

Verb Sense Disambiguation by Measuring Semantic Relatedness between Verb and Surrounding Terms of Context

Arpita Dutta¹, Samir Kumar Borgohain²

Department of Computer Science and Engineering National Institute of Technology Silchar
Assam-788010, India

Abstract—Word sense disambiguation (WSD) is considered an AI complete problem which may be defined as the ability to resolve the intended meaning of ambiguous words occurring in a language. Language has complex structure and is highly ambiguous which has deep rooted relations between its different components specifically words, sentences and paragraphs. Incidentally, human beings can easily comprehend and resolve the intended meanings of the ambiguous words. The difficulty arises in building a highly accurate machine translation system or information retrieval system because of ambiguity. A number of algorithms have been devised to solve ambiguity but the success rate of these algorithms are very much limited. Context might have played a decisive role in human judgment while deciphering the meaning of polysemic words. A significant number of psychological models have been proposed to emulate the way the human beings understand the meaning of words, sentences or text depending on the context. The pertinent question that the researchers want to address is how the meanings are represented by human beings in mental memory and whether it is feasible to simulate with a computational model. Latent Semantic Analysis (LSA), a mathematical technique which is effective in representation of meanings in the form of vectors that closely approximates human semantic space. By comparing the vectors in the LSA generated semantic space, the closest neighbours of the word vector can be derived which indirectly provides lot of information about a word. However, LSA does not provide a complete theory of meaning. That is why psychological process modules are combined with LSA to make the theory of meaning concrete. Predication algorithm with LSA was proposed by Kintch, 2001 which was sufficient to capture various word senses and was successful in homonym disambiguation. Meaning of a word might have multiple senses specifically verbs. For example, verb “run” has 42 senses in WordNet. In order to find the correct sense of a verb is really a daunting task and resolving verb ambiguity using psycholinguistic model is very much limited. The proposed method has exploited the high dimensional vector LSA space resulted from training samples by applying predication algorithm to derive the most appropriate semantic neighbours for the target polysemous verb from the semantic space. Finally the vector space of test samples are checked with the training samples i.e. semantic neighbours to classify the senses of polysemous words in accurate manner.

Keywords—Word sense disambiguation; ambiguous verb; context; semantic space; latent semantic analysis; polysemy; machine translation

I. INTRODUCTION

Even though the research in Word Sense Disambiguation (WSD) has been carried out by researchers from 1940[1] onwards but still the problem is not resolved fully. The ambiguity is present in almost all the natural languages spoken in the world which sometimes makes it difficult to get the correct meaning or sense of a word in the context. Human beings are well organized to understand the meaning of ambiguous words, but in case of machines it requires a mechanism that will help the machine to find out the correct meaning of ambiguous words [2]. For example, ambiguous noun “plane”, “The plane flies like a bird in the sky” where the surrounding terms fly, bird, sky can help to recognize the ambiguous term ‘plane’ is an aeroplane whereas for the example, “the plane is made of paper” where the term paper can identify that “plane” is a geometric plane. Now, if these two examples are given as input text in a computer for machine translation, it is difficult to assume which sense of plane will be considered for translation. If exact meaning of ambiguous term cannot be predicted then the correct meaning of the sentence will be altered. So, the surrounding terms of the ambiguous term must be determined in order to get the true sense of the term. For instance, the word “piggy bank” is related to coin or money that means these terms help to find out the exact sense of bank as it is an ambiguous word.

WSD is one of the most challenging area in the research field of Natural Language Processing dated back to 1940s [3, 4]. There are different approaches to WSD problem such as knowledge base, supervised, unsupervised, semi supervised and hybrid approaches. Knowledge-based approaches were based on different knowledge resources such as machine readable dictionary or thesauri etc. where WordNet [5] is mostly used as a machine readable dictionary in this field. Most of the WSD works are based on different techniques of supervised approach which consists of training and testing dataset. Training dataset of supervised approach is used for classifier to learn and it comprises of target ambiguous words. In contrary, unsupervised approach does not depend on external resources or sense-annotated dataset. Here, word sense discrimination is performed by dividing the occurrences of words into classes to determine the words whether it belongs to the same sense or not. However, Evaluation of unsupervised approach is difficult to measure. Semi supervised approach may be called as minimally supervised approach where unlabelled data is used with the combination

of small quantity of labelled data thereby increasing the machine learning efficiency with much better performance [6]. Hybrid approach is a combination of different types of knowledge resources. In 1950, Kaplan [7] has determined that in a particular context two words on either side of an ambiguous word are equivalent to the whole sentence to the context. Kaplan's work is remarkable in the field of WSD. In 1957, Masterman [8] suggested his theory of finding the actual sense of a word using the headings of the categories present in the Roget's International Thesaurus. In 1964, Yehoshua [9] has pointed out that it is never possible to distinguish ambiguous meaning of a word without a Universal Encyclopaedia as he has used WSD as a part of his machine translation work. In 1980, Searle [10] devised the way in which computer system processed a language. He also highlighted the fact that linguistic symbols are meaningless unless and until it is not grounded or comprehend by someone. In 1990, Miller [11] has invented WordNet which is a revolution in the field of WSD as because there was no such hierarchical organized database of word senses called 'synsets' previously. Later, in 1991, Brown [12] has implemented corpus based WSD for the first time. Needless to say that most of the WSD works are performed on ambiguous noun using different approaches whereas there are very few works available based on ambiguous verb.

