TheAdvisor: A Webservice for Academic Recommendation

Ozil Küçüktunç 1,2, Erik Saule 1, Kamer Kaya 1, and Ümit V. Çatalyürek 1,3

1Dept of Biomedical Informatics, 2Dept of Computer Science and Engineering, 3Dept of Electrical and Computer Engineering, The Ohio State University

kucuktunc.1@osu.edu, {esaule,kamer,umit}@bmi.osu.edu

Introduction

Motivation: Existing tools for literature search (i.e., Citeseer, Google Scholar, etc.) are suitable for a first-level bibliographic search. Personalized search based on the interests of the researcher will improve the process.

Solution: We present a web service called theadvisor which helps the users to build a strong bibliography by extending the document set obtained after a first-level search. The users of the service:

1. receive personalized paper, venue, and potential reviewer recommendations by submitting their bibliographic file (BibTeX, RIS, EndNote XML),
2. may specify the interest in recent developments or older well-known documents in the literature,
3. may guide suggestions by giving relevance feedback,
4. visualize the related part of the citation graph.

Overview of the framework

The framework has four main components:

1. Paper mapper: finds the entities in our citation graph which correspond to the input from the user.
2. Recommendation engine recommends a diversified set of papers/citations, venues, and experts to the user.
3. Visualization uses graph drawing techniques on the recommended papers to visualize their relationships.
4. Relevance feedback obtains the comments on the recommendations and refines the search accordingly.

The service is available at:

http://theadvisor.osu.edu

Acknowledgments

This work was partially supported by the U.S. Department of Energy Sandia Center DE-FG02-06ER25756 and NSF grants CNS-0413069, OCI-0904809 and OCI-0904802.

Recommendation Engine

Direction-aware Random Walk with Restart (DaRW): Personalized PageRank variant, distribute probabilities towards citations and references according to the x parameter, find x = Pr(x) where

\[ P(u, v) = \begin{cases} (1 - \delta^+) + \frac{\delta^-(v)}{\left|Q\right|} & \text{if } (u, v) \in E \\ (1 - \delta^-) + \frac{\delta^+(v)}{\left|Q\right|} & \text{if } (v, u) \in E \\ (1 - \delta^) & \text{otherwise}, \end{cases} \]

Parameter test: The average shortest distance from seed papers (left) and publication year (right) of top-10 recommendations by DaRW based on d and e.

Accuracy: Link prediction experiments on three scenar-ions, mean average precision (MAP@50) is reported

Results: DaRW leads to a higher accuracy when the query is targeted. Their dominance is statistically significant.

Relevance Feedback (RF)

Users of theadvisor are given the option of providing explicit relevance feedback to the recommended papers. The feedback can be either positive or negative, making a recommendation relevant or irrelevant for the query.

Experiments: performance profiles (left) and sample of #papers one has to go through expressed as a percentage of #papers without using any feedback (right)

Efficiency Improvements

Each iteration of DaRW contains a sparse-matrix dense-vector multiplication (SpMV). With a straightforward implementation of SpMV, the ranking takes 3.5 sec.

Observation: Nonzero pattern of the citation matrix is highly irregular and the computation suffers from this irregularity due to the high number of cache misses.

Preprocessing: Reduce the number of cache misses by partitioning and ordering heuristics such as Reverse Cuthill-McKee (RCM), Approximate Minimum Degree (AMD), SlashBurn (SB).

Implementation improvements:

- CRS-Full: straightforward implementation using citation matrix in compressed row storage (CRS).
- COO-Half: use only the reference edges, scale the vector entry p_{u,v}(e) with d^{-} to compute nonzero values.
- COO-Half: improves plus store the row/column coordinates of the nonzero explicitly.

Hybrid: CRS-Full for first iterations, then COO-Half.

Experiments: Execution times in sec for DaRW.

Result Diversification

Parameterized relaxed local maxima (xRLM): Incrementally compute local maxima within top-k results based on DaRW ranks until \(|K| = k\), and remove the selected vertices from the subgraph for the next step.

Visualisation and API

The Advisor: Academic Recommendation on Graphs: Pitfalls, Measures, and Applications, in \\
IEEE/ACM Int’l Conf on Advances in Social Networks Analysis and Mining, August 2013.

References

[1] O. Küçüktunç, E. Saule, K. Kaya, and Ü. V. Çatalyürek. Fast and scalable recommendation on bibliographic networks with approximate re-
dundancy partitioning. Social Network Analysis and Mining 2013.

tions-aware in citation recommendation. In Int’l Workshop on Ranking in Databases in conj. VLDB’12, Aug 2012.

