Camil Demetrescu (Ed.)

LNCS 4525

Experimental Algorithms

6th International Workshop, WEA 2007
Rome, Italy, June 2007
Proceedings

Springer

Camil Demetrescu (Ed.)

Experimental Algorithms

6th International Workshop, WEA 2007
Rome, Italy, June 6-8, 2007
Proceedings

Springer

Volume Editor

Camil Demetrescu
Univ. Roma "La Sapienza"
Facoltà di Ingegneria
Via Eudossiana, 18, 00184 Rome, Italy
E-mail: demetres@dis.uniroma1.it

Library of Congress Control Number: 2007927502

CR Subject Classification (1998): F.2.1-2, E.1, G.1-2, I.3.5, I.2.8

LNCS Sublibrary: SL 1 – Theoretical Computer Science and General Issues

ISSN 0302-9743
ISBN-10 3-540-72844-9 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-72844-3 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springer.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12070992 06/3180 5 4 3 2 1 0

Lecture Notes in Computer Science 4525

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Lecture Notes in Computer Science

For information about Vols. 1–4412

please contact your bookseller or Springer

Vol. 4534: I. Tomkos, F. Neri, J. Solé Pareta, X. Masip Bruin, S. Sánchez Lopez (Eds.), Optical Network Design and Modeling. XI, 460 pages. 2007.

Vol. 4531: J. Indulska, K. Raymond (Eds.), Distributed Applications and Interoperable Systems. XI, 337 pages. 2007.

Vol. 4526: M. Malek, M. Reitenspieß, A. van Moorsel (Eds.), Service Availability. X, 155 pages. 2007.

Vol. 4525: C. Demetrescu (Ed.), Experimental Algorithms. XIII, 448 pages. 2007.

Vol. 4523: Y.-H. Lee, H.-N. Kim, J. Kim, Y. Park, L.T. Yang, S.W. Kim (Eds.), Embedded Software and Systems. XIX, 829 pages. 2007.

Vol. 4521: J. Katz, M. Yung (Eds.), Applied Cryptography and Network Security. XIII, 498 pages. 2007.

Vol. 4519: E. Franconi, M. Kifer, W. May (Eds.), The Semantic Web: Research and Applications. XVIII, 830 pages. 2007.

Vol. 4517: F. Boavida, E. Monteiro, S. Mascolo, Y. Koucheryavy (Eds.), Wired/Wireless Internet Communications. XIV, 382 pages. 2007.

Vol. 4515: M. Naor (Ed.), Advances in Cryptology - EUROCRYPT 2007. XIII, 591 pages. 2007.

Vol. 4514: S.N. Artemov, A. Nerode (Eds.), Logical Foundations of Computer Science. XI, 513 pages. 2007.

Vol. 4510: P. Van Hentenryck, L. Wolsey (Eds.), Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems. X, 391 pages. 2007.

Vol. 4509: Z. Kobti, D. Wu (Eds.), Advances in Artificial Intelligence. XII, 552 pages. 2007. (Sublibrary LNAI).

Vol. 4506: D. Zeng, I. Gotham, K. Komatsu, C. Lynch, M. Thurmond, D. Madigan, B. Lober, J. Kvach, H. Chen (Eds.), Intelligence and Security Informatics: Biosurveillance. XI, 234 pages. 2007.

Vol. 4504: J. Huang, R. Kowalczyk, Z. Maamar, D. Martin, I. Müller, S. Stoutenburg, K.P. Sycara (Eds.), Service-Oriented Computing: Agents, Semantics, and Engineering. X, 175 pages. 2007.

Vol. 4501: J. Marques-Silva, K.A. Sakallah (Eds.), Theory and Applications of Satisfiability Testing – SAT 2007. XI, 384 pages. 2007.

Vol. 4500: N. Streitz, A. Kameas, I. Mavrommati (Eds.), The Disappearing Computer. XVIII, 304 pages. 2007.

