Fairness in Transport Network Design - A Multi-Objective Reinforcement Learning Approach

Dimitris Michailidis
Civic AI Lab, University of Amsterdam
Amsterdam, The Netherlands
d.michailidis@uva.nl

Willem Röpke
AI Lab, Vrije Universiteit Brussel
Brussels, Belgium
willem.ropke@vub.be

Sennay Ghebreab
Civic AI Lab, University of Amsterdam
Amsterdam, The Netherlands
s.ghebreab@uva.nl

Diederik M. Roijers
Urban Innovation and R&D, City of Amsterdam
Amsterdam, The Netherlands
d.roijers@amsterdam.nl

Fernando P. Santos
Civic AI Lab
University of Amsterdam
Amsterdam, The Netherlands
f.p.santos@uva.nl

ABSTRACT

Optimizing urban transportation networks can improve the lives of millions of citizens worldwide. The problem of generating new transportation lines, which maximize the levels of satisfied travel demand is, however, a complex endeavor. This problem is known as the Transport Network Design Problem (TNDP) and it is NP-hard. On top of efficiency concerns, it is nowadays fundamental to also consider the development of transportation systems that contribute to alleviating social inequalities. Which technical approaches can we employ to tackle both efficiency and fairness in TNDP? In this paper, we explore Multi-Objective Reinforcement Learning (MORL) as a tool to design efficient and fair transportation networks. We start by formulating Multi-Objective transport network design problems as Multi-objective Markov decision processes. We highlight the main challenges of introducing multiple objectives in TNDP. Finally, we describe novel methodologies that can be used to tackle this problem. With this paper, we hope to start a line of research that can provide suitable decision support for TNDP, by providing alternative solutions with different trade-offs between (different metrics of) fairness, efficiency, and cost.

1 INTRODUCTION

Developing efficient and inclusive transportation systems is a fundamental challenge of urban planning, whose success can improve the life quality of millions of citizens [13]. Computationally, the problem of designing an efficient transportation network — i.e., generating transportation lines that maximize satisfied travel demand — has been formalized through the so-called transport network design problem (TNDP) [6]. Although having the potential to positively impact the lives of many people, solving TNDPs is challenging: The transport network design problem (TNDP) is an NP-hard optimization problem and advancing methods to solve it is an active research line in computer science and operations research [6].

TNDP has traditionally been addressed through integer optimization [12], simulated annealing [5], genetic [19] and other, heuristic-based algorithms [9, 26]. These methods, while successful, require a long list of expert-derived constraints and are hard to generalize. Researchers have to limit the search space (e.g. by pre-defined corridors and candidate stations) to tackle intractability. Recently, Deep Reinforcement Learning (Deep RL) models have been proposed that outperform previous approaches in satisfied demand, without the need to specify multiple constraints [30]. Given the large state-action spaces considered, Deep RL can further contribute to devising scalable and generalizable solutions to the TNDP: indeed, it was recently demonstrated to achieve state-of-the-art results in terms of optimizing transportation demand (i.e., efficiency) [30]. This, combined with its generalizability capabilities, has made it a prominent method for various real-world optimization applications [11, 16, 31].

Despite Deep RL leading to promising results in efficiency for the TNDP, dealing with fairness issues has only recently started to be explored. In particular, it has been shown that traditional TNDP optimization, which focuses on efficiency, can lead to solutions that disproportionately benefit the societal groups with the highest mobility, while ignoring other, less advantaged groups [15]. Such a design approach might increase average efficiency but can also lead to further enhancing existing patterns of inequality and segregation in cities [17, 18]. It is therefore crucial to look beyond maximizing efficiency and consider solutions that offer desirable trade-offs between multiple objectives when designing public transportation.

In this work-in-progress paper, **we propose studying fairness in transportation network design via a multi-objective reinforcement learning approach**. We argue that fairness is a multifaceted notion that cannot be tackled using a single, scalar utility and propose a formalization of the problem as a Multi-Objective Markov Decision Process (MOMDP). We discuss the main challenges with this formulation and assess the usage of novel methods for tackling it in two real-world environments: Xi'an, China, and Amsterdam, Netherlands (Figure 2).

Multiple objectives in TNDP are not an entirely new concept. In fact, the original problem can arguably be classified as a multi-objective optimization problem, in which the cost of the operator and the user are taken into account. Previous works have studied the trade-offs between these two objectives, using heuristics [4], meta-heuristics [14] and genetic algorithms [1, 19]. However,

Proc. of the Adaptive and Learning Agents Workshop (ALA 2023), Cruz, Hayes, Wang, Yates (eds.), May 9-10, 2023, Online, https://ala2023.github.io/. 2023.

