

# A New Support Measure for Items in Streams \*

Toon Calders

Eindhoven University of Technology

Nele Dexters      Bart Goethals

University of Antwerp

## Abstract

Mining streams is a challenging problem, because the data can only be looked at once, and only small summaries of the data can be stored. We present a new frequency measure for items in streams that does not rely on a fixed window length or a time-decaying factor. Based on the properties of the measure, an algorithm to compute it is shown. Experimental evaluation supports the claim that the new measure can be computed from a summary with very small memory requirements, that can be maintained and updated efficiently. In this extended abstract, the main points of the presentation are discussed.

## 1 Motivation

Mining frequent items over streams received recently a lot of attention. It presents interesting new challenges over traditional mining in static databases. It is assumed that the stream can only be scanned once, and hence if an item is passed, it can not be revisited, unless it is stored in main memory. Storing large parts of the stream, however, is not possible because the amount of data passing by is typically huge.

Different models have already been proposed in literature. The main characteristic is: how must the frequency of an item be measured? There are different types of models. (1) the sliding window model, (2) the time-fading model, or (3) the landmark model. In the sliding window model [1, 3, 6, 8, 10], only the most recent events are used to determine the frequency of an item. In order to avoid having to count the supports on this window all over again in every time point, the algorithm in fact updates the frequency of the items based on the deletion of some transactions and the insertion of other. In the time-fading model, the past is still considered important, but not as important as the present. This is modelled by gradually fading away the past [9]. That is, there is, e.g., a

\*The presentation is based on material presented in the ECML/PKDD'06 workshop International Workshop on Knowledge Discovery from Data Streams (IWKDDs) [2]