One of the significant works on removing ambiguity of verb is predication algorithm [13], which is discussed for homonym disambiguation and similarity judgment by the concept of latent semantic analysis with construction integration model. Another work [14] is highlighted on an interaction between the meaning of a context and vehicle terms of the metaphor where meaning is represented as vectors of a high dimensional semantic space. A new approach [15] has also been discussed on whether multise-mantic-role (MSR) based on selectional preferences could be used to improve the performance of supervised verb sense disambiguation method. Here performance is evaluated on two distinct datasets-lexical sample task of SENSEVAL-2 and the verbs from a movie script corpus. Another paper [16] presented one approach to improve the extraction of meaning from Diagnostic corpus by applying little bit modification in predication algorithm. Furthermore, a new concept is proposed [17] where different methods are used to extract the meaning of a polysemic word without using context by vector sum and existing predication algorithm. Some of the distributional approaches that are discussed [18] in literature for sense disambiguation application as well as reformulating the problem of measuring semantic similarity with respect to a particular context and outline a distributional method for identifying diverse documents that activate the sense of polysemous word. One of the reported works [19] has followed the predication algorithm where various semantic space models are compared and also generalized the predication algorithm for the problem of word-concept mapping model from the child learning which is verified by CHILD corpus. In another recent paper [20] on removing ambiguity of noun and verb together where it has discussed the existing methods and also created a new dataset of over 30,000 naturally-occurring non-trivial examples of noun-verb ambiguity for their experiment. This paper also has reported

errors that are very often using English part-of-speech tagger related to noun-verb ambiguity. In addition to these, new approach of visual word sense disambiguation for verb senses has been presented in a recent paper [21] by introducing the Multi-Sense dataset of 9,504 images annotated with English, German, and Spanish verbs. They have shown the benefits of cross-lingual verb sense disambiguation model over visual context by comparing uni-modal baselines. In order to find the correct sense of a verb is really a challenging task and resolving verb ambiguity using psycholinguistic model is very much limited. Customarily very few works have been reported on the propose topic but none of them is found to be impressive.

In the proposed work, connectionist network with activation function is used to remove the ambiguity of verb by finding out the surrounding terms to disambiguate the particular sense in which a verb is being used. Here the authors have considered the surrounding words as unordered in nature and found which words are commonly occurred around the target word. This technique is considered as supervised since it requires a training corpus where training must classify each word corresponding to a particular sense. In this work, Latent Semantic Analysis [22, 23] also has been used to find the real meaning of words used in a set of documents as because there is ambiguous term in the document. It maps both the words and the documents into a high dimensional semantic space and finds out the relationship between them. For example, when the word "bat" is used with words like ball, player, field then it may infer a cricket bat. Similarly, the word "bat" with words like trees, wings, fly specifies the sense of "animal bat".

Verb Sense Disambiguation method has not received enough attention in the literature survey of WSD since long time. Most of the WSD work has been performed in different languages using various techniques to remove the ambiguity of noun. There are many databases as well as thesaurus available for noun whereas no proper database is available for verb. Also most of the methods to disambiguate verbs are used in the same way as noun. Therefore, the performance of verb sense disambiguation method is not adequate in the state of art. In this paper, the authors attempt to find the sense of ambiguous verb in a context using vector space model with the notion of activation function to classify senses of verb with most probable surrounding terms despite the lack of conventional verb database.

This paper is organized as follows; Section 2 discusses the methodology of the proposed system. Section 3 discloses the experiment with discussion of various results of the proposed system. Section 4 reveals conclusion and future work.

II. PROPOSED METHODOLOGY

The basic approach of the proposed work is to gather distributional information of high-dimensional vectors and define semantic similarity in terms of vector similarity [24, 25]. Here, authors have used a document as a bag of words (BOW) which is commonly used for information retrieval. In BOW, authors count the number of times each word appears in a document which is the frequency of each word of the document and make a frequency histogram from it. The steps

that are followed by architecture of proposed work can be divided into training phase and a testing phase which is illustrated in Fig. 1.

The schematic diagram in Fig. 1 represents the architecture of work. The proposed methodology may be explained under four broad steps: i) Dataset creation ii) Data pre-processing iii) Training and iv) Testing.

