

Mots-clefs, Collaborations, Sentiments: Évolutions de la Conférence EGC depuis 2004

Erick Stattner, Didier Henry, Nathan Jadoul

Université des Antilles
Laboratoire de Mathématiques, Informatique et Applications (LAMIA)
{erick.stattner, didier.henry, nathan.jadoul}@univ-antilles.fr

Résumé. Dans cet article, nous menons un travail d'analyse de données issues de la conférence “*Extraction et Gestion des Connaissances*” (EGC). Ces données concernent les articles acceptés à la conférence EGC depuis 2004 ainsi que les messages publiés sur le réseau social Twitter. Notre objectif est de mettre en évidence, depuis 2004, les grandes évolutions de la conférence EGC en matière (i) de mot-clefs, (ii) de collaborations et (iii) de sentiments exprimés sur Twitter. Les résultats obtenus ont ainsi permis de mettre en lumière l'évolution des thématiques de la conférence depuis 2004, les schémas collaboratifs qui se mettent en place ainsi que leur évolution, et les sentiments qui émergent des messages publiés autour de la conférence sur Twitter.

1 Introduction

L'extraction de connaissances de données hétérogènes est devenue un des axes de recherche les plus prolifiques de ces dernières décennies (Waller et Fawcett, 2013; Van Der Aalst, 2016). Cette explosion des travaux menés autour de l'analyse de données s'explique par plusieurs facteurs. (i) Tout d'abord la collecte de données est aujourd'hui massive. En effet, des données très hétérogènes peuvent aujourd'hui être collectées : données de terrain, données issues des réseaux sociaux en ligne, données capturées par des périphériques mobiles, données individuelles induites ou calculées à partir d'autres données et même des données de projection simulées par des algorithmes. (ii) L'amélioration des méthodes d'analyse est également un facteur qui explique ce succès, puisque de nombreuses méthodes ont été proposées pour tenir compte de cette hétérogénéité des données et rechercher des schémas de connaissances de plus en plus adaptés au contexte. (iii) Enfin, la diversité des problèmes abordés et la précision des résultats sont sans doute l'un des aspects les plus marquants du domaine. En effet, des problèmes très variés peuvent aujourd'hui être abordés par l'analyse des données : impact de phénomènes de diffusion (Barrat et al., 2014), prédiction d'élections (Chung et Mustafaraj, 2011), détection d'événements (Sakaki et al., 2010), etc.

Dans ce travail, nous menons un travail d'analyse de données issues de la conférence “*Extraction et Gestion des Connaissances*” (EGC). Ces données concernent les articles acceptés à la conférence EGC depuis 2004 ainsi que les messages publiés sur le réseau social Twitter. Notre objectif est de mettre en évidence, depuis 2004, les grandes évolutions de la conférence