RC24107 (W0611-055) November 13, 2006
Computer Science

IBM Research Report

A Winner-Loser Labeled Tournament Has at Most Twice as Many Outdegree Misrankings as Pair Misrankings

Nikhil Bansal, Don Coppersmith*, Gregory B. Sorkin
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

*IDA Center for Communications Research
Princeton, NJ 08540

A WINNER–LOSER LABELED TOURNAMENT HAS AT MOST TWICE AS MANY OUTDEGREE MISRANKINGS AS PAIR MISRANKINGS

NIKHIL BANSAL, DON COPPERSMITH*, AND GREGORY B. SORKIN

ABSTRACT. In any tournament, with the players partitioned any way into two groups called Winners and Losers, we define two measures: f is the number of vertex pairs consisting of a labeled "loser" and a "winner" where the loser beats the winner, and similarly g is the number of such pairs where the loser has at least as many total wins as the winner. We prove that $g \leq 2f$, and this bound is tight. The result has a natural interpretation and easy generalization in the domain of majorization.

1. Introduction and Main Result

The paper [BBLS06] shows a reduction from ranking to classification: roughly, a machine which can learn (with classification regret r) which item from a pair dominates the other, can be used to rank items from strongest to weakest (with ranking "AUC" regret at most $4r$). (It also shows that ranking error rate r results in AUC error at most $4r$; see [BBLS06] for definitions and precise statements.) The present paper's Theorem 1 replaces a key technical element of [BBLS06], improving its constant of 4 to 2 (already shown in [BBLS06] to be the best possible), and thus showing that classification regret (or error rate) r results in ranking regret (respectively, error) at most $2r$. Theorems 3 and 4 here are equivalent statements to Theorem 1, while Corollary 5 provides an elegant generalization.

Let T be a tournament, i.e., a complete graph in which each edge is directed one way or the other, so that for every pair of vertices $i \neq j$, either $i \to j$ is an edge or $j \to i$ is an edge, but not both. There are no loops (no edges $i \to i$). We write $d(i)$ for the *outdegree* of vertex i, so $d(i) = \sum_j \mathbf{1}(i \to j)$, where the indicator function $\mathbf{1}(i \to j)$ is 1 if T has an edge $i \to j$ and 0 otherwise.

Let the vertices of T be partitioned into a set W of "winners" and a set L of "losers"; equivalently, each vertex is labeled as either a winner or a loser. Call the triple (T, W, L) a winner–loser partitioned tournament, and denote it by $\mathbf{T}$. (We imagine that typically winners beat losers, so there are many $w \to l$ pairs but few $l \to w$ ones, and correspondingly that winners typically have large outdegree and losers small outdegree. However, we allow arbitrary labelings of the vertices.) Our aim is to show, roughly, that for *any* (T, W, L), the number of winner–loser pairs where the loser outranks the winner is at most twice the number of such winner–loser pairs where the loser beats the winner.

More concretely, define two functions:

$$f(\mathbf{T}) := \sum_{\ell \in L} \sum_{w \in W} \mathbf{1}(\ell \to w) \tag{1}$$

$$= \sum_{\ell \in L} \deg(\ell) - \binom{|L|}{2} \tag{2}$$

$$g(\mathbf{T}) := \sum_{\ell \in L} \sum_{w \in W} \mathbf{1}(\deg(\ell) \geq \deg(w)). \tag{3}$$

The first equality holds because (1), the number of edges from L to W, is equal to (2), the total number of edges out of L minus the number of edges from L to L. Expression (2) will be easier to work with.

Theorem 1. *For any winner–loser partitioned tournament* $\mathbf{T}$, $g(T, W, L) \leq 2f(T, W, L)$.

Rather than working with a (labeled) tournament, a relatively complex object, the reformulation of (1) as (2) allows us to work with a (labeled) degree sequence. Landau's theorem says that there exists a tournament with outdegree sequence $d_1 \leq d_2 \leq \cdots \leq d_n$ if and only if, for all $1 \leq i \leq n$, $\sum_{j=1}^{i} d_j \geq \sum_{j=1}^{i} (j-1)$, with equality for $i = n$.

* Research supported in part by EPSRC grant GR/S26323/01.

