Optimizing Fairness in Transport Network Design using Deep Reinforcement Learning

Michailidis, D.; Ghebreab, S.; Santos, F.P.

Publication date
2022

Document Version
Final published version

Citation for published version (APA):
Michailidis, D., Ghebreab, S., & Santos, F. P. (2022). *Optimizing Fairness in Transport Network Design using Deep Reinforcement Learning*. Paper presented at IJCAI 2022 DSO Workshop, Vienna, Austria. https://drive.google.com/file/d/1XQlhy8NLYD9WtP45s8UYVgidTBVfQ77h/view

Optimizing Fairness in Transport Network Design using Deep Reinforcement Learning

Dimitris Michailidis, Sennay Ghebreab, Fernando P. Santos

Civic AI Lab and Socially Intelligent Artificial Systems
Informatics Institute, University of Amsterdam
{d.michailidis, s.ghebreab, f.p.santos}@uva.nl

Abstract

Transportation systems have a fundamental impact on human well-being, economic productivity, and environmental sustainability. Designing well-functioning transport networks is however non-trivial, given the large space of solutions and constraints - ranging from physical to social and legal. Moreover, different sources of inequality in cities complicates the task even more. It is crucial to address the disproportional benefits that transportation network design can lead to. In this work-in-progress paper, we explore the application of Deep Reinforcement Learning to the Transport Network Design Problem (TNDP) and the goal of optimizing utility (satisfied total demand) and fairness (equity in distribution among groups). We test reward functions based on *Utilitarianism*, *Equal Sharing of Benefits*, *Narrowing the Gap*, and *Rawl's theory of justice*, and our initial results show that different fairness criteria can achieve different compromises between fairness and utility.

1 Introduction

Deep Reinforcement Learning (Deep RL) is increasingly being used to tackle NP-hard combinatorial optimization tasks [Kool *et al.*, 2018], such as Travelling Salesman, [Bello *et al.*, 2017] and Vehicle Routing [Nazari *et al.*, 2018]. Deep RL outperforms integer programming and heuristic methods, due to its power in taking optimal long-term sequential decisions, especially in environments with large search spaces.

One advantage of Deep RL is that it does not need an exhaustive list of constraints to find good solutions. With simple reward mechanisms, the agent is able to learn the dynamics of the environment via trial-and-error, and even transfer learned policies to alternative and broader scenarios [Bello *et al.*, 2017]. This property makes it suitable to tackle the Transport Network Design Problem (TNDP). In TNDP, the goal is to design a new transport line, by connecting areas within the city, so as to maximize the total satisfied mobility demand [Farahani *et al.*, 2013].

Traditionally, TNDP has been tackled using integer programming [Gutiérrez-Jarpa *et al.*, 2018], simulated annealing [Fan and Machemehl, 2006], bee colony, [Szeto and Jiang, 2014] or genetic algorithms [Owais and Osman, 2018]. Despite achieving good results, the previous methods require a lot of computational power and city-specific constraints that prevent developing general solutions. Recently, [Wei *et al.*, 2020] proposed a Deep RL model that outperforms previous approaches in satisfied demand, without the need of specifying multiple constraints. Moreover, their model can be used to learn different environments without many changes. Deep RL methods are also advantageous as they allow to specify reward functions that can implement different objectives relevant to stakeholders and communities. One can explicitly aim to design transit networks such that utility and fairness are both considered. This is particularly relevant given the existing patterns of inequality in cities [Nijman and Wei, 2020; Nicoletti *et al.*, 2022], and the role of transportation in alleviating them. Indeed, in order to achieve the United Nations Sustainable Development Goal of providing accessible, affordable, and sustainable public transport to those in need[1], it is crucial to move beyond optimizing total utility and consider fairness.

In this work-in-progress paper, we evaluate how different reward functions can be employed to achieve fairness objectives in the TNDP. Fairness in this context refers to an equitable distribution of benefits among groups within a city. Specifically, we explore the compromises between the overall satisfied demand (utility) and the satisfied demand of individual groups, based on their area's house price index (fairness/equity). We show that recent approaches struggle to effectively address this trade-off, even when they account for the development index of the connected areas. The proposed methods achieve different fairness objectives without the need to change the original architecture.

