WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Word Sense Disambiguation Using Supervised Machine Learning Model

Saket Fulewar and Akash Shiv

Department of Computer Science, Nagpur Institute of Technology, Katol Road, Nagpur.

-Introduction to WSD

Word-Sense Disambiguation (WSD) refers to the process of identifying the accurate meaning or sense of a word in a certain context, especially when the word has numerous potential meanings. This field has significant significance in the realm of computational linguistics and language processing.

Precise word sense disambiguation (WSD) is essential in natural language processing activities such as ma-chine translation, information retrieval, and sentiment analysis. Without adequate disambiguation, a comput-er or AI software may mistake the intended meaning of a word in a phrase, leading to errors and inaccurate results.

Word-Sense Disambiguation (WSD) is a challenging issue because of the inherent ambiguity present in language. Developing universal Word Sense Disam-biguation (WSD) algorithms is tough due to the dis-tinct word senses and grammatical structures found in different languages. Moreover, discrepancies in grammar and lexical choices across several geograph-ical regions further complicate the process of resolv-ing ambiguity.

-Methods And Applications

Although word-sense disambiguation (WSD) remains an unresolved issue, notable advancements have been achieved in this sector. Scientists are investigating several methodologies, such as supervised machine learning, unsupervised methods, and knowledge-based strategies, to enhance the precision of WSD systems. Word-Sense Disambiguation (WSD) finds extensive use in several domains. For instance, in the field of machine translation, precise word sense dis-ambiguation (WSD) is of utmost importance in order to accurately choose the appropriate translation of a word, taking into account its surrounding context. Within the field of information retrieval, word-sense disambiguation (WSD) plays a crucial role in enhanc-ing the accuracy of search results by comprehending the intended semantic interpretation of search queries. Additionally, WSD contributes to the reliability of sentiment analysis outcomes by facilitating the precise categorization of sentiment words according to their context.

- Overview of the Marathi language and its specific challenges in WSD

There are several options to resolving the sentence's ambiguity. For example, knowledge-based approach-es, machine learning methods, or a combination of the two. In this case, we utilized a machine learning tech-nique that better matches our objective and meets the criteria for solving the issue.

-Previous Work In the field

Many researchers have developed techniques to re-solve ambiguity (Yoong et al., 2004). One of these researchers has conducted research and developed an SVM-based word sense disambiguation utilising the Lesk algorithm, which was first presented by Michael E. Lesk in 1986. The objective of the Lesk algorithm is to ascertain the appropriate meaning of a word in a certain setting by analysing the degree to which the meanings of individual terms overlap with one anoth-er. The concept is predicated on the notion that the meaning of a word may be deduced by examining the words that are located in close proximity to it.

-Support Vector Machine

The main model to be used for training is the support vector machine. The model is a regression model that utilizes a classifier-based approach and has a high level of robustness. The support vector machine (SVM) is a widely used machine learning technique known for its ability to handle high-dimensional data and complex classification tasks. It works by finding the optimal hyperplane that separates different class-es of data points, maximizing the margin between them. SVMs have been successfully applied in vari-ous domains, including natural language processing and text classification. In this study, we propose to use SVM as the main model for training our WSD system in Marathi language, as it offers a promising solution to

- Preprocessing steps for Marathi language text, including tokenization and stemming

The IndoWornet dictionary provides distinct connections between synsets, which are sets of synonyms that reflect individual ideas. It was created at IIT, Bombay, by a group led by Dr. Pushpak Bhat-tacharya.

Word Sense Disambiguation for the target word: The objective of the Word Sense Disambiguation (WSD) system is to resolve the meaning of a certain set of target words, typically one word per phrase. In this context, supervised methods are often used, using a tagged corpus to train the model. Subsequently, the trained model is used to resolve the ambiguities in-cluded in the target document.

Paradigmatic connections, such as synonymy, hy-ponymy, antonymy, and entailment, are used in its construction. Marathi WordNet is a highly used lexi-cal database in current NLP research for the Marathi language. It provides a comprehensive list of syn-onym sets or synsets for open-class words such as nouns, verbs, adjectives, and adverbs. The index_txt file contains information about all the words in Wordnet. The data_txt file provides details about each word in the index file. The onto_txt file provides ontology details about the words in the data file.

Example:

data_txt FileStructure of Data_txt: Example 00054554 03 02

फसवणे:चकवणे 0001 0400 00000183 | फसेलअसे

करणे:"नकली मालाची ववक्री करून दकानदारु लोकांना फसवतात."

Synset id=00054554

POS=03 number of words present in synset=02

Synsets= फसवणे:चकवणे

Number of relations lexical as well as

semantic=0001

Four-digit code relation id=0400 synset_id for

which that relation exists=00000183 gloss= फसेल

असेकरणे

-Training & Evaluation Techniques

We possess a dataset that has been annotated with the appropriate sense placeholder. The SVM model is trained using a word dictionary, which enables the machine model to identify and analyse words that are similar in meaning.

-References

Aparitosh Gahankari*, Dr. Avinash S. Kapse, Mohammad Atique, Dr. V.M. Thakare & Dr. Arvind S. Kapse, “Word Sense Disambiguation in Marathi Language using Fast-Text model and Indo-Wordnet”, CIMS, vol. 28, no. 11, pp. 1607–1617, Dec. 2022.

Aparitosh Gahankari, Dr. Avanish S.Kapse, Dr. V. M. Thakre, " Word Sense Disambiguation - Supervised Approaches: Present Scenario", International Journal of Scientific Research in Science and Technology(IJSRST), Print ISSN : 2395-6011, Online ISSN : 2395-602X, Volume 5, Issue 6, pp.150-154, January-February-2020.

Yoong, Hwee, Tee, & the. (2004). 137140