Appendix: Content Summaries of Selected Best Papers for IMIA Yearbook 2018, Section 'Public Health and Epidemiology Informatics'.

Choi S, Lee J, Kang MG, Min H, Chang YS, Yoon S

Large-scale machine learning of media outlets for understanding public reactions to nation-wide viral infection outbreaks

Methods Inf Med 2017;129:50–9

Analyzing digital media for understanding public reaction is a current hot topic in Public Health informatics. In this paper, Choi et al. studied, in the context of a nation-wide outbreak of Middle East respiratory syndrome (MERS) in Korea in 2015, the relationship between the disease, social/mass media, and public emotions. They used a sophisticated approach collecting data from 153 news media in Korea (articles and comments representing 86 millions words), generating a dictionary, and performing data analysis based on statistical learning methods (including latent Dirichlet allocation). Then, they analyzed the interplay of public reaction with the epidemics using transfer entropy. The methodological approach and the results are very interesting with the proposition of a positive feedback loop created between the mass media and public emotion variables. The first result is an objectivation of the high levels of fear and worries when mining social media. The second result is the causal interpretation starting by an overestimation of the lethal rate of MERS that led to a high number of articles in the media which triggered fear in the public. This public reaction likely motivated reporters to write poor papers leading to the positive loop.

Dernoncourt F, Lee JY, Uzuner O, Szolovits P
De-identification of patient notes with recurrent neural networks

J Am Med Inform Assoc 2017;24:596–606

The paper presents a new methodology to de-identify Electronic Health Record (EHR) based on artificial neural networks. EHRs are representing a fabulous opportunity for researchers and investigators but their use needs de-identification, that is leaving out any information about name, address, coordinates... Manual approaches are time-consuming and present a poor reproducibility. Statistical approaches have been tried and compared among which decision trees, support vector machines, conditional random fields. This last method has been compared in the present paper with a completely new approach based on artificial neural network (Long Short Term Memory Recurrent Neural Networks) through an i2b2 challenge. The artificial neural network approach out-performed the previous ones being better at incorporating context and being more flexible to variations inherent in human languages.