Dataset of Home Care Scheduling and Routing Problems with Synchronized Visits

Mateusz Polnik[a,*], Annalisa Riccardi[a], Kerem Akartunalı[b]

[a] *Mechanical and Aerospace Engineering, University of Strathclyde, 75 Montrose Street, Glasgow, G1 1XJ, UK*
[b] *Management Science, University of Strathclyde, 130 Rottenrow, Glasgow, G4 0GE, UK*

Abstract

The document describes the dataset of the problem instances solved in the article: *A Multistage Optimization Algorithm for the Home Care Scheduling and Routing Problem with Synchronized Visits.*

The data set was built based on anonymized instances of real scheduling and routing problems obtained from a home care provider who operates in a major city in the UK. The data set is released on the Open Data Commons Attribution License (ODC-BY) by Glasgow City Council.

Problem Instances

The data set comprises two groups of problem instances, which we refer to as the real problems and benchmark problems, respectively.

The real problems are examples of the instances the home care organization faced in their daily operations. Visits included in the dataset were performed in a single district in the first two weeks of October 2017. The numbers of available carers and visits delivered depend on the day. Overall, during the selected period the area was supported by 47-71 carers who completed 444-508 visits.

On the other hand, the benchmark problems are small instances of 25 of 50 visits derived from the real problem instances. The benchmark problems come in two variants: with and without multiple-carer visits. The latter group was obtained by relaxing the requirement that two carers must perform a given appointment. The primary motivation for creating the benchmark instances was to demonstrate that a scheduling algorithm can find provably optimal solutions or nearly optimal solutions. See the article for the best solutions found by the

*Corresponding author
Email addresses: mateusz.polnik@strath.ac.uk (Mateusz Polnik), annalisa.riccardi@strath.ac.uk (Annalisa Riccardi), kerem.akartunali@strath.ac.uk (Kerem Akartunalı)

Gurobi Mixed-Integer Solver and the Routing Library of Google Operational Research Tools.

Naming Convention for Files

Three CSV files containing carers, visits, and a distance matrix between visits' locations define each problem instance. The content of the file is determined by a suitable suffix. For example, the 'carers' suffix indicates a file with carers. Other possible suffixes are 'visits' and 'distance' which are self-explanatory.

The files describing the same problem instance share a common prefix which uniquely determinates the problem instance. The 'real' prefix indicates the scheduling problems of the Home Care organization. The 'benchmark' root prefix distinguishes the benchmark instances. The full prefix of a benchmark instance is longer and encodes the day, the number of visits, the number of carers, and the presence of the synchronization constraints. For example, the prefix 'benchmark_20171001_v25c3m5' indicates a problem containing 25 visits to schedule on the 1st of October 2017 using three carers. Five of these appointments are multiple-carer visits.

File Format

File formats are explained below.

Carers

The files with the 'carer' suffix contain a list of time slots in which carers work. A snippet with sample data records is displayed in Table 1.

Each row contains a time slot defined by Begin and End columns in which the carer indicated by the CarerId column is available for work on the day given in the Date column.

Visits

The files with the 'visits' suffix store the visits to schedule. Table 2 displays an example content of such a file.

Each row defines a visit to schedule. Data stored in each column is detailed below.

UserId Identifier of the service user, required to extract the travel time from the distance matrix,

Date Day of the appointment,

Time Most preferred start time for the visit, the center of the time window,

Duration Estimated duration of the visit in seconds,

CarerCount Number of carers required to perform the visit.

Table 1: Snippet of the file with carers. Each row contains a time slot defined by Begin and End columns in which the carer indicated by the CarerId column is available for work on the day given in the Date column.

CarerId	Date	Begin	End
36	2017-10-04	2017-10-04T08:00:00	2017-10-04T13:00:00
36	2017-10-05	2017-10-05T08:00:00	2017-10-05T13:00:00
36	2017-10-06	2017-10-06T08:00:00	2017-10-06T13:00:00
36	2017-10-07	2017-10-07T08:00:00	2017-10-07T14:00:00
36	2017-10-07	2017-10-07T16:00:00	2017-10-07T20:00:00
36	2017-10-08	2017-10-08T08:00:00	2017-10-08T14:00:00
36	2017-10-08	2017-10-08T16:00:00	2017-10-08T20:00:00
36	2017-10-09	2017-10-09T08:00:00	2017-10-09T13:00:00
36	2017-10-10	2017-10-10T08:00:00	2017-10-10T13:00:00
36	2017-10-11	2017-10-11T08:00:00	2017-10-11T13:00:00
36	2017-10-12	2017-10-12T08:00:00	2017-10-12T13:00:00
36	2017-10-13	2017-10-13T08:00:00	2017-10-13T13:00:00
⋮	⋮	⋮	⋮
1749	2017-10-02	2017-10-02T16:30:00	2017-10-02T20:30:00
1749	2017-10-03	2017-10-03T16:30:00	2017-10-03T20:30:00
1749	2017-10-04	2017-10-04T16:30:00	2017-10-04T20:30:00
1749	2017-10-10	2017-10-10T16:00:00	2017-10-10T20:30:00
1749	2017-10-13	2017-10-13T16:00:00	2017-10-13T20:30:00

Table 2: Snippet of a file with visits. Each row defines a visit to schedule. The UserId column indicates a reference location of the appointment used for travel time estimation. The Date column denotes the day in which a visit should be performed. The Time column defines the preferred start time for a visit, which is the center of the time window. The Duration column contains the estimated length of a visit in seconds. The CarerCount column indicates the number of carers required to perform a visit.

VisitId	UserId	Date	Time	Duration	CarerCount
2806455	6819	2017-10-13	11:30:00	1800	1
2806454	6819	2017-10-06	10:00:00	1159	1
2806448	6819	2017-10-11	11:30:00	1800	1
2806450	6819	2017-10-02	10:00:00	1159	1
2806447	6819	2017-10-04	10:00:00	1159	1
2806451	6819	2017-10-09	10:00:00	1159	1
⋮	⋮	⋮	⋮	⋮	⋮
2715231	5068	2017-10-13	09:45:00	1334	2
2715643	5068	2017-10-07	09:45:00	1394	2

Distance Matrix

The files with the 'distances' suffix contain the time required for travel between each pair of visits' locations. Table 3 presents a sample format of the

distance matrix.

Table 3: A snippet of the distances.csv file. The file contains a square symmetric matrix in which each element indicates travel time in seconds between two possible visit locations. The header, as well as the first column of the file, refer to the UserID index which links visits and locations.

UserId	4	38	...	8507	8545
4	0	1280	...	1025	1017
38	1280	0	...	1515	1433
⋮	⋮	⋮	⋱	⋮	⋮
8507	1025	1515	...	0	82
8545	1017	1433	...	82	0

The file contains a square symmetric matrix in which each element indicates travel time in seconds between visit locations represented by rows and columns assigned to given UserID indices.