Data Article

Problem instances dataset of a real-world sequencing problem with transition constraints and asymmetric costs

Nicolás Álvarez-Gil[a],*, Segundo Álvarez García[b], Rafael Rosillo[a], David de la Fuente[a]

[a] Department of Business Administration, University of Oviedo, C\ Luis Ortiz Berrocal s/n Campus de Gijón, Gijón, Asturias 33203, Spain
[b] ArcelorMittal Global R&D Asturias, P.O. Box 90, Avilés, Asturias 33400, Spain

ABSTRACT

This data article describes 30 instances of the real-world problem of sequencing steel coils in a continuous galvanizing line. Each instance is represented by a cost matrix that gives information of the cost of sequencing each pair of coils or items together (e.g. a transition). Some transitions are forbidden due to technical limitations of the line and/or because of the properties of the coils, what makes the problem more challenging. These costs were previously obtained by a cost model that estimates the final cost of each transition for a set of coils to be sequenced in the line. Although the instances come from this real context, the problem can be theoretically seen as finding a minimum cost Hamiltonian path (e.g. a minimum cost feasible production sequence with all the coils appearing just once). It is a well-known NP-Hard combinatorial optimization problem. Since these instances represent real challenges found in the industry, they can be very useful for algorithm development and testing. Due to the cost distributions obtained for the given coils, just finding a feasible sequence can be a challenging task, especially for some types of approximate algorithms (Alvarez-Gil et al., 2022).

Specifications Table

Subject	Business, Management and decision sciences (Management Science and Operations Research)
Specific subject area	Operations Research applications (planning & scheduling, logistics, etc.) and Combinatorial Optimization (algorithm development and benchmarking)
Type of data	• Problem instances in separated text files (.txt) • Figures • Tables
How the data were acquired	The data were acquired using a customized software that helps the scheduling crew with the sequencing tasks. The software receives the information of the coils to be sequenced. The software has an embedded cost model that calculates, for each pair of coils, the cost of sequencing them together based on its properties, generating a cost matrix for the set of coils. Then the software allows to export the cost matrices in the provided format (.txt).
Data format	Raw
Description of data collection	The data were acquired ensuring that: • Each instance corresponds to a daily real production schedule • Each set of coils is sequenceable (e.g. there exist a Hamiltonian path)
Data source location	Spanish steel factory
Data accessibility	Repository name: Problem instances of the sequencing problem of a steel continuous galvanizing line (Mendeley Data) Data identification number: DOI:10.17632/v357z2ncbh.2 Direct URL to data: https://data.mendeley.com/datasets/v357z2ncbh/2
Related research article	N. Alvarez-Gil, S. Alvarez, R. Rosillo, D. de la Fuente, Sequencing Jobs with Asymmetric Costs and Transition Constraints in a Finishing Line: A real case study, Computers & Industrial Engineering 165 (2022). 10.1016/j.cie.2021.107908

Value of the Data

- The dataset provides 30 different instances of a real-world sequencing problem with transition constraints and asymmetric costs that can be used to evaluate and compare algorithms performance. The reference results presented in [1] can be used as benchmark solutions.
- The instances can help in the development, design and testing of combinatorial optimization algorithms. Since the instances are highly constrained, it can help in the development of new approaches able to efficiently explore the solution space and in the design of constraint-handling strategies. Due to the cost distributions obtained from the real sets of items to be sequenced in the line, some of the instances are very complex and can be very challenging for the approximate algorithms, especially for those algorithms that use the cost information to guide the exploration, and this focus on the costs can misguide the feasibility search (see [1]). The use of the presented dataset can be very useful to develop robust solutions able to handle constraints while minimizing the cost.
- Problem instances are directly provided as a cost matrix what makes them easy and ready to use. They represent real challenges of a scheduling problem found in the industry, but they can also be abstracted to the more theoretical problem of finding a minimum-cost Hamiltonian path [2]. Hence, they can be used both to gain insights for the different combinatorial optimization applications (scheduling, routing, etc.) and for theoretical purposes (graph analysis, constraint-handling strategies, approximate algorithms, etc.).

1. Data Description

The dataset consists in 30 different problem instances with sizes ranging from 17 to 114 items (e.g. coils, nodes, etc.). Each instance is named as "*cgl_X.txt*", being *X* the size of the instance. Table 1 shows the name and the size of the problem instances provided in the dataset.

Each instance is represented by a *n* x *n* square matrix, being *n* the size of the instance (e.g. number of items to be sequenced). The cost matrix provides information about the cost of scheduling two pair of coils together in the sequence. The element [i,j] of the matrix represents the cost of processing item *i* right before item *j* (C_{ij}). An element C_{ij} can take two possible values:

- $C_{ij} > 0$, $C_{ij} \in \mathbb{R}$: This means that the transition between item *i* and *j* is allowed (e.g. item *i* can be produced immediately before item *j*) with a cost C_{ij}.
- $C_{ij} = -1$: This means that the transition between item *i* and *j* is not allowed (e.g. item *i* cannot be produced immediately before item *j*) and it represents a hard constraint.

