Word Sense Disambiguation Using Sense Induction

Shipra Shaw

5/28/2019

Jadavpur University,Kolkata

JADAVPUR UNIVERSITY

MASTER DEGREE THESIS

"Word Sense Disambiguation using Sense Induction"

A thesis submitted in partial fulfillment of the requirements
for the degree of Master of Engineering in Computer Science and Engineering

in the

Department of Computer Science & Engineering

by
SHIPRA SHAW
University Roll Number: 001710502022
Examination Roll Number: M4CSE19026
Registration Number: 140761 of 2017-2018

Under the Supervision of
Dr. DIGANTA SAHA
Department of Computer Science and Engineering
Faculty of Engineering and Technology
Jadavpur University
Kolkata-700032

Declaration of Authorship

I, Shipra Shaw, declare that this thesis titled, “Word Sense Disambiguation using Sense Induction” and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely of my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed my-self.

Signed: _____

Date: _____

To whom it may concern

This is to certify that the work in this thesis entitled “Word Sense Disambiguation Using Sense Induction” has been satisfactorily completed by Shipra Shaw, University Roll Number: 001710502022, Examination Roll Number: M4CSE19026 , Registration Number: 140761 of 2017-2018. It is a bona-fide piece of work carried out under my supervision at Jadavpur University, Kolkata-700032, for partial fulfillment of the requirements for the degree of Master of Engineering in Computer Science and Engineering from the Department of Computer Science & Engineering, Jadavpur University for the academic session 2017-2019.

Dr. Diganta Saha
Professor
Department of Computer Science & Engineering,
Jadavpur University
Kolkata-700032.

Prof. Mahan Tapas Kundu

Head of the Department

Department of Computer Science & Engineering,

Jadavpur University

Kolkata-700032.

Prof. Chiranjib Bhattacharjee

Dean, Faculty of Engineering and Technology

Jadavpur University

Kolkata-700032.

Certificate of Approval

(Only in case the thesis is approved)

This is to certify that the thesis entitled “Word Sense Disambiguation Using Sense Induction ” is a bona-fide record of work carried out by Shipra Shaw, University Roll Number: 001710502022, Examination Roll Number: M4CSE19026 , Registration Number: 140761 of 2017-2018 , in partial fulfilment of the requirements for the award of the degree Master of Engineering in Computer Science and Engineering from the Department of Computer Science and Engineering, Jadavpur University for the academic session 2017-2019. It is understood that by this approval the undersigned do not necessarily endorse or approve any statement made, opinion expressed or conclusion drawn there in but approve the thesis only for the purpose for which it has been submitted.

_____

(Signature of the Examiner)
Date:

_____

(Signature of the Examiner)
Date:

Abstract

In this thesis an approach for Word Sense Disambiguation in Bengali language using sense induction technique is proposed .The proposed work is done in two phases , in the first phase it have Sense Induction module by generating unique sense clustered and in the second phase semantic similarity measure technique is used for the word sense disambiguation within the sentences. The data sets are prepared from the corpus developed under the TDIL (Technology Development for Indian Languages) project of the Government of India. After processing the two phases one after another, the meaning of multisensous word is found with the help of semantic similarity method.

Acknowledgements

On the submission of "Word Sense Disambiguation using Sense induction", I wish to express my gratitude to the Department of Computer
Science & Engineering for sanctioning a thesis work under Jadavpur University under which this work has been completed.

I would like to convey my sincere gratitude to **Dr. Diganta Saha**, Professor, Department of Computer Science & Engineering, Jadavpur University for his valuable suggestions throughout the project duration. I am really grateful to him for his constant support which helped me a lot to fully involve myself in this project and develop new approaches in the field of NLP and text analytics.

I would also wish to thank Prof. Mahan Tapas Kundu, Head of the Department of Computer Science and Engineering, Jadavpur University and Prof. Chiranjib Bhattacharjee, Dean, Faculty of Engineering and Technology, Jadavpur University for providing me all the facilities and for their support to the activities of this research.

