Emerging Topic Detection based on LDA Combined with Bibliometrics Indices

Jianxia Ma

majx@lzb.ac.cn
Lanzhou Library of Chinese Academy of Sciences (China)

Abstract

According to the study of features of emerging topic, we proposed a set of bibliometrics indices for emerging topic features. We employ novelty index (NI) , published volume index (PVI) put forward by Tu Yining and Seng Jia Lang, and propose a new index, cited volume index to characterize the emerging topic. Then we proposed a method to identify the features of the emerging topic based on the LDA(Latent Dirichlet Allocation) model. The first step is to extract the topical words of the documents using the LDA model, the next is to build the mapping from topics to documents using the document-topic matrix, and then to visualize the life span of an emerging topic especially with novelty index (NI) , published volume index (PVI), cited volume index and the detection point to characterize the emerging topics. We depict the features of the development of topics in the period of the born, potential emerging, emerging in the topic life cycle. According to the method a toolkit is developed to carry out emerging topic detection. With this method and tool, we carried out an experiment on the corpus covering "machine learning" downloaded from the Web of Science to prove after adding the time dimension into the indicators, we detect emerging topics from the corpus. We verify utilizing LDA to extract topics can avoid the semantic ambiguity of frequency of words. We find combined cited volume index with novelty index and published volume index, we can detect the emerging topic earlier. And we analyze the effectiveness and validity of the indices and method we supposed.

Keywords

LDA(Latent Dirichlet Allocation);Topic Model;Emerging Topic Indices;Emerging Topic Detection; Novelty Index;Published Volume Index;Cited volume Index; Life Cycle