Theoretical Computer Science 636 (2016) 66–76

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Ranking chain sum orders ☆

Franz J. Brandenburg *, Andreas Gleißner

University of Passau, 94032 Passau, Germany

ARTICLE INFO

Article history:
Received 8 January 2015
Received in revised form 7 March 2016
Accepted 18 May 2016
Available online 24 May 2016
Communicated by G.F. Italiano

Keywords:
Total and partial orders
Distance measures
Ranking
$\mathcal{NP}$-hardness
Fixed-parameter complexity

ABSTRACT

Ranking information is an important topic in information sciences, Internet searching, voting systems, and sports. In the full information approach, a ranking is a total order of the candidates. We compare two rankings by pairwise comparisons under the nearest neighbor Kendall tau distance and study the distance and rank aggregation problems.

In many settings, the information is incomplete and a ranking is given by a partial order. A bucket order is obtained if a ranking allows ties and treats tied candidates as equivalent. Then the distance and rank aggregation problems can be solved efficiently in almost linear time. A chain sum order is complementary to a bucket order and consists of a set of disjoint total orders. Its width and height is the number and the maximum size of the total orders, respectively.

We show that the distance and rank aggregation problems of a total order and a chain sum order of bounded width or of height at most two can be solved in polynomial time and are $\mathcal{NP}$-complete for a total and a chain sum order of height at least 12. Both problems remain $\mathcal{NP}$-complete for a total and a heap order which is the partial order obtained from a single-elimination tournament (in sports). However, the problems are fixed-parameter tractable with respect to the distance and the Ulam distance, but are fixed-parameter intractable with respect to the order dimension.

1. Introduction

The problem of comparing inconsistent information from different sources appears in many contexts and disciplines. It arises from conflicting preferences and disagreements on the ranking of candidates. The information comes from an individual voter or an athletic competition and is based on a pairwise comparison of two candidates x and y with the outcome $x < y$ if x is better than y or x defeats y. For ranking problems the $<$-relation is assumed to be transitive, which does not hold in sports. The objective is to measure the disagreement between several rankings and to find a consensus ranking. These problems are called the *distance* and the *rank aggregation problems*, respectively.

Our approach is comparison based, in contrast with score based aggregation, which ranks candidates by the number of votes, points, or won prize money. With full information there is a comparison between any two candidates, which results in a total order of the candidates. Two total orders σ and τ are compared by the *Kendall tau distance* $K(\sigma, \tau)$. Given a set $\mathcal{D}$ of n candidates, $K(\sigma, \tau) = |\{\{x, y\} : x, y \in \mathcal{D}, \sigma(x) < \sigma(y)$ but $\tau(x) > \tau(y)\}$. Thus, $K(\sigma, \tau)$ counts the number of disagreements

☆ Supported by the Deutsche Forschungsgemeinschaft (DFG), Grant Br835/16-1.

* Corresponding author.
E-mail addresses: brandenb@informatik.uni-passau.de (F.J. Brandenburg), gleissner@fim.uni-passau.de (A. Gleißner).

http://dx.doi.org/10.1016/j.tcs.2016.05.026

on the ranking of two candidates. The Kendall tau distance of two total orders can be computed efficiently in almost linear time [6,26]. The currently best bound is $O(n \log n / \log \log n)$ [5]. Clearly, there are other distances for rankings [3,14,19,24].

1.1. Classes of partial orders

Comparing and ranking information has been generalized to *partial orders* where some candidates are unrelated. Then two candidates x and y are *tied* and are treated as equivalent, or they are *incomparable*, such as “apples and oranges”. This distinction has a strong impact on the complexity of the distance and rank aggregation problems.

The case of tied candidates leads to *bucket* orders where the (elements in different) buckets are totally ordered, whereas elements in the same bucket are unrelated and are treated as equivalent [1,24,25]. Then any two candidates are comparable under a $\leq$-relation. Algorithmically, bucket orders behave like total orders, and the distance and rank aggregation problems of two bucket orders can be solved in almost linear time under the Kendall tau distance. This is due to the refinement of one bucket order with respect to another bucket order as studied by Fagin et al. [24,25]. They also compare bucket orders under various metrics. In contrast, the case of incomparable candidates leads to (arbitrary) partial orders, where the distance and the ranking problems are $\mathcal{NP}$-complete for a partial and a total order [3,13,14].