The problem consists in finding a minimum-cost feasible sequence containing all the nodes just once, what can be seen as finding a minimum-cost Hamiltonian path [2]. We refer to a feasible sequence as a sequence for which all the transitions are allowed: a sequence without constraints violations. Just finding a feasible sequence is a challenge itself, but the cost should also be minimized. The lower number of constraints the better. For a feasible sequence, the lower the cost the better. The initial and final items of the sequence are not fixed and thus that decisions are part of the optimization problem, since they can directly impact in the final cost of the sequence.

The problem shares some similarities with other well-known combinatorial optimization problems such as the Asymmetric Traveling Salesman Problem [3,4], the Sequencing Ordering Problem [5,6] and the Hamiltonian Cycle Problem [7,8]. More information about the similarities and differences can be found in [1].

All the cost matrices are provided in the same format, shown in Fig. 1.

Table 1
List of problem instances names, sizes and reference solutions.

Instance name	Size	Instance name	Size
cgl_17.txt	17 items	cgl_51b.txt	51 items
cgl_26.txt	26 items	cgl_57.txt	57 items
cgl_28.txt	28 items	cgl_58.txt	58 items
cgl_32.txt	32 items	cgl_60.txt	60 items
cgl_33.txt	33 items	cgl_66.txt	66 items
cgl_37.txt	37 items	cgl_70.txt	70 items
cgl_38.txt	38 items	cgl_70b.txt	70 items
cgl_43.txt	43 items	cgl_72.txt	72 items
cgl_44.txt	44 items	cgl_73.txt	73 items
cgl_45.txt	45 items	cgl_76.txt	76 items
cgl_47.txt	47 items	cgl_78.txt	78 items
cgl_48.txt	48 items	cgl_81.txt	81 items
cgl_48b.txt	48 items	cgl_88.txt	88 items
cgl_50.txt	50 items	cgl_107.txt	107 items
cgl_51.txt	51 items	cgl_114.txt	114 items

```
-1;300;-1;654;926;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1
305;-1;963;117;237;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1
952;493;-1;704;686;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1
678;123;-1;-1;41;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1
946;237;1155;34;-1;804;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1
-1;-1;-1;-1;-1;-1;71;1721;-1;-1;-1;-1;-1;-1;-1;-1;-1
-1;-1;-1;-1;-1;73;-1;1777;-1;-1;-1;-1;-1;-1;-1;-1;-1
-1;-1;-1;-1;-1;2034;2020;-1;634;634;634;634;634;634;634;634;67
-1;-1;-1;-1;-1;-1;-1;685;-1;0;0;0;0;0;0;0;308
-1;-1;-1;-1;-1;-1;-1;685;0;-1;0;0;0;0;0;0;308
-1;-1;-1;-1;-1;-1;-1;685;0;0;-1;0;0;0;0;0;308
-1;-1;-1;-1;-1;-1;-1;685;0;0;0;-1;0;0;0;0;308
-1;-1;-1;-1;-1;-1;-1;685;0;0;0;0;-1;0;0;0;308
-1;-1;-1;-1;-1;-1;-1;685;0;0;0;0;0;-1;0;0;308
-1;-1;-1;-1;-1;-1;-1;685;0;0;0;0;0;0;-1;0;308
-1;-1;-1;-1;-1;-1;-1;685;0;0;0;0;0;0;0;-1;308
-1;-1;-1;-1;-1;-1;-1;77;293;293;293;293;293;293;293;293;-1
```

Fig. 1. Content of file cgl_17.txt.

Each text file (.txt) has n rows, and each row has n columns delimited by the character ";". Taking instance *cgl_17.txt* for illustration (Fig. 1), the cost matrix should be interpreted in the following way:

From row 1:

- Matrix element [0, 0] ($C_{00}= -1$): not possible, each item must appear just once in the sequence.
- Matrix element [0, 1] ($C_{01}= 300$): cost of producing item 0 right before item 1.
- Matrix element [0, 2] ($C_{02}= -1$): not possible, production constraint.
- Matrix element [0, 3] ($C_{03}= 654$): cost of producing item 0 right before item 3.
- Matrix element [0, 4] ($C_{04}= 926$): cost of producing item 0 right before item 4.
- (...)

From row 2:

- Matrix element [1, 0] ($C_{10}= 305$): cost of producing item 1 right before item 0.
- Matrix element [1, 1] ($C_{11}= -1$): not possible, each item must appear just once in the sequence.
- Matrix element [1, 2] ($C_{12}= 963$): cost of producing item 1 right before item 2.
- Matrix element [1, 3] ($C_{13}= 117$): cost of producing item 1 right before item 3.
- Matrix element [1, 4] ($C_{14}= 237$): cost of producing item 1 right before item 4.
- (...)

