

THE MATHEMATICS OF RANKING

organized by
Shivani Agarwal and Lek-Heng Lim

Workshop Summary

Introduction

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 [Arrow51]. In statistics, the study of ranked data has played an important role in the development of non-parametric statistics in particular [Lehmann75]; there have also been many rich results on studying rankings via probability models defined on permutation groups [Diaconis88]. 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 [Cohen+99, Freund+03, Agarwal+05, Rudin+05, Agarwal06, Clemencon+08, CossockZh08, AgarwalNi09, Jiang+09].

Mathematical topics that arise in ranking problems include representations of the symmetric group [Diaconis88, Diaconis89], discrete geometry of permutahedrons [Morton+06, Morton+09], order relations on posets [Arrow51], combinatorial Hodge theory [Jiang+09], harmonic analysis on finite groups and their homogeneous spaces [Kondor+07, Kondor07], algorithmic graph theory [Ailon+08, Dwork+02], Perron-Frobenius theory (in PageRank) [Page+99, Langville06], singular value decomposition (in HITS) [Kleinberg99], network flows [Hochbaum06], and convergence in probability [Agarwal+05, Rudin+05, Clemencon+08]. There are also studies of ranking that are clearly mathematical but difficult to categorize [Saari94, Saari08, Saaty94, Saaty07].

This workshop brought together researchers from a broad spectrum of both fundamental and applied areas to share their perspectives on ranking, to brainstorm, and to forge collaborations. The subject of ranking is truly multidisciplinary – with mathematics being the common thread. Our list of participants included not only academic colleagues from computer science, economics, mathematics, operations research, statistics, but also industry

colleagues from Google, Microsoft, and Yahoo!; and featured a seniority range going from graduate students to Nobel prize winners. One of the primary goals of the workshop was to enable participants to learn from each other about the key mathematical tools and techniques used to study ranking in other disciplines, and thus facilitate meaningful collaborations that would help to shed new light on some of the existing questions in this area. Another goal was that collaborative work initiated at this AIM workshop should help bridge the gap between classical statistical approaches in ranking and newer machine learning approaches. Yet a third goal was to explore connections with other mathematical areas in which ranking is important, such as social choice theory and economics.

This AIM workshop was 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 our meeting was a small but important first step in helping piece together the jigsaw puzzle that we call 'The Mathematics of Ranking'.

Proceedings

One goal of the workshop was to enable participants to learn from each other the key tools and techniques from other disciplines. To this end, we selected a theme or class of approaches each day (such as classical statistical approaches one day, machine learning approaches another day, etc), and had presentations by participants in the corresponding areas during the mornings. This gave everyone an opportunity to learn about the main approaches in each area, as well as the main open questions in each area. The afternoons then consisted of smaller group interactions focusing on understanding better the connections between different approaches, with some progress made on some of the open questions as a result of these interactions.

Monday.

Kenneth Arrow gave an hour-long lecture titled "Some Remarks on the Aggregation of Rankings," where he introduced his eponymous impossibility theorem. Donald Saari gave the second hour-long lecture titled "The Geometry of Voting."

Tuesday.

Devavrat Shah gave an introductory tutorial on how group representation theory arises when one studies ranking, followed by a brief description of some of the recent results from their group on inferring rankings from partial information. Thomas Saaty gave a talk about the Analytic Hierarchy Process that he invented.

Wednesday.

Amy Langville and David Gleich gave a joint lecture on the various aspects of PageRank. Dorit Hochbaum spoke about optimization algorithms for rank aggregation. Lek-Heng Lim gave a talk on applying Hodge decomposition to the analysis of rank aggregation.

Thursday.

Shivani Agarwal gave an tutorial introduction to ranking methods in machine learning. Yoram Singer described progress on label ranking problems in machine learning. Nicolas Vayatis spoke about statistical analysis of ranking in machine learning, and Lester Mackey gave a short talk on statistical consistency of ranking methods.

Friday.

