

Epi_DCA : Adaptation et mise en œuvre de la théorie du danger pour la veille épidémiologique

Bahdja Boudoua^{1,3}, Mathieu Roche^{1,4}, Maguelonne Teisseire ^{1,3}, Annelise 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.

Le rôle principal des systèmes de surveillance basés 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. Cependant, les facteurs de risque ne sont pas toujours pris en compte par les SBE. Dans ce contexte de veille sanitaire, nous avons posé les premières bases d'une démarche générique (indépendante d'une maladie ou d'un hôte spécifique) afin de renforcer ou non un événement détecté par les SBE en y intégrant les facteurs de risques disponibles. Epi_DCA est une adaptation de l'algorithme des cellules dendritiques (DCA) à la problématique de veille sanitaire.

Les cellules dendritiques permettent l'activation de la réponse immunitaire. Elles passent d'un état "semi-mature" (inhibiteur) à "mature" (activateur) en fonction des deux types de signaux auxquels elles sont exposées : 1) les signaux de danger augmentent à la présence de données représentant une situation "anormale", 2) les signaux sécuritaires augmentent à la présence de données représentant une situation "normale". Ce comportement a inspiré l'algorithme DCA qui se décline en quatre phases : 1) catégorisation du signal (utilisant en amont une réduction des attributs), 2) détection des antigènes (utilisant des pondérations éventuellement empiriques pour obtenir des signaux de sortie cumulés), 3) évaluation du contexte cellulaire, 4) classification finale des antigènes (par un coefficient d'anomalie). (Chelly and Elouedi, 2016). La contribution principale de ce travail est l'intégration de l'information spatio-temporelle dans la méthode. 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. Notre premier cas d'étude autour de l'influenza aviaire (IA) s'appuie sur un jeu de données constitué de 174 articles (87 pertinents et 87 non pertinents) issus du SBE Healthmap (Freifeld et al., 2008) et classés manuellement par un épidémiologiste, pour la région d'Asie du Sud-Est sur une période allant de 2018 à 2019. Les données épidémiologiques issues des articles détectés (source d'information, hôte, maladie) sont utilisées pour générer les signaux de danger. Nous nous référons à la connaissance d'experts afin d'établir un score pour chaque donnée observée. Pour générer les signaux sécuritaires, nous avons créé une carte de risque d'occurrence d'IA selon la méthode de (Stevens et al., 2013) en utilisant des données récentes sur les populations d'hôtes sensibles. Ensuite, les événements ont été associés aux données environnementales par correspondance spatiale

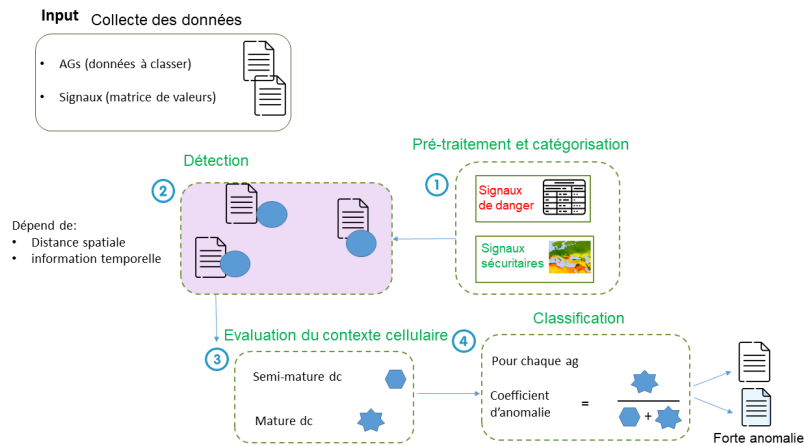


FIG. 1 – Processus en 4 phases pour la veille épidémiologique

à l'aide d'un Système d'Information Géographique (SIG)¹. Les résultats obtenus, fondés sur un F-score de 0.86, suggèrent que la prise en compte du contexte environnemental (dans ses dimensions spatio-temporelles) dans l'analyse des données épidémiologiques permet de renforcer les articles détectés par les SBE.

Afin de comparer la méthode proposée, nous avons évalué 4 méthodes d'apprentissage supervisé (SVM, Naive Bayes, Knn et Random Forest) sur notre jeu de données en effectuant une validation croisée en 5 plis. Nous avons obtenu une F-mesure entre 0.86 (Naive Bayes) et 0.91 (SVM) ce qui montre que notre approche *Epi_DCA*, qui a la caractéristique d'être non supervisée, reste tout à fait compétitive.

Dans cette étude, nous avons réalisé une première évaluation de l'approche qui sera étendue à d'autres jeux de données et sur d'autres cas d'étude pour tester la généralité et la robustesse de la méthode.

Références

- Z. Chelly and Z. Elouedi. A survey of the dendritic cell algorithm. *Knowledge and Information Systems*, 48(3) :505–535, 2016.
- C. C. Freifeld, K. D. Mandl, B. Y. Reis, and J. S. Brownstein. Healthmap : global infectious disease monitoring through automated classification and visualization of internet media reports. *Journal of the American Medical Informatics Association*, 15(2) :150–157, 2008.
- K. B. Stevens, M. Gilbert, and D. U. Pfeiffer. Modeling habitat suitability for occurrence of highly pathogenic avian influenza virus H5N1 in domestic poultry in Asia : a spatial multicriteria decision analysis approach. *Spatial and spatio-temporal epidemiology*, 4 :1–14, 2013.

1. www.qgis.org