2 The Transport Network Design Problem

We consider the TNDP where the goal is to generate an undirected graph $G(N, E)$, which represents a transport line, where N are the locations to place stations on, and E are the connection edges between them. 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

[1]https://www.undp.org/sustainable-development-goals#sustainable-cities-and-communities

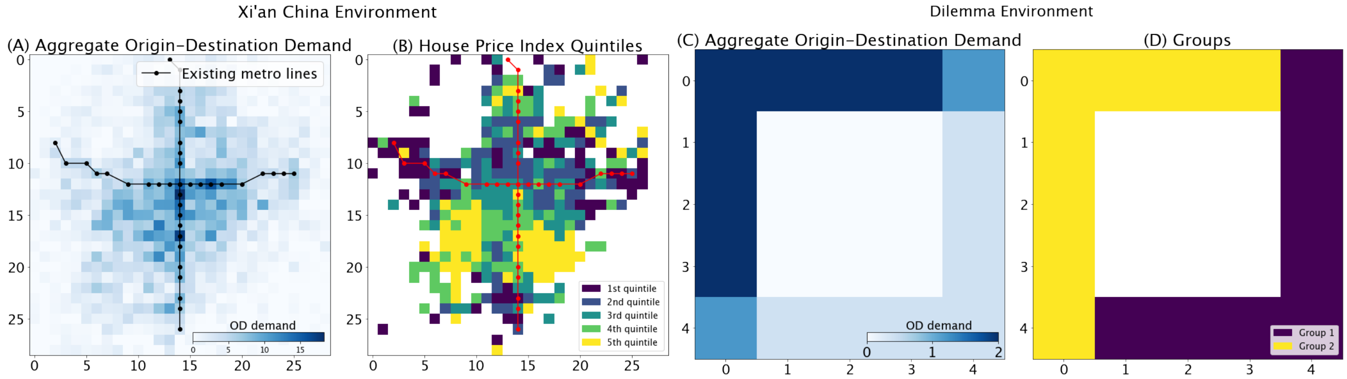

Figure 1: We present two environments where the Transport Network Design Problem (TNDP) can be solved. On the left, we represent a real-life example of Xi'an, China, where the city is split into a 29x29 grid. Each square in the grid is associated with an aggregate origin-destination demand (blue colormap, panel A) and a house price index quintile (panel B). On the right side, a synthetic toy environment where we present an extreme example where there is a clear dilemma between utility and fairness maximization. Specifically, Group 2 (yellow) has a disproportionately higher origin-destination demand than Group 1, as represented by the increased color intensity in panel C.

the estimated Origin-Destination (OD) matrix [Guihaire and Hao, 2008; Farahani *et al.*, 2013]. Finally, the area selection is 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 [Wei *et al.*, 2020].

Given the above, the problem is formalized as follows: Find $G(N, E)$, such that:

$$\begin{aligned} \max \quad & U_{od}(G(N,E)) \\ \text{s.t.} \quad & cost(G) < B \\ & |N| < T \end{aligned} \tag{1}$$

Where:

- $N \subseteq H$
- $E = \{(i,j) : h_i, h_j \in N\}$

The traditional TNDP definition does not address the distribution of benefits. We thus need to adjust the objective function to incorporate different notions of fairness.

Fairness in Transport Network Design

There are multiple theories of social equity and fairness. Here, we adopt the notions outlined by Behbahani *et al.* for transportation network design. We thus define a set of groups A, which can be based on income, race, gender, etc. Each square $h \in H^{n \times m}$ of the environment is associated with a group $a \in A$. We adjust the objective function for each notion as follows (for simplicity, we use U to represent the utility function and G to represent the generated line graph):

- **Utilitarianism**: maximize total benefits (Equation 1 - the classic TNDP with no fairness considerations).

- **Rawls**: maximize benefits of most disadvantaged group.

$$\max \min_a U(G)$$

- **Equal Sharing**: equalize benefits among groups.

$$\min \sum_i \sum_j |U_i(G) - U_j(G)|$$

- **Narrowing the gap**: maximize benefits while narrowing the gap between any two groups.

$$\begin{aligned} & \max U(G) \\ \text{s.t. } & |U_i(G) - U_j(G)| < T \end{aligned}$$

In Figure 1 (A,B), we show a real-life environment for the TNDP, based on [Wei *et al.*, 2020]. The city of Xi'an in China is split into a grid $H^{29 \times 29}$. An origin-destination matrix is estimated from the movement trajectories of 25 million mobile phone users. The grid squares are split into five groups, each representing a quintile of the house price index. There are two existing public transport lines that cover some of the demand, and the goal is to create a third line to cover currently unserviced demand.

