

## Empirical Comparative Study to Some Supervised Approaches

Boshra F. Zopon AL\_Bayaty<sup>1</sup>, Dr. Shashank Joshi<sup>2</sup>

Department of Computer Science, Yashwantrao Mohite College, Bharati Vidyapeeth University, AL-Mustansiriya University, Baghdad, Iraq<sup>1</sup>

Department of Computer Engineering, Engineering College, Bharati Vidyapeeth University, Pune, Maharashtra, India<sup>2</sup>

**ABSTRACT:** Word sense disambiguation is solved with the help of various data mining approaches like Naïve Bayes Approach, Decision List, decision tree, and SVM (Support Vector machine). These approaches help to find out correct meaning of word by referring WordNet 2.1. Experiment performed is discussed in this paper along with the comparison of SVM algorithm with various approaches. In this study Decision List achieved the best result among all other approaches.

**KEYWORDS:** Support Vector Machine, Naive Bayes, Decision List, Decision Tree, Supervised learning approaches, Senseval-3, WSD, WordNet.

### I. INTRODUCTION

Natural language processing is study of word and their meaning role from meaningful language. Most for every system this word acts as an input. While inferring means out of it, if system misinterprets it entire system will get affected. That's why WSD is extremely important to infer correctly meaning of word as per the perception of user or machine who has inserted it.

Word sense disambiguation is a task to identify correct meaning of word by using some algorithm with the help of some or other approach<sup>[1]</sup>. To accomplish this process system is trained to identify correct results meaning of word according to the multiple words like Map. Map is a geographical representation of particular place or it is an association between two terms (Mapping). So problem statement is to identify the meaning of given word as per the requirement of user<sup>[2]</sup>.

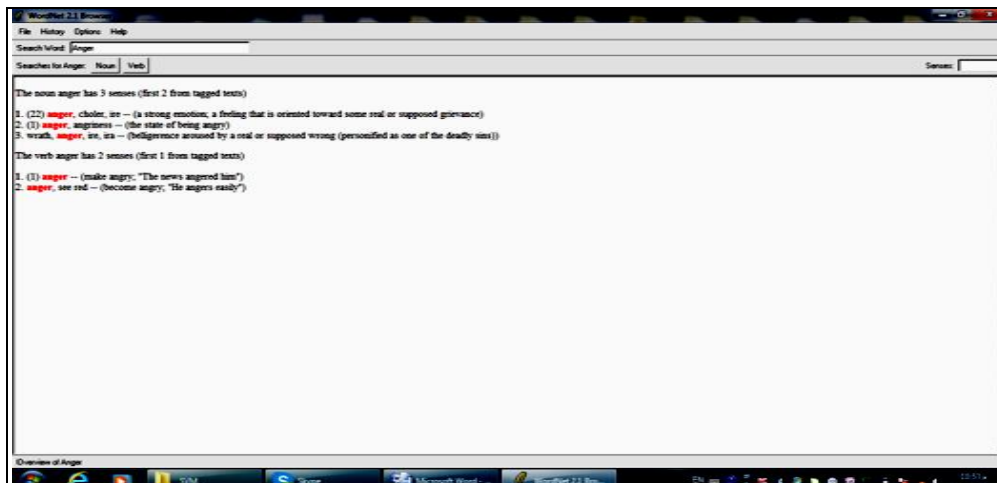


Fig. 1: The Screenshot Shows the Multiple of Anger Word

# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 12, December 2014

## II. BACK GROUND AND RELATED WORK

Many researchers have contributed to this field of disambiguation. There are various approaches to accomplish this task of disambiguation.

- **Support vector machine** is to generate a hyperplan to separate hyperspace by separating them as per the category or group. Distance between closest edges of plan is known as support vector<sup>[3]</sup>.
- **Naive Bayes** approach is a way to calculate posterior probability by using conditional probability. Naïve part of the classifier is to extract features dependency. It is assumed that there is no dependency among the features extracted<sup>[4]</sup>:

$$p(C_i | F_1, F_2) = \sum_{i=1}^n \frac{p(F_1, F_2 | C_i)}{p(F_1, F_2)}$$

Where:

F1, F2 are features

C is category.

