World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:17, No:11, 2023

Contextual SenSe Model: Word Sense Disambiguation using Sense and Sense Value of Context Surrounding the Target

Authors : Vishal Raj, Noorhan Abbas

Abstract : Ambiguity in NLP (Natural language processing) refers to the ability of a word, phrase, sentence, or text to have multiple meanings. This results in various kinds of ambiguities such as lexical, syntactic, semantic, anaphoric and referential am-biguities. This study is focused mainly on solving the issue of Lexical ambiguity. Word Sense Disambiguation (WSD) is an NLP technique that aims to resolve lexical ambiguity by determining the correct meaning of a word within a given context. Most WSD solutions rely on words for training and testing, but we have used lemma and Part of Speech (POS) tokens of words for training and testing. Lemma adds generality and POS adds properties of word into token. We have designed a novel method to create an affinity matrix to calculate the affinity be-tween any pair of lemma_POS (a token where lemma and POS of word are joined by underscore) of given training set. Additionally, we have devised an al-gorithm to create the sense clusters of tokens using affinity matrix under hierar-chy of POS of lemma. Furthermore, three different mechanisms to predict the sense of target word using the affinity/similarity value are devised. Each contex-tual token contributes to the sense of target word with some value and whichever sense gets higher value becomes the sense of target word. So, contextual tokens play a key role in creating sense clusters and predicting the sense of target word, hence, the model is named Contextual SenSe Model (CSM). CSM exhibits a noteworthy simplicity and explication lucidity in contrast to contemporary deep learning models characterized by intricacy, time-intensive processes, and chal-lenging explication. CSM is trained on SemCor training data and evaluated on SemEval test dataset. The results indicate that despite the naivety of the method, it achieves promising results when compared to the Most Frequent Sense (MFS) model.

Keywords : word sense disambiguation (wsd), contextual sense model (csm), most frequent sense (mfs), part of speech (pos), natural language processing (nlp), oov (out of vocabulary), lemma_pos (a token where lemma and pos of word are joined by underscore), information retrieval (ir), machine translation (mt)
Conference Title : ICNLP 2023 : International Conference on Natural Language Processing
Conference Location : Jeddah, Saudi Arabia
Conference Dates : November 20-21, 2023