AIM WORKSHOP: THE MATHEMATICS OF RANKING

Dates: August 16–20, 2010

Organizers:
Lek-Heng Lim, University of California, Berkeley [primary contact]
Shivani Agarwal, Massachusetts Institute of Technology

Invited participants:

Kenneth Arrow, Stanford University, Economics (Social Choice Theory)
Peter Bartlett, University of California, Berkeley, Statistics (Statistical Learning Theory)
Alexander Barvinok, University of Michigan, Mathematics (Combinatorics)
Chiranjib Bhattacharyya, Indian Institute of Science, Computer Science (Machine Learning)
Corinna Cortes, Google Research (Machine Learning, Applications)
Koby Crammer, Technion, Electrical Engineering (Machine Learning)
Persi Diaconis, Stanford University, Mathematics (Statistics)
Dorit Hochbaum, University of California, Berkeley, Operations Research (Optimization)
Susan Holmes, Stanford University, Statistics (Computational Statistics)
Michael Kearns, University of Pennsylvania, Computer Science (Machine Learning, Game Theory)
Ravi Kumar, Yahoo! Research (Algorithms, Applications)
Risi Kondor, California Institute of Technology, Mathematics (Combinatorics, Group Representations)
Amy Langville, College of Charleston, Mathematics (Linear Algebra, Markov Chains)
Tanya Leise, Amherst College, Mathematics (Linear Algebra, Differential Equations)
Jason Morton, Stanford University, Mathematics (Discrete Geometry, Statistical Models)
David Parkes, Harvard University, Computer Science (Game Theory, Economics)
Donald Saari, University of California, Irvine, Mathematics (Voting Theory)
Thomas Saaty, University of Pittsburgh, Business (Decision Theory)
Robert Schapire, Princeton University, Computer Science (Machine Learning, Algorithms)
Anne Shiu, University of California, Berkeley, Mathematics (Algebra, Computational Biology)
Yoram Singer, Google Research (Machine Learning, Algorithms)
Kevin Small, Tufts University, Computer Science (Machine Learning)
Nicolas Vayatis, Ecole Normale Supérieure, Mathematics (Statistical Learning Theory)
Yuan Yao, Peking University (Bioinformatics, Machine Learning)
Tong Zhang, Rutgers University, Statistics (Statistical Learning Theory)

MSC:
05B35, 05E15, 05E18, 06A05, 06A06, 06A07, 06A75, 20B30, 20B35, 20C05, 20C15, 20C30, 20F16, 20P05, 43A40, 43A65, 51G05, 58A14, 60B15, 60C05, 62F07, 60G25, 62G30, 62M20, 68Q32, 68Q87, 68T05, 68W40, 90-08, 91B06, 91B08, 91B12, 91B14, 91B16, 91B18, 91E40

1. Workshop Description

Problems of ranking arise in multiple domains. In elections, one wants to rank candidates based on votes cast by different voters. In web search, one wants to rank web pages according to relevance to a topic or query. In recommendation systems, one wants to rank items according to a user's likes and dislikes. In computational biology, one wants to rank genes by relevance to a disease, and in drug discovery, one wants to rank molecules based on their chances of success as a drug.

Ranking problems have been studied under different guises in many different fields. Each field has asked its own questions about ranking, and has developed its own mathematical tools in an attempt to answer those questions. For example, in social choice theory, which studies voting procedures, it is often of interest to combine or aggregate rankings of different candidates produced by different voters, in order to arrive

at a consensus ranking; many fascinating mathematical results have been shown in this area, including for example the famous impossibility result of Kenneth Arrow [6]. In statistics, the study of ranked data has played an important role in the development of non-parametric statistics in particular [22]; there have also been many rich results on studying rankings via probability models defined on permutation groups [10]. More recently, ranking problems have received much attention in machine learning, a field that brings together techniques from a diverse range of disciplines including computation, mathematics, and statistics in order to analyze complex data [8, 13, 3, 26, 2, 7, 9, 4, 15].

Mathematical topics that arise in ranking problems include representations of the symmetric group [10, 11], discrete geometry of permutahedrons [23, 24], order relations on posets [6], combinatorial Hodge theory [15], harmonic analysis on finite groups and their homogeneous spaces [20, 19], algorithmic graph theory [5, 12], Perron-Frobenius theory (in PageRank) [25, 21], singular value decomposition (in HITS) [16], network flows [14], and convergence in probability [3, 26, 7]. There are also studies of ranking that are clearly mathematical but difficult to categorize [27, 28, 29, 30].