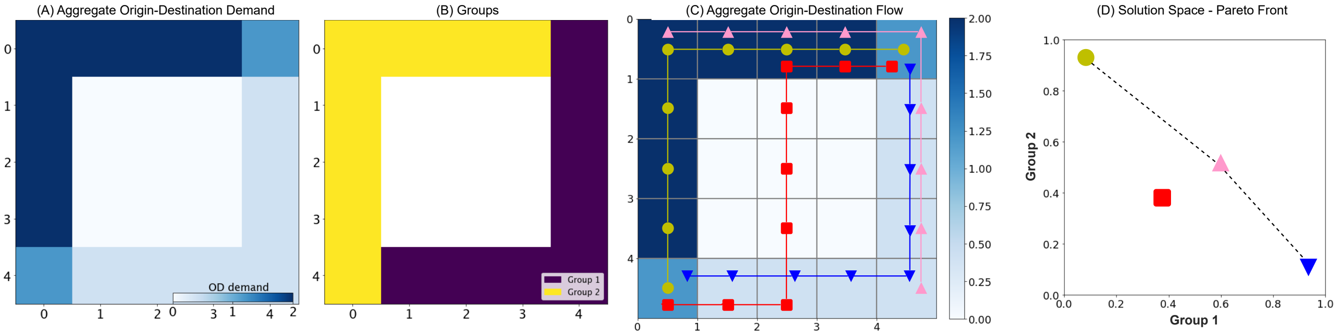

Figure 1: To clarify the multiple conflicting objectives in a TNDP environment, we define a toy example. We consider two groups (panel B) with different origin-destination (OD) demands (panel A). We show that different optimization choices can lead to different results for the two groups. Panel C represents possible new transportation lines and Panel D depicts how such solutions fare in terms of satisfied OD flows for the two groups and outlines the possibly optimal solutions. The yellow line (circle) leads to the highest satisfied demand for group 2, while the pink one (triangle) is the best for group 1, the pink triangle represents a different optimal trade-off between the two, while the red square represents a solution that is dominated by the pink triangle one. We expect that building a model that learns the different trade-offs leads to useful decision support for transport planners.

this approach has yet to be comprehensively extended to studying fairness between groups. In this paper, we adopt recent reinforcement learning formulations and propose extensions and methods to tackle group-based fairness as multiple-objectives [15, 30]. Specifically, we will consider how different societal groups benefit from the generated transportation line, in terms of satisfied origin-destination mobility demand.

The remainder of the paper is structured as follows: first, we state the multi-objective transport network design problem (Section 2) and formulate it as a Multi-Objective Markov Decision Process (Section 3). We continue by presenting environments to apply the formulation (Section 4) and outlining the main challenges of introducing multiple objectives to the TNDP (Section 5). We conclude by describing novel methodologies that can be used to tackle the problem (Section 6).

2 TRANSPORT NETWORK DESIGN PROBLEM (TNDP)

We consider the TNDP where the goal is to generate a graph $G(N, E)$, which represents a transport line, where N (nodes) are locations to place stations on, and E (edges) are connections between them. Depending on the modality of the line, the graph can be directed (bus, tram, etc.) or undirected (metro, subway, etc). Since we deal with metro networks, the graph is undirected — as in previous works [30].

The city is represented as a two-dimensional grid environment $H^{n\times m}$. The traditional optimization objective is defined as the *total captured travel demand* of the created line, expressed as a function U_{od} of the estimated Origin-Destination (OD) matrix [7, 8], where $\{U_{od}\}_{ij}$ represents traveling magnitude from location i to j. The OD matrix is considered deterministic and does not change during the episode. The total number of selected locations is hard-constrained by a construction budget B, a station number limit T, and a set of direction-based constraints, so as to avoid unorthodox line shapes [30]. We use $U_{od}(G(N, E))$ to denote the demand covered by the new transportation graph G. This is calculated by summing the origin-destination demand between all nodes in the graph:

$$U_{od}(G(N,E)) = \sum_{i\in N}\sum_{j\in N}\{U_{od}\}_{ij}, i \neq j \tag{1}$$

Note that despite the transport line being represented as a graph, the environment in which we apply the problem is grid-based. Nodes are grid cells and edges are connections between them. Given the above, the problem is formalized as follows:
Find the transportation graph $G(N, E)$, such that:

$$\begin{aligned} \max \quad & U_{od}(G(N,E)) \\ \text{s.t.} \quad & cost(G) \leq B \\ & |N| \leq T \end{aligned} \tag{2}$$

Where:

- $N \subseteq H^{n\times m}$
- $E = \{(h_i, h_j) : h_i, h_j \in N, h_i \neq h_j\}$

In the traditional objective, the optimizer learns to maximize U_{od}, the sum of satisfied mobility flows captured by the created line. However, this efficiency-based optimization objective does not address how the benefits of the newly designed line are distributed between different groups. By groups here we refer to socio-economic divisions such as income, education, or development index. Optimizing for efficiency can lead to large disparities between low and high-developed areas [15].

