Marwah Alian, Arafat Awajan, Generating Sense Inventories for Ambiguous Arabic Words, The International Arab Journal of Information Technology (IAJIT). 2021, (ISI,Scopus)

Abstract

The process of selecting the appropriate meaning of an ambigous word according to its context is known as word sense disambiguation. In this research, we generate a number of Arabic sense inventories based on an unsupervised approach and different pre-trained embeddings, such as Aravec, Fast text, and Arabic-News embeddings. The resulted inventories from the pre-trained embeddings are evaluated to investigate their efficiency in Arabic word sense disambiguation and sentence similarity. The sense inventories are generated using an unsupervised approach that is based on a graph-based word sense induction algorithm. Results show that the Aravec-Twitter inventory achieves the best accuracy of 0.47 for 50 neighbors and a close accuracy to the Fast text inventory for 200 neighbors while it provides similar accuracy to the Arabic-News inventory for 100neighbors. The experiment of replacing ambiguous words with their sense vectors is tested for sentence similarity using all sense inventories and the results show that using Aravec-Twitter sense inventory provides a better correlation value