1) *Dataset creation*: Here authors have perceived the word-sense disambiguation of verbs as a classification task. In any classification task, the machine learning algorithms are applied to a dataset of training samples and later on tested with testing samples. The accuracy of classification depends on the no. of unknown/testing samples that are correctly classified. Since the author's task is classification in nature, a standard dataset of ambiguous verbs only is in demand. However, due to the non-availability of standard dataset of such type; this major challenge is overcome with the creation of custom oriented dataset. Two versions of datasets are created viz. a training dataset containing sentences of ambiguous verbs which are extracted from WordNet and a test

dataset which is similar to the former one but the sentences are extracted from Babel Net. WordNet consists of 1, 17,000 synsets/classes organised in the form of a hierarchy. Each of the synset/class has its sense id, gloss and an equivalent example sentence. The information is not limited to the above-mentioned attributes, but carries other information too. But authors have extracted only the sense id, gloss and the example sentences for ten ambiguous verbs namely 'run', 'give', 'break', 'call', 'know', 'put', 'take', 'make', 'draw', 'get'. A total of 500 example sentences encompassing 10 ambiguous verbs are considered in the training dataset. Authors have taken 80% of total dataset as training and 20% as test dataset. Here, example sentences which is depicted as a subset of experimental training dataset for ambiguous verb 'run' are shown in the Table I.

Similar, to the above method of creating the training dataset, authors have created the test dataset from Babel net which contains only the example sentences of the ten ambiguous verbs.

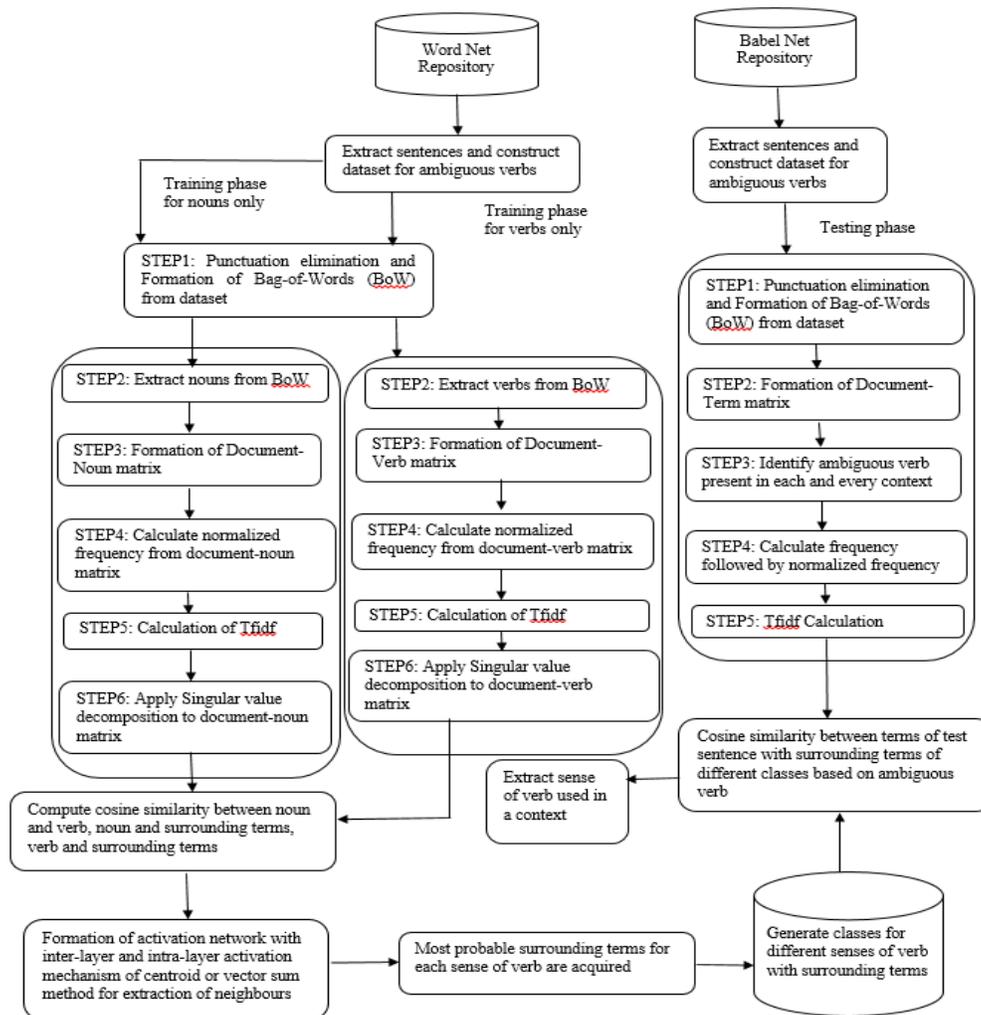


Fig. 1. Block Diagram of Proposed Model.

