Development and evaluation of a hybrid travel time forecasting model

Jinsoo You and Tschangho John Kim

Transportation Research Part C 8 - Emerging Technologies

Presented by Troels V. Larsen

AALBORG UNIVERSITY
Introduction

- Department of Urban and Regional Planning
- University of Illinois at Urbana-Champaign

- Travel time estimation is hard using only a single forecasting method.
- Goal: Implement a hybrid travel time forecasting model
- Based on GIS technologies
  - “...a computer system capable of integrating, storing, editing, analyzing, sharing and displaying geographically-referenced information.”
Department of Urban and Regional Planning
University of Illinois at Urbana-Champaign
Travel time estimation is hard using only a single forecasting method.

Goal: Implement a hybrid travel time forecasting model
Based on GIS technologies

“...a computer system capable of integrating, storing, editing, analyzing, sharing and displaying geographically-referenced information.”
Introduction

- Department of Urban and Regional Planning
- University of Illinois at Urbana-Champaign
- Travel time estimation is hard using only a single forecasting method.
- Goal: Implement a hybrid travel time forecasting model
- Based on GIS technologies
  - “...a computer system capable of integrating, storing, editing, analyzing, sharing and displaying geographically-referenced information.”
Department of Urban and Regional Planning
University of Illinois at Urbana-Champaign
Travel time estimation is hard using only a single forecasting method.
Goal: Implement a hybrid travel time forecasting model
Based on GIS technologies
“...a computer system capable of integrating, storing, editing, analyzing, sharing and displaying geographically-referenced information.”
Study focus

- Historical database development
- Historical database – road network integration
- Hybrid travel time forecasting model
Study focus

- Historical database development
- Historical database – road network integration
- Hybrid travel time forecasting model
Study focus

- Historical database development
- Historical database – road network integration
- Hybrid travel time forecasting model
<table>
<thead>
<tr>
<th>Architecture</th>
</tr>
</thead>
</table>

- Overview
- Scenario specifics
- Network representation
- Historical database

Introduction

Forecasting

Experimental evaluation

Conclusion

Jinsoo You and Tschangho John Kim

Development and evaluation of a hybrid travel time forecasting...
The transmitted raw traffic data are initially stored in the real-time traffic data storage, an intermediate data storage before updating the historical database. Data transmission intervals are every 30 s for highway data and 5 min for arterial data. Thus, the intermediate storage becomes necessary because there is no other place to store the data during the forecasting process, and the historical database is not allowed to alter or update when a travel time forecasting process is activated.

Since the hybrid forecasting model is designed to predict future travel times for a period of 15–60 min, we assume that it is unacceptable if a computational delay is greater than 15 min for a...
Scenario specifics

- Recording intervals:
  - Highway data: 30 seconds
  - Arterial data: 5 minutes
- Computation time: Max 15 minutes, preferably less than 1-2 minutes.
- Usage: Predict travel times 15-60 minutes into the future.
Scenario specifics

- Recording intervals:
  - Highway data: 30 seconds
  - Arterial data: 5 minutes

- Computation time: Max 15 minutes, preferably less than 1-2 minutes.

- Usage: Predict travel times 15-60 minutes into the future.
Scenario specifics

- Recording intervals:
  - Highway data: 30 seconds
  - Arterial data: 5 minutes

- Computation time: Max 15 minutes, preferably less than 1-2 minutes.