Lastly I would like to thank all my teachers, classmates, guardians and well wishers for encouraging and co-operating me throughout the development of this project. I would like to especially thank my parents whose blessings helped me to carry out my project in a dedicated way.

Regards,
Shipra Shaw
University Roll Number: 001710502022
Examination Roll Number: M4CSE19026
Registration Number: 140761 of 2017-2018
Department of Computer Science & Engineering
Jadavpur University

Signed:

Date:

Contents

Chapter 1

1.1 Introduction …….3
1.2 The state of art……..4
1.3 A brief Survey……..6
1.4 Conventional approaches to WSD……...7

Chapter 2

2.1 Pre-processing……..9
2.1.1 Introduction……9
2.1.2 Normalization……10
2.1.3 Lemmatization……...11
2.1.4 Conclusion……12

Chapter 3

3.1 Proposed Work……...14
3.1.1 Introduction……...14
3.1.2 Algorithms used……15
3.1.2.1 Algorithm1: Sense Finding……15
3.1.2.2Algorithm2:Word Sense Disambiguation Classifiers17
3.1.2 Conclusion……19

Chapter 4

4.1 Output and Discussions……21
4.2 Challenges……...23

Chapter 5

5.1 Conclusion……25

5.2 References……26

List of Abbreviations

Word Sense Disambiguation, Sense Induction , WordNet, Semantic Similarity Measure, Context Expansion.

Chapter 1

1.1 Introduction

Word Sense disambiguation (**WSD**) is the task of identifying the most appropriate sense of a multisensous word in a particular syntactic context. A normal human being, due to his/her innate linguistic competence, is able to capture the actual contextual sense of an multisensous word within a specific syntactic frame with the knowledgebase triggered from various intra- and extra-linguistic environments. Since a machine does not possess such capacities and competence, it requires some predefined rules or statistical methods to do this job successfully. Normally, two major learning procedures are used in WSD task. First one is Supervised Learning where a learning set is considered for the system to predict the actual meaning of a multisensous word within a syntactic frame. Other one is Unsupervised Learning where dictionary information (i.e., glosses) of the multisensous word is used to do the same task. In most of the cases, WordNet is used as the online semantic dictionary.

The proposed work is developed using a semi-supervised approach. The overall work is carried out in two phases, in the first phase different sense clusters are created using Sense Induction technique and in the second phase, Word Sense Disambiguation is developed using Semantic Similarity Measure. In Sense Induction [1-6] phase, a context window is taken containing the multisensous word and the words preceding and succeeding it to classify multiple senses from the sentence collection. In Word Sense Disambiguation stage, the similarity of the new test data is evaluated with the previously evaluated sentence classes using semantic similarity measure. This semantic similarity is measured using the Bengali WordNet. The test data is assigned with that sense where the maximum similarities found.

The data sets are prepared from the corpus developed under the TDIL (Technology Development for Indian Languages) project of the Government of India. The developed model is tested on multisensous words commonly used in Bengali language. The challenges and the pitfalls of this work are explained in detail at the end of this thesis paper.

1.2 The state of art

Since the earliest days of computer language treatment in the 1950s, the automatic disambiguation of word senses has been an interest and concern. Wilks and Stevenson(1996) called WSD as an ”intermediate task” which is used at each level in natural language processing .It is clearly necessary for language based applications such as message understanding, man to machine communication, etc. This is the least helpful for those applications, and there are some examples that are not intended for Understanding of language:

- **Machine Translation** : For understanding the proper translation of words like in case of French grille WSD is very useful. It can be translated on the basis of context, like railing, gate, bar, grid, scale, schedule, etc. (weaver [1949]; Yangway [1955], etc.)

