

Exercises for DW & DM

Sheet 3 (until 24.01.2017)

Please drop your solution in the silver homework box (second floor where the IfIS is located) until Thursday, before the lecture (date is also mentioned above). You may answer in either German or English. **You are encouraged to work in teams of 2 students** (not more than 2), and send your solution as a team. Please mention in your email the **name of both students** together with the corresponding **inmatriculation numbers**.

Exercise 1

Simulate the functionality of the Multiple Minimum Supports mining algorithm on the transactions provided in Annex 1, describing each step starting from $k=1$ up to the rules. Minimum support values are also provided in Annex 1. $\varphi = 20\%$ and $\text{minconf} = 60\%$.

Exercise 2

Simulate the functionality of the GSP mining algorithm of the transactions provided in Annex 2, considering a minimum support of 2.

Hint for the generalization step:

Joining:

Two sequences, s_1 and s_2 can be joined if after dropping the first item from s_1 and the last item from s_2 , we obtain the same sequence. E.g.: $\langle bc \rangle$ and $\langle ca \rangle$ can be joined since by dropping b from $\langle bc \rangle$ and a from $\langle ca \rangle$ we obtain $\langle c \rangle$ for both. The joined result is $\langle bca \rangle$. $\langle ba \rangle$ and $\langle (ab) \rangle$ can also be joined and we obtain $\langle b(ab) \rangle$

Pruning:

Similar to the apriori algorithm $\langle bca \rangle$ passes pruning only if $\langle bc \rangle$, $\langle ba \rangle$ and $\langle ca \rangle \in F_2$. $\langle b(ab) \rangle$ passes pruning only if $\langle ba \rangle$, $\langle bb \rangle$ and $\langle (ab) \rangle \in F_2$.

Exercise 3

1. Calculate the moving average of order 4, $MA(4)$, for the following dataset:
4.38, 4.19, 4.65, 6.40, 6.26, 13.51, 4.19, 8.41, 6.50, 8.43, 9.87, 9.56, 6.57, 9.03, 10.18

And then represent the two lines (the dataset and the MA in a chart).

2. Shortly present the role of R-trees in similarity search for time series.

Annex 1:

Item	MIS %
1	70
2	17
3	15
4	30
5	30
6	35

Transactions
1,4,6
1
1,5,6
1,6
4,6
1,2,3,5
1,2,3,5
6
1
1,6

Annex 2:

SID	Sequence
1	<(dc)b(ac)>
2	<bc(bac)>
3	<(ab)a>