Performance Improvement of Context Identification for Human Computer Interaction

Priti Saktel[1], Minal Domke[2], Lekhraj Vilhekar[3]

[1, 2]Department of Computer Application, Shri. Ramdeobaba College of Engineering and Management, Nagpur, MS, India

[3] Department of Information Technology, Government Polytechnic, Nagpur, MS, India

Abstract: ***Context Identification is a task of identifying intended sense (meaning) of word based on context, has been a prominent research work of Natural Language Processing for Word Sense Disambiguation (WSD). Human Computer Interaction (HCI) is useful to improve users and computers interactions by making it more usable. For this improvement, combination of Supervised and Unsupervised WSD methods are used. Under this framework, the words from ambiguous sentences have categorized for finding the appropriate sense of given word, amounts to correct domain of word among the number of domain representing its correct sense. While interacting with the system, sentence or instruction provided to the computer should be well analyzed and understood properly, such that there should be no confusion. It is useful for Human Computer Interaction (HCI) as a self learning process or language which provides people with the ability to explore themselves. For effective disambiguation, these methods find to be more helpful in the various areas that demands human computer interaction. Also, it motivates the people of ruler areas for self learning English language. In this paper, the results of unsupervised learning are reported. Also, the accuracy of this work is calculated with the aim of finding best suitable domain of word for WSD. It shows that combination of supervised and unsupervised approach improves accuracy.***

Keywords: Ambiguity, Context Identification, HCI, Supervised Training, Unsupervised Learning**.**

1. Introduction

Human Computer Interaction is a process which provides user a platform to interact or communicate with the machine. A Self learning Language is helpful to learn the English language in ruler areas where people could not able to go to school for learning English. To resolve an ambiguity in a sentence, natural language processing provides Word Sense Disambiguation which identifies correct sense out of multiple meanings of a word in a sentence [1]. WSD is a process which identifies the correct sense of a word with the help of surrounding words in a sentence. From the context of the sentence, correct sense of word is obtained. Based on context, we associate a different meaning of the single word in each sentence. Thus, if the word relationship appears near the word doctor and patient, we can say that its meaning related to _Doctrines‘ and not _Education‘ which is known as sentential context[2]. At a onetime Computer that read words, must use a process called word sense disambiguation to find the correct meaning of a word [3].

Under this framework, the database is created to store Domains, General words & Meanings. Also, the POS (Part-of-Speech) Tagger process is implemented to separate the content words. In this, the separation of words is done as step 1 and the target word is picked up from content words as step 2.

After, these steps three categories of the words are created as C1, C2 and C3. C1 indicates separated content words, C2 indicates assigned domain of words stored in database and C3describes maximum count of domain based on context. Based on this the paper had published [4]. Afterwards, context based domain identification was done to resolve ambiguity, which was published [18]. Also, various comparisons are performed to obtain correct domain of word. Before performing these comparisons the domain is distributed to words using database. The system is trained using supervised training. Apart from this the spell checker utility is implemented for storing updated entries into the database and it was published [19]. This paper focuses on, unsupervised learning to obtain the correct domain of word by the system automatically. It is described in section 1. The evaluation of this work is discussed in section 2 and section 3 shows the result of unsupervised learning.

2. Experimental Work

In this work, the MySQL database is used for unsupervised learning. This database contains Domain, General Word and Meanings table. The database tables are shown below in Fig.1.

ID	*Domain*
3	Medical
5	Factotum
12	Education
62	Doctrines
86	Social science

ID	*General Word*
70	Is
71	The
72	Was
81	That

ID	**Word**	**FieldID**
147	fruits	1
148	fruits	4
151	doctor	3
152	doctor	62
158	doctor	12
161	Patient	62
162	trust	86
163	relationship	5

Figure 1: MySQL Database (Domains, General words, Meanings)

The Stanford POS Tagger is used to tag the words. The output of this tagger is in the following form.

The|DTR doctor|DTR patient|NN relationship|NN is|NN based IN on|NN trust.|.

From these only the content words were selected [5]. The domains corresponding to these words were obtained from the database. The ID corresponding to the domain was obtained from the Domains table.

This paper describes the techniques used to assign the correct domain using synsets from WordNet domain. To find the meaning of word in the given context from all word domain in database. Here, the synonym relationship is investigated [6].

