Learning To Rank For Information Retrieval And Natural Language Processing Hang Li

Understanding Machine Learning
From Theory to Algorithms
Advances in Information Retrieval
Advances in Information Retrieval
25th Pacific-Asia Conference, PAKDD 2021, Virtual Event, May 11–14, 2021, Proceedings, Part III
Advances in Information Retrieval
Second Edition
Advances in Knowledge Discovery and Data Mining
42nd European Conference on IR Research, ECIR 2020, Lisbon, Portugal, April 14–17, 2020, Proceedings, Part II
Advances in Information Retrieval
Statistical Language Models for Information Retrieval
Proceedings
Learning to Rank for Information Retrieval
Web Information Systems and Applications
Methods, Systems, Challenges
42nd European Conference on IR Research, ECIR 2020, Lisbon, Portugal, April 14–17, 2020, Proceedings, Part I
Relevant Search
Machine-learned Ranking Algorithms for E-commerce Search and Recommendation Applications
An Introduction to Neural Information Retrieval
Advances in Information Retrieval
Case Studies and Algorithms to Get You Started
Machine Learning, Low-Rank Approximations and Reduced Order Modeling in Computational Mechanics
Learning to Rank for Information Retrieval
16th International Conference, WISA 2019, Qingdao, China, September 20-22, 2019, Proceedings
Twenty-fourth International Conference on Machine Learning
Foundations of Machine Learning, second edition
Machine Learning for Text
Direct Optimization of Ranking Measures for Learning to Rank Models
Solving Least Squares Problems
Fast and Reliable Online Learning to Rank for Information Retrieval
18th European Conference on Machine Learning, Warsaw, Poland, September 17-21, 2007, Proceedings

Machine Learning for Hackers
41st European Conference on IR Research, ECIR 2019, Cologne, Germany, April 14–18, 2019, Proceedings, Part II
6th International Conference, MDIS 2019, Sibiu, Romania, October 3–5, 2019, Revised Selected Papers
43rd European Conference on IR Research, ECIR 2021, Virtual Event, March 28 – April 1, 2021, Proceedings, Part I
Learning to Rank for Information Retrieval
Preference Learning
Automated Machine Learning
A Regression Framework for Learning to Rank in Web Information Retrieval

CHASE WERNER

Understanding Machine Learning SIAM
This primer reviews the published literature on search result diversification. In particular, it discusses the motivations for diversifying the search results for an ambiguous query and provides a formal definition of the search result diversification problem. In addition, it describes the most successful approaches in the literature for producing and evaluating diversity in multiple search domains.
From Theory to Algorithms MIT Press
This volume constitutes the refereed proceedings of the 6th International Conference on Modelling and Development of Intelligent Systems, MDIS 2019, held in Sibiu, Romania, in October 2019. The 13 revised full papers presented in the volume were carefully reviewed and selected from 31 submissions. The papers are organized in topical sections on adaptive systems; conceptual modelling; data mining; intelligent systems for decision support; machine learning.
Advances in Information Retrieval Newnes
This book constitutes the refereed proceedings of the 39th European Conference on IR Research, ECIR 2017, held in Aberdeen, UK, in April 2017. The 36 full papers and 47 poster papers presented together with 5 Abstracts, were carefully reviewed and selected from 248 submissions. Being the premier European forum for the presentation of new research results in the field of Information Retrieval, ECIR features a wide range of topics such as: IR Theory and Practice; Deep Learning and IR; Web and Social Media IR; User Aspects; IR System Architectures; Content Representation and Processing; Evaluation; Multimedia and Cross-Media IR; Applications.
Advances in Information Retrieval MDPI
Introduces machine learning and its algorithmic paradigms, explaining the principles behind automated learning approaches and the considerations underlying their usage.
25th Pacific-Asia Conference, PAKDD 2021, Virtual Event, May 11–14, 2021, Proceedings, Part III Academic Press
Deep Learning through Sparse Representation and Low-Rank Modeling bridges classical sparse