- Usage: Predict travel times 15-60 minutes into the future.
Network representation

![Network diagram with nodes and links]

<table>
<thead>
<tr>
<th>Link ID</th>
<th>From Node#</th>
<th>To Node#</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>6</td>
</tr>
</tbody>
</table>
## Historical database

- Time and link are recorded
- Each link is stored twice, unless it is a one way street.
- *(Link ID, Historical DB ID, From Node, To Node)*

<table>
<thead>
<tr>
<th>Time</th>
<th>Link</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>0:00</td>
<td>34</td>
</tr>
<tr>
<td>0:05</td>
<td>33</td>
</tr>
<tr>
<td>0:10</td>
<td>29</td>
</tr>
<tr>
<td>0:15</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Time and link are recorded
- Each link is stored twice, unless it is a one way street.
- (Link ID, Historical DB ID, From Node, To Node)

<table>
<thead>
<tr>
<th>Time</th>
<th>Link</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>0:00</td>
<td>34</td>
</tr>
<tr>
<td>0:05</td>
<td>33</td>
</tr>
<tr>
<td>0:10</td>
<td>29</td>
</tr>
<tr>
<td>0:15</td>
<td>27</td>
</tr>
</tbody>
</table>
Historical database

- Time and link are recorded
- Each link is stored twice, unless it is a one way street.
- (Link ID, Historical DB ID, From Node, To Node)

<table>
<thead>
<tr>
<th>Time</th>
<th>Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>0:00</td>
<td>34 29 ... 27 14</td>
</tr>
<tr>
<td>0:05</td>
<td>33 31 ... 33 12</td>
</tr>
<tr>
<td>0:10</td>
<td>29 27 ... 32 11</td>
</tr>
<tr>
<td>0:15</td>
<td>27 25 ... 29 9</td>
</tr>
</tbody>
</table>
Forecasting

Development and evaluation of a hybrid travel time forecasting algorithm.
Forecasting modules

- Data preprocessing
  - Travel time forecasting
  - Evaluation
### Forecasting modules

- Data preprocessing
- Travel time forecasting
- Evaluation
Forecasting modules

- Data preprocessing
- Travel time forecasting
- Evaluation
Data preprocessing

- Screens and filters noise
  - Wavelet transformation technique
  - Outlier detection algorithm
- Remove noise from probe vehicles such as delivery trucks
Data preprocessing

- Screens and filters noise
  - Wavelet transformation technique
  - Outlier detection algorithm
- Remove noise from probe vehicles such as delivery trucks
Data preprocessing

- Screens and filters noise
  - Wavelet transformation technique
  - Outlier detection algorithm

- Remove noise from probe vehicles such as delivery trucks
Data preprocessing

- Screens and filters noise
  - Wavelet transformation technique
  - Outlier detection algorithm
- Remove noise from probe vehicles such as delivery trucks
Forecasting through method learning

- Parameter learning
  - Relies on k-nearest neighbour
Forecasting through method learning

- Parameter learning
- Relies on k-nearest neighbour
### Parameters

**Domains of model parameters**

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Type</th>
<th>Domain</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forecasting range</td>
<td>Discrete</td>
<td>{15, 30, 45, 60}</td>
<td>Minute</td>
</tr>
<tr>
<td>Search data segment length</td>
<td>Discrete</td>
<td>{15, 30, 45, 60}</td>
<td>Minute</td>
</tr>
<tr>
<td>Day of the week</td>
<td>Binary</td>
<td>{Consider, Ignore}</td>
<td></td>
</tr>
<tr>
<td>Search range</td>
<td>Discrete</td>
<td>{1, 2, 3}</td>
<td>Hour</td>
</tr>
<tr>
<td>Large K</td>
<td>Discrete</td>
<td>{1, 2, 3, 4, 5, 6, 7, 8, 9, 10}</td>
<td></td>
</tr>
<tr>
<td>Small k</td>
<td>Discrete</td>
<td>{1, 2, 3, 4, 5, 6, 7, 8, 9, 10}</td>
<td></td>
</tr>
<tr>
<td>Local estimation method</td>
<td>Binary</td>
<td>{Local averaging, Local fitting}</td>
<td></td>
</tr>
<tr>
<td>Data preprocessing</td>
<td>Binary</td>
<td>{Wavelet, Outlier detection}</td>
<td></td>
</tr>
</tbody>
</table>
Evaluation

- Is activated as actual travel times arrives
  - If the difference between actual and estimated travel time is too large, the parameters are readjusted using the ML module.
- ML Module:
  - Generates training samples
  - Identifies the lowest forecasting error from each parameter
  - Updates the hybrid model with the new parameters
Evaluation

- Is activated as actual travel times arrives
- If the difference between actual and estimated travel time is too large, the parameters are readjusted using the ML module.