To address this we formulate a multi-objective problem, in which we introduce the *group-based satisfied mobility flow*. We define a set A, which represents d different groups based on socio-economic indicators, such as income, development index, or education. Each cell $h \in H^{n\times m}$ of the environment is then associated with a group $a \in A$.

Further, we define the group-based satisfied mobility flow objective $U^a_{od},\ a \in A$. The optimization formulation in Equation 2 is not sufficient under these conditions, as we are not looking to maximize a single objective, but rather to provide the trade-off and allow for different decisions. We, therefore, re-formalize it as a multi-objective optimization problem in Equation 3.

$$\begin{aligned} \max \quad & \left\{U_{od}^{a_1}(G(N,E)), \cdots, U_{od}^{a_d}(G(N,E))\right\} \\ \text{s.t.} \quad & cost(G) \leq B \\ & |N| \leq T \end{aligned} \tag{3}$$

If a (possibly non-linear) utility function for the decision-maker is known, the problem can again be solved using the aforementioned techniques. If, however, the preferences are unknown, the multi-objective formulation will in general not lead to a single optimum. This is because what may be optimal for one group can be bad for a different group. As such, the proposed formulation leads to a set of non-dominated solutions from which a potential decision-maker may choose. To aid a decision-maker in this situation, previous work has studied support systems for multi-objective sequential decision-making problems [32].

When considering multiple conflicting objectives, it is not immediately obvious which solution to choose.

3 MOMDP FORMULATION

To study the Multi-Objective TNDP using Reinforcement Learning, we first adopt the formulation of Wei et al., which transforms it into a sequential decision-making process. Following this framework, there is a single agent that generates a solution (i.e. a transport line) by taking sequential actions, receiving a reward and adapting its policy based on it. In particular, at every time step t, the agent selects a cell $h \in H^{n \times m}$ to place a station on. At the end of the episode, the sequence of selected cells is the generated transport graph $G(N, E)$, where each cell is a node, and there exists an edge between every two sequential nodes. Note that actions are further constrained by feasibility rules that dictate the direction, by preventing the agent from selecting a backward cell or forming unusual metro line shapes (e.g. meandering) [30]. For simplicity, we omit the feasibility rules from the following discussion.

Formulated as a Multi-Objective Markov Decision Process (MOMDP) $\langle \mathcal{S}, \mathcal{A}, \mathcal{P}, \vec{\mathcal{R}} \rangle$, the Transport Network Design Problem is characterized as follows:

- $\mathcal{S}$: the state; sequence of selected grid cells.
- $\mathcal{A}$: the action; selected cell at each time step.
- $\mathcal{P} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow [0, 1]$: the state transition function; in this problem it is deterministic.
- $\vec{\mathcal{R}} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow \mathbb{R}^d$: the vectorial reward the agent receives for taking actions $\mathcal{A}$; $d = |A|$, where d corresponds to the number of groups for which a group-based satisfied mobility flow objective (Equation 3) has been defined.

At each time step, actions are taken according to a policy $\pi \in \Pi$, where $\pi : S \times A \rightarrow [0, 1]$. In the single-objective case, the optimization goal is to find the optimal policy π^*, that maximizes the expected cumulative reward, which is represented by the value function V^π:

$$V^\pi = \mathbb{E}\left[\sum_{t=0}^{H} \gamma^t r_{t+1} | \pi, s_0\right] \tag{4}$$

Where s_0 is the initial state and H is the (finite) horizon, which is constrained by B and T. For TNDP, the discount factor γ is considered constant and will be omitted in future references. The policy-optimization objective can thus be defined as:

$$\pi^* = \arg\max_{\pi} V^\pi \tag{5}$$

This is helpful in that it creates a complete ordering of policies, making comparison and selection a straightforward process [10].

In contrast, on multi-objective cases, the value function is a vector itself, with as many dimensions as there are objectives. This leads to a vector-valued value function:

$$\vec{V}^\pi = \mathbb{E}\left[\sum_{t=0}^{H} \gamma^t \vec{r}_{t+1} | \pi, s_0\right] \tag{6}$$

Comparing policies is therefore not straightforward, and different decision-makers may prefer different policies. As such, we should compute a *coverage set*, i.e., a set of alternative policies containing an optimal policy for any preference function that a decision-maker might have [10, 23]. A common choice for such a coverage set is a Pareto front (or possibly a smaller Pareto coverage set), which is optimal if we do not know anything about the preference function other than that it is monotonically increasing in all objectives, and policies should be deterministic. A different possible choice that could be natural in a TNDP setting is a Lorenz optimal set [20], which is a subset of the Pareto front that incorporates the notion that if we can 'transfer value' from an objective – which corresponds to the utility of a group in a TNDP – that has a higher value to an objective with a lower value, decreasing the sum over all objectives, this ought to be preferred. In this sense, it implements a rather minimal notion of fairness.