and low rank models-those that emphasize problem-specific Interpretability-with recent deep network models that have enabled a larger learning capacity and better utilization of Big Data. It shows how the toolkit of deep learning is closely tied with the sparse/low rank methods and algorithms, providing a rich variety of theoretical and analytic tools to guide the design and interpretation of deep learning models. The development of the theory and models is supported by a wide variety of applications in computer vision, machine learning, signal processing, and data mining. This book will be highly useful for researchers, graduate students and practitioners working in the fields of computer vision, machine learning, signal processing, optimization and statistics. Combines classical sparse and low-rank models and algorithms with the latest advances in deep learning networks Shows how the structure and algorithms of sparse and low-rank methods improves the performance and interpretability of Deep Learning models Provides tactics on how to build and apply customized deep learning models for various applications *Advances in Information Retrieval* Springer This book constitutes the proceedings of the 16th International Conference on Web Information Systems and Applications, WISA 2019, held in Qingdao, China, in September 2019. The 39 revised full papers and 33 short papers presented were carefully reviewed and selected from 154 submissions. The papers are grouped in topical sections on machine learning and data mining, cloud computing and big data, information retrieval, natural language processing, data privacy and security, knowledge graphs and social networks, blockchain, query processing, and recommendations. *Second Edition* Springer Due to the fast growth of the Web and the difficulties in finding desired information, efficient and effective information retrieval systems have become more important than ever, and the search engine has become an essential tool for many people. The ranker, a central component in every search engine, is responsible for the matching between processed queries and indexed documents. Because of its central role, great attention has been paid to the research and development of ranking technologies. In addition, ranking is also pivotal for many other information retrieval applications, such as collaborative filtering, definition ranking, question answering, multimedia retrieval, text summarization, and online advertisement. Leveraging machine learning technologies in the ranking process has led to innovative and more effective ranking models, and eventually to a completely new research area called “learning to rank”. Liu first gives a comprehensive review of the major approaches to learning to rank. For each approach he presents the basic framework, with example algorithms, and he discusses its advantages and disadvantages. He continues with some recent advances in learning to rank that cannot be simply categorized into the three major approaches – these include relational ranking, query-dependent ranking,

transfer ranking, and semisupervised ranking. His presentation is completed by several examples that apply these technologies to solve real information retrieval problems, and by theoretical discussions on guarantees for ranking performance. This book is written for researchers and graduate students in both information retrieval and machine learning. They will find here the only comprehensive description of the state of the art in a field that has driven the recent advances in search engine development.
Advances in Knowledge Discovery and Data Mining Foundations and Trends (R) in Information Retrieval
Provides a practical guide to get started and execute on machine learning within a few days without necessarily knowing much about machine learning.The first five chapters are enough to get you started and the next few chapters provide you a good feel of more advanced topics to pursue.
42nd European Conference on IR Research, ECIR 2020, Lisbon, Portugal, April 14–17, 2020, Proceedings, Part II Springer Nature
This two-volume set LNCS 12035 and 12036 constitutes the refereed proceedings of the 42nd European Conference on IR Research, ECIR 2020, held in Lisbon, Portugal, in April 2020.* The 55 full papers presented together with 8 reproducibility papers, 46 short papers, 10 demonstration papers, 12 invited CLEF papers, 7 doctoral consortium papers, 4 workshop papers, and 3 tutorials were carefully reviewed and selected from 457 submissions. They were organized in topical sections named: Part I: deep learning I; entities; evaluation; recommendation; information extraction; deep learning II; retrieval; multimedia; deep learning III; queries; IR – general; question answering, prediction, and bias; and deep learning IV. Part II: reproducibility papers; short papers; demonstration papers; CLEF organizers lab track; doctoral consortium papers; workshops; and tutorials. *Due to the COVID-19 pandemic, this conference was held virtually.
Advances in Information Retrieval Springer Science & Business Media
Search is one of the most critical functionalities of an e-commerce site. Almost every e-commerce site provides a search box. A customer expresses her intent about a product or a category of products in the form of one or more keywords and enters those keywords in the search box. The underlying search engine for an e-commerce site takes those keywords as input and formulate a search query. It subsequently queries its index and retrieves a ranked list of relevant items. The customer then goes through these from the top one after another. She then can click one of those items and lands on its product view page. That page contains detail information about the product. The customer then can add that product to her shopping cart and can purchase it. The internals of an e-commerce search engine is mostly similar to a web search engine. However, the ranking problem for an e-commerce search engine has several unique challenges. In this thesis we unfold the complexities of designing, optimizing and evaluating an e-commerce search engine. We begin with the aspect of the evaluation

