Road Surface Quality using Smartphone

Domain background

My selected project is road surface quality evaluation using smartphone data. There are already several projects in this field, with some promising results. The main goal of these projects is to have a picture of the condition of the road surface, which can help to develop safer and more comfortable traffic conditions. Among these projects there are methods with cameras, specific external hardwares and with smartphones.

For example Perttunen et. al. used smartphone data to classify two types of anomalies [1]. Not only papers, but there are existing applications on the market. One of these the SmartRoadSense, this application uses smartphone user data as well [2].

In order to monitor the quality of the road surfaces, there are three approches [3].

- Treshold-based
- **Machine Learning based**
- Dinamyc Time Wrapping based

Motivation

I live in Hungary, where the road surface quality is one of the worst in Europe [4][5]. While I was on the road, I thought several times that it would be good to plan my trip depending on the road quality. In this case I could better enjoy the beauties of my country.

[1] Perttunen M., Mazhelis O., Cong F., Kauppila M., Leppänen T., Kantola J., Collin J., Pirttikangas S., Haverinen J., Ristaniemi T., et al. International Conference on Ubiquitous Intelligence and Computing. Springer; Berlin, Germany: 2011. Distributed road surface condition monitoring using mobile phones; pp. 64–78.

[2] https://smartroadsense.it/data/map/

[3] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6263868/

[4] https://internet.kozut.hu/kozerdeku-adatok/orszagos-kozuti-adatbank/az-allami-kozuthalozatrol/

[5] http://www.ksh.hu/docs/hun/xstadat/xstadat_eves/i_odu004b.html

Problem statement

There are time series measurements where the time points should be classified into road surface quality classes. Not only the data acquirement, data transformation, feature engineering and model training, but the proper data representation should be solved as well.

Datasets and inputs

The dataset will be my own measurements with a smartphone. I will use the following sensors of the device:

- accelerometer
- gyroscope
- gps

In order to get useful measurement results. I use the phyphox application.

The process of the measurements is really simple. I will drive on known roads which I can easily classify into 3 quality classes. I use my IPhone on the middle console of the car with a ducktape, so the smartphone can record the neccessery parameters relatively reproducible.

The output is a comma separated csv file, where the file name will represent the road surface quality label and in the file each column represent a sensor signal. The maximal sample rate for the measurements will be 50Hz.

Solution statement

My solution will use supervised learning techniques, to classify raw sensor data with 3 labels.

Each label corresponds to a quality class.

Benchmark model

This is a self created project, with own measurements, this is the reason why I cannot use another project as a benchmark. The variety in the hardwares and the boundary conditions (road surface, car, weather, etc.) are not suitable to compere my solution with another project. Considering this, I use my own labeled data as a benchmark from an individual route. My goal is to classify a route with an accuracy of 85%.

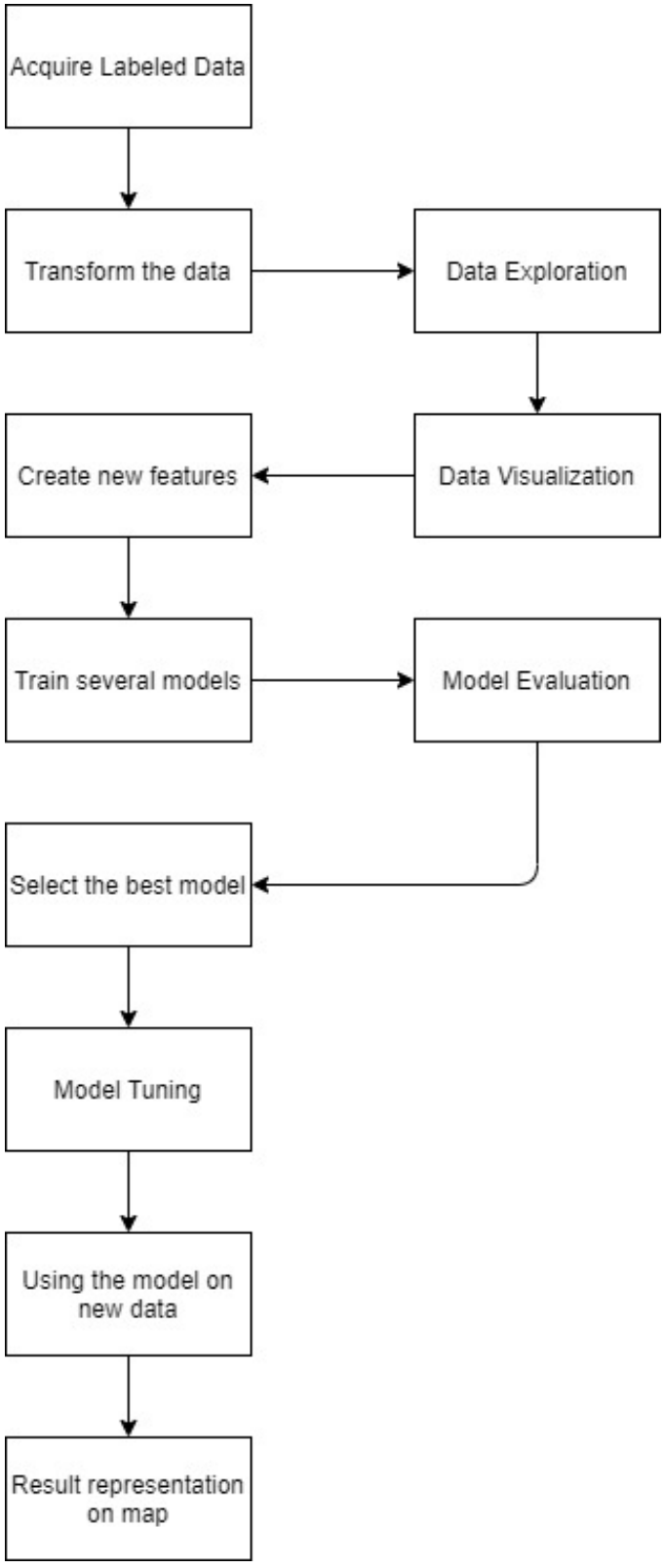
Evaluation metrics

Because of this is a classification problem I can use model accuracy as my metrics to evaluate my model.

Project design

To classify road surface quality into classes, I will use my own data. This means that the first thing what I need to do is to acquire proper labeled data. After that I need to bring this data in a form, which is suitable for machine learning models. I need to explore the sensory data, with data wrangling techniques and data representations. It is important to have the best features which can describe statistically the road surface quality, so I will create new features from the existing signals. To do so, I need to aggregate it, which I can do with time windows. The speed of the car was not stationary, therefore it is considerable to use dynamic window lengths. In the case of using accelerometers and gyroscopes it is a best practice to evaluate the data not only in the time domain, but in the frequency domain as well.

After that I have everything to start training machine learning models. It is important to test several models, then I can select the most suitable for this problem. To establish which is the best model, I will use model accuracy with cross validation. If I will have the most promising model, then I can use hyperparameter tuning, to fine tune my model. After that I have a machine learning model which can classify unknown road surface data. The results should be touchable, so I will represent the classified road surface data in a map.


Acquire Labeled Data
Transform the data
Data Exploration
Create new features
Data Visualization
Train several models
Model Evaluation
Select the best model
Model Tuning
Using the model on new data
Result representation on map