In Figure 1 (Panel D), we show three Pareto-efficient solutions, each with different values for each group. Depending on the decision-makers, any of these three might be preferred. Therefore, all three could be selected to be implemented, and as such should be presented to the decision-maker. The metro line corresponding to the red square in Figure 1 cannot be optimal, and should therefore be removed before presenting the possible solutions to the decision-maker.

4 ENVIRONMENTS

We provide two real-world case study environments, based on the cities of Amsterdam (Netherlands) and Xi'an (China). Figure 2 presents these environments.

Amsterdam environment. We generate and release the Amsterdam environment. The city is split in a $H^{35 \times 47}$ grid, with equally sized cells of $0.5km^2$. Since GPS data are not available for Amsterdam, we estimate the origin-destination travel demand using the recently published universal law of human mobility, which states that the total mobility flow between two areas i, j depends on their distance and the visitation frequency [24]. It is calculated as follows:

$$OD_{ij} = \mu_j K_i / d_{ij}^2 \ln(f_{max}/f_{min}) \tag{7}$$

where K_i is the total area of the origin location i, d_{ij}^2 the (manhattan) distance between i, j and μ_j is the magnitude of flows, calculated as follows:

$$\mu_j \approx \rho_{pop}(j) d_j^2 f_{max} \tag{8}$$

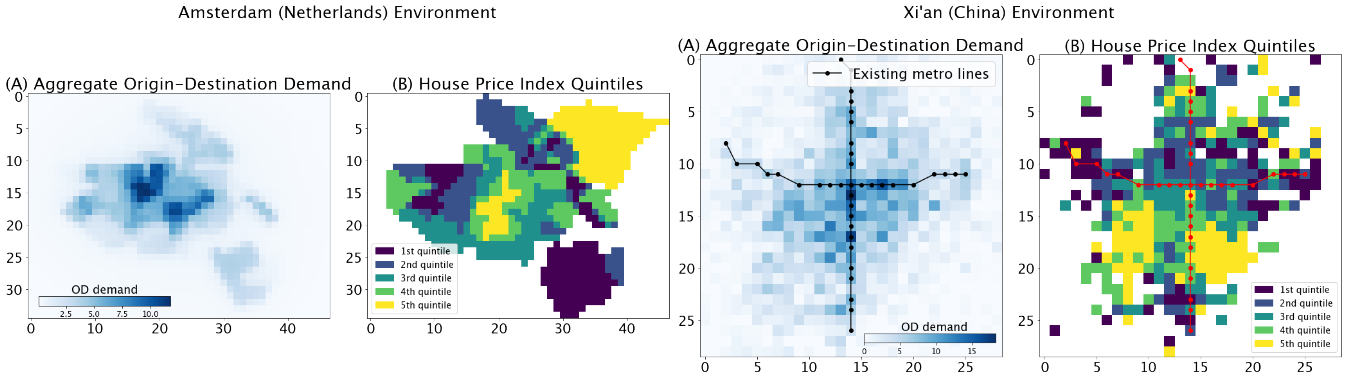

Figure 2: Two real-world environments for the MO-TNDP. On the left, the city of Amsterdam is split into square cells of $0.5km^2$ each, creating a 35x47 grid. Each cell is associated with an aggregate origin-destination demand (blue colormap, panel A) and a house price quintile (panel B). On the right side, the same plots are shown for the city of Xi'an, which is split into a 29x29 grid (cell size $1km^2$). Note that in this environment we also represent the two metro lines already existing in the city. MO-TNDP can be solved in both empty and environments with pre-existing lines.

We estimate the flows for a full week, by setting f_{min}, f_{max} to 1/7 and 7 respectively. Since in our case the grid cells are of equal size, K can be omitted from the calculation.

Every cell is associated with an average house price, which we take from the publicly available statistical bureau of The Netherlands (CBS)[1] dataset. Groups are the five quintiles of the price. We ran experiments on an empty environment (no previously existing lines).