ML Module:
- Generates training samples
- Identifies the lowest forecasting error from each parameter
- Updates the hybrid model with the new parameters
Evaluation

- Is activated as actual travel times arrives
- If the difference between actual and estimated travel time is too large, the parameters are readjusted using the ML module.
- **ML Module:**
  - Generates training samples
  - Identifies the lowest forecasting error from each parameter
  - Updates the hybrid model with the new parameters
3.4.4. Evaluation

The evaluation module will be activated when observed actual travel times become available. There is a time delay because the observed actual travel times are not available at the time when a user orders a forecasting task. Based on the evaluation result, if a priori given acceptable error margin has not been met, the hybrid forecasting model activates the ML module to readjust.
Experimental Evaluation
Experiment:
- 200 randomly selected points from the historical database
- Separated into arterial and highway data
- Each experiment within 24 hours
Experimental evaluation

**Fig. 14** Forecasting based on randomly selected discrete time points with arterial data (origin: B1173; destination: B1179).

**Fig. 15** Forecasting based on randomly selected discrete time points with highway data (origin: PanGyo; destination: SinGal).

<table>
<thead>
<tr>
<th>Case</th>
<th>Observed (sec)</th>
<th>Forecasted (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1</td>
<td>15.47</td>
<td>8.02</td>
</tr>
<tr>
<td>Case 2</td>
<td>19.63</td>
<td>9.88</td>
</tr>
</tbody>
</table>
Experimental evaluation

<table>
<thead>
<tr>
<th>Case</th>
<th>RMSE (sec)</th>
<th>MAPE (%)</th>
<th>$\rho$</th>
<th>Average Observed Travel Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1</td>
<td>14.27</td>
<td>2.22</td>
<td>0.83</td>
<td>404.48</td>
</tr>
<tr>
<td>Case 2</td>
<td>8.46</td>
<td>1.67</td>
<td>0.90</td>
<td>390.52</td>
</tr>
</tbody>
</table>

Fig. 14. Forecasting based on randomly selected discrete time points with arterial data (origin: B1173; destination: B1179).

Fig. 15. Forecasting based on randomly selected discrete time points with highway data (origin: PanGyo; destination: SinGal).
Relation to our project

- Travel time estimation
  - Offline / Online
  - Method learning
  - Evaluation of actual travel time
- GIS
  - Shape files
  - Software built on top of GIS
- Data storage
  - Relational database
  - Data warehouse
Relation to our project

- **Travel time estimation**
  - Offline / Online
  - Method learning
  - Evaluation of actual travel time

- **GIS**
  - Shape files
  - Software built on top of GIS

- **Data storage**
  - Relational database
  - Datawarehouse
Relation to our project

- Travel time estimation
  - Offline / Online
  - Method learning
  - Evaluation of actual travel time
- GIS
  - Shape files
  - Software built on top of GIS
- Data storage
  - Relational database
  - Datawarehouse
Relation to our project

- Travel time estimation
  - Offline / Online
  - Method learning
  - Evaluation of actual travel time

- GIS
  - Shape files
  - Software built on top of GIS

- Data storage
  - Relational database
  - Datawarehouse
Relation to our project

- Travel time estimation
  - Offline / Online
  - Method learning
  - Evaluation of actual travel time
- GIS
  - Shape files
  - Software built on top of GIS
- Data storage
  - Relational database
  - Datasware warehouse
Relation to our project

- Travel time estimation
  - Offline / Online
  - Method learning
  - Evaluation of actual travel time