TABLE I. SUBSET OF TRAINING DATASET FOR AMBIGUOUS VERB 'RUN'

1.	The horse is running in the park.
2.	The horse is running in the race-course.
3.	The horse ran very fast in the race.
4.	The horse runs very fast.
5.	My horse runs last.
6.	The rabbit is running in the garden.
7.	The rabbit runs very fast.
8.	The rabbit is running around in the house.
9.	The kangaroo runs with a baby in its pouch.
10.	The kangaroo is running in the forest.
11.	The machine runs on electricity.
12.	The machine is running in the factory.
13.	The machine runs on crude-oil.
14.	The machine is running very smoothly.
15.	The machine ran properly for many hours.
16.	The colours run.
17.	These dyes and colours are guaranteed not to run.
18.	blood runs in the veins
19.	Blood runs from the heart to all parts of the body through artery.
20.	The bus runs between railway-station and airport.
21.	The bus runs between two states.
22.	The bus runs 100 miles daily.
23.	The bus runs daily throughout the year.
24.	The computer runs the instruction.
25.	The film runs for 3 hours.
26.	The ship runs before the wind.
27.	She runs 10 miles daily.
28.	He runs a new program on the laptop.
29.	They run the tapes over and over again.
30.	He never tires of running that video.
31.	The apple runs large this year.

2) *Data pre-processing*: As already described in step (i), the attributes contained in the training samples are sense id, gloss and example sentence. In the data pre-processing step, the punctuation markers are removed from the example sentences only. Each of the treated example sentences are then converted to bag-of-words (bow). In case of training phase for nouns, bow contains all unique word tokens including stop words, verbs, pronouns, prepositions, adjectives and numeric values whereas in case for verbs, bow contains all the word tokens. The bow for noun are converted to document-noun matrix and similar to the former approach the bow for verbs are converted to document-verb matrix. Each row of the matrices are considered as a vector having n-dimensions.

3) *Training phase*: The training phase is divided into two parts, one for training nouns only and another for verb where same steps are followed in both parts. There is a significance of n-dimensional vector representation in the document-verb and document-noun matrix where both the matrices represents this phase involves separate training of n-dimensional noun vector as well as n-dimensional verb vector. In this phase, the tf-idf values of the n-dimensional vectors in the matrices are calculated out. As the terms or words of the dataset are now become vectors, so authors need to compute the weight of all the vectors present in the dataset. As a result of that, term frequency followed by normalized term frequency is computed from which later tf-idf is calculated.

For instance, if the two sentences from training dataset are considered as follows:

S20: The bus runs between railway-station and airport.

S24: The computer runs the instruction.

The bag-of-words formed from the above sentences are:

[“the”, “bus”, “runs”, “between”, “railway-station”, “and”, “airport”, “computer”, “instruction”]

After finding out the term frequency of Bag-of-words, a count matrix is formed where all the sentences of training data set are considered as rows and words as columns shown in Table II.

These frequencies of words of training data set are normalized since each document of training set are of different size. Normalization of frequency is required as because the frequency of a particular word is much higher in a larger document than the smaller document as it contains few terms. Now, the matrices that are built using Latent Semantic Analysis are very large as well as very sparse because most of the cells are blanks due to small number of words in a document. The sparseness of the matrix is removed to get latent features of the terms of Bag-of-words. After that, here documents are converted to vectors of features and finding out the semantic similarity between two documents without considering word order by measuring the distance between these features by cosine similarity. The normalized frequency is obtained by dividing each term frequency with total number of terms present in a document which is shown in Table III.

In reality, certain words that occur too frequently such as article like a, an, the or some prepositions namely of, for, by etc. have little effect in determining the meaning of word. So by weighing down the effects of too frequently occurring words and vice versa for the less frequently occurring words, Inverse Document Frequency is calculated.

TABLE II. COUNT MATRIX FOR PREVIOUS EXAMPLE SENTENCES S20 AND S24

The bus runs between railway-station and airport. The computer runs the instruction.									
S20	1	1	1	1	1	1	1	0	0
S24	2	0	1	0	0	0	0	1	1

TABLE III. NORMALIZED FREQUENCY FOR THE PREVIOUS EXAMPLE SENTENCES S20 AND S24

The bus runs between railway-station and airport. The computer runs the instruction.									
S20	0.1428	0.1428	0.1428	0.1428	0.1428	0.1428	0.1428	0	0
S24	0.4	0	0.2	0	0	0	0	0.2	0.2

IDF= 1 + log_e (Total number of documents/ Number of documents in which the word is present).

