How to combine spatio-temporal information and Danger theory for animal disease surveillance?

Bahdja Boudoua, Manon Richard, Mathieu Roche, Maguelonne Teisseire, Annelise Tran

Abstract

Event Based Surveillance (EBS) systems such as HealthMap, Promed and PADI-web are used daily to detect outbreak events reported in web articles. Once the articles are collected, these systems rely on human moderation, and implement supervised classification algorithms to classify articles according to their relevance (Valentin et al., 2020). Applying such supervised methods can be challenging, as epidemiological datasets have an imbalanced class distribution, and because the annotation task, which is critical to the success of these models, can be expensive and time consuming. Another important limitation of EBS systems is that the drivers of disease transmission (e.g. disease characteristics, environmental and epidemiological risk factors) are not always found in textual data and are therefore not taken into account by EBS systems (Kim & Ahn, 2021).

In this context, we propose an unsupervised approach that relies on the spatio-temporal information of the reported epidemiological events, to classify articles while taking into account the environmental factors associated with disease onset through risk mapping. This method, called EpiDCA, is an adaptation of the Dendritic Cells Algorithm (DCA), inspired by the danger theory (Greensmith et al., 2008). EpiDCA is characterized by expert-defined parameters, making it applicable to different diseases and environmental contexts. The proposed method was first tested and evaluated using PADI-web and HealthMap datasets related to avian influenza (AI) in Asia between 2018 and 2019, and a suitability map for AI produced for the same area. To measure the accuracy of the model, we calculated the precision, recall and F-score. EpiDCA achieved a very good performance with an F-score of 0.70 and 0.90 for an imbalanced and a balanced dataset respectively. The results confirmed that considering disease risk factors is a good approach in event classification. EpiDCA was then compared with state-of-the-art supervised machine learning methods and appeared to be competitive.

To test the robustness of the method we then applied EpiDCA for the same disease (AI) in a different geographical context. We used PADI-web dataset related to AI in France between 2022 and 2023 and the corresponding risk map. In this case, events were reported at different levels of spatial granularity (country, region, department and city, from the lowest to the highest granularity). The consistent results confirmed that this method is robust, and the best performance in terms of precision, recall and F-score was observed when events were reported at the department level.

In the future, additional tests will be conducted on different datasets and case studies (a cross-border disease such as African Swine Fever, and a vector-borne disease such as West Nile fever disease) to evaluate the genericity of the proposed method.

References

Valentin, S., Arsevska, E., Falala, S., De Goër, J., Lancelot, R., Mercier, A., ... & Roche, M. (2020). PADI-web: A multilingual event-based surveillance system for monitoring animal infectious diseases. Computers and Electronics in Agriculture, 169, 105163.

Kim, J. & Ahn, I. (2021). Infectious disease outbreak prediction using media articles with machine learning models. Scientific reports, 11(1), 4413.

Greensmith, J. & Aickelin, U. (2008). The deterministic dendritic cell algorithm. In International conference on artificial immune systems . Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 291-302.

Keywords

Danger theory; Dendritic cells algorithm; Event-Based Surveillance; Spatio-temporal data; Risk factors