A Scalable Approach for Vehicle Routing Problem with Reinforcement Learning

C.Y. Lo and C. K. M. Lee
Department of Industrial and Systems Engineering
The Hong Kong Polytechnic University
Hong Kong, China
cheung-yu.lo@connect.polyu.hk

Keywords
Vehicle Routing Problem, Logistic Routing, Reinforcement Learning

Abstract
This paper proposed a scalable approach for solving vehicle routing problems by breaking down a large scale problem to small sub-problems where the reinforcement learning model is employed recursively. The approach is inspired by large scale vehicle routing problem in a city with some pre-defined district clusters (E.g. Hong Kong). The result within every sub-problems is obtained using a sequence to sequence deep learning network trained by policy gradient. Despite the time required for training, the model outperforms traditional heuristics models with a comparable amount of time.

Bello et al. (2016) have proposed a framework for solving neural combinatorial problems with reinforcement learning. A pointer network (Vinyals et al., 2015) has been employed to process sequential inputs. Nazar et al. (2018) have replaced the pointer network with an LSTM recurrent neural network that accepts both static and dynamic inputs. Our proposed model broke down the vehicle routing problem to a city level problem (VRP for a group of districts) and many district level problems (VRP for groups of nodes) given the districts labels. Each problem is solved using the reinforcement learning model given its generic properties proven in previous studies.

The model receives an input array $X = \{x_i, i = 1, \dots, M\}$. Each input $x_i = \{l : d,\ n\}$ composes of the concatenation of a 2D array, the location for each node and the demand in the specific location, as well as the corresponding district n. The location of nodes are represented as latitudes and longitudes which will be mapped to cartesian coordinates. The result will be passed to the encoder of the network. An attention mechanism is employed for constraining the amount of computation in the model.

Parameters of the network are updated using the REINFORCE algorithm in which the baseline in the equation is estimated by a critic network which learns the approximated time required for tours travelling. The network will be trained and updated concurrently with the sequence to sequence (actor) network. The rewards generated by the environment equal to the negative amount of tour travelling time. The model objective is approaching the optimal policy π^* that assign high possibility to the tour that require less amount of time. The stochastic policy generated by the network are calculated as follow :

$$p(\pi | s) = \prod_{j=1}^{n} p(\pi(j) \mid \pi(< j), s)$$

where s represents the previous state of the network.

The centroid of the districts will be calculated and pass to the city level model. In every step of the city level model, a district level model with independent parameters is executed. The result of the district level model will be passed back to the city level model and the city level model will proceed to the next step by policy sampling.

The preliminary results indicated that the model outperforms traditional heuristic algorithms such as Clarke-Wright Savings Heuristic and Sweep Heuristic. The significance of the study is that it provides an approach for solving large scale vehicle routing problem which is highly computational expensive with traditional algorithms.

Acknowledgements

The research is supported by The Hong Kong Polytechnic University. The authors would like to express their gratitude to the Department of Industrial and Systems Engineering of the Hong Kong Polytechnic University for the support in the project.

References

Bello, I., Pham, H., Le, Q. V., Norouzi, M., & Bengio, S. (2016). Neural combinatorial optimization with reinforcement learning. *arXiv preprint arXiv:1611.09940*.

Nazari, M., Oroojlooy, A., Snyder, L., & Takac, M. (2018). Reinforcement Learning for Solving the Vehicle Routing Problem. In *Advances in Neural Information Processing Systems*(pp. 9861-9871).

Vinyals, O., Fortunato, M., & Jaitly, N. (2015). Pointer networks. In *Advances in Neural Information Processing Systems* (pp. 2692-2700).

Biographies

C.Y. Lo is an undergraduate student in the Department of Industrial and Systems Engineering at the Hong Kong Polytechnic University, Hong Kong, China. His interested research areas include reinforcement learning and operation research.

C. K. M. Lee is currently an assistant professor in the Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hong Kong. She is the program leader of BSc(Hons) Enterprise Engineering with Management. She obtained her PhD and BEng degree from The Hong Kong Polytechnic University. She was awarded Bronze Award of 16th China National Invention Exhibition Award in 2006 and Outstanding Professional Service and Innovation Award, The Hong Kong Polytechnic University in 2006. Dr Lee has authored or co-authored more than 100 journal papers. Her current research areas include logistics information management, smart manufacturing, application of internet of things and data mining techniques.