Vol. 4496: N.T. Nguyen, A. Grzech, R.J. Howlett, L.C. Jain (Eds.), Agent and Multi-Agent Systems: Technologies and Applications. XXI, 1046 pages. 2007. (Sublibrary LNAI).

Vol. 4493: D. Liu, S. Fei, Z. Hou, H. Zhang, C. Sun (Eds.), Advances in Neural Networks – ISNN 2007, Part III. XXVI, 1215 pages. 2007.

Vol. 4492: D. Liu, S. Fei, Z. Hou, H. Zhang, C. Sun (Eds.), Advances in Neural Networks – ISNN 2007, Part II. XXVII, 1321 pages. 2007.

Vol. 4491: D. Liu, S. Fei, Z.-G. Hou, H. Zhang, C. Sun (Eds.), Advances in Neural Networks – ISNN 2007, Part I. LIV, 1365 pages. 2007.

Vol. 4490: Y. Shi, G.D. van Albada, J. Dongarra, P.M.A. Sloot (Eds.), Computational Science – ICCS 2007, Part IV. XXXVII, 1211 pages. 2007.

Vol. 4489: Y. Shi, G.D. van Albada, J. Dongarra, P.M.A. Sloot (Eds.), Computational Science – ICCS 2007, Part III. XXXVII, 1257 pages. 2007.

Vol. 4488: Y. Shi, G.D. van Albada, J. Dongarra, P.M.A. Sloot (Eds.), Computational Science – ICCS 2007, Part II. XXXV, 1251 pages. 2007.

Vol. 4487: Y. Shi, G.D. van Albada, J. Dongarra, P.M.A. Sloot (Eds.), Computational Science – ICCS 2007, Part I. LXXXI, 1275 pages. 2007.

Vol. 4486: M. Bernardo, J. Hillston (Eds.), Formal Methods for Performance Evaluation. VII, 469 pages. 2007.

Vol. 4485: F. Sgallari, A. Murli, N. Paragios (Eds.), Scale Space Methods and Variational Methods in Computer Vision. XV, 931 pages. 2007.

Vol. 4484: J.-Y. Cai, S.B. Cooper, H. Zhu (Eds.), Theory and Applications of Models of Computation. XIII, 772 pages. 2007.

Vol. 4483: C. Baral, G. Brewka, J. Schlipf (Eds.), Logic Programming and Nonmonotonic Reasoning. IX, 327 pages. 2007. (Sublibrary LNAI).

Vol. 4482: A. An, J. Stefanowski, S. Ramanna, C.J. Butz, W. Pedrycz, G. Wang (Eds.), Rough Sets, Fuzzy Sets, Data Mining and Granular Computing. XIV, 585 pages. 2007. (Sublibrary LNAI).

Vol. 4481: J. Yao, P. Lingras, W.-Z. Wu, M. Szczuka, N.J. Cercone, D. Ślęzak (Eds.), Rough Sets and Knowledge Technology. XIV, 576 pages. 2007. (Sublibrary LNAI).

Vol. 4480: A. LaMarca, M. Langheinrich, K.N. Truong (Eds.), Pervasive Computing. XIII, 369 pages. 2007.

Vol. 4479: I.F. Akyildiz, R. Sivakumar, E. Ekici, J.C.d. Oliveira, J. McNair (Eds.), NETWORKING 2007. Ad Hoc and Sensor Networks, Wireless Networks, Next Generation Internet. XXVII, 1252 pages. 2007.

Vol. 4478: J. Martí, J.M. Benedí, A.M. Mendonça, J. Serrat (Eds.), Pattern Recognition and Image Analysis, Part II. XXVII, 657 pages. 2007.

Vol. 4477: J. Martí, J.M. Benedí, A.M. Mendonça, J. Serrat (Eds.), Pattern Recognition and Image Analysis, Part I. XXVII, 625 pages. 2007.

Vol. 4476: V. Gorodetsky, C. Zhang, V.A. Skormin, L. Cao (Eds.), Autonomous Intelligent Systems: Multi-Agents and Data Mining. XIII, 323 pages. 2007. (Sublibrary LNAI).

