

Biomedical Data Recommendation Using Machine Learning and Crowdsourcing

By

Xiaoqian Jiang, Ph.D.

**Division of Biomedical Informatics
University of California at San Diego**

Abstract:

Recommendation systems have witnessed a lot of successes in movie suggestion, online shopping, content searching, etc. An advanced data recommendation system can promote scientific discovery and improve the healthcare quality.

This project aims at making data recommendation based on content similarity, user background, and the context. We developed a novel framework called DataRank to support personalized presentation of biomedical datasets to researchers. DataRank takes the bipartite citation graph (between datasets and papers from PubMed Central) to enrich the features associated with the dataset (e.g., by aggregating MeSH terms from papers in the bipartite citation graph). For each search query, DataRank first maps the "free text query" to a "MeSH query" and yields an initial ranking of datasets for the MeSH query using a Bayesian approach, where the likelihood is proportional to the Jaccard index and the prior is proportional to number of citations of the dataset. This is called the offline model as it does not depend on users' preference. We further extended the DataRank with an online algorithm by incorporating user-feedbacks regarding the relevance of ranking.

Thursday, April 30, 2015

12:00 – 1:00 p.m.

Rollins School of Public Health

Claudia Nance Rollins Building, Room

1000

Parking available in the Michael Street Visitors Parking Deck
(behind the Wayne Rollins Research Bldg., 2nd deck entrance)
or at the 1525 Clifton Road Visitor pay parking deck

Questions: rwaggon@emory.edu

