution [1, 4].

In this paper, we discuss some heuristics aimed to reduce the complexity of such global optimization. The paper is organized as follows. Section 2 discusses the heuristics for the search algorithm, aimed to fast finding a near-optimal solution. Section 3 introduces a new algorithm for reduction of dimensionality of the search space, useful for computationally expensive approaches. Section 4 presents the experimental results and discussion. Finally, Section 5 concludes the paper.

2 Search Heuristics

We denote N the number of words in the text fragment under disambiguation, n_w the number of senses for the word w, and $m^{uw}_{s_u s_w}$ the relatedness (a numerical value) between the sense s_u of the word u and s_w of the word w. The task in Fig. 1 consists in finding a combination of senses s_w maximizing the value $\sum_{u=1}^{N} \sum_{w=u+1}^{N} m^{uw}_{s_u s_w}$. The complexity of the task is exponential in N, namely, $\prod_{w=1}^{N} s_w$.

To select the best combination, exhaustive search [5], simulated annealing [1], or genetic algorithm [4] can be used. A number of heuristics can be tried as well. The heuristics we suggest resemble those used for other combinatorial problems, such as the traveling salesman problem.

- *A greedy approach* that for each word w chooses the sense s with maximum average relatedness to other words:

$$s = \arg\max_s \sum_{u=1, u \neq w}^{N} \sum_{i=1}^{n_u} m^{uw}_{is}. \tag{1}$$

The complexity of such an algorithm is only $Nn_w n_{avg}$, where n_{avg} is the average number of senses per word $u \neq w$. This approach is based on the hypothesis that the correct sense of the word w_i is not known and the probability for the sense to be the intended one is distributed uniformly. The average relatedness is that of a given sense to the senses of the other word weighted by the probability of those senses.

- *An even more greedy approach* that for each word w chooses the sense s with maximum relatedness to other words:

$$s = \arg\max_s \{m_{is}^{uw} \mid u \neq w; 1 \leq i \leq n_u\}. \tag{2}$$

The complexity of this algorithm is again $Nn_w n_{avg}$. The latter heuristic seems to be more motivated linguistically than the previous one. Indeed, a word is expected to be immediately related to one (or few) words in the context and not to all surrounding words. However, it is logically inconsistent, since the senses selected for other words can be different from those that influenced the decision for the given one. In this case, the order in which the decisions are made (the words are considered) becomes important. This idea corresponds to the insertion strategies for the traveling salesman problem. Different heuristics can be used to choose the optimal order: (1) *Direct order*: from the first to the last word, (2) *Inverse order*: from the last to the first word, (3) *Greedy order*: at each step, choose the word that gives the best increase in the total coherence value, (4) Another greedy order: choose the word that has a sense with a most clear advantage over other senses of the same word, (5) *A genetic algorithm* can be used to find an optimal order.

3   Reduction of Dimensionality

With computationally expensive heuristics discussed above (such as exhaustive search or evolutionary approaches) producing high-quality results, reduction of dimensionality before computation is of great help. For this, some senses can be proven in advance not to give an optimal solution. Namely, a sense s of the word w can be removed from consideration if

$$\max_{s' \neq s} \sum_{u \neq w} \min_{s_u} \left(m_{s_u s'}^{uw} - m_{s_u s}^{uw}\right) \geq 0. \tag{3}$$

Indeed, in this case there exists a sense s' of the same word that gives a better (or at least equal) contribution to the total coherence of the text with any selection of the senses s_u of all other words. In case of equality in the above formula, other criteria can be applied to select one of the two "equal" senses, such as that the senses with smaller numbers in the dictionary are often more frequent and thus more plausible.

The process is repeated iteratively, until no more senses can be excluded. The complexity of this process is polynomial.

4 Experimental Results and Discussion

The preliminary results of our experiments can be summarized as follows:

- The algorithms that look for globally optimized solution, such as genetic algorithms, perform (in terms of quality) some 10% better than heuristic approaches.
- Heuristic approaches perform about twice better than the baseline solutions such as random selection.
- Dimensionality reduction algorithm allowed us to reduce the number of senses in a randomly selected sample from 1138 (search space 5.2×10^{95}) to 433 (1.4×10^{42}), which speeds up the genetic algorithm twice. Still, the search space is too huge for exhaustive search. Hence the importance of the heuristic methods.

We also noted that iterating the dimensionality reduction algorithm did not give a considerable gain: most of the removed senses were removed at the first iteration.

5 Conclusions

We have suggested some heuristics to improve the speed of global coherence optimization WSD algorithms. In particular, we have described a dimensionality reduction algorithm useful for computationally expensive approaches, such as genetic algorithms. Our experiments show that such reduction can speed up such algorithms approximately twice, though still does not allow for exhaustive search.

In the future, we plan to investigate the effects of linguistically-motivated constraints on sense selection, such as the one-sense-per-discourse heuristic [3].

References

1. Cowie, J., J. A. Guthrie, L. Guthrie. Lexical disambiguation using simulated annealing. In Proc. of the International Conference on Computational Linguistics, 1992, 359–365.

2. Edmonds, P., and A. Kilgarriff (Eds.), Journal of Natural Language Engineering, Vol. 9 no. 1, 2003. Special issue based on Senseval-2; www.senseval.org.

3. Gale, W., K. Church and D. Yarowsky. *One sense per discourse*. In proc. of the DARPA Speech and Natural Language workshop, Harriman, NY, February 1992.

4. Gelbukh, A., Grigori Sidorov, San-Yong Han. Evolutionary Approach to Natural Language Word Sense Disambiguation through Global Coherence Optimization. *WSEAS Transactions on Communications,* 1(2):11–19, 2003.

5. Lesk, M., Automatic sense disambiguation using machine-readable dictionaries: how to tell a pine cone from an ice cream cone. Proc. of ACM SIGDOC Conference. Toronto, Canada, 1986, p. 24–26.