Vol. 4475: P. Crescenzi, G. Prencipe, G. Pucci (Eds.), Fun with Algorithms. X, 273 pages. 2007.

Vol. 4472: M. Haindl, J. Kittler, F. Roli (Eds.), Multiple Classifier Systems. XI, 524 pages. 2007.

Vol. 4471: P. Cesar, K. Chorianopoulos, J.F. Jensen (Eds.), Interactive TV: a Shared Experience. XIII, 236 pages. 2007.

Vol. 4470: Q. Wang, D. Pfahl, D.M. Raffo (Eds.), Software Process Dynamics and Agility. XI, 346 pages. 2007.

Vol. 4465: T. Chahed, B. Tuffin (Eds.), Network Control and Optimization. XIII, 305 pages. 2007.

Vol. 4464: E. Dawson, D.S. Wong (Eds.), Information Security Practice and Experience. XIII, 361 pages. 2007.

Vol. 4463: I. Măndoiu, A. Zelikovsky (Eds.), Bioinformatics Research and Applications. XV, 653 pages. 2007. (Sublibrary LNBI).

Vol. 4462: D. Sauveron, K. Markantonakis, A. Bilas, J.-J. Quisquater (Eds.), Information Security Theory and Practices. XII, 255 pages. 2007.

Vol. 4459: C. Cérin, K.-C. Li (Eds.), Advances in Grid and Pervasive Computing. XVI, 759 pages. 2007.

Vol. 4453: T. Speed, H. Huang (Eds.), Research in Computational Molecular Biology. XVI, 550 pages. 2007. (Sublibrary LNBI).

Vol. 4452: M. Fasli, O. Shehory (Eds.), Agent-Mediated Electronic Commerce. VIII, 249 pages. 2007. (Sublibrary LNAI).

Vol. 4451: T.S. Huang, A. Nijholt, M. Pantic, A. Pentland (Eds.), Artifical Intelligence for Human Computing. XVI, 359 pages. 2007. (Sublibrary LNAI).

Vol. 4450: T. Okamoto, X. Wang (Eds.), Public Key Cryptography – PKC 2007. XIII, 491 pages. 2007.

Vol. 4448: M. Giacobini et al. (Ed.), Applications of Evolutionary Computing. XXIII, 755 pages. 2007.

Vol. 4447: E. Marchiori, J.H. Moore, J.C. Rajapakse (Eds.), Evolutionary Computation,Machine Learning and Data Mining in Bioinformatics. XI, 302 pages. 2007.

Vol. 4446: C. Cotta, J. van Hemert (Eds.), Evolutionary Computation in Combinatorial Optimization. XII, 241 pages. 2007.

Vol. 4445: M. Ebner, M. O'Neill, A. Ekárt, L. Vanneschi, A.I. Esparcia-Alcázar (Eds.), Genetic Programming. XI, 382 pages. 2007.

Vol. 4444: T. Reps, M. Sagiv, J. Bauer (Eds.), Program Analysis and Compilation, Theory and Practice. X, 361 pages. 2007.

Vol. 4443: R. Kotagiri, P.R. Krishna, M. Mohania, E. Nantajeewarawat (Eds.), Advances in Databases: Concepts, Systems and Applications. XXI, 1126 pages. 2007.

Vol. 4440: B. Liblit, Cooperative Bug Isolation. XV, 101 pages. 2007.

Vol. 4439: W. Abramowicz (Ed.), Business Information Systems. XV, 654 pages. 2007.

Vol. 4438: L. Maicher, A. Sigel, L.M. Garshol (Eds.), Leveraging the Semantics of Topic Maps. X, 257 pages. 2007. (Sublibrary LNAI).

Vol. 4433: E. Şahin, W.M. Spears, A.F.T. Winfield (Eds.), Swarm Robotics. XII, 221 pages. 2007.

Vol. 4432: B. Beliczynski, A. Dzielinski, M. Iwanowski, B. Ribeiro (Eds.), Adaptive and Natural Computing Algorithms, Part II. XXVI, 761 pages. 2007.

