ESANN 2019 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2019, i6doc.com publ., ISBN 978-287-587-065-0. Available from http://www.i6doc.com/en/.

Feature and Algorithm Selection for Capacitated Vehicle Routing Problems

Jussi Rasku[1], Nysret Musliu[2], and Tommi Kärkkäinen[1]

1- Faculty of Information Technology
University of Jyvaskyla, FI-40014 Jyvaskyla - Finland

2- CD-Lab Artis, Institute of Logic and Computation,
TU Wien, A-1040 Vienna - Austria

Abstract. Many exact, heuristic, and metaheuristic algorithms have been proposed to effectively produce high quality solutions to vehicle routing problems. However, it remains an open question which algorithm is the most appropriate for solving a given problem instance, mostly because the different strengths and weaknesses of algorithms are still not well understood. We propose an extensive feature set for describing capacitated vehicle routing problem instances and illustrate how it can be used in algorithm selection, and how different feature selection approaches can be used to recognize the most relevant features for this task.

1 Introduction

Vehicle routing problem is one of the most intensively studied problems in operations research because of its many applications. Recently we proposed using an autoencoder based approach [6] to recognize interesting Capaciated Vehicle Routing Problem (CVRP) features [7]. These features were proposed earlier in [14], where we demonstrated that the features can be used in algorithm configuration and unsupervised learning. In this study we focus on feature selection and gaining a deeper understanding on feature relevance in meta-optimization. More specifically, we study the predictive accuracy of subsets of the proposed features in a task of algorithm selection of classical CVRP heuristics.

Automatically selecting the most suitable algorithm for solving Traveling Salesman Problems (TSPs) has been studied for example in [16, 5, 10] and for Vehicle Routing Problems (VRPs) in [8, 11, 21, 18]. However, recognizing the most relevant features has received little attention. The questions by Smith-Miles and van Hemert [16] are relevant here: Which of the features prove useful when predicting algorithm performance and solution quality?

According to Rice's framework [15, 17] the requirements for successfully applying algorithm selection on a given problem are: (i) that there are large and diverse problem instance sets available, (ii) there are several competitive algorithms for solving those problem instances, (iii) there is a way to measure the algorithm performance or accuracy, and (iv) access to features that are suitable to describe the problem instances can be recognized. Addressing (i) and (iii) is trivial as the quality of a VRP heuristic algorithm is usually measured with the optimality gap and there are many well-known problem sets for CVRP [22]. For the other requirements we have proposed feature extractors for CVRP in

ESANN 2019 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2019, i6doc.com publ., ISBN 978-287-587-065-0. Available from http://www.i6doc.com/en/.

[14] and have implemented 15 classical heuristics in [13]. To recognize the most relevant features, we propose two algorithm selection scenarios and use them to show how an ensemble of VRP algorithms can yield high quality solutions.

2 VRP Feature Extraction and Selection

In this study, we consider the classical capaciated vehicle routing problems. Through the years, several algorithms have been proposed to solve them effectively [20]. The modern metaheuristics rely on stochasticity to produce high quality results in a reasonable time, but the stochasticity can make experimental study of heuristic algorithms tricky [12]. Furthermore, the computational experiment conventions have been inconsistent and limited, which makes building algorithm selection scenarios from the results published by different authors impossible. To sidestep these issues, we used our open-source library of classical CVRP algorithms [13]. The classical heuristics are ideal for experimentally verifying the suitability of our proposed feature set because of their deterministic behavior and ability to solve also the larger problem instances.

Problem instances come from the CVRPLIB[1] problem sets A, B, E, F, M, P, X, CMT, RT, Golden, and Li, together with additional CVRP instance sets Gaskell, V, and van Breedam. Together, these form a heterogeneous set of 454 problem instances. For descriptions of these sets we refer to [22] and [13]. To characterize the instances, we further extended our comprehensive set of CVRP features proposed in [14]: The measurements of the number of checked and accepted local search moves, as proposed earlier for TSP in [10], are now included. Additionally, the ratio between integer and non-integer values and their distribution from an exact solver probing are recorded similarly to [5]. The problem instances with the maximum route duration constraint in this study warranted additional features describing the tightness of this constraint. This was calculated using a greedy TSP algorithm and the DBSCAN clusters. In total, our feature extractor set[2] produces 433 values for each instance.

