Supplementary Material
Instance Space Analysis of the Capacitated Vehicle Routing Problem with Mixture Discriminant Analysis

Danielle Notice[1], Hamed Soleimani[2], Nicos G. Pavlidis[1], Ahmed Kheiri[3], Mario Andrés Muñoz[4]

[1] Lancaster University, United Kingdom; [2] LaTrobe University, Bundoora, Australia; [3] The University of Manchester, United Kingdom [4] The University of Melbourne, Parkville, Australia

1 CVRP Feature Space

We constructed a set of 105 features across six groups to characterise the CVRP instances. Many of these features have previously been used to characterise TSP and VRP instances [1, 2, 3, 4, 5]. Some features result is sets of values (such as distances, demands), in which case we include several features for different summary statistics (mean, variance, range, etc.).

1. **Customer demand**
   - customer demands (summary)
   - Ratio of total demand to total capacity, Ratio of max customer demand to vehicle capacity
   - Minimum fleet size (k), ratio of customers to vehicles
2. **Distance matrix**
   - Number of customers (n)
   - Edge cost (summary)
   - Proportion of distances lower than mean, fraction of distinct distances, sum of n shortest edges, expected tour length
   - Pairwise distance of customers from centroid (summary), centroid coordinates, distance between depot and centroid
   - Pairwise distance of customers from depot (summary), depot coordinates, distances of k nearest neighbours to depot
3. **Nearest neighbour distance**
   - Distance of nearest neighbour (NN) for each customer (summary)
   - Proportion of nodes with depot as NN, ratio of number of nodes with depot as NN to number of routes, angle between city and 2 NNs
4. **Minimum spanning tree (MST)**
   - Graph edge lengths (summary), normalised sum of edge lengths, number of edges
   - Graph node degree (summary)
5. **Geometric**
   - Area and perimeter of rectangle containing nodes

- Convex hull - area, perimeter, fraction of nodes on convex hull, edge length of convex hull (summary)

6. **Clustering**
   - Number of clusters, number of cluster relative to n
   - Cluster sizes (summary), distances between customers in cluster to cluster centroid (summary)
   - Ratio of outliers, edge and core points
   - Ratio of maximum cluster demand to vehicle capacity, ratio of outlier demand to overall demand

References

[1] Olaf Mersmann et al. "A novel feature-based approach to characterize algorithm performance for the traveling salesperson problem". In: *Annals of Mathematics and Artificial Intelligence* 69.2 (2013), pp. 151–182. DOI: 10.1007/s10472-013-9341-2.

[2] Josef Pihera and Nysret Musliu. "Application of Machine Learning to Algorithm Selection for TSP". In: *2014 IEEE 26th International Conference on Tools with Artificial Intelligence*. 2014, pp. 47–54. DOI: 10.1109/ictai.2014.18.

[3] J. Rasku, T. Kärkkäinen, and N. Musliu. "Feature extractors for describing vehicle routing problem instances". In: *OpenAccess Series in Informatics*. Vol. 50. 2016, pp. 7.1–7.13. DOI: 10.4230/OASIcs.SCOR.2016.7.

[4] K. Smith-Miles and J. van Hemert. "Discovering the suitability of optimisation algorithms by learning from evolved instances". In: *Annals of Mathematics and Artificial Intelligence* 61.2 (2011), pp. 87–104. DOI: 10.1007/s10472-011-9230-5.

[5] Meghan Steinhaus. "The Application of the Self Organizing Map to the Vehicle Routing Problem". In: *Open Access Dissertations* (2015). DOI: 10.23860/diss-steinhaus-meghan-2015.

2 Instance space analysis of initial problem space

We first conducted an instance space analysis on the initial 1242 problem instance described in the paper. 75% of these instances were used to train the models, and the remaining 25% used as test data. Model selection was done using 5-fold cross-validation on the training data. Here we include the parameters for the mixture discriminant analysis, the visualisations of the projected space, and prediction and footprint evaluation metrics.

Table 1: Number of subclasses used in mixture discriminant analysis for initial instance spaces.

Budget	Classifier	Class	No. subclasses
Short	Binary HGS	Good	4
		Bad	4
	Binary FILO	Good	3
		Bad	1
	Selection	HGS	4
		FILO	1
Full	Binary HGS	Good	2
		Bad	1
	Binary FILO	Good	2
		Bad	1
	Selection	HGS	4
		FILO	5

Table 2: Test evaluation metrics of predictions for binary performance and selections for the initial problem space.

