Using a real-time top-k algorithm to mine the most frequent items over multiple streams.


Summary: Some applications such as sensor networks, internet traffic analysis, location-based services, and health measurements are always required for considering unbounded, fast, large-volumes, continuous, even for distributed stream data. It’s a better way to use synopsis as a list of partial summaries of unknown item sets in order to reduce the memory space usage, let it can afford to process so fast and huge incoming data. Normally, different quantity of item set leads to different summaries, especially for Top-k operator which as a partial preprocess over synopsis. Therefore, we proposed smooth synopsis that dynamically assigns a numeral interval to resolve the items set, in order to maintain a more accurate approximate answers’ list from partial Top-k processing. In particular, we proposed an algorithm (called SFI algorithm) to mine the most frequent items by a more adaptive and fast way in specific stream resources. Finally, our experimental results demonstrate the accuracy and efficiency of our approximation techniques.

Keywords: sliding window; synopsis; streaming data; top-k query; SFI algorithm
doi:10.1007/978-3-642-39479-9_36