

Prédiction de la qualité dans les plateformes collaboratives : une approche générique par les graphes hétérogènes

Baptiste de La Robertie*, Yoann Pitarch*, Olivier Teste*

*Université de Toulouse, IRIT UMR5505
prenom.nom@irit.fr

Résumé. La qualité des contenus sur les plateformes collaboratives est très hétérogène. Dans la littérature scientifique, les algorithmes d'analyse structurelle appliqués à la tâche de détection de contenu de qualité reposent généralement sur des graphes définis à partir d'un seul type de nœuds et de relations. Pourtant les graphes sur lesquels reposent ces récentes plateformes présentent de nombreuses sémantiques de nœuds et relations différentes, e.g., producteurs/consommateurs, questions/réponses, etc. Ces solutions souffrent d'un manque de généralité et ne peuvent s'adapter facilement à l'évolution des plateformes. Nous proposons une modélisation générique de ces plateformes par les graphes hétérogènes pouvant intégrer automatiquement de nouvelles sémantiques de nœuds et de relations. Un algorithme de prédiction de qualité des contenus reposant sur ce modèle est proposé. Nous montrons qu'il généralise plusieurs travaux de la littérature. Enfin, en intégrant certaines relations inter-utilisateurs, nous montrons que notre solution, évaluée sur Wikipedia et Stack Exchange, améliore la tâche de détection de contenu de qualité.

1 Introduction

Le web voit apparaître de nombreuses plateformes collaboratives sur lesquelles l'internaute est sollicité pour enrichir différents types de contenus. Comme de nombreuses sociétés¹, les Wikis et sites de questions/réponses (Q&R) constituent les exemples les plus connus de portails collaboratifs tirant profits des collaborations à grande échelle. Le nouveau statut de producteur de l'utilisateur le place au cœur de la plateforme et savoir identifier influenceurs et productions de qualité est rapidement devenu d'un intérêt majeur. L'évolution de ces plateformes est la première motivation de notre travail. Par exemple, certains sites de Q&R comme *Yahoo! Answers*² ou *Stack Exchange*³ ont vu leur modèle évoluer en intégrant successivement de nouveaux types de contenus et donc de nouvelles relations entre les entités du modèle. Les solutions qui n'ont pas anticipé cette évolution ne peuvent tirer profit des nouvelles informations. Il est donc nécessaire de développer des formalisations génériques capables d'une part, d'anticiper la prise en compte de nouvelles sémantiques de nœuds et de relations, c'est à dire pouvant

1. <https://www.forrester.com/The+Collaborative+Economy+Will+Drive+Business+Innovation+And+Growth/fulltext/-/E-res113683>

2. <https://fr.answers.yahoo.com/>

3. <http://french.stackexchange.com/>