

ASSIGNMENT 8
SPECIAL TOPICS IN EDUCATIONAL DATA MINING
PROFESSOR RYAN S.J.d. BAKER
SEQUENTIAL PATTERN MINING
DUE NOON, WEDNESDAY APRIL 24

The goal of this assignment is to find sequential patterns of interest in the data, using a sequential pattern mining algorithm of your choice, using the data in Asgn8-spm-data-v1.xlsx

This data set has five variables:

- Anonid – which student it is
- Moffsetfromstart – how many milliseconds since the beginning of observation
- Obsnum – how many observations have been conducted for this student
- Behavior – what behavior is coded by the coder
- Affect – what affect is coded by the coder

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 and report back on sequential patterns in the data, which are unlikely to simply be due to chance. Also, you will not receive credit for telling me about patterns that are already published in Baker et al. (2011).

You can conduct sequential pattern mining in any software package, or in Excel, using any sequential pattern mining algorithm you like.

Please turn in:

- All modified versions of the data file that you create
- All analysis outputs, whether from a software package or Excel
- All data mining code you used to generate the outputs (if relevant)
- A report of what patterns you found, what they might mean, and relevant evidence that they are unlikely to be due to chance. This report should also explain how you completed the assignment

You will be graded on completeness and comprehensibility of your hand-in, whether you correctly and validly apply the method you choose to this data, and whether the methods you chose fit the requirements of this assignment.

BONUS: The student who succeeds in producing the “most interesting” pattern which is correctly (or reasonably) interpreted from the data will receive the bonus. Please identify which sequential pattern you would like me to consider as most interesting for the bonus, and explain why it is interesting. (Hand-ins which do not do so will not be eligible for the bonus)