

Relevance Ranking for Vertical Search Engines

Bo Long and Yi Chang *Editors*

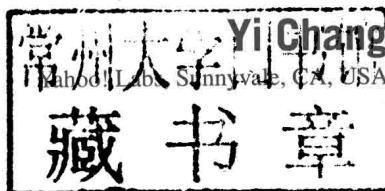
MK
MORGAN KAUFMANN

Relevance Ranking for Vertical Search Engines

Edited by

Bo Long

LinkedIn Inc., Mountain View, CA, USA



AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Morgan Kaufmann is an imprint of Elsevier



Acquiring Editor: Steve Elliot
Editorial Project Manager: Lindsay Lawrence
Project Manager: Punithavathy Govindaradjane
Designer: Maria Inês Cruz

Morgan Kaufmann is an imprint of Elsevier
225 Wyman Street, Waltham, MA 02451, USA

Copyright © 2014 Elsevier Inc. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices

Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data

Relevance ranking for vertical search engines / Bo Long, Yi Chang (Editors).

pages cm

Includes bibliographical references and index.

ISBN 978-0-12-407171-1

1. Text processing (Computer science) 2. Sorting (Electronic computers) 3. Relevance.
4. Database searching. 5. Search engines—Programming. I. Long, Bo, editor of compilation.
II. Chang, Yi (Computer expert)

QA76.9.T48R455 2014

025.04—dc23

2013039777

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library

ISBN: 978-0-12-407171-1

Printed and bound in the United States of America

14 15 16 17 18 10 9 8 7 6 5 4 3 2 1



Working together
to grow libraries in
developing countries

www.elsevier.com • www.bookaid.org

For information on all MK publications visit our website at www.mkp.com

Relevance Ranking for Vertical Search Engines

List of Tables

Number	Table	Page
2.1	Temporal features for URL freshness and query model.	16
2.2	Evaluation corpus.	19
2.3	Feature weights learned by JRFL.	21
2.4	Performance of individual relevance and freshness estimations.	22
2.5	Query intention analysis by the inferred query weight.	23
2.6	Query length distribution under different query categories.	23
2.7	Comparison of random bucket clicks.	24
2.8	Comparison of normal clicks.	24
2.9	Comparison of editorial annotations.	25
2.10	Case study: Degenerated ranking results by JRFL for query "afghanistan."	25
2.11	Performance of offline clustering system.	33
2.12	Results with various simple offline clustering algorithms and the real-time clustering algorithm, which includes the meta-clustering algorithm.	40
2.13	Real-time clustering results with QrySim similarity measure that boosts the weights to the terms that occur close to the query term over the standard similarity measure (OrigSim) with equal weights to all terms.	40
2.14	Q_4 values with the real-time clustering algorithm with various combinations of features. The baselines include features with title and abstract and a single offline clustering algorithm. Although the combined feature set with all the features is the best one, the features with the offline clusters and title and abstract features are comparable to the ones that include body features.	41

2.15	Q_4 values with the real-time clustering algorithm and various granularity settings of offline clusters as features. The baseline feature set includes just title and abstract features. The numbers 1, 2, 3 refer to different settings of the offline clustering algorithm at different granularity settings, specifically varying from coarse to fine representation of clusters. It can be observed that the best accuracy is obtained by combining all the configurations, and individual cluster IDs themselves provide inferior performance.	41
3.1	Distribution of categories of medical concepts in EMERSE queries.	49
4.1	Classification and reranking.	77
4.2	Performance comparison between classification and ranking.	77
5.1	Relative feature importance in baseline.	90
5.2	Comparison of methods for modeling rating scores. Norm+Pred combines methods tagged using *. $b/z/m/c/q/r$ indicates the significance over Baseline, ZeroOneNorm, MeanNorm, AutoNorm-C, AutoNorm-Q, and RatingPred, respectively, at the 0.001 level using the Wilcoxon nondirectional test.	93
5.3	Comparison of methods for modeling review counts. Norm+Pred combines methods tagged using *. The description of notations $b/z/m/c/q/r$ are the same as Table 5.2.	95
5.4	Comparison of methods for modeling distance. Methods with an indicator "+" apply logarithm transformation. $b/z/m/a$ indicates the significance over Baseline, ZeroOneNorm+, MeanNorm, and AutoNorm, respectively, at the 0.05 level using the Wilcoxon nondirectional test.	98
5.5	Comparison of methods for modeling user preference. Methods with an indicator "+" apply logarithm transformation. $b/n/m$ indicates the significance over Baseline, NoNorm, and MeanNorm, respectively, at the 0.01 level using the Wilcoxon nondirectional test.	102
5.6	Sensitivity analysis. These data show that combining the proposed new features (i.e., All) can improve the Baseline over 7%.	102
5.7	Relative feature importance in the final model.	103
6.1	Example entity.	111
6.2	Example facet.	111
6.3	Features.	114

