

An Existential Review of Word Sense Disambiguation Approaches

Jaspreet Kaur¹, Navjot Kaur²

¹SGGSWU, Fatehgarh Sahib, India
aarsi_2086@yahoo.co.in

²Assistant Professor ,SGGSWU, Fatehgarh Sahib, India
navjot_anttal@yahoo.co.in

Abstract: *Word sense disambiguation (WSD) is the process of automatically clarifying the meaning of a word in its context. Various issues like scalability, ambiguity, diversity and evaluation pose great challenges to WSD solutions. The aim of this report is to develop a WSD technique which can easily handle all these issues with better performance. and accuracy. The future work would include experimenting with different variations of the approach. This paper will provide users with general knowledge for choosing WSD algorithms for their specific applications or for further adaptation.*

Keywords: Word Sense Disambiguation, Deep Approach, Shallow Approach, Lesk algorithm.

1. Natural language processing (NLP)

NLP is a field of computer science, artificial intelligence and linguistics concerned with the interactions between computer and human languages. NLP is also being developed to create human readable text and to translate between one human language and another. The ultimate goal of NLP is to build software that will analyze, understand and generate human languages naturally, enabling communication with a computer as if it were a human. It is motivated by its use in many crucial applications such as Information retrieval, Information extraction, Machine Translation, Part of Speech tagging, Automatic Summarization, Natural language generation, Parsing, Optical Recognition and Question Answering.

2. Word sense disambiguation

Word sense disambiguation is the process of automatically figuring out the intended meaning of such a word when used in a sentence [7]. Words can have different senses. Some words have multiple meanings. This is called Polysemy. Word sense disambiguation is the process of finding the correct sense of a word depending on its context. For example word “bank” can mean “a financial institution”, “landform”, “supply” etc. The actual meaning is determined by the context. Word sense disambiguation techniques are often divided into two categories: supervised word sense disambiguation and unsupervised word sense disambiguation. Supervised word sense disambiguation relies on a sense-tagged corpus and uses information from the corpus to perform disambiguation. In contrast, unsupervised word sense disambiguation does not need sense-tagged corpus. It usually relies on a machine-readable dictionary [8].

3. Approaches of WSD

3.1 Deep approaches:

It presumes access to a comprehensive body of world knowledge. Knowledge such as “you can go fishing for a type of fish, but not for low frequency sounds” and “songs have low frequency sounds as parts, but not types of fish” is then used to determine in which sense the word is used. These approaches are not very successful in practice, mainly because we don’t have access to such a body of knowledge, except in very limited domains. But if such knowledge did exist, they would be much better than the shallow approaches [3].

3.2 Shallow approaches:

In this we don’t try to understand the text. They just consider the surrounding words, using information like “if ‘bass’ has words ‘sea’ or ‘fishing’ nearby, it probably is in the fish sense; if ‘bass’ has the words ‘music’ or ‘song’ nearby, it is probably in the music sense.” There are different types of shallow approaches to WSD [3]:

3.2.1 Dictionary-Bases approaches:

These rely primarily on dictionaries, thesauri, and lexical knowledge bases, without using any corpus evidence. A thesaurus is a reference work that lists words grouped together according to similarity of meaning (containing synonyms and sometimes antonyms) , in contrast to a dictionary, which contains definitions and pronunciations. In dictionary each word may have multiple meanings. Some dictionaries include each separate meaning in the order of most common usage while others list definitions in historical order, with the oldest usage first.

3.2.2 Machine Learning Approaches:

It can be further divided as:

a) Supervised methods [3]:

It is based on labeled training set. It makes use of training data that typically consists of a large set of example sentences of the ambiguous word, where each occurrence of the ambiguous word is tagged by a human with the sense in which the word is used. A set of rules is then automatically learned from this data that specify, for example, that if the words “dog” and “bark” both appear in a sentence and the word “tree” does not, then “bark” means a “dog’s call”. Using such rules, this approach can then disambiguate words occurring in new pieces of text.

b) Unsupervised methods [3]:

Unsupervised approaches on the other hand forgo the use of such data, and thereby avoid all the problems associated with the supervised approaches. Instead of hand-tagged data, these approaches typically make use of other sources of information. For example, the Lesk Algorithm uses the information contained in a dictionary to perform word sense disambiguation [6]. This algorithm is based on the intuition that words that co-occur in a sentence are being used to refer to the same topic, and that topically related senses of words are defined in a dictionary using the same words [5].

4. Conclusion

Word sense disambiguation is a key problem to address in many applications in the areas of Natural Language Processing, Information Retrieval and others. The various techniques identify the meaning of sentence like human brain. It disambiguates ambiguous words based on object on which sentence is written as in above example. The more examples must be taken so that its accuracy can be tested for different words. Further, we can improve the earlier proposed solutions, if possible and explore other approaches.

References

- [1] Zhi Zhong and Hwee Tou Ng , “Word Sense Disambiguation Improves Information Retrieval”, Department of Computer Science National University of Singapore 13 Computing Drive, Singapore 117417, In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, pages 273–282, Jeju, Republic of Korea ,July 2012.
- [2] Sulema Torres and Alexander Gelbukh, “Comparing Similarity Measures for Original WSD Lesk Algorithm”, Centro de Investigación en Computación (CIC IPN), Unidad Profesional Adolfo-López Mateos, Av. Juan de Dios Bátiz s/n and M. Othón de Mendizábal, Zacatenco, México, Advances in Computer Science and Applications Research in Computing Science 43., pp. 155-166 ,2009.
- [3] Xiaohua Zhou, Hyoil Han , “Survey of Word Sense Disambiguation Approaches”, College of Information Science & Technology, Drexel University 3401 Chestnut Street, Philadelphia, PA 19104, Appeared in

The 18th FLAIRS Conference, Clearwater Beach, Florida, May 15-17, 2005.

- [4] Rada Mihalcea “Using Wikipedia for Automatic Word Sense Disambiguation” Department of Computer Science, University of North Texas, April 2007.
- [5] Jonas EKEDAHL, Koraljka GOLUB, “ Word sense disambiguation using WordNet and the Lesk algorithm”, Engineering Physics, Lund Univ. Tunav. 39 H537, 223 63 Lund, Sweden.
- [6] Satanjeev Banerjee, “Adapting the Lesk algorithm for Word Sense Disambiguation to WordNet” , Master’s thesis. Dept. of Computer Science, University of Minnesota, USA, 2002.
- [7] <http://www.d.umn.edu/~tpederse/Pubs/banerjee.pdf>.
- [8] Eneko Agirre, Word Sense Disambiguation: Algorithms and applications (springer site, book site) Series: Text, Speech and Language Technology , Vol. 33 .
- [9] http://en.wikipedia.org/wiki/Word-sense_disambiguation.
- [10] Allen, James. “Natural Language Understanding”. 2nd Ed. Addison Wesley, 1995.