Abstract

Dynamic Bayesian Networks (DyBNs) have found application in several areas due to their ability to model temporal and spatial relationships. In this work, we perform several experiments to investigate different DyBN model formulation's ability to contribute to improvements in disease incidence prediction. Estimates of future disease incidence allows time for health workers to devise interventions, averting or reducing on the undesirable consequences a disease outbreak may have on a population. We focus on two diseases cholera and malaria. Cholera is a water-borne disease, while malaria is a vector-borne disease, and therefore these two diseases were separately modelled. DyBN models are a general class of hidden Markov, and Kalman filter models.

In the first part, we explored how to improve prediction of cholera outbreaks in Uganda. We used a hierarchical clustering algorithm to identify districts with similar historical dynamics so that they could be modelled together. We made one step ahead predictions of disease incidences from a DyBNs model with weekly reported cholera disease counts and climate data as inputs. Using outbreak thresholds, we were able to classify onsets of outbreaks plausibly. In the second part of this work, we explored disease incidence predictions with different DyBN model formulations based on two assumptions: homogeneity or heterogeneity of the hidden state; and the probability distribution of the output variables. Results showed that it is better to assume that the hidden state is heterogeneous, and output variables follow a Poisson instead of a Gaussian distribution. In the last part, we proposed a framework of combining the two tasks of automatic disease diagnosis and disease density estimation. The frame allows for both tasks to be carried out seamlessly in realtime and proved improvement in results when performed together as they benefit from each other.

This research makes three significant contributions to the application of computer science in bio-surveillance and epidemiology. These include a new computational model framework that allows coupling of, rather normally separate tasks in practice, disease density estimation and automatic disease diagnosis; insights into different modelling assumptions in both time and space; and inference with data at both individual and population level. This work also extensively discusses different sources and the preprocessing necessary before modelling. This discussion is particularly important for researchers in biosurveillance, especially in data constrained environments like sub-Saharan Africa. All this, we believe, will significantly contribute to future disease incidence modelling for accurate and timely early disease detection and characterisation to support disease response and control.