ENERGETIC MAP DATA IMPUTATION: A MACHINE LEARNING APPROACH

Master Thesis - Final Presentation
Mandy Nagy

Technical University of Munich
Department of Informatics
Chair of Connected Mobility

Supervisors:
Leonardo Tonetto (Chair of Connected Mobility, TUM)
Tobias Straub, Maxim Sidorov (BMW Group)

Garching, 14.01.2020
CONTENTS

Motivation

Problem Statement & Objectives

Methodology

Results

Conclusions & Outlook
1 MOTIVATION
**ELECTRO-MOBILITY CONTEXT**

Impediments [1,2]:
- range anxiety
- inconvenient charging
- higher price

Ease:
- reliable range assistant
- specific route planning
- proactive driving support

Electric vehicle acceptance level among customers

---


**DRIVING PROFILE**

<table>
<thead>
<tr>
<th>vehicle sensory data</th>
<th>spatio-temporal segmentation</th>
<th>Driving Profile Map Attribute Distributions (DPMADs)</th>
</tr>
</thead>
</table>

\[ E_{\text{Route, an Batteriekleine}} = \sum_{\text{Start}}^{\text{Ziel}} \frac{1}{\eta \text{sgn}(F_{\text{Rad}})} \cdot \sum_{x_n}^{x_{n+1}} \left( \begin{array}{c} \alpha(s) \\ v^2(s) \\ \cos(\alpha(s)) \\ \sin(\alpha(s)) \end{array} \right)^T \left( \begin{array}{c} m \cdot e \\ \rho \cdot c_x \cdot A \\ \frac{2}{f_R \cdot m \cdot g} \\ \frac{m \cdot g}{m} \end{array} \right) ds + \frac{\Delta t_n}{\Delta s_n} \cdot P_{NV,n} \right]_n \]

- **Integral acceleration** \( \frac{m}{s^2 \cdot m} \)
- **Integral squared velocity** \( \left( \frac{km^2}{h^2 \cdot m} \right) \)
- **Average velocity** \( \frac{km}{h} \)

**recuperation**

**propulsion**

- **all weekdays**
- **every 30 min**

50 m 50 m 50 m

PROBLEM STATEMENT & OBJECTIVES
PROBLEM STATEMENT

A time interval of 30 min

- Fleet data set at time $t_n$
PROBLEM STATEMENT

fleet data

fleet data set at time $t_n$

time interval 1 of 30 min
PROBLEM STATEMENT

Problem scenarios:

- Missing sensory data for spatio-temporal buckets
- Lost connection to the backend
PROBLEM STATEMENT

Input

- fleet data

all weekdays
every 30 min

50 m 50 m 50 m
PROBLEM STATEMENT

Input
- fleet data

all weekdays
every 30 min

50 m 50 m 50 m

Output
- 5 DPMADs
- a natural number $n$ of a cluster representative (e.g. cluster #2)

fleets data

$I_S$ $V_r$ $e_c$

\[ ISV_{rec} \]
PROBLEM STATEMENT

Input

?  

all weekdays

every 30 min

50 m 50 m 50 m

Output

- 5 DPMADs
- a natural number n of a cluster representative (e.g. cluster #2)

fleet data / connection

\( ISV_{rec} \)

\%
SOLUTION APPROACH

Input
Map features:
- slope
- curvature
- functional road class
- speed limit

Output
- 5 DPMADs
- a natural number n of a cluster representative (e.g. cluster #2)

Data imputation
supervised Machine Learning (ML) methods:
- Regression
- Classification

Fleet data / connection

all weekdays
every 30 min

50 m 50 m 50 m

\[ ISV_{rec} \]

\[ % \]
3 METHODOLOGY
EXPERIMENTAL SETUP

feature extraction
- slope, curvature: \{min, max, average\}

outlier detection
- Inter Quartile Range (IQR) method
- Q1 - 1.5·IQR < filtered output < Q3 + 1.5·IQR
- reconsider the structure of DPMADs

clustering
- addressed in previous Master Thesis [3]
- robustness of the model: DPMADs based on 3 or more measurements

regression
- Linear Regression algorithm
- main parameters:
  fit-intercept = true, normalise = false

classification
- Decision Tree algorithm
- main parameters:
  max-depth = 10, splitter="best"

Data Set | Traces(km) | Traces (count)
---|---|---
training | 3.503.958 | 95% of Munich
test (Munich) | 556.135 | 5% of Munich
test (Leipzig) | 554.366 | similar to Munich

## EVALUATION SCENARIOS

<table>
<thead>
<tr>
<th>test scenario</th>
<th>Lost Backend Connection</th>
<th>Missing Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leipzig area</td>
<td>regression ↔ classification</td>
<td></td>
</tr>
<tr>
<td>Munich area</td>
<td>regression ↔ classification</td>
<td>regression ↔ classification</td>
</tr>
</tbody>
</table>
4 RESULTS
MACHINE LEARNING PERSPECTIVE

**Evaluation region**

<table>
<thead>
<tr>
<th></th>
<th>Munich area</th>
<th>Leipzig area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integral Acceleration (IA)</td>
<td>Clsf (-1.5%)</td>
<td>Clsf (-2.1%)</td>
</tr>
<tr>
<td>Integral Squared Velocity (ISV)</td>
<td>Clsf (-3.2%)</td>
<td>Clsf (-1.3%)</td>
</tr>
<tr>
<td>Average Velocity (AV)</td>
<td>Regr (-27.2%)</td>
<td>Regr (-28.9%)</td>
</tr>
</tbody>
</table>

- Classification model: perform better in most cases
- Regression model: bigger performance advantages

\{ need problem-specific interpretability \}
Problem scenarios

- Lost connection:
  - the regression model better for both cross validation and application testing sets
  - no relevant overfitting

- Missing data:
  - closest to real world situation
  - regression model outcomes within the values in literature (4-8%) [4,5]

CONCLUSIONS & OUTLOOK
### System Parameter

<table>
<thead>
<tr>
<th>Input Features</th>
<th>Outlook</th>
</tr>
</thead>
<tbody>
<tr>
<td>Include further available map features</td>
<td></td>
</tr>
<tr>
<td>Real time features, e.g. weather, traffic</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feature Engineering</th>
<th>Outlook</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross correlations</td>
<td></td>
</tr>
<tr>
<td>Higher polynomial degree</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Machine Learning Algorithms</th>
<th>Outlook</th>
</tr>
</thead>
<tbody>
<tr>
<td>More sophisticated algorithms, e.g. Neural Networks, Support Vector Machines</td>
<td></td>
</tr>
<tr>
<td>Further parametric optimization</td>
<td></td>
</tr>
</tbody>
</table>

### Conclusions

- Regression model can be deployed in the vehicle given the achieved performances:
  - Lost connection scenario (worst case): 12.6% error
  - Missing data scenario: 7.2% error comparable to related works
- Reliable and precise energy prediction
- Raise BEVs acceptance level
THANK YOU FOR YOUR ATTENTION!