Vol. 4431: B. Beliczynski, A. Dzielinski, M. Iwanowski, B. Ribeiro (Eds.), Adaptive and Natural Computing Algorithms, Part I. XXV, 851 pages. 2007.

Vol. 4430: C.C. Yang, D. Zeng, M. Chau, K. Chang, Q. Yang, X. Cheng, J. Wang, F.-Y. Wang, H. Chen (Eds.), Intelligence and Security Informatics. XII, 330 pages. 2007.

Vol. 4429: R. Lu, J.H. Siekmann, C. Ullrich (Eds.), Cognitive Systems. X, 161 pages. 2007. (Sublibrary LNAI).

Vol. 4427: S. Uhlig, K. Papagiannaki, O. Bonaventure (Eds.), Passive and Active Network Measurement. XI, 274 pages. 2007.

Vol. 4426: Z.-H. Zhou, H. Li, Q. Yang (Eds.), Advances in Knowledge Discovery and Data Mining. XXV, 1161 pages. 2007. (Sublibrary LNAI).

Vol. 4425: G. Amati, C. Carpineto, G. Romano (Eds.), Advances in Information Retrieval. XIX, 759 pages. 2007.

Vol. 4424: O. Grumberg, M. Huth (Eds.), Tools and Algorithms for the Construction and Analysis of Systems. XX, 738 pages. 2007.

Vol. 4423: H. Seidl (Ed.), Foundations of Software Science and Computational Structures. XVI, 379 pages. 2007.

Vol. 4422: M.B. Dwyer, A. Lopes (Eds.), Fundamental Approaches to Software Engineering. XV, 440 pages. 2007.

Vol. 4421: R. De Nicola (Ed.), Programming Languages and Systems. XVII, 538 pages. 2007.

Vol. 4420: S. Krishnamurthi, M. Odersky (Eds.), Compiler Construction. XIV, 233 pages. 2007.

Vol. 4419: P.C. Diniz, E. Marques, K. Bertels, M.M. Fernandes, J.M.P. Cardoso (Eds.), Reconfigurable Computing: Architectures, Tools and Applications. XIV, 391 pages. 2007.

Vol. 4418: A. Gagalowicz, W. Philips (Eds.), Computer Vision/Computer Graphics Collaboration Techniques. XV, 620 pages. 2007.

Vol. 4416: A. Bemporad, A. Bicchi, G. Buttazzo (Eds.), Hybrid Systems: Computation and Control. XVII, 797 pages. 2007.

Vol. 4415: P. Lukowicz, L. Thiele, G. Tröster (Eds.), Architecture of Computing Systems - ARCS 2007. X, 297 pages. 2007.

Vol. 4414: S. Hochreiter, R. Wagner (Eds.), Bioinformatics Research and Development. XVI, 482 pages. 2007. (Sublibrary LNBI).

Preface

This volume contains the papers presented at the 6th Workshop on Experimental Algorithms (WEA 2007), held at the School of Engineering of the University of Rome "La Sapienza" on June 6–8, 2007. The conference is devoted to fostering and disseminating high quality research results focused on the experimental analysis of algorithms and aims at bringing together researchers from the computer science and operations research communities. Papers were solicited from all areas of algorithmic engineering research.

The preceding workshops were held in Riga (Latvia, 2001), Ascona (Switzerland, 2003), Angra dos Reis (Brazil, 2004), Santorini (Greece, 2005), and Menorca Island (Spain, 2006). The proceedings of the previous WEAs were published as Springer volumes LNCS 2138 (in conjunction with the 13th International Symposium on Fundamentals of Computation Theory, FCT 2001), LNCS 2647 (2003), LNCS 3059 (2004), LNCS 3503 (2005), and LNCS 4007 (2006).

The conference received 121 submissions. Each submission was reviewed by at least three program committee members and evaluated on its quality, originality, and relevance to the conference. Overall, the program committee wrote 440 reviews with the help of 100 trusted external referees. The committee selected 30 papers, leading to an acceptance rate of 24.8%. On average, the authors of each submitted paper received 800 words of comments. The decision process was made electronically using the EasyChair conference management system.

