

# Improvement of Word Sense Disambiguation Using MINION

Romika Yadav

Banasthali University, Rajasthan, India

Rashmi Manker

Banasthali University, Rajasthan, India

## Abstract

*Word Sense Disambiguation problem is comes from NLP (Natural Language Processing); it is basically to select an appropriate sense of a word in the given context. This paper shows the word sense disambiguation improvement using MINION i.e. a constraint solver. In this paper the Word Sense Disambiguation problem is solved by collecting the aligned meaning of a Word with the help of MINION (A Tool) and then the RULES are formed using CLIPS language to get correct sense of a word.*

**Keyword:** Word Sense Disambiguation (WSD), CHAMPOLLION, MINION, C Language Integrated Production System (CLIPS), Natural Language Processing.

## 1.Introduction

Word sense disambiguation (WSD) is the ability to identify the correct sense of a words based on the context in a computational manner. Some words have multiple meanings; these types of words are called Polysemy. For example: word "Bank" can be a financial institute or to depend/trust. Sometimes two completely different words are spelled the same; these types of words are called Homonymy. For example: word "Can", can be used as model verb: You can do it, or as container: She brought a can of soda. By forming the WSD rules using CLIPS language with the help of MINION. So we can get the correct sense of the same kinds of words i.e. used in different meaning in different context.

A word having more than one sense depending on their context, which is Word Sense Disambiguation task to determine correct sense of a word in a given context[3]. Through Machine Translation a word having different senses in source language gives different translations in target language it means a word have multiple senses in the various contexts [6]. Word sense disambiguation will give the correct sense

of a word by the Machine Translation in the particular context. Word Sense Disambiguation used in the various applications such as Machine Translation and Information retrieval [7]. The usefulness of Word Sense Disambiguation in statistical based machine translation, which is more popular challenge now a day's [8].

In Word Sense Disambiguation there are two approaches i.e. Deep approach and Shallow approach. Deep approaches always presume access to a comprehensive body of world knowledge but these approaches are not successful in practice because the body of world knowledge is not in a computer readable format. While, the Shallow approach do not understand the complete text, but only considering the surrounding text.

Aim of Word sense Disambiguation is to find the correct sense of a word in a given context in which the word exists. The repository sense can come from WordNet (Computation lexicon), Dictionary (Machine readable) and a thesaurus [9]. When an ambiguous word is pronounced the sense of that word is correctly understood by the humans according to the situation and the context of that sentence, but it is difficult for a machine to decide the correct sense of the word in a given context. So where the machine processes the natural language application the problem of an ambiguity will arise. Consider an example take a word "bat" one sense is used: Bat hit the ball and another one is: Bat is a flying mammal [11]. Word Sense Disambiguation is described as an 'AI-Complete' problem it means first it solve the Artificial intelligence problems like encyclopedic knowledge and then sense word [14]. To identify the particular sense is a very difficult task for linguistics. The method used in word sense disambiguation is applied as to solve the problem of identifying the correct sense [15].

Word sense disambiguation facing the problem of a word having different senses and problem is, to find what the correct sense is. There are different types of dictionaries present, provide different sense on a given

word, which is mostly create a problematic situation to identify the correct sense. Other difficulties are part of speech tagging, inter-judge variance, discreteness of senses and common sense. The most obvious application of word sense disambiguation is machine translation. But WSD used in various other applications such as Information Retrieval, Web Semantic, knowledge mining and Bioinformatics.

## 2. Related Work

Giving an approach to improve WSD using topic features that consist of latent dirichlet allocation (LDA) algorithm on the data. This paper incorporated the features in the modification in Naive Bayes network like syntactic patterns, part of speech words, and single word context. This modified method achieved improvement and also more accuracy over the simple Naive Bayes network. The contribution of this paper includes the improvement of word sense disambiguation through the LDA algorithm and by the performance of Naive Bayes network in the WSD [3]. Giving an approach to improve WSD using Lexical Chaining that is mainly improving the accuracy of word sense disambiguation. In this paper a Linear – Time Algorithm is used for lexical chaining that gives one sense per discourse. Lexical chaining is based on the semantically connecting related words to create a chain that represent the cohesion through the text. The algorithm consists of three steps firstly collect all representation of a text, disambiguate all the words and then finally build a lexical chain [4].

Giving an approach improving the Impact of Subjectivity on Word Sense Disambiguation on Contextual Opinion Analysis. It provides an integration of Subjectivity Word Sense Disambiguation (SWSD) into the contextual opinion analysis for improve the performance. For SWSD the objective and subjective senses of a word is needed. In SWSD classifying the data which is firstly train for each target word. So it means the SWSD system having mainly SWSD classifiers as with all the targeted words [5].

