Journal of Discrete Algorithms 31 (2015) 2–13

Contents lists available at ScienceDirect

Journal of Discrete Algorithms

www.elsevier.com/locate/jda

On the hardness of maximum rank aggregation problems [☆]

Christian Bachmaier [a], Franz J. Brandenburg [a,*], Andreas Gleißner [a], Andreas Hofmeier [b]

[a] University of Passau, 94030 Passau, Germany
[b] DocuWare GmbH, 82110 Germering, Germany

ARTICLE INFO

Article history:
Available online 16 October 2014

Keywords:
Center consensus
Rank aggregation
Maximum ranking
Permutation metrics
Computational complexity
FPT

ABSTRACT

The rank aggregation problem consists in finding a consensus ranking on a set of alternatives, based on the preferences of individual voters. The alternatives are expressed by permutations, whose pairwise distance can be measured in many ways.

In this work we study a collection of distances, including the Kendall tau, Spearman footrule, Minkowski, Cayley, Hamming, Ulam, and related edit distances. Unlike the common median by summation, we compute the consensus against the maximum. The maximum consensus attempts to minimize the discrimination against any voter and is a smallest enclosing ball or center problem.

We provide a general schema via local permutations for the **NP**-hardness of the maximum rank aggregation problems under all distances which satisfy some general requirements. This unifies former **NP**-hardness results for some distances and lays the ground for further ones. In particular, we establish a dichotomy for rank aggregation problems under the Spearman footrule and Minkowski distances: The median version is solvable in polynomial time whereas the maximum version is **NP**-hard. Moreover, we show that the maximum rank aggregation problem is 2-approximable under any pseudometric and fixed-parameter tractable under the Kendall tau, Hamming, and Minkowski distances, where again a general schema via modification sets applies.

1. Introduction

The task of ranking a list of alternatives is encountered in many situations. A major goal is to find the best consensus. This task is known as the rank aggregation problem, and was widely studied in recent years [1,2,5,7,9,10,17,26,28]. The problem has numerous applications in sports, voting systems for elections, search engines, and evaluation systems on the web [17].

From mathematical and computational perspectives, the rank aggregation problem is given by a set of m permutations on a set of size n, and the goal is to find a consensus permutation with minimum distance to the given permutations. There are many ways to measure the distance between two permutations and to aggregate the cost by an objective function. Various

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

* Corresponding author.
E-mail addresses: bachmaier@fim.uni-passau.de (C. Bachmaier), brandenb@informatik.uni-passau.de (F.J. Brandenburg), gleissner@fim.uni-passau.de (A. Gleißner), andreas.hofmeier@docuware.com (A. Hofmeier).

http://dx.doi.org/10.1016/j.jda.2014.10.002
1570-8667/

distances are based on primitive operations on permutations, as they are used in sorting algorithms and string matching. Aggregation is by taking the sum or the maximum.

For the rank aggregation problem Kemeny, [27] proposed to count the pairwise disagreements between the orderings of two items, which is commonly known as the Kendall tau distance. For permutations it is the "bubble sort" distance, i.e., the number of pairwise adjacent transpositions needed to transform one permutation into the other, or the number of crossings in a two-layered drawing of a bipartite graph with vertices of 1 to n on each layer and edges $\{i, i\}$ for $i = 1, \dots, n$ [8]. Another popular measure is the Spearman footrule distance [15], which is the L_1-norm of two n-dimensional vectors and expresses the total movement of items.

The geometric median of the input permutations is commonly taken for the aggregation, which means *summing* up the cost of comparing each input permutation with the consensus. From the computational perspective this makes an essential difference between the Spearman footrule and the Kendall tau distances, since the further allows a polynomial time solution via weighted bipartite matching [17], whereas the latter leads to an **NP**-hard rank aggregation problem [5], even for four voters [8,17]. It has a PTAS [28] and is fixed-parameter tractable [7,26].

Here we study the *maximum version*, which is also known as a smallest enclosing ball or center problem. The aim is to avoid a discrimination of a single voter or permutation against the consensus. The objective is a minimum k such that all permutations are within distance k from the consensus. Biedl et al. [8] studied this version for the Kendall tau distance and showed that it is **NP**-hard to determine whether there is a permutation τ which is within a distance of at most k to all input permutations, even for any $m \geq 4$ permutations. The **NP**-hardness was independently proven by Popov [34] and further investigated by Schwarz [35]. The smallest enclosing ball problem is a famous mathematical problem. It dates back to Sylvester in 1857 [37] and has been intensively studied in computational geometry [36], production planning [23], and stringology [25].

