Abstract

Language is one of many tools used by humans to communicate with each other. Languages in different areas usually differ between one and another, which certainly leads in every language having their own unique patterns, sequences, and structures. Because of its significant function and complex structure, many researches have been made to study all about language. One of the important things in finding a meaning of a word in context of a sentence is to identify its tag/word type.

POS Tag approach based on Graph Clustering method specifically using the Chinese Whisper algorithm is used to find the clusters/groups of these words where the weight of each word reflects on the similarity of each pair. Form this process the granularity of clusters are usually finer than average tag sets hence needing an extra mapping process using the many-to-1 accuracy. To gain the tag sequence of an input text a Viterbi algorithm is utilized where the values of the matrix used are gained from the clustering process.

Clustering using the Chinese Whisper method is not satisfactory if the corpus used contains many words of low frequency so when the HFW and MFW threshold are increased, to gain higher relevancy words in one cluster, will effect in many words being left out making the tag sequence sparse.

Keywords: POS Tagging, clustering, Graph Clustering, Tagging, Chinese Whisper