Budget	Algorithm	Accuracy	Precision	Recall
Short	**HGS**	87.1	84.8	90.3
	FILO	75.6	75.6	87.0
	Best	83.9	84.6	84.3
Full	**HGS**	80.4	89.0	79.7
	FILO	73.6	67.0	85.5
	Best	80.1	80.3	79.3

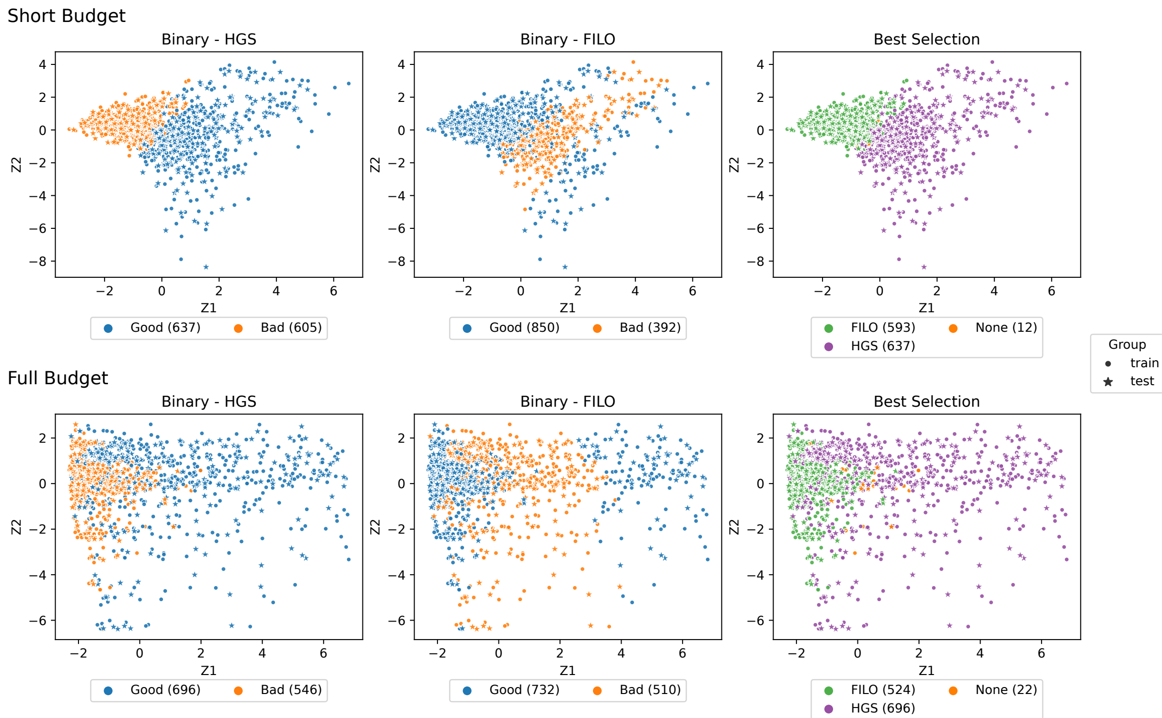

Figure 1: Two dimensional projections and classification for MDA models of good/ bad performance for HGS and FILO as well as the selector. Classifiers for the short budget case appear on the top row and for the full budget in the bottom row.

Table 3: Footprint evaluation metrics for initial instance spaces.

		Good			**Best**		
Budget	Algorithm	Area	Density	Purity	Area	Density	Purity
Short	**HGS**	0.832	0.719	0.777	0.562	0.739	0.896
	FILO	0.496	0.983	0.938	0.228	2.081	0.936
Full	**HGS**	0.924	0.732	0.801	0.814	0.574	0.869
	FILO	0.269	1.905	0.761	0.157	3.347	0.738

