Selection of Auxiliary Objectives in the Travelling Salesman Problem using Reinforcement Learning

Irina Petrova
ITMO University
49 Kronverkskiy ave.
Saint-Petersburg, Russia
petrova@rain.ifmo.ru

Arina Buzdalova
ITMO University
49 Kronverkskiy ave.
Saint-Petersburg, Russia
abuzdalova@gmail.com

ABSTRACT

Auxiliary objectives may be used to reduce number of iterations of an evolutionary algorithm (EA). The corresponding approach is called multi-objectivization.

We consider two multi-objectivization methods: EA+RL and MOEA+RL, where MOEA is a multi-objective EA, RL is reinforcement learning. In these methods, RL is used to select an objective during optimization process. In EA+RL only the selected objective is optimized, so a single-objective EA is used. In MOEA+RL the selected objective is optimized together with the target objective. Previously in these methods, RL for stationary environments was used. Recently, a new non-stationary RL algorithm was proposed. This algorithm was specially developed for the case when behaviour of auxiliary objectives changes during optimization process. However, this RL algorithm was tested only with EA+RL on some simple problems.

In the present work we apply EA+RL and MOEA+RL with stationary and non-stationary RL to the travelling salesman problem (TSP) and compare them with the previously used multi-objectivization methods. We also analyze different types of auxiliary objectives for TSP. For the most of the considered problem instances, EA+RL and MOEA+RL for non-stationary environment perform better than the other considered methods.

CCS Concepts

•**Computing methodologies** → *Genetic algorithms; Reinforcement learning;* •**Mathematics of computing** → *Paths and connectivity problems;*

Keywords

multi-objectivization; helper-objectives; non-stationarity

1. INTRODUCTION

Consider multi-objectivization approaches [3, 4]. The approaches proposed by Knowles et al. [4] and Jähne et al. [2] are based on decomposition of the target objective into several auxiliary objectives. These auxiliary objectives are optimized simultaneously instead of the target objective.

Another approach proposed by Jensen [3] is to use some auxiliary objectives and optimize one of them together with the target objective. An auxiliary objective is randomly selected at each step of the algorithm [3]. Another way to select an objective is to use an ad-hoc heuristic [5]. The first approach is general, but does not use information about the problem. The second one is proposed for the specific problem and may not be applicable to other problems. The MOEA+RL method deals with these issues [1, 7]. In this method reinforcement learning [9] is used to select an objective. Also there exists the EA+RL approach which is similar to MOEA+RL. The difference is that in EA+RL only one objective (target or auxiliary) is optimized at a time.

In RL, an agent applies an action to an environment, then the environment returns some representation of its state and a numerical reward to the agent, and the process repeats. In EA+RL and MOEA+RL, EA is treated as an environment. An action corresponds to selection of an objective. Reward is based on the difference of the target objective values in two consecutive iterations. In early studies, it was implied that the environment was stationary and stationary RL algorithms were used. The environment is stationary if the obtained reward depends only on the applied action and the state of the environment [9]. However, if properties of the auxiliary objectives change during optimization, the reward for the same action can be different in the same state. In this case non-stationary RL algorithms should be used. In our previous work a non-stationary RL algorithm was proposed [8]. It was used in EA+RL to solve a test problem. In this work we apply this RL approach to solve TSP.

2. SOLVING TSP

There are multi-objectivization approaches proposed by Knowles et al. [4], Jensen [3] and Jähne et al. [2] which were used to solve TSP. In all of them for different individuals the same auxiliary objective may or may not help in optimizing the target objective, which leads to non-stationarity.

We compared EA+RL and MOEA+RL with these three approaches. Stationary and non-stationary RL algorithms were considered. Description and results of the experiment with EA+RL as well as description of the experiment with MOEA+RL are presented in supplementary materials[1].

GECCO '15 July 11-15, 2015, Madrid, Spain

ACM ISBN 978-1-4503-3488-4/15/07.

