ORIENTAL JOURNAL OF
COMPUTER SCIENCE & TECHNOLOGY
An International Open Free Access, Peer Reviewed Research Journal
Published By: **Oriental Scientific Publishing Co., India.**
www.computerscijournal.org

ISSN: 0974-6471
August 2015,
Vol. 8, No. (2):
Pgs. 99-102

Empirically Implementation Adaboost to Solve Ambiguity

BOSHRA F. ZOPON AL-BAYATY and SHASHANK JOSHI

Department of Computer Science,Yashwantrao Mohite College,
Bharati Vidyapeeth University, Al-Mustansiriya University, Baghdad, Iraq.

(Received: June 01, 2015; Accepted: July 28, 2015)

ABSTRACT

Word sense disambiguation is process of identifying correct meaning based on algorithm used. Many more research is carried out in this domain popular dataset referred is wordnet. This paper discuss about word sense disambiguation using adaboost algorithm. In this work wordnet data and senseval standards are used resolve meaning of word with the help of given context.

Key words: WSD, Supervised learning approaches, Senseval-3, WSD, WordNet.

INTRODUCTION

One of natural language processing applications is word sense disambiguation. There are two main ways to identify meaning of word correctly:

Supervised Approach

Where along with the algorithm context is used to train system to identify word correctly. Adaboost, is theoretical approach for learning model called probably Approximately correct (PAC). Adaptive Boosting constructs a strong classifier by taking a linear combination of a number of weak classifier. This approach is known as adaptive boosting, because classifier technique helps to classify those words which were not classified correctly.

Unsupervised Approach

In these approaches acquire information from unannotated raw text. Always the performance of unsupervised approaches is been lower than that of the other approaches used for word sense disambiguation

Problem Definition

To identify meaning of word correctly using adaptive boosting approach to improve overall classification. In this case algorithms are used to report their classification and then overall accuracy of classification is improved.

Excremental Setup

To address the problem statement discussed so far experiment is preformed and set up for that is as below.

1. Data set: 10 nouns, 5 verbs.
2. Reference for meaning and POS: WordNet ver. 2.1.
3. Algorithm: Adaboost.
4. Dictionary file: To specify meaning.
5. Training: To train system with given context.
6. Senseval format: Representation in the form of XML.
7. IDE: Eclipse kepler 6.0.
8. P.L.: J2SE 6.0.
9. O.S.: Windows 7 32 bit.

Implementation and algorithm used

Adaptive boosting approach identifies week learner (classifier) and boosts performance of these classifiers. The actual process carried out is as mentioned below.

Box (1): Adaboost Algorithm implemented

For x =1; x< m; x++)
{
Fetch weight áx from classifier cx

$$H(x) = sign\sum^{y} \text{a } x(x(X$$

}

Where H(x) sign is function for linear combine of weak learner to boost the performance.

To make learning process easier members of training data are weighted equally. Adaboost Algorithm treats it as an input. For X components, it is iterated y times one turn is allotted for each classifier.

The training phase

Data set of 10 nouns and 5 verbs is used. To make understanding of senses, system is trained by referring senseval-3 structure to map word with sense by using surrounding context. This entire structure uses XML format to represent and process data using semi structured approach.

