

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

Camille Schaeffer¹, Roberto Interdonato², Renaud Lancelot³,
Mathieu Roche², Maguelonne Tisseire¹

¹ INRAE, UMR TETIS, Montpellier, France

² CIRAD, UMR TETIS, Montpellier, France

³ CIRAD, UMR ASTRE, Montpellier, France

Purpose

Event Based Surveillance (EBS) systems detect and monitor diseases by analysing articles from online news papers 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 with 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 0.7).

Conclusion

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

Acknowledgement

This work has been funded by the “Monitoring outbreak events for disease surveillance in a data science context” (MOOD) project from the European Union’s Horizon 2020 research and innovation program under grant agreement No. 874850 (<https://mood-h2020.eu>).