In addition to the accepted contributions, this volume also contains the invited lectures by Corinna Cortes (Google Research), Peter Sanders (Universität Karlsruhe), and Maria Serna (Universitat Politècnica de Catalunya).

We would like to thank all the authors who responded to the call for papers, the invited speakers, the members of the program committee, as well as the external referees and the organizing committee members.

We gratefully acknowledge support from the University of Rome "La Sapienza" and the University of Rome "Tor Vergata".

April 2007 Camil Demetrescu

Conference Organization

Program Committee

Susanne Albers	U. Freiburg
Eric Angel	U. Evry Val D'Essonne
Giorgio Ausiello	U. Roma "La Sapienza"
Ricardo Baeza-Yates	Yahoo! Research
Jon Bentley	Avaya Labs Research
Adam Buchsbaum	AT&T Labs–Research
Ioannis Chatzigiannakis	U. Patras
Camil Demetrescu (Chair)	U. Roma "La Sapienza"
Cid de Souza	U. Campinas
Rolf Fagerberg	U. Southern Denmark, Odense
Rudolf Fleischer	Fudan U. Shanghai
Monika Henzinger	Google & EPFL
Dorit S. Hochbaum	U. California at Berkeley
Michel Gendreau	U. Montréal
Giuseppe F. Italiano	U. Roma "Tor Vergata"
Riko Jacob	ETH Zürich
Richard Ladner	U. Washington
Kurt Mehlhorn	MPII Saarbrücken
Bernard Moret	EPFL
S. Muthukrishnan	Google
Petra Mutzel	U. Dortmund
Giri Narasimhan	Florida International U.
Robert Sedgewick	Princeton U.
Thomas Stuetzle	U. Libre de Bruxelles
Jan Vahrenhold	U. Dortmund
Renato Werneck	Microsoft Research Silicon Valley
Norbert Zeh	Dalhousie U., Halifax

Steering Committee

Edoardo Amaldi	Politecnico di Milano
David A. Bader	Georgia Tech
Josep Díaz	T.U. Catalonia
Giuseppe F. Italiano	U. Roma "Tor Vergata"
David S. Johnson	AT&T Labs–Research
Klaus Jansen	U. Kiel
Kurt Mehlhorn	MPII Saarbrücken
Ian Munro	U. Waterloo

Sotiris Nikoletseas	U. Patras & CTI
Jose Rolim (Chair)	U. Geneva
Pavlos Spirakis	U. Patras & CTI

Organizing Committee

Vincenzo Bonifaci	U. Roma "La Sapienza"
Saverio Caminiti	U. Roma "La Sapienza"
Fabio Dellutri	U. Roma "Tor Vergata"
Camil Demetrescu	U. Roma "La Sapienza"
Irene Finocchi	U. Roma "La Sapienza"
Luigi Laura	U. Roma "La Sapienza"
Andrea Vitaletti	U. Roma "La Sapienza"

External Reviewers

Ilan Adler
Nina Amenta
Diogo Andrade
Prasanna Balaprakash
Evripidis Bampis
Jørgen Bang-Jensen
Luca Becchetti
Nicola Beume
Immanuel Bomze
Vincenzo Bonifaci
Markus Borschbach
Prosenjit Bose
Patricia Buendia
Luciana Buriol
Pippo Cattaneo
Victor Cavalcante
Bala Chandran
Marco Chiarandini
Markus Chimani
Edmund Christiansen
Alexandre Cunha
Florian Diedrich
David Eppstein
Udo Feldkamp
Carlos Ferreira
Esteban Feuerstein
Michele Flammini
Paolo Giulio Franciosa
Markus Gärtner
Michael Gatto
Fabian Gieseke
Aristides Gionis
Andrew V. Goldberg
Laurent Gourves
Fabrizio Grandoni
Carsten Gutwenger
Edna Hoshino
Frank Hutter
David S. Johnson
Allan Jørgensen
Marcin Kaminski
Maria Kandyba
George Karakostas
Athanasios Kinalis
Karsten Klein
Stephen Kobourov
Sven O. Krumke
Alexander Kröller
Kim S. Larsen
Orlando Lee
Asaf Levin
Vassiliki Liagkou
Andrea Lodi
Manuel López-Ibáñez
Abilio Lucena
Charles (Chip) Martel
Peter Marwedel
Jens Maue