B. Unsupervised Learning Method

In this, the sentence is given as input. The domain from domain distribution table (MySQL Database) is assigned to words of sentence. Afterwards comparison is done for best suitable domain. Comparison is done based on context of sentence and if comparison gives maximum count for intended domain then that domain of word is displayed. If it is correct that is considered as correct domain of word (disambiguation) and this entry is updated in the database. Else, user has given the chance to input the sentence again. This flow is shown in Fig. 2. The knowledge acquisition bottleneck problem is overcome by unsupervised learning, since it is independent of manual work.

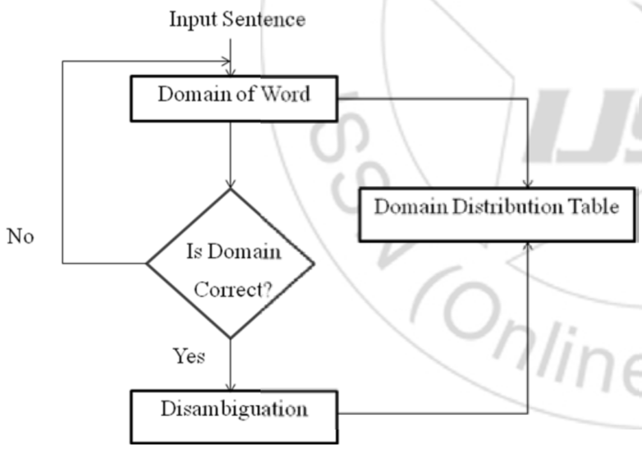

Figure 2: Unsupervised Learning Flow

The experimental setup is done by following steps and accuracy of unsupervised, supervised and proposed hybrid method is evaluated using mathematical formula as

C. Synonym Relationship Approach

The doctor patient relationship is based on trust. After, the processing of POS tagger will pick two words as doctor and patient. Here, the ambiguity is in word doctor and patient, it has 2 FieldIDs which is shown in Table 1 below.

The table describes multiple domains for a word. Table 1 clearly showing, the word doctor has 3 domains as Education, Medical and Doctrines. Similarly, for patient domain is Doctrines. So, doctor and patient are related to ‗Doctrines' since context of the sentence [15].

Table 1: Domains Comparison

FieldID	Word	Domain
12	Doctor	Education
3	Doctor	Medical
62	Doctor	Doctrines
62	Patient	Doctrines
5	Relationship	Factotum
86	Trust	Social_science

$$\frac{\sum_t \text{Number of Correct terms}}{\sum_i \text{Number of Input}}$$

Where, t=correct terms (Correctly disambiguated)
& i= input (Number of sentences)

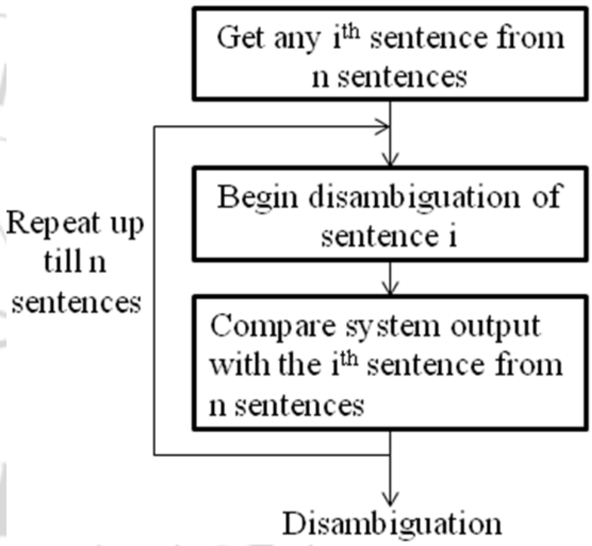

Figure 3: Experimental Setup Steps

Repeat the above steps for:
i=1…number of sentences (n), n=1….15
Where,
i indicate sentence and n indicates number of sentences.

3. Results and Discussions

The result of unsupervised learning as ―Identification of domain" is described in Figure 5 below. When the sentence is entered by user firstly, the separation of each word is done. Then target word is picked for domain distribution. To find correct domain the comparison is done.

For example

Max Value: 2
For field ID: 4

After this, the domain is checked for correctness, if it is correct then system is displaying correct domain or the sentence is entered again.

Sentence: The play of the imagination.

Step 1: Separating All Words
Word: The
Word: play
Word: of
Word: the
Word: imagination.

Step 2: Finding Matching Domain
Match - play: play
Match - play: play
Match - play: play
Match - play: play
Match - play: play
Match - imagination: imagination.
Match - imagination: imagination.

