

Gujarati Language Model: Word Sense Disambiguation using Supervised Technique

Tarjni Vyas, Amit Ganatra

Abstract: Word Sense Disambiguation (WSD) is a complex problem as it entirely depends on the language convolutions. Gujarati language is a multifaceted language which has so many variations. In this paper, the debate has advanced two methodologies for WSD: knowledge-based and deep learning approach. Accordingly, the Deep learning approach is found to perform even better one of its shortcoming is the essential of colossal data sources without which getting ready is near incomprehensible. On the other hand, uses data sources to pick the implications of words in a particular setting. Provided with that, deep learning approaches appear to be more suitable to manage word sense disambiguation; however, the process will always be challenging given the ambiguity of natural languages.

Keywords: Word Sense Disambiguation, Gujarati Language, Deep learning, Natural language processing, Lesk Algorithm, Wordnet.

I. INTRODUCTION

Natural language, such as English are indistinguishable when contrasted with programming languages. Human language is considered to be ambiguous, with the goal that numerous words can be deciphered in different ways relying upon the setting wherein they happen. Take, for instance, the following two sentences: (a) He paid a ten per cent interest on the loan he received. (b) The interest he paid her led many to assume much. The occurrences of the word interest in the two sentences unmistakably indicate various implications: financial value and a feeling of admiration, respectively. Unluckily, the recognizable proof of the particular implying that a word assumes in a setting is just clearly straightforward. This signifies that the importance of the word is chosen by its encompassing setting in a sentence. While more often than not people don't consider the ambiguities of language, machines need to process unstructured printed information and change them into information structures that must be dissected to identify the hidden meanings.

Revised Manuscript Received on July 22, 2019.

* Correspondence Author

Tarjni Vyas*, Department of Computer Science and Engineering (CSE), Institute of Technology, Nirma University, Ahmedabad, India. Email: tarjni.vyas@nirmauni.ac.in

Amit Ganatra, Dean-Faculty of Technology and Engineering, Devang Patel Institute of Advance Technology & Research (DEPSTAR), Charotar University of Science and Technology. Email: amitganatra.ce@charusat.ac.in

The computational distinguishing proof of significance for words in a setting is what is referred to as Word Sense Disambiguation (WSD).

In line with Chen, Liu, and Sun^[1], words do not have well-characterized limits between their meanings of senses, which brings about the issue of figuring out which sense of the word is intended in a given setting. Accordingly, this is the absolute first problem that is experienced by Natural Language Processing (NLP) systems, which is alluded to as lexical-semantic equivocality. Navigli^[2] provides that WSD is an exploration field in NLP, which is helpful these days. For instance, it is the method of NLP. Also, it can be represented by assignment, execution, learning source; computational multifaceted nature, conventions and application for WSD calculations.

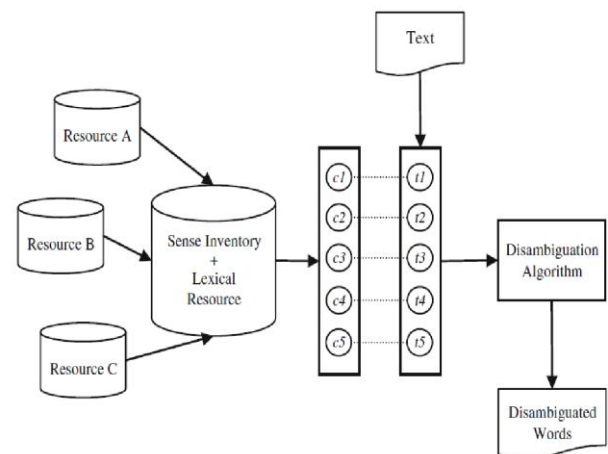


Figure 1 Knowledge based word sense disambiguation techniques

Zhong and Ng^[3] claim that WSD includes more words and word learning or presence of mind, which recognizes Dictionary or Thesauri. In this manner, it is additionally useful in numerous application, for example, information extraction (IE), information recovery (IR) together with speech recognition (SR)^[3]. As such, word sense disambiguation is significant for Lexical learning and word information. Tentatively, there are different approaches when it comes to dealing with WSD. In any case, this paper will focus on a comparative study between the knowledge-based approach and deep learning or machine learning approach.

