

Comment combiner informations sanitaires et environnementales dans les systèmes de veille épidémiologiques?

B. Boudoua^{1,3}, M. Roche^{1,4}, M. Teisseire^{1,3}, A. Tran^{1,2,4}

¹ UMR TETIS, Univ. Montpellier, AgroParisTech, CIRAD, CNRS, INRAE, Montpellier, France.

² UMR ASTRE, Univ. Montpellier, CIRAD, INRAE, Montpellier, France.

³ INRAE, UMR TETIS, Montpellier, France.

⁴ CIRAD, UMR TETIS, F-34398 Montpellier, France.

Résumé

Le rôle principal des systèmes de surveillance fondés sur les événements (SBE) est de détecter les nouvelles épidémies (événements) en explorant les informations sanitaires publiées en ligne dans un large éventail de sources formelles et informelles. Les facteurs de risques (environnementaux, climatiques, liés aux pratiques d'élevage etc.) n'étant pas pris en compte par ces systèmes, le but de ce travail est d'apporter une démarche générique qui permet de renforcer ou non un événement détecté par les systèmes de veille en y intégrant des facteurs de risques environnementaux.

Dans ce contexte, nous proposons une nouvelle méthode inspirée de la théorie du danger qui permet d'intégrer et pondérer des indicateurs environnementaux. Ce travail préliminaire sera expérimenté sur trois cas d'étude différents afin de tester la robustesse et la généralité de l'approche proposée.

Mots-clés

Veille sanitaire, grippe aviaire, facteurs de risque, théorie du danger.

Abstract

The primary role of event-based surveillance (EBS) systems is to detect news outbreaks (events) by mining health information published online from a wide range of formal and informal sources. The risk factors (environmental, climatic, related to the practice of breeding, etc.) not being taken into account by these systems, the aim of this work is to provide a generic approach which makes it possible to reinforce or not an event detected by event-based surveillance systems by integrating environmental risk factors.

In this context, we provide a new method inspired by the danger theory which makes it possible to integrate and weight environmental indicators. This preliminary work will be experimented on three different case studies in order to test the robustness and generation of the proposed approach.

Keywords

Epidemic intelligence, avian influenza, risk factors, danger theory.

1 Introduction

Pour faire face aux maladies émergentes qui représentent un risque croissant pour la santé publique, de nombreux pays adoptent une stratégie de veille sanitaire [9]. Celle-ci repose sur deux composantes : la surveillance fondée sur les indicateurs (SBI) (issus de sources officielles telles que l'OIE, l'OMS, la FAO) et la surveillance fondée sur les événements (SBE) issus de sources non-officielles (média en ligne, réseaux sociaux, etc.), assurée par des systèmes de SBE tels que ProMed [4], HealthMap [3], PADI-web [12].

Le but de ce travail est d'apporter une démarche générique (indépendante d'une maladie ou d'un hôte spécifique) qui permette d'identifier les descripteurs épidémiologiques tout en prenant en compte les facteurs de risque (environnementaux, météorologiques, liés aux pratiques d'élevage, etc.) permettant ainsi de renforcer ou non un événement détecté par un système de veille.

2 État de l'art

Les systèmes de SBE sont utilisés quotidiennement afin de détecter les événements de santé inhabituels. Ils collectent et analysent un flux quotidien de données textuelles non structurées (articles) à partir d'internet, en utilisant des mots-clés ou des combinaisons de mots-clés [1]. Par la suite, ces articles sont triés selon leur pertinence et classés par date, localisation géographique, source, et maladie. La surveillance basée sur les événements permet ainsi l'obtention de nombreuses informations mais présente certaines limites. En particulier, les facteurs de risque liés à l'apparition des maladies ne sont pas toujours retrouvés dans les données textuelles et ne sont pas pris en compte par les SBE.

Parallèlement, la cartographie du risque en épidémiologie permet de mettre en évidence les zones favorables à l'apparition d'une maladie en s'appuyant sur la répartition spatiale des facteurs de risque associés [6]. La connaissance de ces facteurs de risque est essentielle pour mieux cibler

les zones de surveillance et adapter les mesures de lutte et de prévention [2].

Dans cette étude nous proposons une approche inspirée de la théorie du danger, afin de combiner les facteurs de risque aux données épidémiologiques issues des systèmes de veille.