Xi'an environment. The Xi'an environment was generated and released to the public by Wei et al.[2] The city is split in a $H^{29\times 29}$ grid, with equally sized cells of $1km^2$ (this is because Xi'an is a much bigger city). An origin-destination demand matrix was created using GPS data from 25 million mobile phones, whose movements were tracked during a period of one month. Each cell is also associated with an average house price index, which we use to split the grid into five equally sized quintiles. This represents a setting where detailed data on mobility demand is available (contrasting with the Amsterdam case study before). We ran experiments on an environment with two previous existing metro lines.

We chose the average house price as a proxy for the development of a neighborhood, as it is universal and available for multiple cities without raising privacy concerns.

5 CHALLENGES IN MULTI-OBJECTIVE TNDP

Designing public transportation with fairness considerations poses multiple challenges, both on technical and decision-making aspects. In this section, we outline the most important ones.

5.1 Unknown Decision-Maker Preference

The biggest advantage of using Reinforcement Learning to tackle optimization tasks like the TNDP is its generalization properties [3, 15]. A single agent can learn on one environment and be used to generate lines on a different environment.

However, given the nature of public transportation design, knowing the optimal preference of the decision-maker beforehand is impossible. Different cities have different objectives and acceptable trade-offs, while there are multiple notions of fairness in TNDP [2, 15]. For example, one might be interested in building a line that maximizes the benefits of the most-disadvantaged group, according to Rawls' theory of justice [2]. Their utility function is defined as follows [15]:

$$\max U_{od}^{a_{min}}(G(N,E)) \tag{9}$$

Where a_{min} is the most disadvantaged group. For example, in cases where some transportation lines exist already, a_{min} can be defined as $a_{min} = \underset{a\in A}{\arg\min}\, U_{od}$, meaning the group with the lowest satisfied OD flows before the line was created.

In another example, a decision-maker might decide it is best to follow an equal-sharing notion, where all groups should receive an equal share of the added benefits [2]. This can be expressed as minimizing the disparities in satisfied ODs between all groups [15]:

$$\min \sum_i \sum_j |U_{od}^{a_i}(G(N,E)) - U_{od}^{a_j}(G(N,E))|, \;\; a_i, a_j \in A, i \neq j \tag{10}$$

Evidently, while the above notions can arguably both be characterized as fair, they lead to completely different solutions. This disparity is illustrated in Figure 1 (D); the pink triangle follows an equal sharing utility while the blue arrow a Rawlsian utility.

5.2 Non-Linear Decision-Maker Preference

In the previous section, we discussed the challenges that arise from the lack of prior knowledge on the utility function of the decision-maker. We offered some examples, based on well-known notions of social good. Should the acceptable notion be known beforehand, then the multi-objective problem is reduced to a single objective, with a scalarized reward, using functions similar to those we presented. However, in many cases, utility functions are non-linear, and better approximation methods need to be applied [21]. This is especially the case when considering group fairness, where

[1] https://www.cbs.nl/nl-nl/maatwerk/2019/31/kerncijfers-wijken-en-buurten-2019
[2] https://github.com/weiyu123112/City-Metro-Network-Expansion-with-RL

a high control over the balance of the utilities of different groups is desired.

A common case of a non-linear utility function is the Generalized Gini Index (GGI), which is a measure used to take decisions that maximize a weighted sum of utilities, by giving more weight to those with lower utilities. The GGI is very versatile and generalizable; it can be used to achieve a wide range of fairness notions [15, 25] and is defined in Equation 11.

$$GGI_w = \sum_{\forall a \in A} w_d u^a_\sigma \tag{11}$$

Where σ is a permutation that sorts the utilities $u^a, a \in A$ in descending order and weights w_d are non-increasing weights, i.e., $w_1 > w_2 > ... > w_d$. In the context of transportation network design, the GGI can be used to generate lines that achieve equity or equality in satisfied origin-destination flows between all defined societal groups.

The GGI has been used to balance rewards between groups in the TNDP, but applying single-objective optimization [15]. To achieve this, the reward is scalarized using the index at every time step. However, this approach does not optimize the non-linear GGI. In contrast, to achieve the non-linear GGI optimization, traditional RL algorithms like Q-Learning and Policy Gradient need to be modified. In Section 6.1 we propose methods to tackle this challenge using multi-objective policies.

5.3 Large Action Space

Q-Learning has been effectively used to tackle a wide range of reinforcement learning problems, either on single or multiple objectives [29]. Formalized as a sequential decision-making process, the TNDP problem is reduced to selecting a grid cell to place the next station on, at every time step. This leads to a very large action space $\mathcal{A}$. In the real-world examples we propose in this paper (Figure 2), the cities of Amsterdam and Xi'an are divided into grids of size 35×47 and 29×29 respectively. Station distance and shape-based constraints are being applied to accommodate feasible transport line solutions, but in order to ensure model generalizability, they are preferably applied on the output of the policy [15, 30]. This poses a major computation challenge for applying multi-objective Q-learning to the TNDP and can make it intractable in large environments.