The last day included a selection of short talks on recent topics in ranking in different fields, including machine learning (Yacov Crammer, Cynthia Rudin, Kevin Small, Purnamrita Sarkar), game theory (Yinyu Ye), representation theory (Michael Orrison, Risi Kondor), statistics (Tong Zhang), and finance (Peter Bartlett).

Problems discussed

The following is a selection of some of the more important problems that were discussed at the workshop (descriptions are based on scribe notes by Ngoc Mai Tran).

Statistical procedure to compare different rankings.

Posed by D. Gleich and A. Langville

Problem description

Suppose two different methods applied to a graph $G = (V, E)$ give rankings σ_1, σ_2 of the nodes. For instance, σ_1 is the output produced by a new ranking method, and σ_2 is the output produced by PageRank. In the current literature, one present some statistics (for instance, Kendall's tau) to measure the similarity of σ_1 and σ_2 , and thereby make claims on the similarity between the new method and PageRank. But this is just a measure of similarity of the output on one graph.

Question

Can we find rigorous statistical tests to compare two graph ranking methods?

Discussions

The question can be phrased more concretely as follows: given a graph $G = (V, E)$, let $f : V \rightarrow \mathbb{R}$ be a centrality measure, that is, a function which is invariant under graph isomorphisms. This function induces the ranking on the node. (For example, the PageRank function, the node degree function). Consider the behavior of this function over all possible graph: this gives us a distribution over all possible rankings \mathbb{S}_n where $n = |V|$ is the number of nodes. That is, f induces a ranking function $f^R : G \rightarrow \mathbb{S}_n$. Now the question is equivalent to finding statistical tests to compare the distributions of two functions f_1^R, f_2^R over all possible graphs G on n nodes.

We identified two issues: firstly, one may want to look at a subset of these graphs only. That is, given an instance of the graph (eg: the graph G), we only compare f_1^R, f_2^R over a subset of graphs on n nodes which are "similar" to G . Then the question is what is a reasonable subset of graphs to look at? We suggested, as an example, that one can look at random graphs of the same degree distribution. This can be obtained by simulations.

Secondly, one may want to model the distribution of f^R as one of the location-scale family, where a scale parameter measures the spread, and the location parameter corresponds to the “median permutation”, that is, the permutation that minimizes the sum of the distances to all the other permutations on the range of f^R . This problem is NP-hard in general, but if we make assumptions about the classes of the permutations to be summarized, it’s potentially tractable.

Clearly the choice of metric on \mathbb{S}_n is important. For example, if $f_1^R(G) = \sigma$ and $f_2^R(G) = \sigma'$ are distinctively different rankings (for example, one is a reverse of the other), then we want to say that f_1^R and f_2^R are very different (on G), while, if σ and σ' only differ by one transposition, then we want to say that f_1^R and f_2^R are very similar. Anne Shiu and Jason Morton pointed out that they encountered a similar problem of detecting permutations with cycles from random permutation in genetics, in which themselves and their coauthors have solved using techniques from algebraic geometry.

Continuing with a geometric approach, we noted that one can inscribe the permutahedron in the smallest sphere containing it. Then we can view the distribution over the permutahedron as distribution over the sphere, and one can measure the distance between two permutations by their geodesic distance on the sphere. This gives a natural metric on \mathbb{S}_n . Then, suppose that f_1^R and f_2^R have the same scale parameter. To compare their location parameters, note that if we assume rotational symmetry of the induced distributions, the permutation closest to the intrinsic mean approximates the minimizer of average Kendall tau distance. Thus an intrinsic or directional statistics based test for equality of the means can be used to decide if two distributions, and hence ranking procedures, are indistinguishable.

Outcome summary

Through the discussions, we learned of related work of Morton and Shiu. The problem was broken into smaller questions in which we made progress in solving, and naturally raised open questions one can consider and follow up.

References

<http://www.stanford.edu/group/datamining/mathofranking/problems/david-amy.txt> Problem write-up by D. Gleich and A. Langville.

J. Morton, L. Pachter, A. Shiu, B. Sturmfels, O. Wienand, <http://arxiv.org/abs/math/0702564> Convergence of rank tests and semigraphoids, *SIAM J. Discrete Math.*, 23 (2009), no. 3, pp. 11171134

Analyzing and modeling rank data, John I Marden, Chapman & Hall, 1995.