of the ranking algorithms for an e-commerce search and provide guidelines for conducting online randomized controlled experiments on a large e-commerce site. In this regard, we discuss managing biases, understanding of the metrics and the population, and the use of right statistical tests. Second, we define a formal framework for designing learning to rank (LTR) algorithms for e-commerce search optimizing the ranking for relevance, revenue, and discovery. We define a measure for discovery and describe the importance of that for an e-commercebusiness. We then propose a practical algorithm for integrating a discovery mechanism with trained learning to rank model. We also describe an approach for evaluating ranking algorithms involving exploration using offline simulation and counterfactuals. We furthermore, design variants of exploration algorithms using our formal framework. We show that a class of such explore algorithms can be designed to be monotonic sub-modular. It is thus possible to construct simple greedy algorithms with theoretical guarantees that maximize the exploration minimizing the loss of relevance and revenue. We conduct simulation studies following our proposed evaluation methodologies to show the effectiveness of our algorithms. Third, we address the problem of incorporating diversity in e-commerce search. We design a knapsack based semi-bandit optimization algorithm for simultaneously learning to diversify and maximizing the revenue. We show that the regret of the algorithm is similar to an existing algorithm for learning to diversify web search although our algorithm is much more straightforward and efficient to realize. We further show that improving diversity in e-commerce search increases the median customer lifetime value (CLV) for an e-commerce business. Fourth, we address the problem of multi-objective learning to rank. We use the LambdaMart algorithm to realize our multi-objective algorithms. LambdaMart algorithm is widely used in Industry, won some recent "learning to rank'' challenges. The authors of the algorithm showed that it could empirically optimize any non-smooth information retrieval ranking metrics such as normalized discounted cumulative gradient (NDCG). It is thus an efficient choice for designing multi-objective learning to rank algorithm. We also extend the physics-based intuition behind the design of lambda gradients to construct a dynamic update for one of the algorithms. We, furthermore, design a second more general algorithm that maintains a fixed number of solutions at each iteration and use a utility function given by a decision maker for that. The algorithm selects the best path that maximizes the NDCG scores of multiple objectives based on the preferences used in the utility function once it completes all the iterations to generate the model. This algorithm can achieve Pareto optimality with the given utility function if such a solution exists. Fifth, we address the problem of quantification and visualization of the excess supply and unmet demand using the contents of thequeries and items. We show the impact of such content gap in search experience.

We quantify the content gap defining a distance between the topic distribution of text of the items and the queries. E-commerce companies can use the insight obtained from such topical content gaps in formulating strategies for demand generation and assortment planning. Finally, we consider the problem of optimization of ranking for the recommender systems for two-sided marketplaces by constructing a two-layer based learning architecture. We show that such two-layer models work better compared to using one model. Although this thesis focuses on the problems in e-commerce search, the algorithms that we developed are general and can have applications in many other domains. We expect to see this thesis serves as a guideline for several new directions of research in information retrieval and machine learning. Notably, we believe this thesis can propel the research for the problems that require integrating online learning with learning to rank, multi-objective learning to rank and regression algorithms, variants of application-specific multi-armed bandit algorithms, and search and recommendation applications for any marketplaces. Statistical Language Models for Information Retrieval Springer Nature Class-tested and coherent, this textbook teaches classical and web information retrieval, including web search and the related areas of text classification and text clustering from basic concepts. It gives an up-to-date treatment of all aspects of the design and implementation of systems for gathering, indexing, and searching documents; methods for evaluating systems; and an introduction to the use of machine learning methods on text collections. All the important ideas are explained using examples and figures, making it perfect for introductory courses in information retrieval for advanced undergraduates and graduate students in computer science. Based on feedback from extensive classroom experience, the book has been carefully structured in order to make teaching more natural and effective. Slides and additional exercises (with solutions for lecturers) are also available through the book's supporting website to help course instructors prepare their lectures. *Proceedings* Synthesis Lectures on Human La The main challenge in learning-to-rank for information retrieval is the difficulty to di- rectly optimize ranking measures to automatically construct a ranking model from training data. It is mainly due to the fact that the ranking measures are determined by the order of ranked documents rather than the specific values of ranking model scores, thus they are non-convex, nondifferentiable and discontinuous. To address this issue, listwise approaches have been proposed where loss functions are defined either by exploiting a probabilistic model or by optimizing upper bounds or smoothed approximations of ranking measures. Even though very promising results have been achieved, there is still a mismatch between target cost and optimization cost. In this work, we present a novel learning algorithm that directly optimizes the ranking measures without resorting to any upper bounds or approx-imations. Our approach is

