Une approche sémantique hybride pour la recommandation des articles d'actualité à large échelle

Hemza Fichel,
Mohamed Ramzi Haddad,
Hajer Baazaoui Zghal

Laboratoire Riadi , École Nationale des Sciences de l'Informatique,
Université de la Manouba, La Manouba 2010, Tunisie
hemza.fichel@ensi-uma.tn
haddad.medramzi@gmail.com
hajer.baazaouizghal@riadi.rnu.tn

Résumé. Les portails d'actualités en ligne produisent un flux d'information ayant un volume et une vélocité importants. Dans ce contexte, il devient plus difficile de proposer en temps réel des recommandations dynamiques adaptées aux intérêts de chaque utilisateur. Dans cet article, nous présentons une approche hybride pour la recommandation des articles d'actualité reposant sur l'analyse sémantique du contenu disponible. L'approche est basée sur l'hybridation de plusieurs approches personnalisées et non personnalisées pour remédier au problème de démarrage à froid. L'expérimentation de notre approche dans un environnement à large échelle et à fortes contraintes temps réel dans le cadre du challenge NEWSREEL a permis d'évaluer la qualité de ses recommandations et de confirmer l'apport de la sémantique dans le processus de recommandation.

1 Introduction

C'est à travers des messages du type « La rédaction vous conseille », « Lire aussi », « Sur le même sujet », « Vous pourriez également être intéressé par », que les portails d'actualités en ligne proposent aux utilisateurs des sélections personnalisées de contenus pouvant les intéresser afin de les assister dans leur exploration des articles d'actualités. Dans ce domaine d'application, les systèmes de recommandation font face à un flux abondant et continu d'actualités qu'il faut collecter, analyser et mettre à disposition des utilisateurs selon leurs préférences et dans les plus brefs délais. Par ailleurs, la recommandation des articles d'actualité diffère des domaines d'application traditionnels des systèmes de recommandation qui se caractérisent par une parcimonie de données reflétant les comportements des utilisateurs à cause de l'inaccessibilité de certains produits (produits non disponibles ou ayant un coût d'acquisition élevé). En effet, les articles d'actualité étant gratuits et accessibles à tous les utilisateurs, les approches de recommandation à adopter doivent être capables de prendre en considération un flux abondant d'interactions entre les utilisateurs et le contenu. À titre d'exemple, sur la plateforme de