Exercise 8.5

zaterdag 7 oktober 2023 19:33

This algorithm can work by running two instances of the regular frequent-items problem in parallel.

More specifically, we run two instances, each of which reset after scanning over 2W items, where one instance resets starting at 0 items and the other starting at W items. Furthermore, each of these instances determines the set of $\frac{\epsilon}{2}$-frequent items. Then, whenever an answer needs to be provided at a given point, the current set of $\frac{\epsilon}{2}$-frequent items maintained by the longest-running instance will be returned.

The reason this works is that, at any point in time, any item which is ϵ-frequent over the last W items will also be $\frac{\epsilon}{2}$-frequent among the last 2W items. Since there is an instance which has run over at least the past W items *(but not more than 2W items)*, there will always be one instance which has a set of at most $\frac{2}{\epsilon}$ items which are $\frac{\epsilon}{2}$-frequent among the last (at most) 2W items; all ϵ-frequent items over the last W items must be included in this set.