Abstract

Word Sense Disambiguation (WSD) aims to disambiguate the words which have multiple sense in a context automatically. Sense denotes the meaning of a word and the words which have various meanings in a context are referred as ambiguous words. WSD is vital in many important Natural Language Processing tasks like MT, IR, TC, SP etc. This research paper attempts to propose a supervised Machine Learning approach- Decision Tree for Word Sense Disambiguation task in Assamese language. A Decision Tree is decision model flow-chart like tree structure where each internal node denotes a test, each branch represents result of a test and each leaf holds a sense label. J48 a Java implementation of C4.5 decision tree algorithm is taken for experimentation in our case. A few polysemous words with different real occurrences in Assamese text with manual sense annotation was collected as the training and test dataset. DT algorithm produces average F-measure of .611 when 10-fold crossvalidation evaluation was performed on 10 Assamese ambiguous words.

References
2. Sarmah, J. and Sarma, S.K., Survey on Word Sense Disambiguation: an initiative towards an Indo-Aryan Language. Accepted in IJEM, March 2016, ISSN: 2305-3631 (Print), ISSN:2306-5982 (Online)

Index Terms

Computer Science  Information Sciences

Keywords

Word Sense Disambiguation, Decision Tree, Assamese, Supervised approach