TDT4300 Datavarehus og datagruvedrift - Spring 2013
Assignment 2: Apriori Rule Generation

1 Association Rules

This assignment will deal with association rules, frequent itemsets, and the apriori algorithm to create both.

1.1 Apriori Frequent Itemset Generation

Three different methods for apriori frequent itemset generation are described in the book:

1. Brute-Force method
2. $F_{k-1} \times F_1$ method
3. $F_{K-1} \times F_{k-1}$ method

Your task in this exercise is to implement all three variants in Java and submit the results. Starting from the attached class files, implement the missing methods. Further your code should count the number of candidates and pruning steps for each of the methods.

Attached you find the following classes:

- ItemSet represents an itemset, the elements of which are sorted alphabetically. Use the methods of this class in your implementation.
- AprioriItemSetGeneratorAndRuleGeneration is the main class for running your implementations. It reads the dataset described in the book and runs all three different apriori implementations.

It is your task to implement the following classes:

1. *BruteForce*, implement the method *apriori*
2. *FKMinus1F1Apriori*, implement *aprioriGen*

3. *FkMinus1FKMinus1*, implement *aprioriGen*

Put your focus on the complexities of the different solutions. Explain which method scales best and why.

1.2 Apriori Rule Generation

Once your frequent itemset generation is opertional, implement the apriori-rule-gen method which is described in the book. Implement *generateRulesBase* in *AbstractApriori*. This method is first called by *generateAllRules*, and will call itself recursively in the following.

1.3 Datasets

There are two available datasets that you should use. A small one ("smallDataset.txt") a larger one ("supermarket.arff"). The small one is similar to the ones used in the book and you can use it to verify if your code runs correctly. Experiment with both datasets in order to draw results for the complexity and the scalability of the methods.

Notes

Code template

Extract the jar file with name "aprioriTemplate" and import the project to Eclipse (File→ Import → Import existing projects in workspace).

Submission to it's learning

Your submission in its learning is a **jar** file with your implementation (**executable jar**, in Eclipse: export/export jar file/add sources and class files, mark main class) and a **pdf description** of what you did (tables of the different numbers of pruning steps of the different methods). Your submission should be between two and three pages.