This workshop aims to bring together researchers from a broad spectrum of both fundamental and applied areas to share their perspectives on ranking, to brainstorm, and to forge collaborations. We recognize that the subject of ranking is truly multidisciplinary – with mathematics being the common thread. And so our list includes not only academic colleagues from computer science, economics, mathematics, operations research, statistics, but also industry colleagues from Google, Microsoft, and Yahoo!; and features a seniority range going from graduate students to Nobel prize winners. One of the primary goals of the workshop will be to enable participants to learn from each other about the key mathematical tools and techniques used to study ranking in other disciplines, which will hopefully facilitate meaningful collaborations that would help to shed new light on some of the existing questions in this area. In particular, it is our hope that collaborative work initiated at this AIM workshop would help bridge the gap between classical statistical approaches in ranking and newer machine learning approaches. We also hope to explore connections with other mathematical areas in which ranking is important, such as social choice theory and economics.

Given the current interest in ranking in many fields, a number of events related to ranking have been organized in recent years, including for example a workshop at the Neural Information Processing Systems (NIPS) conference in 2005 titled 'Learning to Rank'; a series of workshops at the ACM SIGIR Conference on Research and Development in Information Retrieval over the last three years titled 'Learning to Rank for Information Retrieval'; and two workshops at this year's NIPS 2009 conference titled 'Advances in Ranking' and 'Learning with Orderings'. However, these events have focused largely on individual disciplines: machine learning in the case of the NIPS workshops, and information retrieval in the case of the SIGIR workshops. The proposed AIM workshop will be the first event to bring together researchers from different disciplines to discuss questions related to ranking and to share their knowledge of the mathematical tools currently used in their disciplines in a collegial and collaborative setting. Indeed, the need for such an event was highlighted recently in Don Saari's 2008 article titled 'Mathematics and Voting' in the *Notices of the AMS* ('voting' is, of course, but a very special case of 'ranking'). It is our hope that this meeting will be the first step in helping us to piece together the jigsaw puzzle that we call 'The Mathematics of Ranking'.

1.1. **Examples of topics/approaches we hope to discuss:**

- Classical statistical approaches to ranking [22, 10]
- Algebraic approaches to ranking [10, 15]
- Geometric approaches to ranking [24]
- Markov chain and other graph-based approaches to ranking [25, 16, 2, 1]
- Machine learning approaches to ranking, including
  - ranking algorithms in machine learning
  - approaches for studying generalization properties of ranking algorithms using tools such as uniform convergence and algorithmic stability [3, 26, 4]
  - approaches for studying statistical convergence properties of ranking algorithms using U-statistics and other tools [7]
- Rank aggregation methods, including
  - probabilistic approaches [17, 18]
  - algorithmic and complexity-theoretic considerations [12]

- Connections with other approaches to ranking, such as those studied in social choice theory, economics, and game theory

1.2. **Participants:** As indicated above, the goal of this workshop is to bring together researchers from a broad range of disciplines relevant to ranking, for five days of intensive discussions, brainstorming, and forging collaborations. We have received overwhelming support for the idea of such an event; in particular, we have talked to all the potential participants listed in this proposal, and everyone has expressed enthusiasm and interest in participating.

1.3. **Format:** One goal of the workshop will be to enable participants to learn from each other the key tools and techniques from other disciplines. To this end, we will select a theme or class of approaches each day (such as classical statistical approaches one day, machine learning approaches another day, etc), and have presentations by participants in the corresponding areas during the mornings. This will give everyone an opportunity to learn about the main approaches in each area, as well as the main open questions in each area. The afternoons will then consist of smaller group interactions focusing on understanding better the connections between different approaches, and possibly making progress on some of the open questions as a result of these interactions.

1.4. **Expected outcomes:** At the end of the workshop, we hope that all participants will have learned about many different views of ranking; we also hope that the interactions themselves will lead to new and potentially more useful views of ranking than those currently understood. We expect that the focused nature of the interactions will lead to reformulations and solutions of some open questions related to ranking. More importantly, the bridges built during this workshop will play an important role in facilitating dialogue and collaboration between different disciplines related to ranking; this will be of tremendous value not only to researchers investigating mathematical aspects of ranking, but also to the increasing number of practitioners using ranking techniques in their applications.

1.5. **About the organizers:** Lek-Heng Lim [primary organizer] is a Morrey assistant professor in the Department of Mathematics at the University of California, Berkeley. He received his PhD in Computational and Mathematical Engineering from Stanford University in 2007 and was a Clare Hall Fellow at Cambridge University during 2000–2001; prior to this, he received an MS in Mathematics from Cornell University in 2000, and a BSc with Honors in Mathematics from the National University of Singapore in 1996. His primary research interests are in computational mathematics and in the analysis of large-scale data. He has organized several mathematical meetings and workshops in the past, including a number of SIAM minisymposia on various topics in algebra and data mining, an MSRI graduate workshop, and an AIM workshop in 2008 titled 'Geometry and Representation Theory of Tensors for Computer Science, Statistics, and Other Areas'.

Shivani Agarwal is a postdoctoral associate in the Computer Science and Artificial Intelligence Laboratory at MIT. She received her PhD in Computer Science from the University of Illinois at Urbana-Champaign in 2005, a BA with Honors in Computer Science from the University of Cambridge, where she was a Nehru Scholar at Trinity College, in 2000, and a BSc with Honors in Mathematics from the University of Delhi in 1998. Her primary research interests are in machine learning and statistical learning theory. She organized an early workshop on ranking methods in machine learning at the NIPS conference in 2005, and is organizing a second workshop titled 'Advances in Ranking' at NIPS 2009.