Victor Milenkovic
Michele Monaci
Gabriel Moruz
Arnaldo Moura
Georgios Mylonas
Umberto Nanni
Gonzalo Navarro
Marc Nunkesser
Marco A. Montes de Oca
Rasmus Pagh
Konstantinos Panagiotou
Luis Paquete
Fanny Pascual
Wolfgang Paul
Christian Nørgaard Storm Pedersen
Leon Peeters
Wei Peng
Ulrich Pferschy
Olivier Powell
Kirk Pruhs
Joe Qranfal
Knut Reinert
Andrea Ribichini
Giovanni Rinaldi
Adi Rosen
Domenico Saccà
Claus P. Schnorr
Meera Sitharam
Martin Skutella
Michiel Smid
Damien Stehle
Etsuji Tomita
Jan van der Veen
Andrea Vitaletti
Yoshiko Wakabayashi
Michael Waschbüsch
Anthony Wirth
Hans-Christoph Wirth
Hoi-Ming Wong
Eduardo Xavier
Tallys Yunes
Martin Zachariasen

Table of Contents

Invited Lectures

Session 1 (Route Planning)

Session 2 (Dynamic Trees, Skip Lists, and Bloom Filters)

Session 3 (Crossing Minimizationm, TSP, and Vehicle Routing

Session 4 (Network Routing and Stability)

Session 5 (Strings and Range Searching)

Session 6 (Matching, Flows, and Spanners)

Session 7 (Covering, Coloring, and Partitioning)

Session 8 (Applications)

Session 9 (Spanning Trees)

Session 10 (Packing and Auctions)

An Alternative Ranking Problem for Search Engines

Corinna Cortes[1], Mehryar Mohri[2,1], and Ashish Rastogi[2]

[1] Google Research,
76 Ninth Avenue,
New York, NY 10011
[2] Courant Institute of Mathematical Sciences,
251 Mercer Street
New York, NY 10012

Abstract. This paper examines in detail an alternative ranking problem for search engines, movie recommendation, and other similar ranking systems motivated by the requirement to not just accurately predict pairwise ordering but also preserve the magnitude of the preferences or the difference between ratings. We describe and analyze several cost functions for this learning problem and give stability bounds for their generalization error, extending previously known stability results to non-bipartite ranking and magnitude of preference-preserving algorithms. We present algorithms optimizing these cost functions, and, in one instance, detail both a batch and an on-line version. For this algorithm, we also show how the leave-one-out error can be computed and approximated efficiently, which can be used to determine the optimal values of the trade-off parameter in the cost function. We report the results of experiments comparing these algorithms on several datasets and contrast them with those obtained using an AUC-maximization algorithm. We also compare training times and performance results for the on-line and batch versions, demonstrating that our on-line algorithm scales to relatively large datasets with no significant loss in accuracy.

1 Motivation

The learning problem of ranking has gained an increasing amount of interest in the machine learning community over the last decade, in part due to the remarkable success of web search engines and recommender systems (Freund et al., 1998; Crammer & Singer, 2001; Joachims, 2002; Shashua & Levin, 2003; Cortes & Mohri, 2004; Rudin et al., 2005; Agarwal & Niyogi, 2005). The recent Netflix challenge has further stimulated the learning community fueling its research with invaluable datasets (Netflix, 2006).

The goal of information retrieval engines is to return a set of documents, or clusters of documents, ranked in decreasing order of relevance to the user. The order may be common to all users, as with most search engines, or tuned to individuals to provide personalized search results or recommendations. The accuracy of this ordered list is the key quality measure of theses systems.