$$TF*IDF=TF*(1+\log_e (N/df)) \quad (1)$$

After training phase of the training dataset taken from Word net ,the words that are very close to the particular sense of ambiguous verb are determined based on their cosine similarity value between nouns and verb. In this way, all the words which are most related to particular sense of a verb are acquired in a class. So, for example, if authors dataset containing ten different senses of an ambiguous verb ‘run’ then there are ten classes such as ‘moving’, ‘working’, ‘diffusion’, ‘flowing’ etc. consisting of surrounding words belonging to that class. Authors have chosen sentences for *run* for 10 different senses from Word Net as it is a database which resembles thesaurus. In the similar way, the classes are obtained for other ambiguous verbs present in training dataset. Now, most probable surrounding terms of any particular sense can be increased by adding more sentences in the training dataset. Now, Cosine similarity is calculated between two non-zero vectors after removing sparseness of matrix by singular value decomposition with this formula,

$$similarity = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i * B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} * \sqrt{\sum_{i=1}^n (B_i)^2}} \quad (2)$$

Henceforth, singular value decomposition method is applied because most of the cells are blanks due to small number of words in a document. Cosine similarity is used to find out the similarity between two terms. So, the similarity or distance between ambiguous verbs with surrounding terms, nouns and surrounding terms and ambiguous verb and noun are calculated by cosine similarity. Here, centroid or sum of vector method [15] is used to extract most probable surrounding terms or neighbours. The activation function is tailored by the work of Kintsch [15]. Now, an activation network is formed with three layers where first layer and third layer consist of one node and central layer consists of many nodes. Node in first layer represents ambiguous verb and node in third layer represents noun with which the sense of ambiguous verb will be changed. Nodes in the central layer or middle layer denotes surrounding terms of the contexts which are activated by two activation mechanism such as inter-layer and intra layer activation mechanism [17] .In the inter-layer activation mechanism, nodes in the central layer are activated

by both ambiguous verb and noun. In this case, some parameters are used in the formula. Ambiguous verb can be represents as V, noun as N and other surrounding terms in the central layer as O. Therefore, cosine similarity between verb and surrounding terms is Cos (V, O) and cosine similarity between noun and surrounding terms as Cos (N, O).So, activation function for inter-layer of the network would be Inter-layer Activation=Cos (V, O) +Cos (N, O).

In the similar way, Intra-layer activation is calculated where each node in the central layer is inhibited by every one of its neighbours. After computing inter-layer and intra-layer activation of each node in the central layer, an ordering has been done from highest to lowest and first n nodes are chosen accordingly as most probable surrounding terms for the ambiguous verb and particular noun with which verb is used in the context. The activation network is illustrated in Fig. 2.

4) *Testing phase:* The work has been performed on training data set where sentences consisting of ten different ambiguous verbs such as ‘run’, ‘give’, ‘break’, ‘call’, ‘know’, ‘put’, ‘take’, ‘make’, ‘draw’, ‘get’ which are taken from Word net. Now, bag-of-words [10] is formed which is a collection of all the words present in the documents. This training data set is prepared with taking care of punctuation and its multiplicity. This bag-of-words concept is used to find term frequency with which a word is appearing in a sentence is considered as a feature point of training. Therefore, a test data set is prepared using Babel net which is a multilingual encyclopaedic dictionary where documents containing ambiguous verbs those are present in training data set. Authors have used Babel Net for formation of testing dataset since it is linked to the computational lexicon of the English language, Word Net. The method which is used to extract feature vectors from training data set, same is used for extracting features from test data set. These features contains nouns and other surrounding terms of the ambiguous verb. To aim is to find the meaning of an ambiguous verb in context as well as the relevant words, semantic similarity of training and test sentences is measured. Whenever, a test sentence comes, the bag-of-words is calculated and ambiguous verb is identified. Therefore, rest terms of the bag-of-words are compared with the terms belong to senses available for the verb. If there is a matching between bag-of-words except verb of test sentence with at least few surrounding terms of same verb of training dataset then cosine similarity score is high.

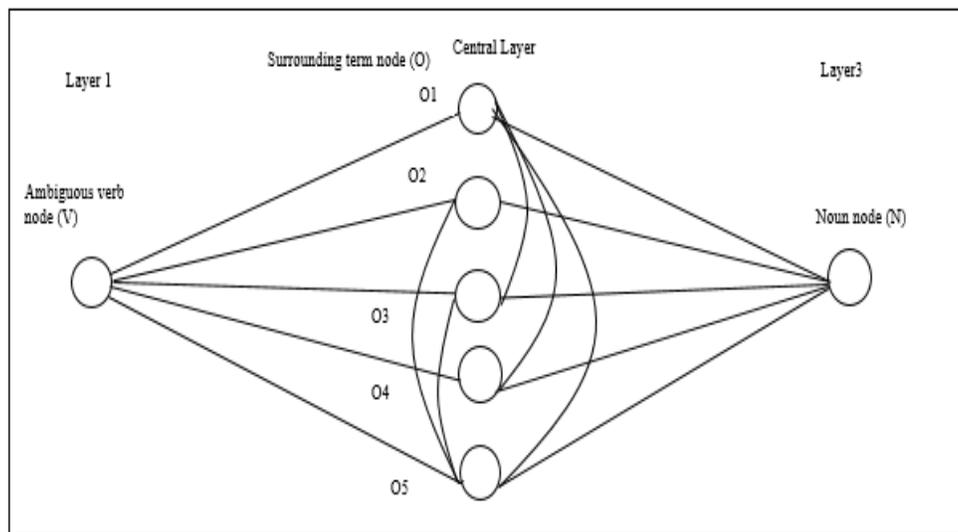


Fig. 2. Layers and Nodes in different Layers of Activation Network.

