Finding Robust Train Paths in Dense Corridors

Holger Flier (ETH)
Thomas Graffagnino (SBB)
Marc Nunkesser (ETH)
Motivation and Problem Statement

- Demand for passenger train transportation increases faster than infrastructure can be extended
- Traffic gets denser
- Risk of delay increases

Problem: Find a schedule for an additional train on a dense corridor that has a low risk of being delayed upon arrival at the final station

12.02.2009 H.Flier, T.Graffagnino, M.Nunkesser
Basic Idea

- Historical delay data contains implicit information
  - Dependencies between trains
  - Resource bottlenecks
  - Dispatching decisions
  - ...

- Learn linear regression models from the data
  - to predict the risk of delay of a train

- Combine statistical and combinatorial models
  - to find optimal suggestions for train paths
LINEAR REGRESSION MODELS
Linear Regression Models

- Goal: “predict” the delay of a train for *a day in the past*

- Linear regression models use predictors from the data

\[ \hat{\delta} = \alpha + \sum_{i}^{n} \beta_{i} \text{ predictor}_i \]
Linear Regression Models:
Predictors from the Data

- Train density around planned arrival/departure time
- Timetable measures
  - Scheduled time difference to previous / next trains in the same / opposite direction
  - Slack time
- Delays of neighbor trains
- Track properties
  - Average increase/decrease of delay of all trains between two operating points, scheduled one hour around the predicted train
- Previous delay
Linear Regression Models

- Goal: “predict” the delay of a train for a day in the past

- Linear regression models use predictors from the data
  \[ \hat{\delta} = \alpha + \sum_{i}^{n} \beta_i \text{ predictor}_i \]

- Set up a series of in-station and between-station models

- Special predictor previous delay concatenates models
  \[ \hat{\delta}_i = \text{model}(\Theta, d, \pi, \hat{\delta}_{i-1}) \]
Linear Regression Models: Residual Plots

Wauwil – Sursee

Olten
Sequence of Linear Regression Models: Quality of Prediction
COMBINATORIAL MODEL
Shortest Path Model
Shortest Path Model: Tracks and Trapezoids
Minimum Risk Problem

Minimize the predicted delay of the additional train upon arrival at the final station for a

a) single day (P)

b) set of days:
   i. Average of the predicted delays (P)
   ii. Median of the predicted delays (NP-hard)

(using basic prediction models)
Travel Time vs. Risk: Pareto Optimal Solutions

- Time (minutes)
- Risk (minutes)

12.02.2009
H. Flier, T. Graffagnino, M. Nunkesser
Travel Time vs. Risk: Pareto Optimal Solutions

ZF - LZ

12.02.2009
H.Flier, T.Graffagnino, M.Nunkesser
Conclusion and Outlook

- Novel approach to minimize risk of delay
  - Combination of linear regression and a shortest path algorithm yielding Pareto optimal solutions w.r.t. travel time vs. risk of delay
- Profit from existing delay data
- Keep infrastructure modeling efforts to a minimum
- Reasonable risk estimates instead of simulation
  - Detailed simulation only for most promising suggestions

- Outlook and future research:
  - Integrate approach into existing planning tools at SBB
  - Modified risk measure, including delay of follow-up trains
  - Option to remove trains of less priority

12.02.2009

H. Flier, T. Graffagnino, M. Nunkesser
Thank you!
## Linear Regression Models: Achievable Quality

<table>
<thead>
<tr>
<th>From</th>
<th>To</th>
<th>SE</th>
<th>DoF</th>
<th>Min</th>
<th>1Q</th>
<th>2Q</th>
<th>3Q</th>
<th>Max</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZF</td>
<td>BRIT</td>
<td>9.2</td>
<td>11903</td>
<td>-27.8</td>
<td>-5.1</td>
<td>-1.2</td>
<td>3.7</td>
<td>126</td>
<td>0.9931</td>
</tr>
<tr>
<td>BRIT</td>
<td>DAG</td>
<td>12.6</td>
<td>11903</td>
<td>-61.5</td>
<td>-4.8</td>
<td>-1.3</td>
<td>3.5</td>
<td>295.9</td>
<td>0.9874</td>
</tr>
<tr>
<td>DAG</td>
<td>NEB</td>
<td>8.2</td>
<td>11902</td>
<td>-57.5</td>
<td>-3.5</td>
<td>-0.8</td>
<td>2.3</td>
<td>219.2</td>
<td>0.9948</td>
</tr>
<tr>
<td>NEB</td>
<td>WAU</td>
<td>6.2</td>
<td>11902</td>
<td>-33.8</td>
<td>-3.4</td>
<td>-0.5</td>
<td>2.7</td>
<td>177.2</td>
<td>0.9971</td>
</tr>
<tr>
<td>WAU</td>
<td>SS</td>
<td>18.3</td>
<td>11898</td>
<td>-105.1</td>
<td>-8.5</td>
<td>-2.3</td>
<td>5.4</td>
<td>370.5</td>
<td>0.9778</td>
</tr>
<tr>
<td>SS</td>
<td>SEM</td>
<td>27.8</td>
<td>14273</td>
<td>-108.0</td>
<td>-14.6</td>
<td>-2.3</td>
<td>11.4</td>
<td>708.9</td>
<td>0.9403</td>
</tr>
<tr>
<td>SEM</td>
<td>RBG</td>
<td>21.1</td>
<td>14274</td>
<td>-141.3</td>
<td>-7.9</td>
<td>-2.2</td>
<td>3.5</td>
<td>1032</td>
<td>0.9686</td>
</tr>
<tr>
<td>RBG</td>
<td>HUEB</td>
<td>25.2</td>
<td>14274</td>
<td>-102.1</td>
<td>-11.4</td>
<td>-0.8</td>
<td>8.3</td>
<td>382.9</td>
<td>0.9582</td>
</tr>
<tr>
<td>HUEB</td>
<td>EBR</td>
<td>16.3</td>
<td>19964</td>
<td>-58.5</td>
<td>-7.7</td>
<td>-1.7</td>
<td>4.6</td>
<td>663.7</td>
<td>0.9820</td>
</tr>
<tr>
<td>EBR</td>
<td>GTS</td>
<td>49.0</td>
<td>19966</td>
<td>-223.1</td>
<td>-27.4</td>
<td>-5.9</td>
<td>18.9</td>
<td>587.5</td>
<td>0.8342</td>
</tr>
<tr>
<td>GTS</td>
<td>LZ</td>
<td>41.2</td>
<td>37761</td>
<td>-202.4</td>
<td>-22.5</td>
<td>-7.3</td>
<td>14</td>
<td>760.5</td>
<td>0.8769</td>
</tr>
</tbody>
</table>

12.02.2009

H.Flier, T.Graffagnino, M.Nunkesser