IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 1 Issue 9, November 2014.
www.ijiset.com
ISSN 2348 – 7968

Implementation SVM to Solve Multiple Meaning of Word Problem

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

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

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

Abstract

Word sense disambiguation is a task to resolve or to identify correct meaning of word. There are many approaches to identify word with exact sense like Naïve Bayes approach, Decision Tree, Decision List, SVM etc. Support vector machine is nonlinear binary classifier used to separate the data with hyper plane. In this case separate classifier is designed for each case that is each sense of given word. Classifier is responsible for separating favorable data and unfavorable data drawing hyper plane. In this approach meaning is selected in such a way to maximize the distance between two instances belong to separate categories. The result reported in this work achieved it was (56.11%) accuracy according to the senseval 3.

Keywords: *Support Vector Machine, Supervised learning approaches, Senseval-3, WSD, WordNet.*

1. Introduction

Word sense disambiguation is to resolve meaning of given word with the help of an algorithm selected and approached used to decide meaning collectively. In the experiment which is implemented used to provide set of such disambiguated word along with feature and POS "part of speech". To perform the connectivity jwnl java WordNet library is used [1], and to connect java program with WordNet and to communicate with WordNet environment [2]. To influence scoring there must be some way to decide correct meaning of a word. In our case senseval-3 is used to frame context to extract meaning of word senseval-3 is nothing but XML presentation at of word, meaning, POS and text [3].

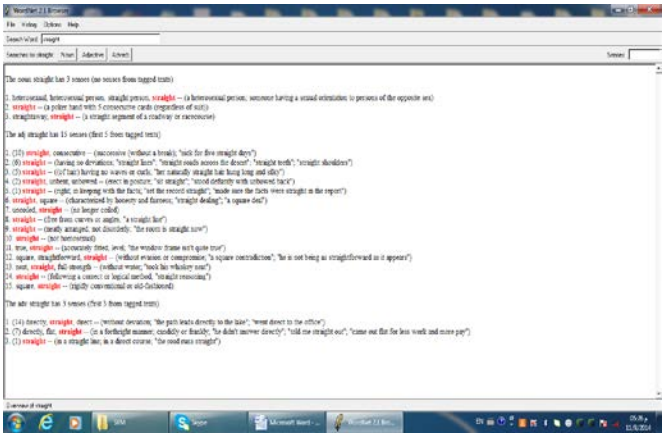

Fig. 1: The Screenshot Shows the Multiple of Straight Word

2. Support Vector Machines(SVM)

Support vector machine is one of 10 top algorithms according to the IEEE International conference on data mining (ICDM) in December 2006[4]. SVM technique is to construct N dimensional hyper plane to separate data into two categories, SVM is used to solve quadratic problem related with programming using linear constraints. Some important terms pertaining to SVM:

1- ***Attribute:*** Variable used for prediction is known as attribute.
2- ***Feature:*** Transformed attribute used to mention hyperplane is known as feature.
3- ***Support Vector:*** Vectors which are near to hyperplane are a support vector which helps to separate one category from other with the help of margin. In this way categories are made with the help of plan by separating vectors from various categories from each other [5].

ISSN 2348 – 7968

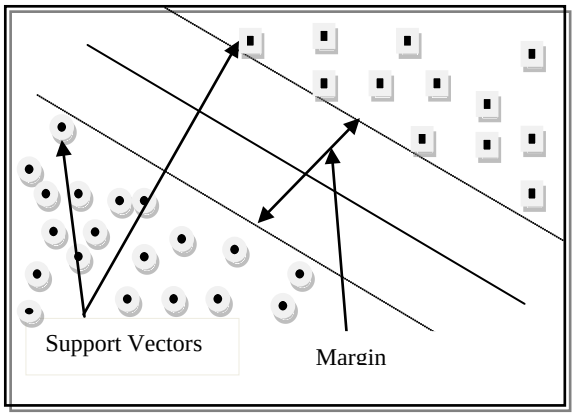

Fig. 2: Dispersion of vectors

3. Application of SVM in NLP

SVM is binary classification got huge application in machine learning Knowledge Engineering discovery medical pattern recognition systems.
SVM is ways learner but provide accurate result (comparatively) so it generally combined with decision tree [hybrid SVM based decision tree]. Important advantage is boundaries clearly defined SVM. SVM could be optimized by using kernel increaser the accuracy and deliver result in less time [6]. TSVM: Transduction (SVM) is to train system with labeled example to increase the scopes of SVM by introducing pulse.