As one can see, we followed the recommendation of [3] and erred on the side of being inclusive instead of risking to omit useful information in the feature construction stage. Unfortunately, this means that the proposed feature set is quite extensive and the issues due to the curse of dimensionality must be addressed, especially since we only have relatively few problem instances. The canonical way is to use principal component analysis (PCA), but also many *feature selection* (FS) methods can be used. These allow recognizing a relevant subset F_r of the set of features [3], which has an additional advantage of being able to omit the computation of any unnecessary features.

[1] http://vrp.atd-lab.inf.puc-rio.br/index.php/en/

[2] http://users.jyu.fi/\%7Ejuherask/selection/FEtable.pdf

ESANN 2019 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2019, i6doc.com publ., ISBN 978-287-587-065-0. Available from http://www.i6doc.com/en/.

3 Experimental Procedure

Infinite feature values were replaced with a value 10 times the largest real valued measurement and the features were normalized by scaling the values to the [0, 1] range. Furthermore, similarly to Pihera and Musliu [10], we did our experiments also with discretized feature data. As a discretization algorithm we used the MDLP multi-interval algorithm of Fayyad and Irani [1].

Using the algorithm performance and accuracy data from [13] we set the classification label to be the best algorithm for each problem instance. Here, the total solution cost was the primary comparison criteria and the ties were resolved by solution time. Hence, the class label of each problem instance is determined by the best algorithm. Predicting this class among the 14 alternatives (see Fig. 3 and [13] for a list) is our first scenario. Foreseeing that this is a difficult task, we also prepared an easier scenario where the class is predicted among the three most successful algorithms: GAP, PTL, and PS|G2P.

In addition to PCA we applied three different FS methods: minimal-redundancy-maximal-relevance criterion (mRMR) [9], Correlation-based Feature Selection (CFS) [4], and feature boosting with extremely randomized trees [2] (ERTB). Then, four classifiers: 3-nearest neighbor (3NN); multilayer perceptron Softmax classifier (MPL) with hidden layer of $|F_r|$ neurons, sigmoid activation functions, and quasi-Newton backpropagation; random forest with 100 trees and max. depth of 10 (RF); and C-support vector machine (SVM) with $C = 1.0$ and $\gamma = 1/|F_r|$ were trained and evaluated. We used leave-one-out cross validation with and without MDLP discretization, and GNU parallel [19] and Scikit-learn on a 72 core F-series Azure VM to compute the results.

4 Results and Conclusions

It turned out that many of the proposed features do not have any MDLP [1] cut candidates that satisfy the minimum description length criterion. For these, all values belong to the same bin and they no longer contribute to the algorithm selection task. With those features removed, we are left with 163 features on the 14 algorithm scenario and 190 on the three algorithm scenario.

According to the feature importance analysis (see Fig. 4), there is a sharp change in feature importance after 10 features. These 10 features together with the ones recognized by other feature selection methods are presented in Fig. 1 together with the earlier results from the autoencoder (AE) based method [7]. Unsurprisingly, the final scoring reveals that the local search probing (LSP) features are highly relevant in the heuristic selection task, but also features related to exact solving attempt (BCP), constraints (DC), and nearest neighbor digraph (NN) are important.

For the scenario of selecting between the 14 heuristics the baseline accuracy (majority class heuristic) is 21.6%. The best measured accuracy was 48.5% with the RF classifier and 100 features from the mRMR feature selection. This was also the upper limit for the number of features in our experiments, and it

FS \ Rank	1	2	3	4	5	6	7	8	9	10
CFS	BCP2 (rt.)	LSP2 (avg)	NN3 (avg)	DC12 (LB)	LSP4 (SD)	NN11 (min)	ND2 (CV)	DC3 (dist.)	LSP7 (2q.m.)	DC6 (rt.)
mRMR	BCP2 (rt.)	LSP1 (avg)	NN3 (avg)	LSP7 (2q.m.)	DC6 (rt.)	LSP1 (CV)	DC3 (dist.)	NN11 (min)	ND2 (avg)	LSP2 (avg)
ERTB	LSP1 (CV)	LSP2 (avg)	LSP1 (avg)	LSP12 (CV)	NN3 (CV)	ND4 (min)	LSP10 (SD)	LSP2 (SD)	DC3 (dist.)	DC6 (rt.)
AE [7]	LSP12 (SD)	LSP7 (1q.c.)	LSP7 (1q.w.)	LSP4 (kurt)	LSP2 (CV)	LSP12 (CV)	LSP1 (SD)	ND1 (avg)	ND4 (avg)	LSP3 (skew)
Score	BCP2 (rt.) 20 pts.	LSP2 (avg) 19 pts.	LSP1 (avg) 17 pts.	NN3 (avg) 16 pts.	LSP12 (SD/CV) 16 pts.	LSP1 (CV/SD) 14 pts.	DC3 (dist.) 9 pts.	LSP7 (1q.c.) 9 pts.	LSP7 (2q.m.) 9 pts.	DC6 (rt.) 8 pts.

