報告番号	甲 第 11467 号

主 論 文 の 要 旨

論文題目 **Activity-Travel Pattern Analysis Based on Mobile Phone GPS Data**
(携帯電話の GPS 位置情報を活用した交通行動分析)

氏 名 宮 磊

論 文 内 容 の 要 旨

There is consensus that traditional person trip survey (PT survey) cannot avoid the disadvantages such as underreported trips, inaccuracies in times, surrogate reporting and sometimes confusion of appropriate trip purpose. The occurrence of Global Positioning System (GPS) technology, especially the popularity of smart phone with GPS censor, makes it possible to collect the PT data to overcome the disadvantages of traditional PT survey. However, the raw GPS data is a series of temporal trajectory points with coordinates (at least longitude and latitude). So efficient techniques and methodologies are needed to extract the PT data with a high accuracy. The objective of this thesis is to propose a series of methodologies to extract PT data, such as activity type (same meaning as "trip purpose" in this thesis), from continuous GPS trajectories and analyze the activity-travel pattern based on the extracted PT data.

The first step of extracting PT data from continuous GPS trajectories is to segment GPS points into separate trip and the activity engaged in at the trip end. A density-based algorithm is developed to distinguish the GPS points into moving points and stop points. It is an improved version of DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm. Two constraints from temporal and spatial are added as improvements to the original algorithm. Then a machine learning method, Support Vector Machine (SVM) is applied to distinguish the activity stop from all the identified stops.

After obtaining the distinguished activity stops, the next step is to identify the specific type of activity. Although some heuristic-rule-method has been applied to this field before, it is machine learning methods that avoid resetting the rules manually when data set changes, because machine learning methods can finish the task by automatically learning the inherent relationship between dependent variable and independent variables with efficiency. In the field of identifying activity types, it is still unknown which machine learning method has the best performance. That is why several machine learning methods are tested and compared in this research with the same dataset. It is found that classification trees outperforms SVM, Neural Networks (NN), and Discriminant Analysis (DA) from the perspective of time cost and accuracy.

Improving the accuracy of activity type identification are very necessary for the follow-up research, such as activity-travel pattern analysis and activity-based models. Due to this concern, several techniques of improving the accuracy of activity type identification are discussed. These techniques include data selection for training set and test set from seasonal data, and inclusion of effective additional variables.

One advantage of PT data obtained from GPS trajectories is gathering the multi-day information. Due to this, analyzing what kind of factors influence the activity-travel pattern during a long period become possible. Ordered logit models are used to estimate the impact of variables from time dimension, trip dimension and weather on trips and trip chains from a prospective of temporal heterogeneity. The number of trips in a trip chain, the number of trip chains in a day, and the number of trips are used as dependent variables individually in the analysis.

Since the accuracy of activity type identification will influence the analysis in the next step analysis, such as activity-travel pattern analysis and activity-based models. Until the activity type identification reaches 100%, it is necessary to consider the activity type identification accuracy in the follow-up study. An activity sequence generation model with consideration of accuracy of activity type identification is developed. The model can be decomposed as a series of discrete choice model in sequence and the activity identification accuracy can be converted as part of the constant value in the function. The results show the prediction of activity frequency forecasted by the activities identified with error and forecasted by the real activities are not significantly different from each other.