To better demonstrate the trade-off between utility and fairness maximization in TNDP, we have created a small synthetic $H^{5 \times 5}$ environment (Figure 1, C,D), where we present an extreme case of the utility-fairness dilemma. The grid is split into two groups, separated geographically, with Group 1 having a disproportionately higher origin-destination travel demand than Group 2 (depicted by the higher blue color intensity). Figure 2 shows three possible lines, alongside their performance on utility and equity. Optimizing for maximum utility leads to a much higher total covered demand, but also to the biggest disparity between the two groups. On the other hand, optimizing for maximum equity leads to almost no disparity in covered demand, but total performance drops significantly.

3 Model

We use Deep Reinforcement Learning (Deep RL) to explore the trade-off described above and modify the reward function of the agent to achieve outcomes with different degrees of fairness.

3.1 Deep RL For Transport Network Design

The base for the agent is the architecture proposed by [Wei *et al.*, 2020]. The problem is a sequential decision-making prob-

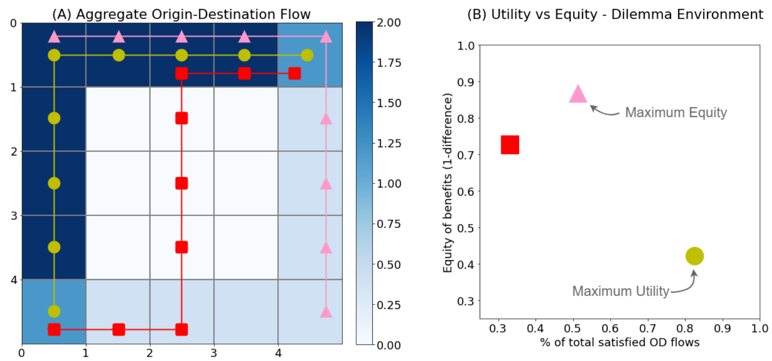

Figure 2: We demonstrate how different optimization choices in the dilemma environment can lead to different results in utility and fairness. The yellow line (circle) leads to the highest satisfied demand, while the pink one (triangle) leads to the fairest distribution of the satisfied demand between the two groups, but performs worse in overall utility.

lem, where in each step a grid square is selected to place a station on. Formulated as a Markov Decision Process (MDP) $\langle \mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R} \rangle$, it is characterized as follows:

- $s_t \in \mathcal{S}$: the sequence of selected grid squares up to timestep $t-1$.
- $a_t \in \mathcal{A}$: the selected station at timestep t. Determined by a parametrized policy π_θ
- $\mathcal{P}$: a deterministic state transition function.
- $r(s_t, a_t) \in \mathcal{R}$: the reward the agent receives for taking action a_t

At each timestep, every possible next action is assigned a probability, using a parametrized policy network $\pi(a_t|s_t, \theta) = Pr\{a|s_t, \theta_t\}$. The agent is trained via a policy gradient advantage actor-critic framework (A2C). In practice, the actor is a recurrent neural network that takes as input the current state (selected squares at t) and computes a probability distribution over all actions (square to be selected in $t+1$) [Bello *et al.*, 2017]. It employs a convolutional neural network as the encoder, a Long-Short Term Memory (LSTM) network as the decoder, and an attention layer as the final layer before using softmax to calculate the probability distribution. The critic is a deep neural network that takes as input the initial state and estimates the future value function. The action space is constrained by a set of feasibility rules that dictate the direction, by preventing the agent from selecting a backward square or forming unusual transport line shapes. The feasibility rules are updated at each time step.

3.2 Fair Reward Functions

In the original implementation by Wei *et al.*, the authors use a reward function that is a weighted sum of the satisfied travel demand and the house price index of the connected areas. By tuning the weights, they attempt to balance equity and utility in their design. Later on we show that even when the house price index weight is maximized, this reward leads to inequality in the satisfied travel demand between groups. We use their reward as a baseline and propose new ones for achieving the fairness goals outlined in Section 2.

Baseline

The intuition behind this design is that equity is achieved when low and high developed areas are connected with each other via the new line. Setting different weights helps to prioritize between utility and equity considerations.

$$r = w_1 U(G) + w_2 D(G) \tag{2}$$

where $D(G)$ is a distance discounted function of the house price index of the connected squares, $w_1, w_2 \in [0,1]$, and $w_1 + w_2 = 1$.

Variance Regularization

This reward function aims to maximize total benefits while narrowing the gap between groups. It achieves this by regularizing the variance of the groups' satisfied travel demand with a hyperparameter λ. It has been successfully used to equalize treatment among neighborhoods in the ride-pooling assignment problem [Raman *et al.*, 2021].

$$r = \sum_a U_a(G) - \lambda \text{var}(U(G)) \tag{3}$$

Lowest Quintile

This reward function is based on Rawl's theory of justice and aims at maximizing the total benefits received by the least privileged group [Behbahani *et al.*, 2019].

$$r = U_{a_0}(G) \tag{4}$$

where a_0 is the group with the least utility before the new line is generated.