Fig. 1: Top 10 discretized features on the three algorithm scenario.

is possible that RF could have benefited from additional features. The most problematic decision boundaries seem to be between GAP, PLT, and PS|G2P heuristics (see Fig. 3). Please note that in this scenario the best algorithm can be determined with a very small margin. If we accept predictions where the heuristic is among the three best for each instance the accuracy rises to 75.3%.

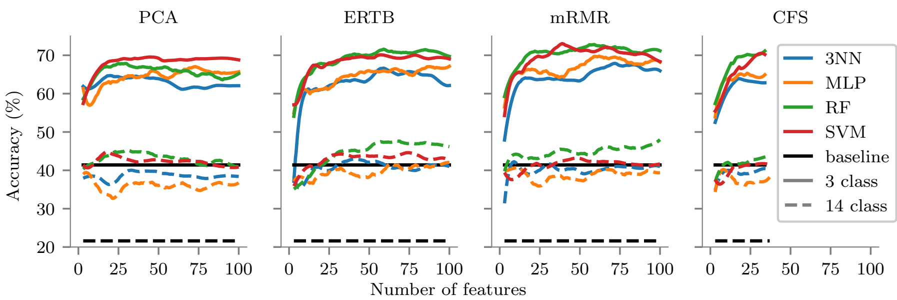

Fig. 2: Savitzky-Golay smoothed classifier accuracy on the discretized data.

For the three algorithm scenario, the baseline is 41.4% and the best accuracy is 74.0% with 75 features chosen using ERTB (see Fig. 4). The accuracy level and the improved the selection accuracy when using discretization are similar to the 3-class experiments in [10] where the accuracy ranged from 64% to 69%. Steinhaus [18] managed to choose the best performing algorithm with 84 % accuracy between three alternatives. However, in our study the algorithms were chosen according their proven performance, which makes our three class scenario more challenging than the one of Steinhaus.

ESANN 2019 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2019, i6doc.com publ., ISBN 978-287-587-065-0. Available from http://www.i6doc.com/en/.

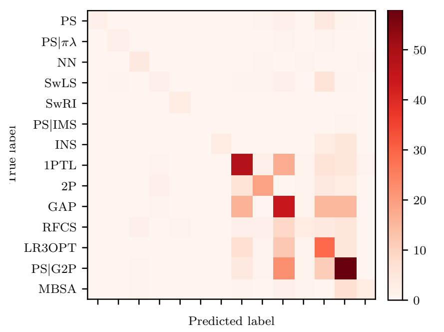

Fig. 3: Confusion matrix for the 14 classical CVRP heuristics.

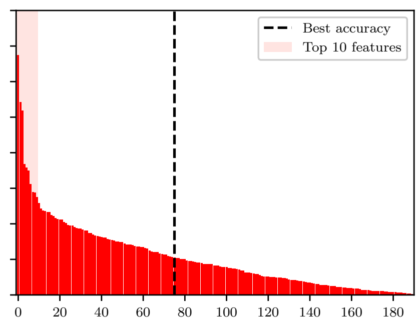

Fig. 4: Feature importances on discretized values in three class scenario.

Regarding the validity of our results, please note that we did not use a separate training set for the feature selection or discretization due to the limited number of samples. We acknowledge that this may induce some positive bias to the selection accuracy. In an extended study it would be advisable to use nested cross-validation, although it would probably necessitate the use of problem generator with the related pitfalls [12, p. 271-273]. Furthermore, while our feature extraction framework is already quite comprehensive, it would be possible to further extend it to describe other VRP variants, and, e.g., to include additional metrics typically used in TSP and VRP solution space analysis. Here, algorithm selection of automatically configured modern metaheuristics would be the most natural, albeit computationally intensive, direction to extend our study to. Aforementioned extensions to the experiments would enable a more in depth analysis of the discriminatory power of the most important features.

Taken together, feature and algorithm selection allowed us to recognize the most interesting and relevant features among the extensive set of 433 features proposed in this study. Additionally, we have shown that the feature set has a good discriminating power and it can be used to leverage an ensemble of vehicle routing heuristics with good results.