Using this formulation it is easy to construct an example achieving the bound given by Theorem 1.

Example 2. *With n odd, let every vertex have degree $(n-1)/2$; note that the degree sequence $\langle \frac{n-1}{2}, \ldots, \frac{n-1}{2} \rangle$ does indeed respect Landau's condition. Label $(n-1)/2$ of the vertices as winners and $(n+1)/2$ as losers. This gives $f = \frac{n+1}{2} \cdot \frac{n-1}{2} - \binom{(n+1)/2}{2} = (n+1)(n-1)/8$, while $g = \frac{n+1}{2} \cdot \frac{n-1}{2} = (n+1)(n-1)/4 = 2f$.*

Recall that a sequence $\langle a_1, \ldots, a_n \rangle$ is *majorized* by $\langle b_1, \ldots, b_n \rangle$ if the two sums are equal and if, when each sequence is sorted in non-increasing order, the prefix sums of the b sequence dominate those of the a sequence. (For a comprehensive treatment of majorization, see [MO79].) Landau's condition is precisely that $\langle d_1, \ldots, d_n \rangle$ is majorized by $\langle 0, \ldots, n-1 \rangle$. (With the sequences sorted in increasing order, Landau's condition is that prefix sums of the degree sequence dominate those of the progression, which is the same as saying that the suffix sums of the degree sequence are dominated by the suffix sums of the progression.) This allows us to take advantage of well-known properties of majorization, notably that if A' is obtained by averaging together any elements of A, then A majorizes A'.

This allows us to re-state Theorem 1 in terms of a sequence and majorization, rather than a tournament, but first we relax the constraints. First, where the original statement requires elements of the degree sequence to be non-negative integers, we allow them to be non-negative reals. Second, the original statement requires that we attach a winner/loser label to each element of the degree sequence. Instead, we will aggregate equal elements of the degree sequence, and for a degree d_i of (integral) multiplicity m_i, assign arbitrary non-negative (but not necessarily integral) portions to winners and losers: $w_i + \ell_i = m_i$. Note that the majorization condition applies only to the "compressed sequence" $\{d_i, m_i\}$, not the labeling.

Theorem 3. *For any winner–loser labeled compressed sequence $\mathbf{D} = (D, W, L)$ where D is majorized by $\langle 0, \ldots, n-1 \rangle$, $g(\mathbf{D}) \le 2f(\mathbf{D})$.*

Proof. We begin with an outline of the proof. Define a compressed sequence $\mathbf{D}$ as being *canonical* if it consists of at most three degrees: a smallest degree d_1 having only losers ($w_1 = 0$), a middle degree d_2 potentially with both winners and losers ($w_2, \ell_2 \ge 0$), and a largest degree d_3 having only winners ($\ell_3 = 0$). We first establish that any canonical sequence has $g(\mathbf{D}) - 2f(\mathbf{D}) \le 0$. We then show how to transform *any* degree sequence to a canonical one with a larger (or equal) value of $g - 2f$, which will complete the argument.

We first show that a canonical sequence $\mathbf{D}$ has $g - 2f \le 0$. For the canonical configuration, $g = w_2 \ell_2$ and $f = \ell_1 d_1 + \ell_2 d_2 - \binom{\ell_1 + \ell_2}{2}$, and hence our goal is to show that

$$\ell_1 d_1 + \ell_2 d_2 \ge (\ell_1 + \ell_2)(\ell_1 + \ell_2 - 1)/2 + w_2 \ell_2 / 2 \tag{4}$$

By Landau's condition applied to ℓ_1 and $\ell_1 + w_2 + \ell_2$, we have the following two relations. Observe that ℓ_1 and $\ell_1 + \ell_2 + w_2$ are both integers.

$$\ell_1 d_1 \ge \binom{\ell_1}{2} \tag{5}$$

and

$$\ell_1 d_1 + (\ell_2 + w_2) d_2 \ge \binom{\ell_1 + w_2 + \ell_2}{2} \tag{6}$$

Multiplying (5) by $w_2/(\ell_2 + w_2)$ and (6) by $\ell_2/(\ell_2 + w_2)$ and adding them, we obtain that

$$\ell_1 d_1 + \ell_2 d_2 \ge \frac{1}{\ell_2 + w_2} \left(w_2 \binom{\ell_1}{2} + \ell_2 \binom{\ell_1 + \ell_2 + w_2}{2} \right). \tag{7}$$

A simple calculation shows that the right side of inequality (7) is exactly equal to the right hand side of (4). This proves that $g \le 2f$ for a canonical sequence.

