2017 Fall Data Mining

Homework 1

Mining association rules

Due Date: 23:59, October 27, Friday, 2017

TA: 蔡子豪 stu8978@gmail.com

In this homework, the major task is to find association rules, the most basic concept is “co-occurrence”, try to apply it with your imagination.

Like HW0, we have a platform to discuss this homework.(except answers)
Here you are. Hackmd

Dataset:

We‘ll use “201707-citibike-tripdata.csv.zip” (after preprocessed in HW0)

Schema:

The preprocessed data should contains the following

Every station’s information
id, name, lat, lng
Every stations’ flow data
id, time, in-flow, out-flow

Tasks:

- What is a transaction?
  Suppose I define a transaction is the in-flow and out-flow for station_id=519, then the transactions probably are (20,56),(38,57)...
- What rules should be discover?
  But we only have limited data, co-occurrence may not happened in the real value, so we need to do discretization,that is, convert the real value to some symbols, maybe (2,5),(3,5)...(divided by 10)
  and how many symbol should you use, depends on what’s your application.

What you need to do is try to define two more your transactions ,and conduct some experiments on the three mining tasks to find some rules behind them.(including my example)

- my example
- your1
- your2

For each task, you should try at least two discretization methods(divided by 10,divided by 20...) and two algorithms(Apriori,FP-growth...)to find association rules, then compare the difference between them.

Report:

In this homework, you have three tasks to do
for every tasks, you need to clearify

1. what is a transaction
2. what rules should be discovered(and discretization method)
3. what algorithm you use(Apriori or FP-growth or something else)
    a. algorithm code from github is allowed(cite the repository)
4. top 3 rules
5. what did you learned, or comparison between different methods you use

Your code should be submitted with the report.(jupyter notebook HTML, R markdown).

If you have any questions or suggestions, feel free to contact me:)