

Un système de détection de thématiques populaires sur Twitter

Adrien Guille*, Cécile Favre*

*Laboratoire ERIC, Université Lumière Lyon 2
<http://mediamining.univ-lyon2.fr/people/guille/egc2014.php>
<http://mediamining.univ-lyon2.fr/people/guille/twitterstream.php>

1 Introduction

Twitter offre des fonctionnalités de microblogging qui sont utilisées par des millions de personnes à travers le monde pour publier des messages courts. Ces personnes créent et partagent de l'information liée à divers types d'événements, allant d'événements personnels banals à des événements importants et/ou globaux, quasiment en temps-réel. L'explosion du nombre d'utilisateurs de ce service de réseautage social a entraîné l'apparition d'un phénomène de surcharge informationnelle. Pour lutter contre cela, il est nécessaire de doter les utilisateurs de moyens leur permettant d'identifier plus facilement les éléments d'information les plus intéressants et de se tenir au courant des derniers événements significatifs.

L'information brute produite par Twitter est délivrée sous la forme d'un flux de messages courts. Par conséquent la manière dont ceux-ci arrivent au fil du temps recèle une part importante de leur signification. La dynamique temporelle des thématiques les plus populaires sur ces réseaux est constituée d'une succession de focus et dé-focus, autrement dits, une succession de *pics* de popularité (Leskovec et al., 2009). C'est pourquoi de nombreuses approches – allant de méthodes basées sur la fréquence des mots (Benhardus et Kalita, 2013) jusqu'à des méthodes plus complexes reposant sur des modèles de thématiques probabilistes dynamiques (Lau et al., 2012) – ont été proposées dans le but d'identifier ce genre de thématiques. Dans cet article, nous présentons un système implémentant la méthode décrite par Guille et Favre (2014). Contrairement à la majorité des méthodes existantes, celle-ci prend en compte l'aspect social du flux traité en considérant la fréquence de création de liens dynamiques entre utilisateurs. Un utilisateur crée un lien dynamique en insérant une ou plusieurs mentions (*i.e.* « @pseudonyme ») dans un tweet. Ce lien entre utilisateurs est dynamique car lié au contenu du tweet et sa durée de vie. Par ailleurs, cette méthode localise plus précisément dans le temps les thématiques que les méthodes existantes et traite les grands volumes de données plus efficacement que celles à base de modèles de thématiques probabilistes.

La suite de cet article est organisée comme suit. Dans la section 2 nous décrivons le fonctionnement du système puis dans la section 3 nous détaillons le cadre de la démonstration. Enfin dans la section 4 nous concluons.