

A Method for Detecting and Characterizing Multiple Outbreaks of Infectious Diseases

John M. Aronis*, Nicholas E. Millett, Michael M. Wagner, Fuchiang Tsui, Ye Ye and Gregory F. Cooper

University of Pittsburgh, Pittsburgh, PA, USA

Introduction

We describe an automated system that can detect multiple outbreaks of infectious diseases from emergency department reports. A case detection system¹ obtains data from electronic medical records, extracts features using natural language processing, then infers a probability distribution over the diseases each patient may have. Then, a multiple outbreak detection system (MODS) searches for models of multiple outbreaks to explain the data. MODS detects outbreaks of influenza and non-influenza influenza-like illnesses (NI-ILI).

Methods

MODS searches over models of multiple outbreaks to maximize the probability of the data². We start with Bayes' Theorem which states $P(M|data) = (P(data|M)P(M))/P(data)$ where $P(M)$ is the prior probability of an epidemiological model of zero, one, or more outbreaks, $P(data|M)$ is the probability of the data given a model, and $P(data)$ is the probability of the data. The method searches for the MAP (maximum a posteriori) model that maximizes the above numerator.

A model consists of baseline levels of non-outbreak influenza and NI-ILI, zero, one, or more influenza outbreaks, and zero, one, or more NI-ILI outbreaks. Searching for the MAP model requires MODS to search over the set of basic parameters for multiple influenza and NI-ILI outbreaks. We search over combinations of multiple influenza and NI-ILI outbreaks; our current implementation (running on a single 1.6GHz processor) takes about 24 hours to adequately search the space of models for a single dataset for one year.

Results

We conducted a battery of experiments with simulated outbreaks consisting of zero, one or two influenza outbreaks and zero, one, or two NI-ILI outbreaks. We start with records of actual influenza, NI-ILI, and other patients, construct a model of multiple outbreaks, then instantiate the model by:

- Randomly sampling patients without influenza or NI-ILI according to a Poisson distribution.
- Randomly sampling influenza and NI-ILI patients according to the postulated outbreak model.

We then search over models consisting of zero, one, or two influenza outbreaks and zero, one, or two NI-ILI outbreaks and report the MAP model. In general, MODS can accurately predict starting dates and peaks for multiple influenza outbreaks, as well as the start, duration, and level of co-occurring NI-ILI outbreaks.

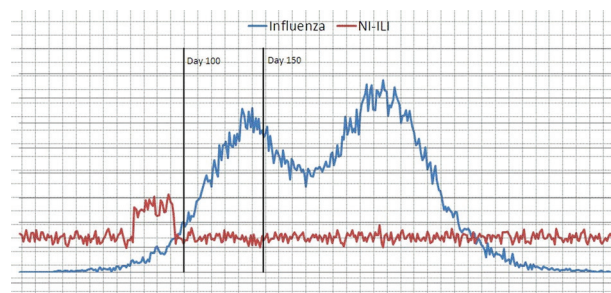
For instance, we constructed a model with two influenza outbreaks with start/peak days 20/138 and 80/222 and a NI-ILI outbreak with start/duration 70/25 (shown in diagram "Simulated Outbreaks").

On day 100, before either influenza peak, MODS predicted one influenza outbreak with start/peak days of 20/138 and one NI-ILI outbreak with start/duration days of 71/19. On day 150, after the first influenza peak but before the second, MODS predicted two influenza outbreaks with start/peak days of 20/138 and 76/215 and one NI-ILI outbreak with start/duration days of 71/19.

We also ran our system on data from Allegheny County for the 2009-2010 influenza season. Given data starting on June 1, 2009, by September 1 the system predicted a peak at about October 15 with a small NI-ILI outbreak in June. Previous analysis using thermometer sales as a measure indicate the actual peak was around October 19³.

Conclusions

This work demonstrates that we can predict and characterize multiple, overlapping outbreaks from clinical data. In particular, it shows that the data have the required fidelity to detect and characterize multiple, overlapping outbreaks.



Simulated Outbreaks

Keywords

outbreak detection; Bayesian modeling; influenza

Acknowledgments

This work was supported by NIH grant R01 LM011370 on "Probabilistic Disease Surveillance." John Aronis was supported by the National Library of Medicine Training Grant T15LM007059 to the University of Pittsburgh.

References

1. Tsui F-C, Wagner M, Cooper G, Que J, Harkema H, Dowling J, Sriburadej T, Espino J, Vorhees R. Probabilistic case detection for disease surveillance using data in electronic medical records. *Online Journal of Public Health*. 2011;3(3).
2. Cooper G, Villamarin R, Tsui F-C, Millett N, Espino J, Wagner M. A method for detecting and characterizing outbreaks of infectious disease from clinical reports. *Journal of Biomedical Informatics*, 2015, Volume 53, 15-26.
3. Villamarin R, Cooper G, Tsui F-C, Wagner M, Espino J. Estimating the incidence of influenza cases that present to emergency departments. *Proceedings of the International Society for Disease Surveillance*. 2010.

*John M. Aronis

E-mail: jma18@pitt.edu