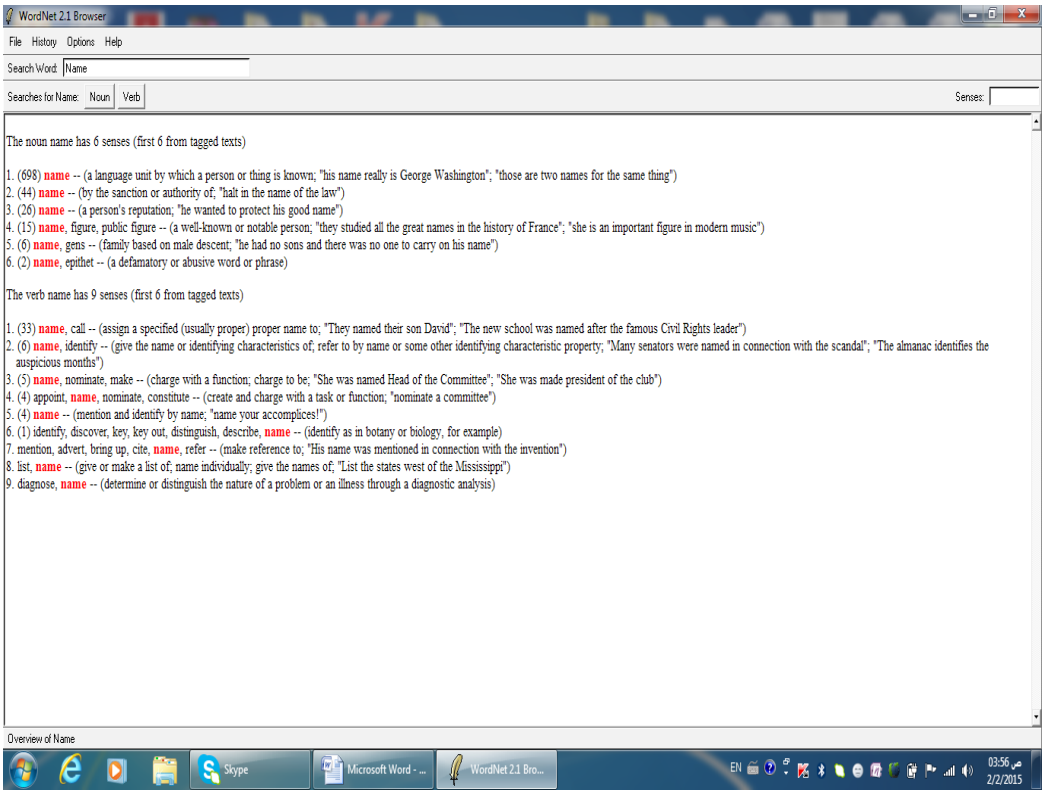

Fig. 1: The Screenshot Shows the Multiple of Name Word

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 Adaboost algorithm implemented

RESULT

The results for our dataset shown in table (1) below:

CONCLUSION

After performing this experiment for some words adaboost delivers more accurate results, for example {Day, Recompense, Owner, Lord, Worlds, Name, and Praise}. But for other words accuracy is not maintained this accuracy need to be modified to increase the probability of identifying word with correct meaning. In this part of our work Adaboost achieved 65.27% accuracy according to the data set using WordNet andSenseval-3.

ACKNOWLEDGMENT

The first author thanks the ministry of higher education/Iraq; also I would like to thank my research guide Dr. Shashank Joshi (Professor at Bharati Vidyapeeth University, College of Engineering) for submitted his advices within preparing this work.

Table 1: Data Set of Words and Results of Adaboost Classifier

Word	POS	# Senses	Score	Accuracy
Praise	n	2	812	1000
Name	n	6	1000	1000
Worship	v	3	450	485
Worlds	n	8	143	1000
Lord	n	3	500	1000
Owner	n	2	811	1000
Recompense	n	2	815	1000
Trust	v	6	167	167
Guide	v	5	371	431
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

Overall accuracy of adaboost is 65.27%, which is quite good.

