

Research Statement

Sorelle Friedler

I am primarily interested in research on computational geometry. My current focus is on developing geometric analyses of kinetic data collected by sensor networks. My goal is to provide the theoretical background so that scientists working with moving data can analyze and understand the underlying statistical trends. I will continue to examine computational geometry problems resulting from application areas and other pure algorithmic questions.

A Sensor-Based Framework for Kinetic Data

Large sensor networks collect and analyze kinetic data generated by moving objects. In order to automate the statistical and geometric analyses of such data, appropriate theoretical and practical frameworks must be developed. This work provides insight into motion trends and automates their identification, leading to applications from many areas in many disciplines, such as traffic detection, species migration, and general analysis of data as it changes over time.

I have developed a sensor-based framework that provides a structure from within which it is possible to create algorithms to analyze kinetic data and formally guarantee the efficiency of those algorithms [1]. My work is significant because it presents a framework that is theoretically sound while also being practical. Previous applied work in this area is highly specialized based on the specific practical context and provides no performance guarantees, only experimental evaluations. Experimentally evaluated work has been done in related areas within the database, networking, computer vision, and robotics communities. My work considers the theoretical background to these questions.

Compression and Retrieval of Kinetic Sensor Data

Due to the vast amounts of data collected by sensor networks, compression of the data is necessary. In order to analyze the data without losing the space savings due to compression, retrieval must be performed over compressed data. I developed a new algorithm that compresses the output of a large sensor system without loss of information. I proved that my algorithm achieves a compression rate that is within a constant factor of the information-theoretic lower bound, as given by the joint entropy of the sensor system. This is the first algorithm that achieves asymptotically optimal lossless compression of collected sensor data. The algorithm was analyzed with respect to both statistical entropy, which assumes an underlying random process, and empirical entropy, which relies only on the observed data. In order to justify the relative storage efficiency of the new framework versus the established framework (kinetic data structures), I proved that under certain realistic assumptions the frameworks perform similarly [1, 2].

In order to perform retrieval over the compressed data, I considered the problem of retrieving the aggregated values of sensor readings within a spherical range over a given time period [2]. My approach involves retrieving counts for individual sensors over the given time period and using a hierarchical method to combine the underlying time instants and spherical regions into an approximate larger spherical range. My retrieval algorithm operates only over the compressed data and I proved that it takes space on the order of the compressed size and query time logarithmic to it, where the compressed size is the optimal space bound assuming that the entirety of the compressed data must be stored.

Robust Statistical Estimators for Kinetic Data

Robust statistical estimators, such as robust clustering, allow data to be analyzed while allowing for outliers. When considered on kinetic data, robust estimators allow for outliers as data changes over time. I developed an efficient algorithm that solves the kinetic robust k -center problem. The kinetic robust k -center problem maintains a clustering as points move over time, where the clustering partitions the points into k groups so as to minimize the maximum distance from any point to the center of its containing group, and the number of points to be clustered is specified by a user-given robustness threshold. My solution uses the kinetic data structures model to maintain a clustering of points as they move through time [3]. This clustering is robust to outliers and is analogous to traffic detection. Solving this problem exactly is known to be computationally intractable. I proved

a bound on the quality of the solution relative to an optimal solution. This bound improves upon the approximation factor previously achieved by Gao, Guibas, and Nguyen¹ for a more restricted version of the same problem. I also proved that my algorithm for calculating this approximate solution is efficient under the kinetic data structures model.

My current research addresses the same robust clustering problem on my sensor-based framework. My solution will serve as a demonstration that the sensor-based framework can be used effectively to examine statistical trends. Since finding an exact solution is computationally intractable, my goal is to mathematically establish an upper bound on the quality of the solution relative to an optimal solution. In conjunction with this, I will show that my algorithm for calculating this approximate solution is efficient.

Future Work

There are many aspects of this work that I plan to continue to explore, including:

- Extensions of my sensor-based framework and further comparisons to other models.
- Experimental evaluations of the compression and retrieval algorithms, possibly on data collected by researchers within the broader scientific community.
- Further development of algorithms to calculate robust estimators within the sensor-based framework.
- Development of online algorithms to analyze collected sensor data under a streaming model.

Collaborators and Other Research Interests

I enjoy working with collaborators, and am interested in continuing my research through collaborations within the department and also with scientists outside of computer science. While I have described my research as pertaining to sensor networks, it could in fact apply to any data streams generated at locations defined with an underlying distance metric, such as occurrences of HIV strains in a population over time. I believe that such collaborations would be interesting for my own research, and also provide a useful perspective for computer science in general.

I believe that it is crucial for undergraduate students to have the possibility of being part of a research project. I'm interested in integrating students into my research plans. I think that one good way to help students enter research of this kind is to begin with an experimental evaluation of an algorithm. This helps the students to gain a deeper understanding of the issues involved, which can lead to future ideas for algorithm design improvement. While my main research area is computational geometry, I have interests and experience in programming languages [4], integer programming [5], and software for teachers [6]. I am open to and interested in continuing to work in these general areas, and see these other areas as another source for potential research collaboration with students as well as faculty.

References

- [1] S. A. Friedler and D. M. Mount. Compressing kinetic data from sensor networks. In *Proc. of the Fifth Intl. Workshop on Algorithmic Aspects of Wireless Sensor Networks (AlgoSensors)*, pages 191 – 202, 2009.
- [2] S. A. Friedler and D. M. Mount. Spatio-temporal range searching over compressed kinetic sensor data. Submitted for publication, *Latin American Theoretical Informatics Symposium*, 2010.
- [3] S. A. Friedler and D. M. Mount. Approximation algorithm for the kinetic robust k -center problem. Accepted subject to minor revisions, *Computational Geometry: Theory and Applications*, 2009.
- [4] S. A. Friedler, C. Castillo, R. Alford, and J. S. Foster. Dynamic tracking of type qualifiers in Java. In preparation, 2009.
- [5] A. Archer, D. Applegate, and S. A. Friedler. An implementation of Jain's algorithm for survivable network design. In preparation, 2009.
- [6] S. A. Friedler, Y. L. Tan, N. J. Peer, and B. Shneiderman. Enabling teachers to explore grade patterns to identify individual needs and promote fairer student assessment. *Computers & Education*, 51(4):1467–1485, December 2008.

¹Jie Gao, Leonidas J. Guibas, and An Nguyen. Deformable spanners and applications. In *Computational Geometry: Theory and Applications*, 2006.