- **Information retrieval and hypertext navigation** : Words which are used in improper sense are eliminated when a specific keywords is searched in a document(Salton[1968]; Salton and McGill [1983]; Krovetz and Croft [1992]; Voorhees [1993]; Schütze and Pedersen [1995])

- **Content and thematic analysis**: For analysing the distribution of predefined words content thematic approach is used. In this kind of analysis the need WSD has long been recognized. (Stone, et al. [1966]; Stone[1969]; Kelly and Stone [1975]; Litowski [1997])

- **Grammatical analysis:**
  For part of speech tagging, sense disambiguation is useful. Some syntactic analyzes, such as prepositional phrase attachment, also requires word sense disambiguation (Jensen and Binot [1987]; Whittemore *et al.*[1990]; Hindle and Rooth [1993])

- **Speech Processing:** The correct phonetization of words in speech synthesis requires disambiguation of senses. In speech recognition for homophone discrimination and word segmentation also WSD is used. (Connine[1990]; Seneff [1992])

- **Text Processing:** In spelling correction also sense disambiguation is necessary to determine when diacritics should be inserted, for case changes, for the literal use of Semitic languages(when vowels are not written).

1.3 A brief Survey

Ramanathan and Rao [1] designed a lightweight stemmer for Hindi language in 2003. The approach was tested on 35977 unique words taken from different sources like politics, sports, business, health, etc. The error in this method was 4.68% for under-stemming and 13.84% for over-stemming task.

Satanjeev Banerjee and Ted Pedersen in 2002[2], evaluated the Lesk algorithm on English lexical sample data from the SENSEVAL-2 Word Sense Disambiguation exercise and achieved overall accuracy of 32%.

Satanjeev Banerjee in 2002[3], adapted a modification on this algorithm by considering a very short window of only three words, means the target word and the first meaningful word before and after that target word. This idea believes that people generally use the immediate words to represent the sense of the multisensous word.

In Hindi language, Prity Bala proposed a knowledge based approach [4] for Word Sense Disambiguation using Hindi WordNet. She considered a test set of 100 words and among them 85 words were responded by the system, out of which 50 words were correctly disambiguated.

1.4 Conventional approaches to WSD

- Dictionary- and knowledge-based methods: These mainly depend on the basis of dictionaries, thesauri and lexical knowledge, without any corroborative evidence.

- Semi-supervised method: It uses secondary knowledge source or corpus of seeded data in a word aligned bilingual corpus.

- Supervised method: It uses meaningful corpus from which to train.

- Unsupervised methods: This relies completely on external information and process on unknown corpus .This is also known as word sense discrimination.

Chapter 2

2.1 Pre-processing

2.1.1 Introduction:

The corpus is first pre-processed to create the knowledge base from where the answer is to be found. Text pre-processing is an essential part of any NLP system, since the characters, words, and sentences identified at this stage are the fundamental units passed to all further processing stages, from analysis and tagging components, such as semantic analysers and part-of-speech taggers, through applications, such as information retrieval and machine translation systems. It is a collection of activities in which Text Documents are pre-processed. Because the text data often contains some special formats like number formats, date formats and the most common words that unlikely to help text mining such as prepositions, articles, and pro-nouns can be eliminated.

2.1.2 Normalization

The texts collected from the Bengali corpus are not adequately normalized. So, a text normalization procedure has been used before the work, as –

Detachment of punctuation marks like single quote, tilde, double quote, parenthesis, comma, etc. that are attached to the words.

Removal of angular brackets, uneven spaces, broken lines, slashes, etc.from sentences.

Identification of sentence terminal markers (i.e., dāri, note of exclamation, and note of interrogation), etc.

A screenshot of a non-normalized text is given below.

<BAC০১><Commerce..><Accountanc><১৯৮৩><Book.><হিসাবশাস><ভট্টাচার><০৬
০১>

আধুনিক হিসাবশাস্ত্র ক্রমশঃই জটিল হইতে জটিলতর আকার ধারণ করিতেছে । শতাব্দীর পর শতাব্দী ধরিয়া ব্যবসায়িগণ ও রাষ্ট্রের রাজস্ববিভাগ কর্তৃক ব্যবহারিক ক্ষেত্রে হিসাবের সর্বাধিক প্রয়োগের ফলে বর্তমানে হিসাবশাস্ত্র ব্যবহারিক বিজ্ঞানে (Applied Science) মর্যাদা লাভ করিয়াছে । হিসাবশাস্ত্রের ক্রমবিকাশের পিছনে রাষ্ট্র ও ব্যবসায়িগণের ভূমিকা অনস্বীকার্য । পরিবর্তিত পরিস্থিতিতে অর্থাত্ সময়ের পরিবর্তনের সঙ্গে সঙ্গে মূলতঃ ইহাদের ব্যবহারিক সুবিধা এবং উপযোগের (Practical convenience and utility) দিকে লক্ষ্য রাখিয়াই নূতন নূতন নীতি ও সূত্র হিসাবশাস্ত্রে সংযোজিত হইতেছে । এই সকল নীতি ও সূত্রগুলি শিক্ষকগণ এবং শিক্ষাপ্রাপ্ত ব্যক্তিগণের অভিজ্ঞতালব্ধ মূল্যবান ফসল । হিসাবশাস্ত্র সম্পর্কে পূর্ণাঙ্গ জ্ঞান লাভ করিতে হইলে ইহার ক্রমবিকাশের ইতিহাস , ইহার প্রকৃতি , কার্যাবলী ইত্যাদি সম্পর্কে পরিষ্কার ধারণা থাকা প্রয়োজন । সেইজন্য এই অধ্যায়ে উক্ত বিষয়াদি সম্পর্কে বিস্তারিত আলোচনা করা হইল । হিসাবশাস্ত্রের ভিত্তি (Foundation of Accountancy) : মানবসভ্যতার ক্রমবিবর্তনের ইতিহাসের মতই হিসাবশাস্ত্রের ক্রমবিবর্তনের ইতিহাস অতি প্রাচীন ও

Fig1: Partial view of non-normalized text

After normalization process the following text is found as output.

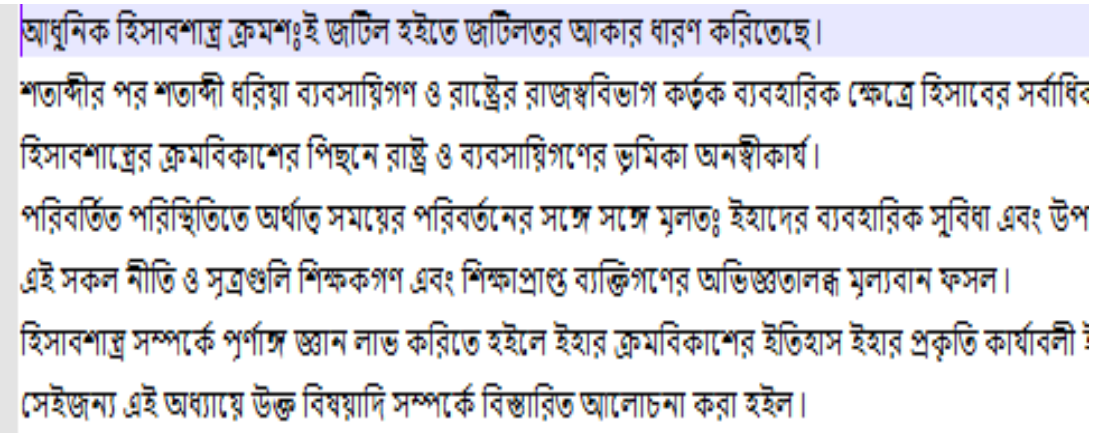

আধুনিক হিসাবশাস্ত্র ক্রমশঃই জটিল হইতে জটিলতর আকার ধারণ করিতেছে।
শতাব্দীর পর শতাব্দী ধরিয়া ব্যবসায়িগণ ও রাষ্ট্রের রাজস্ববিভাগ কর্তৃক ব্যবহারিক ক্ষেত্রে হিসাবের সর্বাধি
হিসাবশাস্ত্রের ক্রমবিকাশের পিছনে রাষ্ট্র ও ব্যবসায়িগণের ভূমিকা অনস্বীকার্য।
পরিবর্তিত পরিস্থিতিতে অর্থাত্ সময়ের পরিবর্তনের সঙ্গে সঙ্গে মূলতঃ ইহাদের ব্যবহারিক সুবিধা এবং উপ
এই সকল নীতি ও সূত্রগুলি শিক্ষকগণ এবং শিক্ষাপ্রাপ্ত ব্যক্তিগণের অভিজ্ঞতালব্ধ মূল্যবান ফসল।
হিসাবশাস্ত্র সম্পর্কে পূর্ণাঙ্গ জ্ঞান লাভ করিতে হইলে ইহার ক্রমবিকাশের ইতিহাস ইহার প্রকৃতি কার্যাবলী
সেইজন্য এই অধ্যায়ে উক্ত বিষয়াদি সম্পর্কে বিস্তারিত আলোচনা করা হইল।

Fig2: Partial view of normalized text.