- **Decision tree** deals with information gained during the experiment. In decision tree processing is from top to bottom that is from root to leaf. So if a length or tree is higher probability of data storage or information gain is comparatively higher. This also calculates the error rate in terms of entropy. Maximum is entropy minimum will be the accuracy and vice versa<sup>[5]</sup>.
- **Decision list** works on condition like (If-else) structure. If condition is satisfied visit the node deal with data otherwise leave it. Repeat the process till desired data or conditions are not meet<sup>[6]</sup>.

These approaches and their comparison is discussed in this paper based on the experiment which is performed to meet the goal of word sense disambiguation using effective approach for empirical retrieval of information.

### 1. Motivation

To address the challenge discussed earlier resinous efforts are needed because every approach facer some or other drawback. The figure below represent support vector machine approach implemented in this paper:

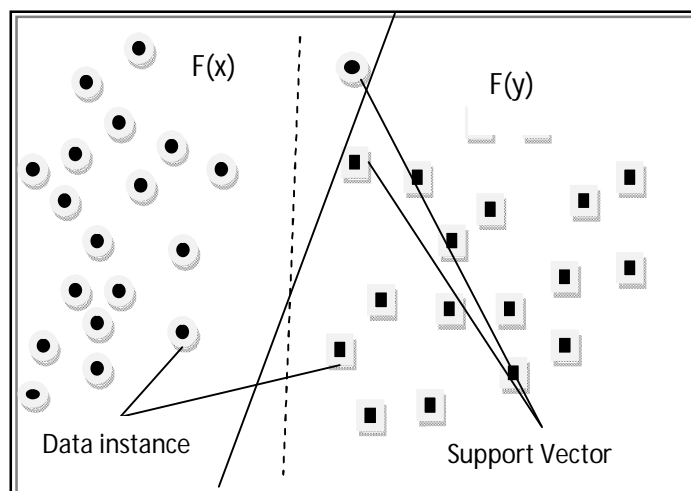


Fig 2. Support Vector Machine with Hyperplane

Where x and y are various categories on which the data instances are separated. So the motivation to conduct this experiment is to increase the overall accuracy, address word sense disambiguation by considering some classifier, which will train the database and identify meaning of word correctly out of total list of meanings which are provided. This task is carried out by referring the context to resolve disambiguation.



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 12, December 2014

## II. EXPERIMENTAL SETUP

### 1.1 Data

Experiment is conducted by using a WordNet repository, 10 nouns and 5 verbs<sup>[7]</sup>. To know the accuracy of sense context is designed by following senseval norms. This representation is made by using XML representation. With the help of algorithm and context mentioned in a database meaning of word is calculated. To accomplish this task semistructured and unstructured representation is used, because of the latency; that is the required to store and retrieve the data to and from database<sup>[8]</sup>.

### 1.2 Implementation Supervised Machine Learning Techniques

To identify meaning of word two types of techniques used, Supervised, unsupervised techniques. If a data is identified on the best is of frequency of occurrence then it is unsupervised approach; But all the time we cannot completely rely on unsupervised approach, because meaning could vary as per the context used and perception. Supervised technique, because system is trained with some defined context to predict meaning based on the surrounding word. Their predictions are made with suitable data mining algorithm like, Naïve Bayes, Decision tree algorithm, Decision List algorithm, and Support vector machine. These algorithms are munitions and empirically implemented in this paper, and the comparative analysis based on the accuracy of that algorithms to predict the meaning.

#### 4.4.1 Naïve Bayes

Naïve Bayes approach works on conditional probability. In some approaches it gives better result while in other approaches it does not deliver appropriate results.

There are few scores where Naïve bayes provide better result and these top 3 results according to the accuracy are: {Name: 1000, worlds: 1000, Day: 1000}.

In some cases performance of Naïve Bayes algorithm is not satisfactory lowest three such cases are: {Worship: 414, Trust: 167, Help: 414}

#### Box.(1): Naive Bayes Algorithm implemented on Our Data Set

1. Initialize context  $c$ , sense  $s$ , and ambiguous word  $w$ .
2. As per training context
3. 
$$p(s | w, c) = \frac{p(w | s, c) p(s | c)}{p(w | c)}$$

Calculate Maximized  $p(s | w, c)$
4. Select one with highest value
5. Map sense according to the highest accuracy.