Generalized Gini Index (GGI)

This reward function is based on the Generalized Gini Index (GGI) coefficient. It is a weighted sum of the groups' utility, where the lowest weight is assigned to the group with the highest utility. The weights are assigned via a hyperparameter and depending on their values, it can achieve both Rawl's and equal sharing fairness. It has been used successfully in equalizing the waiting times of the traffic light control problem [Siddique *et al.*, 2020].

$$r = GGI_w(U(G)) = \sum_a w_a v_a^{\uparrow} \tag{5}$$

where $w_a \in [0,1]$ is a weight vector such as $w_1 > w_2 > \ldots. > w_{|A|}$.

Table 1: Results of different Reward Functions on Utility and Fairness objectives.

Reward Function	**Fairness Notion**	**Total OD %**	**Gini Index**	**Lowest Quintile OD %**
Maximize OD (baseline, $w_1 = 1$)	Utilitarianism	**6.01**	0.21	2.90
Maximize Equity (baseline, $w_2 = 1$)	Utilitarianism	4.32	0.22	2.41
Variance Regularization ($\lambda = 5$)	Narrowing the Gap	4.58	0.08	4.16
Lowest Quintile	Rawls	2.47	0.39	**6.77**
Generalized Gini Index (GGI) ($w_a = \frac{1}{2^a}$)	Equal Sharing	4.34	**0.05**	4.24

Bold face indicates best result in the corresponding metric.

4 Preliminary Results

We applied the fair reward functions in the Xi'an China environment, split in a 29×29 grid. For the model, we used the optimized architecture and hyperparameters proposed by [Wei *et al.*, 2020]. The actor's encoder consists of two one-dimensional convolutional neural networks (CNN) with a filter size of 128 and the decoder of an LSTM with a hidden size of 128. The critic consists of two one-dimensional CNNs and a fully connected output layer of size $29 \times 29 = 841$. Both actor and critic learning rates are set to 10^{-4} and the batch size is set to 128. Each selected square has a station cost of 5 units and each connection between stations a cost of 1 unit. The total budget for all experiments is set to 210 and the station limit is set to 45. The agent is trained for a maximum of 3500 epochs.

We evaluated the reward functions on total utility (satisfied OD demand %), equality of benefits (Gini index of satisfied OD demand %), and performance of the lowest house price quintile. Wherever a line connects origin and destination squares that belong to different groups, the satisfied demand of both groups is counted. This means that the sum of all groups' ODs does not equal the total satisfied OD. Detailed results of all reward functions presented on Table 1.

As expected, the baseline utility maximization reward function performs the best on total satisfied demand. However, as Figure 3 (blue) shows, it is unevenly distributed between the five quintiles. Even when setting the equity weight to a maximum of 1, inequality is still present. While this function aims at connecting low-developed areas with high-developed ones, it does not optimize for the distribution of benefits between the groups and achieves maximum reward when the line connects high-developed areas next to each other.

The proposed reward functions outperform the baselines in their respective fairness goals. In Figure 3 (striped red) we show that the GGI reward function achieves near equality of the added benefits of the new line and a Gini index of 0.05. In accordance with Rawl's theory of justice, Lowest Quintile reward achieves the best performance in the most disadvantaged group. Naturally, it also leads to the highest total inequality of benefits, as it only focuses on one group. Finally, Variance Regularization , while not being the best in a single metric, accomplishes the lowest gap between utility and fairness.

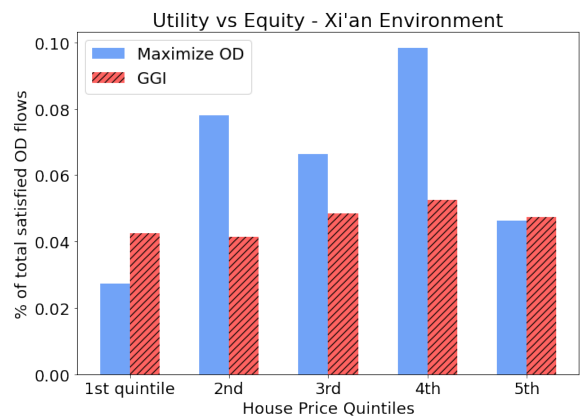

Figure 3: We show the results of applying two distinct reward functions to design a transportation line in the Xi'an environment. In blue we apply a Maximize OD baseline reward (Equation 1) which leads to an uneven distribution of satisfied OD over different groups (i.e., areas with different wealth levels, here represented by house prices). In red (striped), the GGI reward function (Equation 5) generates a line that achieves near-equal distribution of the added utility between groups but comes at the expense of total utility. In Table 1 we report full results for all reward functions.