6 TACKLING THE MULTI-OBJECTIVE TNDP

In this section, we discuss methodologies for tackling the Multi-Objective Transport Network Design Problem. Specifically, we outline single-policy (6.1) and multi-policy learning (6.2). Note that this is not an exhaustive set of methodologies.

6.1 Single-Policy Methods

When the preferences of the decision-maker are known exactly, it may be possible to employ methods from single-objective RL. Specifically, when the utility function of the decision-maker is linear, the MOMDP can be reduced to an equivalent MDP [23]. When utility functions are non-linear, however, this is generally not possible [22]. Fortunately, the TNDP can be represented as a purely deterministic MOMDP, which can still be reduced to a single-objective MODMP by augmenting the state space [27].

A concrete non-linear preference function that a decision-maker may aim to optimize in the TNDP is the generalized Gini index (Equation 11). Assuming a vectorized $\vec{Q}_\theta \in \mathbb{R}^{|A|}$ function parameterized by θ, the objective is defined as follows:

$$\vec{Q}_\theta = r + \gamma \vec{Q}_{\theta'}(s', a^*) \tag{12}$$

Where $a^* = \arg\max_{a' \in \mathcal{A}} GGI_w(r + \gamma \vec{Q}_{\theta'}(s', a'))$. It is important to note here that this formulation violates the additive returns assumption of the Bellman equation for non-linear $\vec{Q}$ [10]. Therefore, it is necessary to extend the state, concatenating the previously received rewards. Nevertheless, this approach is intractable in problems with large action spaces, like the TNDP.

Alternatively, one can use a policy-search approach, where a mapping π_θ from states to actions is learned, without relying on Bellman returns. Recently, a novel method was proposed to tackle fairness in MO problems using policy gradient [25]. It is defined as follows:

$$\nabla_\phi GGI_w(J(\pi_\phi) = \nabla) = w_\sigma^\intercal \nabla_\phi J(\pi_\phi) \tag{13}$$

Where $J(\pi_\phi) \in \mathcal{R}^{|A|}$ and σ is the sorting permutation over J. J is used here to associate the objective with the original policy-gradient definition.

Siddique et al. leverage Equation 13 to develop a GGI-based Advantage Actor Critic (A2C) and Proximal Policy Optimization (PPO). They apply their methods in two optimization domains and achieve a fair division of outcomes between different objectives. We propose to use this method to tackle the single-policy MO-TNDP, because of its efficiency on large action spaces.

6.2 Multi-Policy Methods

A complicating factor in the proposed multi-objective formulation is that the environments may have a large state and action space as well as a relatively high number of objectives, depending on the decision-maker's criteria. This makes popular tabular methods such as Pareto Q-learning [29] or scalarized multi-objective Q-learning [28] intractable. Recently, however, a novel deep RL method called Pareto Conditioned Networks (PCN) was proposed for multi-objective RL in deterministic environments [21]. We believe that this algorithm can be leveraged to solve the multi-objective TNDP.

PCN is a multi-policy algorithm for sequential decision-making that concurrently learns which policies lead to non-dominated solutions and how to execute these policies again when prompted. In addition, it leverages quality metrics that ensure the learned Pareto front represents a wide range of diverse solutions. As such, the resulting set of policies may be used to effectively inform decision-making.

PCNs attempt to stabilize learning using a supervised method to parameterize the policy, instead of traditional TD-learning. The training dataset is dynamic and generated via a collection of trajectories $\left\langle s, \hat{h}, \hat{\mathcal{R}} \right\rangle$ it encounters while learning; s represents the state and $\hat{\mathcal{R}}$ the reward obtained in horizon $\hat{h}$. The action $a_i \in \mathcal{A}$ taken at time step i is used as the label. A datapoint $\langle s_t, h_t, r_t \rangle$ is created for each time step in the (finite) trajectory. The policy π is updated using a cross-entropy loss function:

$$H = -\sum_{a \in \mathcal{A}} y_a log\pi(a|s_t, h_t, r_t) \tag{14}$$

PCN learns to generate solutions on the Pareto front by applying a pruning function to keep only tuples that lead to non-dominated returns on the different objectives. This process ensures that the boundaries of the coverage set are constantly being extended. PCNs have been proven to perform better than baselines even in settings with a large number of objectives, which is important for the MO-TNDP [21].

7 CONCLUSION

In this work-in-progress paper, we proposed a Multi-objective Reinforcement Learning formulation for studying fairness in the Transport Network Design Problem. We provided environments and outlined the most important challenges in applying the formulation. Finally, we proposed state-of-the-art methodologies for single and multiple policies to tackle the problem. In the future, we plan to run experiments based on these algorithms.

