Arabic Part-of-Speech Tagger, an Approach Based on Neural Network Modelling

Rabab Ali Abumalloh, Hasan Muaidi Al-Serhan, Othman Bin Ibrahim, Waheeb Abu-Ulbeh

POS-tagging gained the interest of researchers in computational linguistics sciences in the recent years. Part-of-speech tagging systems assign the proper grammatical tag or morpho-syntactical category labels automatically to every word in the corpus per its appearance on the text. POS-tagging serves as a fundamental and preliminary step in linguistic analysis which can help in developing many natural language processing applications such as: word processing systems, spell checking systems, building dictionaries and in parsing systems. Arabic language gained the interest of researchers which led to increasing demand for Arabic natural language processing systems. Artificial neural networks has been applied in many applications such as speech recognition and part of speech prediction, but it is considered as a new approach in Part-of-speech tagging. In this research, we developed an Arabic POS-tagger using artificial neural network. A corpus of 20,620 words, which were manually assigned to the appropriate tags was developed and used to train the artificial neural network and to test the part of speech tagger systems' overall performance. The accuracy of the developed tagger reaches 89.04% using the testing dataset. While, it reaches 98.94% using the training dataset. By combining the two datasets, the accuracy rate for the whole system is 96.96%.