

Identification and Extraction of Binary, Ternary, Transitive associations and Frequent Patterns from Text Documents in an Interactive Way

Shruti Dilip Waranashiwar¹

¹Department of Computer and Information Science, School Of Science, IUPUI

As the amount of electronically accessible textual material has been growing exponentially, Text mining is a new and exciting research area that tries to solve the information overload problem. It is a promising and automated approach for extracting knowledge from unstructured textual documents. The purpose of this research in text mining area is to find compact but high quality associations from Neuroscience related text documents. Here, we try to find the relationships (binary, ternary and transitive) between the terms related to some of the common disorders in neuroscience like Alcoholism and Schizophrenia from a database PubMed, using Vector Space Model (VSM) and the Artificial Neural Network (ANN). We also use Graphviz to visualize these associations. This research reveals many stronger and weaker associations between the different terms in different comorbidities, which are otherwise difficult to understand by reading articles or journals manually. Once the model is developed, it can be generalized to different terms and can be used to study different combinations of terms and comorbidities. As response time of these models is very fast, it will greatly contribute towards speeding up medical research. In such light, extracting associations between keywords could provide very interesting insights into their roles in various diseases and other biological processes.

We also try to prove that instead of mining all frequent patterns, all of which may not be interesting to user, interactive method to mine only desired and interesting patterns is far better approach in terms of utilization of resources. We find the compact but high-quality frequent patterns in an interactive way using MCMC sampling method. In interactive patterns mining, a user gives feedback on whether a pattern is interesting or not. The discovery of interesting Associations has application in many fields. Few of them are business decision-making processes, web usage mining, intrusion detection and bioinformatics.

Mentor : Dr. Snehasis Mukhopadhyay, Department of Computer and Information Science, IUPUI