Modeling travel behavior using socio-demographic characteristics and statistical matching methods

Tim Hilgert, Martin Kagerbauer, Christine Weiss, and Peter Vortisch

Karlsruhe Institute of Technology

Abstract:

Travel behavior data are useful for understanding current mobility needs and transport demands. Surveying these data is costly and time-consuming thus data sources are often not available. Nevertheless, these data are required to investigate recent research questions in transport planning, transport economic and transport policy. Alternative methods generating travel behavior data are helpful to reduce costs and fulfil needs. In this paper, we present a modeling approach to reproduce travel behavior data synthetically using socio-demographic and commuting similarities.

In our investigated use case we had socio-demographic and commuting characteristics of a sample available only and wanted to enrich the knowledge about their travel behavior. Therefore we developed a modeling approach that generates travel behavior data synthetically. The model is based on the assumption that travel behavior information can be predicted by using statistical similarities of peoples' socio demographic and commuting characteristics only.

Besides the sample data of the investigated use case, the modeling framework used data from the German Mobility Panel (MOP). This national household panel survey is carried out every year and collects data about daily travel behavior representative for the German population over the course of one week.
In this model we estimated travel behavior by the idea of statistical matching methods. These methods are used in different research areas; e.g. in market research. In this field often limited information is available for a subset of people while disjunctive information is available for another subset only. Statistical matching methods link these subsets and help to enhance relevant information for the entire sample. The principle of linking subsets bases on common variables that exist in all different subsets. These common variables are used to match the elements of different datasets by finding statistical twins. A statistical twin of a person is defined as being similar to the person in at least some of the common variables.
In our modeling approach we used the database of the MOP and the sample data derived from the use case we investigated. For this second database, we only knew socio-demographic and commuting characteristics of people but no travel behavior information. The socio-demographic and commuting characteristics available in both databases were taken as common variables. For the matching process, we used different of these common variables such as age group, gender etc. We determined statistical twins for a person by selecting those people from the MOP sample who are identical in selected variables. Finally, the determined statistical twins were used to synthetically reproduce the travel behavior.

The mayor task for the model calibration was the selection of the “best-fitting” variables’ combination to create a matching as best as possible. The problem consisted in finding the variables that created the sets of statistical twins which fitted the travel behavior best. In total we chose the variables’ combination of 12 common variables. This led to 2^12 = 4,096 possible combinations of variables for the matching process. For example, a matching using the gender variable only is possible. This results in a high number of statistical twins found for every person but also in an insufficient quality of the modeled behavior. In contrast taking all 12 variables for the matching probably no statistical twins are found to reproduce travel behavior.

To find the “best-fitting” variables’ combination we calibrated the model by using data of the MOP only. Thus we wanted to reproduce the reported travel behavior by using the statistical matching approach. An algorithm was developed which iteratively performed the matching for any possible combination of variables. We applied this algorithm for an artificial subsample of 500 people randomly drawn from the MOP. This gave us the opportunity of evaluating the quality of a variables’ combination. Therefore, we compared the daily travel behavior of the statistical twins to the reported travel behavior of every person drawn out from the MOP sample. We defined the quality of a variables’ combination mainly by comparing the travel behavior in terms of trips made and distance traveled per week. Comparing these two values we tried to minimize the mean deviation of these values. Therefore, we performed the matching for all possible variables’ combinations and finally selected the combination with the minimal mean deviation. The “best-fitting” combination of variables is gender, employment status, commuting distance, commuting time, number of persons in the household and number of cars in the household.

To evaluate the calibration results the modal split of the randomly drawn MOP participants subsample is calculated. We compare these values to the average modal split of the determined statistical twins who represent the modeled travel behavior. For the 500 MOP participants and the final variables’ combination modeled data (statistical twins) contain 25.9 trips per person per week compared with 25.8 trips per person per week (randomly drawn MOP participants). This results in a deviation of about 0.5%. Concerning the distance traveled per person per week, the deviation is about 9% – 317 km (statistical twins) and 345 km (randomly drawn MOP participants).

The model presented in this paper was developed and used in the CROME (cross-border mobility for electric vehicles) research project. In this project, participants of the field study within the project used electric vehicles for business trips. Information about the travel behavior in general was needed to evaluate the potential of using electric vehicles in their daily travel behavior. We assessed the potential of substituting trips by electric vehicles in the everyday travel using the statistical matching approach described.

Our approach shows that reproducing travel behavior is possible using statistical matching methods, limited socio-demographic and commuting characteristics. Our results indicate that the modeling of average modal split for a group of people has only a small deviation and the method used is suitable for application. Statistical matching methods can provide certain information about the travel behavior of a group of people without carrying out elaborate travel surveys. The application of statistical matching methods in travel behavior research is an alternative to travel diary surveys when an inexpensively alternative is required.