Overall accuracy of Naïve Bayes algorithm is (58.32 %) which need to be improved to find but desired word correctly<sup>[9]</sup>.

#### 4.4.2 Decision Tree

Decision tree is based on storage of result or meaning at node. As far as WSD is concerned for data set that we are referring overall accuracy of decision tree is not satisfactory.

Overall accuracy is (45.14%).

# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 12, December 2014

## Box.(2) : C4.5 Algorithm implemented on Our Data Set.

1. Read data set and calculation POS (e.g. recompense.)
2. Prepare context containing various senses of word (e.g. Recompense- reward)
3. Calculate frequency at context (i.e. - p- and +P+)  
-P- Negative  
-P+ Positive
4. Calculate information gain for calculating entropy (S) = -  
 $P+\log_2 P+ - P-\log_2 P-$
5. Gain (S,A)= Entropy(S) -  $\sum_{v \in DA} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$
6. Select highest (Entropy, Attribute ratio)

Though overall accuracy of decision tree is not up to the mark but for few cases it gives better results, such top 2 cases are: {Name: 1000, Worlds:1000}. On this contrary, there are some result its where performance is not satisfactory such lowest three cases are: {Trust: 167, Day: 109, Help: 125}<sup>[10]</sup>.

### 4.4.3 Decision List

Among the approaches discussed, so for decision list provides more accurate result by forming if else ladder. The efficiency and accuracy would be noted by few cases where results are better are mentioned below: {Praise: 1000, Name: 1000, Worlds: 1000, Lord: 1000, Recompense: 1000, Day: 1000}.

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.

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

## Box (3): Decision List Algorithm implemented

Though overall accuracy is better in case of decision list there are some cases where the performance is not according to the expectation is not satisfactory are as below:

{Trust: 167, Help: 125, Favored: 250, Path: 333}<sup>[11]</sup>.

### 4.2.4 Support Vector Machine

Support vector machine is a technique to separate a data in a hyperspace with the help of hyperplane. This separation is done creating hyperplane by maximizing the distance between the data instances which are located at the edge. If we observer working of SVM carefully it is observed that it is practically difficult to sprat data instances clearly, so this gap is known as slack. This slack is to be maximized to separate data instances and categorize them under one heading. Support vector machine is an idea example of binary classifier but when it comes to word sense disambiguation performance or the results are not up to the mark.

Such top 4 “four cases” where results are at pick are mentioned below:

{Name: 1000, Worlds: 1000, Guide: 1000, Day: 1000}.

In some cases performance of Support vector machine algorithm is not satisfactory lowest three such cases are: {Worship: 414, Lord: 431, Trust, 167, Path: 318, Favored: 250 Help: 125}<sup>[12]</sup>.



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 12, December 2014

## Box.(4):7636SVM Algorithm implemented on Our Data Set

1. Data Set: five verbs, ten nouns <sup>[7]</sup>, and WordNet repository *with*  $x_i \in Pd$  and  $y_i \in \{-1,1\}$  learn a classification

2. Training data: Part of speech, user, context, and words are used to train the systems.

$$f(x_i) = \{ \geq 0 \quad y_i = \pm 1; < 0 \quad y_i = -1 \}$$

3. For linear approach

$$F(x) = w^T x + b$$

$w \longrightarrow$  Normal line

$b \longrightarrow$  base line

$w^T \longrightarrow$  weight vector

4. Processing of Algorithm application<sup>[8]</sup>:

For all sense of word decide meaning with high score

Where  $m \longrightarrow$  meaning

5. Testing: By delivering final result with accuracy and score helps to decide performance of algorithm<sup>[9]</sup>.

## III. THE RESULTS

Disambiguation is performed in this paper via a four supervised approaches, using WordNet and Senseval-3. Table (1), shows the results of four approaches, Naïve Bayes, Decision tree, Decision List, and Support vector machine, which's has been given based on their score and accuracy.ich shows a comparative of those different approaches has been given based on their score and accuracy.



