

INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY

WINGS TO YOUR THOUGHTS.....

Decision List Master Approach to Solve WSD Problem

Boshra F. Zopon AL_Bayaty¹, Dr. Shashank Josh²

¹Bharati Vidyapeeth University, ²Bharati Vidyapeeth University

¹Department of Computer Science, Yashwantrao Mohite College,

²Department of Computer Engineering, Engineering College

¹AL- Mustansiriyah University, Baghdad, Iraq

India, Pune, Pin no.411046

¹bushraalbayaty123@gmail.com, ²Sdj@live.in

Abstract: Word sense disambiguation is central problem in NLP, and this problem has a long history in computational linguistics, there are many algorithms deal with this problem, one of supervised approaches is Decision List. In this paper, we will select decision list as a robust algorithm to be master approach in Master-Slave technique. In addition its obtained highest accuracy among several approaches, decision list has a good reputation in this field. Decision list obtained the highest accuracy among several supervised algorithms, its achieved (69.12) using WordNet and Senseval-3.

Keywords: Decision List, Supervised learning approaches, Senseval-3, WSD, WordNet, Master-Slave technique.

1. INTRODUCTION

There are a lot of words have multiple meaning, such as “Worship” word in screenshot below, The task to remove the ambiguity and select proper meaning or sense called Word Sense Disambiguation.



Figure 1: The Screenshot from WordNet Shows the Multiple meaning of Worship Word

We can distinguish two main approaches in machine learning, supervised and unsupervised approaches. Supervised approaches use machine learning technique to learn a classifier from labeled training sets, and supervised always obtained good results than unsupervised one^[1]. This work is a part of our research to implement Master – Slave technique, which discussed in more detail in section 5.

2. METHODOLOGY

The Information gathering about our study collected from survey, different books, Thesis, Web Sites and research journals and study different Supervised WSD Approaches. We started with dataset provided by the <http://www.e-quran.com/language/english>. Since one of particular steps is to train the dataset and to disambiguate the words by selecting the proper

meaning in context, we used WordNet 2.1 [Miller al.1990; fellbaum 1990], which is available at <http://wordnet.princeton.edu>, to know POS and provide the sense of words information and provides the mapping of word and different sense associated with that word. And to make sure, test, evaluation the our supervised approaches study and properly assigned to word, we used practically senseval-3 [Mihalcea and Edmonds, 2004] computation to preparation of data set using XML ”extensible Markup Language”, (which is referred to the restructure the context in the form XML, empirically our work, senseval XML mapping technique is used, where the given data set and senses are expressed with XML). And to ensure effective working of approaches training and testing file is used. Job of file is to provide the context which will be extremely useful exactly to know meaning of particular word. For implementing algorithms eclipse ID2, is used. At final we developed and implemented algorithm as Design and Creation phase using Java Language.

3. SUPERVISED APPROACHES IMPLEMENTED

We can broadly distinguish two main approaches to WSD, Supervised WSD and Unsupervised WSD. Supervised systems have obtained better results than the unsupervised ones, Supervised WSD approaches use machine learning techniques to learn a classifier from labeled training sets, which are sets of examples encoded in terms of a number of features together with their appropriate sense label.

We implemented empirical five supervised approaches, Naïve Bayes^[2], Decision list^[3], Support Vector Machine^[4], Decision Tree^[5] and AdaBoost^[6],

INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY

WINGS TO YOUR THOUGHTS.....

as in figure (2) above, and the results reported of these approaches are mentions in table (1) below.

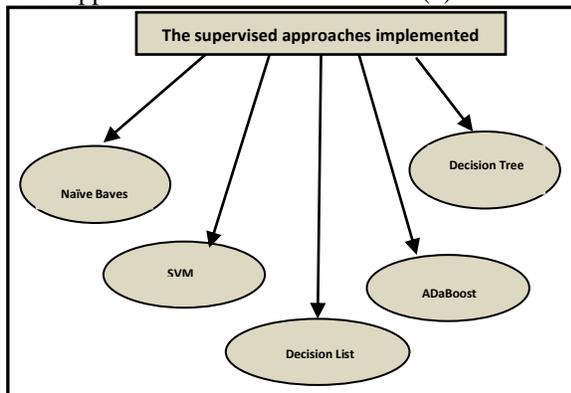


Figure 2: The supervised approaches implemented

Table 1: The final average of results of approaches research implemented

No.	Approach Name	Precision	Recall	F-measure
1.	Decision List	69.12	44.03333	20.738
2.	Adaboost	65.27	45.92	19.582
3.	Naive Bayes	62.86	30.57333	18.858
4.	SVM	56.11	32.1	16.834
5.	Decision Tree	45.14	33.22667	13.544

