Investigation of Topic Modeling-Based Semi-Supervised Interpretable Document Classifier

Authors : Dasom Kim, William Xiu Shun Wong, Yoonjin Hyun, Donghoon Lee, Minji Paek, Sungho Byun, Namgyu Kim

Abstract : There have been many researches on document classification for classifying voluminous documents automatically. Through document classification, we can assign a specific category to each unlabeled document on the basis of various machine learning algorithms. However, providing labeled documents manually requires considerable time and effort. To overcome the limitations, the semi-supervised learning which uses unlabeled document as well as labeled documents has been invented. However, traditional document classifiers, regardless of supervised or semi-supervised ones, cannot sufficiently explain the reason or the process of the classification. Thus, in this paper, we proposed a methodology to visualize major topics and class components of each document. We believe that our methodology for visualizing topics and classes of each document can enhance the reliability and explanatory power of document classifiers.