

# **Learning to Rank for Information Retrieval and Natural Language Processing**

**Second Edition**

# Synthesis Lectures on Human Language Technologies

Editor

**Graeme Hirst**, *University of Toronto*

**Synthesis Lectures on Human Language Technologies** is edited by Graeme Hirst of the University of Toronto. The series consists of 50- to 150-page monographs on topics relating to natural language processing, computational linguistics, information retrieval, and spoken language understanding. Emphasis is on important new techniques, on new applications, and on topics that combine two or more HLT subfields.

Learning to Rank for Information Retrieval and Natural Language Processing, Second Edition

Hang Li

2014

Ontology-Based Interpretation of Natural Language

Philipp Cimiano, Christina Unger, and John McCrae

2014

Automated Grammatical Error Detection for Language Learners, Second Edition

Claudia Leacock, Martin Chodorow, Michael Gamon, and Joel Tetreault

2014

Web Corpus Construction

Roland Schäfer and Felix Bildhauer

2013

Recognizing Textual Entailment: Models and Applications

Ido Dagan, Dan Roth, Mark Sammons, and Fabio Massimo Zanzotto

2013

Linguistic Fundamentals for Natural Language Processing: 100 Essentials from Morphology and Syntax

Emily M. Bender

2013

Semi-Supervised Learning and Domain Adaptation in Natural Language Processing

Anders Søgaard

2013

Semantic Relations Between Nominals

Vivi Nastase, Preslav Nakov, Diarmuid Ó Séaghdha, and Stan Szpakowicz

2013

Computational Modeling of Narrative

Inderjeet Mani

2012

Natural Language Processing for Historical Texts

Michael Piotrowski

2012

Sentiment Analysis and Opinion Mining

Bing Liu

2012

Discourse Processing

Manfred Stede

2011

Bitext Alignment

Jörg Tiedemann

2011

Linguistic Structure Prediction

Noah A. Smith

2011

Learning to Rank for Information Retrieval and Natural Language Processing

Hang Li

2011

Computational Modeling of Human Language Acquisition

Afra Alishahi

2010

Introduction to Arabic Natural Language Processing

Nizar Y. Habash

2010

Cross-Language Information Retrieval

Jian-Yun Nie

2010

Automated Grammatical Error Detection for Language Learners  
Claudia Leacock, Martin Chodorow, Michael Gamon, and Joel Tetreault  
2010

Data-Intensive Text Processing with MapReduce  
Jimmy Lin and Chris Dyer  
2010

Semantic Role Labeling  
Martha Palmer, Daniel Gildea, and Nianwen Xue  
2010

Spoken Dialogue Systems  
Kristiina Jokinen and Michael McTear  
2009

Introduction to Chinese Natural Language Processing  
Kam-Fai Wong, Wenjie Li, Ruifeng Xu, and Zheng-sheng Zhang  
2009

Introduction to Linguistic Annotation and Text Analytics  
Graham Wilcock  
2009

Dependency Parsing  
Sandra Kübler, Ryan McDonald, and Joakim Nivre  
2009

Statistical Language Models for Information Retrieval  
ChengXiang Zhai  
2008

© Springer Nature Switzerland AG 2022

Reprint of original edition © Morgan & Claypool 2015

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means—electronic, mechanical, photocopy, recording, or any other except for brief quotations in printed reviews, without the prior permission of the publisher.

Learning to Rank for Information Retrieval and Natural Language Processing, Second Edition

Hang Li

ISBN: 978-3-031-01027-9      paperback

ISBN: 978-3-031-02155-8      ebook

DOI 10.1007/S00607ED2V01Y201410HLT026

A Publication in the Springer series

*SYNTHESIS LECTURES ON HUMAN LANGUAGE TECHNOLOGIES*

Lecture #26

Series Editor: Graeme Hirst, *University of Toronto*

Series ISSN

Synthesis Lectures on Human Language Technologies

Print 1947-4040    Electronic 1947-4059

# Learning to Rank for Information Retrieval and Natural Language Processing

Second Edition

Hang Li  
Huawei Technologies

*SYNTHESIS LECTURES ON HUMAN LANGUAGE TECHNOLOGIES #26*

## ABSTRACT

Learning to rank refers to machine learning techniques for training a model in a ranking task. Learning to rank is useful for many applications in information retrieval, natural language processing, and data mining. Intensive studies have been conducted on its problems recently, and significant progress has been made. This lecture gives an introduction to the area including the fundamental problems, major approaches, theories, applications, and future work.

