Assignment 3

Task I: Business understanding: business goals

- problemdefinition:
  In case of changing a module we want to know the relations between modules

- reduce costs for development process:
  → through minimal amount of module
  → less complexity
  → most efficient usage of the resources

- improve customer satisfaction through no bugs in his software

Task I : Business Understanding business goals

- who has the problem?
  → the company who want to use the system

- definition in company term:
  → detect the degree of component coupling to be able to ensure the most efficient usage of the resources

- succes: if we can prevent bugs during the change of modules
  → more efficient development, reduced costs for development

Task I : Business Understanding data mining goals

- data mining goal : create a model which is able to show the coupling between modules

- success criteria:

  → coupling estimated by the model equals coupling in real projects

  → when rules are based on TRUE-Values

Task II : Data Understanding

- - The Data show an extract of the development of the X-Window system
- - In the first column are given the change-id's
- - For every change-id, there are the sets of the modules containing the changes
  which were made, with 1 and 0 as possible values
- - 10875 changes
- - 14 different modules

Task III : Data Preparation

- It is necessary to transform the data for weka

- We convert the data as hereinafter: 1 -> t 0 -> FALSE

- We also have to take off the change-id column, because the numerical entries here don't give us information about the component coupling

Task IV : Modeling

- no split of the data set, using every information
- using the apriori-algorithm to do association
- running the algorithm gives us as output these 19 rules as rules with the highest probability

Task V : Evaluation

1. Module_id_4=t Module_id_5=t Module_id_10=t 1114 ==> Module_id_7=t 1114 conf:(1)
2. Module_id_4=t Module_id_5=t 1156 ==> Module_id_7=t 1155 conf:(1)
3. Module_id_1=t Module_id_5=t 1670 ==> Module_id_10=t 1651 conf:(0.99)
4. Module_id_1=t Module_id_5=t Module_id_7=t 1434 ==> Module_id_10=t 1416 conf:(0.99)

...

Task V : Evaluation

- We've detected component coupling
- (easy) Interpretation possible
- makes sense according to our problem

Task VI: Plan deployment

- Usage in development process to receive information about the influence the change has on coupled modules (get knowledge about the relation)
- remove of coupling between modules could be difficult though
- could affect performance