Recovering distribution over \mathbb{S}_n given partial information on λ -coefficients.

Posed by Srikanth Jagabathula and Devavrat Shah

Let $\mathcal{F} := \{f : \mathbb{S}_n \rightarrow \mathbb{R}_+\}$ be the space of functions with non-negative values on \mathbb{S}_n . Let $f \in \mathcal{F}$ be the (unknown) function we are interested in learning. Suppose that we know the Fourier transform of f : $\hat{f}_\lambda = \sum_{\sigma \in \mathbb{S}_n} f(\sigma) \rho_\lambda(\sigma)$ (note that ρ_λ denotes everything up to λ , not just one single irreducible representation). Let $|f|_0$ denote the sparsity of f , that is, the support of f on \mathbb{S}_n :

$$|f|_0 := \text{supp}(f) = \{\sigma \in \mathbb{S}_n : f(\sigma) \neq 0\}.$$

Question

Suppose that $|f|_0 = K$. For what values of K is the recovery of f from \hat{f}_λ not possible?

Background

The authors who posed this problem have proved the following bound on K . With respect to random model $R(K, T)$, the probability of error is uniformly bounded away from 0 for all n large enough and any λ , if

$$K \geq \frac{3D_\lambda^2}{n \log n} \left[\log \left(\frac{D_\lambda^2}{n \log n} \vee T \right) \right]. \quad (1)$$

However, they conjectured that a sharper bound $K \approx D_\lambda$ can be obtained.

Below is a detailed write-up by one of the authors, Srikanth Jagabathula, starting with some necessary information theory preliminaries and ending with a proof of the above theorem. *Information Theory Preliminaries*. Here we recall some necessary Information Theory preliminaries. Further details can be found in the book by Cover and Thomas.

Consider a discrete random variable X that is uniformly distributed over a finite set \mathcal{X} . Let X be *transmitted* over a *noisy* channel to a receiver; suppose the receiver receives a random variable Y , which takes values in a finite set \mathcal{Y} . Essentially, such “transmission over noisy channel” setup describes any two random variables X, Y defined through a joint probability distribution over a common probability space.

Now let $\hat{X} = g(Y)$ be an estimation of the transmitted information that the receiver produces based on the observation Y using some function $g : \mathcal{Y} \rightarrow \mathcal{X}$. Define probability of error as $p_{\text{err}} = \Pr(X \neq \hat{X})$. Since X is uniformly distributed over \mathcal{X} , it follows that

$$p_{\text{err}} = \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} \Pr(g(Y) \neq x | x). \quad (2)$$

Recovery of X is called asymptotically reliable if $p_{\text{err}} \rightarrow 0$ as $|\mathcal{X}| \rightarrow \infty$. Therefore, in order to show that recovery is not asymptotically reliable, it is sufficient to prove that p_{err} is bounded away from 0 as $|\mathcal{X}| \rightarrow \infty$. In order to obtain a lower bound on p_{err} , we use Fano’s inequality:

$$H(X|\hat{X}) \leq 1 + p_{\text{err}} \log |\mathcal{X}|. \quad (3)$$

Using (3), we can write

$$H(X) = I(X; \hat{X}) + H(X|\hat{X}) \quad (4)$$

$$\leq I(X; \hat{X}) + p_{\text{err}} \log |\mathcal{X}| + 1 \quad (5)$$

$$\stackrel{(a)}{\leq} I(X; Y) + p_{\text{err}} \log |\mathcal{X}| + 1 \quad (6)$$

$$= H(Y) - H(Y|X) + p_{\text{err}} \log |\mathcal{X}| + 1 \quad (7)$$

$$\leq H(Y) + p_{\text{err}} \log |\mathcal{X}| + 1, \quad (8)$$

where we used $H(Y|X) \geq 0$ for a discrete¹ valued random variable. The inequality (a) follows from the data processing inequality: if we have Markov chain $X \rightarrow Y \rightarrow \hat{X}$, then

¹The counterpart of this inequality for a continuous valued random variable is not true. This led us to study the limitation of recovery algorithm over model $R(K, T)$ rather than $R(K, \mathcal{C})$.

