Business understanding

Data mining and visualizations in driving along Rally Estonia SS20 Elva stage

Ehar Kala

Identifying your business goals

Background

We are a team of 3 people and this project is done within a project required by Data Science course in Tartu University. Our initial goals were based on data mining and visualizations in driving recordings of a Bolt project but in the latest mails by Ardi Tampuu we were informed that our goals were changed because of the access difficulties to the Bolt data folder. Our new goal is analysis, preparation and visualization of driving along Rally Estonia SS20 Elva stage.

Business goals

Our business goals are:

- Visualization
  - The route of the car
  - Driving speed (with colors)
  - Driving whiteness (formula provided)
  - Mistakes made by AI (compared to human driving)
  - On a map (optional)
- Analysis
  - Data quality
  - AI performance
  - How clean is the driving
- Preparation
  - Finding and removing unnecessary data (Empty, corrupt etc)
  - Changing data types if needed for the visualization (dummies for example)

Business success criteria

We can consider our project to be successful in case we are able to prepare and analyze data in a way it is comfortable to use it for needed visualization. Criteria for successful visualization is a moving trajectory of all given data about the driving and within provide visualized information about driving speed, whiteness and mistakes made by AI.

Assessing your situation

Inventory of resources

Our team consists of 3 people and Our supervisor is Ardi Tampuu. We currently have one csv file containing approximately 2.3MB of data, information about Estonian coordinates system and multiple sources for downloading maps representing different information. Given csv file contains following columns:

- **index** - timestamp with high precision
- **autonomous** - was this recording driven by human (0) or AI (1)
- **lidar_filename** - not relevant
- **steering_angle** - car's wheels' angle
- **vehicle_speed** – speed (given in m/s)
- **turn_signal** - is the turn signal on
- **position_x** - x-corrdinate in some frame of reference (not yet specified by coordinator)
- **position_y** - y-coordinate other coordinate (so either north-south pos or east-west pos)
- **position_z** - height from sea level (or height compared to the center of frame of reference)
- **roll, pitch, yaw** - rotation of the car along the 3 axis. I'm assuming roll is the orientation with respect to north.

Requirements, assumptions, and constraints

We need recorded csv files of both human and AI driving, formula to determine clean driving, good quality maps of the track and python modules capable of visualizing continuous change in the data. Finished work should provide simple but informative visuals of driving data provided. We currently have access to one csv file of human driving recordings. Only data we still lack is recordings of AI driving.

Risks and contingencies

In case of internet outage, we can keep contact with our team by phone and still continue to work on our data processing and visualizing side of the project.

Terminology

- **ADL** - Autonomous Driving Lab
- **Python** – Programming language used in our language.
- **Driving data** – Csv files of recorded AI and human driving.
- **Supervisor** – Ardi Tampuu
- **AI** – Artificial Intelligence

Visualization of the data – continuously moving visuals of the car driving through-out the track consisting of trajectory, speed, whiteness, height (optional) and AI driving mistakes.

Costs and benefits

The cost of our project could only be measured in work hours. This mentioned, the cost is at least 90 hours of work (30 hours per student as required).

Our project will support analysis of AI driving performance compared to human driving.

Defining your data-mining goals

Data-mining goals

- Initial presentation consisting of business understanding, data understanding and the planning of the project.
- Prepared and cleaned csv files
- Initial model of our project
- Finding a way to determine differences in human and AI driving
- Recognize AI driving mistakes
- Tested and improved version of our project model

Data-mining success criteria

Data mining is successful if all eligible data can be visualized in a continuously moving picture and our predictive techniques are at least 95% accurate.

Data understanding

Data mining and visualizations in driving along Rally Estonia SS20 Elva stage

Kristjan Paavel
Henri Tein

D-13 2021

Gathering data

Outline data requirements

Types of data necessary:

- Csv files on nonautonomous vehicle
- Csv files on autonomous vehicle
- Python module files
- Mathematical formulas (Driving whiteness formula e.g.)
- Module usage instructions

In order to understand the position of the vehicle we need access to map where we can convert x, y and z positions into location on map, to better visualize data.

The time range of processing the data is unmeasurable for now, since we don't have all the data that is required for this project.

Verify data availability

We have looked at the csv file of nonautonomous vehicle driving data and made sure that we have access to it.

Also, we have found a way to use coordinates in website that is called "geoportaal.maaamet.ee".

Python modules that are required for example are turtle, pandas, NumPy, most of these we already have, since we have used them before or during this course.

The instructions for these modules can be found by googling those modules instructions.

At the moment we don't have data on AI driving, but we should be getting it in the future.

Driving whiteness formula will be provided to us by our project supervisor.

Define selection criteria

In the csv file there are many columns that we will use, such as index, vehicle_speed, roll, pitch, yew, position x, y and z.

The use of these columns will be to visualize movement of vehicle.

For the time being use of other columns is not necessary unless we are given specific tasks that require the use of these columns.

Python modules mentioned above were select to visualize the movement of vehicle.

Describing data

The csv data we currently have was given to us by our project supervisor.

Data shows what are vehicles x, y and z positions with timestamps, also it shows vehicles current speed.

For the moment we only have information of one recording of driving.

Exploring data

Our current csv data is recorded in 19 minutes.

We can see that this data is only for a man-driven car, because the autonomous column shows false.

The speed of vehicle varies from 0-18 m/s (0-65km/h).

The x, y coordinates define horisontal position of the vehicle.

The z coordinate defines the height of the car.

We were given 2 constants for the x and y coordinates from csv file. Constant for x = +650000 and for y=+6465000. They were subtracted in the first place to make the .csv more compact.

Turn_signal column shows if the driver is using a turning signal or not. If the turn_signal = 1 the signal is off. Otherwise (0 or 2), the turning signal is on.

Roll - Rotation around the front-to-back axis

Pitch - Rotation around the side-to-side axis

yaw - Rotation around the vertical axis

Verifying data quality

We don’t have all the data that is required for our project since it hasn’t been provided to us by our project supervisor.

But we have data on one recording of the vehicle that is good enough to support our goals.

Planning

Data mining and visualizations in driving along Rally Estonia SS20 Elva stage

Kristjan Paavel

Our project is mainly distributed into 6 tasks, 2 for each team member. The total time that has to be contributed on this project must be at least 90 hours, it means 30 hours for each team member.

Gathering and analyzing - 31 hours total.

- **Business understanding – 10 hours** Ehar Kala
- **Data understanding – 10 hours** Henri Tein
- **Planning – 6 hours** Kristjan Paavel
- **Analyzing the data quality – 5 hours** Kristjan Paavel

Preparation and modeling - 12 hours total.

- **Preparing the data for usage - we need to find and remove all the unnecessary data (Empty, corrupt etc) and change the data types (dummies for example) if needed for the visualization - 6 hours** Ehar Kala
- **Modeling - 6 hours (2 hours per member)**

Coding and building - 58 hours total.

- **Coding initial model (evaluation needed) - 12 hours for each member (36h total)**
- **Testing – 4 hours** Henri Tein
- **Visualizing the graphs (route, speed, height, driving whiteness) - 8 hours** Kristjan Paavel
- **Analyzing the AI performance and driving - 5 hours** Henri Tein
- **Visualizing AI mistakes – 5 hours** Ehar Kala

Final improvements - 21 hours total.

- **Evaluating results - 1 hour** (Team meeting)
- **Code cleaning and improving usability – 1 hour** (Team meeting)
- **Creating a PDF and mp4 (if needed) for our presentation - 15 hours total (5 per member)**

Total – 122 hours.

Repository - https://github.com/EharK/Project-D13