- GIS
  - Shape files
    - Software built on top of GIS

- Data storage
  - Relational database
  - Data warehouse
Relation to our project

- Travel time estimation
  - Offline / Online
  - Method learning
  - Evaluation of actual travel time
- GIS
  - Shape files
  - Software built on top of GIS
- Data storage
  - Relational database
  - Datawarehouse
Relation to our project

- Travel time estimation
  - Offline / Online
  - Method learning
  - Evaluation of actual travel time
- GIS
  - Shape files
  - Software built on top of GIS
- Data storage
  - Relational database
  - Data warehouse
Relation to our project

- Travel time estimation
  - Offline / Online
  - Method learning
  - Evaluation of actual travel time

- GIS
  - Shape files
  - Software built on top of GIS

- Data storage
  - Relational database
  - Data warehouse
Relation to our project

- Travel time estimation
  - Offline / Online
  - Method learning
  - Evaluation of actual travel time

- GIS
  - Shape files
  - Software built on top of GIS

- Data storage
  - Relational database
  - Datawarehouse
Relation to our project

- Travel time estimation
  - Offline / Online
  - Method learning
  - Evaluation of actual travel time
- GIS
  - Shape files
  - Software built on top of GIS
- Data storage
  - Relational database
  - Datawarehouse
Strengths

- Interesting ideas
- Sensible work
- Possibly a good average error rate
Strengths

- Interesting ideas
- Sensible work
- Possibly a good average error rate
Strengths

- Interesting ideas
- Sensible work
- Possibly a good average error rate
Weaknesses

- Illogical structure
- Lacks a good overview
- Spends too much time discussing subjects that are irrelevant to the solution
- Figures are not used optimally - some should be explained better
- Inconclusive results
- Bad running time for highway data
- Nothing mentioned about time or space complexity. (Not a CS article)
Weaknesses

- Illogical structure
- Lacks a good overview
  - Spends too much time discussing subjects that are irrelevant to the solution
  - Figures are not used optimally - some should be explained better
- Inconclusive results
- Bad running time for highway data
- Nothing mentioned about time or space complexity. (Not a CS article)
Weaknesses

- Illogical structure
- Lacks a good overview
- Spends too much time discussing subjects that are irrelevant to the solution
  - Figures are not used optimally - some should be explained better
- Inconclusive results
- Bad running time for highway data
- Nothing mentioned about time or space complexity. (Not a CS article)
Weaknesses

- Illogical structure
- Lacks a good overview
- Spends too much time discussing subjects that are irrelevant to the solution
- Figures are not used optimally - some should be explained better
- Inconclusive results
- Bad running time for highway data
- Nothing mentioned about time or space complexity. (Not a CS article)
Weaknesses

- Illogical structure
- Lacks a good overview
- Spends too much time discussing subjects that are irrelevant to the solution
- Figures are not used optimally - some should be explained better
- Inconclusive results
  - Bad running time for highway data
  - Nothing mentioned about time or space complexity. (Not a CS article)
Weaknesses

- Illogical structure
- Lacks a good overview
- Spends too much time discussing subjects that are irrelevant to the solution
- Figures are not used optimally - some should be explained better
- Inconclusive results
- Bad running time for highway data
- Nothing mentioned about time or space complexity. (Not a CS article)
Weaknesses

- Illogical structure
- Lacks a good overview
- Spends too much time discussing subjects that are irrelevant to the solution
- Figures are not used optimally - some should be explained better
- Inconclusive results
- Bad running time for highway data
- Nothing mentioned about time or space complexity. (Not a CS article)
Related work

- **A simple and Effective Method for Predicting Travel Times on Freeways** – John Rice and Erik van Zwet
  - Travel-time prediction on freeways
  - Uses linear regression
- **Integration of GPS and GIS for traffic congestion studies** – Taylor, Wooley and Zito
  - Relies on several GIS layers
  - Same journal
- **Traffic variable estimation and traffic signal based soft computation** – Conglin, Wu and Yuejin
  - Uses neural networks
  - Based on loop detectors
Related work

