

Mission: We create novel techniques for knowledge discovery from large scale data streams.

Current Focus Areas:

- Massive-scale sequence modeling of data streams with applications in hurricane intensity prediction and metagenomic sequence classification and characterization.
- Fast probabilistic sequence matching using locality sensitive hashing for abundance estimation and security applications.
- Simulation data analytics for earth quake induced liquefaction.
- Modeling the progression of type-2 diabetes using virtual screening based on electronic health care records.

Sustainable Research: All code is published as open source R packages (CRAN/BioConductor). We currently maintain 15 very popular packages with 35,000+ downloads per month.

Team: 6 SMU faculty, 10 students, 7 collaborators

Director: M. Hahsler <mhahsler@lyle.smu.edu>

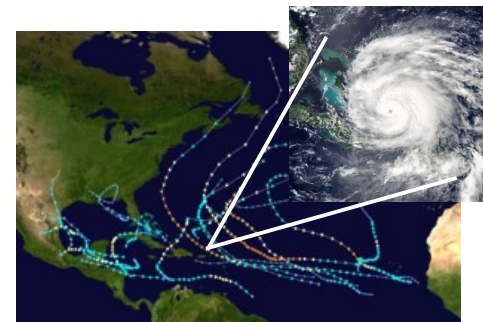
Supported by



Simulation Data Analytics



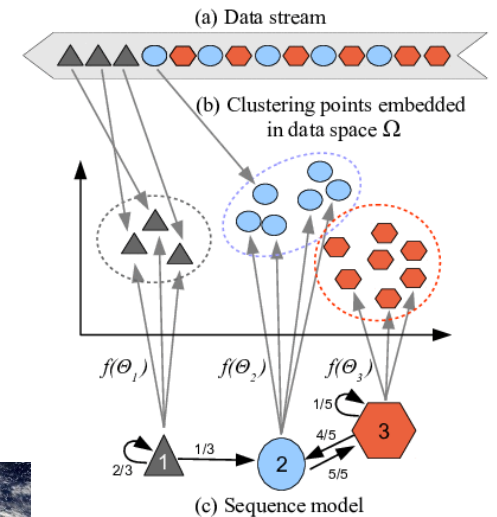
Meteorology



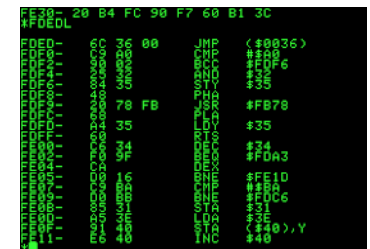
Metagenomics



Massive-scale Sequence Modeling & Data Stream Mining



Cybersecurity



Health Care Analytics

