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):
https://drive.google.com/file/d/1XQIhy8NLYD9WtP45s8UYVgidTBVfQ77h/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, 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

\[ U_{od} = \sum_{(i,j) \in E} d_{ij} \]

where \( d_{ij} \) is the demand between nodes \( i \) and \( j \) and \( E \) is the set of edges. However, this objective does not account for fairness or equity in the distribution of benefits among groups within the city.

One way to incorporate fairness is to use reward functions that can implement different objectives relevant to the stakeholders and communities. For example, one can use reward functions based on Utilitarianism, Equal Sharing of Benefits, Narrowing the Gap, and Rawl’s theory of justice. These reward functions can be employed to achieve fairness objectives.

By using Deep RL, we can explore different reward functions and their impact on the satisfaction of different groups within the city. This allows us to understand the trade-offs between utility and fairness and to design transport networks that are fair and efficient.
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:

$$\max \ U_{od}(G(N, E))$$

s.t. $$\text{cost}(G) < B$$

and $$|N| < T$$

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 Beshabani 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.

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

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

$$\max \ U(G)$$

s.t. $$|U_i(G) - U_j(G)| < T$$

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-
3.2 Fair Reward Functions

The feasibility rules are updated at each time step: for each optimization function, the direction the agent takes is dictated by the probability distribution over all actions (square to be selected in the current state \( s_t \) and estimated at \( t+1 \)).

- \( s_t \in S \): the sequence of selected grid squares up to timestep \( t - 1 \).
- \( a_t \in A \): the selected station at timestep \( t \). Determined by a parametrized policy \( \pi_\theta \).
- \( P: \) 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 \( \pi(a_t|s_t, \theta) = Pr(a_t|s_t, \theta) \). 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.

**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.

\[
r = w_1 U(G) + w_2 D(G)
\]

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 \( \lambda \). It has been successfully used to equalize treatment among neighborhoods in the ride-pooling assignment problem [Raman et al., 2021].

\[
r = \sum_a U_a(G) - \lambda \text{var}(U(G))
\]

**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].

\[
r = U_{a_0}(G)
\]

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].

\[
r = GGI_w(U(G)) = \sum_a w_a v_a^t
\]

where \( w_a \in [0, 1] \) is a weight vector such as \( w_1 > w_2 > ... > w_{|A|} \).
Table 1: Results of different Reward Functions on Utility and Fairness objectives.

<table>
<thead>
<tr>
<th>Reward Function</th>
<th>Fairness Notion</th>
<th>Total OD %</th>
<th>Gini Index</th>
<th>Lowest Quintile OD %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximize OD (baseline, $w_1 = 1$)</td>
<td>Utilitarianism</td>
<td>6.01</td>
<td>0.21</td>
<td>2.90</td>
</tr>
<tr>
<td>Maximize Equity (baseline, $w_2 = 1$)</td>
<td>Utilitarianism</td>
<td>4.32</td>
<td>0.22</td>
<td>2.41</td>
</tr>
<tr>
<td>Variance Regularization ($\lambda = 5$)</td>
<td>Narrowing the Gap</td>
<td>4.58</td>
<td>0.08</td>
<td>4.16</td>
</tr>
<tr>
<td>Lowest Quintile</td>
<td>Rawls</td>
<td>2.47</td>
<td>0.39</td>
<td>6.77</td>
</tr>
<tr>
<td>Generalized Gini Index (GGI)</td>
<td>Equal Sharing</td>
<td>4.34</td>
<td>0.05</td>
<td>4.24</td>
</tr>
</tbody>
</table>

**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 $\times$ 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.

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


