Appendix: Content Summary of Selected Best Papers for the IMIA Yearbook 2023, section Public Health and Epidemiology Informatics (PHEI)

Valentin S, Arsevska E, Vilain A, De Waele V, Lancelot R, Roche M

Elaboration of a new framework for fine-grained epidemiological annotation

This paper introduces a novel annotation framework for epidemiological surveillance of newspaper articles by proposing a sentence-level annotation approach that deviates from the traditional document-level annotation. By categorizing sentences based on their epidemiological types, the framework enables precise identification and extraction of critical information, ultimately facilitating effective outbreak detection. The authors' focus on meticulous annotation guidelines has significantly improved accuracy and agreement among annotators in four progressive steps. Rigorous evaluation of inter-annotator agreement at each stage further enhances the reliability and robustness of the final annotations. The resulting annotated corpus offers high-quality annotations and specializes specifically in animal health and online news articles, providing a superior alternative to existing benchmark corpora in this domain.

Ji H, Wang J, Meng B, Cao Z, Yang T, Zhi G, Chen S, Wang S, Zhang J

Research on adaption to air pollution in Chinese cities: Evidence from social media-based health sensing

Pollution has been associated with a variety of diseases, including respiratory diseases, cardiovascular diseases, and mental illness. Geographical and socioeconomic factors affect the air quality and health effects in cities differently. Using social media data, researchers can now examine the health impacts of air pollution. It has been demonstrated that user-generated data can be used to track diseases. This is exactly what the study in this paper accomplished. The authors mine health sensing from social media data and examine its relationship to air pollution in 70 Chinese cities in order to fill the aforementioned research gaps. The objectives of this study are: 1) to quantitatively correlate health sensing with air pollution utilizing the generalized additive model, and 2) to examine air pollution sensitivities across 70 Chinese cities. Weibo, the Chinese equivalent of Twitter, is the source of the used social data. Weibo data can detect the real-time health of urban residents. In terms of methodology, a generalized additive model was used to quantify the impact of air pollution on public health sensing from Weibo data. Resulting from the integration of geographical conditions, China is divided into seven distinct regions. In these seven regions, the sensitivity to air pollution varied. Cities with good air quality in the south and east of China were more sensitive to air pollution than those with poor air quality in the northwest and north. The study provides new perspectives and methodologies for monitoring health and the health effects of air pollution.