II. LITERATURE SURVEY

A. KNOWLEDGE-BASED APPROACH

This kind of approach is based on learning assets of machine-readable lexicons in a type of WorldNet, corpus and so forth. In line with Miller et al.^[4], they may utilize sentence structure rules when it comes to disambiguation. Provided with that, the point of knowledge-based approach – also termed as Dictionary-based methodology – in WSD is to make use of learning assets to induce the meanings of words in a given setting. These learning assets include word references, thesauri, ontology's, collocations and so forth^[4]. Gomaa and Fahmy^[5] are in agreement and state that there are three knowledge-based methodologies which are as follows:

i) Lesk Algorithm

In line with Gomaa and Fahmy^[5], M. Lesk suggested a way to establish the overlap occurring in words in the sense meanings of equivocal words and the meanings of context words encompassing the equivocal words in a given setting. Basile et al.^[6] mention that the greatest disadvantage of this type of knowledge-based algorithm is that lexicon definitions are regularly short and need more words for the algorithm to function admirably.

Algorithm	Accuracy
WSD using Selectional Restrictions	44% on Brown Corpus
Lesk's algorithm	50-60% on short samples of "Pride and Prejudice" and some "news stories"
WSD using conceptual density	54% on Brown corpus
Walker's algorithm	50% when tested on 10 highly Polysemous English words

Table 1 Knowledge based approaches

Nonetheless, Wei et al.^[7] point out that modifications have been proposed that aim to manage the mentioned issue by adjusting the algorithm to a semantically sorted out lexical database that is referred to as WordNet. Other than storing words and their sense like a typical lexicon, WordNet likewise interfaces related words together. In doing so, this overcomes the issue of short definitions by searching for regular words not just between the meanings of the words being disambiguated, yet additionally between the meanings of words that are firmly identified with them in WordNet^[7].

ii) WSD using conceptual density

Golkar et al.^[8] position that the notion of conceptual density relates to the proportion of how the idea that the word speaks to is identified with the idea of the words in its unique situation. This means that the meaning of a word is identified within the context it is placed in. In the example of the word interest, the two meanings are differentiated given the unique setting of the sentence.

iii) Walker's Algorithm

Apart from the two mentioned, Walker's algorithm also forms part of the knowledge-based approach. Kalita and

Barman^[9] articulate that Walker proposed a basic algorithm by fusing subject codes. The algorithm, in turn, depends on the presumption that the subject codes appointed to a word mirror the meaning of the word. On the off chance that a word has more than one subject code, at that point it will have more than one meaning. For instance: "Longman's Dictionary of Contemporary English" incorporates the subject code EC to signify the budgetary meaning associated with a bank. This subject code is useful in realizing that another term, such as "depositing" is identified with the budgetary meaning of a bank.

B. DEEP LEARNING APPROACH

Deep learning or machine learning approaches make use of frameworks that are prepared and skilled to deal with word sense disambiguation. Borah et al.^[10] direct that in this technique, a classifier is designed and prepared, which is then used to assign meanings to concealed examples. In this methodology, the underlying input comprises of the words to be disambiguated, alongside content in which it is installed - which is referred to as its setting. Accordingly, this underlying input is prepared to utilize grammatical feature labelling or any morphological handling. After this initial preparing, Nameh^[11] states that a fixed arrangement of linguistic highlights is extricated applicable to the learning task. These highlights can be either of two classes: co-occurrence or collocation^[12]. At the onset, co-occurrence highlights comprise of information about neighboring words. In this methodology, words themselves fill in as highlights. The value of a highlight is the occasions the word happens in the region encompassing the objective word. On the other hand, collocation highlights encode data about expressions of explicit positions that are situated to left or right of the objective word^[12]. This way, typical highlights incorporate the word, the root type of word, together with the word's grammatical form.

i) Supervised techniques

This technique makes use of Artificial Intelligence (AI) systems for actuating a classifier from informational collections that are manually annotated. For the most part, Maas et al.^[13] mention that the classifier, which is frequently termed as a word expert is more worried about a solitary word and performs an arrangement task to allocate the proper meaning to each occurrence of that word. This way, the preparation set used to teach the classifier regularly contains a lot of models where a given objective word is physically labelled with a meaning from the meaning inventory of a reference lexicon^[14].

In this technique, Khan et al.^[15] claim that words can be named with their meanings. For instance: (a) He paid ten per cent"interest"- "money" on the loan he received. (b) He demonstrated a great deal of "interest"- "curiosity/admiration" in her. In supervised approaches, a meaning disambiguation framework is taught from are the presentative arrangement of marked cases drawn from the same circulation as the test set to be utilized^[15].

