

Fast Algorithms for Mining Association Rules

This paper presents (2+1) new algorithms for association rule discovery in sales transaction databases. These new algorithms are Apriori, AprioriTid and the hybrid of these two, AprioriHybrid. The authors demonstrate that Apriori algorithms have better calculation speeds than the existing algorithms like AIS and STEM. The experiment results for different number of transaction show that Apriori and AprioriTid are significantly faster than the AIS and STEM. Also their performance gap is increasing with the number of transactions. Finally, the authors combine Apriori and AprioriTid to create a more scalable and faster AprioriHybrid algorithm.

The subject of the paper is creating faster algorithm for a specific task. And the authors developed not only one algorithm but 2 (actually 3) new algorithms for the same task. Paper uses mathematical expressions such as lemmas and theorems for better explanation of the new algorithms. Since the authors try to present faster algorithms, they compare them with current algorithms. Authors describe how AIS and STEM algorithms' work and they also explain their data structures and buffer managements. Paper also mentions that they used best data structure for AIS algorithm to perform as fast as possible. This shows that, new algorithms can even beat the previous algorithms' best state. And finally, experiment results/graphs indicate that Apriori algorithms are way faster. The new algorithms also perform even better for lower minimum support values such as 0.5 and 0.25.

When I read the paper for the first time, I couldn't understand some of the statements. I think the authors should have explained the expressions like candidate sets, minimum support and minimum confidence more clearly. In the introduction section, they should have used more examples to introduce the topic to the readers. The other weakness is not using visuals when describing the algorithms. For instance, they mentioned they used the hash tree structure for Apriori algorithm. I expected a visual/graph of that hash tree but, the paper has no visuals other than the experiment result graphs and some tables. Apriori algorithms are faster than the previous ones but they require larger memory space than the old algorithms. So, if the system memory isn't sufficient, algorithm performance may not be that fast.

As I mentioned in weakness part, better explanations for some of the terms would be better for people who are new to the field/problem. Also, visuals/graphs for the set operations or data structures (like hash tree) could improve readers' motivation and understanding. The last thing that the authors could improve is examples. They used some examples for algorithms but those were basic set examples. I would like to see some real world examples with shopping transactions.

2019-01-20

Baran Kaya, kayab@mcmaster.ca, 400284996