

Recommandations et prédictions de préférences basées sur la combinaison de données sémantiques et de folksonomie

Pierre-René Lhérisson*,** Fabrice Muhlenbach* et Pierre Maret*

*Univ. Lyon, UJM-Saint-Etienne, CNRS,
Laboratoire Hubert Curien UMR 5516, F-42023 Saint Etienne, France
{pr.lherisson / fabrice.muhlenbach / pierre.maret}@univ-st-etienne.fr

**1D Lab, 5 rue Javelin Pagnon, F-42000 Saint Etienne, France
pierrere.lherisson@1d-lab.eu

Résumé. Dans les systèmes de recommandation, l'approche du filtrage sur le contenu est revenue en force face à celle du filtrage collaboratif grâce à l'arrivée du paradigme de l'apprentissage profond et des techniques de *word embedding*. Dans cette même veine, l'avènement des folksonomies et du *web* sémantique a apporté une meilleure compréhension des profils des utilisateurs et des caractéristiques des articles à recommander. Dans cet article, nous nous intéressons au domaine musical et nous introduisons un nouveau calcul de mesure de préférence intégrée dans un système de recommandations basées sur le contenu. En testant notre approche sur le jeu de données *Last.fm*, nous montrons que l'utilisation de termes issus d'une folksonomie associés à des informations issues du *web* sémantique permet d'améliorer le processus de recommandation musicale.

1 Introduction

Dans les plates-formes d'e-commerce, il est nécessaire de filtrer la grande masse d'informations disponibles pour proposer automatiquement à chaque consommateur les articles qui sont susceptibles de l'intéresser. Les systèmes de recommandation, qui permettent d'automatiser ce processus (Ricci et al., 2015), procèdent le plus souvent soit suivant le filtrage collaboratif, où les recommandations se font à partir des évaluations déjà effectuées par d'autres utilisateurs sur des articles, soit suivant une approche basée sur le contenu (Ricci et al., 2011) qui utilise les descriptions des articles et des utilisateurs pour créer des profils d'articles et des profils d'utilisateurs.

L'approche de recommandation basée sur le contenu connaît un regain d'intérêt aujourd'hui avec l'arrivée de nouvelles technologies. Néanmoins, pour qu'une approche telle que le *word embedding* soit efficace, il faut que le contenu soit disponible en très grosses quantités sur les articles à recommander. Ceci n'est pas toujours possible suivant les domaines d'application. À titre d'exemple, considérons un système de recommandation associé à un site de *streaming* diffusant de la musique en ligne. Dans cette situation, la masse de contenu textuel exploitable sera importante si les artistes musicaux proposés sont célèbres. Cela ne sera plus vrai dans le cas où le site propose dans son catalogue des artistes musicaux issus de labels indépendants pour lesquels les informations textuelles sont faibles et peu structurées.