La théorie du danger [8] est basée sur le fonctionnement des cellules immunitaires dendritiques. Elle stipule que la reconnaissance d'un antigène par une cellule ne réside pas dans la distinction entre le soi et le non-soi mais dépend plutôt du contexte environnemental (signaux) dans lequel il est identifié. Cette théorie a servi d'inspiration pour le développement d'algorithmes de classification [5] qui ont été appliqués avec succès par exemple dans le domaine de la sécurité informatique [7, 10], ou encore en sismologie [13].

Cette méthode permet de classer des éléments appelés "antigènes" en intégrant des données hétérogènes par le biais de deux types de signaux, tous deux appartenant généralement à l'intervalle [0; 100].

- Les signaux de danger (*danger signals*) : augmentent proportionnellement à la présence de données représentant une situation "anormale".
- Les signaux sécuritaires (*safe signals*) : augmentent proportionnellement à la présence de données représentant une situation "normale".

L'algorithme applique une fonction de somme pondérée afin de générer un signal sortant.

Les signaux sortants servent à définir un contexte cellulaire (normal ou anormal) qui sera par la suite utilisé pour calculer le coefficient d'anomalie de chaque antigène.

3 Transposition à la veille épidémiologique

Dans le contexte de nos travaux, les événements extraits d'articles détectés par les systèmes SBE représentent nos antigènes (ce que l'on veut classer), associés, par correspondance spatiale, à des données environnementales (Figure 1 Étape 1).

Dans une étape de pré-traitement et catégorisation (Figure 1 Étape 2), ces données d'entrées sont converties en deux catégories de signaux : signaux de danger (*danger signals*) et signaux sécuritaires (*safe signals*). Les données épidémiologiques issues de ces articles (source de l'information, hôte, maladie) constituent les signaux de danger. Nous nous référons à la connaissance d'experts afin d'établir un barème et donner une note à chaque donnée observée par exemple : +20 si la source de l'article est officielle, +15 si l'hôte est un animal domestique, etc. Si aucune de ces données n'est présente, le signal de danger est nul et l'article n'est pas pris en compte. Les données environnementales représentent les signaux sécuritaires. Un signal sécuritaire maximal indique que l'environnement est défavorable

à l'apparition de la maladie, un signal sécuritaire égal à 0 indique au contraire un environnement favorable où tous les facteurs de risque identifiés sont présents.

L'exposition des cellules (Figure 1 Étape 3) est ensuite calculée en combinant les signaux pondérés de danger et sécuritaires. La pondération est adaptée à la maladie étudiée : pour une maladie fortement influencée par les facteurs environnementaux (par exemple une maladie à transmission vectorielle comme la fièvre du Nil Occidental, ou impliquant un réservoir sauvage, comme la grippe aviaire) un poids plus important sera attribué aux signaux sécuritaires qui caractérisent le contexte environnemental dans notre approche.

Enfin, dans l'étape de classification finale des événements détectés par les SBE (Figure 1 Étape 4), les signaux sortants sont utilisés pour générer un coefficient d'anomalie propre à chaque antigène et qui prend ainsi en compte à la fois les informations sanitaires et le contexte environnemental.

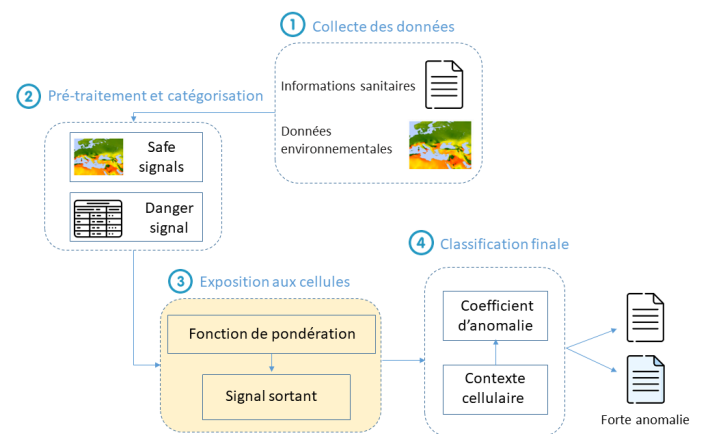


FIGURE 1 – Vers un processus en 4 étapes pour la veille épidémiologique

4 Perspectives

Le premier cas d'étude auquel nous nous intéressons est la grippe aviaire, d'une part parce qu'il s'agit d'une maladie médiatisée et de ce fait, un nombre conséquent d'articles est détecté par les SBE et d'autre part parce que l'émergence et la diffusion de cette maladie dépendent de différents facteurs de risque (proximité des zones humides, population d'oiseaux sauvage, population d'oiseaux domestiques, etc.) [11].