6.4	Entity popularity feature values of top entities.	116
6.5	CTR on category of the query entity vs. facet entity. Each row represents the category of the query entity and the column represents the category of the facet entity; each cell represents aggregate CTR at the intersection. The CTR values are normalized for each row such that the category with the highest CTR in each row is given 1.0. The missing entries indicate that the data for the intersection of those particular categories are not available.	123
6.6	Relevance improvements with various intercategory weights over the baseline. The smaller α , the more the intracategory relationships between facets are emphasized. DCG is computed for each group of facets with the same category.	124
6.7	DCG gain of various sets of features over the baseline.	125
7.1	The aspect relevance mapping function for local search.	135
7.2	Statistics of multi-aspect relevance data.	138
7.3	Statistics of overall relative preference datasets obtained through side-by-side comparison.	140
7.4	Evaluation of aggregation functions on label aggregation accuracy. Linear and Joint are significantly better than Rule (p-value < 0.01).	141
7.5	Evaluation of aggregation functions on ranking accuracy. Statistically significant differences (p-value < 0.01) compared to Rule are highlighted in bold.	142
9.1	Data summary for one source domain and three target domains.	193

List of Figures

Number	Figure	Page
2.1	Scatter plot of CTR versus editor's relevance judgments.	11
2.2	Intuitive illustration of the proposed Joint Relevance and Freshness Learning (JRFL) model. The user-issued query and the corresponding returned URLs are represented by their features on the left part. Dashed-arrow lines in the middle indicate the assumed user's judging process before she clicks. The check boxes on the right record the clicked URLs.	13
2.3	Coordinate descent for JRFL.	15
2.4	Scatter plot of CTR versus JRFL's prediction. (a) Object function value update. (b) Pairwise error rate (PER) update. (c) Query weight α^Q update.	20
2.5	Scatter plot of CTR versus JRFL's prediction.	26
2.6	The architecture of the news search result clustering.	29
2.7	Overview of the offline clustering system.	30
3.1	Snapshot of the new EHR Search Engine (EHR-SE). Medical terms in the query are expanded with synonyms under the same medical concepts. A user can choose to turn on or turn off the query suggestions. The retrieved results are ranked in the lower-right panel with the matched concepts highlighted. The left panel provides a functionality of social search so that the users can store, share, and adopt effective queries.	53
3.2	Network of search bundles presenting the diffusion of search knowledge across departments. Every circle node is a search bundle, sized proportionally to its frequency of use. Every rectangle node refers to a department, with colored convex hulls encompassing all search bundles created by the same department. Every red edge connects a bundle to the departments of the users who "borrowed" them in search; every gray edge connects a bundle to the department of the user who created it. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this book.)	56

3.3	Network of users and bundles presenting the diffusion of search knowledge. Every circle node is a user, sized proportionally to the number of search bundles he has created. Orange edges connect users in the same department. Every other (rectangle) node is a search bundle. A colored convex hull encompasses all users and search bundles created by the users in the same department. Every red edge connects a bundle to the users who "borrowed" them in search; every gray edge connects a bundle to the user who created it.	57
4.1	A generic visual search system.	60
4.2	The four paradigms for visual search reranking.	72
4.3	Examples of top 30 ranked results from a commercial Web image search engine.	73
5.1	Sample search result page of Bing Mobile Local.	82
5.2	Probability of clicks for businesses from a bin, plotted against the mean rating score of this bin.	91
5.3	Rating of clicked businesses, plotted against the mean rating score of the corresponding business category.	92
5.4	Probability of click for businesses from a bin, plotted against the mean distance and $\log(\text{distance})$ scores of this bin.	97
5.5	Distance of clicked businesses, plotted against the mean (traveling) distance of the corresponding business category.	98
5.6	Probability of click for businesses from a bin plotted against the mean user preference of this bin.	101
6.1	Screenshot of entity ranking results embedded into the left-side rail of the search results page.	108
6.2	Distribution of the number of facets per entity. The number of entities is in logarithmic scale. Although some entities have large numbers of facets (typically these are location entities that are connected to many other points of interest), some entities have fewer related facets.	110
6.3	Feature sources for entity ranking.	113
6.4	Entity graph: An example.	115
6.5	Comparison of positionwise CTR between traditional Web search results and entity ranking. All the CTRs in both presented results have been normalized so that CTR at position #1 is 1.	122