C. Demetrescu (Ed.): WEA 2007, LNCS 4525, pp. 1–22, 2007.

In most previous research studies, the problem of ranking has been formulated as that of learning from a labeled sample of pairwise preferences a scoring function with small pairwise misranking error (Freund et al., 1998; Herbrich et al., 2000; Crammer & Singer, 2001; Joachims, 2002; Rudin et al., 2005; Agarwal & Niyogi, 2005). But this formulation suffers some short-comings.

Firstly, most users inspect only the top results. Thus, it would be natural to enforce that the results returned near the top be particularly relevant and correctly ordered. The quality and ordering of the results further down the list matter less. An average pairwise misranking error directly penalizes errors at both extremes of a list more heavily than errors towards the middle of the list, since errors at the extremes result in more misranked pairs. However, one may wish to explicitly encode the requirement of ranking quality at the top in the cost function. One common solution is to weigh examples differently during training so that more important or high-quality results be assigned larger weights. This imposes higher accuracy on these examples, but does not ensure a high-quality ordering at the top. A good formulation of this problem leading to a convex optimization problem with a unique minimum is still an open question.

Another shortcoming of the pairwise misranking error is that this formulation of the problem and thus the scoring function learned ignore the magnitude of the preferences. In many applications, it is not sufficient to determine if one example is preferred to another. One may further request an assessment of how large that preference is. Taking this magnitude of preference into consideration is critical, for example in the design of search engines, which originally motivated our study, but also in other recommendation systems. For a recommendation system, one may choose to truncate the ordered list returned where a large gap in predicted preference is found. For a search engine, this may trigger a search in parallel corpora to display more relevant results.

This motivated our study of the problem of ranking while preserving the magnitude of preferences, which we will refer to in short by *magnitude-preserving ranking*.[1] The problem that we are studying bears some resemblance with that of ordinal regression (McCullagh, 1980; McCullagh & Nelder, 1983; Shashua & Levin, 2003; Chu & Keerthi, 2005). It is however distinct from ordinal regression since in ordinal regression the magnitude of the difference in target values is not taken into consideration in the formulation of the problem or the solutions proposed. The algorithm of Chu and Keerthi (2005) does take into account the ordering of the classes by imposing that the thresholds be monotonically increasing, but this still ignores the difference of target values and thus does not follow the same objective. A crucial aspect of the algorithms we propose is that they penalize misranking errors more heavily in the case of larger magnitudes of preferences.

We describe and analyze several cost functions for this learning problem and give stability bounds for their generalization error, extending previously known stability results to non-bipartite ranking and magnitude of preference-preserving algorithms. In particular, our bounds extend the framework of (Bousquet &

[1] This paper is an extended version of (Cortes et al., 2007).

Elisseeff, 2000; Bousquet & Elisseeff, 2002) to the case of cost functions over pairs of examples, and extend the bounds of Agarwal and Niyogi (2005) beyond the bi-partite ranking problem. Our bounds also apply to algorithms optimizing the so-called *hinge rank loss*.

We present several algorithms optimizing these cost functions, and in one instance detail both a batch and an on-line version. For this algorithm, MPRank, we also show how the leave-one-out error can be computed and approximated efficiently, which can be used to determine the optimal values of the trade-off parameter in the cost function. We also report the results of experiments comparing these algorithms on several datasets and contrast them with those obtained using RankBoost (Freund et al., 1998; Rudin et al., 2005), an algorithm designed to minimize the exponentiated loss associated with the Area Under the ROC Curve (AUC), or pairwise misranking. We also compare training times and performance results for the on-line and batch versions of MPRank, demonstrating that our on-line algorithm scales to relatively large datasets with no significant loss in accuracy.

The remainder of the paper is organized as follows. Section 2 describes and analyzes our algorithms in detail. Section 3 presents stability-based generalization bounds for a family of magnitude-preserving algorithms. Section 4 presents the results of our experiments with these algorithms on several datasets.

