

Learning-to-Rank in research paper C F recommendation: Leveraging irrelevant papers

Anas Alzoghbi

Victor A. Arrascue Ayala

Peter M. Fischer

Georg Lausen

Department of Computer Science University of Freiburg Germany
{alzoghba arrascue peter.fischer lausen}@informatik.uni-freiburg.de

BSTR CT

Suggesting relevant literature to researchers has become an active area of study, typically relying on content-based filtering (CBF) over the rich textual features available. Given the high dimensionality and the sparsity of the training samples inherent to this domain, the focus has so far been on heuristic-based methods. In this paper, we argue for the model-based approach and propose a learning-to-rank method that leverages publicly available publications' meta-data to produce an effective prediction model. The proposed method is systematically evaluated on a scholarly paper recommendation dataset and compared against state-of-the-art model-based approaches as well as current, domain-specific heuristic methods. The results show that our approach clearly outperforms state-of-the-art research paper recommendations utilizing only publicly available meta-data.

CCS Concepts

•Information systems → Learning to rank; Recommender systems;

Keywords

Research paper recommendation; Learning-to-Rank; Content-based Recommendation; Model-based user profile

1. INTRODUCTION

Scholars and researchers are confronted with an overwhelming number of newly published research papers in their domain of expertise. Although advantageous in restricting the domain, keyword-based search tools typically available in digital libraries offer a limited help to researchers in locating the relevant content. As a result, researchers need to manually search within unspecific search results to identify paper(s) of interest. This is the situation where recommendation systems have great potential, and indeed plenty

of works adopted different techniques to tackle this problem. A recent extensive survey in this domain [3] identified content-based filtering (CBF) as the predominant approach for research paper recommendation because of the rich textual features available. For learning user profile, almost exclusively the focus was on relevance feedback approaches, building on the assumption that papers appearing in user's preference list have an equal (or a presumed extent) share in the underlying user taste. Thus, user profiles are constructed as aggregation of relevant papers' keywords. Based on the classification suggested by Adomavicius et al. in [1], these approaches are referred to as heuristic-based. In contrast, model-based approaches depend on a learning method to fit the underlying user model (profile). This enables constructing a better modeling of researcher-keywords relation in user profiles. But they require a large body of training data which is not intuitively available in this domain. As a result, little work on applying model-based approaches exists for this problem.

In this paper, we employ pairwise learning-to-rank [4] as a model-based technique for learning user profile. We incorporate both relevant and irrelevant "peer" papers -papers published in relevant papers' conferences- to formulate pairwise preferences and enrich the training set. Our main contributions include:

- We investigate and customize learning-to-rank for CBF research paper recommendation.
- We incorporate only a small set of data, restricted to publicly available metadata of papers. This makes our approach suitable for a much larger domain than previous approaches which require papers' full-text.
- We perform an initial, yet systematic study on a real-world dataset in which we show that our approach clearly outperforms existing heuristic- and model-based algorithms.

The rest of this paper is organized as following: the second section provides an overview of existing related work. In section 3 we present our approach and in section 4 we demonstrate experimental setup and results. Finally, we conclude in section 5 by summarizing our findings and situate this work within our future plan.

2. RELATED WORK

A rich amount of related work tackled the problem of research paper recommendation. collaborative filtering (CF) approaches [8, 13, 14] showed a successful application of