Giving an approach of relation structure to measure the relationship between words by using the WordNet. Through the relation structure the exact sense is achieved. Semantic processing is maintained by the word sense disambiguation. In this paper algorithm is compared with the other algorithms and proofed that it gives the high precision [10].

Giving an approach of a research perspective of various methods of word sense disambiguation and their brief evaluation. The observation only stated that the word sense disambiguation is an open problem in natural language processing. The comparisons of

algorithm, hierarchical partitions and multilingual senses are distinct accordingly [12].

Giving an approach of k-way clustering method in which some authors having the same name, so it leads difficulty in web document retrieval and in web search information. This thing is improved by the k-way clustering method [13].

## 3. Word Sense Disambiguation Methods

### 3.1 Dictionary and knowledge based method

Dictionary Based or Knowledge based method utilize the information from the explicit lexicon or knowledge base is to disambiguate a word. Lexicon is a machine readable dictionary, ontology or a thesaurus. LESK algorithm is the first algorithm that is developed by the researchers for the Word sense disambiguation task. The LESK algorithm firstly introduced by the Michael E. Lesk [16]. LESK algorithm is based on the assumption of given neighborhood, which gives the common topics. A version of LESK algorithm is adapted from the WordNet [17], which means the word having different senses, in a dictionary and should count the amount of words that are neighborhood of disambiguated word. Finally the sense is chosen on the basis of highest count number. The Simplified LESK algorithm [18], in which each word is taken individually by locating the sense between the numbers of dictionary definitions according to a given context.

**Merits:** Provide a simple approach, which is easy to understand and do not need any trained data, Accuracy of the LESK algorithm is 50 to 70 percent i.e. purely depend on the word.

**Demerits:** Lesk Algorithm is more sensitive towards the word definition, if word is absent so the result changes rapidly. It is based on the dictionary so if sometime dictionary do not provides the sufficient meaning to match with the fine-grained senses then it creates the problem.

**Example -** Let take a word “ash” and this word having many senses let take 3 senses.

Sense s1 – Remains of burned thing.

Sense s2 – Leafs of ash tree.

Sense s3 – wood of various ash trees that is used to make furniture.

Sentences:

1. This house was burnt completely while the fire brigade reached here.
2. This chair made up of ash wood.

The LESK algorithm scores are listed in Table 1.

Context (Sentences)	Sense s1	Sense s2	Sense s3
1	1	0	0
2	0	0	1

**Table 1**

The above example shows how the LESK algorithm works it gives 50 to 70 percent accuracy on the disambiguated word. Three senses are assumed on the “ash” word and two sentences are taken of different context then score will evaluate by the LESK algorithm that is illustrated in Table 1.

### 3.2 Supervised Method

In supervised method the trained data is used to perform a machine learning tasks. The trained data is present in the form of training examples for the machine learning. Supervised algorithms produced the inferred function on the trained data, if the output comes in the discrete form so it is called classifier. If the output comes in the continuous form so it is called regression function. The various algorithms are available in the wide range but none of the algorithm will overcome the problems of supervised learning. Most popular approach we discussed here that is Naive Bayes Approach.

Naive Bayes approach basically works on the trained data as it is the supervised learning method, which is assuming the features of independence. ArgmaxP means their senses over feature vector (senses|feature vector).

$$\operatorname{argmax}_{s \in S} P(s) \prod_{j=1}^n P(v_j | s)$$

P(s) defines prior of the sense. P(v<sub>j</sub>|s) is the conditional probability of any particular feature; both are come from the corpus with the encoded features.

**Merits:** Simple approach with the trained data, the accuracy of the approaches is 70 to 80 percent.

**Demerits:** Requires the trained data; Less applicable on the high dimensional data.

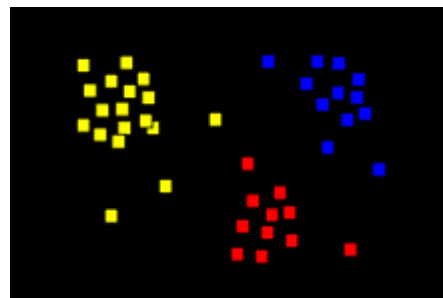
**Example -** Take a word “Line” test on the corpus of examples to check the accuracy of Naive Bayes Approach. The accuracy is achieved by the Naive approach is 73 percent correct.