2.1.3.Lemmatization

Lemmatization is the grouping together of different forms of the same word. In search queries, lemmatization allows end users to query any version of a base word and get relevant results. Stemming is a part of the lemmatization process where the base word of the particular word is found. The base word broadens the sense (search domain).

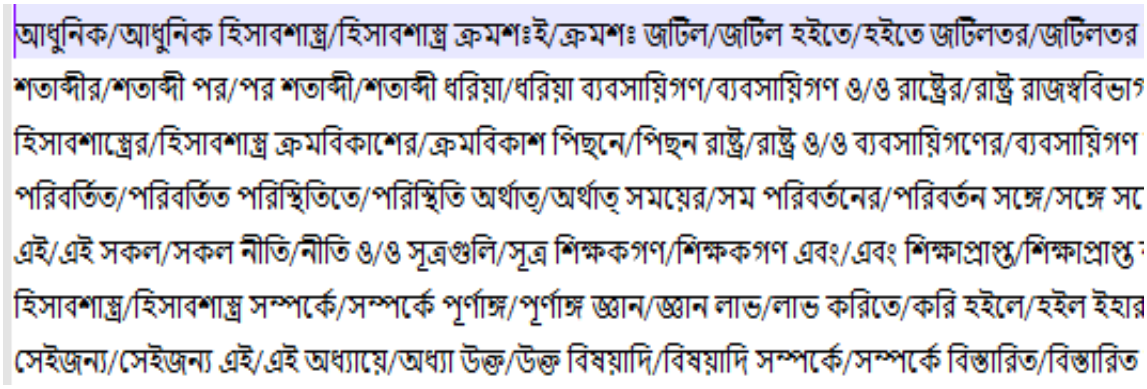

আধুনিক/আধুনিক হিসাবশাস্ত্র/হিসাবশাস্ত্র ক্রমশঃই/ক্রমশঃ জটিল/জটিল হইতে/হইতে জটিলতর/জটিলতর
শতাব্দীর/শতাব্দী পর/পর শতাব্দী/শতাব্দী ধরিয়া/ধরিয়া ব্যবসায়িগণ/ব্যবসায়িগণ ও/ও রাষ্ট্রের/রাষ্ট্র রাজস্ববিভাগ
হিসাবশাস্ত্রের/হিসাবশাস্ত্র ক্রমবিকাশের/ক্রমবিকাশ পিছনে/পিছন রাষ্ট্র/রাষ্ট্র ও/ও ব্যবসায়িগণের/ব্যবসায়িগণ
পরিবর্তিত/পরিবর্তিত পরিস্থিতিতে/পরিস্থিতি অর্থাত্/অর্থাত্ সময়ের/সম পরিবর্তনের/পরিবর্তন সঙ্গে/সঙ্গে সঙ
এই/এই সকল/সকল নীতি/নীতি ও/ও সূত্রগুলি/সূত্র শিক্ষকগণ/শিক্ষকগণ এবং/এবং শিক্ষাপ্রাপ্ত/শিক্ষাপ্রাপ্ত
হিসাবশাস্ত্র/হিসাবশাস্ত্র সম্পর্কে/সম্পর্কে পূর্ণাঙ্গ/পূর্ণাঙ্গ জ্ঞান/জ্ঞান লাভ/লাভ করিতে/করি হইলে/হইল ইহার
সেইজন্য/সেইজন্য এই/এই অধ্যায়ে/অধ্যা উক্ত/উক্ত বিষয়াদি/বিষয়াদি সম্পর্কে/সম্পর্কে বিস্তারিত/বিস্তারিত

Fig3: Partial view of the lemmatized text.

