Research Proposal Group 2, Mode Choice, ME44312 - Machine Learning for Transport and Multi-Machine Systems

Pans, Gaetan[5858372], Sajith, Navneet[5432804], Sasidharan, Siddharth[5788919], van Maanen, Koen[4855086], and Tournoy, Aster[4862031]

Delft University of Technology, Delft, The Netherlands

1 Data Set and Motivation

This research will focus on the data set: mode choice behaviour in Switzerland. This data set is chosen, because it is very extensive. With more than 50 variables and more than 50 attitude questions, this data set gives many different options for research. Furthermore people's choice behaviour of transportation modes is an interesting topic the team would like to research. By examining this data set, it is possible to gain deeper understanding of people's decisions in transportation and using this knowledge more effective and sustainable transportation systems can be designed for the future.

2 Research Question

Using the data set of Vincent Kaufmann, Michel Bierlaire and Martin Schuler regarding the preferred mode choice in Switzerland, it is possible to predict the transport mode of people living in different areas in Switzerland [1]. Based on that data, it can be analysed what possible investments in transport are the most beneficiary for the Swiss government for both environmental and economical concerns. These investments can for example be in public transport or in a better infrastructure for electric vehicles or in better infrastructure for cyclists, etc. To ensure which investments are the most valuable, the following questions that are listed below are to be answered.

- Are different mode choices preferred in different areas of the country of Switzerland?
- Are different mode choices preferred in different wealth classes and does that correspond to different areas of Switzerland?
- Is there a difference in preferred mode choice based on environmental concerns and does that correspond to different areas of Switzerland?
- Does the amount of kids have an effect on the preferred mode choice and does that correlate to different areas of Switzerland?

After answering these sub-questions, the final research question can be answered with valid data. This research question is as follows: "What is the best investment for the Swiss government, for environmental and economical impact, based on the different areas in the country of Switzerland?"

3 Plan to Answer Research Question

To answer the research question, five sub-questions need to be answered, as explained in Section 2.

For each of the questions, the following machine learning methodologies will be used:

- Question 1: Clustering to determine type of transport user
- Question 2: Clustering to determine whether a person is wealthy
- Question 3: Clustering by logistic regression to determine environmental concern
- Question 4: Classification by neural networks

4 Type of Experiments to test and evaluate

Question 1: A 3D plot will be generated, where each axis represents the frequency at which the user uses each type of transport. Clustering will then be used to group each user into a single mode of transport based on their preferences. This will then be repeated for both rural and urban citizens to then see if there are certain modes of transport that are preferred in the respective 2 areas.

Question 2: Perform an extra clustering to figure out if a person is wealthy based on the amount of kids, income, the number of TVs etc.

Question 4: Initially, the qualitative questions of each user regarding their care for the environment will be converted into a single quantitative value. Using this number, a logistic regression will be used to classify each user to whether they care for the environment or don't. From here it can then be seen whether there is a correlation between concern for the environment and mode of transport.

Question 5: From the data provided, a plot will be generated of number of kids against income. Each point (representing a user) will be identified to use a type of transport (from the clustering in question 1) which will be represented by different markers on the plot. From here, classification will be used to determine the boundaries of the classes.

Overall Approach: From the plots and relationships determined by the individual questions, a better understanding can be made on the factors that have a role into the mode of transport used. This will then be used to answer the final research question.

References

[1] Bierlaire, M., Curchod, A., Danalet, A., Doyen, E., Faure, P., Glerum, A., Kaufmann, V., Tabaka, K., and Schuler, M. (2011). Projet de recherche sur la mobilité combinée, rapport définitif de l'enquête de préférences révélées. page 232.