It's clear from the results shows in table above the decision List got the highest accuracy, In the given list of graph various graphs are drawn by considering precision, recall and f- measure. The comparative analysis of all words by considering precision, recall, and f-measure gives the summarized representation of accuracy of all algorithms. This helps to decide the efficiency of an algorithm. Once we know efficiency of an algorithm and the historical background of the algorithm, selection of Master will be very easy similarly selection of slaves can be done. After looking at the said parameters Decision list is better choice as Master and AdaBoost, Naïve Bayes are better choice as Slaves. Figure (3), represents the comparative analysis of graph to the supervised approaches implemented.

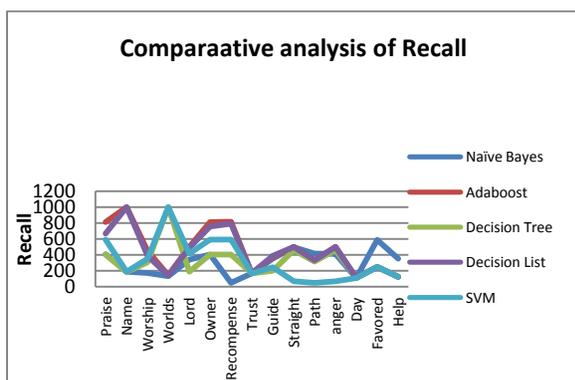


Figure 3: Comparative Analysis of Precision Graph

4. DECISION LIST FOR WSD HISTORICAL

Decision List is one of the robust approaches in Word Sense disambiguation field to address sense disambiguation. It's used for WSD by (Yarowsky, 1994). It has long history back ground e.g. - Kelly and stone, 1975, Block, 1988. Decision list is one of reputed algorithm with considerable historic background. History performance it is very important parameter which plays vital role in deciding algorithm as master or slave in our model suggested. Decision list has a good reputation in WSD field from the results previous work reported.

5. MASTER – SLAVE TECHNIQUE

The Master-Slave technique [7] is a technique to achieve important in web in search engine results, by combination one or more of supervised classifiers, there are many methods for WSD, such as voting, stacking. The voting method can be weighted or non-weighted; the weighted can be done by adding more weight to classifier which achieves high accuracy. Figure (4), below shows the master-slave technique [8], we called this technique (Master - Slave), because the master still control to select better classifier suggested by the slaves according to their results. In this technique decision list selected to be as Master approach, and Adaboost, Naïve Bayes will be as Slaves. Decision list empirically investigated good result among all supervised approaches research suggested, like Adaboost, Naïve Bayes, Decision Tree, and Support Vector Machine. Its reported (69.12%), in addition in WSD field as robust algorithm t has a good reputation in WSD field as robust algorithm to remove ambiguity from the words.

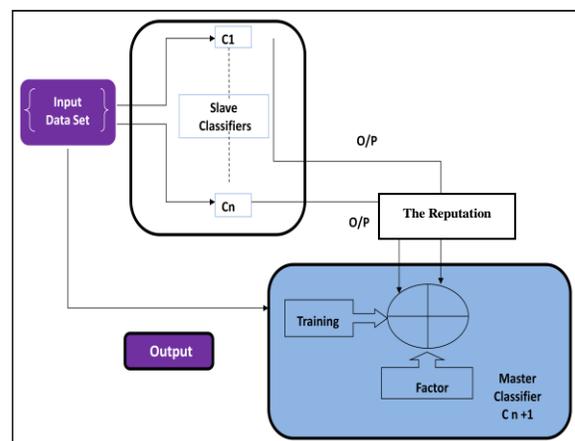


Figure 4: Mater – Slave technique

INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY

WINGS TO YOUR THOUGHTS.....