After that, the sense of the verb can be assumed based on the surrounding terms with which verb is used. Now for example, run is an ambiguous verb as it has multiple meaning depending on the context. It can be used as sense like moving, working, diffusion, flowing, executing, covering a certain distance etc. If in a test sentence containing verb runs with others terms such as machine, electricity, crude-oil which are also present in training document then the result of cosine similarity must be close to 1 indicating the sense for verb 'runs' is working in the test document with those surrounding terms. In the similar way, the terms like blood, artery, heart, veins, body of the test document with verb 'runs' has close to 1 cosine similarity value with the training document indicating that the meaning of runs is diffusion. Now using this method authors can remove the ambiguity of verb in a particular context whenever a new test sentence will come. Ambiguous verb along with the surrounding terms together can solve the problem where the notion of activation function is used in the way that the sense of verb can be achieved if verb is used with those specific surrounding terms of context. The meaning of verb will be completely different depending on the changes of those surrounding terms. In our work, multiple meaning or sense of ambiguous verb as well as training data set has been prepared from Word Net. In the same way, test data set for ambiguous verb is made ready from Babel Net. In our work, both the Training and Test data set are hand-crafted as because there is no such customized verb data set for work till date. So, dictionary cum thesaurus have been used for preparing training and testing dataset.

III. RESULTS AND DISCUSSION

The result of the experiment based on the example sentences for ten ambiguous verbs namely 'run', 'give', 'break', 'call', 'know', 'put', 'take', 'make', 'draw', 'get'. Now, for ambiguous verb run can have different meaning such as moving, working, flowing, diffusion, covering a certain distance etc. in different context. Here, authors have shown few example training sentences for run as 'moving' and 'working' sense and also corresponding test context. Now,

Table IV shows example sentences for run as moving sense with noun horse and Table V shows example testing sentence. Authors need to find most probable surrounding terms which can appear with run and horse in the same context.

In Table VI, It can be seen that race-course, race, park and last terms have obtained activation value above the range of threshold value 0.5 after applying activation function for the network which consists of verb run, noun horse and surrounding terms like race-course, race, park, fast, last etc. So, it can be concluded that these are the most probable surrounding terms in a context if ambiguous verb run is used with noun horse. These surrounding terms can be increased if more number of sentences are trained which is containing verb run with noun horse with it. At the same time, run and horse together can infer the sense of run is moving provided race or race-course or last appears together in the same context.

If any test context or query sentence which is containing verb run with horse, then similarity between surrounding terms of test context with any of the surrounding terms which have obtained in Table VI can help to find the sense of ambiguous verb run. So, different classes with surrounding terms can be formed after training the dataset containing verb run. Authors can also find cosine similarity between query document with any document of training dataset containing same verb and noun which is shown in the above Table VII. If cosine similarity value is close to 1 that indicates the surrounding terms of test context is similar with training document. Table VIII and Table IX display example training sentences and testing sentence for 'run' as moving sense, respectively.

TABLE IV. EXAMPLE TRAINING SENTENCES FOR RUN AS 'MOVING' SENSE

1. The horse is running in the park.
2. The horse is running in the race-course.
3. The horse ran very fast in the race.
4. The horse runs very fast.
5. My horse runs last.

TABLE V. EXAMPLE TEST SENTENCE FOR RUN AS ‘MOVING’ SENSE

1. <i>The horse runs faster than zebra in race. The horse runs the race for the cup.</i>
--

TABLE VI. ACTIVATION VALUE TO FIND SURROUNDING TERMS FOR MOVING SENSE OF VERB RUN

neighbours of run for ‘moving’ sense	Cosine neighbour of run	cosine neighbour of horse	activation value
race-course	0.350785733	0.358330308	0.709116041
race	0.658504608	0.31353902	0.972043628
park	0.24804297	0.253377791	0.501420761
fast	0.266184711	0.126740637	0.392925348
last	0.24804297	0.450906457	0.697721584
very	0.148172217	0.085139467	0.233311684

TABLE VII. COSINE SIMILARITY TO FIND SENSE OF AMBIGUOUS VERB IN TEST SENTENCE

Neighbourhood of run in training dataset of Table IV	Neighbourhood of run in test sentence of Table V	Semantic similarity between training sentences and test sentence 1
horse	horse	0.9800
race	race	
fast	faster	
last	zebra	
park	cup	
race-course		

