A NEW ACCURATE HEURISTIC ALGORITHM TO SOLVE THE RANK AGGREGATION PROBLEM WITH A LARGE NUMBER OF OBJECTS

Maurizio Romano [1] and Roberta Siciliano[2]

[1] Department of Business and Economics, University of Cagliari, (e-mail: romano.maurizio@unica.it)

[2] Department of Electrical and Information Technology, University of Naples Federico II, (e-mail: roberta@unina.it)

ABSTRACT: The analysis of preference rankings has become an important topic in the general field of data analysis in recent years. The classic meaning of preference rankings understood as orders expressed by a series of judges have been joined by the concept of judges is no longer always understood as human beings, but as resulting from automatic evaluation procedures. This paper provides a particle swarm-based optimization algorithm that provides an accurate solution to the rank aggregation problem, namely producing a ranking that best synthesizes the orderings stated by each judge, when the number of items to be evaluated is large

KEYWORDS: Kemeny problem, tied rankings, heuristics, particle swarm optimization

1 Introduction

The rank aggregation problem is known to be a NP hard problem. For this reason, several heuristic solutions have been proposed over the years. For example:

- Amodio *et al.*, 2016, proposes FAST, an heuristic algorithm based on QUICK that estimates consensus rankings from aggregate preferences. Computational efficiency and accuracy are shown with simulations and real data case studies;
- D'Ambrosio *et al.*, 2017, introduces DECoR, a Differential Evolution algorithm for Consensus Ranking, able to work with full, partial, and incomplete rankings. It outperforms previous proposals in both accuracy and speed while handling large datasets;

- Aledo *et al.*, 2017, considers the Optimal Bucket Order Problem (OBOP) by proposing improvements to the standard greedy algorithm (BPA), resulting in improved accuracy and reduced output variance;
- Aledo *et al.*, 2018, presents $(1+\lambda)$ evolution strategies for solving OBOP, with specific mutation operators and initialization methods. Accuracy improvement with respect to the state-of-the-art algorithm is shown with simulated data;
- Aledo *et al.*, 2021, proposes greedy algorithms based on sort-first and cluster-second strategies to efficiently solve OBOP. Accuracy and scalability improvements of the proposed algorithms with respect to the state-of-the-art algorithm are shown with simulated data;
- Acampora *et al.*, 2021, introduces a memetic algorithm combining genetic algorithms with hill-climbing search for rank aggregation. In particular, results are compared with the DeCoR algorithm (D'Ambrosio *et al.*, 2017).

We follow Kemeny's axiomatic approach (Kemeny, 1959; Kemeny & Snell, 1962), according to which the median (or consensus) ranking is that ranking, or those rankings, that minimize the sum of the distances between a candidate ranking belonging to the universe of rankings and the orderings expressed by a set of judges. Moreover, especially when the number of items to be ranked is large, we assume that all possible tied rankings are allowed either in the data matrix containing the orderings or in the final solution. In other words, we assume that tied rankings are not indifference declarations, but they are 'positive statements of agreement' (Emond & Mason, 2002).

2 Particle swarm optimization for preference rankings

We propose a particle swarm optimization algorithm for the rank aggregation problem (PSORaP). We compared the solutions achieved by the DECoR algorithm (Differential Evolution algorithm for Consensus Ranking, D'Ambrosio *et al.*, 2017) and our PSOPaR on the USA ranks data set (O'Leary Morgan & Morgon, 2010), that contains rankings of the 50 US states with respect to various aspects about the economic and social situation, security, etc., included in the internal repository of the `R` package `ConsRank` (D'Ambrosio, 2021). Table 1 shows the solutions obtained by the DECoR and our PSO algorithm, evaluated through the τ_X rank correlation coefficient (Emond & Mason, 2002). The solutions are really similar (DECoR $\tau_X = 0.2976688$, PSORaP $\tau_X = 0.297449$).