Besides the Kendall tau and the Spearman footrule distances there are other distance measures on permutations [17,21,27]. Many of them are edit distances, which can be expressed as the minimum number of specific primitive operations to transform one permutation into the other. Some operations are local, others operate globally on singletons, and the most powerful ones manipulate blocks or subsequences in a single step. The swap of two adjacent items, a unit movement of an item and a substitution are local operations and are used for the Kendall tau, Spearman footrule, and Hamming distances, respectively, whereas the Cayley distance allows the exchange of two items at arbitrary positions. The block reversal distance counts the reversal of a block in a permutation as a unit step. In consequence, the distance between two permutations often varies by a factor of $\mathcal{O}(n)$, e.g., if the first and last candidates are interchanged or if the second is the reversal of the first permutation. Such permutations are within unit distance for the block reversal distance and $\mathcal{O}(n^2)$ for the Kendall tau and Spearman footrule distances. As shown by Diaconis and Graham [15], these two distances are within a factor of two. The same applies to the Hamming and Cayley distances. Thus, these pairs meet the *metric boundedness property* [19]. For a broad discussion of distances we refer to [21]. Since computing the block reversal or the block transposition distance is **NP**-hard [11,12], we do not expect that maximum ranking under these distances is efficiently solvable and refrain from treating them any further.

We extend the collection of distances on permutations by Swap-and-Mismatch, Damerau–Levenshtein, and Lee distances, which are used in combinatorics for genome comparisons [21]. Our main contribution is a general schema for the complexity analysis of maximum rank aggregation problems, which allows us to prove **NP**-hardness and fixed-parameter tractability under any metric which satisfies some requirements. These requirements are met by our collection of distances. We associate the maximum rank aggregations on permutations and the string consensus problem on strings. Permutations on a set of size n can be seen as strings on an alphabet of size n, where each element occurs exactly once. However, the alphabet must scale with the length of the permutation and the uniqueness of the elements makes them special as strings.

For the association we use the generalization of total to bucket orders and local permutations as extensions of bucket orders. The technique of local permutations was first used implicitly by Popov [34] for Kendall tau and Cayley distances and with the main focus on the string consensus problem. Thereafter we obtain the **NP**-hardness results by reductions from the Closest Binary String and Hitting String problems, which is more general than the previous reductions [5,8,17,34].

The paper is organized as follows. After some preliminaries in Section 2 we show in Section 3 that Maximum Ranking (MR) is tractable under the Maximum distance, whereas MR is intractable under many other distances as shown in Section 4. In Section 5 we establish that MR is 2-approximable for pseudometrics. Finally, in Section 6, we present fixed-parameter algorithms to solve MR under various distances.

In a preliminary version of this paper [4] presented at IWOCA 2013 we consider only a subset of the distances, but our generalized schema applies to a broader set.

2. Preliminaries

For a binary relation ρ on a domain $\mathcal{D}$ and for each $x, y \in \mathcal{D}$, we write $x <_\rho y$ if $(x, y) \in \rho$ and $x \not<_\rho y$ if $(x, y) \notin \rho$. A binary relation κ is a (strict) *partial order* if it is *irreflexive*, *asymmetric* and *transitive*, i.e., $x \not<_\kappa x$, $x <_\kappa y \Rightarrow y \not<_\kappa x$, and $x <_\kappa y \wedge y <_\kappa z \Rightarrow x <_\kappa z$ for all $x, y, z \in \mathcal{D}$. Candidates x and y are called *unrelated by* κ if $x \not<_\kappa y \wedge y \not<_\kappa x$, which we denote by $x \not\gtrless_\kappa y$. The intuition of $x <_\kappa y$ is that κ *ranks x before y*, which means a preference for x. If $x <_\kappa y$ or $y <_\kappa x$, we speak of a *constraint of* κ *on x and y*. For $\mathcal{X}, \mathcal{Y} \subseteq \mathcal{D}$ we denote $\mathcal{X} <_\kappa \mathcal{Y}$ if $x <_\kappa y$ for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, and define $x <_\kappa \mathcal{Y}$