TABLE VIII. EXAMPLE TRAINING SENTENCES FOR RUN AS ‘WORKING’ SENSE

11. <i>The machine runs on electricity.</i>
12. <i>The machine is running in the factory.</i>
13. <i>The machine runs on crude-oil.</i>
14. <i>The machine is running very smoothly.</i>
15. <i>The machine ran properly for many hours.</i>

TABLE IX. EXAMPLE TEST SENTENCE FOR RUN AS ‘WORKING’ SENSE

1. The machine runs on electricity in the factory for whole day.
--

In Table X, It can be seen that crude-oil, electricity, factory etc. terms have obtained activation value above the range of threshold value 0.5 after applying activation function for the network which consists of verb run, noun machine and surrounding terms like crude-oil, electricity ,hours, many, factory, properly etc. Now, run and machine together can infer the sense of run is working provided factory, crude-oil, electricity etc. appear together in the same context. Table XI shows the cosine similarity between training and testing sentence for verb ‘run’ as working sense.

Table XII, XIII and XIV show example training sentences, testing sentence and activation value to find surrounding terms

respectively for verb ‘run’ as covering a certain distance. If any test context or query sentence which is containing verb run with machine, then similarity between surrounding terms of test context with any of the surrounding terms which we have obtained in Table X can help to find the sense of ambiguous verb run. Confusion matrix is used to measure the performance of machine learning classification which is around 0.8235. It can be improved by increasing number of sentences in the dataset for ambiguous verb with all available sentences.

TABLE X. ACTIVATION VALUE TO FIND SURROUNDING TERMS FOR WORKING SENSE OF VERB RUN

neighbours of run for ‘working’ sense	Cosine neighbour of run	cosine neighbour of machine	activation value
crude-oil	0.274597701	0.51883493	0.793432631
electricity	0.274597701	0.51883493	0.793432631
smoothly	0.409250021	0.432362441	0.841612463
factory	0.350785733	0.370596378	0.721382111
properly	0.752576695	0.370596378	1.123173073
hours	0.145619993	0.202143479	0.347763472
many	0.398588494	0.1962796	0.594868094
very	0.148172217	0.156540252	0.304712469

TABLE XI. COSINE SIMILARITY TO FIND SENSE OF AMBIGUOUS VERB IN TEST SENTENCE

Neighbourhood of run in training dataset of Table VIII	Neighbourhood of run in test sentence of Table IX	Semantic similarity between training sentences of Table VIII and test sentence 2	Semantic Similarity between training sentences of Table IV and test sentence 2
machine	machine	1.0000	0
smoothly	electricity		
electricity	factory		
factory	whole		
properly	day		
hours			
crude-oil			

TABLE XII. EXAMPLE TRAINING SENTENCES FOR RUN AS ‘COVERING A CERTAIN DISTANCE’ SENSE

20. <i>The bus runs between railway-station and airport.</i>
21. <i>The bus runs between two states.</i>
22. <i>The bus runs 100 miles daily.</i>
23. <i>The bus runs daily throughout the year.</i>

TABLE XIII. EXAMPLE TEST SENTENCE FOR RUN AS ‘COVERING A CERTAIN DISTANCE’ SENSE:

3. The bus ran 10 miles that day.

TABLE XIV. ACTIVATION VALUE TO FIND SURROUNDING TERMS FOR COVERING A CERTAIN DISTANCE SENSE OF VERB RUN

neighbours of run for 'covering a certain distance' sense	cosine neighbour of run	Cosine neighbour of bus	activation value
miles	0.2745	0.4511	.7256
daily	0.1140	0.5941	.7081
states	0.2288	0.4511	.6799
railway-station	0.1716	0.3866	.5582
airport	0.1716	0.3866	.5582
year	0.1548	0.2245	.3793

IV. CONCLUSION AND FUTURE WORK

Removal of ambiguity of polysemous verb is very hard as it depends on the context. If the context of the same verb is altered then the meaning of the verb will be different. Since, ambiguity of verb needs to be removed in machine translation as inappropriate translation of source always leads to misprediction of information. In this work authors have used supervised machine learning approach by using centroid or vector sum method which helps in finding the meaning of ambiguous verb by classifying different senses of an ambiguous verb with most probable surrounding terms with it. So an ambiguous verb can be used with particular words for the specific sense of that verb and surrounding terms are changed if the sense of that verb in context is different. Therefore, sense of the verb can be predicted based on most probable surrounding terms only. Efficiency of the proposed method can be increased with larger set of data as more surrounding terms can be obtained which are feature points of the ambiguous verbs. The authors have used combination of dictionary and thesaurus for acquiring senses available for ambiguous verb as well as preparing hand crafted dataset for training and testing since there is no dataset available for verb. Future work may include increasing the accuracy of this method with the creation of database for ambiguous verb with all available senses.

