学　位　論　文　内　容　の　要　旨

博士の専攻分野の名称　　博士（情報科学）　　氏名　ニレンダ　マユンボ

学　位　論　文　題　名

Studies on Approximate Bayesian Computation and Speedup of Spatial Data Access Using Distributed Quadtrees

（近似ベイズ計算と分散 4 分木を用いた空間データアクセスの高速化）

Rapid advances in information technology are driving epidemiological research towards a focus on the predication of diseases dynamics. Among these advances is the extremely large amounts of observational data that can be analyzed computationally to reveal interesting patterns of information. Another is the increase in computational power which allows for simulations using agent based models. Coupled with these advances is improvements in mathematical models for epidemics and the need for efficient resource planning to prevent major disease outbreaks. To this effect data assimilation is increasingly becoming important in offering a way to forecast and estimate epidemic dynamics by using initial conditions and model parameters to constrain a mathematical model used for simulation to yield results that approximate the real world fairly well.

Many Mathematical models describing the epidemiological dynamics of diseases have been proposed. However the dynamics for multistrain diseases such as influenza B are complex. Computation of likelihoods for such models is often intractable and thus rendering most studied and used approaches inapplicable. This problem is further amplified by the computational cost of simulating such dynamics. Furthermore data assimilation involving large amounts of spatial data faces a data access bottleneck. In this thesis we therefore propose a framework that uses likelihood-free approaches for epidemic estimation. Further, We propose an architecture based on distributed quadtrees for accelerating access to large amounts of spatial data to alleviate the data access bottleneck.

In Chapter 3 we propose and discuss a framework for epidemic estimation. We also propose a stochastic individual based model for complex multi-strain epidemic dynamics. We then use our framework and model to estimate the dynamics of influenze B. Using the estimated parameters, we predicted the dominant lineage in 2015-2016 season in Japan. The accuracy of this prediction is 68.8

In Chapter 4 we consider the use of distributed quadtrees in a shared-nothing memory approach to reduce the data access bottleneck in data assimilation systems. We distribute data across nodes and construct a directory for the distributed nodes by using a quadtree built from sampled points. We discuss approaches for partitioning and allocating data and queries across the distributed nodes. Results from the experiments we conducted using a scale-down parallel data load and search distributed processor system show that a collection of small indices of distributed shared-nothing memory is more efficient than the conventional approach with a single processor with a large external index.

In Chapter 5 we take into account the process of query redirection during the construction of the distributed quadtree as well as query redirection during a data retrieval process. We propose taking advantage of the static nature of the sample points of the data and use of hashmaps and dilated inte-

gers to speed up traversal of the directory. Results from the experiments conducted show a threefold improvement in performance and also show less sensitive to data skewness.

Finally, in Chapter 6 we conclude this thesis and discuss future researches. The main point and focus of this thesis is the application of already existing techniques to improve data assimilation processes with application to real-world problems. We use Approximate Bayesian Computation as a likelihood free data assimilation tool to estimate the dynamics of influenza B. Results from the experiments conducted show that our proposed approach is capable of learning the essential parameters of influenza B required to predict the dominant lineage of the following year. Furthermore our proposed architecture for acceleration of spatial data access in data assimilation systems results in significant gains in performance.