4. Feacher Selection

To resolve meaning of word external factor apart from algorithm is required for word sense disambiguation of given meaning used that external factor as a context.
This context follows standards meaning by snseval-3, which contains:

- Word.
- Meaning.
- Context.
- POS.

5. Methodology

1. Data Set: five verbs, ten nouns [7], and WordNet repository

   $with\ xi \in Pd\ and\ yi \in \{-1,1\}\ learn\ a\ classification$

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

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

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[8]:

   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 [9].

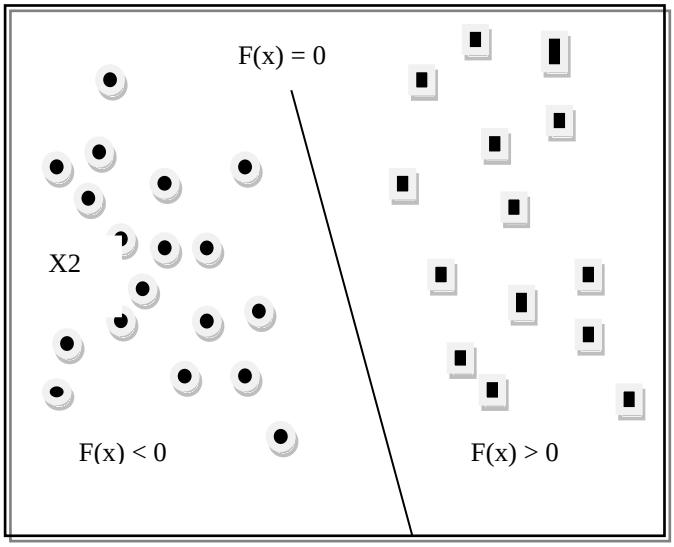

Fig. 3: Geometric Distribution of Vectors

6. Implementation using Java

This is part of java program which is an implementation of dynamic classification using SVM. It developed algorithm comprises of training model, calculation model, output delivering module. Below part of training phase Java code:

```
void train(String category, char[] cs, int
start, int end) {
        train(category,new
String(cs,start,end-start));
    }

void train(String category, CharSequence
sampleCSeq) {
```

ISSN 2348 – 7968

```
        train(category,sampleCSeq,1);
    }
public    void    train(String    category,
CharSequence sampleCSeq, int count) {
        if (count < 0) {
            String msg = "Counts must be
non-negative."
                + " Found count=" + count;
            throw                          new
IllegalArgumentException(msg);
        }
        if (count == 0) return;

languageModel(category).train(sampleCSeq,c
ount);

categoryDistribution().train(category,coun
t);
    }
```

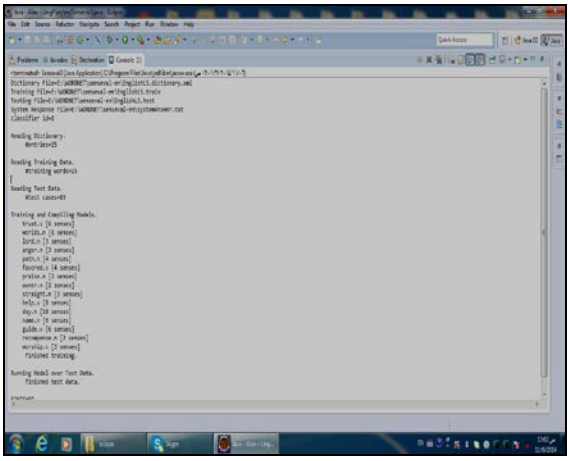

Fig 4: The Screenshot Shows Training and compilation Model

7. Error Ret Calculation

Error rat a observed for SVM for given dataset is as mentioned below (which is minimum error rate).

Given some training data D where D is context in the implementation:

$$D = \{(x_i, y_i) \mid x_i \in R^p, y_i \in \{-1,1\}^n{}_{i=1} \text{-----} (1)$$

Where: $k(x, y)$ is Kernel function, yi ranges 1 or -1 denotes the category from which xi belongs, xi is a real vector. Aim is to find margin with maximum value.

Kernel function is function which is accepting.

$$\frac{meaning1}{w.x - b = 1} \text{-----} (2)$$

$$\frac{meaning2}{w.x - b = -1} \text{-----} (3)$$

For given error rate (e) ------ error rate.

$$y_i(w.x_i - b) \geq 1 - e \text{ Where } i \leq i \leq n \text{-----} (4)$$

Overall error rate of two cases is equal to zero day, name, and guide. Average error rate for given words = 0%, which is excellent.

Disambiguation based on the meaning can be with kelp of training dataset (context), algorithm (SVM) and it is optimized by reducing error.