### 3.3 Unsupervised Method

Unsupervised learning is trying to find out the hidden structure from the unlabeled data. Unsupervised learning is the biggest challenge for the researchers, which do not, contains trained data. Various approaches are used to select an appropriate sense of a word in a given context. Unsupervised learning used the Clustering method to select an appropriate sense on the basis of similarity in the context. The similar type of data become cluster and other types of data becomes another cluster. Take an example of clustering in illustrated in Figure 1. The yellow points showing one type of cluster data, blue showing another type of data and red also showing different type of data.

**Merits:** Unsupervised Learning Do not required any trained data; similar senses are creating a group or cluster that is called homogenous data.

**Demerits:** Unsupervised Learning algorithms sometime do not identify the correct patterns for a specific problem because unsupervised is an unguided method.



**Figure 1**

### 3.4 Semi-Supervised Method

Semi-supervised learning is an oldest approach of self-training or self labeling [19], application examples are given by the Scudder (1965) [20]. The framework of semi-supervised learning transductive learning is introduced by the Vladimir Vapnik (1970) [21]. A probability based semi-supervised learning i.e. Gaussian mixture was given by Ratsaby and Venkatesh

(1995) [22]. Semi-supervised learning methods lies between the supervised learning methods and unsupervised methods. It's not totally depended on the labeled data i.e. trained data and also not depend on the unlabeled data i.e. untrained data. Semi-supervised learning method needs annotated corpus. The various algorithms are available in the wide range but, most popular approach we discussed here that is Bootstrapping Approach.

Bootstrap basically a Semi- supervised learning approach to select an appropriate meaning of a disambiguate word sense in a particular context. In Bootstrapping Approach by taking a word and try to co-occur with the target word in the given sense then get the target word through the corpus and finally assume the target tag in the correct sense.

**Merits:** Eliminates the need of large trained data, Sense of the word is more clearly define with high accuracy.

**Demerits:** It is a repetitive process so the training corpus grows and the untagged instances are reduced, it requires trained data when it works on the labeled data.

**Example -** Take a word "bass" it's a musical instrument in one sense and in another sense it is a type of fish. Let assume "Play" (i.e. sense 1) in the sense of music and "fish" (i.e. sense 2) comes in the sense of fish. So the small number instances are labeled in the sense 1 and in the sense 2. These labeled are used to extract a large number of labeled instances. The bass result shown in Figure 2.

We need more good teachers - right now, there are only a half a dozen who can play the free bass with ease.
An electric guitar and bass player stand off to one side, not really part of the scene, just as a sort of nod to gringo expectations perhaps.
When the New Jersey Jazz Society, in a fund-raiser for the American Jazz Hall of Fame, honors this historic night next Saturday, Harry Goodman, Mr. Goodman's brother and bass player at the original concert, will be in the audience with other family members.
The researchers said the worms spend part of their life cycle in such fish as Pacific salmon and striped bass and Pacific rockfish or snapper.
And it all started when fishermen decided the striped bass in Lake Mead were too skinny.
Though still a far cry from the lake's record 52-pound bass of a decade ago, "you could fillet these fish again, and that made people very, very happy," Mr. Paulson says.

Figure 2

#### 4. Our Contribution

Improvement of word sense disambiguation already being done by various methods like LESK Algorithm. We are giving another approach to improve the WSD using MINION by getting alignment of English to Hindi sentences from Champollion tool. Now the word to word alignment is done by the MINION tool so a English word is aligned to a Hindi word which give the correct meaning of a word and then by writing the rules in the CLIPS language on those words to get the correct sense of a particular word. Area of this paper focused only on the sense of data set.

#### CHAMPOLLION

CHAMPOLLION tool is a parallel text Aligner , which is freely available for the public [1]. Champollion tool works on the noisy parallel text; it gives higher weights to the words those are less translated words. Champollion tool can easily ported on any language. Initially it was developed for aligning Chinese to English parallel text and later in pairs, including Arabic to English and Hindi to English. Parallel text plays a very essential role in Statistical Machine Translation (SMT) which includes machine translation, word sense disambiguation and Information Retrieval. Champollion tool is different from the other aligner tools; it is sentence based parallel text aligner [23].

#### MINION

Minion is a solver for constraint satisfaction problems. Constraints are the powerful and natural means of knowledge representation and inference in many areas of industry and academia. Constraint solving of a combinatorial problem proceeds in two phases. Firstly, the problem is modeled into the set of decision variables, and then the constraints are applied on those set of variables that a solution must satisfy. A decision variable represents a choice that it must be made in order to solve the problem. The domain potential values associated with each decision variable corresponds the options for that choice. Minion is Open Source Software, which is licensed under GNU General Public License Version 2. Minion is maintained by a Source Forge [2].