$I(X; \hat{X}) \leq I(X; Y)$. Since $H(X) = \log |\mathcal{X}|$, from (8) we obtain

$$p_{\text{err}} \geq 1 - \frac{H(Y) + 1}{\log |\mathcal{X}|}. \quad (9)$$

Therefore, to establish that probability of error is bounded away from zero, it is sufficient to show that

$$\frac{H(Y) + 1}{\log |\mathcal{X}|} \leq 1 - \delta, \quad (10)$$

for any fixed constant $\delta > 0$.

Proof of the theorem.

Proof. Our goal is to show that when K is large enough (in particular, as claimed in the statement of Theorem), the probability of error of *any* recovery algorithm is uniformly bounded away from 0. For that, we first fix a recovery algorithm, and then utilize the above setup to show that recovery is not asymptotically reliable when K is large. Specifically, we use (10), for which we need to identify random variables X and Y .

To this end, for a given K and T , let f be generated as per the random model $R(K, T)$. Let random variable X represent the support of function f i.e., X takes values in $\mathcal{X} = S_n^K$. Given λ , let $\hat{f}(\lambda)$ be the partial information that the recovery algorithm uses to recover f . Let random variable Y represent $\hat{f}(\lambda)$, the $D_\lambda \times D_\lambda$ matrix. Let $h = h(Y)$ denote the estimate of f , and $g = g(Y) = \text{supp} h$ denote the estimate of the support of f produced by the given recovery algorithm. Then,

$$\Pr(h \neq f) \geq \Pr(\text{supp} h \neq \text{supp} f) \quad (11)$$

$$= \Pr(g(Y) \neq X). \quad (12)$$

Therefore, in order to uniformly lower bound the probability of error of the recovery algorithm, it is sufficient to lower bound its probability of making an error in recovering the support of f . Therefore, we focus on

$$p_{\text{err}} = \Pr(g(Y) \neq X).$$

It follows from the discussion in Section that in order to show that p_{err} is uniformly bounded away from 0, it is sufficient to show that for some constant $\delta > 0$

$$\frac{H(Y) + 1}{\log |\mathcal{X}|} \leq 1 - \delta. \quad (13)$$

Observe that $|\mathcal{X}| = (n!)^K$. Therefore, using Stirling's approximation, it follows that

$$\log |\mathcal{X}| = (1 + o(1))Kn \log n. \quad (14)$$

Now $Y = \hat{f}(\lambda)$ is a $D_\lambda \times D_\lambda$ matrix. Let $Y = [Y_{ij}]$ with $Y_{ij}, 1 \leq i, j \leq D_\lambda$, taking values in $\{1, \dots, KT\}$; it is easy to see that $H(Y_{ij}) \leq \log KT$. Therefore, it follows that

$$H(Y) \leq \sum_{i,j=1}^{D_\lambda} H(Y_{ij}) \quad (15)$$

$$\leq D_\lambda^2 \log KT = D_\lambda^2 (\log K + \log T). \quad (16)$$

For small enough constant $\delta > 0$, it is easy to see that the condition of (13) will follow if K satisfies the following two inequalities:

$$\frac{D_\lambda^2 \log K}{Kn \log n} \leq \frac{1}{3}(1 + \delta) \quad \Leftrightarrow \quad \frac{K}{\log K} \geq \frac{3(1 - \delta/2)D_\lambda^2}{n \log n}, \quad (17)$$

$$\frac{D_\lambda^2 \log T}{Kn \log n} \leq \frac{1}{3}(1 + \delta) \quad \Leftrightarrow \quad K \geq \frac{3(1 - \delta/2)D_\lambda^2 \log T}{n \log n}. \quad (18)$$