2.1.4 Conclusion:

In this phase, first the overall data sets are passed through a series of pre-processing steps, as-Normalization and Lemmatization. At the end of this phase each and every sentences are normalized and lemmatized.

Chapter 3

3.1 Proposed Work

3.1.1 Introduction:

In this work, first the sentences are retrieved from the corpus. As the sentences are real life data, those are passed through normalization and lemmatization process.

Next, using a separate program all the sentences containing a particular multisensous word are retrieved from the corpus.

After that, sentences are passed through Sense Induction phase. In this phase, different word windows are considered containing the multisensous word and its surrounding word/s. In this experiment, following word windows are considered:

$(w_k w_{k+1})$, $(w_k w_{k+1} w_{k+2})$, $(w_{k-1} w_k)$, $(w_{k-2} w_{k-1} w_k)$, $(w_{k-2} w_{k-1} w_k w_{k+1})$, $(w_{k-2} w_{k-1} w_k w_{k+1} w_{k+2})$ and $(w_{k-1} w_k w_{k+1})$, where w_k is the targeted multisensous word.

Using all of these word windows, the sentences in a data set are classified. After generation of all possible sense classes, the redundant classes are discarded. In this process, different sense carrying subclasses of a particular multisensous word are generated. As every subclass carries a distinct sense, these are used as the sense inventory for WSD task.

In the second phase, semantic similarity is measured between a test sentence and the sense tagged classes. A test sentence is tagged with a particular sense based on the closeness of that sentence with the sense tagged classes. This semantic similarity is measured with the help of Bengali WordNet.

In real life scenario, there is a very least chance of lexical similarity between two sentences. So, in this experiment contexts of the test sentence and the sentences of every class are expanded considering synonyms and gloss of every word from the WordNet. As the contexts of the sentences are expanded the chance of finding semantic similarity is increased.

3.1.2 Algorithms Used

3.1.2.1 ALGORITHM 1: SENSE FINDING

Input: Sentences containing multisensous words.

Output: Sense classes.

This module involves the following steps:

Step 1: Sentences are collected from the corpus

Step 2: Data sets are normalized.

Step 3: Words in every sentence are lemmatized.

Step 4: Different sense classes are generated using all possible word windows.

Step 5: Redundant sense classes are discarded to generate unique sense classes.

Step 6: Stop.

In the following figure first sentences are collected from corpus then all the data sets are normalized means removal of angular brackets, uneven spaces, broken lines, slashes, etc. from sentences. After that all the words are lemmatized and create all the possible word windows. Then redundant sense classes are discarded to generate unique sense classes.

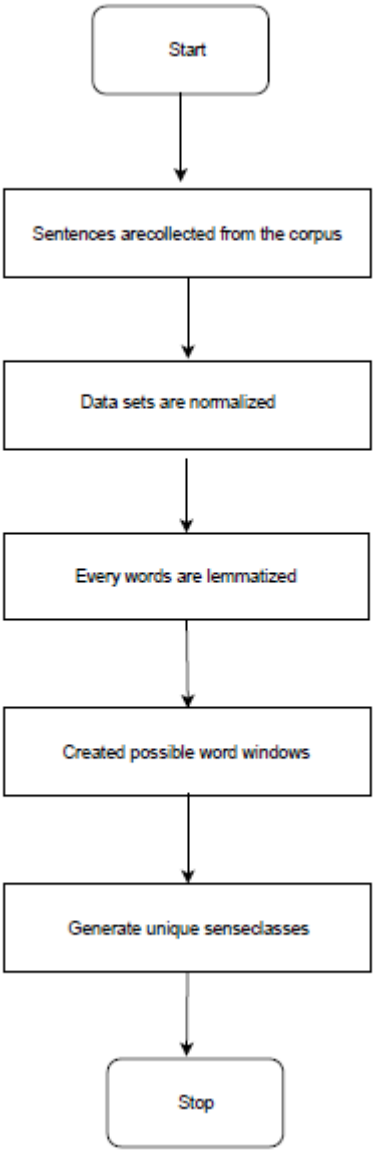


Fig 1. Sense Finding

3.1.2.2 ALGORITHM 2: WORD SENSE DISAMBIGUATION USING CLASSIFIERS

Input: Sentence whose sense have to find i.e. Test set.