This way, occurrences to these methodologies are highlight encoded alongside their proper names. The yield of the framework is a classifier framework fit for assigning labels to new component encoded inputs. Ting et al.^[16] specify that the following are some of the existing supervised techniques:-

a) Naive Bayes Classifier

It depends on the reasoning that picking the best meaning for the input vector adds up to picking the most likely meaning of the word in its context^[16].

b) Decision Lists

These classifiers are proportionate to straightforward case explanations in most programming dialects. When it comes to decision lists, Mihalcea^[17] mentions that a classifier is usually a succession of tests is applied to every vector encoded input. If a test succeeds, at that point, the meaning related to that test is returned. On the off chance that test falls flat, at that point next test in the succession is applied. This proceeds until the list is finished, where a default test essentially returns the majority meaning. Support Vector Machine, exemplar based approach and Perceptron trained Hidden Markov Model are another supervised techniques.

no.	Approach	Corpus used	Accuracy reported
1	Naïve bayes	Sense val -3	60.9%
2	Exemplar based approach	WSJ6* containing 191 contents words	63.7%
3	Decision list based approach	Tested on a set of 12 highly polysemous Words.	63.9%
4	Support vector machine	Senseval 3-lexical sample task used for disambiguation of 57 words.	55.2%
5	Perceptron trained HMM	Senseval 3-all words task	60.9%

Table 2 Comparison between supervised approaches

ii) Unsupervised techniques

Tsatsaronis et al.^[18] show that dissimilar to the supervised techniques, unsupervised techniques do not need the hand-named learning of sense data in huge scale assets for the disambiguation. Simply put, the technique depends on the way that words having comparable meanings will have comparable encompassing words. In turn, word meanings are determined by forming bunches of events of words and the undertaking is to classify the new event to the inferred groups. This way, the methodology as opposed to appointing the meanings of names, distinguishes the groups instead. There are two types of unsupervised techniques

a) Context Clustering

This strategy depends on clustering methods in which first setting vectors are designed and after that, they will be

gathered into groups to distinguish the importance of the word. This technique utilizes vector space as word space and its measurements are words as it were^[19]. Likewise in this strategy, a word which is in a corpus will be meant as a vector and how frequently it happens will be tallied inside its setting.

b) Word Clustering

In this method, words having comparable implications are assigned to a similar bunch. As a result, this enables the grouping of words that are the same as the objective world. The similitude between the words is given by linguistic dependency^[20].

Some another approaches are also there which are related to unsupervised algorithms. (Table 3)

no	Approach	Corpus used	Accuracy reported
1	Linn's algorithm	Trained using WSJ corpus containing 25 million words.	64.2%
2	Hyperlex	Tagged on a set of 10 highly polysemous French words.	73%

Table 3 Comparison between unsupervised approaches

III. COMPARISON BETWEEN DIFFERENT APPROACHES

As per the chart shown below we can see that supervised technique performs better when it comes to accuracy but also it depends on the availability of the training datasets.

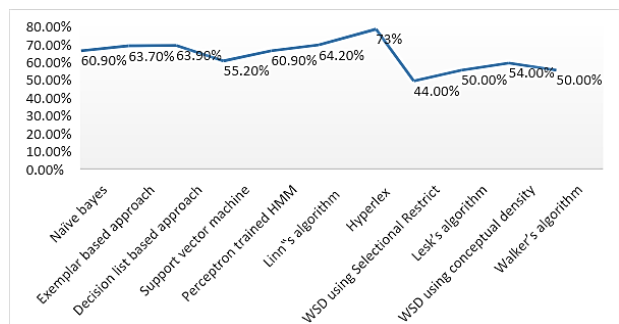


Figure 2 Comparison of different word sense disambiguation techniques

IV. PROPOSED APPROACH FOR GUJARATI WSD:

In Proposed approach, Gujarati language model has been trained using Gujarati Wordnet and Wikipedia.

- Database: Gujarati WordNet (35677 Synsets) and Wikipedia (27,800 pages)
- Algorithm Used: LSTM model
- Input: Gujarati Corpus
- Output: Disambiguated Sense class