6. SYSTEM MODELING AND COMBILATION

When this application is executed output is delivered in two formats. One output, is in the form of .txt, which gives the weigh age, and the second output, gives the number of sense of given word on command prompt, as shown in figure (4) below:-

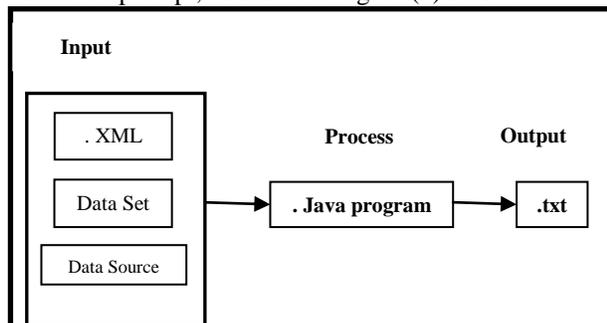


Figure 5: Modeling of files

7. DECISION LIST IMPLEMENTED

Decision List is one of reported algorithm with considerable historic background. It is combination of rules in a type of format. i.e. ordered list format. Using decision list we can obtain positive answer after considering senses [condition or situations]. In case there is multiple condition or sense by using decision list we get series of answers. For obtaining result it have some rules, why it is necessary? Because, after considering sense were getting particular result that result is calculated by using [if-then-else] rule. It is more supervised or used for decision algorithms. The senses calculation is done by testing phase by using decision list algorithm^[9].

7.1 Advantages of Decision List

There is a possibility that single instance of context could be matched with multiple collections in order to identify correct word, most significant proof is used. The process of categorization clubs large amount of information need not to be independent with complex modeling.

7.2 How Decision List can be build?

It is order which is collections in tests by using log of conditional probabilities. One is used to indicative meaning of words. We defined the formula below:

$$Abs(\log \frac{p(s = 1 | Fi = collection i)}{p(s = 2 | Fi = collection i)})$$

But there is one disadvantage of this approach that is, in most of the cases accuracy of words senses disambiguation gets hampered. Box 1. Below, shows you the Decision list algorithm applied to meet our

challenge and remove the ambiguity from the words of the data set.

Box .1: Decision List Algorithm implemented

1. Identify and calculate feature (f).
2. Calculate value of sense (Si).
3. Identify collocation one value per collocation basis.
4. Repeat this process for multiple senses.
5. Calculate absolute (log) of P (Si | f) for all sense.
6. Select maximum value out of it.

$$Abs \log \left(\frac{P(s_1 | f)}{P(s_2 | f)} \right)$$

8. EXPERIMENTAL SETUP

For word sense disambiguation following set up is performed:

1. Dataset provided by the <http://www.e-quran.com/language/english>, the dataset composed 15 English words, 10 nouns and 5 verbs, such as path, help which are have ambiguity.
2. WordNet 2.1 to refer the words senses and part of speech WordNet dictionary is used.
3. Senseval 3: To convert data into semi structured, i.e. XML format senseval is referred. In this for every meaning of given word different context is mapped. Sense delivering high accuracy is considered.
4. Java programming language is used to fetch the data and process it with the help of JWNL (java WordNet library) using decision List algorithm.

9. THE RESULT

The experiment over given dataset to resolve disambiguation, after conducting is as mentioned below: Day name worlds, lord, recompense, gives accurate result by providing 100% accuracy with Accuracy 1000/1000. While Owner, trust, path, favored, help, have value or score less than 50%. The results for our dataset shown in table (2) below:

TABLE 2: Data Set of Words and Results of Decision List Classifier

Word	POS	# Senses	Score	Accuracy
Praise	n	2	668	1000
Name	n	6	1000	1000
Worship	v	3	387	500
Worlds	n	8	142	1000
Lord	n	3	489	1000
Owner	n	2	755	999
Recompense	n	2	791	1000
Trust	v	6	167	167
Guide	v	5	387	995
Straight	n	3	500	500
Path	n	4	333	333
anger	n	3	500	500
Day	n	10	111	1000
Favored	v	4	250	250
Help	v	8	125	125

INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY

WINGS TO YOUR THOUGHTS.....

The graph represents the comparison among the accuracy levels of various words by considering their meanings and this variation is there for precision and recall for the data set selected. The comparative analysis of all words by considering precision, recall, and f-measure gives the summarized representation of accuracy of all algorithms as shown in table (1). This helps to decide the efficiency of an algorithm, figure (5) below represents the decision list separate graph.