Step 3: Checking for Best Probable Field
Field 11 found 2 times
Field 2 found 2 times
Max Value: 9 For field ID: 69
The Domain is Free_time

Step4: Checking for Correctness
Is this the type of the sentence at input? Y/N
The new elements with selected domains have been updated.

Figure 4: Unsupervised Learning

Figure 6 (See at the end of the paper)

The ambiguity resolvation was evaluated using 15 sentences. These sentences are manually disambiguated using MySQL database based on WordNet domain functionality. These sentences contained 31 target words. Out of which we could disambiguate 30 target words. The accuracy of our approach was 63%, which means that our system disambiguated correctly 19 out of 31 target words. Table 2 shows the accuracy of unsupervised learning method of our system and Fig.5 is the column graph showing this method accuracy across the sentences. The red color columns are indicating accuracy and blue color column indicates sentences.

Table 2: Results of Unsupervised learning method accuracy of 15 sentences

Sentence	*Target word*	*Disambiguated*	*Correctly disambiguated*	*Accuracy (%)*
1	2	2	1	50
2	3	3	2	67
3	1	1	1	100
4	1	1	1	100
5	2	2	1	50
6	2	2	2	100
7	2	2	1	50
8	1	1	1	100
9	1	1	1	100
10	3	3	1	33
11	3	3	2	67
12	2	2	1	50
13	3	3	2	67
14	3	2	1	50
15	2	2	1	50
Total	31	30	19	63

In Supervised learning method, these sentences contained 29 target words. Out of which we could disambiguate 29 target words. The accuracy of this method was 76%, which means that our system disambiguated correctly 22 out of 29 target

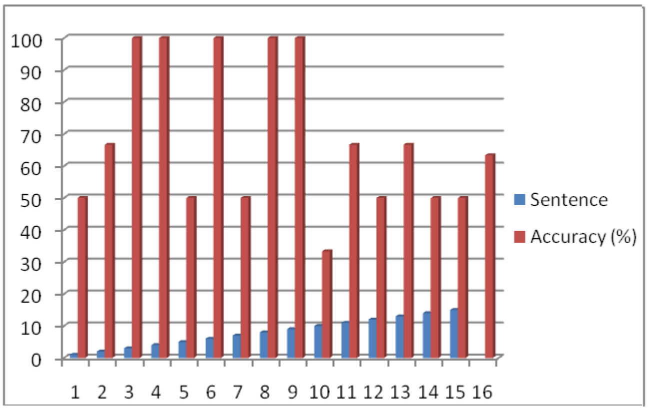

Figure 5: Unsupervised learning method graph accuracy

words. Table 3 shows the accuracy of supervised method and Fig.6 below shows graph of this method.

Table 3: Results of Supervised learning method accuracy of 15 sentences

Sentence	*Target word*	*Disambiguated*	*Correctly disambiguated*	*Accuracy (%)*
1	2	2	2	100
2	3	3	2	67
3	1	1	1	100
4	1	1	1	100
5	2	2	1	50
6	2	2	2	100
7	2	2	1	50
8	1	1	1	100
9	1	1	1	100
10	2	2	1	50
11	3	3	2	67
12	1	1	1	100
13	3	3	2	67
14	3	3	3	100
15	2	2	1	50
Total	29	29	22	76

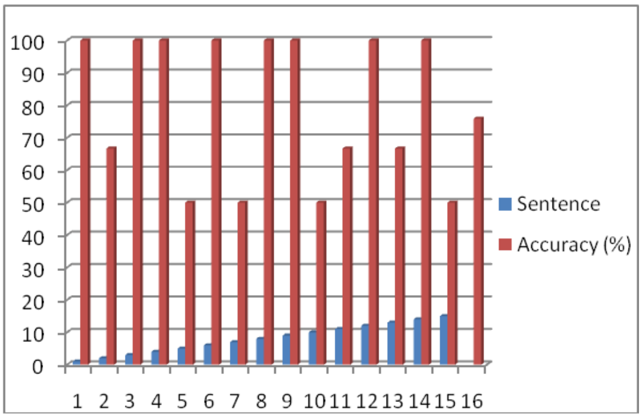

Figure 6: Supervised Learning method graph accuracy

The sentences for hybrid method contained 27 target words. Out of which we could disambiguate 25 target words. The accuracy of hybrid approach was 80%, which means that our system disambiguated correctly 20 out of 27 target words. Table 4 shows the hybrid accuracy of our system and Fig.7 describes the graph of hybrid method accuracy.

