

T-61.5060 EXERCISE 4/2005

In T3 on 19 October 2005 at 12 o'clock.

Notice that there is no exercise session on 26 October 2005 and no lecture on 27 October 2005 due to the Examination Period.

1. Consider episodes α , β and γ (figure 1) in an event sequence $s = \langle (A, 1), (B, 2), (C, 4), (A, 9), (A, 13), (B, 20), (A, 22), (C, 23), (A, 28), (B, 32), (C, 35) \rangle$.
 - (a) Find frequencies of the episodes (e.g. $fr(\alpha, s, W)$) for some time windows W .
 - (b) What can be said of the frequencies of the episodes α , β and γ for a general event sequence?
 - (c) Find the sets of minimal occurrences, $mo(\alpha)$, $mo(\beta)$ and $mo(\gamma)$, and the corresponding supports.
 - (d) What is the confidence of a rule $\beta[3] \Rightarrow \gamma[5]$?
2. Consider a very long event sequence s containing events of types A and C distributed uniformly at random over the time interval. Let ν_A and ν_C be the expected number of events A and C occurring in a time window W , respectively. What is the expectation of the frequency $fr(\alpha, s, W)$, where the episode α is given by figure 1?
3. So far we have applied the levelwise algorithm to analyze sets of items and time sequences. Select some data structure other than sets or sequences to be analyzed with the levelwise algorithm. Design a suitable set of patterns and define the concepts and elements (support etc.) needed to apply the levelwise algorithm. How do you think your method would perform in some practical application?
4. Would the episode discovery framework be useful in looking for patterns in documents? (Consider a document as a sequence of events. I.e., each word is an event type, and an event is the occurrence of that word at a particular position in the document. Time of the event can be equated with the number of the word in the document. What would the (frequent) patterns mean in this case?)

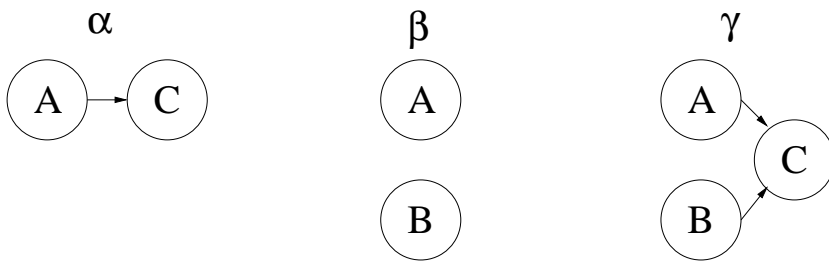


Figure 1: Some episodes.