#### CLIPS

CLIPS is abbreviated as C Language Integrated Production System and it is a delivery expert system tool which provides a complete environment for the construction of rule and/or object based expert systems. CLIPS is now widely used in fields of academia, government and industries.

CLIPS support only forward-chaining rules. The OOP capabilities of CLIPS are referred to as CLIPS Object-Oriented Language (COOL). The procedural language capabilities of CLIPS are similar to languages such as C, Ada, Pascal, and Lisp. Facts consist of Relation name (symbolic field) and Zero or more slots with associated values.

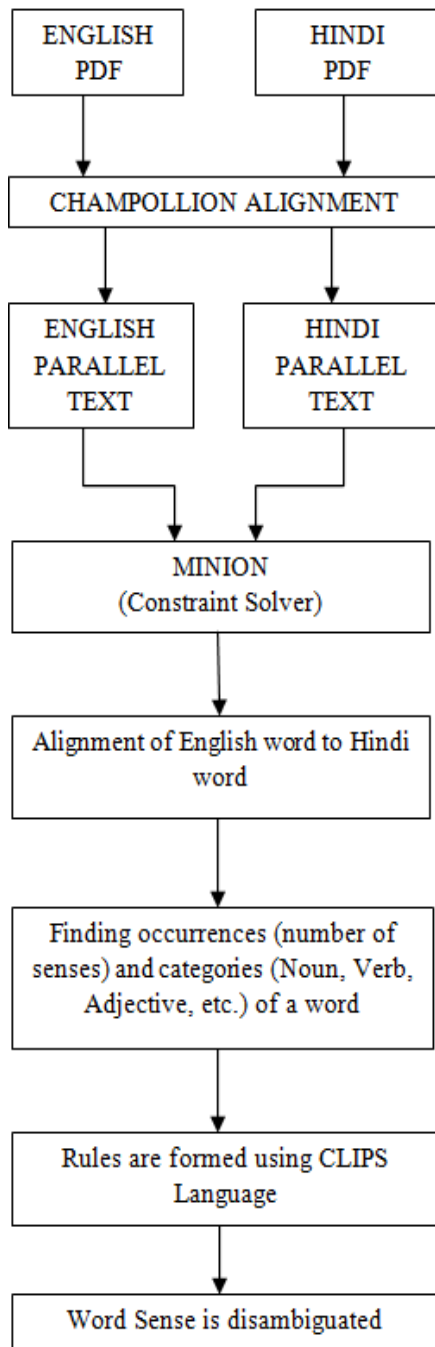


Figure 3

Figure 3 shows the two input file i.e. English PDF and Hindi PDF for the sentence alignment using Champollion tool. Then the output comes in the form of ENGLISH TEXT and HINDI TEXT files. Now these two files i.e. a parallel corpora and it is treated as input in MINION tool. Minion is a constraint solver that receives the inputs and applies constraints on them to find the appropriate meaning and solve them accordingly. Basically Minion gives Alignment of English word to Hindi word which is the correct meaning of those Words. Then correct sense of the word is achieved on the basis of its Occurrence (number of senses) and categories (Noun, Verb, Adjective etc.). First the meaning is decided with respect to the given context then the count is calculated on that particular meaning. The meaning which consist the highest count value on that sense of a word then rule is formed using CLIPS language. Finally the correct sense of the particular word is achieved. Some examples are given below:

**Example 1-** A word “top” having different meaning in the different context.

Sentence 1 - You were on top priority.

Sense: top refers to सर्वोच्च .

Sentence 2 - The Movement of axis is top around the vertical is called procession.

Sense: top refers to लड़.

**Example 2-** Another word “matter” having different meaning in the different context.

Sentence 1 – Matter is considered to be a substance that has rest mass and volume.

Sense: Matter refers to द्रव्य.

Sentence 2 – Why it matter to you if you do not care about it?

Sense: Matter refers to विषय.

Both sentences having different meaning in different context, so by writing the Rules on the word “top” and “Matter” which contains all the different senses in which only one meaning gives the correct sense to it. Firstly we sense the correct meaning of the word and then on the basis of its occurrence (number of senses) and categories (Noun, Verb, Adjective etc.), we form the rule using CLIPS language. Rules are providing the Hybrid approach which uses the mixture of trained and untrained data. In this approach WSD rules covers all the sense in a broader perspective. These are error free



but gathering the data collection is a time consuming task. Rules provide the correct sense accordingly and the purpose of WSD is solved.