essentially an iterative greedy coordinate descent method in optimization. For each iteration, we only update one parameter along one coordinate with all others fixed. Since the ranking measure is a stepwise function of a single parameter, we exploit an exhaustive line search algorithm to locate the interval with the smallest ranking measure along each coordinate. We pick the coordinate that leads to the largest reduction of ranking measure. In order to determine the optimal value of the parameter for the selected coordinate, we construct a probabilistic framework for the permutation, and maximize the likelihood of top-m ranked documents. This iterative procedure is continued until conver- gence. We conduct experiments of five datasets selected from Microsoft LETOR datasets, our experimental results show that the proposed direct rank algorithm outperforms several well-known state-of-the-art ranking algorithms.

Learning to Rank for Information Retrieval "O'Reilly Media, Inc." "The amount of digital data we produce every day far surpasses our ability to process this data, and finding useful information in this constant flow of data has become one of the major challenges of the 21st century. Search engines are one way of accessing large data collections. Their algorithms have evolved far beyond simply matching search queries to sets of documents. Today's most sophisticated search engines combine hundreds of relevance signals to provide the best possible results for each searcher. Current approaches for tuning the parameters of search engines can be highly effective. However, they typically require considerable expertise and manual effort. They rely on supervised learning to rank, meaning that they learn from manually annotated examples of relevant documents for given queries. Obtaining large quantities of sufficiently accurate manual annotations is becoming increasingly difficult, especially for personalized search, access to sensitive data, or search in settings that change over time. In this thesis, I develop new online learning to rank techniques, based on insights from reinforcement learning. In contrast to supervised approaches, these methods allow search engines to learn directly from users' interactions. User interactions can typically be observed easily and cheaply, and reflect the preferences of real users. Interpreting user interactions and learning from them is challenging, because they can be biased and noisy. The contributions of this thesis include a novel interleaved comparison method, called probabilistic interleave, that allows unbiased comparisons of search engine result rankings, and methods for learning quickly and effectively from the resulting relative feedback. The obtained analytical and experimental results show how search engines can effectively learn from user interactions."--Omslag.

Web Information Systems and Applications Morgan & Claypool Publishers As online information grows dramatically, search engines such as Google are playing a more and more important role in our lives. Critical to all search engines is the problem of designing an effective retrieval model that can rank documents

accurately for a given query. This has been a central research problem in information retrieval for several decades. In the past ten years, a new generation of retrieval models, often referred to as statistical language models, has been successfully applied to solve many different information retrieval problems. Compared with the traditional models such as the vector space model, these new models have a more sound statistical foundation and can leverage statistical estimation to optimize retrieval parameters. They can also be more easily adapted to model non-traditional and complex retrieval problems. Empirically, they tend to achieve comparable or better performance than a traditional model with less effort on parameter tuning. This book systematically reviews the large body of literature on applying statistical language models to information retrieval with an emphasis on the underlying principles, empirically effective language models, and language models developed for non-traditional retrieval tasks. All the relevant literature has been synthesized to make it easy for a reader to digest the research progress achieved so far and see the frontier of research in this area. The book also offers practitioners an informative introduction to a set of practically useful language models that can effectively solve a variety of retrieval problems. No prior knowledge about information retrieval is required, but some basic knowledge about probability and statistics would be useful for fully digesting all the details. Table of Contents: Introduction / Overview of Information Retrieval Models / Simple Query Likelihood Retrieval Model / Complex Query Likelihood Model / Probabilistic Distance Retrieval Model / Language Models for Special Retrieval Tasks / Language Models for Latent Topic Analysis / Conclusions