If a sequence is not canonical then there are two consecutive degrees d_i and d_j ($j = i+1$) such that one of the following cases holds. In each case we will apply a transformation producing from the degree sequence $\mathbf{D}$ a new degree sequence $\mathbf{D}'$, where:

- the total weight of winners in $\mathbf{D}'$ is equal to that of $\mathbf{D}$;
- the same is true for losers (and thus for the total weight);
- the total weight on each degree remains integral;
- $\mathbf{D}'$ maintains the majorization needed for Landau's theorem;
- the value of $g - 2f$ is at least as large for $\mathbf{D}'$ as for $\mathbf{D}$; and

- either the number of nonzero values w_i and ℓ_i or the number of distinct degrees d_i is strictly smaller for $\mathbf{D}'$ than for $\mathbf{D}$, and the other is no larger for $\mathbf{D}'$ than for $\mathbf{D}$ (for Transformation 2, this may first require application of another transformation).

We first sketch the cases and then detail the transformations.

Case 1a: d_i has only winners ($l_i = 0$).
Apply Transformation 1a, combining the two degrees into one.

Case 1b: d_j has only losers ($w_j = 0$).
Apply Transformation 1b, combining the two degrees into one.

Case 2: All of w_i, l_i, w_j and l_j are nonzero.
Apply Transformation 2, leaving the degrees the same but transforming the weights so that one of them is equal to 0 and one of the preceding cases applies, or the weights obey an equality allowing application of Transformation 3, which combines the two degrees into one.

Transformation 1a: In Case 1a, where d_i has only winners, change $\mathbf{D}$ to a new sequence $\mathbf{D}'$ by replacing the pair $(d_i, w_i, 0)$, (d_j, w_j, l_j) by their "average": the single degree (d', w', l'), where

$$w' = w_i + w_j, \quad l' = l_j, \quad d' = \frac{(w_i)d_i + (w_j + l_j)d_j}{w_i + w_j + l_j}.$$

The stated conditions on a transformation are easily checked. The total weight of winners is clearly preserved, as is the total weight of losers and the total degree (out-edges). Summing weights preserves integrality. The number of distinct degrees is reduced by one, and the number of nonzero weights may be decreased by one or may remain unchanged. The Landau majorization condition holds because $\mathbf{D}'$, as an averaging of $\mathbf{D}$, is majorized by it, and majorization is transitive. The only non-trivial condition is the non-decrease in $g - 2f$. The number of losers outranking winners is increased by $l_i w_j$, so $g(\mathbf{D}) \leq g(\mathbf{D}')$. Also, f depends only on the total weight of losers (which is unchanged) and on the average degree of losers. This average degree would be unchanged if w_i were 0; since $w_i \geq 0$, the average degree may decrease. Thus $f(\mathbf{D}) \geq f(\mathbf{D}')$, and $(g - 2f)(\mathbf{D}) \leq (g - 2f)(\mathbf{D}')$, as desired.

Transformation 1b: Symmetrically to Transformation 1a, obtain $\mathbf{D}'$ by replacing the pair of labeled weighted degrees (d_i, w_i, l_i) and $(d_j, 0, l_j)$ with a single one (d', w', l'), where

$$w' = w_i, \quad l' = l_i + l_j, \quad d' = \frac{(l_i + w_i)d_i + (l_j)d_j}{l_i + w_i + l_j}.$$

Transformation 2: Where w_i, l_i, w_j and l_j are all nonzero, we will begin with one case, which will lead to one other. In the usual case, we transform $\mathbf{D}$ to $\mathbf{D}'$ by replacing the pair (d_i, w_i, l_i), (d_j, w_j, l_j) with

$$(d_i, w_i + x, l_i - x), \quad (d_j, w_j - x, l_j + x),$$

for some value of x (positive or negative) to be determined. This affects only the labeling, not the weighted degree sequence itself, and is therefore legitimate as long as the four quantities $w_i + x$, $l_i - x$, $w_j - x$ and $l_j + x$ are all non-negative.

