

## Online Topic Model for Broadcasting Contents Using Semantic Correlation Information

**Authors :** Chang-Uk Kwak, Sun-Joong Kim, Seong-Bae Park, Sang-Jo Lee

**Abstract :** This paper proposes a method of learning topics for broadcasting contents. There are two kinds of texts related to broadcasting contents. One is a broadcasting script which is a series of texts including directions and dialogues. The other is blogposts which possesses relatively abstracted contents, stories and diverse information of broadcasting contents. Although two texts range over similar broadcasting contents, words in blogposts and broadcasting script are different. In order to improve the quality of topics, it needs a method to consider the word difference. In this paper, we introduce a semantic vocabulary expansion method to solve the word difference. We expand topics of the broadcasting script by incorporating the words in blogposts. Each word in blogposts is added to the most semantically correlated topics. We use word2vec to get the semantic correlation between words in blogposts and topics of scripts. The vocabularies of topics are updated and then posterior inference is performed to rearrange the topics. In experiments, we verified that the proposed method can learn more salient topics for broadcasting contents.

**Keywords :** broadcasting script analysis, topic expansion, semantic correlation analysis, word2vec

**Conference Title :** ICAIPR 2016 : International Conference on Artificial Intelligence and Pattern Recognition

**Conference Location :** Singapore, Singapore

**Conference Dates :** January 07-08, 2016