Application of Big Data Analytics for Understanding the Complexity of Vehicle Routing Problems

Jussi Rasku[†] and Tommi Kärkkäinen[†*]

[†] Faculty of Information Technology, University of Jyväskylä
P.O.Box 35 (Agora), FI-40014 Jyväskylä, Finland
e-mail: `tommi.karkkainen@jyu.fi` - web page: `https://www.jyu.fi/it/en/`

ABSTRACT

Recently we set out to improve the current methods of preprocessing, analyzing, and inferring from vehicle routing problems. We aim to better understand the idiosyncrasies of the problems in order to effectively apply state-of-the-art machine learning methods in solving them. The deeper understanding of these problems is necessary to advance economical, cost effective and environmentally friendly transportation planning.

In this study we concentrate on Capacitated Vehicle routing problem (CVRP), which has been under intensive academic study [see e.g. 7, 3, 10] since it was formulated by Dantzig and Ramser [2] in 1959. Yet, it is not well understood which features make a problem instance (that is, a single task of assigning the deliveries to the vehicles) hard. Similarly, due to the thousands of academic papers reviewing, improving, and introducing new solution methods, there are several possible VRP algorithms to choose from. It is not usually clear which one to choose, or how the configurable parameters of the selected algorithm should be set. Increasing our understanding in how the problem instances, algorithms, and their parameters are interconnected calls for advances in: (i) describing the problem instances and solutions numerically as a feature vector (ii) developing and applying advanced pre-processing and data-analysis methods for the feature data, and (iii) empirical experimentation involving fitting different predictive machine learning models in the domain of vehicle routing problem research. In this study we concentrate on the second task by proposing an application of a novel dimensionality reduction technique for the pre-processing phase.

The VRP solutions are typically represented as a high-dimensional vector of binary decision variables. In a recent study [9] we proposed a set of 74 feature extractors for VRP instances. Using these we can produce 386 statistical attributes that can be used in algorithm performance prediction, algorithm selection, or in related tasks. Unfortunately, the dimensionality of this feature space is too large for many of the popular machine algorithms. This can be experienced by a poor predictive performance of the resulting models. In machine learning this effect is known as 'the curse of dimensionality' of high dimensional datasets [1].

This 'curse' is usually lifted with dimensionality reduction, which relies on the fact that the data is not usually uniformly distributed in the high dimensional space but is concentrated close to a lower-dimensional manifold. In practice, dimensionality reduction can be achieved through with feature selection, or with, as it the case in the method proposed in this study, feature extraction. The popular techniques are principal component analysis (PCA) and linear discriminant analysis (LDA). However, in this study we propose a novel combination of linear [see, e.g. 6] and nonlinear feedforward autoencoders [see, e.g. 4] (that is, single and multilayered perceptron neural networks), whose joint contribution is to estimate the importance of VRP problem features similarly to [5].

We evaluate the performance of the proposed dimensionality reduction method in tasks connected to analyzing and inferring from VRP instances and solutions. Also, besides feature selection, dimensionality reduction allows visualization of the high dimensional data. In addition to applications in feature extraction and selection, we provide a comparison of the proposed method against the multidimensional scaling based method for VRP solution space visualization presented in [8].

---

[*] the presenting author

The results illustrate the different possibilities of the proposed dimensionality reduction method in analyzing and solving vehicle routing problems. However, further experimental and analytical work is required to fully explore the limitations and applicability of the method. The proposed dimensionality reduction method is one tool in the toolbox that is needed when building next generation of route optimization systems. Novel techniques, like the one presented here, help academics and optimization solution providers to envision a path towards greener and more sustainable transportation.

References

[1] Bellman, R. E. (1961). *Adaptive control processes: a guided tour.* Princeton university press.

[2] Dantzig, G. B. and Ramser, J. H. (1959). The truck dispatching problem. *Management Science*, 6(1):80–91.

[3] Eksioglu, B., Vural, A. V., and Reisman, A. (2009). The vehicle routing problem: A taxonomic review. *Computers & Industrial Engineering*, 57(4):1472–1483.

[4] Hänninen, J. and Kärkkäinen, T. (2016). Comparison of four- and six-layered configurations for deep network pretraining. In *Proceedings of the 24th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2016), Bruges, Belgium, April 27-29, 2016.*

[5] Kärkkäinen, T. (2015). Assessment of feature saliency of MLP using analytic sensitivity. In *Proceedings of the 23th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2015), Bruges, Belgium, April 22-24, 2015.*

[6] Kärkkäinen, T. and Saarela, M. (2015). Robust principal component analysis of data with missing values. In P., P., editor, *Proceedings of the 11th International Workshop on Machine Learning and Data Mining in Pattern Recognition (MLDM 2015), Hamburg, Germany, July 20-21, 2015*, Lecture Notes in Artificial Intelligence (9166), pages 140–154. Springer.

[7] Laporte, G. (2009). Fifty Years of Vehicle Routing. *Transport. Sci.*, 43(4):408–416.

[8] Rasku, J., Kärkkäinen, T., and Hotokka, P. (2013). Solution space visualization as a tool for vehicle routing algorithm development. In Collan, M., Hämälainen, J., and Luukka, P., editors, *Proceedings of the Finnish Operations Research Society 40th Anniversary Workshop (FORS40)*, volume 13, pages 9–12. LUT Scientific and Expertise Publications.

[9] Rasku, J., Kärkkäinen, T., and Musliu, N. (2016). Feature Extractors for Describing Vehicle Routing Problem Instances. In Hardy, B., Qazi, A., and Ravizza, S., editors, *5th Student Conference on Operational Research (SCOR 2016)*, volume 50 of *OpenAccess Series in Informatics (OASIcs)*, pages 1–13, Dagstuhl, Germany. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

[10] Toth, P. and Vigo, D. (2014). *Vehicle routing: problems, methods, and applications.* MOS-SIAM Series on Optimization. SIAM, Philadelphia, US.

Keywords: vehicle routing problem, dimensionality reduction, fitness landscape analysis