Technical Perspective: A Framework for Adversarially Robust Streaming Algorithms

Graham Cormode
University of Warwick, UK
G.Cormode@warwick.ac.uk

Over the past two decades the data management community has devoted particular attention to handling data that arrives as a stream of updates. This captures a number of "big data" scenarios, ranging from monitoring networks to processing high volumes of transactions in commerce and finance. This has led to data streams becoming a mainstream data management topic, with many systems offering explicit support for handling such inputs. Within these systems, streaming algorithms are used to approximate various statistical and modeling queries, which would traditionally require random access to the full data to compute exactly.

The area of streaming algorithms is by now very mature, with effective and efficient techniques known for many core analytics problems – tracking frequencies of fluctuating quantities, monitoring statistical distributions – supporting applications like anomaly and change detection over high speed, multidimensional streams. These algorithms have been deployed deep within the infrastructure of major data processors – governments, search engines and cloud service providers – to flag anomalies and changes in distribution. These applications consequently demand a high level of reliability to ensure that the results can be trusted.

High reliability comes assured when an algorithm provides a deterministic guarantee. However, many of the widely used streaming algorithms are randomized in nature: for any input, there is some small probability that they will give an incorrect answer. This prompts the question of how reliable can these algorithms be when they are queried multiple times in succession? In particular, what happens if the input to the algorithm might be set by actions taken in response to the algorithm's earlier output – then the prior fixed input guarantees may not hold.

The following paper, "A Framework for Adversarially Robust Streaming Algorithms", makes a substantial advance on this question by addressing the notion of "adversarially robust" streaming algorithms. The idea is to show a strong guarantee on a streaming algorithm by assuming that the input is generated by an adaptive adversary, who observes the outputs of the algorithm, and sends new stream elements with the intent of provoking an erroneous response. If we can prove an algorithm is robust to such a malicious adversary, then we can be assured that it will be highly reliable on any input distribution, no matter how it is formed. A significant feature is that rather than consider the robustness of individual algorithms, the paper provides a framework so that *any* streaming algorithm for any problem can be made robust, by executing it with modified parameters or by running multiple independent instances. This is a powerful first step in measuring the "price of robustness", compared to solutions for the same problems which lack a robustness guarantee.

The topic of robust streaming has led to a flurry of activity in a short time. Recent interest in robust streaming stems in part from work of a subset of the paper's authors, who addressed the question of sampling, and showed that, after appropriate setting of parameters, several sampling algorithms are inherently robust [2]. Due to a temporal quirk, the results on sampling and streaming both appeared in PODS 2020, where the streaming work won the best paper award.

Subsequent work has sought to better understand the generality of robustness results, and to reduce the space needed to give these guarantees. New papers have sketched out an intriguing connection with differential privacy, to show that adding suitable noise to the output of an algorithm can ensure robustness [3]; they have provided tighter bounds for specific questions in the infinite window and sliding window model of streams [5]; and, most recently, proved a separation between the adversarial and non-adversarial model by showing that the "adaptive data analysis" problem which requires exponentially more space in the adversarial setting [4]; and exhibited a deep connection between adversarial sampling and the theory of online learning [1]. These rapid advances demonstrate the interest in this work, and suggest that there is still yet more to discover about this challenging topic.

1. REFERENCES

[1] N. Alon, O. Ben-Eliezer, Y. Dagan, S. Moran, M. Naor, and E. Yogev. Adversarial laws of large numbers and optimal regret in online classification. *CoRR*, abs/2101.09054, 2021. To appear in STOC 2021.

[2] O. Ben-Eliezer and E. Yogev. The adversarial robustness of sampling. In *ACM PODS*, 2020.

[3] A. Hassidim, H. Kaplan, Y. Mansour, Y. Matias, and U. Stemmer. Adversarially robust streaming algorithms via differential privacy. In *NeurIPS*, 2020.

[4] H. Kaplan, Y. Mansour, K. Nissim, and U. Stemmer. Separating adaptive streaming from oblivious streaming. *CoRR*, abs/2101.10836, 2021.

[5] D. P. Woodruff and S. Zhou. Tight bounds for adversarially robust streams and sliding windows via difference estimators. *CoRR*, abs/2011.07471, 2020.