Han-joon Kim et al., J Comput Eng Inf Technol 2016, 5:4(Suppl)
http://dx.doi.org/10.4172/2324-9307.C1.006

Building cluster-based word networks from textual data

Han-joon Kim and **Han-mook Ryu**
University of Seoul, Korea

This paper describes a new way of producing more significant word networks from textual data by combining text clustering and keyword association techniques. Basically, one of crucial aspects in text mining is the analysis of concept relationships, where concepts originate from keywords. The problem is to discover more reasonable set of keywords and their relationships called 'word network'. In general, the word networks can be built by using the frequency of co-occurrence of words indexed. However, only the co-occurrence frequency is not enough to measure the strength of associations among words because significant associations with relatively low frequency are ignored. In our work, to overcome the problem, we intend to perform the word association task over the clustered results for incoming documents instead of a whole document. Rather than building a word network from the entire set of documents, it is likely to extract more meaningful word associations from the clustered results of the documents. Our proposed method is performed broadly in two steps: Firstly, a given documents collection is partitioned into a set of clusters, each of which is represented as a minimum spanning tree by conducting a priori association mining. Here, we note that each cluster includes a set of documents with similar word occurrence patterns, and thus it would have cluster-specific words and their strong associations. Thus, as a next step, our method iteratively compute weighted mutual information that evaluates the degree of significance between two word nodes, and extracts the top-N significant words and their word associations hidden in each cluster.

Biography

Han-joon Kim has received his BS and MS degrees in Computer Science and Statistics from Seoul National University, Seoul, Korea in 1994 and 1996, respectively. He has done his PhD degree in Computer Science and Engineering from Seoul National University, Seoul, Korea in 2002. He is currently a Professor at the School of Electrical and Computer Engineering, University of Seoul, Korea. His current research interests include Data/Text Mining, Database Systems, and Intelligent Information Retrieval.

khj@uos.ac.kr

Notes: