

ASSOCIATION ANALYSIS

tdt4300-undass@idi.ntnu.no

Spring 2022

1 Apriori Algorithm

Given the data in Table 1 (market basket transaction), your task is to describe the purchasing behavior of customers in the form of association rules. Use the Apriori algorithm to find all frequent itemsets with minimum support 33.33% (i.e. minimum support count is 2).

- (a) Show thoroughly the steps on how the frequent itemsets are generated.
- (b) one of the frequent itemsets is $\{H, C, I\}$. Find all association rules based on this set, given confidence threshold $c = 60\%$ (it is not necessary to use Apriori to find the association rules, but show how confidence is calculated for each of the candidate rules that are based on $\{H, C, I\}$.)

TID	Items
T1	H, B, K
T2	H, B
T3	H, C, I
T4	C, I
T5	I, K
T6	H, C, I, U

Table 1: Market basket transactions1.

2 FP-Growth Algorithm

Use the Frequent Pattern Growth algorithm to discover the frequent itemsets in the given transaction dataset (Table 2). You are now going to use the FP-growth algorithm in order to find all frequent itemsets with minimum support of 22 % (i.e., minimum support count is 2). Construct an FP-tree and mine the frequent itemsets by creating conditional (sub-)pattern bases. Use the table notation with columns: item, conditional pattern base, conditional FP-tree, frequent patterns generated. The recursive steps of the FP-Growth algorithm must be clearly captured using the aforementioned table notation. Sort items alphabetically in case of ties in the item support. **Describe thoroughly the process and the outcome of each step.**

TID	Items
T1	b, e, g
T2	b, d, i
T3	b, d, e, f
T4	a, d, e
T5	d, e
T6	b, d, j
T7	b, c, d, e, f
T8	b, d, e, f
T9	b, e, h

Table 2: Market basket transactions 2.

3 KNIME

For this task you will need to install and use the KNIME¹ data analytics platform. You are given a file *market_basket_transactions.arff*, Your task is to implement two simple workflows for mining association rules, one implementing Apriori algorithm and second implementing FP-Growth algorithm. Use the WEKA nodes both for Apriori and FP-Growth. Use $minsup = 0.5$ and $minconf = 0.8$. **Present pictures of your workflows, and the outputs from both Apriori and FP-Growth nodes. Deliver also the exported KNIME workflows.**

Note: In Knime 4.X Weka nodes are not automatically included. Go to this link² and follow

¹<http://www.knime.org/downloads/overview>

²https://hub.knime.com/knime/extensions/org.knime.features.ext.weka_3.7/latest

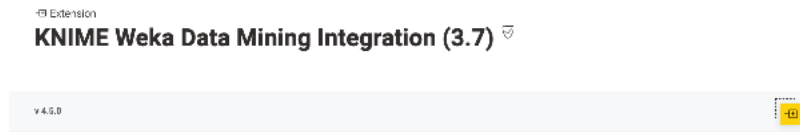


Figure 1: Add Weka rules to KNIME

the directions under *Add to KNIME Analytics Platform* (drag the extension to your Workflow board in KNIME to install). *Click the add symbol on the right and drag it to KNIME. You need to accept cookies first to see this button.*

4 Compact Representation of Frequent Itemsets

Given the compact representation of frequent itemsets in Table 3, use the appropriate algorithm to generate all frequent itemsets including the support counts. **Describe thoroughly each step of the algorithm and present the final result.**

Closed Frequent Itemsets	Support Count
{b}	10
{d}	13
{a, d}	11
{b, d}	7
{b, e}	8
{d, e}	6
{a, b, e}	7
{a, c, d}	6
{b, d, e}	4
{a, c, d, e}	5

Table 3: Closed frequent itemsets.

Submission Requirements

In this assignment we expect you to submit following artifacts:

- A PDF file with the report.
 - Text must not be handwritten.
 - Make sure that the document follows the usual conventions (**names**, assignment/task number, etc.).
- A ZIP file with KNIME workflows.

All assignment artifacts are to be delivered using ***BlackBoard***. You are allowed to **work in pairs**, however, the identical artifacts must be delivered individually.