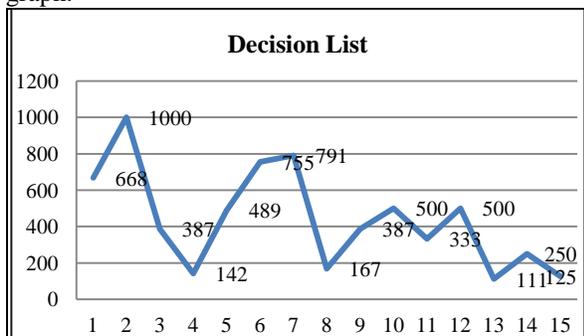


Figure 6: Decision List Graph

10. CONCLUSION

In this paper, we have selected Decision List algorithm using WordNet 2.1 and senseval-3, among several supervised approaches. The most important of our study regarding WSD it still open and Determining the proper meaning in a specific context is still such as challenge of WSD problem. The result of decision list reported in this work achieved it was (69.12%) accuracy according to the senseval-3. From this result, decision List algorithm is useful to be Master approach to implement Master- Slave technique suggested.

ACKNOWLEDGMENT

I would like to thank the ministry of higher education/Iraq; also I'm grateful to my research guide respected Dr. Shashank Joshi (Professor at Bharati Vidyapeeth University, College of Engineering) for his support to me always.

References

- [1] Nitin Indurkha and Fred J. Damerou "HANDBOOK OF NATURAL LANGUAGE PROCESSING" SECOND EDITION. Chapman & Hall/CRC, USA, 2010.(book style).
- [2] Boshra F. Zopon AL_Bayaty, Shashank Joshi, Empirical Implementation Naive Bayes Classifier for WSD Using WordNet., Bharati Vidyapeeth University, international journal of computer engineering & technology (IJCET), ISSN 0976 – 6367(Print), ISSN 0976

– 6375, Volume 5, Issue 8, August (2014), pp. 25-31,© IAEME: www.iaeme.com/IJCET.asp, Journal Impact Factor (2014): 8.5328 (Calculated by GIS), [Online]. Available: www.jifactor.com.

- [3] Boshra F. Zopon AL_Bayaty, Shashank Joshi, SENSE IDENTIFICATION FOR AMBIGUOUS WORD USING DECISION LIST, International Journal of Advance Research In Science And Engineering. [Online]. Available: <http://www.ijarse.com>, Vol. No.3, Issue No.10, October 2014 ISSN-2319-8354(E).
- [4] Boshra F. Zopon AL_Bayaty, Shashank Joshi, Implementation SVM to Solve Multiple Meaning of Word Problem, IJISSET - International Journal of Innovative Science, Engineering & Technology, Vol. 1 Issue 9, November 2014. www.ijiset.com ISSN 2348 – 7968
- [5] Boshra F. Zopon AL_Bayaty, Shashank Joshi, Empirical Implementation Decision Tree Classifier to WSD Problem, International Conference on Emerging Trends Science and Cutting Edge Technology (ICETSCET), YMCA, 28,Sep, 2014, International Journal of Advanced Technology in Engineering and Science. [Online]. Available: www.ijates.com,Volume No.02, Special Issue No. 01, September 2014 ISSN: 2348 – 7550.(conference style).
- [6] Boshra F. Zopon AL_Bayaty, Shashank Joshi, Empirically mplementation Adaboost to Solve Ambiguity, "National Conference on, Modelling, Optimization and Control, 4th -6th March 2015, NCMOC – 2015", Pune, India. (Conference style).
- [7] Ahmed H. Aliwy. Arabic Morphosyntactic Raw Text part of Speech Tagging System. PhD dissertation, University of Warsaw, 2013.
- [8] Boshra F. Zopon AL_Bayaty, Dr. Shashank Joshi, Conceptualisation of Knowledge Discovery from Web Search, Bharati Vidyapeeth University, International Journal of Scientific & Engineering Research, Volume 5, Issue 2, February-2014, pages 1246- 1248.
- [9] Arindam Chatterjee, Roll No: 09305905, Under the guidance of Prof. Pushpak Bhattacharyya, Word Sense Disambiguation: Literature Survey, Chapter 1, page 1-2, Department of Computer Science and Engineering, Indian Institute of Technology, Bombay, 2012.
- [10] <http://www.e-quran.com/language/english>

INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY

WINGS TO YOUR THOUGHTS.....

- [11] <http://wordnet.princeton.edu>.
- [12] <http://www.senseval.org/senseval3>.
- [13] Patrick Niemeyer and Jonathan Knudsen,
Learning Java, O'REILLY, Second Edition,
2002.
- [14] Steve Holzner, Eclipse, O'RILLY, 2007.