SENSE IDENTIFICATION FOR AMBIGUOUS WORD USING DECISION LIST

Boshra F. Zopon AL_Bayaty[1], Dr. Shashank Joshi[2]

[1] *Department of Computer Science, Yashwantrao Mohite College, Bharati Vidyapeeth University AL-Mustansiriya University, Baghdad, Iraq*

[2] *Department of Computer Engineering, Engineering College, Bharati Vidyapeeth University*

ABSTRACT

Meaning is to be identified to avoid disambiguation. Decision List performs this job by identifying collection data (feature. s) from context. This technique is based on the disambiguation based on the context and occurrence of meaning in that context. In this approach the score and accuracy is calculated and one with highest value is considered as a final value or result. Our study in this work achieved (69.12 %) accuracy.

Keywords: Decision List, Supervised Learning Approaches, Senseval-3WSD, Wordnet.

I INTRODUCTION

Word sense disambiguation is one of the main challenges in natural language processing. WSD one word many meaning, could cause problem during information retrieval, e.g. search engine could not interpret correct meaning as per user's perception. In information retrieval process it is difficult to identify correct meaning this problem is known as "Word Sense Disambiguation".

Fig. 1. The Screenshot from WordNet Shows the Multiple meaning of Worship Word.

[1] Lecturer in AL-Mustansiriya University, Baghdad, Iraq.

A word having multiple meanings is known as ambiguous words. For example port is point through which transport at sea is carried out and port is a channel or way on which process is executed in a computer. So meaning could be analyzed by looking at the context. The figure below showed the multiple meaning for verb "Worship".

II DECISION LIST FOR WSD

Decision List is one of robust approaches to address sense disambiguation. In this approach identifier collocation data (features) from sense tagged data based on conditional probability.

$$S = \text{argmax}_{Si \in SensesD(w)} score(Si).$$

$$\text{Argmax}(\text{abs log} \left(\frac{(si \mid f)}{(si \mid f)} \right))$$

Consider same example for sense 1 of worship mentioned in figure 1. Above, we can apply the formula as below:

Sense 1:

Abs log (3 / 1023) = log (3)
= log (3) – log (1.23)
= 0.477 – 0.089
= 0.387

Sense 2:

abs log (6 / 2.46)
= log (6) – log (2.46)
= 0.778 – 0.390 = 0.387

Sense 3:

Abs log (9 / 2.82)
= log (9) – log (2.84)
= 0.954 – 0.453
=500

The Sorted list as per accuracy high accuracy is selected as final result = 0.500, the table below shows the calculation result to worship word.

TABEL1. Apply decision list formula for worship word

Accuracy	Sense
0.500	3
0.387	1
0.387	2

Similarly for sense 2 = 0.425, and for sense 3 = 0.308. Box (1) shows the algorithm implemented in this work.

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.

$$\text{Max log}\left(\frac{P(si \mid f)}{P(s2 \mid f)}\right)$$

III EXPERIMENTAL SETUP

For word sense disambiguation following set up is performed:

1- Data source: Wordnet 2.1 to refer the words, senses and part of speech wordnet dictionary is used.

2- Java programming language is used to fetch the data and process it with the help of JWNL (java wordnet library) using decision List algorithm.

3- Senseval 3: To convert data into semistructured, 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.

IV TRAINING AND COMPILATION MODEL

To train a system dictionary in XML for**mat** providing context is used, once system understand the process of extraction of meaning from context meaning could be identified with the help of context. The screenshot below showed the training and compilation model.

