[P17] Network of the words in Korea presidential speech using Word2Vec

Young-Jai Park, Hanyang university

Presidential speech is always in the middle of the public concern since it reflects domestic and foreign situation and government policies as its response. In this study, we analyze the total 6851 Korean presidential speeches, about 9 million words, from 1948, the first year of the first president Syngman Rhee, to 2015 of the current president Geun-hye Park to quantify the impact of the social issues in specific periods. Using the word embedding algorithm in machine learning, we find the locational patterns of words in a text of a speech as well as the basic stylometric analysis such as Zipfs law [1] by counting word frequencies and the correlations between the most-frequently spoken 100 words. From the spatial locations of word vectors, we extract the spoken word networks for each president using the angular distance between word vectors [2]. Through the spoken word networks, we try to understand the intended purpose of the governmental policies and the domestic and international situations of Korea directly and intuitively.

[1] Ruokuang Lin, Chunhua Bian and Qianli D. Y. Ma, ”Scaling laws in human speech, decreasing emergence of new words and a generalized model”, arXiv:1412.4846v2 [cs.CL] 7 Jan 2015.
[2] Tomas Mikolov, Kai Chen, Greg Corrado and Jeffrey Dean, ”Efficient Estimation of Word Representations in Vector Space”, arXiv:1301.3781v3 [cs.CL] 7 Sep 2013.