

INTISARI

Model Rekomendasi Artikel Ilmiah Menggunakan *Modified* PageRank Berbasis Fitur Komunitas Multi-topik pada Academic Citation Network

oleh
Agung Hadhiatma
18/435399/SPA/00637

Digital library artikel ilmiah yang direpresentasikan dalam Academic Citation Network mempunyai nilai potensi informasi yang tinggi karena mengandung sekaligus informasi teks dan *graph*. Salah satu penelitian pada Academic Citation Network yang berkembang saat ini adalah *academic recommendation*. Metode *personalized* Pagerank yang merupakan pendekatan Graph-Based Filtering (GF) saat ini sudah cukup efektif untuk memeringkat artikel ilmiah yang penting dan berpengaruh dengan memperhatikan konsep *centrality*. Namun metode ini masih menghasilkan pemeringkatan yang bias karena aspek kemiripan topik, multi-topik dan relasi topik.

Penelitian ini mengembangkan metode *personalized* PageRank dengan tidak hanya mempertimbangkan kemiripan teks dan bobot pengaruh artikel ilmiah saja, tetapi juga memperhatikan aspek multi-topik dan relasi antar topik. Rekomendasi artikel dengan aspek relasi topik merupakan artikel yang berbeda topik dengan kueri namun masih relevan. Selain itu usulan model juga dapat melakukan pembatasan ruang pencarian kandidat rekomendasi untuk meningkatkan efektivitas pencarian. Model yang dikembangkan meliputi ekstraksi fitur komunitas multi-topik pada Academic Citation Network, seleksi komunitas kandidat artikel ilmiah, dan pemeringkatan kandidat artikel ilmiah menggunakan *modified* PageRank. *Modified* Pagerank adalah metode *personalized* PageRank yang dikembangkan untuk bekerja berdasarkan fitur komunitas multi-topik.

Percobaan menunjukkan bahwa ekstraksi komunitas multi-topik dapat digunakan secara efektif untuk seleksi komunitas dan pemeringkatan kandidat rekomendasi artikel ilmiah. Komunitas terseleksi telah dapat mengurangi ruang pencarian sebesar 82% dengan akurasi lebih 90%. Dengan membatasi ruang pencarian dan memperhatikan aspek kemiripan topik, relasi topik yang berbeda dan *impact factor*, hasil percobaan menunjukkan bahwa 3 usulan metode rekomendasi mempunyai tingkat akurasi lebih tinggi dibandingkan dengan 6 metode lain sebagai *baselines* sebesar untuk Doc2Vect(73%), TF-IDF Cosine Similarity (21%), Bert Similarity(17%), PPR Restart (66%), Edge Weight PPR(18%), dan Query-PPR(15%). Selain itu usulan model tersebut mempunyai akurasi yang lebih konsisten terhadap berbagai grup kueri pada bidang berbeda. Usulan model mempunyai sensitivitas yang lebih kecil terhadap perubahan volume dataset dan bekerja lebih efektif dibandingkan dengan *baselines* pada *large dataset*.

Kata kunci: rekomendasi artikel, *modified* PageRank, komunitas multi-topik, *impact factor*, relasi topik.

ABSTRACT

A Scientific Paper Recommendation Framework Using Modified PageRank Based on Multi-topic Community Features on An Academic Citation Network

by

Agung Hadhiatma
18/435399/SPA/00637

The digital library of scientific research articles represented in an academic citation network has a high potential information value as it contains text and graph information. One of the studies on the Academic Citation Network is an academic recommendation. Some personalized PageRank methods as a graph-based Filtering (GF) approach, are effective enough to rank essential and influential scientific articles by considering centrality measures. However, these methods still produce biased rankings due to multi-topic, similar, and related topics.

This research develops a personalized PageRank method by not only considering the similarity of texts and the impact factor of scientific articles but also paying attention to multi-topic aspects and related topics. The related topics of recommended papers can be papers with different topics to queries but still relevant. In addition, the proposed model can also limit the search space of recommended paper candidates to enhance the recommendation's accuracy. To accomplish these challenges, we propose a framework of scientific citation recommendation methods, including extracting multi-topic community features, retrieving recommended paper candidates of multi-topic communities, and ranking the recommended paper candidates with modified PageRank.

Experiments show that multi-topic community extraction can be used effectively for community selection and ranking of scientific article recommendation candidates. Selected communities have been able to reduce search space by 82% with over 90% accuracy. By limiting the search space and paying attention to similar topics, related topics, and high-impact factors, the experimental results show that the three proposed recommendation methods have a higher level of accuracy compared to the six other methods as baselines such as Doc2Vect(73%), TF-IDF Cosine (21%), Bert Similarity(17%), PPR Restart (66%), Edge Weight PPR(18%), and Query-PPR(15%). In addition, the proposed models have a more consistent performance against various query groups in different fields. Moreover, the proposed models are less sensitive to dataset volume changes and work more effectively than baselines on large datasets.

Keywords: recommended articles, modified PageRank, multi-topic communities, impact factor, related topic.