L'exploitation des techniques de régression pour l'évaluation de la crédibilité des tweets

Hamda Slimi, Ibrahim Bounhas
Yahya Slimani

Laboratoire LISI
<http://www.jarir.tn>
École Nationale des Sciences de l'Informatique
Université de la Manouba, La Manouba 2010, Tunisie
hamda.slimi@ensi-uma.tn,
ibrahim.bounhas@gmail.com
yahya.slimani@gmail.com

Résumé. La crédibilité de l'information diffusée sur les réseaux sociaux est de plus en plus suspectée. En effet, accusés d'avoir servi de plate-formes de propagande pendant des événements de grande envergure (Elections américaine, BREXIT, etc.), ces réseaux ne cessent pas de perdre en crédibilité. Dans ce contexte, plusieurs travaux de recherche ont été proposés pour faire face à ces manipulations de l'information. Dans ce travail, nous présentons une approche permettant l'évaluation de la crédibilité des informations sur le réseau social Twitter. L'approche modélise la crédibilité comme une valeur continue qui prend en compte plusieurs caractéristiques des tweets, afin d'analyser et de prédire la crédibilité de ces derniers. Plusieurs variantes de notre approche ont été testées afin de valider l'apport de notre proposition. Les résultats obtenus ont montré la qualité et l'intérêt de notre approche.

1 Introduction

Twitter était considéré comme un réseau social régulier où ses utilisateurs partagent leurs activités quotidiennes. Toutefois, plusieurs facteurs ont contribué à l'évolution de Twitter en une plateforme de renommée mondiale. La nature des interactions sociales, où Twitter permet aux utilisateurs de se suivre mutuellement sans demande de permission explicite, contrairement à Facebook qui exige l'approbation d'une demande d'amitié entre utilisateurs. C'est ainsi que Twitter est devenu plus centré autour d'une structure appelée graphe d'intérêt, où les utilisateurs suivent d'autres qui leurs fournissent des informations ayant une valeur ajoutée (et qui ont donc un intérêt). De plus, certaines caractéristiques d'un tweet, tels que le hashtag (topic) et sa taille (280 caractères) permettent d'obtenir une description correcte et brève d'informations dignes d'intérêt. Dans (Kwak et al., 2010) les auteurs ont collecté et analysé 106 millions de tweets décrivant 4262 sujets de tendances. Cette analyse leur a permis de constater que 85 % des sujets ont fait la une des journaux et tout tweet retweeté à atteint jusqu'à 1000 utilisateurs, quel que soit le nombre d'abonnés du compte de son auteur originel. Ces caractéristiques