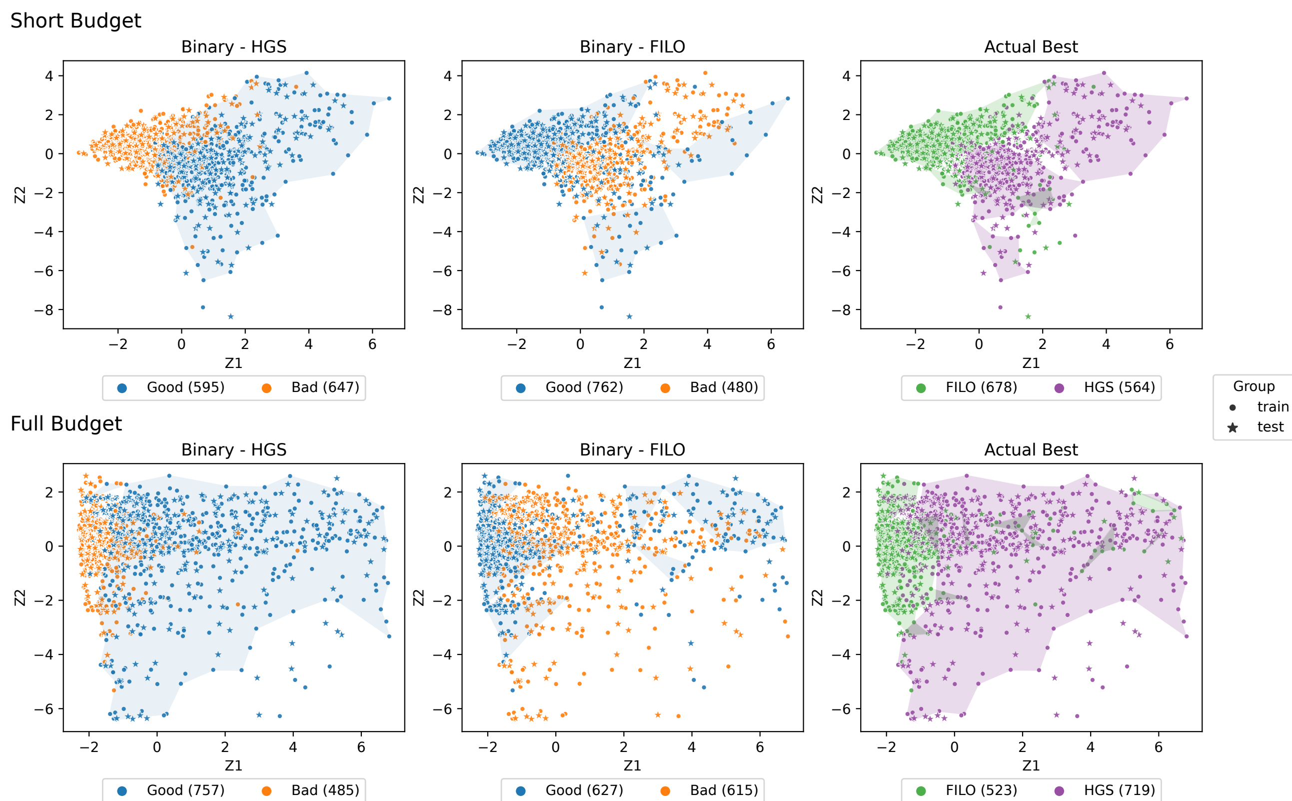

Figure 2: Two dimensional MDA projections and observed binary performance for HGS and FILO, as well as the actual best algorithm. Highlighted are the good/best algorithm footprints. Projections for the short budget case appear on the top row and for the full budget in the bottom row.

3 Additional results for ISA of expanded problem space

Table 4: Test evaluation metrics of predictions for binary performance and selections for the expanded problem space.

Budget	**Algorithm**	**Accuracy**	**Precision**	**Recall**
Short	**HGS**	85.8	84.4	84.0
	FILO	78.6	82.9	82.6
	Best	80.6	80.7	79.5
Full	**HGS**	83.8	92.5	77.8
	FILO	76.0	71.7	84.6
	Best	82.6	83.4	83.1

Table 5: Number of subclasses used in mixture discriminant analysis for expanded instance spaces.

Budget	**Classifier**	**Class**	**No. subclasses**
Short	Binary HGS	Good	1
		Bad	2
	Binary FILO	Good	2
		Bad	2
	Selection	HGS	1
		FILO	2
Full	Binary HGS	Good	2
		Bad	1
	Binary FILO	Good	2
		Bad	2
	Selection	HGS	3
		FILO	2