Online Learning to Rank for Information Retrieval: Project Proposal

TINGDAN LUO (TL3XD), XUEYANG HAN (XH9CZ), and XIAO LIU (LX3Y), University of Virginia

CCS Concepts: • Information systems — Learning to rank;

Additional Key Words and Phrases: Information retrieval; Learning to rank

ACM Reference format:

1 INTRODUCTION

There are more than 45 billion web pages on the Internet, and this number is increasing dramatically every year. Ranking of query results has always being a central problem in information retrieval (IR). Learning to rank is the application of machine learning in the construction of ranking models for information retrieval systems. It aims to rank documents by relevance, placing the most relevant documents at the top. There are two types of learning to rank model – offline learning to rank and online learning to rank. Offline learning to rank depends on a set of human labeled data, whereas online learning to rank depends on human interactive feedback in real time.

The purpose of this project is to build a ranking model to rank documents by relevance when given a query. In this project we will implement and compare offline learning and online learning, as well as offline training and online learning schemes. We will also implement state-of-the-art online learning to rank algorithms. After implementing and comparing different models and algorithms, we will come up with new exploration strategy and analyze and implement it.

2 BACKGROUND

To begin this project, it is necessary to have a thorough understanding of learning to rank, which is an application of machine learning to solve the problem of ranking for information retrieval systems. In general, a learning to rank method learns the weights of a function that maps a document-query pair described by a feature vector to a value that is used to rank documents for a given query.

Both offline learning and online learning has its own advantages and disadvantages. Although it is easy to conduct experiment for offline learning and it has absolute feedback. However, it is expensive to create a complete dataset with manually labeled relevance tags for smaller search engines and impractical in personalized search. Online learning to rank improves this problem based on users' feedback in real time, however, online learning is more difficult because it is necessary to balance exploration and exploitation.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2017 ACM. XXXX-XXXX/2017/3-ART1 $15.00
DOI: 10.1145/nmnnnn.nmnnnn

1:2  •  Tingdan Luo (tl3xd), Xueyang Han (xh9cz), and Xiao Liu (lx3y)

3 METHODOLOGY

The implementation of our project contains 4 important steps.

1. Implement and compare offline learning, online learning, offline training with online learning schemes.
2. Implement state-of-the-art online learning to rank algorithms within the following three papers.
   - Constructing Reliable Gradient Exploration for Online Learning to Rank. [4]
   - Interactively Optimizing Information Retrieval Systems as a Dueling Bandits Problem. [3]
   - Multileave Gradient Descent for Fast Online Learning to Rank. [2]
3. Propose our idea.
4. Theoretical analysis of our idea.

4 EVALUATION

Precision, recall, nDCG, DCG, Map and ERR will be calculated to evaluate offline learning to rank. Then, we will use Lerot [1] to evaluate our online ranking model. Lerot is an online learning to rank framework, and could offer a solution for evaluating and experimenting with online learning to rank algorithms in living labs and simulations.

4.1 Minimal Success

Successfully implement offline learning, online learning, offline training with online learning schemes and state-of-the-art online learning.

4.2 Medium Success

Propose and implement our own idea to build a ranking model.

4.3 Good Success

Prove our own ranking model has excellent performance.

5 SUMMARY

After finishing this project, we will have a deep understanding of ranking in information retrieval. We will know the implementation, advantages and disadvantages of several ranking models. We will also know what are the metrics for ranking evaluation and how to design and implement a learning to rank algorithm.

There are many applications for online learning to rank, such as optimization of search engine, personalized advertisement and personalized recommendation. We can apply what we learn in this project into real world applications in the future.

REFERENCES