DOI: http://dx.doi.org/10.1145/2739482.2764646

[1] https://github.com/iruuunechka/papers/blob/master/GECCO-2015/tsp.pdf

Table 1: **Average target values. The dark (light) background corresponds to the first (second) best result.**

Instance	Optimum	NS MOEA+RL	S MOEA+RL	Jähne	Jensen-Jähne	Jensen
kroB100	22141	22144	22145	22150	22158	22155
kroD100	21294	21342	21353	21344	21349	21347
kroE100	21294	22093	22095	22169	22095	22100
eil101	629	641.39	641.84	641.50	641.59	641.95
pr124	59030	59030	59030	59030	59032	59052
bier127	118282	118324	118394	118387	118408	118394
pr136	96772	96975	97000	96980	97193	97063
kroA150	26524	26540	26558	26533	26557	26558
kroB150	26130	26153	26166	26170	26166	26174
pr152	73682	73693	73702	73904	73820	73821
pr439	107217	107675	107677	107748	108035	107743
rat575	6773	6869	6872	6874	6863	6877
pr1002	259045	263158	263318	263425	263184	263189

Experiment results of MOEA+RL are shown in Table 1. For each problem, the average target objective value is presented. The first two columns contain names of the instances and their best known solutions. The next four colums contain results of MOEA+RL with the non-stationary RL algorithm (NS MOEA+RL), MOEA+RL with stationary ε-greedy Q-learning (S MOEA+RL), Jähne et al. (Jähne) and Jensen (Jensen-Jähne) approaches. In all these approaches, two auxiliary objectives proposed by Jähne et al. were used. The last column contains results of Jensen approach which was run on ten auxiliary objectives proposed by Jensen.

According to the multiple signed test, MOEA+RL with non-stationary RL is distinguishable from the other methods at the level of statistical significance $p = 0.05$. To sum up, non-stationary MOEA+RL with the auxiliary objectives proposed by Jähne et al. turns to be the most efficient approach for the considered instances.

3. CONCLUSION

We applied the recently proposed non-stationary RL algorithm together with EA+RL and MOEA+RL for solving TSP. This approach outperformed other considered methods. The obtained results confirm that auxiliary objectives proposed by Jähne et al. are efficient for solving TSP.

We considered two major ways of using auxiliary objectives. The first way is to simultaneously optimize the auxiliary objectives instead of the target objective [4]. Most of the recent research is focused on this approach [2, 6]. The second way is to optimize the target objective together with a dynamically selected auxiliary objective [3]. The results of the present work suggest that the second approach may be more efficient than the first one when a proper selection method is used. Particularly, for the considered instances of TSP, the second approach with the non-stationary RL based selection outperformed the other methods.

This work was partially financially supported by the Government of Russian Federation, Grant 074-U01.

4. REFERENCES

[1] A. Buzdalova and M. Buzdalov. Increasing Efficiency of Evolutionary Algorithms by Choosing between Auxiliary Fitness Functions with Reinforcement Learning. In *Proceedings of the International Conference on Machine Learning and Applications*, volume 1, pages 150–155, 2012.

[2] M. Jähne, X. Li, and J. Branke. Evolutionary algorithms and multi-objectivization for the travelling salesman problem. In *Proceedings of the 11th Annual Conference on Genetic and Evolutionary Computation*, GECCO '09, pages 595–602, New York, NY, USA, 2009. ACM.

[3] M. T. Jensen. Helper-Objectives: Using Multi-Objective Evolutionary Algorithms for Single-Objective Optimisation: Evolutionary Computation Combinatorial Optimization. *Journal of Mathematical Modelling and Algorithms*, 3(4):323–347, 2004.

[4] J. D. Knowles, R. A. Watson, and D. Corne. Reducing Local Optima in Single-Objective Problems by Multi-objectivization. In *Proceedings of the First International Conference on Evolutionary Multi-Criterion Optimization*, pages 269–283. Springer-Verlag, 2001.

[5] D. F. Lochtefeld and F. W. Ciarallo. Deterministic Helper-Objective Sequences Applied to Job-Shop Scheduling. In *Proceedings of Genetic and Evolutionary Computation Conference*, pages 431–438. ACM, 2010.

[6] D. F. Lochtefeld and F. W. Ciarallo. An analysis of decomposition approaches in multi-objectivization via segmentation. *Appl. Soft Comput.*, 18:209–222, 2014.

[7] I. Petrova, A. Buzdalova, and M. Buzdalov. Improved Helper-Objective Optimization Strategy for Job-Shop Scheduling Problem. In *Proceedings of the International Conference on Machine Learning and Applications*, volume 2, pages 374–377. IEEE Computer Society, 2013.

[8] I. Petrova, A. Buzdalova, and M. Buzdalov. Improved selection of auxiliary objectives using reinforcement learning in non-stationary environment. In *Proceedings of the International Conference on Machine Learning and Applications*, pages 580–583, 2014.

[9] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA, USA, 1998.