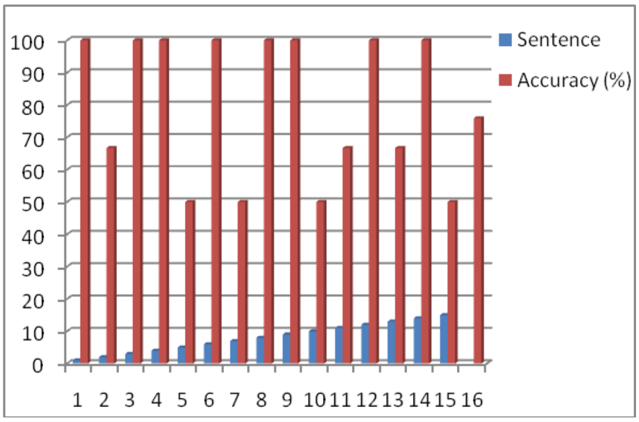

Figure 7: Hybrid learning method accuracy

Table 4: Results of Hybrid learning method accuracy of 15 sentences

Sentence	*Target word*	*Disambiguated*	*Correctly disambiguated*	*Accuracy (%)*
1	2	2	2	100
2	3	3	2	66.67
3	1	1	1	100
4	1	1	1	100
5	2	2	1	50
6	2	2	2	100
7	2	2	1	50
8	1	1	1	100
9	1	1	1	100
10	2	1	1	100
11	3	3	2	66.67
12	1	1	1	100
13	3	3	2	66.67
14	1	1	1	100
15	2	1	1	50
Total	27	25	20	80

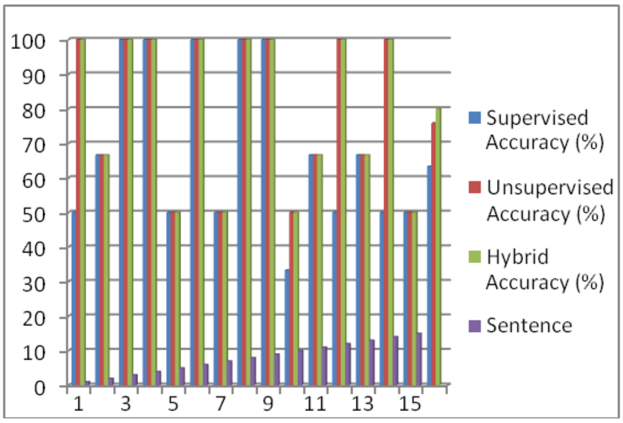

Figure 8: Comparison of unsupervised, supervised, hybrid accuracy

The comparison of unsupervised, supervised and hybrid learning method with the same sentences are shown in Table 5 and graph of this method shown in Fig.8. In which unsupervised method gives 63%, supervised method gives 76% and hybrid method gives 80% of accuracy.

Table 5: Results of comparison of unsupervised, supervised and Hybrid learning method accuracy of 15 sentences

Sentence	*Target word*	*Disambiguated*	*Correctly disambiguated*	*Supervised Accuracy (%)*	*Unsupervised Accuracy (%)*	*Hybrid Accuracy (%)*
1	2	2	1	50	100	100
2	3	3	2	67	67	67
3	1	1	1	100	100	100
4	1	1	1	100	100	100
5	2	2	1	50	50	50
6	2	2	2	100	100	100
7	2	2	1	50	50	50
8	1	1	1	100	100	100
9	1	1	1	100	100	100
10	3	3	1	33	50	50
11	3	3	2	67	67	67
12	2	2	1	50	100	100
13	3	3	2	67	67	67
14	3	2	1	50	100	100
15	2	2	1	50	50	50
Total	31	30	19	63	76	80

Sentence	*Separation of Words*	*Target Word*	*Domain Identification*	*Comparison*	*Final Domain*
The doctor patient relationship is based on trust.	The doctor Patient relationship is based on trust	Match – doctor: doctor Clustered under Match –patient: patient Clustered under Match – relationship: relationship Clustered under Match – trust: trust. Match – trust: trust.	Education Medical Doctrines Doctrines Factotum Social_science	Max Value :02 For field ID: 62	Doctrines