- *A simple and Effective Method for Predicting Travel Times on Freeways* – John Rice and Erik van Zwet
  - Travel-time prediction on freeways
  - Uses linear regression

- *Integration of GPS and GIS for traffic congestion studies* – Taylor, Wooley and Zito
  - Relies on several GIS layers
  - Same journal

- *Traffic variable estimation and traffic signal based soft computation* – Conglin, Wu and Yuejin
  - Uses neural networks
  - Based on loop detectors
Related work

- **A simple and Effective Method for Predicting Travel Times on Freeways** – John Rice and Erik van Zwet
  - Travel-time prediction on freeways
  - Uses linear regression

- **Integration of GPS and GIS for traffic congestion studies** – Taylor, Wooley and Zito
  - Relies on several GIS layers
  - Same journal

- **Traffic variable estimation and traffic signal based soft computation** – Conglin, Wu and Yuejin
  - Uses neural networks
  - Based on loop detectors
Related work

- **A simple and Effective Method for Predicting Travel Times on Freeways** – John Rice and Erik van Zwet
  - Travel-time prediction on freeways
  - Uses linear regression

- **Integration of GPS and GIS for traffic congestion studies** – Taylor, Wooley and Zito
  - Relies on several GIS layers
  - Same journal

- **Traffic variable estimation and traffic signal based soft computation** – Conglin, Wu and Yuejin
  - Uses neural networks
  - Based on loop detectors
Related work

- A simple and Effective Method for Predicting Travel Times on Freeways – John Rice and Erik van Zwet
  - Travel-time prediction on freeways
  - Uses linear regression

- Integration of GPS and GIS for traffic congestion studies – Taylor, Wooley and Zito
  - Relies on several GIS layers
  - Same journal

- Traffic variable estimation and traffic signal based soft computation – Conglin, Wu and Yuejin
  - Uses neural networks
  - Based on loop detectors
**Related work**

- *A simple and Effective Method for Predicting Travel Times on Freeways* – John Rice and Erik van Zwet
  - Travel-time prediction on freeways
  - Uses linear regression

- *Integration of GPS and GIS for traffic congestion studies* – Taylor, Wooley and Zito
  - Relies on several GIS layers
  - Same journal

- *Traffic variable estimation and traffic signal based soft computation* – Conglin, Wu and Yuejin
  - Uses neural networks
  - Based on loop detectors
Related work

- **A simple and Effective Method for Predicting Travel Times on Freeways** – John Rice and Erik van Zwet
  - Travel-time prediction on freeways
  - Uses linear regression
- **Integration of GPS and GIS for traffic congestion studies** – Taylor, Wooley and Zito
  - Relies on several GIS layers
  - Same journal
- **Traffic variable estimation and traffic signal based soft computation** – Conglin, Wu and Yuejin
  - Uses neural networks
  - Based on loop detectors
Related work

- **A simple and Effective Method for Predicting Travel Times on Freeways** – John Rice and Erik van Zwet
  - Travel-time prediction on freeways
  - Uses linear regression
- **Integration of GPS and GIS for traffic congestion studies** – Taylor, Wooley and Zito
  - Relies on several GIS layers
  - Same journal
- **Traffic variable estimation and traffic signal based soft computation** – Conglin, Wu and Yuejin
  - Uses neural networks
  - Based on loop detectors
Related work

- **A simple and Effective Method for Predicting Travel Times on Freeways** – John Rice and Erik van Zwet
  - Travel-time prediction on freeways
  - Uses linear regression

- **Integration of GPS and GIS for traffic congestion studies** – Taylor, Wooley and Zito
  - Relies on several GIS layers
  - Same journal

- **Traffic variable estimation and traffic signal based soft computation** – Conglin, Wu and Yuejin
  - Uses neural networks
  - Based on loop detectors