Advanced Topics in Medical Modeling

Biosurveillance and Data Fusion for Early Outbreak Detection and Classification

Siân Jenkins, Riskaware Ltd
Martyn Bull, Riskaware Ltd
Suzie Abbs, Dstl
Simon Maskell, University of Liverpool
Paul Horridge, University of Liverpool

Riskaware, together with the University of Liverpool and Dstl, who also funded and managed the project, have developed an information fusion algorithm for bio-surveillance (BSV), which fuses data from multiple healthcare and non-healthcare sources, such as hospital data, primary healthcare provider data and social media data. The algorithm aims to alert healthcare providers to the presence of a disease outbreak and identify it before they become aware of it through traditional means, such as diagnostic testing. This will allow healthcare providers to respond faster and prevent the further spread of infection. The capability can consider infections that are endemic in the population, such as influenza, as well as those that could be the result of a biological attack, such as anthrax or pneumonic plague. The algorithm combines a particle filter, an epidemiological model and a Bayesian network. The particle filter performs sequential Bayesian inference to provide a probabilistic estimate for the populations in the compartments of the epidemiological model, which is an extended SEIR (Susceptible, Exposed, Infected, Recovered) model, to infer whether there is an infection present in the population. Both Marginal and Fixed-Lag Sequential Monte Carlo (FLSMC) particle filters have been trialled within the system, with the FLSMC algorithm providing improved results by overcoming the problematic effects of “model-lag”, where the signal is detected days after the initial outbreak. If an outbreak is present in the population, the algorithm estimates the probability of each considered infection being the source of the observed signal in the data. The Bayesian network is used to account for seasonal effects and long-term trends in the observations. Probabilistic forecasts can then be generated from the results of the algorithm, to determine the likely future progress of the outbreak. The algorithm has been developed within an extensible and flexible framework that allows daily execution on new data, as well as a batch mode. Statistical process control (SPC) algorithms have also been included within the framework for comparison. Applications for the algorithm exist in both military and civilian scenarios and include detecting the effects of biological weapons, as well as the emergence of new strains of a communicable disease. The incorporation of a feature that allows for changing populations allows the algorithm to function correctly in the presence of large population changes due to troop deployment and recall. The algorithm has already been tested on synthetic data and shown to be able to identify and track infected populations through the different stages of their infections. Testing is currently being conducted using real-world data obtained from Public Health Wales. Evaluation metrics used include probability of detection, probability of false alarm and conditional expected delay. Receiver operating characteristic (ROC) curves have been plotted to explore the effects of different parameters on algorithm performance. Riskaware now has a licence agreement with Dstl to develop and distribute BSV.