Defining $\Delta = (g - 2f)(\mathbf{D}') - (g - 2f)(\mathbf{D})$, we wish to choose x to make $\Delta > 0$.

$$\begin{aligned}\Delta &= \Big\{\big[(l_j + x)(w_i + x + w_j - x) + (l_i - x)(w_i + x)\big] - \big[l_j(w_i + w_j) + l_i w_i\big]\Big\} \\ &\quad - 2\Big\{\big[(l_i - x)d_i + (l_j + x)d_j\big] - \big[l_i d_i + l_j d_j\big]\Big\} \\ &= x(w_j + l_i - 2(d_j - d_i) - x) \\ &= x(a - x),\end{aligned}$$

where $a = w_j + l_i - 2(d_j - d_i)$. This is a simple quadratic expression with negative coefficient of x^2, so its value increases monotonically as x is varied from 0 to $a/2$, where the maximum is obtained. (Note that a may be negative.) If $a = 0$ then this transformation makes no change, and instead we use Transformation 3, below. Otherwise, vary x from 0 to $a/2$ stopping when x reaches $a/2$ or when any of $w_i + x$, $l_i - x$, $w_j - x$ and $l_j + x$ becomes 0; call this value $x^\star$.

If any of $w_i + x$, $l_i - x$, $w_j - x$ and $l_j + x$ is 0 then the number of nonzero weights is decreased (while the number of distinct degees is unchanged). Otherwise, $x^\star = a/2$. In that case, the new $\mathbf{D}'$ has $a = 0$ (because another application of the same procedure would give $x^\star = 0$). With $a = 0$ we apply Transformation 3, which reduces the number of nonzero weights.

Transformation 3. Similar to Cases 1a and 1b, transform $\mathbf{D}$ to $\mathbf{D}'$ by replacing the pair (d_i, w_i, l_i), (d_j, w_j, l_j) with a single degree (d', w', l') that is their weighted average,

$$w' = w_i + w_j, \quad l' = l_i + l_j, \quad d' = \frac{(w_i + l_i)d_i + (w_j + l_j)d_j}{w_i + l_i + w_j + l_j}.$$

This gives

$$\begin{aligned} \Delta &:= (g - 2f)(\mathbf{D}') - (g - 2f)(\mathbf{D}) \\ &= (l_i w_j) - 2(l_i d' + l_j d' - l_i d_i - l_j d_j) \\ &= l_i w_j + \frac{2(d_j - d_i)(w_i l_j - w_j l_i)}{w_i + l_i + w_j + l_j}. \end{aligned}$$

We will apply this transformation only in the case where Transformation 2 fails to give any improvement because its "a" expression is equal to 0, i.e., $d_j - d_i = (w_i + l_j)/2$. Making the corresponding substitution gives

$$\begin{aligned} \Delta &= l_i w_j + \frac{(w_i + l_j)(w_i l_j - w_j l_i)}{w_i + l_i + w_j + l_j} \\ &= \frac{(l_i w_j)(l_i + w_j) + (l_j w_i)(l_j + w_i)}{w_i + l_i + w_j + l_j} \\ &> 0. \end{aligned}$$

This reduces the number of distinct degrees by one, without increasing the number of nonzero weights.

Concluding the argument, we have shown that any non-canonical configuration $\mathbf{D}$ can be replaced by a configuration with a strictly smaller total of distinct degrees and nonzero weights, and at least as large a value of $g - 2f$. Since $\mathbf{D}$ had at most n distinct degrees and $2n$ nonzero weights originally, a canonical configuration $\mathbf{D}^\star$ is reached after at most $3n - 1$ transformations. (All that is important is that the number of transformations is finite: that a canonical configuration is eventually reached.) Then, $(g - 2f)(\mathbf{D}) \leq (g - 2f)(\mathbf{D}^\star) \leq 0$. □

2. Another Interpretation and a Generalization

We begin with another interpretation of the main theorem, and then state and prove an elegant generalization.