5 Conclusion

In this work in progress paper, we used deep reinforcement learning to explore the trade-off between utility and fairness maximization in the transport network design problem. Our preliminary analysis and results shows that different fairness goals can be achieved by changing the reward functions accordingly, without the need of introducing new constraints.

In the future, we plan to further explore this trade-off and expand our definitions of fairness to include more theories (including *sufficientarianism*) and attempt to find the fair reward function that achieves better results on overall utility. We further plan to test the generalizability of the model by applying it to the city of Amsterdam. Finally, we plan to evaluate our methodology on transport accessibility alongside satisfied travel demand, as previous research on transportation science points accessibility as central for measuring urban inequality [Martens and Bastiaanssen, 2019].

Acknowledgments

This research was supported by the Innovation Center for AI (ICAI).

References

[Behbahani *et al.*, 2019] Hamid Behbahani, Sobhan Nazari, Masood Jafari Kang, and Todd Litman. A conceptual framework to formulate transportation network design problem considering social equity criteria. *Transportation Research Part A: Policy and Practice*, 125:171–183, July 2019.

[Bello *et al.*, 2017] Irwan Bello, Hieu Pham, Quoc V. Le, Mohammad Norouzi, and Samy Bengio. Neural Combinatorial Optimization with Reinforcement Learning. *arXiv:1611.09940 [cs, stat]*, January 2017. arXiv: 1611.09940.

[Fan and Machemehl, 2006] Wei Fan and Randy B. Machemehl. Using a Simulated Annealing Algorithm to Solve the Transit Route Network Design Problem. *Journal of Transportation Engineering*, 132(2):122–132, February 2006. Publisher: American Society of Civil Engineers.

[Farahani *et al.*, 2013] Reza Zanjirani Farahani, Elnaz Miandoabchi, W. Y. Szeto, and Hannaneh Rashidi. A review of urban transportation network design problems. *European Journal of Operational Research*, 229(2):281–302, September 2013.

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

[Gutiérrez-Jarpa *et al.*, 2018] Gabriel Gutiérrez-Jarpa, Gilbert Laporte, and Vladimir Marianov. Corridor-based metro network design with travel flow capture. *Computers & Operations Research*, 89:58–67, January 2018.

[Kool *et al.*, 2018] Wouter Kool, Herke van Hoof, and Max Welling. Attention, learn to solve routing problems! In *International Conference on Learning Representations*, 2018.

[Martens and Bastiaanssen, 2019] Karel Martens and Jeroen Bastiaanssen. An index to measure accessibility poverty risk. In *Measuring Transport Equity*, pages 39–55. Elsevier, 2019.

[Nazari *et al.*, 2018] MohammadReza Nazari, Afshin Oroojlooy, Lawrence Snyder, and Martin Takac. Reinforcement learning for solving the vehicle routing problem. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.

[Nicoletti *et al.*, 2022] Leonardo Nicoletti, Mikhail Sirenko, and Trivik Verma. Disadvantaged communities have lower access to urban infrastructure. *arXiv preprint arXiv:2203.13784*, 2022.

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

[Owais and Osman, 2018] Mahmoud Owais and Mostafa K. Osman. Complete hierarchical multi-objective genetic algorithm for transit network design problem. *Expert Systems with Applications*, 114:143–154, December 2018.

[Raman *et al.*, 2021] Naveen Raman, Sanket Shah, and John Dickerson. Data-Driven Methods for Balancing Fairness and Efficiency in Ride-Pooling. *arXiv:2110.03524 [cs]*, October 2021. arXiv: 2110.03524.

[Siddique *et al.*, 2020] Umer Siddique, Paul Weng, and Matthieu Zimmer. Learning Fair Policies in Multiobjective (Deep) Reinforcement Learning with Average and Discounted Rewards. *arXiv:2008.07773 [cs]*, August 2020. arXiv: 2008.07773.

[Szeto and Jiang, 2014] W. Y. Szeto and Y. Jiang. Transit route and frequency design: Bi-level modeling and hybrid artificial bee colony algorithm approach. *Transportation Research Part B: Methodological*, 67:235–263, September 2014.

[Wei *et al.*, 2020] Yu Wei, Minjia Mao, Xi Zhao, Jianhua Zou, and Ping An. City Metro Network Expansion with Reinforcement Learning. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2646–2656, Virtual Event CA USA, August 2020. ACM.