A *partial order* κ is a binary relation over a set of candidates $\mathcal{D}$, which is irreflexive, asymmetric and transitive. It can be represented by its directed acyclic *associate graph*, where the candidates are the vertices and the edges represent the binary relations of κ. Since partial orders are transitive, a more compact representation is the *Hasse diagram* $HD(\kappa)$ which is the transitive reduction of the associate graph, see also [29]. $HD(\kappa)$ consists of the defining relations of κ. A *chain* is a sequence of related candidates which lie on a path of the Hasse diagram, whereas unrelated candidates belong to an *anti-chain*.

Hasse diagrams are used to define and classify partial orders. Obviously, a partial order κ over a set of candidates $\mathcal{D}$ is a *total order* if and only if its Hasse diagram is a path. It is a *bucket order* if $\mathcal{D}$ is partitioned into pairwise disjoint sets $B_1, \ldots, B_r$ for some r and $HD(\kappa)$ consists of a sequence of complete bipartite graphs $G_i = (B_i \cup B_{i+1}, B_i \times B_{i+1})$ for $1 \leq i < r$. Bucket orders are also known as weak orders, complete preorders, or partial rankings [24]. The buckets are totally ordered whereas candidates in a bucket are tied and form a maximal anti-chain. The complementary relation, which orders a pair of candidates if they were unrelated, is a *chain sum order*. Here $HD(\kappa)$ consists of a set of w disjoint paths of length at most h, where w is called the *width* and h the *height* of κ. The set of candidates is partitioned into w pairwise disjoint sets $P_1, \ldots, P_w$ with $|P_i| \leq h$ for $1 \leq i \leq w$, such that the candidates of each P_i are totally ordered whereas candidates in two distinct P_i are unrelated. Loosely speaking, a chain sum order is an anti-chain of chains and a bucket order is a chain of anti-chains. A chain sum order is *uniform* (of height h) if all chains have length exactly h and $|P_i| = h$ for $1 \leq i \leq w$. Bucket and chain sum orders are a special case of *series parallel orders* whose Hasse diagrams are series-parallel graphs. A partial order κ is a *hierarchical order* if $HD(\kappa)$ is a tree and is a *heap order* if $HD(\kappa)$ is a complete binary tree. Various other classes of partial orders have been studied in scheduling and are used to express precedence constraints [22]. For a systematic treatment of partial orders we refer to Trotter [29].

A partial order κ has a set of *extensions*, which are the total orders that do not disagree with κ. In graph theoretic terms, this is the set of topological sorts from the Hasse diagram or the associate graph. This set of total orders is used to define the distance of two partial orders. In this work, we use the *nearest neighbor distance*, which uses a min–min definition and chooses an extension with least distance to a given total order. At other places the Hausdorff distance is used [24], which uses a min–max definition.

1.2. Applications

Ranking information has been studied long ago in voting theory by Borda [12] and Condorcet [16]. The work of Dwork et al. [23] initiated recent research on the rank aggregation problem. On the theoretical side, they improved upon the $\mathcal{NP}$-hardness result of Bartoldi et al. [7], and on the practical side, they demonstrated the use of the rank aggregation problem for web site ranking and spam reduction.

Sports is another important application area of rankings. It sets highlights, such as the Olympics, world and national championships, and famous tournaments. There are hundreds of sports and millions of athletes, teams, and competitions, but there are primarily two ways to determine a winner: by a point system or by a (single-elimination) tournament. Point systems are score based and are used, e.g., in national football leagues. On the other hand, there are famous tournaments in tennis or at the Olympics, which are comparison based and proceed by eliminating the loser. In many sports there is a ranking list although all competitions are (single-elimination) tournaments. Score based rankings are vulnerable, since the outcome strongly depends on the score system. These problems are discussed in social science theory [7] with a strong impact from computational complexity [9,17,30]. Who is the better athlete or team, candidate a who won twice and ended last in a third competition or candidate b who ended second in all three competitions? In tennis, golf, and many other sports a would earn more prize money and would be ranked top, whereas b would win in skiing or Formula 1 racing, and a wins if the Condorcet criterion holds [16]. Because of such drawbacks it is generally accepted that a comparison based system has advantages over a score based system. However, computations on comparison based systems are generally much harder than on score based systems [7,9,23,30]. In addition, score based systems are easier to predict. As an example, consider football, where the winning team gets three points. Suppose that two teams A and B are close and team A has four more points than team B. Then A is champion even if it loses the last match against B. However, in a comparison based