ABSTRACT (English)

Confidence Value Analysis of FP-Growth Algorithm on Oracle RDBMS using Set Theory and Unsupervised Learning

Boby Siswanto

Supervisor: Prof. Dr. The Houw Liong

Co-Supervisor: Shaufiah, MT

Values used as a reference of the association rule mining are support value and confidence value. The higher the value the support and confidence value, the better the resulting rules. Association rule mining algorithms apply unsupervised learning because the resulting rule is not determined to be a certain class. Performance of association rule mining algorithms rely heavily on the dataset size / dimensions are used. Performance can be measured from the time processing is generated. The larger the dataset, the dimensions will be greater and the processing time will be longer. If the dimensionality of dataset can be pruned, the processing time will be faster and the performance will be better, with confidence values relatively unchanged.

Intersection is a kind of set theory that can reduce the number of attributes on related sets. Oracle is one of the RDBMS, related sets can be applied to the Oracle RDBMS as the related tables. IST-EFP algorithm is a proposed algorithm that combines the EFP (Expand FP-Growth) with set theory.

In this study, IST-EFP algorithm can reduce the dimension of the dataset to 87.5% with a 26.6% improvement on time processing. The confidence value obtained was relatively unchanged, as an example of FP-18 $\Rightarrow$ FKUE72 rule, previously the rule have 23.4% confidence value, after IST-EFP implemented confidence value change to 23.41%.

This results can be used for business actions.

Keyword: association rule mining, unsupervised learning, apriori, fp-growth, dataset, set theory, pruned, intersections, Oracle RDBMS, business actions.