ACKNOWLEDGMENTS

The authors are highly thankful to the Ministry of Electronics and Information Technology, Government of India initiated "Visvesvaraya PhD Scheme for Electronics and IT" for the support to carry out the research work.

REFERENCES

[1] Daniel Jurafsky, James H. Martin., "Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition", 2nd ed. Pearson Prentice Hall, Inc., Upper Saddle River, NJ, 2009.

[2] A.R.Pal, D.Saha, "Word sense disambiguation: A survey," International Journal of Control Theory and Computer Modeling, vol.5, no.3, July 2015.

[3] Word Sense Disambiguation; Algorithms and Applications, Edited by Eneko Agirre and Philip Edmonds, Springer, VOLUME 33.

[4] Zipf, George Kingsley. "Relative frequency and dynamic equilibrium in phonology and morphology", In Proceedings of the 6th international congress of linguists. Paris, 391-408, 1949.

[5] George A. Miller. Word net: A Lexical Database for English Communication, ACM, 38(11):39-41, 1995.

[6] Nandanwar, Lokesh, and Kalyani Mamulkar. "Supervised, semi-supervised and unsupervised WSD approaches: An overview." International Journal of Science and Research (IJSR) 4.2 (2015): 1684-1688.

[7] A. Kaplan. An experiment study of ambiguity and context. Mechanical Translation, 2:39-46, 1955.

[8] Masterman, Margaret. "The thesaurus in syntax and semantics." Mechanical Translation 4.1-2 (1957): 35-43.

[9] Y. Bar-Hillel. The present status of automatic translation of languages. In Advances in computers, volume 1, pages 91-163. Elsevier, 1960.

[10] J. R. Searle. Minds, brains, and programs. Behavioral and brain sciences, 3(3):417-424, 1980.

[11] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. J. Miller. Introduction to word net: An on-line lexical database. International journal of lexicography, 3(4):235-244, 1990.

[12] P. F. Brown, S. A. D. Pietra, V. J. D. Pietra, and R. L. Mercer. Word-sense disambiguation using statistical methods. In Proceedings of the 29th annual meeting on Association for Computational Linguistics, pages 264-270. Association for Computational Linguistics, 1991.

[13] Kintsch, W. (2001), "Predication", Cognitive Science, 25, 173-202.

[14] Kintsch, Walter. "Metaphor comprehension: A computational theory." Psychonomic bulletin & review 7.2 (2000): 257-266.

[15] Ye, Patrick, and Timothy Baldwin. "Verb sense disambiguation using selectional preferences extracted with a state-of-the-art semantic role labeler." Proceedings of the Australasian Language Technology Workshop 2006. 2006.

[16] Jorge-Botana, Guillermo, Ricardo Olmos, and José Antonio León. "Using latent semantic analysis and the predication algorithm to improve extraction of meanings from a diagnostic corpus." The Spanish journal of psychology 12.2 (2009): 424-440.

[17] Jorge-Botana, Guillermo, et al. "Visualizing polysemy using LSA and the predication algorithm." Journal of the American Society for Information Science and Technology 61.8 (2010): 1706-1724.

[18] Rumshisky, Anna. "Resolving polysemy in verbs: Contextualized distributional approach to argument semantics." Distributional Models of the Lexicon in Linguistics and Cognitive Science, special issue of Italian Journal of Linguistics/Rivista di Linguistica (2008).

[19] Kievit-Kylar, Brent, George Kachergis, and Michael Jones. "Naturalistic word-concept pair learning with semantic spaces." Proceedings of the Annual Meeting of the Cognitive Science Society. Vol. 35. No. 35. 2013.

[20] Elkahky, Ali, et al. "A Challenge Set and Methods for Noun-Verb Ambiguity." Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018.

[21] Gella, Spandana, Desmond Elliott, and Frank Keller. "Cross-lingual Visual Verb Sense Disambiguation." arXiv preprint arXiv: 1904.05092 (2019).

[22] Dumais, Susan T., et al. "Using latent semantic analysis to improve access to textual information." Proceedings of the SIGCHI conference on Human factors in computing systems. 1988.

[23] Landauer, Thomas K., Peter W. Foltz, and Darrell Laham. "An introduction to latent semantic analysis." Discourse processes 25.2-3 (1998): 259-284.

[24] Kintsch, W. (2007) Meaning in context, in T. K. Landauer, D. McNamara, S. Dennis and W. Kintsch (eds) Handbook of latent semantic analysis. Mahwah, NJ: Erlbaum.

[25] Soumya George K, Shibily Joseph, Text Classification by Augmenting Bag of Words (BOW) Representation with Co-occurrence Feature, IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p- ISSN: 2278-8727 Volume 16, Issue 1, Ver. V (Jan. 2014), PP 34-38.