Improving Route Choice Models by Incorporating Contextual Factors via Knowledge Distillation

Qun Liu, Supratik Mukhopadhyay, Yimin Zhu, Ravindra Gudishala Sanaz Saeidi, Alimire Nabijiang

Louisiana State University
Introduction

- The total cost and time loss associated with traffic congestion in the US has been reported to be more than $121 billion per year and 38 hours per person, respectively [1].
- Most of the route choice models link route characteristics of alternative routes to those chosen by the drivers.
- **High fidelity** route choice models are required to predict traffic levels with higher accuracy. Existing route choice models use revealed preference behavior to model route choice.
- Adding subjective or contextual requires availability of the data at individual or disaggregate level. Stated Choice Experiments (SCEs) are a scientific methodology to capture the effect of context sensitive factors in route choice.
Related Work

- Route choice behavior theories began to evolve in the late eighties and early nineties as engineers' understanding of route choice behavior improved by studying data about empirical route choice behavior.
- Pursula and Talvite [2] developed a mathematical route model by postulating that drivers do consider other factors apart from travel time in making a route choice.
- In [3-4], the authors discovered that commuters prefer to use habitual routes when traveling in familiar areas as opposed to choosing a route that provides them with maximum utility.
- Hinton et al. [5] proposed a different compression technique for knowledge distillation in a neural network.
Contributions

- It presents a novel approach using knowledge distillation for developing high-fidelity route choice models by augmenting existing baseline models with information about drivers' reaction to contextual factors acquired from SCEs in IVEs.

- We present a general end-to-end knowledge distillation framework that uses a multilayer perceptron as a feature extraction network to provide a feature learning architecture for teacher and student networks and then transfers knowledge from the former to the latter by optimizing distillation loss.
Proposed Method

Overview of our framework architecture
Proposed Method

- **Basic Route Choice Model**: the baseline model predicts the probability of exiting a highway through a given exit using the following equation,

\[ P_b = \alpha_b T \]  

(1)

- **Multilayer Perceptron Models**: the feature vector \( F \) from feature extraction network \( M \) can be described as,

\[ F = M(g(X); \theta_m) \]  

(2)

- The probability that an individual driver exits through highway exit \( k \) for a feature vector \( F \) is given by the following equation,

\[ P(y = k \mid F) = f^k(F; \theta_f) \]  

(3)
Proposed Method

- Adopted the cross-entropy loss, described by the following equation,

$$\mathcal{L}_F(F, y) = - \sum_{k} y^k \log P(y = k \mid F)$$  \hspace{1cm} (4)

- **Knowledge Distillation**: aims to learn a student model $S(\cdot)$ from a well-trained teacher model $T(\cdot)$ by minimizing the distillation, given by equations,

$$P_i = \frac{\exp(z_i)}{\sum_j \exp(z_j)}$$  \hspace{1cm} (5)

$$P_i = \frac{\exp(z_i/T)}{\sum_j \exp(z_j/T)}$$  \hspace{1cm} (6)

$$\mathcal{L}_d(X, Y) = \alpha \sum_{n} \mathcal{L}_1(y_n, S(x_n; \theta_s)) + \beta \sum_{n} \mathcal{L}_2(S(x_n; \theta_s), T(x_n; \theta_t))$$  \hspace{1cm} (7)

$$\mathcal{L}(x, y) = - \frac{1}{N} \sum_{n=1}^{N} \sum_{m=1}^{M} [x_{nm} \log(y_{nm}) + (1 - x_{nm}) \log(1 - y_{nm})]$$  \hspace{1cm} (8)
Proposed Method

- **Algorithm**: for each iteration, the algorithm computes the logits for the input data. The cross-entropy loss is computed for backpropagating the gradients in the student network. The parameters in the student network are updated using gradient descent,

\[
\nabla_{\theta_s} \frac{1}{n} \sum_{i \in D} \mathcal{L}_d(i) \tag{9}
\]

After training, the student network is tested standalone on the test set.

---

**Algorithm 1: Training Algorithm**

1. Pretrain(teacher)
2. Initialize(student)
3. for epoch = 1, 2, ..., K do
   4. for number of batches do
      5. \( D \leftarrow \) Sample \( n \) examples with its labels \( y \)
      6. \( z \leftarrow \) Forward(\( \text{teacher}, \text{student}, D \))
      7. \( \mathcal{L}_d \leftarrow \) Loss(\( z, y \))
      8. \( \text{grad} \leftarrow \) Backward(\( \mathcal{L}_d \))
      9. Update the parameters of student network regarding the gradients, \( \nabla_{\theta_s} \frac{1}{n} \sum_{i \in D} \mathcal{L}_d(i) \)
   10. end
4. end
Experiment Evaluation

- **IVE Experimental Setting**: based on the I-10, starting off the Mississippi River bridge all the way to the intersection of Perkins Rd and Staring Ln. Along the way, four alternate routes were introduced to the participants, Exits 1, 2, 3, and 4, the latter of which would be College Dr.,

- **41 volunteers** (20 male and 21 females; age: 31.44 ± 7.97) participate in the SCE.