In order to obtain a bound on K from (17), consider the following: for large numbers x, y , let $y = (c + \varepsilon)x \log x$, for some constants $c, \varepsilon > 0$. Then, $\log y = \log x + \log \log x + \log(c + \varepsilon)$ which is $(1 + o(1)) \log x$. Therefore,

$$\frac{y}{\log y} = \frac{c + \varepsilon}{1 + o(1)} x \geq cx, \quad (19)$$

for $x \rightarrow \infty$ and constants $c, \varepsilon > 0$. Also, observe that $y/\log y$ is a non-decreasing function; hence, it follows that for $y \geq (c + \varepsilon)x \log x$, $y/\log y \geq cx$ for large x . Now take $x = \frac{D_\lambda^2}{n \log n}$, $c = 3$, $\varepsilon = 1$ and $y = K$. Note that $D_\lambda \geq n$ for all λ of interest; therefore, $x \rightarrow \infty$ as $n \rightarrow \infty$. Hence, (17) is satisfied for the choice of

$$K \geq \frac{4D_\lambda^2}{n \log n} \left(\log \frac{D_\lambda^2}{n \log n} \right). \quad (20)$$

From (13), (17), (18), and (20) it follows that the probability of error of any algorithm is at least $\delta > 0$ for n large enough and any λ if

$$K \geq \frac{4D_\lambda^2}{n \log n} \left[\log \left(\frac{D_\lambda^2}{n \log n} \vee T \right) \right]. \quad (21)$$

□

Outcome summary

There were some discussions on this problem amongst Konder, Orrison and Jagabathula, however, there was no definitive outcome. The problem remains opened. Further description of the problem can be found at the AIM workshop website.

References

<http://www.stanford.edu/group/datamining/mathofranking/problems/devavrat.pdf> Problem write-up by Srikanth Jagabathula and Devavrat Shah

Rademacher complexity of kernel classes on \mathbb{S} .

ⁿ Posed by Risi Kondor

Question

What is the Rademacher complexity of kernel classes on \mathbb{S}_n ?

Background

A kernel K on \mathbb{S}_n is a symmetric positive semidefinite function $K : \mathbb{S}_n \times \mathbb{S}_n \rightarrow \mathbb{R}$. For example, $K(\sigma_1, \sigma_2)$ is the diffusion kernel induced by adjacent transpositions on the Caley Graph of S_n . Given a set $F := \{f : X \rightarrow \mathbb{R}\}$ consisting real-valued functions on some space

X , the Rademacher complexity of F is defined as

$$R_m(F) := E_{X_i, \epsilon_i} \left[\sup_{f \in F} \frac{1}{m} \sum_{i=1}^m \epsilon_i f(X_i) \right]$$

where ϵ_i are i.i.d Bernoulli. Reproducing kernel Hilbert spaces (RKHS) induced by K is the set of functions $F_K = \{f : S_n \rightarrow \mathbb{R} : f(x) = \sum_{i=1}^m \alpha_i K(x, x_i) \text{ for some } m, x_i, \alpha_i\}$. It is known that if $F_K^B = \{f \in F_K : \|f\|_K \leq B\}$, then

$$E[R_m(F_K^B)] \leq \frac{2B}{m} E\left(\sqrt{\sum_i K(X_i, X_i)}\right)$$

Discussions

There were discussions on the role of kernel in ranking. Risi Kondor gave some motivations in his talk, and also in the following small example: suppose we have n partial rankings: $\sigma_1, \dots, \sigma_n$ of n different people ranking on movies. Now suppose we have a new user, who told us that she prefers A to B to C . Then, taking $A > B > C$ as the input, should we recommend $D > E$ or $E > D$? This problem leads to the problem of learning on permutations with sparse data, in which one approach is regularization. Specifically, a kernel induce a RKHS, which gives a sense of smoothness, and has been used for regularization for other set of functions in machine learning. In this particular instance, the space of functions we want to regularize is the space of functions on the symmetric group $\{f : S_n \rightarrow \mathbb{R}\}$. By regularization, we mean replacing the original loss function $\sum_i L(f(\sigma_i))$ by its regularized version $R(f) := \sum_i L(f(\sigma_i)) + \|f\|_{\mathcal{H}}^2$ where $\|f\|_{\mathcal{H}}^2$ is the norm of f induced by the norm on the RKHS corresponding to a chosen kernel K .

