

A Framework for Bibliographic Recommendation System based on Heterogeneous Retrieval Model*

Poonam Anthony and Plaban Kumar Bhowmick

Center for Educational Technology, Indian Institute of Technology Kharagpur
anthonypoonam@gmail.com plaban@cet.iitkgp.ac.in

Abstract. In this paper, we propose an architectural framework for recommending heterogeneous resources in a digital library. We present an outline of our proposed recommendation framework, and discuss briefly its performance over SpringerNature SciGraph¹ dataset.

Keywords: Recommendation System · Heterogeneous Information Networks · meta-path based similarity

1 Introduction and Brief Overview

With the escalating popularity of online learning, one of the popular features incorporated in most contemporary digital libraries, is providing recommendations. The recommended resources are generally of the same type as the resource being viewed. However, a user reading an article may be interested in other authors or journals related to it. Therefore, in our work we lay emphasis on generating different types of recommendation items. The heterogeneity in the resource and property types of the bibliographic data of a digital library allows us to model it as a Heterogeneous Information Network (HIN) network. Thus, we map our recommendation task to the problem of finding related entities in a HIN. We leverage on the paths connecting a pair of entities, formally termed as meta-paths [3], for determining the relatedness between them. Hence the objectives of our work are formulated as - enumerating the meta-paths in a given bibliographic HIN, and generating recommendations by exploiting these meta-paths to find related resources.

Given the afore-stated objectives, our proposed recommendation framework is comprised of two broad modules, namely - the meta-path generation module, and the recommendation engine. The meta-path generation module takes as input a set of example pairs of entities, and the entire HIN, and outputs a set of meta-paths along-with their respective weights. Our approach for meta-path generation is derived from the GreedyTree strategy proposed in [1]. We adopt a bi-directional meta-path search algorithm by incorporating HeteSim [2] similarity measure. For a given item and meta-paths relevant to that item generated

* Supported by National Digital Library of India and SUR Grant of IBM Research

¹ <https://www.springernature.com/in/researchers/scigraph>

by the previous module, the recommendation engine finds those resources from the HIN which satisfy the meta-paths and produces a list of recommendations ranked by their HeteSim similarity score, to be presented to the user.

2 Preliminary Results and Discussion

In the present study, we have used a portion of the SpringerNature SciGraph (SN SciGraph) dataset as a sample Bibliographic HIN. We form the input set of entity pairs by selecting resource pairs of two types, namely, *(Article, Grant)* and *(Article, Subject)*, comprised of 10 positive and 10 negative example pairs each. The resulting meta-paths obtained, along-with their weights are as follows:
(Article, Grant) - $\langle \text{sg}^2:\text{hasContributingOrganization}, \hat{\text{sg}}:\text{hasRecipientOrganization} \rangle$ (2.30647), $\langle \text{sg}:\text{hasContribution}, \text{sg}:\text{hasAffiliation}, \text{sg}:\text{hasOrganization}, \hat{\text{sg}}:\text{hasRecipientOrganization} \rangle$ (0.044108)
(Article, Subject) - $\langle \text{sg}:\text{hasCategory} \rangle$ (0.121286),
 $\langle \text{sg}:\text{hasCategory}, \text{skos}^3:\text{broader}, \text{skos}:\text{narrower} \rangle$ (0.094166)

For measuring the performance of the framework, we computed the Mean Average Precision (MAP) and Mean Reciprocal Rank (MRR) scores for two recommendation types, namely *Grant* and *Subject*, for a test set of 20 articles. For *Grant*, we obtained a MAP score of 0.69 and MRR of 0.85, while for *Subject* these were 0.75 and 0.90 respectively. The higher MRR indicates that relevant items were ranked higher. However, we noted that some false positives are produced in a mid-to-lower range of ranks. This may be attributed to the generic nature of the meta-paths in a Bibliographic network like SN SciGraph, resulting in resource pairs that may not always be related. Hence, augmenting the meta-path based score with the contextual information of the nodes may produce better results from the perspective of recommendation. We intend to examine this aspect in our future studies. Future scope also includes generating meta-path consisting of node classes. Further, we also intend to evaluate our framework over other data-sets like Wikidata⁴.

References

1. Meng, C., Cheng, R., Maniu, S., Senellart, P., Zhang, W.: Discovering meta-paths in large heterogeneous information networks. In: Proceedings of the 24th International Conference on World Wide Web. pp. 754–764. International World Wide Web Conferences Steering Committee (2015)
2. Shi, C., Kong, X., Huang, Y., Philip, S.Y., Wu, B.: Hetesim: A general framework for relevance measure in heterogeneous networks. *IEEE Transactions on Knowledge and Data Engineering* **26**(10), 2479–2492 (2014)
3. Shi, C., Li, Y., Zhang, J., Sun, Y., Philip, S.Y.: A survey of heterogeneous information network analysis. *IEEE Transactions on Knowledge and Data Engineering* **29**(1), 17–37 (2017)

² sg: <http://scigraph.springernature.com/ontologies/core/>

³ skos: <http://www.w3.org/2004/02/skos/core#>

⁴ https://www.wikidata.org/wiki/Wikidata:Main_Page