

## COMP8400 – Tutorial 2 questions

### *Mining association rules and patterns*

---

Q1: Can you find out what the impact of this paper is in the scientific data mining community? (Hint: Impact of a paper is commonly measured by the number of times a paper has been cited by others – you can get citation counts for example from <http://scholar.google.com>, or <http://portal.acm.org/dl.cfm>).

Additional question: Can you find a paper that has a similar impact?

Q2: The second sub-problem of association rules mining (described in Section 1.1 in the paper) is to generate the rules for each of the large item-sets found. How many rules with non-empty left- and right-hand sides can be generated from an item-set that contains  $k$  items ( $k > 1$ )?

Q3: What is the main drawback of the earlier algorithms *AIS* and *SETM*?

Q4: The *AprioriHybrid* algorithm aims to combine the advantages of *Apriori* and *AprioriTid*. Compared to 1994 (when the paper was published), how do you think modern computer hardware is influencing the way *AprioriHybrid* would be run today?

Q5: In Section 4 the authors describe as one future work dimension the investigation of considering the *quantities* of the items in a transaction. Can you imagine a situation where these quantities are important, and a situation where they are not useful?

Q6: Do you have any criticism of this paper?

**Questions / issues you like to have discussed in this tutorial:**