For example, $L(f(\sigma_i))$ can be the multi class hinge loss. This is a multi class support vector machine (SVM). Since f is in the RKHS, the solution is of the form $f(\sigma) = \sum_i \alpha_i K(\sigma_i, \sigma)$. This approach has the advantage that the kernel ignores the symmetric group S_n , and converts the problem to a usual SVM problem.

In order to get a bound on the convergence rate of the estimator, one needs a VC or Rademacher bound on the RKHS. Here, Lester Mackay pointed out that from lecture notes of Peter Bartlett, the problem $\min_{f \in \mathcal{H}} \frac{1}{n} \sum_i L(f(\sigma_i)) + \|f\|_{\mathcal{H}}^2$ is equivalent to $\min_{\|f\|_{\mathcal{H}}^2 \leq B} \frac{1}{n} \sum_i L(f(\sigma_i))$, and one has

$$\hat{R}_n(F_B) \leq \frac{B}{\sqrt{n}} \sqrt{\frac{1}{n} \sum_i K(\sigma_i, \sigma_i)}, \quad \text{and } R_n(F_B) \leq B \sqrt{\sum_{i=1}^{\infty} \lambda_i / n}.$$

This result, together with the bound

$$E \sup_{f \in \mathcal{F}} [E(L(f)) - 1/m \sum_i L(f, \sigma_i)] \leq 2R_n(\mathcal{F}_B)$$

in Kondor's notes leads to a solution to the problem.

Outcome summary

With the discovery of a result from Peter Bartlett's lecture notes coupled with Kondor's own results, the problem has been solved. Kondor said this will appear in his next paper.

References

<http://www.cs.berkeley.edu/~bartlett/courses/281b-sp08/Stat281B> Lecture Notes, Spring 2009, Peter Bartlett

Non convex surrogates for ranking (and resulting information patterns).

Posed by Yoram Singer

The goal is to improve on the bound appeared in the joint work of Peter Bartlett with Sylvain Arlot, Mike Jordan, Jon McAuliffe and Mikhail Traskin.

Background

Suppose we have a loss function we are interested in minimizing. For example:

$$\sum_{i,j} a_{ij} 1_{f(x_i) - f(x_j) < 0}.$$

Finding f^* which optimizes the loss is a NP hard problem. So we consider minimizing a surrogate (a relaxation of the loss). Suppose that the minimizer of the surrogate function is \hat{f} . The aforementioned work of Bartlett and his coauthors gave conditions for choices of the surrogate function in which one have $\hat{f} = f^*$ (in some asymptotic sense).

Outcome summary

Peter Bartlett gave a talk on his result in classification. Lester Mackay gave a talk on his work in a similar problem in ranking context. However, due to time constraint, there was little further discussions on this topic. The problem remains open.

References

Sylvain Arlot, Peter Bartlett, Mike Jordan, Jon McAuliffe and Mikhail Traskin

<http://www.stat.berkeley.edu/~bartlett/talks/ims2008.pdf> Convex methods for classification. IMS Medallion Lecture. June 2008.

Connection between AHP (analytic hierarchical process) and Borda Count.

Posed and commented on by Donald Saari

Question

What is the relationship between AHP and pairwise voting?

Discussions

Recall that AHP works with matrices with weights

$$\begin{bmatrix} 1 & a_{12} & a_{13} & \dots \\ \frac{1}{a_{12}} & 1 & a_{23} & \dots \\ \frac{1}{a_{13}} & \frac{1}{a_{23}} & 1 & \dots \end{bmatrix}$$

Can we connect these weights in AHP and pairwise voting? The answer is *yes*. Suppose, that i, j are candidates, and candidate i gets one million votes, candidate j gets 500 thousand votes. Then by defining a_{ij} to be is the exponential of the difference in tally

$$a_{ij} = \exp(1000000 - 500000) = \exp(500000),$$

we can form a matrix of weights in AHP. The consistency condition in AHP ($a_{ij}a_{jk} = a_{ik}$) naturally corresponds to the consistency in voting by taking log:

$$((\text{tally for } i) - (\text{tally for } j)) \times (\text{tally for } j - \text{tally for } k) = (\text{tally for } i - \text{tally for } k).$$

So this means a question on AHP can be translated into a question on social choice (by log) and then mapped back (by exp).