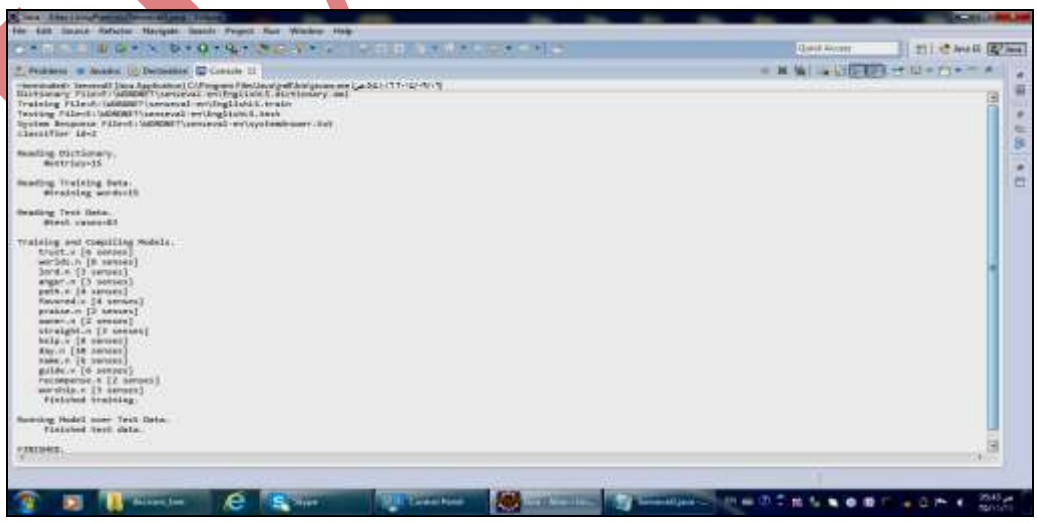

Fig. 2. The Screenshot Shows Training and compilation Model

V THE SYSTEM ANSWER FILE

This file provide accuracy related with various senses and meaning with high accuracy is identified and considered as a final answer by refering context. The screenshot below shows the System Answer. Txt file for decision List implemented

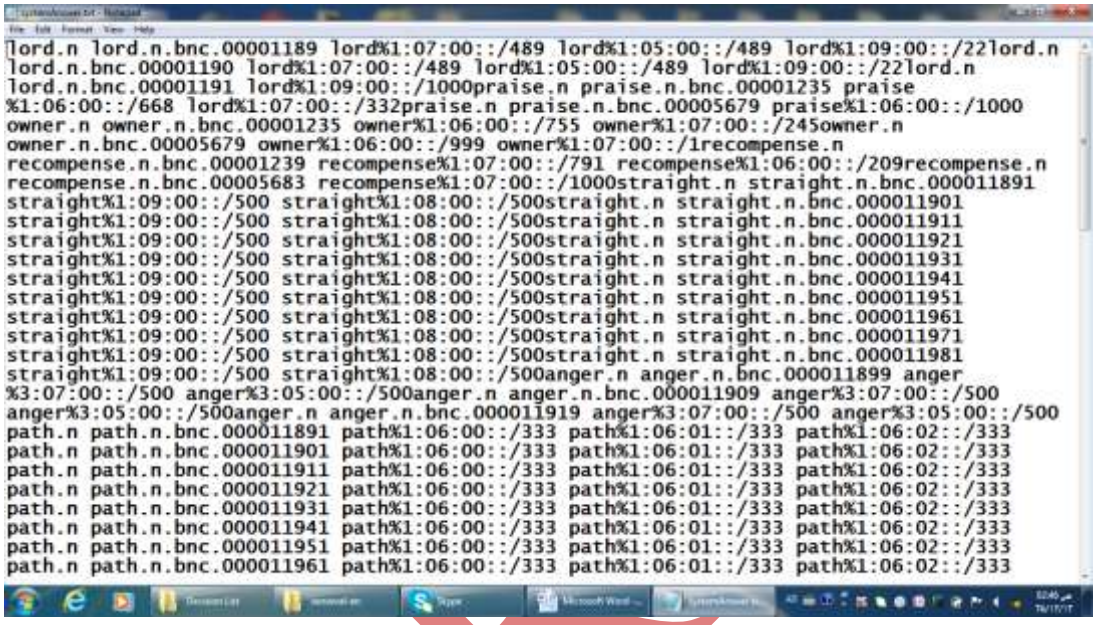

Fig. 3. The Screenshot Shows The System Answer.Txt File Compilation Model

VI 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 (1) 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

VII CONCLUSIONS

We have implemented Naive Bayes algorithm using WordNet 2.1. 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 reported in this work achieved it was (69.12%) accuracy according to the senseval-3. From this result, decision List algorithm is useful in some scenarios but not all.

Acknowledgment

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

Journal Papers

[1] 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.

[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(Online), Volume 5, Issue 8, August (2014), pp. 25-31,©

IAEME: ww.iaeme.com/IJCET.asp, Journal Impact Factor (2014): 8.5328 (Calculated by GISI), www.jifactor.com.

[3] 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 www.ijates.com,Volume No.02, Special Issue No. 01, September 2014 ISSN (online): 2348 – 7550.

[4] Boshra F. Zopon AL_Bayaty, Shashank Joshi, Word Sense Disambiguation (WSD) and Information Retrieval (IR): Literature Review, Bharati Vidyapeeth University, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 2, February-2014, pages 722- 726.

[5] Mark Alan Finlayson, Java Libraries for Accessing the Princeton WordNet: Comparison and Evaluation, Computer Science ad Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 2007.

[6] Oi Yee Kwong, Psycholinguistics, Lexicography, and Word Sense Disambiguation, Department of Chinese, Translation and Linguistics, copyright 2012 by Oi Yee Kwong, 26th Pacific Asia Conference on Languge, Information and Computation pages 408-417, 2012.

[7] Miller, G. et al., 1993, Introduction to WordNet: An On-line Lexical Database, ftp://ftp.cogsci.princeton.edu/pub/wordnet/5papers.pdf, Princeton University.

[8] Navigli, R. 2009.Word sense disambiguation: A survey. ACM Compute. Survey. 41, 2, Article 10 (February 2009), 69 pages DOI = 10.1145/1459352.1459355 http://doi.acm.org/10.1145/1459352.1459355.

[9] Approaches for Word Sense Disambiguation – A Survey, Pranjal Protim Borah, Gitimoni Talukdar, Arup Baruah, International Journal of Recent Technology and Engineering (IJRTE), ISSN:2277-3878, Volume-3, Issue-1, March2014.

[10] Ted Pedersen, Evaluation the Effectiveness of Ensembles of Decision Trees in Disambiguation Senseval Lexical Samples, department of computer science, university of Minnesota Duluth, Duluth, MN 55812 USA.

[11] Esha Palta, Word Sense Disambiguation, First stage report, Kanwal Rekhi School of Information Technology, Indian Institute of Technology, Powai, Mumbai, 2006-2007.

Books

[12] Nitin Indurkhya and Fred J. Damerau "HANDBOOK OF NATURAL LANGUAGE PROCESSING" SECOND EDITION. Chapman & Hall/CRC, USA, 2010.

[13] Daniel Jurafsky and James H. Martin, Naïve Bayes Classifier Approach to Word Sense Disambiguation, chapter 20, Computational Lexical Semantics, Sections 1 to 2, University of Groningen, 2009.

[14] Patrick Niemeyer and Jonathan Knudsen, Learning Java, O'REILLY, Second Edition, 2002.

[15] Steve Holzner, Eclipse, O'RILLY, 2007.

Theses

[16] Ahmed H. Aliwy. Arabic Morphosyntactic Raw Text part of Speech Tagging System. PhD dissertation, University of Warsaw, 2013.

[17] M. en C. Francisco Viveros Jimenez, Word Sense Disambiguation through associative dictionaries, PhD Thesis, Mexico, D.F. Jonio2014.

Links:

[18] http://www.senseval.org/senseval3.

[19] http://www.e-quran.com/language/english.

[20] http://wordnet.princeton.edu.