The author begins by showing that various ranking problems in information retrieval and natural language processing can be formalized as two basic ranking tasks, namely ranking creation (or simply ranking) and ranking aggregation. In ranking creation, given a request, one wants to generate a ranking list of offerings based on the features derived from the request and the offerings. In ranking aggregation, given a request, as well as a number of ranking lists of offerings, one wants to generate a new ranking list of the offerings.

Ranking creation (or ranking) is the major problem in learning to rank. It is usually formalized as a supervised learning task. The author gives detailed explanations on learning for ranking creation and ranking aggregation, including training and testing, evaluation, feature creation, and major approaches. Many methods have been proposed for ranking creation. The methods can be categorized as the pointwise, pairwise, and listwise approaches according to the loss functions they employ. They can also be categorized according to the techniques they employ, such as the SVM based, Boosting based, and Neural Network based approaches.

The author also introduces some popular learning to rank methods in details. These include: PRank, OC SVM, McRank, Ranking SVM, IR SVM, GBRank, RankNet, ListNet & ListMLE, AdaRank, SVM MAP, SoftRank, LambdaRank, LambdaMART, Borda Count, Markov Chain, and CRanking.

The author explains several example applications of learning to rank including web search, collaborative filtering, definition search, keyphrase extraction, query dependent summarization, and re-ranking in machine translation.

A formulation of learning for ranking creation is given in the statistical learning framework. Ongoing and future research directions for learning to rank are also discussed.

## KEYWORDS

learning to rank, ranking, ranking creation, ranking aggregation, information retrieval, natural language processing, supervised learning, web search, collaborative filtering, machine translation.

# Contents

|   |             |
|---|-------------|
| <b>Preface</b> .....                            | <b>xiii</b> |
| <b>1 Learning to Rank</b> .....                 | <b>1</b>    |
| 1.1 Ranking .....                               | 1           |
| 1.2 Learning to Rank .....                      | 4           |
| 1.3 Ranking Creation .....                      | 5           |
| 1.4 Ranking Aggregation .....                   | 6           |
| 1.5 Learning for Ranking Creation .....         | 7           |
| 1.6 Learning for Ranking Aggregation .....      | 8           |
| <b>2 Learning for Ranking Creation</b> .....    | <b>11</b>   |
| 2.1 Document Retrieval as Example .....         | 11          |
| 2.2 Learning Task .....                         | 12          |
| 2.2.1 Training and Testing .....                | 12          |
| 2.2.2 Training Data Creation .....              | 15          |
| 2.2.3 Feature Construction .....                | 16          |
| 2.2.4 Evaluation .....                          | 17          |
| 2.2.5 Relations with Other Learning Tasks ..... | 21          |
| 2.3 Learning Approaches .....                   | 22          |
| 2.3.1 Pointwise Approach .....                  | 22          |
| 2.3.2 Pairwise Approach .....                   | 25          |
| 2.3.3 Listwise Approach .....                   | 27          |
| 2.3.4 Evaluation Results .....                  | 29          |
| <b>3 Learning for Ranking Aggregation</b> ..... | <b>33</b>   |
| 3.1 Learning Task .....                         | 33          |
| 3.2 Learning Methods .....                      | 35          |
| <b>4 Methods of Learning to Rank</b> .....      | <b>37</b>   |
| 4.1 PRank .....                                 | 37          |
| 4.1.1 Model .....                               | 37          |
| 4.1.2 Learning Algorithm .....                  | 37          |