Outcome

This is an important revelation, and Donald Saari plans to publish a paper on this topic.

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. Indeed, the interactions at the workshop already led to reformulations and solutions of some open questions related to ranking, and we anticipate some publications will result from these. More importantly, the bridges built during this workshop will play an important role in facilitating dialogue and collaboration between researchers in 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.

Acknowledgement

The organizers thank Ngoc Mai Tran for her excellent scribing.

Bibliography

- [AgarwalCh07]A. Agarwal and S. Chakrabarti. Learning random walks to rank nodes in graphs. In *Proceedings of the 24th International Conference on Machine Learning*, 2007.
- [Agarwal06]S. Agarwal. Ranking on graph data. In *Proceedings of the 23rd International Conference on Machine Learning*, 2006.
- [Agarwal+05]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.
- [AgarwalNi09]S. Agarwal and P. Niyogi. Generalization bounds for ranking algorithms via algorithmic stability. *Journal of Machine Learning Research*, 10:441–474, 2009.
- [Ailon+08]N. Ailon, M. Charikar, and A. Newman. Aggregating inconsistent information: ranking and clustering. *J. ACM*, 55(5):Art. 23, 27, 2008.
- [Arrow51]K. Arrow. *Social Choice and Individual Values*. 1951, 2nd ed., 1963.
- [Clemencon+08]S. Clemencon, G. Lugosi, and N. Vayatis. Ranking and empirical minimization of U-statistics. *Annals of Statistics*, 36:844–874, 2008.
- [Cohen+99]W. W. Cohen, R. E. Schapire, and Y. Singer. Learning to order things. *Journal of Artificial Intelligence Research*, 10:243–270, 1999.
- [CossockZh08]D. Cossock and T. Zhang. Statistical analysis of Bayes optimal subset ranking. *IEEE Transactions on Information Theory*, 54(11):5140–5154, 2008.
- [Diaconis88]P. Diaconis. *Group Representations in Probability and Statistics*. Institute of Mathematical Statistics, Hayward, California, 1988.
- [Diaconis89]P. Diaconis. A generalization of spectral analysis with application to ranked data. *Ann. Statist.*, 17(3):949–979, 1989.

- [Dwork+02]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.
- [Freund+03]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.
- [Hochbaum06]D. S. Hochbaum and A. Levin. Methodologies and algorithms for group-rankings decision. *MANAGEMENT SCIENCE*, 52(9):1394–1408, 2006.
- [Jiang+09]X. Jiang, L.-H. Lim, Y. Yao, and Y. Ye. Statistical ranking and combinatorial Hodge theory. Preprint, 2009.
- [Kleinberg99]J. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604–632, 1999.
- [Klementiev+08]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.
- [Klementiev+09]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.
- [Kondor07]R. Kondor. The skew spectrum of functions on finite groups and their homogeneous spaces. Preprint, 2007.
- [Kondor+07]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.
- [Langville06]A. N. Langville and C. D. Meyer. *Google’s PageRank and beyond: the science of search engine rankings*. Princeton University Press, Princeton, NJ, 2006.
- [Lehmann75]E. L. Lehmann. *Nonparametrics: Statistical Methods Based on Ranks*. Holden-Day, San Francisco, California, 1975.
- [Morton+06]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.
- [Morton+09]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.
- [Page+99]L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab, 1999.
- [Rudin+05]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.
- [Saari94]D. G. Saari. *Geometry of voting*, volume 3 of *Studies in Economic Theory*. Springer-Verlag, Berlin, 1994.
- [Saari08]D. G. Saari. Mathematics and voting. *Notices Amer. Math. Soc.*, 55(4):448–455, 2008.
- [Saaty94]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.

[Saaty07]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.