Social Network Data and Epidemiological Intelligence: A Case Study of Avian Influenza

Camille Schaeffer, Roberto Interdonato, Renaud Lancelot, Mathieu Roche, Maguelonne Teisseire

Methods & Materials: From ProMED-mail spanning 1994 to 2019, we established a unique dengue corpus using professionals' annotations which achieved near perfect agreement (90% Cohen's Kappa statistic). To generate ProMED-mail summaries, a dual-channel bidirectional long-short term memory with an attention mechanism that infuses latent-syntactic features was developed to identify key sentences from the alerts.

Results: Our method outperformed many well-known machine learning and neural network approaches in identifying important sentences, achieving a macro average F_1-score performance of 93%. In addition to verifying the model, we also recruited 5 experts from related fields to conduct a satisfaction survey on the generated summaries, and 83.6% of the summaries received high satisfaction ratings.

Conclusion: The proposed approach successfully fuses latent-syntactic features into a deep neural network to analyze the syntactic content information in the text. It then exploits the derived information to identify key sentences. When a new alert arrives, we can quickly identify the case-relevant-information that is essential for reference or further analysis.

https://doi.org/10.1016/j.ijid.2021.12.232

PS21.03 (551)

Social Network Data and Epidemiological Intelligence: A Case Study of Avian Influenza

C. Schaeffer [1,*], R. Interdonato [1], R. Lancelot [2], M. Roche [1], M. Teisseire [1]

[1] *UMR TETIS, Montpellier, France*
[2] *UMR ASTRE, Montpellier, France*

Purpose: Event Based Surveillance (EBS) systems detect and monitor diseases by analysing articles from online newspapers and reports from health organizations (e.g. FAO, OIE, etc.). However, they partially integrate data from social networks, even though these data are present in large quantities on the web. The purpose of this study is to exploit social network data, such as Twitter and YouTube, to provide epidemiological and additional information for Avian Influenza surveillance.

Methods & Materials: In this context, we propose new text-mining approaches combining lexical rules and statistical approaches, in order to normalise textual data from Social Network ('h5 n8'→'H5N8') and to correct errors from YouTube transcriptions (e.g. 'birth flu'→'bird flu'). Another challenge consists of extracting epidemiological events automatically by identifying spatial entities (Where?), thematic entities (What?), and temporal information (When?). For this, we extended Named Entity Recognition (NER) tools like spaCy.

Results: We collected 100 automatic transcripts of YouTube videos and 268 tweets, in English, dealing with avian influenza, thanks to dedicated API. We obtain encouraging results (i.e. accuracy around 0.6) in order to recognise automatically epidemiological information (e.g. hosts, symptoms etc.) in textual data contents. Extraction of spatial information obtains better results (i.e. accuracy around 0.8).

Conclusion: The final objective of the study consists of linking social media data based on these entities with official information from health organisations, for the improvement of epidemiological monitoring.

https://doi.org/10.1016/j.ijid.2021.12.233

PS21.04 (556)

Understanding Outbreak Data Dissemination In Event Based Surveillance Systems. Application On Avian Influenza Using PADI-web

B. Boudoua [*], C. Hautefeuille, E. Arsevska, S. Valentin

CIRAD, Montpellier, France

Purpose: Epidemic intelligence (EI) has been adopted by several countries to reach fast detection of new and emerging infectious diseases. EI collects information from two types of sources: official sources (i.e. health reports from OIE or FAO) and unofficial sources (i.e. online media outlets, scientific publications, etc.). In France, the EI system PADI-web (Platform for Automated extraction of Disease Information from the Web) is used since 2014 to detect signals of animal health events with risk of introduction to France.

The objective of this work was to understand how health information (signal) is disseminated from a primary source (transmitter) to a final source (EI system) through quantitative and qualitative network analysis methods.

Methods & Materials: We analysed all English reports related to avian influenza detected by PADI-web between August 2018 and June 2019. We used the sources cited in the detected reports to trace the path of each signal. Signals were categorized as official and unofficial according to the source. We have built a directed network where the nodes represented the sources (characterized by type, location and geographical focus) and the edges represented the signal flow. To describe the network, we used network centrality measurements (degree, betweenness and eigenvector) to determine which nodes were important in the data dissemination. We also included the reactivity, calculated as the difference (in days) between the detection of an outbreak by PADI-web and its official notification by Empres-i (gold-standard) with a distinction between wild and domestic birds.

Results: PADI-web detected 202 official signals and 26 unofficial signals. The OIE occupies a central position in the PADI-web information network. National veterinary authorities were the major primary sources. Online news outlets followed by press agencies were the main secondary sources. A significant portion of the signals was detected early in wild birds (41%) and in domestic birds (13%).

Conclusion: This work showed PADI-web's capacity to detect early signals and can be used to define priority sources to improve this tool in terms of reactivity and data quality. In the future, similar work will be conducted on other diseases and EI systems to improve these systems.

https://doi.org/10.1016/j.ijid.2021.12.234

PS21.05 (775)

Detection of Increased Scarlet Fever Incidence Using Digital Surveillance Data

H. Chua [1,2,*], W. Zhou [1,2], B.J. Cowling [1,2], E.H. Lau [1,2]

[1] *School of Public Health, University of Hong Kong, Hong Kong, Hong Kong*
[2] *Laboratory of Data Discovery for health, Hong Kong Science and Technology Park, Hong Kong, Hong Kong*

Purpose: Digital data could be used to monitor disease activity and to detect emerging infections, and is motivated by its abundance, affordability and accessibility. Studies show that Google Trend data alone may not accurately predict incidence of influenza-like-illnesses, but there is a potential to capture the epi-