## 5. Conclusion

We conclude that our approach is to improve the word sense disambiguation through Minion is simple and provide correct sense of word in a given context. We are using CHAMPOLLION and MINION tools to get the correct sense on a particular word. Through WSD rules, the correct sense is disambiguated. This paper is helpful for the beginners to get an idea about the word sense disambiguation approaches, their merits and demerits. The future work is to improve the efficiency of this proposed work.

## References:

- [1] <http://champollion.sourceforge.net>
- [2] <http://minion.sourceforge.net/>
- [3] Jun Fu Cai, Wee Sun Lee, Yee Whye Teh. Improving Word Sense Disambiguation Using Topic Features. Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 1015–1023, Prague, June 2007.
- [4] Michel Galley and Kathleen McKeown. Improving Word Sense Disambiguation in Lexical Chaining (2004).
- [5] Cem Akkaya, Janyce Wiebe, Alexander . Improving the Impact of Subjectivity Word Sense Disambiguation on Contextual Opinion Analysis. Proceedings of the Fifteenth Conference on Computational Natural Language Learning, pages 87–96, Portland, Oregon, USA, 23–24 June 2011.
- [6] Yee Seng Chan and Hwee Tou Ng, David Chiang. Word Sense Disambiguation Improves Statistical Machine Translation (2006).
- [7] Zhi Zhong and Hwee Tou Ng. Word Sense Disambiguation Improves Information Retrieval. Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, pages 273–282, Jeju, Republic of Korea, 8-14 July 2012.
- [8] Marine CARPUAT, Dekai WU. Improving Statistical Machine Translation using Word Sense Disambiguation. Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 61–72, Prague, June 2007.
- [9] Deepesh Kumar Kimtani, Jyotirmayee Choudhury, Alok Chakrabarty. Improvement in Word Sense Disambiguation by introducing enhancements in English WordNet Structure. International Journal on Computer Science and Engineering (IJCSSE) ISSN: 0975-3397 Vol. 4 No. 07 July 2012.
- [10] Myunggwon Hwang, Chang Choi, Byungsu Youn, Pankoo Kim. Word Sense Disambiguation based on Relation Structure. International Conference on Advanced Language Processing and Web Information Technology.
- [11] Mark Sanderson. Word Sense Disambiguation and Information Retrieval.
- [12] Philip Resnik, David Yarowsky. A perspective of word sense disambiguation Methods and Their Evaluation.
- [13] Hui Han, Hongyuan Zha, Lee Giles. Name Disambiguation in Author Citations using a Kway Spectral Clustering Method (2003).
- [14] Nancy Ide, Jean Veronis. Introduction to the Special Issue on Word Sense Disambiguation: The state of the Art.
- [15] Roberto Navigli and Paola Velardi, Structural Semantic Interconnections: A Knowledge-Based Approach to Word Sense Disambiguation. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 27, NO. 7, JULY 2005.
- [16] Lesk, M. (1986). Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone. In SIGDOC '86: Proceedings of the 5th annual international conference on Systems documentation, pages 24-26, New York, NY, USA. ACM.
- [17] Satanjeev Banerjee and Ted Pedersen. *An Adapted Lesk Algorithm for Word Sense Disambiguation Using WordNet*, Lecture Notes in Computer Science; Vol. 2276, Pages: 136 - 145, 2002. ISBN 3-540-43219-1.
- [18] Kilgarriff and J. Rosenzweig. 2000. English SENSEVAL: Report and Results. In Proceedings of the 2nd International Conference on Language Resources and Evaluation, LREC, Athens, Greece.
- [19] Chapelle, Olivier; Schölkopf, Bernhard; Zien, Alexander (2006). *Semi-supervised learning*. Cambridge, Mass.: MIT Press. ISBN 978-0-262-03358-9.
- [20] Scudder, H.J. Probability of Error of Some Adaptive Pattern-Recognition Machines. IEEE Transaction on Information Theory, 11:363–371 (1965). Cited in Chapelle et al. 2006, page 3.
- [21] Vapnik, V. and Chervonenkis, A. Theory of Pattern Recognition [in Russian]. Nauka, Moscow (1974). Cited in Chapelle et al. 2006, page 3.
- [22] Ratsaby, J. and Venkatesh, S. Learning from a mixture of labeled and unlabeled examples with parametric side information. In *Proceedings of the Eighth Annual Conference on Computational Learning Theory*, pages 412-417 (1995). Cited in Chapelle et al. 2006, page 4.
- [23] Xiaoyi Ma , Chanpollion: A Robust Parallel Text Sentence Aligner.