Acknowledgements

WR is supported by the Research Foundation – Flanders (FWO), grant number 1197622N. This research was supported by funding from the Flemish Government under the "Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen" program.

REFERENCES

[1] Renato Oliveira Arbex and Claudio Barbieri da Cunha. 2015. Efficient transit network design and frequencies setting multi-objective optimization by alternating objective genetic algorithm. *Transportation Research Part B: Methodological* 81 (Nov. 2015), 355–376. https://doi.org/10.1016/j.trb.2015.06.014

[2] Hamid Behbahani, Sobhan Nazari, Masood Jafari Kang, and Todd Litman. 2019. A conceptual framework to formulate transportation network design problem considering social equity criteria. *Transportation Research Part A: Policy and Practice* 125 (July 2019), 171–183. https://doi.org/10.1016/j.tra.2018.04.005

[3] Irwan Bello, Hieu Pham, Quoc V. Le, Mohammad Norouzi, and Samy Bengio. 2017. Neural Combinatorial Optimization with Reinforcement Learning. *arXiv:1611.09940 [cs, stat]* (Jan. 2017). http://arxiv.org/abs/1611.09940 arXiv: 1611.09940.

[4] Lang Fan, Christine L. Mumford, and Dafydd Evans. 2009. A simple multi-objective optimization algorithm for the urban transit routing problem. In *2009 IEEE Congress on Evolutionary Computation*. 1–7. https://doi.org/10.1109/CEC.2009.4982923 ISSN: 1941-0026.

[5] Wei Fan and Randy B. Machemehl. 2006. Using a Simulated Annealing Algorithm to Solve the Transit Route Network Design Problem. *Journal of Transportation Engineering* 132, 2 (Feb. 2006), 122–132. https://doi.org/10.1061/(ASCE)0733-947X(2006)132:2(122) Publisher: American Society of Civil Engineers.

[6] Reza Zanjirani Farahani, Elnaz Miandoabchi, Wai Yuen Szeto, and Hannaneh Rashidi. 2013. A review of urban transportation network design problems. *European journal of operational research* 229, 2 (2013), 281–302.

[7] Reza Zanjirani Farahani, Elnaz Miandoabchi, W. Y. Szeto, and Hannaneh Rashidi. 2013. A review of urban transportation network design problems. *European Journal of Operational Research* 229, 2 (Sept. 2013), 281–302. https://doi.org/10.1016/j.ejor.2013.01.001

[8] Valérie Guihaire and Jin-Kao Hao. 2008. Transit network design and scheduling: A global review. *Transportation Research Part A: Policy and Practice* 42, 10 (2008), 1251–1273.

[9] Gabriel Gutiérrez-Jarpa, Gilbert Laporte, and Vladimir Marianov. 2018. Corridor-based metro network design with travel flow capture. *Computers & Operations Research* 89 (Jan. 2018), 58–67. https://doi.org/10.1016/j.cor.2017.08.007

[10] Conor F. Hayes, Roxana Rădulescu, Eugenio Bargiacchi, Johan Källström, Matthew Macfarlane, Mathieu Reymond, Timothy Verstraeten, Luisa M. Zintgraf, Richard Dazeley, Fredrik Heintz, Enda Howley, Athirai A. Irissappane, Patrick Mannion, Ann Nowé, Gabriel Ramos, Marcello Restelli, Peter Vamplew, and Diederik M. Roijers. 2022. A practical guide to multi-objective reinforcement learning and planning. *Autonomous Agents and Multi-Agent Systems* 36, 1 (April 2022), 26. https://doi.org/10.1007/s10458-022-09552-y

[11] Wouter Kool, Herke van Hoof, and Max Welling. 2018. Attention, Learn to Solve Routing Problems!. In *International Conference on Learning Representations*.

[12] Gilbert Laporte and Marta M. B. Pascoal. 2015. Path based algorithms for metro network design. *Computers & Operations Research* 62 (Oct. 2015), 78–94. https://doi.org/10.1016/j.cor.2015.04.007

[13] Karel Martens. 2016. *Transport Justice: Designing fair transportation systems.* Routledge. Google-Books-ID: m0yTDAAAQBAJ.

[14] Antonio Mauttone and María E. Urquhart. 2009. A multi-objective metaheuristic approach for the Transit Network Design Problem. *Public Transport* 1, 4 (Nov. 2009), 253–273. https://doi.org/10.1007/s12469-010-0016-7

[15] Dimitris Michailidis, Sennay Ghebreab, and Fernando P. Santos. 2023. Balancing Fairness and Efficiency in Transport Network Design through Reinforcement Learning. In *Proceedings of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS '23)*. IFAAMAS, London, United Kingdom.