References

[1] Fayyad, U. M. and Irani, K. B. (1993). Multi-interval discretization of continuous valued attributes for classification learning. In *In Proceedings of the 13th International Joint Conference on Artificial Intelligence*, page 1022–1029. Morgan-Kaufmann.

[2] Geurts, P., Ernst, D., and Wehenkel, L. (2006). Extremely randomized trees. *Machine learning*, 63(1):3–42.

[3] Guyon, I. and Elisseeff, A. (2006). An introduction to feature extraction. In *Feature extraction*. Springer.

[4] Hall, M. A. (2000). Correlation-based feature selection for discrete and numeric class machine learning. In *Proceedings of the Seventeenth International Conference on Machine Learning (ICML '00)*, pages 359–366. Morgan-Kaufmann.

ESANN 2019 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2019, i6doc.com publ., ISBN 978-287-587-065-0. Available from http://www.i6doc.com/en/.

[5] Hutter, F., Xu, L., Hoos, H. H., and Leyton-Brown, K. (2014). Algorithm runtime prediction: Methods & evaluation. *Artificial Intelligence*, 206:79–111.

[6] Kärkkäinen, T. (2015). Assessment of feature saliency of MLP using analytic sensitivity. In *23rd European Symposium on Artificial Neural Networks (ESANN 2015)*, pages 273–278.

[7] Kärkkäinen, T. and Rasku, J. (2019). Application of a knowledge discovery process to study instances of capacitated vehicle routing problems. In Diez, P., Neittaanmäki, P., Periaux, J., Tuovinen, T., and Jordi, P.-P., editors, *Computation and Big Data for Transport - Digital innovations in surface and air transport systems*. Springer. (To appear in).

[8] Nygard, K. E., Juell, P., and Kadaba, N. (1990). Neural networks for selective vehicle routing heuristics. *ORSA Journal on Computing*, 2(4):353–364.

[9] Peng, H., Long, F., and Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on pattern analysis and machine intelligence*, 27(8):1226–1238.

[10] Pihera, J. and Musliu, N. (2014). Application of machine learning to algorithm selection for TSP. In *Tools with Artificial Intelligence (ICTAI), IEEE 26th International Conference on*, pages 47–54. IEEE.

[11] Potvin, J.-Y., Lapalme, G., and Rousseau, J.-M. (1990). Integration of AI and OR techniques for computer-aided algorithmic design in the vehicle routing domain. *Journal of the Operational Research Society*, 41(6).

[12] Rardin, R. L. and Uzsoy, R. (2001). Experimental evaluation of heuristic optimization algorithms: A tutorial. *Journal of Heuristics*, 7(3):261–304.

[13] Rasku, J. (2019). Meta-survey and implementations of classic capacitated vehicle routing heuristics with reproductions of earlier results. Manuscript, University of Jyväskylä, Finland. (to appear).

[14] Rasku, J., Kärkkäinen, T., and Musliu, N. (2016). Feature Extractors for Describing Vehicle Routing Problem Instances. In *5th Student Conference on Operational Research (SCOR 2016)*, volume 50 of *OpenAccess Series in Informatics (OASIcs)*, pages 7:1–7:13, Dagstuhl, Germany.

[15] Rice, J. (1976). The algorithm selection problem. *Advances in Computers*, 15:65–118.

[16] Smith-Miles, K. and van Hemert, J. (2011). Discovering the suitability of optimisation algorithms by learning from evolved instances. *Annals of Mathematics and Artificial Intelligence*, 61(2):87–104.

[17] Smith-Miles, K. A. (2009). Cross-disciplinary perspectives on meta-learning for algorithm selection. *ACM Computing Surveys (CSUR)*, 41(1):6.

[18] Steinhaus, M. (2015). *The Application of the Self Organizing Map to the Vehicle Routing Problem*. PhD thesis, University of Rhode Island, US.

[19] Tange, O. (2011). GNU parallel - the command-line power tool. *;login: The USENIX Magazine*, 36(1):42–47.

[20] Toth, P. and Vigo, D. (2014). *Vehicle Routing: Problems, Methods, and Applications*. SIAM, PA, USA.

[21] Tuzun, D., Magent, M. A., and Burke, L. I. (1997). Selection of vehicle routing heuristic using neural networks. *International Transactions in Operational Research*, 4(3):211–221.

[22] Uchoa, E., Pecin, D., Pessoa, A., Poggi, M., Vidal, T., and Subramanian, A. (2017). New benchmark instances for the capacitated vehicle routing problem. *European Journal of Operational Research*, 257(3):845–858.