6.6	Precision vs. recall of several click models and the pairwise comparison model (PCM). The tradeoff between precision and recall is obtained by different thresholds used for a pairwise preference prediction.	124
7.1	An example of local search result.	128
7.2	An example of an editorial guideline for local search. Note that the overall relevance question is <i>NOT</i> needed in our aggregation methods. The overall relevance is used only for the conventional method.	132
7.3	Distribution of aspect relevance labels. The 1, 0.5, and 0 correspond to the aspect labels in Table 7.1. (a) Category queries; (b) business name queries.	139
7.4	The clickthrough rate (CTR) comparisons based on online experiments for a commercial local search engine. CTRs are normalized not to show absolute values. Each comparison is done for a different test period due to online experiment constraints. (a) CTR5 comparison of Rule and Linear; (b) CTR5 comparison of Linear and ModAgg.	144
8.1	Given the query “new york style pizza,” an aggregated search system decides to blend results from the <i>local</i> , <i>images</i> , and <i>news</i> verticals into the core Web results.	148
8.2	Feature availability during different stages of decision making. Pre-retrieval features, $\phi_{q,v}^{\text{pre-ret}}$, are easy to compute, are available to all verticals, and help with vertical selection decisions. Post-retrieval features, $\phi_{q,v}^{\text{post-ret}}$, are computed only for those verticals from which we request results and are useful for making final presentation decisions. Post-presentation features, $\phi_{q,v}^{\text{post-pres}}$, are available only after the presentation decision has been made and is useful for evaluating performance and predicting the values of post-presentation features for future issuances of q . Features in gray are not computed or logged during an individual search session because of upstream decisions.	151
8.3	Feature caching. The left subfigure represents the logged features computed for previous issuances of q , including post-presentation features. The right subfigure represents the availability of cached feature values (red, in boxes) during runtime at the pre-retrieval stage. Note that these are approximations to downstream feature values and may be sensitive to corpus change or user differences.	151

9.1	A PCDF Bayesian network.	186
9.2	A PCDF model for homogeneous data.	187
9.3	The pointwise cross-domain factor model.	188
9.4	The effects of relative weights of source domain data on DCG5 performance.	195
9.5	DCG5 comparisons of 10 algorithms on the three target domains.	196
9.6	DCG1 comparisons of 10 algorithms on the three target domains.	197

About the Editors

Bo Long is currently a staff applied researcher at LinkedIn Inc., and was formerly a senior research scientist at Yahoo! Labs. His research interests lie in data mining and machine learning with applications to web search, recommendation, and social network analysis. He holds eight innovations and has published peer-reviewed papers in top conferences and journals including ICML, KDD, ICDM, AAAI, SDM, CIKM, and KAIS. He has served as reviewer, workshop co-organizer, conference organizer, committee member, and area chair for multiple conferences, including KDD, NIPS, SIGIR, ICML, SDM, CIKM, JSM, etc.

Yi Chang is a principal scientist and sciences director in Yahoo Labs, where he leads the search and anti-abuse science group. His research interests include web search, applied machine learning, natural language processing, and social computing. Yi has published more than 60 conference/journal papers, and has served as workshop co-organizer, conference organizer, committee member, and area chair for multiple conferences, including WWW, SIGIR, ICML, KDD, CIKM, etc.

List of Contributors

Jaime Arguello

University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Jiang Bian

Microsoft Inc., Beijing, P.R. China

Yi Chang

Yahoo! Labs, Sunnyvale, CA, USA

Fernando Diaz

Microsoft Inc., New York, NY, USA

Anlei Dong

Yahoo! Labs, Sunnyvale, CA, USA

David Hanauer

University of Michigan, Ann Arbor, MI, USA

Yoshiyuki Inagaki

Yahoo! Labs, Sunnyvale, CA, USA

Changsung Kang

Yahoo! Labs, Sunnyvale, CA, USA

Yuan Liu

Microsoft Inc., Beijing, P.R. China

Bo Long

LinkedIn Inc., Mountain View, CA, USA

Yuanhua Lv

Microsoft Inc., Mountain View, CA, USA

Qiaozhu Mei

University of Michigan, Ann Arbor, MI, USA

Tao Mei

Microsoft Inc., Beijing, P.R. China

Belle Tseng

Apple Inc., Cupertino, CA, USA

Srinivas Vadrevu

Microsoft Inc., Mountain View, CA, USA

Hongning Wang

University of Illinois at Urbana-Champaign, Urbana, IL, USA

Xuanhui Wang

Facebook Inc., Menlo Park, CA, USA

Kai Zheng

University of Michigan, Ann Arbor, MI, USA