Rank the vertices of T in order of increasing outdegree, breaking ties arbitrarily. For a vertex v, let $\text{rank}(v)$ denote the number of vertices that appear before v in this ordering. Then, per (3), $g = \sum_{\ell \in L} \text{rank}(\ell) - \binom{|L|}{2}$. Also, recall (2), $f = \sum_{\ell \in L} \deg(\ell) - \binom{|L|}{2}$. This leads to the following equivalent statement of Theorem 1.

Theorem 4. *For any tournament T and any subset L ("losers") of its vertices,*

$$\sum_{\ell \in L} \deg(\ell) \geq \frac{1}{2} \left(\sum_{\ell \in L} \text{rank}(\ell) + \binom{|L|}{2} \right). \tag{8}$$

Corollary 5. *Let D be any element-wise non-decreasing vector majorized by $I = \langle 0, 1, \ldots, n-1 \rangle$. Let H be any non-negative n-vector (not necessarily increasing), and let I' be the re-ordering of I which minimizes the dot product IH. Then*

$$HD \geq \frac{1}{2}(HI + HI').$$

Proof. We first show that if H is a 0-1 vector then the corollary follows directly from our main theorem, and then show how to reduce the general case to the case of 0-1 vectors.

Given an arbitrary 0-1 vector H, let L (the set of losers) be the positions j where $H(j) = 1$. Then $HD = \sum_{\ell \in L} \deg(\ell)$, $HI = \sum_{\ell \in L} \text{rank}(\ell)$, and $HI' = \binom{|L|}{2}$, and Theorem 4 implies the desired inequality (8).

If H is an arbitrary non-negative vector, it is uniquely expressible as $H = \sum_i a_i v_i$, where each $a_i > 0$ and the v_i are 0-1 vectors such that $v_1 > v_2 > \cdots$. Here $v_i > v_j$ indicates that v_i is at least as large as v_j

coordinate-wise, and the strict inequality means that there is some coordinate k where $v_i(k) = 1$ but $v_j(k) = 0$. (Interpreting the vectors v_i as indicators, the corresponding sets $S(v_i)$ satisfy $S(v_1) \supsetneq S(v_2) \supsetneq \cdots$.) Such a decomposition of H follows by subtracting the smallest non-zero entry a_1 of $H_1 := H$ from all the non-zero entries of H_1 (indicated by v_1) to yield H_2; subtracting H_2's smallest non-zero entry a_2 from its non-zeros (indicated by v_2) to yield H_3; and so forth until the 0 vector is obtained.

Note that HI' is identical to sorting H in non-decreasing order (call it H') and multiplying it by I. The crucial property of our decomposition is that the *same* permutation applied to sort H as H', applied to each v_i to yield v_i', also makes each v_i' non-decreasing; this property just follows from the nestedness of the vectors v_i. Thus, $H' = \sum_i a_i(v_i)'$, where each v_i' is a non-increasing 0-1 vector.

Since v_i is a 0-1 vector, we know that $v_i D \geq \frac{1}{2}(v_i I + v_i' I)$. Thus

$$HD = \sum_i a_i v_i D \geq \sum_i a_i \frac{1}{2}(v_i I + v_i' I) = \frac{1}{2}(HI + H'I) = \frac{1}{2}(HI + HI').$$

□

References

[BBLS06] Maria-Florina Balcan, Alina Beygelzimer, John Langford, and Gregory B. Sorkin, *Robust reductions from ranking to classification*, Tech. report, IBM Research Report, November 2006.

[MO79] Albert W. Marshall and Ingram Olkin, *Inequalities: theory of majorization and its applications*, Mathematics in Science and Engineering, vol. 143, Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1979. MR MR552278 (81b:00002)

(Nikhil Bansal) Department of Mathematical Sciences, IBM T.J. Watson Research Center, Yorktown Heights NY 10598, USA

E-mail address: bansal@us.ibm.com

(Don Coppersmith) IDA Center for Communications Research, Princeton NJ 08540, USA

E-mail address: dcopper@idaccr.org

(Gregory B. Sorkin) Department of Mathematical Sciences, IBM T.J. Watson Research Center, Yorktown Heights NY 10598, USA

E-mail address: sorkin@watson.ibm.com