Optimal part-of-speech tagging have great importance in various field of natural language processing such as machine translation, information extraction, word sense disambiguation, speech recognition and others. Due to the special nature of the Nepali language, Tagset used and Size of the corpus (training data), getting accurate part-of-speech tagger is one of the challenging task. This study is oriented to build an analytical machine learning model based on which it can be possible to determine the attainable accuracy. To complete this task, the support vector machine based part-of-speech tagger has been developed and tested for various instances of input to verify the accuracy level. The SVM tagger construct the feature vectors for each word in input and classify the word into one of two classes (One Vs Rest).

References

Support Vector Machines based Part of Speech Tagging for Nepali Text


Index Terms

Computer Science Artificial Intelligence

Keywords

Support Vector Machine POS Tagging HMM Supervised Machine Learning