Figure 9: Result OF Unsupervised Learning

4. Conclusion

This paper improves the accuracy of identifying the correct domain of word. As per the Table 5 it shows that self learning language is improved by obtaining correct sense of a word by removing ambiguity from a sentence with full automation. Also, improves disambiguation process by obtaining appropriate sense of a word. Hence, sentence comprised of various content words. The synonym relationship approach is used to identify context of the sentence. The system is trained using supervised training to check correctness of domain which gives 76% of accuracy; an unsupervised learning is used to update the database with the selected sentences and word-meaning pairs automatically. It gives 63% of accuracy. The hybrid method improves this accuracy up to 80% from Table 5.In this, when the number of target word is correctly disambiguated system gives 100% accuracy. Else, the accuracy may be 66% or 50%. Hence, the overall 80% accuracy is evaluated. These results are beneficial for Human Computer Interaction as it is motivating people to learn the language by themselves. Additionally, the spell checker utility is implemented to avoid mistakes in words.

References

[1] Roberto Navigli, Mirella Lapata (2010), ―An Experimental Study of graph connectivity for unsupervised word sense disambiguation", IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 32, No. 4.
[2] Myunggwon Hwang, Chang Choi (2011)," Automatic Enrichment of Semantic Relation Network and Its application to Word sense Disambiguation",IEEE transaction on knowledge and data engineering, vol. 23, no. 6.
[3] Francisco Tacoa, Hiroshi Uchida (2010),"A Word Sense Disambiguation approach for converting Natural LanguageText into a Common Semantic Description", Fourth International Conference on Semantic Computing, IEEE.
[4] Priti Saktel, Urmila Shrawankar (2012), ―Context based Meaning Extraction for HCI Using WSD Algorithm", IEEE-International Conference on Advances in Engineering, Science and Management ICAESM, March 30, 31, pg. no. 208-212.
[5] Ping Chen, Wei Ding (2010), ―Word Sense Disambiguation with Automatically Acquired Knowledge",IEEE Intelligent Systems.
[6] Ling Che Yangsen Zhang (2011),"Study on Word Sense Disambiguation Knowledge base based on Multi-Sources", IEEE.
[7] E. Agirre and P. Edmonds (2006)," Word sense Disambiguation: Algorithms and Applications", Springer.
[8] Johan Bos and Malvina Nissim (2009)," From shallow to deep Natural language processing: A hands-on tutorial", Springer.
[9] Yousif, J.H. Sembok, T.(2008),"Arabic part-of-speech tagger based Support Vectors Machines", Information Technology,2008. ITSim 2008. International Symposium.
[10] Diana McCarthy, Rob Koaling(2007),"Unsupervised Acquisition of Predominant Word Senses," Computational Linguistics, Vol.33, No.4.
[11] Andrei Mincă, Ştefan Diaconescu (2011),"An Approach to Knowledge-Based Word Sense Disambiguation Using Semantic Trees Built on a WordNet Lexicon Network",IEEE.
[12] Reza Soltanpoor, Mehran Mohsenzadeh,Morteza Mohaqeqi (2010),"A New Approach for Better Document Retrieval and Classification Performance Using Supervised WSD and Concept Graph", First International Conference on Integrated Intelligent Computing, IEEE.
[13] Jerome R. Bellegarda, Fellow (2010)," Part-of-Speech Tagging by Latent Analogy", IEEE Journal of Selected Topics In Signal Processing, Vol. 4, No. 6, Dec. 2010.
[14] Yuhang Guo, Wanxiang Che (2010),"Semi-Supervised Domain Adaptation for WSD: Using a Word-by-Word Model Selection Approach, ―Proceedings 9th IEEE International Conference on Cognitive Informatics, ICCR10.
[15] Wanxiang Che, Ting Liu(2010),"Using Word Sense Disambiguation for Semantic Role Labeling", IUCS 2010, IEEE.
[16] Mirella Lapata, Chris Brew (2004),"Verb Class Disambiguation Using Informative Priors", Computational Linguistics, Vol.30, No.1, Springer.
[17] Diana McCarthy, Rob Koaling(2007),"Unsupervised Acquisition of Predominant Word Senses," Computational Linguistics, Vol.33, No.4.
[18] Priti Saktel, Urmila Shrawankar (2012), ―Context based Domain Identification for Resolving Ambiguity", IEEE-

Third International Conference on Computing Communication & Networking Technologies ICCCNT, July 26-28, pg. no. 1-5.

[19] Priti Saktel, Urmila Shrawankar (2012), ―An Improved Approach For Word Ambiguity Removal", International Journal of Human Computer Interaction IJHCI, Vol. 3:Issue 3, pg. no.71-82.