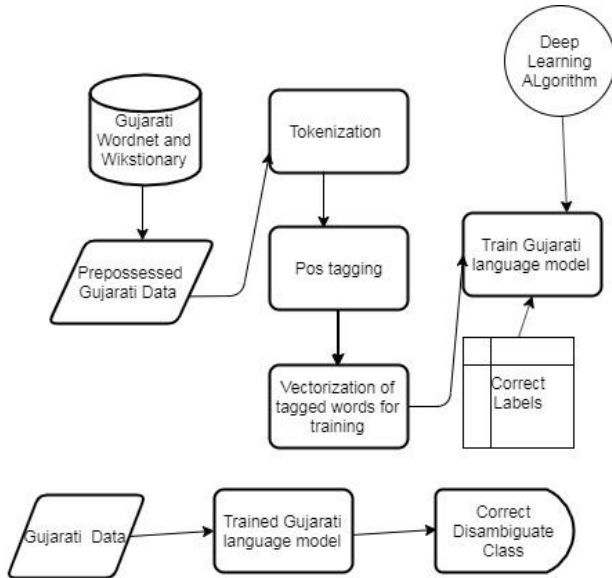


Figure 3 Proposed approach for Gujarati WSD

V. RESULT ANALYSIS

For classification, Gujarati language model has to be trained first so Gujarati statements with class are given as training data. Figure 4, 5 and 6 represent how Gujarati Language model is being trained. Learning rate is the rate at which weights are updated. As shown in figure 4 if learning rate is too high, model never reaches global optima and if too low then also model saturates at a cost higher than lowest. Learning rate is decided by experimenting with different values as shown in figure 5. We can see that learning rate of 0.01 converges to minimum cost. In figure 6 all the scenarios about learning rates are discussed.

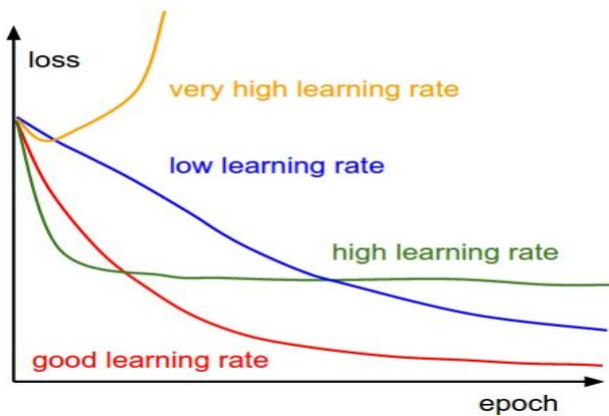


Figure 4 Learning rate V/S loss function

As per given in Table 5, Results are given which describes accuracy and Kappa_score of given data which states if Kappa score is higher than accuracy of disambiguated class is good.

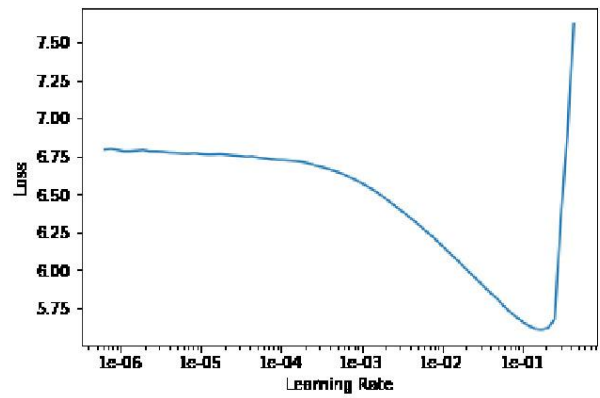


Figure 5 Learning rate V/S loss function

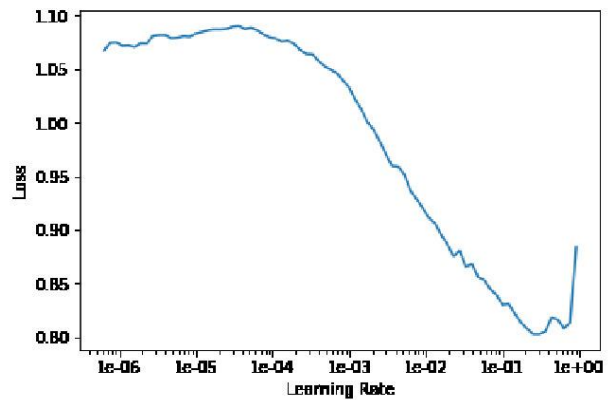


Figure 6 Learning rate V/S loss function

epoch	train_loss	valid_loss	kappa_score	accuracy
1	0.495619	0.301899	0.810159	0.879363
2	0.411036	0.271846	0.836352	0.895296
3	0.329943	0.247750	0.853615	0.906677
4	0.253060	0.261082	0.862736	0.911988
5	0.206439	0.265496	0.856568	0.908194

Table 4 Accuracy reported on Test data

VI. CONCLUSION

As per result shown in result analysis section table 6 compare to other techniques studied this approach is giving better accuracy. Hence we see that the problem of Gujarati word sense disambiguation is a rather hard problem to tackle as the nature of the Devanagari script and the English language is considerably different. The scarcity of labeled data also adds to the problem. Here solutions are proposed but they are just a starting step in the vast majority of the approaches are yet to be explored. Also a lot of help can be obtained by POS Tags and with the Wordnet set up we can also explore such models in the future.

