

Informativité dans les Contenus Textuels Twitter pour la Phytosurveillance Centrée sur l'observation des Agriculteurs

Shufan Jiang^{*,**}, Rafael Angarita^{*}, Stéphane Cormier^{**},
Julien Orensanz^{***}, Francis Rousseaux^{**}

^{*}Institut Supérieur d'Electronique de Paris, LISITE, Paris, France
name.lastname@isep.fr,

^{**} Université de Reims Champagne Ardenne, CReSTIC, Reims, France
name.lastname@univ-reims.fr

^{***} CAP2020, Gradignan, France

Face au défi de la croissance démographique et de l'évolution des habitudes alimentaires, l'agriculture de précision a été introduite pour augmenter durablement la production alimentaire. Un sujet important en agriculture de précision est de mesurer les risques naturels au niveau global ou local, puis d'améliorer les tâches de prévention des risques tout en limitant l'usage et l'impact des pesticides. Suite au paradigme *social sensing* Wang et al. (2015), les individus -qu'ils soient agriculteurs ou non- ont de plus en plus de connectivité à l'information lors de leurs déplacements, au niveau du terrain. Chaque individu peut devenir un diffuseur d'informations. En ce sens, des informations sur les dangers en temps réel sont publiées sur les réseaux sociaux tels que Twitter. En effet, Twitter permet aux agriculteurs d'échanger leurs expériences entre eux, de s'abonner à des sujets d'intérêt à l'aide d'hashtags et de partager des informations en temps réel sur les risques naturels. Par rapport aux applications payantes, les informations sur Twitter, présentées sous forme de texte, d'image, de son, de vidéo ou une association de ces entités, sont plus accessibles au public, mais moins formalisées ou structurées. En effet, de plus en plus d'agriculteurs s'impliquent dans les communautés Twitter en ligne en ajoutant des hashtags, comme #AgriChatUK¹ ou #FrAgTw². Ainsi, Twitter pourrait être considéré comme une source ouverte d'échange de connaissances d'agriculteur à agriculteur.

Compte tenu de la nature de ces publications, il n'est pas simple d'exploiter efficacement les informations qu'elles contiennent, encore moins de le faire automatiquement et de relier ces données à des données provenant d'autres sources telles que des capteurs ou d'autres systèmes d'information. Dans l'état de l'art sur social sensing, le Web de données et les technologies de traitement automatique des langues (TAL) sont mobilisés. Le projet de recherche Vigi4Med Karapetiantz et al. (2019) vise à utiliser les commentaires des patients sur les effets indésirables susceptibles d'être liés aux médicaments rapportés dans les réseaux sociaux comme source d'information complémentaire pour renforcer la pharmacovigilance. Les expériences de Vigi4med prouvent que le réseau social contient suffisamment de données pour l'évaluation de certains problèmes médicaux. Cependant, le nombre de faux positifs détectés

1. <http://www.agrichatuk.org/>

2. <https://franceagritwittos.com/>

dans Vigi4med souligne également l'importance de l'annotation manuelle par des experts du domaine. La plateforme suricate-nat Auclair et al. (2019) propose un outil basé sur les champs aléatoires conditionnels (CRFs) pour la détection de la localisation des municipalités et une méthode semi-supervisée pour l'inférence de géolocalisation. Les travaux de Suricate-NAT peuvent être réutilisés pour la détection de la localisation dans la surveillance des nuisibles.

Cet article pose les questions suivantes : quelles informations phytosanitaires peuvent être extraites des contenus textuels sur Twitter à l'aide des technologies de TAL, et la qualité de ces informations.

Pour répondre à cette question, en collaboration avec les experts d'Arvalis - institut du végétal ³ et de Cap2020, fabriquant des pièges connectés ⁴, nous avons collecté des Tweets pour plusieurs cas d'observation :

1. "pyrale du maïs", représentant un insecte ravageur
2. "JNO" (abréviation de "jaunisse nanisante de l'orge"), représentant respectivement les maladies des plantes
3. le rendement des céréales à paille, représentant une activité agricole

En fonction des dimensions de chaque cas, nous avons appliqué différentes méthodes pour extraire l'information.

Références

- Auclair, S., F. Boulahya, B. Birregah, R. Quique, R. Ouaret, et E. Soulier (2019). SURICATE-Nat : Innovative citizen centered platform for Twitter based natural disaster monitoring. In *2019 International Conference on Information and Communication Technologies for Disaster Management (ICT-DM)*, Paris, France, pp. 1–8. IEEE.
- Karapetiantz, P., A. Lillo-Le Louët, et C. Bousquet (2019). Informativité des forums de discussion français pour l'évaluation des effets indésirables du baclofène. *Thérapies* 74(6), 569–578.
- Wang, D., T. Abdelzaher, et L. Kaplan (2015). *Social sensing : building reliable systems on unreliable data*. Morgan Kaufmann.

Summary

Data mining in social media has been widely applied in different domains for monitoring and measuring social phenomena, such as opinion analysis towards popular events, sentiment analysis of a population, tracing early secondary effects of medicinal drugs and earthquake detection. In the other side, social media also brings new forms of bias and shadings while it attracts people to circulate information by its openness. As farmers are increasingly getting engaged in sharing their observation and their knowledge on social media, we want to valorize these data. In this work, we chose several scenarios to collect Tweets, then we applied different natural language processing technologies to measure the informativeness of Tweets in French as a complementary source for phytosanitary monitoring.

3. <https://www.arvalisinstitutduvegetal.fr/index.html>

4. <https://www.cap2020.online/>