1. **Summary.** The paper proposes a novel streaming algorithm to mine the top-$k$ episodes in a stream of events. The frequency of the episodes is computed over a sliding window which length is defined by the user. The key idea in this paper is based on two new concepts related to the stream: *maximum rate of change* and *top-$k$ separation*. The sliding window is decomposed into batches and the previous two concepts are used in tracking the top-$k$ elements in every batch in the stream. The author show how an exact and approximate solution to the top-$k$ problem can be found in the stream by mining only a fraction of the frequent episodes in each batch. This approach is the first solution for mining events in a stream and represents a generalization of the existing streaming pattern discovery algorithms.

2. **Positive Points.**

   - I like the notion of episode used in the problem statement. This makes the approach more general than the classical frequent itemsets miner algorithms.
   - The authors propose a formal study of the properties of the event stream: rate of change and top-$k$ separation, and I like how they employ these properties to develop their solution and characterize the quality of the mined results.
   - In the experiment section the authors show how effective their approach is in different scenarios (real and synthetic data) and with respect to similar streaming approaches. From their results we can see that this approach provide a considerable improvement in running time and memory usage (up to 4 time better) while providing better utility results with respect to the tested algorithms.
3. Negative Points.

- It turns out that the error in the estimation of the top-\(k\) events is linear with the length of the sliding window considered. This is also evident from the experiment where the \(F_1\) score is decreasing when the length of the window is increasing. This could considerably degenerate the utility results when the sliding window is large.

- The maximum rate of change plays a central role in the algorithm. As the authors state this parameter is highly data dependent and therefore it has to be estimated using the 75th percentile. It could be interesting if the authors could provide some indication how the performance of their algorithm change with this rate and how far are from those that could be obtained if the real rate of change is used.

- Although the approximation algorithm is very interesting it seems that the approximation is not feasible when the stream does not present enough top-\(k\) separation or the rate of change is too high. Therefore this makes it not effective in stream where the highly dynamic events.

4. Discussion.

- I think that this approach is very interesting and it leaves open several future research questions, for example how to privately mine the events from episode streams or how to improve the efficacy of the miner for specific events.

- It could interesting and maybe beneficial for the algorithm to use variable length batches, so that we can better control the rate of change of the frequency and provide a better approximation.

- The intuition in this approach bases on the assumptions that the persistent episodes in the window they do not get too mixed with non frequent episodes in successive batches. But what if this does not hold and these persistent episodes are not enough?