Approach to Correctly Distinguish the Meaning of Word in a Context

Nishakant Ojha

Abstract: The paper discusses the task of automatic selection of correct sense of word in the given context. In the Natural language processing, WSD is difficult task. Basically it used in the field of information retrieval, machine translation, text processing etc. For solving the problem of ambiguous words, different approach are used. In machine learning approaches Naive bayes classifier of superwise technique is used.

Key Terms: ambiguous, context, sense, word sense disambiguation.

I. INTRODUCTION

Word Sense Disambiguation (WSD) is a subfield of computational linguistics in which computer systems are designed to determine the appropriate meaning of a word as it appears in the linguistic context. In computational linguistics, the sense of a word is its meaning in a part of speech. A word may have different senses and it is difficult for the machine to determine what sense to refer to in a sentence. Word sense disambiguation is a task of removing ambiguities and selecting the closest sense of the word in context. The words which have many meanings is also called as polysemous word. These polysemous words contain ambiguity. Approach for remove the disambiguate the words are knowledge based approach, machine learning based approach, hybrid approach. Consider the example, in English language sentence "He scored a goal", "It was his goal in life" The word 'goal' carry different meaning in both the sentence.
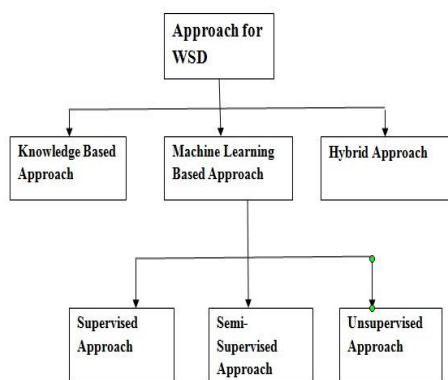

Fig: Approaches for Word Sense Disambiguation

The Figure shows the tree structure of different approach for solving WSD problem. Which contains the approaches to WSD with Machine Learning approaches in details.

Knowledge Based Approach

Knowledge-based approaches based on different knowledge sources as machine readable dictionaries or sense inventories, thesauri etc. Wordnet is the mostly used machine readable dictionaries in this research field. The knowledge based approach contains different algorithm such as Lesk algorithm, Conceptual density algorithm, Walker's algorithm etc.

Machine Based Approach

The machine learning approach is base on machine training. The machine based algorithm is again divided into 3 parts, that are supervised approach, semi-supervised approach and unsupervised approach. Every approach has different algorithm. This project is based on the supervised based approach. Navie Bayes algorithm which is based on the supervised approach is used in this project.

Hybrid Approach

These approaches can be neither properly classified as knowledge or corpus based but use part of both approaches. A good example of this is Luk's system which uses the textual definitions of senses from a machine readable dictionary (LDOCE) to identify relations between senses. He then uses a corpus to calculate mutual information scores between these related senses in order to discover the most useful.

Motivation

Word sense disambiguation a task of removing the ambiguity of word in given context, is important for many NLP applications such as:

1. Machine Translation: Every language have different context with different meaning of their use. Consider the example, in English language sentence "He scored a goal", "It was his goal in life" The word 'goal' carry different meaning in both the sentence.
2. Information Retrieval: Resolving ambiguity in the query is most important issue. For extracting data from database we have to write query. For getting the correct information from table, it should understand the query without ambiguity.
3. Information Extraction and text processing: WSD plays important role in natural language processing for extraction of information.

4. Question-Answering: Where the user asks a question and expects the system to search a document collection and then collate the information to provide an appropriately formed answer to the question.

5. Machine Translation: Where the system must translate a source text into the target language.

Problem Definition

There are many words in every language which carry more than one meaning for the same word. Those words, we called as Polysemous words. The word sense disambiguation is task of identifying the correct sense of polysemous words. The word sense disambiguation (WSD) is one of the oldest task in computational linguistics. There are different methodologies for solving the WSD problems such as knowledge based, machine learning and hybrid approach. The primary focus of knowledge based is on external knowledge dictionary to tackle WSD problem. The machine learning approach is depends on the training. It contain contains supervised and unsupervised technique. In the category of supervised machine learning algorithm, we would like to Navie Bayes algorithm. This very simplistic and high efficiency technique based in Bayesian decision theory. The objective of the proposed system is to disambiguate, analysis and evaluate the word sense disambiguation to disambiguate the sense of some words by using the Navie Bayes approach.

Proposed system of Paper

Section I of the paper involves the basic introduction of Word Sense Disambiguation, it's different approach to solve ambiguity of words. **Section II** describes the approach for the Word Sense Disambiguation. **Section III** describes the implementation details and implementation environment .

II.   APPROACH FOR SOLVING WSD

The proposed system is based on the machine learning approach. Sentence is given as input. Firstly, the sentence is scan by the system. After that task is to find out the ambiguous word. Then after find the exact sense of the word in the given context. The input sentence is pre-processed by elimination of the stop words technique. Identify the ambiguous word. After that sense of the word in the given context is find out. The exact sense of the word is displayed to the user along with the all possible sense of the word. Current paper shows the Naive Bayes of machine learning approach for find out the sense of the word.

Stop word filtering

In context, the feature i.e. ambiguous words are taken by removing the stop words. it will consist of the Stop words which are randomly collected from the various sources.