Output: Derived sense of the multisensous word in that sentence.

Step 1: Start

Step 2: The test sentence is normalized and lemmatized.

Step 3: Contexts of the test sentence and the sentences of every class are expanded considering synonyms and gloss of every word from the WordNet.

Step 4: Semantic similarity between sentences is measured on the expanded contexts.

Step 5: Maximum similarity with a sense class represents respective sense of the multisensous word.

Step 6: Stop.

In the following figure user input the sentence, every sentence is normalized and lemmatized first then each and every word of the sentence is split and find the meaning of each and every word from the word-net. After that find semantic similarity between the sentences and find the most similar word.

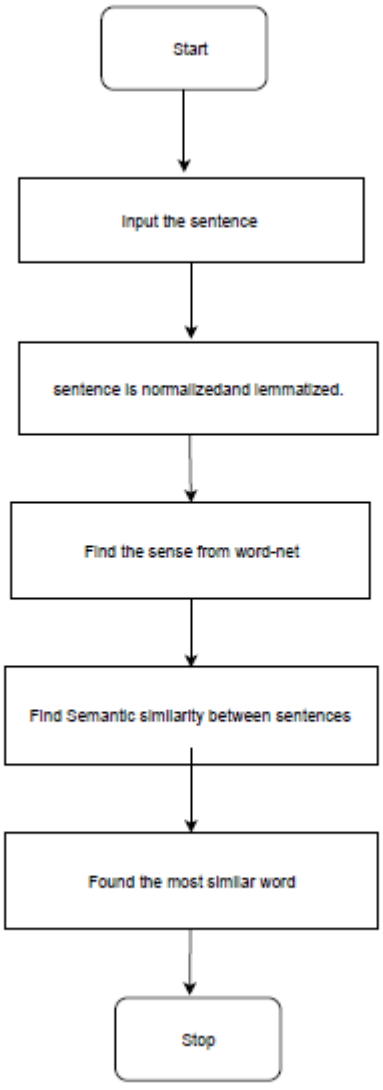


Fig.2 Finding most similar word

3.1.5 Conclusion

Here the similarity of the new sentences containing the multisensous word is evaluated with respect to the evaluated classifiers using Semantic Similarity Measure technique. The test data will be assigned to the sense where it found maximum similarities. Finally we get the sense resolved sentence as the output.

Chapter 4

4.1 Output and Discussions

The model is tested on multisensous words commonly used in Bengali language. Here the different senses of the multisensous word containing word হাত (hāt) are evaluated.

Table 1: Number of sentences retrieved from the corpus carrying multisensous words.

Sl. No.	Multisensous words	Number of sentences retrieved from the corpus
1	হাত (hāt)	2383

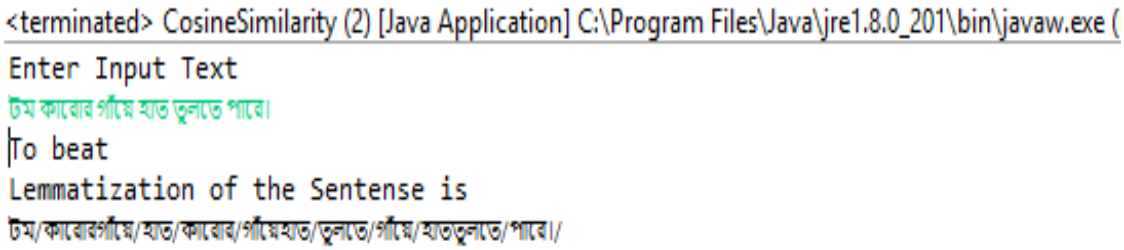
Although, during the experiment all the possible words windows are considered for generating the sense classes, during evaluation of the results it is observed that the ($w_k w_{k+1}$) word sequence produces the most acceptable results. One sample output for the multisensous word **'হাত'** (hāt) is given in Table 2.