<table>
<thead>
<tr>
<th>ID</th>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Traffic</td>
<td>1 Normal, 2 Medium, 3 Heavy</td>
</tr>
<tr>
<td>2</td>
<td>Urgency</td>
<td>1 Urgency, 2 Non-Urgency</td>
</tr>
<tr>
<td>3</td>
<td>Social impact</td>
<td>1 No, 2 Yes</td>
</tr>
<tr>
<td>4</td>
<td>Age</td>
<td>1 less than 25, 2 greater than or equal to 25</td>
</tr>
<tr>
<td>5</td>
<td>Gender</td>
<td>1 Male, 2 Female</td>
</tr>
<tr>
<td>6</td>
<td>Race</td>
<td>1 Middle Eastern, 2 White, 3 Other</td>
</tr>
<tr>
<td>7</td>
<td>Education</td>
<td>1 Post graduate degree, 2 High school graduate, 3 College graduate</td>
</tr>
<tr>
<td>8</td>
<td>Employment status</td>
<td>1 Employed part time, 2 Employed full time, 3 Student, 4 Unemployed looking for work</td>
</tr>
<tr>
<td>9</td>
<td>Concern while stuck in traffic</td>
<td>1 Hours of extra travel time, 2 Chaos, 3 Monetized value of delay, 4 Speed reduction due to congestion</td>
</tr>
<tr>
<td>10</td>
<td>Familiarity with the environment</td>
<td>1 Once a week, 2 Once a year, 3 Once a month, 4 More than once a week, 5 Never</td>
</tr>
<tr>
<td>11</td>
<td>Financial Concerns</td>
<td>1 Sometimes, 2 Always, 3 Most of the time, 4 About half the time, 5 Never</td>
</tr>
<tr>
<td>12</td>
<td>Choice</td>
<td>1 First exit, 2 Second exit, 3 Third exit, 4 Fourth exit, 5 Fifth exit</td>
</tr>
</tbody>
</table>
Experiment Evaluation

- **VR Data**: total of 410 driving records collected. Since the data collected is limited, to better train the teacher network, we augmented it using a Gaussian mixture model (GMM).
  - After data augmentation, 10,000 synthetic driving records were generated.
  - The VR data was divided by 80% for training and 20% for testing and the training set was used to train the teacher.
- **Basic Data**: uniformly randomly sampled 10,000 driving records based on the probability distribution predicted by the baseline route choice model.
  - We divided this dataset into training (80%) and testing sets (20%). During knowledge distillation, the teacher model, pretrained on the augmented VR data, provides the prior knowledge for our framework.
- **Real Data**: Given the real traffic volumes captured at the four exits, the probability of taking an exit is computed by,

  \[ P_e = \frac{V_e}{\sum_{i \in E} V_i} \]  

  (10)
## Experiment Evaluation

### Implementation Details

Table shows the architectures of the different teacher networks that we considered in our experiments.

<table>
<thead>
<tr>
<th>Network Architecture</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>10n-0.25DP-30n-0.35DP-20n</td>
<td>92.02</td>
</tr>
<tr>
<td>10n-0.25DP-50n-0.35DP-30n-0.15DP</td>
<td>93.65</td>
</tr>
<tr>
<td>10n-0.25DP-50n-0.35DP-50n-0.25DP-50n-0.15DP</td>
<td>94.37</td>
</tr>
<tr>
<td>10n-0.25DP-30n-0.35DP-20n-0.25DP-50n-0.45DP</td>
<td>94.95</td>
</tr>
</tbody>
</table>
Evaluation

(a) Prediction accuracy of teacher network on VR data  
(b) Prediction accuracy of student network on Basic data  
(c) Prediction accuracy of teacher-student network on Basic data  
(d) The comparisons with the predictions from our framework

Fig. 7. The prediction accuracy of our framework. Top two: shows the prediction accuracy of teacher and student networks on testing set. Bottom: shows the prediction accuracy of our framework (teacher-student network) on the left; on the right shows the probabilities of leave at exits calculated based on the predictions from our framework. We compared our results with the results from the Basic model and Real data, we also plotted our VR data for showing better insights.

<table>
<thead>
<tr>
<th>Network</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher</td>
<td>97.93</td>
</tr>
<tr>
<td>Student</td>
<td>77.45</td>
</tr>
<tr>
<td>Our Model</td>
<td>95.20</td>
</tr>
</tbody>
</table>

TABLE III
Comparison of classification accuracy (%) of our framework with teacher and student networks on Basic data
Conclusion

- We proposed a novel approach for developing high-fidelity route choice models with increased predictive power by augmenting existing aggregate level models with contextual information obtained from SCE carried out in an IVE through the use of knowledge distillation.

- We presented a general end-to-end knowledge distillation framework that uses a multilayer perceptron as a feature extraction network to provide a feature learning architecture for teacher and student networks and then transfers knowledge from the former to the latter by optimizing distillation loss.

- Experimental results have demonstrated that that the predictions of the augmented models produced by our approach are much closer to reality than that of the baseline.
References


Thank You!