

Raymond Martin
Data Mining 8331 - Professor Hahsler
February 25, 2015

Frequent Pattern Mining in Data Streams Abstract

Frequent pattern mining is an essential task in data mining and is also a critical building point for many other data mining techniques. Currently most techniques for frequent pattern mining have exponential time and space complexities while also requiring many passes on the dataset. Due to these downfalls, frequent pattern mining over a potentially infinite data stream is nearly impossible. Some state-of-the-art algorithms have been developed to combat these challenges. These algorithms include Lossy Counting, FPDM, MOMENT, extDec, and more. All of these algorithms suffer from their own downfalls. A new algorithm called Stream-Mining has been introduced into the data mining space which has many advantages compared to the ones mentioned above. KPS, an extremely memory and time efficient algorithm, is leveraged in the Stream-Mining algorithm to make it a one pass algorithm with a false positive orientation, making it extremely attractive for large sets of data. Some of Stream-Mining's extraordinary capabilities include that it does not require an out-of-core memory data structure, has deterministic bounds on the accuracy, and does not produce false negatives [1].

Sources:

1. Jin, Rouming, and Gagan Argawal. "Frequent Pattern Mining In Data Streams." *Data Streams*. Boston/Dordrecht/London: Kluwer Academic, 2007. 61-81. Print.