Table 1. *Direct comparison of the consensus generated by DECoR and PSO*

Rank	**DECoR**	**PSORaP**	**Rank**	**DECoR**	**PSORaP**
1	California	California	26	Colorado	{Colorado
2	New.York	New.York	27	Connecticut	Minnesota}
3	Florida	Florida	28	Minnesota	Alabama
4	Maryland	Maryland	29	Alabama	{Connecticut
5	Louisiana	Louisiana	30	South.Carolina	South.Carolina}
6	Illinois	New.Mexico	31	Oregon	Oregon
7	New.Mexico	{Illinois	32	Oklahoma	Oklahoma
8	Delaware	Texas}	33	Mississipi	Mississipi
9	Texas	Pennsylvania	34	Arkansas	Arkansas
10	Pennsylvania	Michigan	35	Hawaii	Hawaii
11	Michigan	{Georgia	36	Kentucky	Kentucky
12	Georgia	North.Carolina}	37	{Kansas	{Kansas
13	North.Carolina	New.Jersey	38	Rhode.Island}	Rhode.Island}
14	New.Jersey	{Massachusetts	39	Utah	Utah
15	Massachusetts	Washington}	40	{Iowa	{Iowa
16	Washington	Nevada	41	Nebraska}	Nebraska}
17	Ohio	Delaware	42	Wyoming	Wyoming
18	Virginia	Ohio	43	West.Virginina	West.Virginina
19	Tennessee	{Arizona	44	Idaho	Idaho
20	Nevada	Virginia}	45	Maine	Maine
21	Arizona	Tennessee	46	Montana	Montana
22	Missouri	Missouri	47	New.Hampshire	New.Hampshire
23	Indiana	{Alaska	48	South.Dakota	South.Dakota
24	Alaska	Indiana}	49	Vermont	Vermont
25	Wisconsin	Wisconsin	50	North.Dakota	North.Dakota

The differences between the solutions are mainly that DECoR returns less tied US states in the first part, with Delaware ranked 8 for DECoR and 17 for PSORaP.

3 Concluding remarks

In this paper, a particle swarm optimization heuristic algorithm for the rank aggregation problem has been introduced. A comparison with an already proposed differential evolution algorithm shows that the results are encouraging. A deeper study of the behavior of PSORaP will be carried out in the future to better understand how setting the tuning parameters for improving the performance of the algorithm in terms of the accuracy of the solution.

References

ACAMPORA, GIOVANNI, IORIO, CARMELA, PANDOLFO, GIUSEPPE, SICILIANO, ROBERTA, & VITIELLO, AUTILIA. 2021. A memetic algorithm for solving the rank aggregation problem. *Algorithms as a Basis of Modern Applied Mathematics*, 447–460.

ALEDO, JUAN A, GÁMEZ, JOSÉ A, & ROSETE, ALEJANDRO. 2017. Utopia in the solution of the bucket order problem. *Decision Support Systems*, **97**, 69–80.

ALEDO, JUAN A, GÁMEZ, JOSÉ A, & ROSETE, ALEJANDRO. 2018. Approaching rank aggregation problems by using evolution strategies: the case of the optimal bucket order problem. *European Journal of Operational Research*, **270**(3), 982–998.

ALEDO, JUAN A, GÁMEZ, JOSÉ A, & ROSETE, ALEJANDRO. 2021. A highly scalable algorithm for weak rankings aggregation. *Information Sciences*, **570**, 144–171.

AMODIO, SONIA, D'AMBROSIO, ANTONIO, & SICILIANO, ROBERTA. 2016. Accurate algorithms for identifying the median ranking when dealing with weak and partial rankings under the Kemeny axiomatic approach. *European Journal of Operational Research*, **249**(2), 667–676.

D'AMBROSIO, ANTONIO. 2021. *ConsRank: Compute the Median Ranking(s) According to the Kemeny's Axiomatic Approach*. R package version 2.1.2.

D'AMBROSIO, ANTONIO, MAZZEO, GIULIO, IORIO, CARMELA, & SICILIANO, ROBERTA. 2017. A differential evolution algorithm for finding the median ranking under the Kemeny axiomatic approach. *Computers & Operations Research*, **82**, 126–138.

EMOND, EDWARD J, & MASON, DAVID W. 2002. A new rank correlation coefficient with application to the consensus ranking problem. *Journal of Multi-Criteria Decision Analysis*, **11**(1), 17–28.

KEMENY, J.G. 1959. Mathematics without numbers. *Daedalus*, **88**.

KEMENY, J.G., & SNELL, J.L. 1962. Preference rankings: An axiomatic approach. *Pages 9–23 of:* KEMENY, J.G., & SNELL, J.L. (eds), *Mathematical models in the social sciences*. New York: Blaisdell.

O'LEARY MORGAN, K., & MORGON, S. 2010. *State Rankings 2010: A Statistical view of America; Crime State Ranking 2010: Crime Across America; Health Care State Rankings 2010: Health Care Across America.* CQ Press.