2 Algorithms

Let S be a sample of m labeled examples drawn i.i.d. from a set X according to some distribution D:

$$(x_1, y_1), \ldots, (x_m, y_m) \in X \times \mathbb{R}. \tag{1}$$

For any $i \in [1, m]$, we denote by S^{-i} the sample derived from S by omitting example (x_i, y_i), and by S^i the sample derived from S by replacing example (x_i, y_i) with an other example (x'_i, y'_i) drawn i.i.d. from X according to D. For convenience, we will sometimes denote by $y_x = y_i$ the label of a point $x = x_i \in X$.

The quality of the ranking algorithms we consider is measured with respect to pairs of examples. Thus, a cost functions c takes as arguments two sample points. For a fixed cost function c, the empirical error $\widehat{R}(h, S)$ of a hypothesis $h : X \mapsto \mathbb{R}$ on a sample S is defined by:

$$\widehat{R}(h, S) = \frac{1}{m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} c(h, x_i, x_j). \tag{2}$$

The true error $R(h)$ is defined by

$$R(h) = \mathrm{E}_{x,x' \sim D}[c(h, x, x')]. \tag{3}$$

2.1 Cost Functions

We introduce several cost functions related to magnitude-preserving ranking. The first one is the so-called *hinge rank loss* which is a natural extension of the pairwise misranking loss (Cortes & Mohri, 2004; Rudin et al., 2005). It penalizes a pairwise misranking by the magnitude of preference predicted or the nth power of that magnitude ($n = 1$ or $n = 2$):

$$c^n_{\mathrm{HR}}(h, x, x') = \begin{cases} 0, \text{ if } (h(x') - h(x))(y_{x'} - y_x) \geq 0 \\ \big|(h(x') - h(x))\big|^n, \text{ otherwise.} \end{cases} \tag{4}$$

c^n_{HR} does not take into consideration the true magnitude of preference $y_{x'} - y_x$ for each pair (x, x') however. The following cost function has this property and penalizes deviations of the predicted magnitude with respect to the true one. Thus, it matches our objective of magnitude-preserving ranking ($n = 1, 2$):

$$c^n_{\mathrm{MP}}(h, x, x') = \big|(h(x') - h(x)) - (y_{x'} - y_x)\big|^n. \tag{5}$$

A one-sided version of that cost function penalizing only misranked pairs is given by ($n = 1, 2$):

$$c^n_{\mathrm{HMP}}(h, x, x') = \begin{cases} 0, \text{ if } (h(x') - h(x))(y_{x'} - y_x) \geq 0 \\ \big|(h(x') - h(x)) - (y_{x'} - y_x)\big|^n, \text{ otherwise.} \end{cases} \tag{6}$$

Finally, we will consider the following cost function derived from the ϵ-insensitive cost function used in SVM regression (SVR) (Vapnik, 1998) ($n = 1, 2$):

$$c^n_{\mathrm{SVR}}(h, x, x') = \begin{cases} 0, \text{ if } \big|\big[(h(x') - h(x)) - (y_{x'} - y_x)\big]\big| \leq \epsilon \\ \big|(h(x') - h(x)) - (y_{x'} - y_x) - \epsilon\big|^n, \text{ otherwise.} \end{cases} \tag{7}$$

Note that all of these cost functions are convex functions of $h(x)$ and $h(x')$.

2.2 Objective Functions

The regularization algorithms based on the cost functions c^n_{MP} and c^n_{SVR} correspond closely to the idea of preserving the magnitude of preferences since these cost functions penalize deviations of a predicted difference of score from the target preferences. We will refer by MPRank to the algorithm minimizing the regularization-based objective function based on c^n_{MP}:

$$F(h, S) = \|h\|^2_K + C \frac{1}{m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} c^n_{\mathrm{MP}}(h, x_i, x_j), \tag{8}$$

and by SVRank to the one based on the cost function c^n_{SVR}

$$F(h, S) = \|h\|^2_K + C \frac{1}{m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} c^n_{\mathrm{SVR}}(h, x_i, x_j). \tag{9}$$

For a fixed n, $n = 1, 2$, the same stability bounds hold for both algorithms as seen in the following section. However, their time complexity is significantly different.