Enhanced Frequent Itemsets Based on Topic Modeling in Information Filtering

Than Than Wai
thanwai85@gmail.com
University of Computer Studies, Mandalay, UCSM

Sint Sint Aung
ssaung@gmail.com
University of Computer Studies, Mandalay, UCSM

In order to generate user's information needs from a collection of documents, many term-based and pattern-based approaches have been used in Information Filtering. In these approaches, the documents in the collection are all about one topic. However, user's interests can be diverse and the documents in the collection often involve multiple topics. Topic modeling is useful for the area of machine learning and text mining. It generates models to discover the hidden multiple topics in a collection of documents and each of these topics are presented by distribution of words. But its effectiveness in information filtering has not been so well explored. Patterns are always thought to be more discriminative than single terms for describing documents. The major challenge found in frequent pattern mining is a large number of result patterns. As the minimum threshold becomes lower, an exponentially large number of patterns are generated. To deal with the above-mentioned limitations and problems, in this paper, a novel information filtering model, EFITM (Enhanced Frequent Itemsets based on Topic Model) model is proposed. Experimental results using the CRANFIELD dataset for the task of information filtering show that the proposed model outperforms over state-of-the-art models.