|        |  |    |
|--------|--|----|
| 4.2    | OC SVM .....                             | 38 |
| 4.2.1  | Model .....                              | 38 |
| 4.2.2  | Learning Algorithm .....                 | 39 |
| 4.3    | McRank .....                             | 40 |
| 4.3.1  | Model .....                              | 41 |
| 4.3.2  | Learning Algorithm .....                 | 41 |
| 4.4    | Ranking SVM .....                        | 41 |
| 4.4.1  | Linear Model as Ranking Function .....   | 41 |
| 4.4.2  | Ranking SVM Model .....                  | 43 |
| 4.4.3  | Learning Algorithm .....                 | 44 |
| 4.5    | IR SVM .....                             | 46 |
| 4.5.1  | Modified Loss Function .....             | 46 |
| 4.5.2  | Learning Algorithm .....                 | 48 |
| 4.6    | GBRank .....                             | 48 |
| 4.6.1  | Loss Function .....                      | 49 |
| 4.6.2  | Learning Algorithm .....                 | 50 |
| 4.7    | RankNet .....                            | 50 |
| 4.7.1  | Loss Function .....                      | 50 |
| 4.7.2  | Model .....                              | 52 |
| 4.7.3  | Learning Algorithm .....                 | 52 |
| 4.7.4  | Speed up of Training .....               | 54 |
| 4.8    | ListNet and ListMLE .....                | 54 |
| 4.8.1  | Plackett-Luce model .....                | 54 |
| 4.8.2  | ListNet .....                            | 56 |
| 4.8.3  | ListMLE .....                            | 58 |
| 4.9    | AdaRank .....                            | 59 |
| 4.9.1  | Loss Function .....                      | 59 |
| 4.9.2  | Learning Algorithm .....                 | 60 |
| 4.10   | SVM MAP .....                            | 60 |
| 4.10.1 | Loss Function .....                      | 60 |
| 4.10.2 | Learning Algorithms .....                | 62 |
| 4.11   | SoftRank .....                           | 64 |
| 4.11.1 | Soft NDCG .....                          | 64 |
| 4.11.2 | Approximation of Rank Distribution ..... | 65 |
| 4.11.3 | Learning Algorithm .....                 | 67 |
| 4.12   | LambdaRank .....                         | 67 |

|          |   |            |
|----------|---|------------|
| 4.12.1   | Loss Function                           | 67         |
| 4.12.2   | Learning Algorithm                      | 69         |
| 4.13     | LambdaMART                              | 70         |
| 4.13.1   | Model and Loss Function                 | 70         |
| 4.13.2   | Learning Algorithm                      | 71         |
| 4.14     | Borda Count                             | 71         |
| 4.15     | Markov Chain                            | 73         |
| 4.16     | Cranking                                | 74         |
| 4.16.1   | Model                                   | 75         |
| 4.16.2   | Learning Algorithm                      | 75         |
| 4.16.3   | Prediction                              | 76         |
| <b>5</b> | <b>Applications of Learning to Rank</b> | <b>77</b>  |
| <b>6</b> | <b>Theory of Learning to Rank</b>       | <b>81</b>  |
| 6.1      | Statistical Learning Formulation        | 81         |
| 6.2      | Loss Functions                          | 82         |
| 6.3      | Relations between Loss Functions        | 84         |
| 6.4      | Theoretical Analysis                    | 85         |
| <b>7</b> | <b>Ongoing and Future Work</b>          | <b>87</b>  |
|          | <b>Bibliography</b>                     | <b>93</b>  |
|          | <b>Author's Biography</b>               | <b>107</b> |

# Preface

This book presents a survey on learning to rank and describes methods for learning to rank in detail. The major focus of the book is *supervised learning for ranking creation*.

The book targets researchers and practitioners in information retrieval, natural language processing, machine learning, data mining, and other related fields. It assumes that the readers of the book have basic knowledge of statistics and machine learning.

Chapter 1 gives a formal definition of learning to rank. Chapter 2 describes learning for ranking creation, and Chapter 3 describes learning for ranking aggregation. Chapter 4 explains in details about state-of-the-art learning to rank methods. Chapter 5 presents applications of learning to rank. Chapter 6 introduces theory of learning to rank. Chapter 7 introduces ongoing and future research on learning to rank.

I would like to express my sincere gratitude to my former colleagues, Tie-Yan Liu, Jun Xu, Tao Qin, Yunbo Cao, and Yunhua Hu. We worked together on learning to rank. Many thanks go to our former intern students, Zhe Cao, Ming-Feng Tsai, Xiubo Geng, Yanyan Lan, Fen Xia, Ming Li, Xin Jiang, and Wei Chen, who also participated in the research. I am very grateful to Wei-Ying Ma, Hsiao-Wuen Hon, and Harry Shum for their encouragement and guidance.

I also thank our collaborators in Microsoft, Chris Burges, Stephen Robertson, Michael Taylor, John Guiver, Dmitriy Meyerzon, Victor Poznanski, Rangan Majumder, and Steven Yao, and our collaborators in other organizations, Rong Jin, Zhi-Hua Zhou, Hamed Valizadegan, Cheng Xiang Zhai, and Thorsten Joachims.

I am deeply indebted to Jun Xu and Tao Qin who provided materials for writing this book.

Many thanks also go to Chen Wang, Wei Wu, Wei Chen, Jun Xu, Xin Jiang, Shuxin Wang, and two anonymous reviewers, who read the draft of this book and made many valuable comments.

I would also like to thank Graeme Hirst and Michael Morgan. Without their support, this book would not have been published.

Hang Li  
October 4, 2014