Methods, Systems, Challenges Cambridge University Press

In plain, uncomplicated language, and using detailed examples to explain the key concepts, models, and algorithms in vertical search ranking, Relevance Ranking for Vertical Search Engines teaches readers how to manipulate ranking algorithms to achieve better results in real-world applications. This reference book for professionals covers concepts and theories from the fundamental to the advanced, such as relevance, query intention, location-based relevance ranking, and cross-property ranking. It covers the most recent developments in vertical search ranking applications, such as freshness-based relevance theory for new search applications, location-based relevance theory for local search applications, and cross-property ranking theory for applications involving multiple verticals. Foreword by Ron Brachman, Chief Scientist and Head, Yahoo! Labs Introduces ranking algorithms and teaches readers how to manipulate ranking algorithms for the best results Covers concepts and theories from the fundamental to the advanced Discusses the state of the art: development of theories and practices in vertical search ranking applications Includes detailed examples, case studies and real-world

situations
42nd European Conference on IR Research, ECIR 2020, Lisbon, Portugal, April 14–17, 2020, Proceedings, Part I Now Publishers
This compact book explores standard tools for text classification, and teaches the reader how to use machine learning to decide whether a e-mail is spam or ham (binary classification), based on raw data from The SpamAssassin Public Corpus. Of course, sometimes the items in one class are not created equally, or we want to distinguish among them in some meaningful way. The second part of the book will look at how to not only filter spam from our email, but also placing "more important" messages at the top of the queue. This is a curated excerpt from the upcoming book "Machine Learning for Hackers."
Relevant Search
Springer Nature
Learning to Rank for Information RetrievalSpringer
Machine-learned Ranking Algorithms for E-commerce Search and Recommendation Applications Springer Nature
Machine learning approaches for learning ranking functions have been generating much interest from both the web information retrieval community and the machine learning community recently. It has the promise of improved relevancy of search engines and reduced demand for manual parameter tuning. We focus on developing a regression framework for learning to rank with complex loss functions. More specifically, this framework first applies functional iterative or boosting algorithm to compute updates for a given loss function and then fit the updates with a standard regression base learner. We explore supervised learning methodology from machine learning, and we distinguish two types of relevance judgments used as the training data: (1) absolute relevance judgments arising from explicit labeling of query-document pairs; and (2) relative relevance judgments extracted from user click-throughs of search results or converted from the absolute relevance judgments. Within the framework, we propose three novel ranking algorithms and illustrate their application to web search ranking. The first one is to calibrate the existing point-wise(univariant) regression loss to incorporate query difference in terms of introducing nuisance parameters in the statistical models, and we present an alternating optimization method to simultaneously learn the retrieval function and the nuisance parameters. It is an improvement over the existing approach within the category of learning to rank using point-wise regression loss. The second is an extension of gradient boosting methods for point-wise regression loss to complex(multi-variant) loss functions. It is based on optimization of quadratic upper bounds of the loss functions which allows us to present a rigorous convergence analysis of the algorithm. We illustrate an application of this approach in pair-wise preference learning to rank for Web search by combining both preference data and labeled data. The third one is a list-wise approach based on minimum effort optimization that takes into account the entire training data within a query at each iteration.

We tackle this optimization problem using functional iterative methods where the update at each iteration is computed by solving an isotonic regression problem. This more global approach results in faster convergency and signficantly improved performance of the learned ranking functions over existing state-of-the-art methods. Experimental results are carried out using both data sets obtained from a commercial search engine and widely used IR benchmarking data, namely OHSUMED and TREC. Our results show significant improvements of our proposed methods over existing state-of-the-art methods.

An Introduction to Neural Information Retrieval Cambridge University Press

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. Table of Contents: Learning to Rank / Learning for Ranking Creation / Learning for Ranking Aggregation / Methods of Learning to Rank / Applications of Learning

to Rank / Theory of Learning to Rank / Ongoing and Future Work

Advances in Information Retrieval

Springer Nature

Learning to rank refers to machine learning techniques for training the 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 the problem recently and significant progress has been made. This lecture gives an introduction to the area including the fundamental problems, existing 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 SVM, Neural Network based approaches. The author also introduces some popular learning to rank methods in details. These include PRank, OC SVM, Ranking SVM, IR SVM, GBRank, RankNet, LambdaRank, ListNet & ListMLE, AdaRank, SVM MAP, SoftRank, 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. Table of Contents: Introduction / Learning for Ranking Creation / Learning for Ranking Aggregation / Methods of Learning to Rank / Applications of Learning to Rank / Theory of Learning to Rank / Ongoing and Future Work

Related with Learning To Rank For Information Retrieval And Natural Language Processing Hang Li:

• Constant Of Proportionality Table Worksheet Answer Key : click here