Création d'un corpus de tweets en français pour la détection automatique de positionnement (stance)

Rémi Uro, Marc Evrard,
Nicolas Hervé, Béatrice Mazoyer

Institut National de l'Audiovisuel (Ina)
18 Avenue des Frères Lumière, 94360 Bry-sur-Marne
{ruro,mevrard,nherve,bmazoyer}@ina.fr

1 Contexte

Le problème des infox (ou *Fake News*) est de plus en plus présent sur les réseaux sociaux, notamment lors d'évènements médiatiques importants. Leur détection est une tâche complexe, requérant une connaissance du monde à la fois large et précise, ce qui rend cette tâche difficile à réaliser de manière complètement automatique. Dans ce projet, nous nous concentrons sur une tâche intermédiaire : la détection du positionnement (*stance*) d'un énoncé par rapport à un autre. En effet, les réseaux sociaux permettent des échanges dans lesquels les interlocuteurs émettent des commentaires qui sont par nature argumentatifs. On s'attend donc à retrouver différents positionnements dans un fil de discussion. Les polarités de ces positionnements peuvent être des indicateurs très intéressants pour identifier des échanges associés à une controverse importante et donc potentiellement liés à des infox. Des travaux étudiant ces phénomènes existent déjà, Zubiaga et al. (2018) présente notamment un inventaire détaillé de ces études. Nous proposons ici de créer un nouveau corpus de *tweets* en français, annotés pour l'apprentissage automatique de la détection du positionnement.

Le projet OTMedia, dans lequel s'inscrit la réalisation de ce projet, a pour but d'étudier les évènements médiatiques dans le paysage français. Le projet Observatoire Transmedia (OTMedia), initié à l'Ina en 2010, est une plateforme de recherche permettant d'analyser d'importants volumes de données transmedia multimodales, hétérogène et liées à l'actualité française et francophone (Hervé et al., 2013).

Nous présenterons ici les problématiques inhérentes à la réalisation ce corpus. Nous avons choisi d'annoter la position de ces tweets selon les 4 catégories proposées dans la littérature — notamment chez Procter et al. (2013) — et utilisées dans la plupart des corpus existants dans d'autres langues : Soutient, Contredit, Questionne, Commente (*Support, Deny, Query, Comment*).

Nous pensons que ces catégories ne sont pas optimales ; elles ne recouvrent en effet pas tous les cas possibles — on peut par exemple montrer son désaccord tout en s'interrogeant — et elles ne forment pas des classes complètement dissociées. Nous les utilisons tout de même pour nos annotations par soucis de compatibilité avec les conventions adoptées par la communauté.

2 Annotation des tweets

Les tweets que nous annotons proviennent d'un corpus francophone de 80 000 tweets collectés au cours de l'été 2018 et annotés comme correspondant à des évènements médiatiques identifiés. Nous nous intéressons ici au positionnement d'un tweet par rapport à ceux présents dans le fil de discussion. Nous annotons le positionnement du tweet courant par rapport au parent dans le fil et par rapport à la racine du fil de discussion. Nous avons choisi de garder uniquement les tweets racines ayant au moins un nombre minimum de réponses. Il nous a semblé plus pertinent de nous baser sur le nombre de réponses d'un tweet, plutôt que sur la longueur des fils de discussion qui en sont issus. Aussi, afin d'éviter les phénomènes de dérives au cours des discussions, nous ne gardons les réponses que jusqu'au deuxième niveau (réponse de réponse au tweet racine du fil de discussion). Après ce filtrage, il reste environ 15 000 tweets à annoter.

Pour faciliter la tâche des annotateurs, nous avons construit une interface d'annotation affichant l'ensemble des fils au départ d'un tweet racine. Chaque tweet est accompagné d'un champ correspondant aux 4 catégories, complété d'une valeur par défaut correspondant à l'incapacité de l'annotateur à faire un choix. Pour les tweets qui ne sont pas une réponse directe au tweet racine, les champs sont affichés 2 fois : un pour l'annotation du tweet courant par rapport au parent et l'autre pour l'annotation par rapport au tweet initial. Un affichage en colonne permet de présenter clairement ces deux niveaux d'annotation pour l'utilisateur.

Afin d'annoter en priorité les fils de discussion les plus intéressants, les paquets de tweets sont présentés par ordre décroissant de nombre de tweets. Les tweets ayant le plus de réponses sont en effet ceux les plus probablement sujets à controverse. Il est aussi plus aisé pour les annotateurs d'analyser de long fils de tweets, qu'une succession de petits fils de discussion n'ayant potentiellement pas de rapport entre eux.

Bien qu'étant en apparence une tâche relativement simple, nous avons soulevé de nombreuses interrogations durant la phase de développement du système au sujet du protocole d'annotation. Nous avons par exemple choisi de ne pas annoter les fils de discussion dont le tweet racine n'est pas rédigé en français. Nous nous sommes aussi interrogés sur la question de l'ironie et du *trolling*, ou encore sur les réponses nuancées dont le positionnement n'est pas clairement ou globalement identifiable.

Références

Hervé, N., M.-L. Viaud, J. Thièvre, A. Saulnier, J. Champ, P. Letessier, O. Buisson, et A. Joly (2013). Otmedia : the french transmedia news observatory. In *ACM Multimedia*, pp. 441–442.

Procter, R., F. Vis, et A. Voss (2013). Reading the riots on twitter : methodological innovation for the analysis of big data. *International Journal of Social Research Methodology 16*(3), 197–214.

Zubiaga, A., A. Aker, K. Bontcheva, M. Liakata, et R. Procter (2018). Detection and resolution of rumours in social media : A survey. *ACM Comput. Surv. 51*(2), 32 :1–32 :36.

Summary

In this paper, we present the construction of a French Twitter thread corpus, labeled for the automatic detection of stance. The context of the research is presented and some questions regarding the annotation process are discussed.