

Tri des actualités sociales: État de l’art et Pistes de recherche

Sami Belkacem*, Kamel Boukhalfa*, Omar Boussaid**

*Laboratoire LSI, USTHB-Alger, Algérie - {s.belkacem,kboukhalfa}@usthb.dz

**Laboratoire ERIC, Université de Lyon, Lyon 2, France - omar.boussaid@univ-lyon2.fr

Résumé. En raison de la grande quantité d’informations (messages, articles, vidéos, musiques, images, etc.) produites et partagées sur les réseaux sociaux, les utilisateurs se retrouvent submergés d’informations générées chronologiquement dans leur fil d’actualité. De plus, la majorité des informations peuvent s’avérer non pertinentes. Le tri, par ordre de pertinence des actualités sociales, est proposé comme une solution pour aider les utilisateurs à consulter et interagir rapidement avec les informations susceptibles de les intéresser. Dans ce travail, nous étudions les approches existantes dans le domaine du tri des actualités sociales, et exposons leurs limites et quelques pistes de recherche selon plusieurs axes: les facteurs influençant la pertinence des informations, les modèles de prédiction de la pertinence, l’apprentissage et l’évaluation des modèles de prédiction, etc.

1 Introduction

Les réseaux sociaux occupent de plus en plus de place dans notre quotidien (Zhan et al., 2016). Les données manipulées sur ces réseaux sont connues pour leurs volumes qui peuvent atteindre des Pétaoctets (10^{15} octets) voire plus, leur hétérogénéité (messages, articles, vidéos, musiques, images, etc.), leur variété (pouvant provenir de différentes sources), et leur vélocité (arrivant en temps réel ou presque) (Xu et al., 2016). Ces caractéristiques ont fait que les technologies de gestion et de traitements classiques se trouvent dans l’incapacité de traiter ce type de données (Krishnan, 2013). Les données sociales constituent un des volets ayant contribué à l’apparition du concept de *big data* qui est défini par les 5 “V” : Volume, Variété, Vélocité, Valeur et Véracité (Lomotey et Deters, 2014). Sur ces réseaux, en raison de la grande quantité d’informations produites et partagées (Guy et al., 2011), les utilisateurs se retrouvent submergés d’informations générées chronologiquement dans leur fil d’actualité¹ (Berkovsky et al., 2012). Pour un utilisateur standard de *Facebook*² par exemple, 1500 nouvelles informations sont générées chaque jour dans son fil d’actualité³. En outre, plusieurs travaux de recherche ont montré que la majorité de ces informations sont considérées comme non pertinentes (Hong

1. Liste d’informations récentes (publications, posts, tweets, etc.) qui permet à un utilisateur de suivre l’actualité des membres de son réseau social.

2. www.facebook.com

3. www.slate.fr/story/112681/qui-controle-ce-qui-apparaît-sur-votre-fil-facebook