Synopses for Summarizing Spatial Data Streams

Jacco Kiezebrink, Wieger R. Punter [1], Odysseas Papapetrou [1] and Kevin Verbeek [1]

[1] M&CS, Eindhoven University of Technology, Eindhoven, The Netherlands
{w.r.punter,o.papapetrou,k.a.b.verbeek}@tue.nl

In today's data-driven landscape, geospatial streams are pivotal in diverse fields, ranging from sociology to network engineering and to meteorology. A key challenge in utilizing these streams is to efficiently compute aggregates over ad-hoc spatial ranges, possibly with additional predicates on the stream items. For each application scenario, different aggregates become relevant, such as the number of distinct items, the frequency of each item, or even the variance of the frequencies of the items that fall within a spatial range.

Storing the entire stream for computing these aggregates is impractical in scenarios that involve fast-paced and unbounded streams, due to prohibitive storage costs and query execution delays. To address this, we propose two sketches, SpatialSketch and DynSketch, that support aggregate queries with different types of aggregates (e.g., frequency estimation, L2 norm, membership queries) by incorporating the functionality of other nested sketches like Count-min sketches [2] and Bloom filters [1]. Both sketches require small space, and they can summarize fast-paced streams and estimate the aggregates, with accuracy guarantees. Importantly, they support *new diverse functionalities, in a plug-and-play manner, without requiring novel theoretical analysis*. In addition to the theoretical contribution, we evaluate SpatialSketch and DynSketch experimentally. Our experiments demonstrate that the two sketches outperform the state of the art, and that they can be used for addressing novel functionalities for which there exist no small-space solutions to date.

This work will appear at EDBT 2025 [3].

References

[1] Broder, A.Z. and Mitzenmacher, M. (2003). Survey: Network Ap- plications of Bloom Filters: A Survey. *Internet Mathematics* 1 (4), 485–509.

[2] Cormode, G. and S. Muthukrishnan. (2005). An improved data stream summary: the count-min sketch and its applications. *Journal of Algorithms* 55 (1), 58–75.

[3] Kiezebrink, J. and Punter, W.R. and Papapetrou, O. and Verbeek, K. (2025). Synopses for Summarizing Spatial Data Streams. *Proceedings of the 28th International Conference on Extending Database Technology (EDBT 2025)*, Barcelona, Spain.