Abstract

This research paper combines the data mining with natural language processing to extract the nuggets of knowledge from massive volume of student feedback dataset on faculty performance. The main objective is to compare two renowned association rule mining and sequential pattern mining algorithms namely Apriori and Generalized Sequential Pattern (GSP) mining in the context of extracting frequent features and opinion words. Student feedback data crawled, pre-process and tagged, then convert in tri-model data files. Both algorithms are applied on prepared data through WEKA 3. 7. 10 (a machine learning tool) to extract the rules. Mined rules are applied on testing files to extract frequent features and opinion words. Evaluated Results show that GSP is more significant to use for textual data mining than Apriori.
References

- Zaki, M. J. "Efficient enumeration of frequent sequences", in &apos;7th International Conference on Information and Knowledge Management&apos;, ACM Press, Bethesda, Maryland, United States, pp. 68–75 in 1996.
- N, Anwer and A, Rashid, "Feature Based Opinion Mining of Online Free Format..."


Index Terms

Computer Science
Data Mining

Keywords

Opinion Mining  Sentence Level Sentiment Classification  Sequential Pattern Mining  Apriori
Generalized Sequential Pattern

Opinion Words

Frequent Features