Department of Computer Engineering
University of Peradeniya

Data Mining and Machine Learning
Lab 08

September 12, 2017

1 Association Rule

In data mining, Association Rule Learning is a popular and well researched method for discovering interesting relations between variables in large databases. Piatetsky-Shapiro describes analysing and presenting strong rules discovered in databases using different measures of interestingness. Based on the concept of strong rules, Agrawal introduced association rules for discovering regularities between products in large scale transaction data recorded by point -of-sale (POS) systems in supermarkets. For example, the rule **onion, potatoes** $=>$ **burger** found in the sales data of a supermarket would indicate that if a customer buys onions and potatoes together, he or she is likely to also buy burger. Such information can be used as the basis for decisions about marketing activities such as promotional pricing or product placements. In addition to the above example from market basket analysis association rules are employed today in many application areas including web usage mining, intrusion detection and bioinformatics.

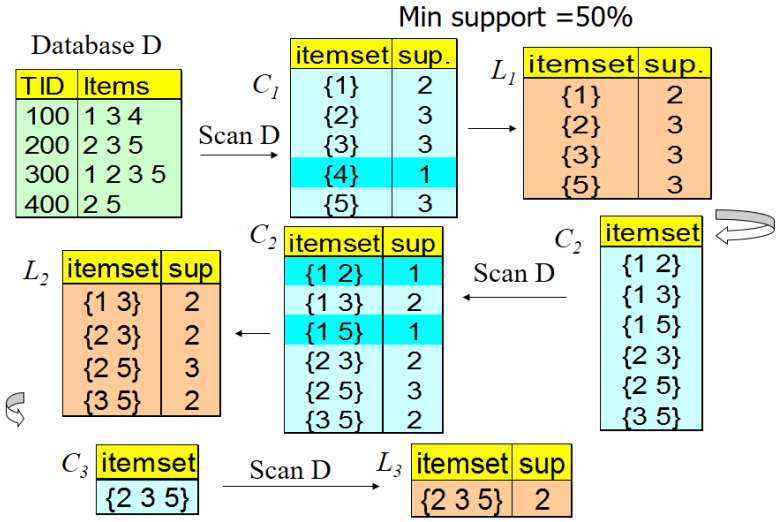

Figure 1: The Apriori Algorithm — Example

- Support :- Fraction of the transactions in which an itemset appears

2 Try Out

2.1 Association Rule Learning

In computer science and data mining, Apriori is a classic algorithm for learning association rules. Apriori is designed to operate on data sets which contain categorical (nominal) data. To illustrate concepts we use a small example from the supermarket domain. This dataset contains 15 instances with four attributes milk, bread, butter and beer which are typical products sold in a supermarket. Each instance represents a completed transaction and 0 indicates that the corresponding product is not sold whereas 1 indicates the corresponding product is sold under the completed transaction which is taken into account.

1. Load the 'mini-supermarket.arff' dataset into the WEKA tool and observe each transaction and its respective attribute values.

2. Go to the 'Associate' tab and choose 'Apriori' as the algorithm under the 'Associator' panel. Left click on the 'Apriori' label and obtain the 'Generic Object Editor' window. Click the 'More' button and identify what is indicated by each parameter in the 'Generic Object Editor'.

3. Briefly describe each of the 4 metric types provided in WEKA tool.

4. Apply the 'Apriori' learner with the default values. Describe the meaning of each rule with its corresponding support and confidence.

5. Apply the 'Apriori' learner with the two different numbers of rules. One less than the default number and one greater than that. Comment on the three different set of rules.

6. Suppose you are working in a supermarket which sells the above set of products and your company wants to increase the sales of the product milk. Using 'Apriori' built a set of rules that predict in which situations the product milk will be sold or not sold. From that rule select a best way to arrange the placement of the products to increase the sales of the milk.

3 Lab Exercise

1. Now let us examine a real-world dataset 'supermarket.arff', which is provided with the WEKA tool. This dataset has been collected from an actual New Zealand supermarket

2. Take a look at this file using a text editor to verify that you understand the structure. The main point of this section is to show you how difficult it is to find any interesting patterns in this type of data.

3. Experiment with 'Apriori' and learn a suitable set of association rules. Write a brief description on the main findings of your investigation.

4 Submission

Submit a single text file as [12|13]xxxlab08.txt where xxx is your registration number. Add answers for questions 2 and 3 as comments in the same file.

5 Important

Make sure that you have the basic understanding of association rules. This lab is really important for successive labs. If you do not understand any concepts, make sure you get some help from instructors.

6 Deadline

September 12, 23:59:59 GMT+5:30.