CORE METHODS IN EDUCATIONAL DATA MINING
FALL 2014
BASIC ASSIGNMENT 8
SEQUENTIAL PATTERN MINING I

In this assignment, your goal is to find sequential patterns of interest in spmI-data-v1.csv

This data set has five variables:

anonid – which student it is

obsnum – how many observations have been conducted for this student

behavior-ontask – in this observation, the student was coded as on task (solitary)

behavior-ontaskconv -- in this observation, the student was coded as on task (conversation)

behavior-offtask-- in this observation, the student was coded as off task

affect-frustrated – in this observation, the student was coded as frustrated

affect-concentrating – in this observation, the student was coded as being in engaged concentration

affect-confused -- in this observation, the student was coded as being confused

affect-bored -- in this observation, the student was coded as being bored

This data set was previously published in
Baker, R.S.J.d., Moore, G., Wagner, A., Kalka, J., Karabinos, M., Ashe, C., Yaron, D. (2011) The Dynamics Between Student Affect and Behavior Occuring Outside of Educational Software. *Proceedings of the 4th bi-annual International Conference on Affective Computing and Intelligent Interaction.*

The goal of this assignment is to find sequential patterns in the data, which are unlikely to simply be due to chance. Please use RapidMiner 5.3 to complete this assignment, as other packages will be likely to produce slightly different results.

Question 1: Set up a RapidMiner process using Read CSV and the GSP operator (Generalized Sequential Patterns – *not* the WEKA version W-GeneralizedSequentialPatterns). What should your customer id be?

A) anonid
B) obsnum
C) behavior
D) affect

Question 2: What should your time attribute be?

A) anonid

B) obsnum
C) behavior
D) affect

Question 3: Set min support = 0.6, window size = 0.0, max gap = 5.0, min gap = 0.0, positive value = 1. Which of these association rules has the highest support?

A) behavior-ontask → behavior-ontask
B) behavior-offtask → behavior-offtask
C) behavior-ontask → affect-concentrating
D) affect-concentrating → affect-concentrating → affect-concentrating → affect-concentrating → behavior-ontask AND affect-concentrating

Question 4: If you set window size = 2.0, what is the association rule with the highest support that now is created (but was not created in question 3's settings)?

A) affect-concentrating → behavior-ontask → behavior-ontask AND affect-concentrating
B) affect-concentrating → behavior-ontask → affect-concentrating
C) affect-concentrating→behavior-ontask →affect-concentrating → affect-concentrating
D) behavior-ontask → behavior-ontask → behavior-ontask AND affect-concentrating

Question 5: Set window size back to 0.0. Set max gap to 1.0. Which is the rule with the most items?

A) behavior-ontask AND affect-concentrating → behavior-ontask AND affect-concentrating
B) affect-concentrating → behavior-ontask → affect-concentrating
C) behavior-ontask → behavior-ontask → behavior-ontask AND affect-concentrating
D) affect-concentrating → affect-concentrating

Question 6: Which of these is a reason why you might want to create a window size above 0?

A) Related events may be linked but separated by a few seconds
B) Unrelated events may be separated by a few seconds
C) Related events may occur at exactly the same time
D) Unrelated events may occur at exactly the same time

Question 7: How many students had the sequential rule behavior-ontask -> affect-concentrating at least once? (Hint: RapidMiner may not be the easiest tool to compute this with)

Question 8: What is the confidence for sequential rule behavior-ontask -> affect-concentrating? (Hint: RapidMiner may not be the easiest tool to compute this with)

Question 9: What is the cosine for sequential rule behavior-ontask -> affect-concentrating (Hint: RapidMiner may not be the easiest tool to compute this with)

Question 10: What is the lift for sequential rule behavior-ontask -> affect-concentrating? (Hint: RapidMiner may not be the easiest tool to compute this with)

Question 11: Would Merceron & Yacef say that this is an interesting association rule?

A) Yes, because cosine is over threshold
B) Yes, because lift is over threshold
C) Yes, because both lift and cosine are over threshold
D) No, because cosine is over threshold
E) No, because lift is over threshold
F) No, because cosine is below threshold
G) No, because lift is below threshold