Data Mining Mandatory Assignment Feedback for Jens Andersson Grøn

Pre-processing

Your choice of pre-processing techniques – Normalization, discretization and missing value replacement – is good and meets the requirement.

Apart from having not that many comments, your pre-processing code looks good. It is evident that you have spent quite some time on pre-processing the data. You also explain well in the report what you have done in regards to pre-processing and why. Good work!

Data Mining Algorithms

Your choice of data mining algorithms – ID3, Apriori and k-medoids – is good and meets the requirement.

I think it is great that you included your ID3 implementation and have described in the report, even though you did not get it to work. But it shows your effort. About the ConcurrentModificationException – it sounds like you are changing the contents of a list while iterating through it via Java's version of the C# foreach loop (i.e. java loops that start like this "for(String av : attributeValues)" – in other words you are changing the contents of attributeValues inside the loop.).

Your apriori implementation looks a bit odd. It looks like you have it hardcoded to only find frequent item sets with a length up to 3. The algorithm should be implemented in a manner that enables it to find frequent item sets until none can be found meeting the support threshold. However, it is great to see you have remembered to use the apriori property to prune candidate sets. Even though your implementation is bit too hard coded, you have still been able to produce some tangible results, which you present well in the report.

Your clustering code shows that you clearly understand the basics of k-medoids, but as you write there is a bug somewhere resulting in duplicates. I think it is good that you have included your k-medoids implementation in your submission, as it show the effort you have put into the assignment.

Code & Report

Your code (apart from the bugs and the hard-coded nature of your apriori) is okay – but some more comments in the code would have been great.

Final Remarks

Although you didn't succeed entirely in your data mining algorithm implementations, you have managed to show you understand the theory and this combined with some good pre-processing makes for an alright submission.

Grade

Pass.