A possible solution for this instance is $S = [2, 0, 1, 3, 4, 5, 6, 7, 16, 11, 15, 10, 8, 12, 14, 13, 9]$. It is a valid sequence since all the items appears just once, and the length of the sequence is n. In order to get the final cost of the sequence, all the transition costs should be computed, and the final cost is the sum of them. For a sequence of length n, the number of transitions is $n-1$. For sequence S, the cost C_S will be:

$$Cs = C_{2,0} + C_{0,1} + C_{1,3} + C_{3,4} + C_{4,5} + C_{5,6} + C_{6,7} + C_{7,16} + C_{16,11} + C_{11,15} + C_{15,10} + C_{10,8}$$
$$+ C_{8,12} + C_{12,14} + C_{14,13} + C_{13,9}$$

$$Cs = 952 + 300 + 117 + 41 + 804 + 71 + 1777 + 67 + 293 + 0 + 0 + 0 + 0 + 0 + 0$$
$$+ 0 = 4422$$

It is worth mentioning that, for all the instances of the dataset, the sequence [0, 1, 2, 3, ..., n-1] is a feasible sequence and can be used as a result reference. The only exception is instance *cgl_38.txt*, for which that solution contains one constraint violation.

2. Experimental Design, Materials and Methods

The instances come from the real-world problem of sequencing steel coils (e.g. final format of flat steel products) in a continuous galvanizing line. The galvanization is a finishing process that provides the coils with a zinc layer to protect them against air and moisture [9]. The line operators use a stand-alone scheduling application installed in their computer that receives, for each day, the set of coils that should be sequenced and their main properties (width, thickness, steel grades, zinc coating, etc.). This information is obtained connecting directly to the plant Manufacturing Execution System (MES). Then, the software calculates the cost of processing each pair of coils and creates the cost matrix, which is the only input required for the sequencing problem.

The galvanizing process is continuous, being all the coils welded to the others in order to create a continuous strip. The cost of sequencing two coils together depends on their properties, seeking to avoid sudden changes in the main properties (width, thickness, zinc coating, etc.) when processing a new coil. For example, if two consecutive coils have different widths, some meters of strip near the welding zone may not meet the quality requirements and may be sold as scrap at a much lower price. Additionally, if the change in some properties is very sharp, there exists risk of strip breakage and that transition is penalized with a high cost. The scheduling software also calculates the constraints. Some coils cannot be sequenced together because their steel grades are not weldable or because the difference in some of their properties exceeds a threshold defined by the expert schedulers.

In order to obtain the problem instances of the dataset, we looked for daily schedules processed in the line in the past, so we were sure that the set of is sequenceable. Then, we used the scheduling software to load that schedules, generate the cost matrices and exported them in the provided format.

Ethics Statement

N/A.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT Author Statement

Nicolás Álvarez-Gil: Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Writing – original draft, Writing – review & editing, Data curation; **Segundo Álvarez García:** Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Writing – review & editing, Data curation; **Rafael Rosillo:** Conceptualization, Methodology, Visualization, Investigation, Supervision, Writing – review & editing; **David de la Fuente:** Conceptualization, Investigation, Supervision.

Acknowledgments

N/A.

References

[1] N. Alvarez-Gil, S. Alvarez, R. Rosillo, D. de la Fuente, Sequencing jobs with asymmetric costs and transition constraints in a finishing line: a real case study, Comput. Ind. Eng. 165 (2022), doi:10.1016/j.cie.2021.107908.
[2] M. Sohel Rahman, M. Kaykobad, On Hamiltonian cycles and Hamiltonian paths, Inf. Process. Lett. 94 (2005) 37–41, doi:10.1016/j.ipl.2004.12.002.
[3] D.L. Applegate, R.E. Bixby, V. Chvátal, W.J. Cook, The Traveling Salesman Problem: A Computational Study, Princeton University Press, 2007, doi:10.1515/9781400841103.
[4] K. Helsgaun, An effective implementation of the Lin–Kernighan traveling salesman heuristic, Eur. J. Oper. Res. 126 (2000) 106–130, doi:10.1016/S0377-2217(99)00284-2.
[5] L.F. Escudero, An inexact algorithm for the sequential ordering problem, Eur. J. Oper. Res. 37 (1988) 236–249, doi:10.1016/0377-2217(88)90333-5.
[6] N. Ascheuer, M. Jünger, G. Reinelt, A branch & cut algorithm for the asymmetric traveling salesman problem with precedence constraints, Comput. Optim. Appl. 17 (200) (2000) 61–84, doi:10.1023/A:1008779125567.
[7] M.R. Garey, D.S. Johnson, R.E. Tarjan, The planar Hamiltonian circuit problems is NP-complete, SIAM J. Comput. 5 (1976) 704–714, doi:10.1137/0205049.
[8] M.R. Garey, D.S. Jhonson, Computers and Intractability A Guide to the Theory of NP-Completeness, W.H. Freeman and Company, 1979.
[9] S. Fernández, S. Álvarez, D. Díaz, M. Iglesias, B. Ena, Scheduling a galvanizing line by ant colony optimization, in: ANTS 2014: Lecture Notes in Computer Science, 8667, 2014, pp. 146–157, doi:10.1007/978-3-319-09952-1_13.