

CSCI 4144, Winter 2017
Assignment 3: Association Rule Mining Algorithm Implementation
(Issue: Feb 03, Due: Feb 22)

- **TA:** Virlla Devi Soothar (vr265712@dal.ca)
 - **Ass3 Tutorial:** Feb 09, 6:00-7:30PM, CS 127.
 - **Help Hours:** Feb 07 & 14, 1:00-2:30PM, CS 233; Feb 16, 6:00-7:30PM, CS 127
-

1. Objectives:

- 1) To gain an in-depth understanding on association rule mining.
- 2) To learn the main technical issues of implementing the Apriori algorithm for mining on relational data.
- 3) You may also gain a team work experience by working in a group of 2 students.

2. Programming language, computer system, etc:

The implementation is required to use a serious production implementation language, such as Java family/C family, etc. (but not a script language, such as R), which should have a compiler on bluenose.

3. Data sets and interface design requirements:

- 1) Your program should be able to handle data files with the following format: the first line contains column headings (i.e., attribute names); and every following row contains the values that represent a tuple. There are three data files available for developing and testing your program. The file data1 is a small data set which may be used for debugging your program. The files data2 and data3 are from real life databases for testing your program.
- 2) The interface should allow the user to choose (a) a data file, (b) a minimum support rate, and (c) a minimum confidence rate.
- 3) The mined rules should be placed into an external file named as "Rules". The format of the Rules file may like the output example attached at the end of the assignment statement.

4. Submit your assign3 electronically:

- 1) Create a directory assign3 in your bluenose account. This directory should include (a) the developed source code, (b) Makefile, and (c) README file.
- 2) The README file should provide: (a) the instructions how to compile and run the program, (b) a brief description of the overall design of the code (the functions and the call relationships, etc.), (c) the task partition for a team work.
- 3) Submit assign3 directory from your home directory by the command line: submit assign3.
- 4) Do not submit any data.

5. Evaluation:

- Your assignment will be evaluated based upon the overall quality of the work including user interface, functionality, modularity and readability of the program, and the clarity of the README file.
- If your program includes some adopted code, it must be clearly stated: 1) which parts of your program are adopted (mark them clearly in your program), 2) from where, i.e. the reference details of the open-source website.

Plagiarism and Intellectual Honesty: (<http://plagiarism.dal.ca>) Dalhousie University defines "plagiarism as the presentation of the work of another author in such a way as to give one's reader reason to think it to be one's own." Plagiarism is considered a serious academic offence which may lead to loss of credit, suspension or expulsion from the University, or even the revocation of a degree.

Appendix: An example of "Rules" file (Demo program: Ass/Ass3-demo)

Summary:

Total rows in the original set: 14

Total rules discovered: 12

The selected measures: Support=.25, Confidence=.60

Discovered Rules:

Rule#1: (Support=0.29, Confidence=0.67)
{PlayTennis=P, Windy=false} ---> {Humidity=normal}
Rule#2: (Support=0.29, Confidence=1.00)
{Humidity=normal, Windy=false} ---> {PlayTennis=P}
Rule#3: (Support=0.29, Confidence=0.67)
{Humidity=normal, PlayTennis=P} ---> {Windy=false}
Rule#4: (Support=0.29, Confidence=0.80)
{PlayTennis=N} ---> {Humidity=high}
Rule#5: (Support=0.29, Confidence=0.67)
{temperature=mild} ---> {Humidity=high}
Rule#6: (Support=0.43, Confidence=0.75)
{Windy=false} ---> {PlayTennis=P}
Rule#7: (Support=0.43, Confidence=0.67)
{PlayTennis=P} ---> {Windy=false}
Rule#8: (Support=0.29, Confidence=1.00)
{outlook=overcast} ---> {PlayTennis=P}
Rule#9: (Support=0.29, Confidence=0.67)
{temperature=mild} ---> {PlayTennis=P}
Rule#10: (Support=0.43, Confidence=0.67)
{PlayTennis=P} ---> {Humidity=normal}
Rule#11: (Support=0.43, Confidence=0.86)
{Humidity=normal} ---> {PlayTennis=P}
Rule#12: (Support=0.29, Confidence=1.00)
{temperature=cool} ---> {Humidity=normal}