References

[1] A. Agarwal and S. Chakrabarti. Learning random walks to rank nodes in graphs. In *Proceedings of the 24th International Conference on Machine Learning*, 2007.

[2] S. Agarwal. Ranking on graph data. In *Proceedings of the 23rd International Conference on Machine Learning*, 2006.

[3] S. Agarwal, T. Graepel, R. Herbrich, S. Har-Peled, and D. Roth. Generalization bounds for the area under the ROC curve. *Journal of Machine Learning Research*, 6:393–425, 2005.

[4] S. Agarwal and P. Niyogi. Generalization bounds for ranking algorithms via algorithmic stability. *Journal of Machine Learning Research*, 10:441–474, 2009.

[5] N. Ailon, M. Charikar, and A. Newman. Aggregating inconsistent information: ranking and clustering. *J. ACM*, 55(5):Art. 23, 27, 2008.

[6] K. Arrow. *Social Choice and Individual Values*. 1951, 2nd ed., 1963.

[7] S. Clemencon, G. Lugosi, and N. Vayatis. Ranking and empirical minimization of U-statistics. *Annals of Statistics*, 36:844–874, 2008.

[8] W. W. Cohen, R. E. Schapire, and Y. Singer. Learning to order things. *Journal of Artificial Intelligence Research*, 10:243–270, 1999.
[9] D. Cossock and T. Zhang. Statistical analysis of Bayes optimal subset ranking. *IEEE Transactions on Information Theory*, 54(11):5140–5154, 2008.
[10] P. Diaconis. *Group Representations in Probability and Statistics*. Institute of Mathematical Statistics, Hayward, California, 1988.
[11] P. Diaconis. A generalization of spectral analysis with application to ranked data. *Ann. Statist.*, 17(3):949–979, 1989.
[12] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar. Rank aggregation methods for the web. In *Proceedings of the 10th International Conference on the World Wide Web*, 2002.
[13] Y. Freund, R. Iyer, R. E. Schapire, and Y. Singer. An efficient boosting algorithm for combining preferences. *Journal of Machine Learning Research*, 4:933–969, 2003.
[14] D. S. Hochbaum and A. Levin. Methodologies and algorithms for group-rankings decision. *MANAGEMENT SCIENCE*, 52(9):1394–1408, 2006.
[15] X. Jiang, L.-H. Lim, Y. Yao, and Y. Ye. Statistical ranking and combinatorial Hodge theory. Preprint, 2009.
[16] J. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604–632, 1999.
[17] A. Klementiev, D. Roth, and K. Small. Unsupervised rank aggregation with distance-based models. In *Proceedings of the 25th International Conference on Machine Learning*, 2008.
[18] A. Klementiev, D. Roth, K. Small, and I. Titov. Unsupervised rank aggregation with domain-specific expertise. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 2009.
[19] R. Kondor. The skew spectrum of functions on finite groups and their homogeneous spaces. Preprint, 2007.
[20] R. Kondor, A. Howard, and T. Jebara. Multi-object tracking with representations of the symmetric group. In *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics*, 2007.
[21] A. N. Langville and C. D. Meyer. *Google's PageRank and beyond: the science of search engine rankings*. Princeton University Press, Princeton, NJ, 2006.
[22] E. L. Lehmann. *Nonparametrics: Statistical Methods Based on Ranks*. Holden-Day, San Francisco, California, 1975.
[23] J. Morton, L. Pachter, A. Shiu, B. Sturmfels, and O. Wienand. Geometry of rank tests. In *Proceedings of the Third European Workshop on Probabilistic Graphical Models*, 2006.
[24] J. Morton, L. Pachter, A. Shiu, B. Sturmfels, and O. Wienand. Convex rank tests and semigraphoids. *SIAM Journal on Discrete Mathematics*, 23:1117–1134, 2009.
[25] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab, 1999.
[26] C. Rudin, C. Cortes, M. Mohri, and R. E. Schapire. Margin-based ranking meets boosting in the middle. In *Proceedings of the 18th Annual Conference on Learning Theory*, 2005.
[27] D. G. Saari. *Geometry of voting*, volume 3 of *Studies in Economic Theory*. Springer-Verlag, Berlin, 1994.
[28] D. G. Saari. Mathematics and voting. *Notices Amer. Math. Soc.*, 55(4):448–455, 2008.
[29] T. L. Saaty. A ratio scale metric and the compatibility of ratio scales: the possibility of Arrow's impossibility theorem. *Appl. Math. Lett.*, 7(6):51–57, 1994.
[30] T. L. Saaty. Multi-decisions decision-making: in addition to wheeling and dealing, our national political bodies need a formal approach for prioritization. *Math. Comput. Modelling*, 46(7-8):1001–1016, 2007.