Cette méthode sera ensuite testée sur d'autres cas d'étude pour tester sa généralité : pour une maladie à transmission vectorielle comme la fièvre du Nil Occidental et pour une maladie transfrontalière comme la peste porcine africaine, et ce dans différents contextes géographiques.

La dimension spatio-temporelle des événements épidémiologiques sera également prise en compte dans le traitement des données.

Remerciements

Ce travail est financé par le projet « Monitoring outbreak events for disease surveillance in a data science context » (MOOD) du programme de recherche et d'innovation Horizon 2020 de l'Union européenne dans le cadre de la convention de subvention n° 874850 (<https://mood-h2020.eu/>)

Références

- [1] Philippe BARBOZA. « Evaluation des systèmes d'intelligence épidémiologique appliqués à la détection précoce des maladies infectieuses au niveau mondial. » Thèse de doct. Université Pierre et Marie Curie-Paris VI, 2014 (cf. p. 1).
- [2] Hannes BERGMANN, Katja SCHULZ, Franz J CONRATHS et Carola SAUTER-LOUIS. « A review of environmental risk factors for African Swine Fever in European wild boar ». In : *Animals* 11.9 (2021), p. 2692 (cf. p. 2).
- [3] John S BROWNSTEIN, Clark C FREIFELD, Ben Y REIS et Kenneth D MANDL. « Surveillance Sans Frontieres : Internet-based emerging infectious disease intelligence and the HealthMap project ». In : *PLoS medicine* 5.7 (2008), e151 (cf. p. 1).
- [4] Malwina CARRION et Lawrence C MADOFF. « ProMED-mail : 22 years of digital surveillance of emerging infectious diseases ». In : *International health* 9.3 (2017), p. 177-183 (cf. p. 1).
- [5] Zeineb CHELLY et Zied ELOUEDI. « A survey of the dendritic cell algorithm ». In : *Knowledge and Information Systems* 48.3 (2016), p. 505-535 (cf. p. 2).
- [6] A HESS, JK DAVIS et MC WIMBERLY. « Identifying environmental risk factors and mapping the distribution of West Nile virus in an endemic region of North America ». In : *GeoHealth* 2.12 (2018), p. 395-409 (cf. p. 1).
- [7] Lincy E JIM, Nahina ISLAM et Mark A GREGORY. « Enhanced MANET security using artificial immune system based danger theory to detect selfish nodes ». In : *Computers & Security* 113 (2022), p. 102538 (cf. p. 2).
- [8] Polly MATZINGER. « The danger model : a renewed sense of self ». In : *science* 296.5566 (2002), p. 301-305 (cf. p. 2).
- [9] Christoph PAQUET, Denis COULOMBIER, Reinhard KAISER et Massimo CIOTTI. « Epidemic intelligence : a new framework for strengthening disease surveillance in Europe ». In : *Eurosurveillance* 11.12 (2006), p. 5-6 (cf. p. 1).
- [10] Aakanksha SHARAFF, Chandramani KAMAL, Sidhartha PORWAL, Surbhi BHATIA, Kuljeet KAUR et Mohammad Mehendi HASSAN. « Spam message detection using Danger theory and Krill herd optimization ». In : *Computer Networks* 199 (2021), p. 108453 (cf. p. 2).
- [11] Kim B STEVENS, Marius GILBERT et Dirk U PFEIFFER. « Modeling habitat suitability for occurrence of highly pathogenic avian influenza virus H5N1 in domestic poultry in Asia : a spatial multi-criteria decision analysis approach ». In : *Spatial and spatio-temporal epidemiology* 4 (2013), p. 1-14 (cf. p. 2).
- [12] Sarah VALENTIN, Elena ARSEVSKA, Sylvain FALALA, Jocelyn DE GOËR, Renaud LANCELOT, Alizé MERCIER, Julien RABATEL et Mathieu ROCHE. « PADI-web : A multilingual event-based surveillance system for monitoring animal infectious diseases ». In : *Computers and Electronics in Agriculture* 169 (2020), p. 105163 (cf. p. 1).
- [13] W ZHOU, YW LIANG, Z MING et HB DONG. « Earthquake prediction model based on danger theory in artificial immunity ». In : *Neural Network World* 30.4 (2020), p. 231 (cf. p. 2).