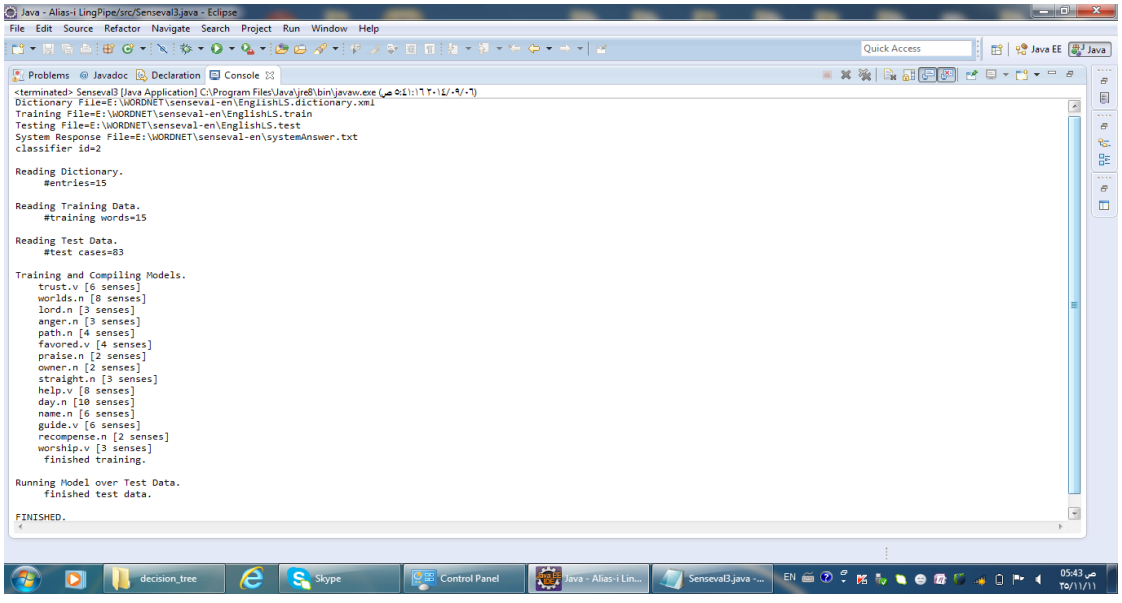

Fig. 2: The Screenshot Shows Taraining and Compilation Model

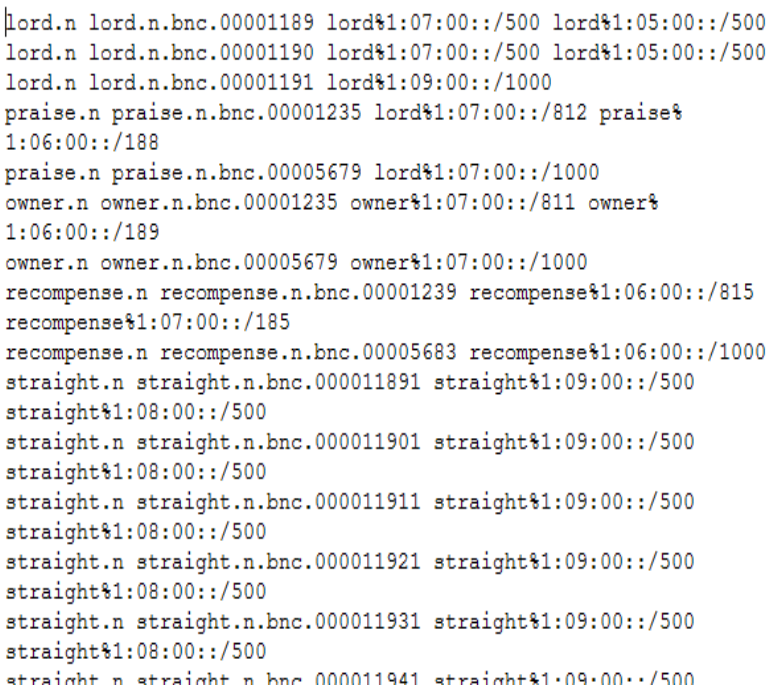

Fig. 3: The Screenshot Shows The System Answer.Txt File Compilation Model

REFERENCES

1. Nitin Indurkhya and Fred J. Damerau "Handbook of natural language processing" Second edition. Chapman & Hall/CRC, USA, 2010.
2. 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.
3. Patrick Niemeyer and Jonathan Knudsen, Learning Java, O'REILLY, Second Edition, USA 2002.
4. Steve Holzner, Eclipse, O'Rilly, Third Indian reprinted, 2007.
5. Zhi-Hua Zhou, Yang Yu, *National Key Laboratory for Novel Software Technology,Nanjing University,* Nanjing 210093, China, 2008.
6. Boshra F. Zopon AL_Bayaty, Dr. Shashank Joshi, Conceptualisation of Knowledge Discovery from Web Search, Bharati Vidyapeeth University, *International Journal of Scientific & Engineering Research,* **5**(2): 1246- 1248 (2014).
7. Miller, G. *et al*., Introduction to WordNet: An On-line Lexical Database,*ftp://ftp.cogsci.princeton.edu/pub/wordnet/5papers.pdf,* Princeton University (1993).
8. http://www.senseval.org/senseval3.
9. http://www.e-quran.com/language/english.
10. http://wordnet.princeton.edu.
11. https://code.google.com/p/pr-toolkit/source/browse/applications/postagging/trunk/src/edlin/classification/AdaBoost.java?r=5