Pareto front clustering for real time rescheduling in dense railway systems

Abdelkrim R. Yelles-Chaouche[1], Hugo Belhomme[1,2],
Stéphane Dauzère-Pérès[1], Mathieu Gagnon[2]

[1] Mines Saint-Étienne, LIMOS, Gardanne, France
[2] SNCF, Hub of Technologies, Innovation and Group Projects, Saint-Denis, France

Keywords : *Railway transportation, Pareto clustering, multi-criteria decision analysis, K-means*

1 Context

Train is one of the most widely used means of urban transport, especially in large and densely populated areas like the Paris region where millions of public trips are performed everyday. This is possible thanks to a large infrastructure of 6,200 trains operating on 1,300 km of tracks and carrying more that 3.2 million passengers per day. Operating such a complex dense railway system is challenging, as the smallest disturbance on a line can quickly amplify and spread, causing major train delays that can significantly impact passengers.

In this context, [1] propose a decision support tool for real time rescheduling in dense railway systems, which was applied at SNCF Transilien. More precisely, the authors have developed a multi-objective optimization-simulation approach to deal with disturbances that occur on the lines. The output is a series of rescheduling decisions that can be applied on the line in order to optimize different conflicting objectives. Possible decisions, that can be combined, include canceling or short-turning trains and skipping or adding stops. The objectives to minimize are primarily the recovery time and and passenger dissatisfaction, but also the sum of train delays and the number of decisions. Based on the state of the considered railway line, an optimization-simulation approach provides the decision maker with a set of non-dominated solutions in terms of Pareto optimality, generally referred to as the Pareto Front. It is worth noting that only a limited number of Pareto optimal solutions should be proposed to the decision maker. A process that selects three solutions based on some ranking rules is introduced in [1]. This process may not guarantee that the most appropriate solutions are selected. Moreover, as it is mainly based on the objectives, this selection process does not take into account decision types. This can lead to proposing solutions that are not acceptable in practice, e.g. a train is never canceled by a decision maker during rush hour.

Hence, our goal to propose a new selection process that provides the decision maker with a subset of non-dominated solutions while taking into account both the objectives and decision types. This means that the selected solutions must be sufficiently diversified to offer the decision maker multiple alternatives with different types of decisions.

2 An approach for Pareto front clustering

To address the above mentioned problem, a clustering approach is proposed that relies on an unsupervised learning technique, more precisely the K-means method, to identify distinct groups from data points. Each identified cluster therefore contains a subset of points that share similar characteristics. In the context of multi-criteria decision analysis, clustering is generally used when it is not possible for the decision maker to articulate preferences with respect to each of the objectives [2]. In this case, a clustering approach allows the Pareto optimal solutions to be partitioned into groups, and thus a representative solution from each cluster to be identified. Interestingly, this approach not only reduces the Pareto front to a limited subset of representative solutions, but also manages to select diversified solutions.

In this context, the present work consists in studying the efficiency of a clustering approach to select K solutions from a Pareto front that is obtained by solving an industrial instance. The steps of the proposed Pareto clustering approach are described below:

1. Set the number of clusters K,
2. Apply the K-means methods on the set of non-dominated solutions,
3. Identify from each cluster $i \in [1, \cdots, K]$ a representative solution s_i, using a selection strategy,
4. Return the final set $S = [s_1, \cdots, s_K]$ with the representative solutions.

In the literature, Pareto clustering is usually performed based on the objectives, which often results in creating clusters that include solutions with similar objective values. This approach is efficient and usually allows interesting solutions to be selected. However, when applied to our problem, two selected solutions may include the same types of decision, thus offering little choice to the decision maker.

To overcome this problem, we applied the same clustering approach, but based on decision types instead of objectives. As a result, the newly defined clusters contain solutions that share the same types and number of decisions. In doing so, we ensure that the selected representative solutions are diverse in terms of decision types, providing the decision maker with a relevant choice of possible decisions.

To select from each cluster a representative solution, four different strategies were developed and tested. The first, and most commonly used strategy is to choose the solution closest to the center of the cluster. In the second strategy, the best solution is selected in each cluster according to the lexicographic order of the objectives. Similarly, the third strategy is based on the ascending sorting of the solutions in each cluster based on the sum of the normalized values of the objectives. Finally, the fourth strategy consists in selecting, in each cluster, the solution with the longest euclidean distance from the centers of the other clusters.

The numerical results show that our approach for Pareto front clustering identifies a subset of diversified representative solutions, selected in a set of non-dominated solutions, that should be more relevant for the decision maker than the approach in [1]. A detailed analysis and a comparison between the different strategies for K-means clustering will be presented at the conference.

References

[1] E. Altazin, S. Dauzère-Pérès, F. Ramond, and S. Tréfond. A multi-objective optimization-simulation approach for real time rescheduling in dense railway systems. *European Journal of Operational Research*, 286(2):662–672, 2020.

[2] S. Petchrompo, D. W. Coit, A. Brintrup, A. Wannakrairot, and A. K. Parlikad. A review of pareto pruning methods for multi-objective optimization. *Computers & Industrial Engineering*, 167:108022, 2022.