REFERENCES

1. Chen, X., Liu, Z., & Sun, M. (2014). A Unified Model for Word Sense Representation and Disambiguation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 1025-1035).
2. Navigli, R. (2012). A Quick Tour of Word Sense Disambiguation, Induction and Related Approaches. In International Conference on Current Trends in Theory and Practice of Computer Science (pp. 115-129). Springer, Berlin, Heidelberg.
3. Zhong, Z., & Ng, H. T. (2012). Word Sense Disambiguation Improves Information Retrieval. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1 (pp. 273-282). Association for Computational Linguistics.
4. Miller, T., Biemann, C., Zesch, T., & Gurevych, I. (2012). Using Distributional Similarity for Lexical Expansion in Knowledge-Based Word Sense Disambiguation. In Proceedings of COLING 2012 (pp. 1781-1796).
5. Gomaa, W. H., & Fahmy, A. A. (2013). A Survey of Text Similarity Approaches. International Journal of Computer Applications, 68(13), 13-18.
6. Basile, P., Caputo, A., & Semeraro, G. (2014). An Enhanced Lesk Word Sense Disambiguation Algorithm through a Distributional Semantic Model. In Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers (pp. 1591-1600).
7. Wei, T., Lu, Y., Chang, H., Zhou, Q., & Bao, X. (2015). A Semantic Approach for Text Clustering Using WordNet and Lexical Chains. Expert Systems with Applications, 42(4), 2264-2275.
8. Golkar, A., Jafari, S., Fakhrahmad, S. M., & Dashti, S. M. S. (2014). Word Sense Disambiguation Based on Number of Lexica's Senses Weighting in Conceptual Density. International Journal of Artificial Intelligence and Mechatronics, 3(1), 12-15.
9. Kalita, P., & Barman, A. K. (2015). Implementation of Walker Algorithm in Word Sense Disambiguation for the Assamese Language. In 2015 International Symposium on Advanced Computing and Communication (ISACC) (pp. 136-140). IEEE.
10. Borah, P. P., Talukdar, G., & Baruah, A. (2014). Approaches for Word Sense Disambiguation—A Survey. International Journal of Recent Technology and Engineering, 3(1), 35-38.
11. Nameh, M., Fakhrahmad, S. M., & Jahromi, M. Z. (2011). A New Approach to Word Sense Disambiguation Based On Context Similarity. In Proceedings of the World Congress on Engineering (Vol. 1).
12. Moro, A., Cecconi, F., & Navigli, R. (2014). Multilingual Word Sense Disambiguation and Entity Linking For Everybody. In International Semantic Web Conference (Posters & Demos) (pp. 25-28).
13. Maas, A. L., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., & Potts, C. (2011). Learning Word Vectors for Sentiment Analysis. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. Volume 1 (pp. 142-150). Association for Computational Linguistics.
14. Crawford, M., Khoshgoftaar, T. M., Prusa, J. D., Richter, A. N., & Al Najada, H. (2015). Survey of Review Spam Detection Using Machine Learning Techniques. Journal of Big Data, 2(1), 23.
15. Khan, A., Baharudin, B., Lee, L. H., & Khan, K. (2010). A Review of Machine Learning Algorithms for Text-Documents Classification. Journal of advances in information technology, 1(1), 4-20.
16. Ting, S. L., Ip, W. H., & Tsang, A. H. (2011). Is Naive Bayes A Good Classifier For Document Classification? International Journal of Software Engineering and Its Applications, 5(3), 37-46.
17. Mihalcea, R. (2010). Word sense disambiguation. Encyclopedia of machine learning, 1027-1030.
18. Tsatsaronis, G., Varlamis, I., & Nørvg, K. (2010). An Experimental Study on Unsupervised Graph-Based Word Sense Disambiguation. In International Conference on Intelligent Text Processing and Computational Linguistics (pp. 184-198). Springer, Berlin, Heidelberg.
19. Ghosh, A., Mishra, N. S., & Ghosh, S. (2011). Fuzzy Clustering Algorithms for Unsupervised Change Detection in Remote Sensing Images. Information Sciences, 181(4), 699-715.
20. Schnabel, T., Labutov, I., Mimno, D., & Joachims, T. (2015). Evaluation Methods for Unsupervised Word Embeddings. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing.