International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-2, Issue-10 E-ISSN: 2347-2693

The best performance method to Solve WSD Problem: Comparative Study

Boshra F. Zopon AL_Bayaty[1*] and Shashank Joshi[2]

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

www.ijcaonline.org

Received: 12 Oct 2014 Revised: 18 Oct 2014 Accepted: 25 Oct 2014 Published: 31 Oct 2014

***Abstract*—** Word is used to convey or extract meaning of particular information. If data that is meaning associated with word is misinterpreted then it will lead to incorrect data. To avoid this problem these is need to resolve meaning of given word correctly. This task can be performed with the help of repository of ambiguous word WordNet2.1 which gives meaning and POS of given word. Now with the help of some other parameter this data could be utilized. That parameter is nothing but context around given word.

***Keywords*—** Decision List, Decision Tree, Naïve Bayes, supervised learning approaches, WSD, WordNet, and Senseval-3

I. INTRODUCTION

Word sense disambiguation is to perform extraction correct meaning from given set or word. To perform this disambiguation these are number of ways to identify the meaning of word [1]:

1-Supervised Approaches: In this approach system is trained to identify correct meaning where input is given data. Data set and algorithm and output is score, accuracy.

2-Unsupervised approach: In this approach the system is not trained, but based on the data accurate meaning is predicate.

There are some important tools available like WordNet, SensEval, and Corpus.

II. BACKGROUND

Word sense disambiguation is one of the open problem in NLP. Many experiment or evaluation techniques are developed to solve WSD like WordNet, Corpus, Senseval. There are also various categories to address WSD as below [2]:

1. **Supervised:** System is trained to identify correct meaning, For example :

A- Naive Bayes: This approach evaluates individual probability to contribute final
B- Decision Tree: This approach is to represent data in terms of tree score with high value is considered.
C- Decision List: Yes/No format is used to select or reject the values.
D- SVM: This approach divides data into acceptable or reject able value by plotting hyper plane.
E- Adaboost: Iterative approach helps to identify correct, meaning.

2. **Unsupervised:** System is trained to tack decision, decisions are made by available data [3].

A- KNN
B- Cosine Distance Approach

Both approaches above, the meaning distance between meaning and correct value meaning with smallest distance is considered as a final result.

III. PROBLEM DEFINITION

To know correct meaning of word based on context (around the word)[4].

IV. REQUIREMENT TO ADDRESS THE ALGORITHM

1. Data Set: Sample data (combination verb and nouns) [5].
2. Training: For how to identify correct meaning we use context for training.
3. Word meanings mapping: We refer format suggested by senseval[6]l.
4. Data Repository: WordNet- contains word and their meaning with part of speech [7].

Corresponding Author: *bushraalbayaty123@gmail.com*

5. Algorithm: To fined weight or score for given instance (meaning) of word.

V. THE THREE APPROACHES IN THIS STUDY

We implemented empirically three supervised approaches, in this section briefly of each one:

1. **Naïve Bayes:** This classifier works on bayes theorem. Bayes theorem stats that every feature is independent of each other. These individual contributions meet final probability (Value) [8].
2. **Decision Tree:** Decision tree train system to divide the data in the form of tree actual value lies at leaf node and non-leaf nodes contains useful information used to derive final value. There some popular algorithms like ID3 and C4.5 which known as an example of decision tree [9].
3. **Decision List:** Decision list works on (If - else) roles. In this case based on the feature extracted from collection provide value per sense and log of their sense value and feature will give a final value which need be maximum value out of all values received[10].

VI. EXPERIMENTAL SETUP

Results are fetched by performing an experiment where decision list is used to resolve meaning by referring the context. WordNet repository is referred as dictionary to know POS, sense. Senseval is referred to restructure the context in the form of XML [11]. Training file is used train the system to identify meaning by using algorithm and context; which is supervised approach. In this work a bag- of words, we selected (A synonym set "synset") consist of 10 nouns and 5 verbs as below: {Praise, Name, Lord, Worlds, Owner, Recompense, Straight, Path, Anger, Day, Worship, Rely, Guide, Favored, Help}. As data source we select the WordNet lexicon [Miller al.1990; fellbaum 1990], is a great computational electronic lexicon database of English (noun, verbs adjective and adverb) which grouped as synonym sets. WordNet version 2.1 composed 207016 word sense pairs and 78695 polysemous senses [12]. In order evaluating our supervised study, we used the third addition of senseval computation to preparation of data set using XML collected from WordNet 2.1 adopted as a sense inventory for nouns and verbs [13, [14].

VII. RESULT

From the table1, there is no best algorithm as such. But based on the overall performance naïve bayes and decision list seems to be useful approaches because of the accurate values. Decision tree is not delivering the better performance as per as summation of overall result is concerned. This accuracy could be increased or decreased with the help of data set referred and context used to resolve to meaning [15], [16], [17].