[16] Grigory Neustroev, Sytze P. E. Andringa, Remco A. Verzijlbergh, and Mathijs M. De Weerdt. 2022. Deep Reinforcement Learning for Active Wake Control. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS '22)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 944–953.

[17] Leonardo Nicoletti, Mikhail Sirenko, and Trivik Verma. 2022. Disadvantaged Communities Have Lower Access to Urban Infrastructure. *arXiv preprint arXiv:2203.13784* (2022).

[18] Jan Nijman and Yehua Dennis Wei. 2020. Urban inequalities in the 21st century economy. *Applied Geography* 117 (2020), 102188.

[19] Mahmoud Owais and Mostafa K. Osman. 2018. Complete hierarchical multi-objective genetic algorithm for transit network design problem. *Expert Systems with Applications* 114 (Dec. 2018), 143–154. https://doi.org/10.1016/j.eswa.2018.07.033

[20] Patrice Perny, Paul Weng, Judy Goldsmith, and Josiah Hanna. 2013. Approximation of Lorenz-Optimal Solutions in Multiobjective Markov Decision Processes. In *Proceedings of the 27th AAAI Conference on Artificial Intelligence*. 92–94.

[21] Mathieu Reymond, Eugenio Bargiacchi, and Ann Nowé. 2022. Pareto Conditioned Networks. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems* (Virtual Event, New Zealand) *(AAMAS '22)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1110–1118.

[22] Mathieu Reymond, Conor Hayes, Diederik M Roijers, Denis Steckelmacher, and Ann Nowé. 2021. Actor-Critic Multi-Objective Reinforcement Learning for Non-Linear Utility Functions. In *Multi-objective decision making workshop (MODeM 2021)*. 9.

[23] Diederik M Roijers, Peter Vamplew, Shimon Whiteson, and Richard Dazeley. 2013. A survey of multi-objective sequential decision-making. *Journal of Artificial Intelligence Research* 48 (2013), 67–113.

[24] Markus Schläpfer, Lei Dong, Kevin O'Keeffe, Paolo Santi, Michael Szell, Hadrien Salat, Samuel Anklesaria, Mohammad Vazifeh, Carlo Ratti, and Geoffrey B. West. 2021. The universal visitation law of human mobility. *Nature* 593, 7860 (May 2021), 522–527. https://doi.org/10.1038/s41586-021-03480-9 Number: 7860 Publisher: Nature Publishing Group.

[25] Umer Siddique, Paul Weng, and Matthieu Zimmer. 2020. Learning Fair Policies in Multiobjective (Deep) Reinforcement Learning with Average and Discounted Rewards. *arXiv:2008.07773 [cs]* (Aug. 2020). http://arxiv.org/abs/2008.07773 arXiv: 2008.07773.

[26] W. Y. Szeto and Y. Jiang. 2014. Transit route and frequency design: Bi-level modeling and hybrid artificial bee colony algorithm approach. *Transportation Research Part B: Methodological* 67 (Sept. 2014), 235–263. https://doi.org/10.1016/j.trb.2014.05.008

[27] Peter Vamplew, Cameron Foale, and Richard Dazeley. 2022. The Impact of Environmental Stochasticity on Value-Based Multiobjective Reinforcement Learning. *Neural Computing and Applications* 34, 3 (Feb. 2022), 1783–1799. https://doi.org/10.1007/s00521-021-05859-1

[28] Kristof Van Moffaert, Madalina M. Drugan, and Ann Nowé. 2013. Scalarized Multi-Objective Reinforcement Learning: Novel Design Techniques. In *2013 IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning (ADPRL)*. 191–199. https://doi.org/10.1109/ADPRL.2013.6615007

[29] Kristof Van Moffaert and Ann Nowé. 2014. Multi-Objective Reinforcement Learning Using Sets of Pareto Dominating Policies. *Journal of Machine Learning Research* 15, 107 (2014), 3663–3692.

[30] Yu Wei, Minjia Mao, Xi Zhao, Jianhua Zou, and Ping An. 2020. City Metro Network Expansion with Reinforcement Learning. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, Virtual Event CA USA, 2646–2656. https://doi.org/10.1145/3394486.3403315

[31] Ziyi Xu, Xue Cheng, and Yangbo He. 2022. Performance of Deep Reinforcement Learning for High Frequency Market Making on Actual Tick Data. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS '22)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1765–1767.

[32] Luisa M Zintgraf, Diederik M. Roijers, Sjoerd Linders, Catholijn M Jonker, and Ann Nowé. 2018. Ordered Preference Elicitation Strategies for Supporting Multi-Objective Decision Making. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, Stockholm, Sweden, 1477–1485.