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 12, December 2014

**TABLE 1.**  
**DATA SET OF WORDS AND RESULTS OF SUPERVISED LEARNING MACHINE CLASSIFIERS**

Word	POS	# Sense	Naïve Bayes		Decision Tree		Decision List		SVM	
			Score	Accuracy	Score	Accuracy	Score	Accuracy	Score	Accuracy
Praise	n	2	0.408	0.592	405	593	668	1000	592	594
Name	n	6	0.189	1.0	184	1000	1000	1000	189	1000
Worship	v	3	0.172	0.414	308	425	387	500	352	414
Worlds	n	8	0.137	1.0	1000	1000	142	1000	1000	1000
Lord	n	3	0.341	0.681	187	426	489	1000	418	431
Owner	n	2	0.406	0.594	405	595	755	999	592	594
Recompe-nse	n	2	0.48	0.594	405	595	791	1000	592	594
Trust	v	6	0.167	0.167	167	167	167	167	167	167
Guide	v	5	0.352	0.648	199	247	387	995	244	1000
Straight	n	3	0.496	0.504	462	462	500	500	69	465
Path	n	4	0.415	0.585	316	316	333	333	47	318
anger	n	3	0.412	0.588	462	462	500	500	69	465
Day	n	10	0.109	1.0	109	109	111	1000	109	1000
Favored	v	4	0.587	0.648	250	250	250	250	250	250
Help	v	8	0.352	0.414	125	125	125	125	125	125

**TABLE.2**  
**THE FINAL RESULTS OF SUPERVISED LEARNING MACHINE CLASSIFIERS**

Approaches	Accuracy (%)
Naïve Bayes	58.32
Decision Tree	45.14
Decision List	69.12
SVM	56.11

## IV. CONCLUSION

We have presented a comparative study for four supervised learning machine algorithms, using WordNet, and Senseval-3, table (2), below shows the final results and accuracy for each approach. In conclusion, Decision List algorithm, obtained high accuracy.



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 12, December 2014

## ACKNOWLEDGMENT

I'm grateful to my research guide respected Dr. Shashank Joshi (Professor at Bharati Vidyapeeth University, College of Engineering) for his support and cooperation all the time.

## REFERENCES

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. 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. Yoong Keok Lee and Hwee Tou Ng and Tee Kiah Chia, Supervised Word Sense Disambiguation with Support Vector Machines and Multiple Knowledge Sources, 2007.
4. Ruben Izquierdo and Armando Suarez, German Rigau. An Empirical Study on Class-based Word Sense Disambiguation. Paper has been supported by the European Union under the projects QALL-ME (FP6 IST-033860) and KY-OTO (FP7 ICT-211423), and the Spanish Government under the project Text-Mess (TIN2006-15265-C06-01) and KNOWS (TIN2006-15049-C03-01).
5. Mahesh Joshi, Ted Pedersen and Richard Maclin. A Comparative Study of Support Vector Machines Applied to the Supervised Word Sense Disambiguation Problem in the Medical Domain. Bhanu Prasad (Editor): IICAI-05, pp.3449-3468, 2005.
6. Vicky Panagiotopoulou, Iraklis Varlamis, Ion Androutsopoulos, and George Tsatsaronis, Word Sense Disambiguation as an Integer Linear Programming Problem., SETN 2012, 7<sup>th</sup> Hellenic Conference on Artificial Intelligence, May 28-31, 2012, Lamia.
7. <http://www.e-quran.com/language/english>.
8. <http://www.senseval.org/senseval3>.
9. 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: [www.iaeme.com/IJCET.asp](http://www.iaeme.com/IJCET.asp), Journal Impact Factor (2014): 8.5328 (Calculated by GIS), [www.ijfactor.com](http://www.ijfactor.com).
10. 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](http://www.ijates.com), Volume No.02, Special Issue No. 01, September 2014 ISSN (online): 2348 – 7550.
11. Boshra F. Zopon AL\_Bayaty, Shashank Joshi, SENSE IDENTIFICATION FOR AMBIGUOUS WORD USING DECISION LIST, International Journal of Advance Research In Science And Engineering <http://www.ijarse.com>, Vol. No.3, Issue No.10, October 2014 ISSN-2319-8354(E).
12. 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](http://www.ijiset.com) ISSN 2348 – 7968