Target word selection

Ambiguous word get selected by removing the stop words, reaming words are compare with the ambiguous words. Consider the following example of ambiguous word: operation, process, fire, bat etc.

Finding the Sense of the word

By applying the Naive Bayes formula for finding the correct sense of the word among the all senses of the ambiguous word.
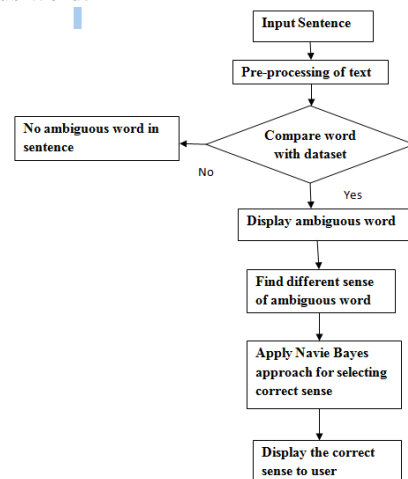
Fig: flowchart for Word Sense Disambiguation

Figure shows the flow chart for getting the sense of word in the sentence. Let "Operation" word has 5 senses, "Process" word has 4 senses, "Fire" word has 2 senses and "Bat" word has 2 senses. Apply Naive Bayes approach for finding different sense of word. The Bayes theorem and classifier find out the correct sense of the word . It calculates the probability of the words. A sense having greater probability than the other senses then it will displayed that sense is based on Baye's theorem. Following probability formula is used to calculate the probability of sense. From that we will calculate the correct sense of the word.

$$P(sense/feature) = \frac{P(sense)* P(feature/sense)}{P(feature)}$$

III.   IMPLEMENTATION

The proposed solution is designed to remove the ambiguity. The system will work with database and have destination storage. So the first job is to create the connection with database. The system is handling database i.e. Oracle1 0g. So java class connection is written to make connection with the database. The connection class will obtain the details for connection like name of database, its password and host string from frame which is written in view class. Once connection class obtains the details require to make connection, the connection class makes the connection to database through java program and returns the status of the connection. If connection is successfully done then it store data in database and retrieve the data at the time of comparison with the ambiguous, stop and meaning of word. First the replicas of source tables are created at intermediate database using Handler class. The tables in database are handle by simple SQL queries. The user can select the In implementation environment project is compatible on operating system Windows 8 or 7. The tool requirements are JDK 1.6.0 for java environment support, which is used for create common platform to run a java and Oracle 10g database connectivity program. For data storage Oracle 10g database is used and it is comfortable for window platform.

IV. RESULT

Finding the sense of the word based on the extraction of the stop word. Shows the ambiguous word from the given context after that it shows the option of the meaning of the word to the user. which will gives the meaning of the word.

Ambiguous words	No of times tested	Display correct sense	Display incorrect sense	Percentage Accuracy
Process	120	97	23	80%
Operation	165	133	32	80.6%
Fire	66	53	13	80.30%
Bat	57	48	9	84.21%
Total	408	331	77	81.27%

The proposed system provides 81.27% accuracy for identification of proper meaning of given word.

V. CONCLUSION

Disambiguation is the process of finding of correct meaning of the word, proposed system design gives the scenario of the disambiguation process. To remove the ambiguity of the words mentioned steps are used on documents, by understanding the scenarios of the existing system to overcome their problems and construct the propose system, by using the Naive Bayes classifier..

VI. SCOPE AND FUTURE WORK

WSD deals with the complexity of natural language. WSD is a hard task of Artificial Intelligence. There are many related problems such as annotating the untagged raw corpus, exact sense of words when present in combinations such as noun-noun, verb-noun, verb-verb and ranking different senses. Scope of the proposed system is up to some words which are disambiguated in the project. Furthermore, improvements can be implemented in order to increase the accuracy of the words.

This concept may enable users to find out the meaning of the word from the given context but in future, by merging the more than one approach of machine learning it will gives the more accuracy and can be extended to enable users get meaning more meaning fully and elaborately. Also, using in education and all other fields of user, it can be provide different types of template and themes. Also provide the security to the users. As we mentioned that by merging the approaches it will increases the complexity of the approach so complexity will be reduces as we apply the minimization to the system. Also it implements for the pdf documents.

REFERENCES

1. Ping Chen and Chris Bowes, Wei Ding and Max Choly, \Word Sense Disambiguation with Automatically Acquired Knowledge". July /august 2012. (Research Paper), pp. 3035.
2. Gurinder Pal Singh Gosal, \International Journal of Advanced Research in Computer Science and Software Engineering. A Nave Bayes Approach for Word Sense Disambiguation Available online at : www.ijarcsse.com ". Volume 5, Issue 7, July 2015.(Research Paper), pp. 336-340.
3. Cuong Anh Le and Akira Shimazu, \High WSD accuracy using Naive Bayesian classifier with rich features." December 8th-10th, 2004. (Research Paper).
4. Abhishek Fulmari, Dr.Manoj B. Chandak, \International Journal of Innovative Research in Computer and Communication Engineering. An Approach for Word Sense Disambiguation using modified Nave Bayes Classifier Available online at: www.ijarcsse.com ". Vol. 2, Issue 4, April 2014. (Research Paper).