From that corresponding dataset 9 different sense classes are generated.

Table 2: Different sense classes for the multisensous word 'হাত' (hāt).

Sl. No.	Senses considering ($w_k w_{k+1}$) word window	**Meaning**
1	হাত করা (hāt karā)	To control
2	হাতের কাজ (hāter kāj)	Handicraft
3	হাত জোড় (hāt jor)	Pray
4	হাত তোলা(hāt tolā)	To beat
5	হাত পাতা(hāt pātā)	Ask for help
6	হাত বাড়ানো(hāt bārāno)	Help someone
7	হাত মেলানো (hāt melāno)	Co-operate
8	হাত লম্বা (hāt lambā)	Type of measurement
9	হাতের লেখা(hāter lekhā)	Handwriting

Few sentences created individual class (consisting of one sentence) because of no lexical match. These classes were discarded. The model is tested using 4-cross validation.



In the above output screenshot user first input the sentence then the sentence is lemmatized and normalized then find the meaning of each and every word from the word-net. After that find semantic similarity between the sentences and find the most similar word, and show as an output.

4.2 Challenges

A lot of challenges appeared in every phase of the experiment, like –

The wide range of morphological inflections of the Bengali words which created a major problem lemmatization task.

Absence of an accurate all word lemmatizer tool in Bengali language affected the performance of the system.

The vast semantic varieties of Bengali sentences are sometimes impossible to track programmatically.

Few sentences have been encountered which carry similar sense but there are no similarities in contextual words.

Spelling errors in words are also the obstacles in the sense retrieval process.

As the Bengali WordNet is under developing stage, scarcity of information in the WordNet is a big issue in sense retrieval task.

Chapter 5

9.Conclusion

In this work, an attempt is made to design an effective algorithm for word sense disambiguation using sense induction in Bengali language. Different senses of a particular multisensous word is derived using sense induction process and WSD is implemented through semantic similarity measure using Bengali WordNet. The experiment is carried out on multisensous words commonly used IN Bengali.

10.Reference:

[1]. Ramanathan and D.D. Rao. A Lightweight Stemmer for Hindi. Workshop on Computational Linguistics for South-Asian Languages, EACL, 2003.

[2]http://www.d.umn.edu/~tpederse/Pubs/cicling2002-b.pdfvisited on: 15/11/2018

[3]. http://www.duluth.umn.edu/cs/thesis/satanjeev_banerjee_ms.pdfvisited on: 15/11/2018

[4]. Prity Bala, "Knowledge Based Approach for Word Sense Disambiguation using Hindi WordNet," The International Journal of Engineering and Science (IJES), Volume 2 Issue 4, pp. 36-41, 2013.

[5] Banerjee, S. & Pedersen, T. 2002. An adapted Lesk algorithm for word sense disambiguation using WordNet. In Proceedings of the 3rd International Conference on Intelligent Text Processing and Computational Linguistics, Mexico City, February.

[6] Lesk, M. 1986. Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice cream cone. Proceedings of SIGDOC.

[7] http://www.dlsi.ua.es/projectes/srim/publicaciones/CICling-2002. Accessed 14 May 2019.

[8] Mittal, K. & Jain, A. 2015. Word sense disambiguation method using semantic similarity measures and Owa Operator. ICTACT Journal on Soft Computing: Special Issue on Soft-Computing Theory, Application and Implications in Engineering and Technology, 5/2.

[9] http://www.d.umn.edu/~tpederse/Pubs/cicling2003-3.pdf. Accessed 14 May 2019.

[10] http://www.aclweb.org/anthology/U04-1021. Accessed 14 May 2019.

[11] Jurafsky, D. & Martin, J. H. 2000. Speech and language processing. Pearson Education Asia, pp. 604.21

[12] Ide, N. &Véronis, J. 1998. Word sense disambiguation: The state of the art. Computational Linguistics, 24/1, 1-40.

[13] Navigli, R. 2009. Word sense disambiguation: A survey. ACM Computing Surveys, 41/2, 1-69.