TABLE 1.
DATA SET OF WORDS AND RESULTS OF NAÏVE BAYES AND DECISION TREE CLASSIFIERS

Word	POS	# Sense	Naïve Bayes		Decision Tree		Decision List	
			Score	Accuracy	Score	Accuracy	Score	Accuracy
Praise	**n**	**2**	**0.408**	**0.592**	**405**	**593**	668	1000
Name	**n**	**6**	**0.189**	**1.0**	**184**	**1000**	**1000**	**1000**
Worship	**v**	**3**	**0.172**	**0.414**	**308**	**425**	**387**	**500**
Worlds	**n**	**8**	**0.137**	**1.0**	**1000**	**1000**	**142**	**1000**
Lord	**n**	**3**	**0.341**	**0.681**	**187**	**426**	**489**	**1000**
Owner	**n**	**2**	**0.406**	**0.594**	**405**	**595**	755	999
Recompe-nse	**n**	**2**	**0.48**	**0.594**	**405**	**595**	**791**	**1000**
Trust	**v**	**6**	**0.167**	**0.167**	**167**	**167**	**167**	**167**
Guide	**v**	**5**	**0.352**	**0.648**	**199**	**247**	387	**995**
Straight	**n**	**3**	**0.496**	**0.504**	**462**	**462**	**500**	**500**
Path	**n**	**4**	**0.415**	**0.585**	**316**	**316**	**333**	**333**
anger	**n**	**3**	**0.412**	**0.588**	**462**	**462**	**500**	**500**
Day	**n**	**10**	**0.109**	**1.0**	**109**	**109**	**111**	**1000**
Favored	**v**	**4**	**0.587**	**0.648**	**250**	**250**	250	**250**
Help	**v**	**8**	**0.352**	**0.414**	**125**	**125**	**125**	**125**

TABLE.2
THE FINAL RESULTS OF NAÏVE BAYES AND DECISION TREE CLASSIFIERS

Approaches	Accuracy (%)
Naïve Bayes	58.32
Decision Tree	45.14
Decision List	69.12

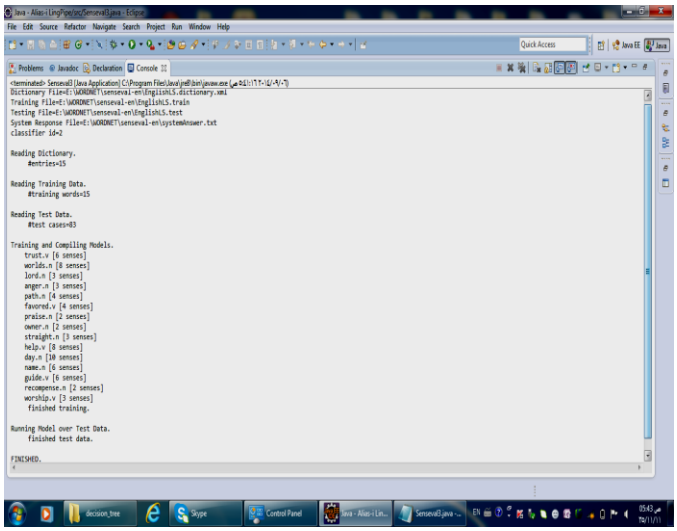

Fig. 1 The Screenshot Shows Training and compilation Model

VIII. CONCLUSIONS

After performing an experiment three different approaches, Naïve Bayes, Decision Tree, and Decision List, none of these algorithms provided accurate values throughout for all words [18]. For some words methods is useful for some approaches methods two or three is useful, but by considering overall accuracy decision list provide higher 69.12% of accuracy.

Acknowledgment

Boshra AL_Bayaty thanks her research guide Dr. Shashank Joshi (Professor at Bharati Vidyapeeth University, College of Engineering) for his support to her all the time.

REFERENCES

[1] 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.

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

[3] Ted Pedersen, A Decision Tree of Bigrams is an Accurate Predictor of Word Sense, department of computer science, university of Minnesota Duluth, Duluth, MN 55812 USA, 2004.

[4] Boshra F. Zopon AL_Bayaty, 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.

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

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

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

[8] 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.

[9] 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.

[10] Boshra F. Zopon AL_Bayaty, Shashank Joshi, Sense Identification for Ambiguous Word Using Decision List" in International Journal of Advance Research in Science & Engineering (ISSN 2319-8354), Volume 03, Issue 10, October 2014.

[11] David Yarowsky, Hierarchical Decision Lists for Word Sense Disambiguation, Computers and the Humanities 34: 197-186, 2000, Kluwer Academic Publishers. Printed in the Netherlands, 2000.

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

[13] A Combative Study of Support Vector Machines Applied to he Supervised Word Sense Disambiguation Problem in the Medical Domain, Mahesh Joshi, Ted Pedersen and Richard Maclin, Department of Computer Science, University of Minnesota, Duluth, MN 55812, USA.

[14] 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.

[15] Learning Rules for Large Vocabulary Word Sense Disambiguation, Georgios Paliouras,

Vangelis Karkaletsis, Constantine D. Spyropoulos, Institute of Informatics & Telecommunications, NCSR "Demokritos" Aghia Paraskevi Attikis, Athens, 15310, Greece.

[16] 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.

[17] Mahesh Joshi, MS, Serguei Pakhomov, PhD, […], and Christopher G. Chute, MD, DrpH. A Comparative Study of Supervised Learning as Applied to Acronym Expansion in Clinical Reports.

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

AUTHORS PROFILE

Boshra F. Zopon AL_Bayaty received her B.E degree in computer science from AL_Mustansiriya University, College of Education in 2002. And received her M.S.C degree in computer science from Iraqi Commission for Computers and Informatics, Informatics Institute for Postgraduate Studies. Doing her the PH.D. Computer Science at Bharati Vidyapeeth Deemed University, Pune.
She is currently working in the Ministry of Higher Education & Scientific Research, AL_Mustansiriyah University in Iraq/ Baghdad. Her research interests include software engineering.

Shashank Joshi, received his B.E. degree in Electronics and Telecommunication from Govt. College of Engineering, Pune in 1988, the M.E. and Ph. D. Degree in Computer Engineering from Bharati Vidyapeeth Deemed University Pune. He is currently working as the Professor in Computer Engineering Department Bharati Vidyapeeth Deemed University College of Engineering, Pune. His research interests include software engineering. Presently he is engaged in SDLC and secure software development methodologies. He is innovative teacher devoted to Education and Learning for the last 23 yrs.