TABLE.1: Error Rate Calculation for Each Word

No.	Word	Error
1	Lord	569
2	Praise	406
3	Owner	406
4	Recompense	406
5	Straight	535
6	Anger	535
7	Path	682
8	Worlds	0
9	Day	0
10	Name	0
11	Worship	586
12	Guide	0
13	favored	750
14	Trust	833
15	help	875

8. Evaluation

SVM approach is popular approach in data mining [10]. In the experiment conducted meaning extraction according to the context provided is (56.11%). In the extraction, context and training phase plays very vital role, so it carrier reflexes on the accuracy as well, so at this stage it is difficult to talk about correctness.

ISSN 2348 – 7968

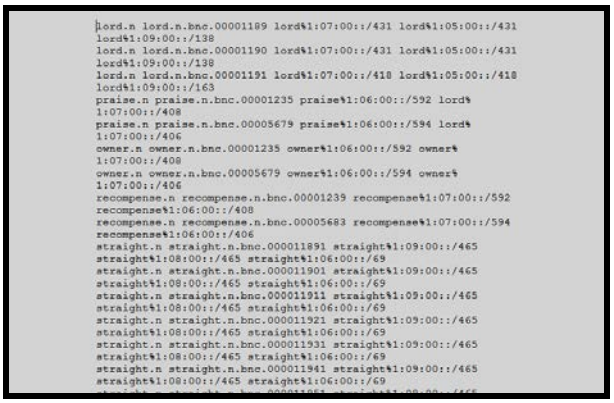

Fig 5. The Screenshot Shows The System Answer.Txt File Compilation Model

9. The Result

By looking at result acquired some of the results are 100% accurate like World (100%), Day (100%), Name (100%), Guide (100%), where Trust (167%), Help (125%), Favored (250%), gives comparatively less results overall accuracy is (56.11%) and score is approximately (32%). Still overall accuracy of SVM is above (50%) which is acceptable. The results for our dataset shown in table (2) below:

TABLE.2: Data Set of Words and Results of SVM Classifier

Word	POS	# Senses	Score	Accuracy
Praise	n	2	592	594
Name	n	6	189	1000
Worship	v	3	352	414
Worlds	n	8	1000	1000
Lord	n	3	418	431
Owner	n	2	592	594
Recompense	n	2	592	594
Trust	v	6	167	167
Guide	v	5	244	1000
Straight	n	3	69	465
Path	n	4	47	318
Anger	n	3	69	465
Day	n	10	109	1000
Favored	v	4	250	250
Help	v	8	125	125

10. Conclusions

We have implemented Support Vector Machine algorithm using WordNet 2.1. Word sense Disambiguation gives score and accuracy which is based on external factor like context. This context is nothing but paragraph in XML format according to senseval format. Overall accuracy of SVM approach is (56.11%) according to the senseval-3, which is acceptable still there is scope for modification to correctly identify meaning of word.

Acknowledgment

I would like to thank my research guide respected Dr. Shashank Joshi (Professor at Bharati Vidyapeeth University, College of Engineering) for his cooperation with me all this period.

References

[1] Steve Holzner, Eclipse, programming Java applications, Third Indian Reprint, 2007.

[2] http://wordnet.princeton.edu.
[3] http://www.senseval.org/senseval3.

[4] XindongWu · Vipin Kumar · J. Ross Quinlan · Joydeep Ghosh · Qiang Yang, Hiroshi Motoda · Geoffrey J. McLachlan , Angus Ng , Bing Liu, Philip S. Yu , Zhi-Hua Zhou , Michael Steinbach, David J. Hand, Dan Steinberg, © Springer-Verlag London Limited 2007.

[5] M.Arun kumar , M.Gopal, A comparative study on multiple binary class SVM methods for unable text categorization, control group, department of electronic engineering, India,2010.

[6] www.si.upc.edu%~scudero/wsd/06-teji.pdf
[7] http://www.e-quran.com/language/english.
[8] http://www.ox.ac.uk.
[9] Miller, G. et al., 1993, Introduction to WordNet: An On-line Lexical Database, ftp://ftp.cogsci.princeton.edu/pub/wordnet/5papers.pdf, Princeton University.
[10] Yoong Keok Lee and Hwee Tou Ng and Tee Kiah Chia, Supervised Word Sense Disambiguation with Support Vector Machines and Multiple Knowledge Sources, 2007.

First Author : 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

ISSN 2348 – 7968

University in Iraq/ Baghdad. Her research interests include software engineering.

Second Author: 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