Lecture 1.2 (11:55-12:20)

Probabilistic Latent Semantic Scaling

Daichi Mochihashi (ISM)

There are many situations in which we want to continuously measure a text according to a certain polarity, e.g., right vs left, positive vs negative. Traditionally, Latent Semantic Scaling (LSS, Watanabe(2020)) is often used together with the associated R package developed in the field of political science methodology. LSS is based on a classical vector space model, thus its estimation results are usually very noisy. In this presentation, we present a PLSS (Probabilistic Latent Semantic Scaling), which views LSS as a probabilistic model based on item response theory (IRT), where modern word embeddings such as word2vec can be leveraged together. The latent polarity of the text is modeled as a θ in IRT, and the parameter is computed by integrating it out by an adaptive Gauss-Hermite quadrature. Experiments were conducted on the published scaled data of Young and Soroka (2012), and PLSS is shown to compute the latent scale of the text much better than LSS.