



Thesis Defense

Computer Science Master's Program

"Implementation of Area-Based Aggregation on Multivariate Continuous Uncertain Data"

By Rahul Nair

Abstract:

Uncertain data is incredibly widespread - from sensor data to AI-based learned information, there exists a need to associate information with a certain probability of its veracity. Traditional relational databases lack a built-in functionality to support uncertain data, instead assigning them boilerplate values. Probabilistic databases tackle this problem by assigning non-deterministic data with an associated, often discrete, probability. Variants of probabilistic databases, namely continuous uncertain databases, are used to better model data represented through ranges and distributions. This is especially applicable with sensor-based geographic data as most commercial equipment contains some inherent margin of error. While uncertain and probabilistic databases contain incredible potential, aggregation on these forms of databases remains inefficient, namely due to the need to consider all present probabilities to compute an exact aggregate. This problem is compounded should we wish to aggregate across some uncertain data. Accordingly, approximation algorithms are often used to compute aggregates within a reasonable bound of accuracy at a much faster rate. Despite extensive prior research on both exact and approximate aggregates, work on computing aggregates of uncertain data across differing uncertain data is lacking; such a situation is especially pertinent with uncertain GIS data due to its numerous data forms and dimensionalities. Specifically, this thesis explores aggregating across regions of uncertain data modeled by two-dimensional distributions, which can be deemed as an aggregation by area. This thesis presents a multi-step approach to aggregating across multidimensional uncertain data. First, we introduce a framework to model varying forms of uncertain data obtained from a singular source. We then propose multiple algorithms to efficiently conduct aggregation of uncertain data across other uncertain data of differing types, applying them in a GIS-centric case study. More accurate insights, which specifically take uncertainty into account, can be obtained as a result.

Date: Tuesday, June 10th, 2025

Time: 2:00 PM – 4:00 PM

Location: 14-238B

Committee: Dr. Stanchev, Dr. Hristov, Dr. Dekhtyar