P. Koomsap et al. (Eds.)

doi:10.3233/ATDE230680

A Simulation of Human Mobility That Reproduces the Behavioral Characteristics

Yuri MIZUNO[1], Daishi SAGAWA, Yuya KIMURA and Kenji TANAKA
School of Engineering, The University of Tokyo, Japan

Abstract. As interest in smart cities grows, foot traffic data is expected to solve diverse problems cities have. However, it has a fragmented nature in time and space, and this incompleteness obstructs its utilization. Therefore, simulation of human mobility is necessary not only for measuring policy effectiveness but also for identifying urban realities that are not observed directly from foot traffic data. Although previous research has proposed various simulations of human mobility, they are inadequate in that they do not take behavioral characteristics into account. This paper presents a behavioral characteristics classification method using foot traffic data and a method for designing a human mobility forecasting simulator that implements a decision-making model to complement the incompleteness of the foot traffic data through a transdisciplinary approach. The proposed method is divided into two phases: classification of behavioral characteristics and designing a human mobility forecasting simulator. We conducted a case study using data from an actual urban city. As a result of the classification, differences were observed in the trends of facilities used and time spent in each cluster. The constructed simulator used Bayesian optimization to adjust parameters, and the results showed a significant correlation between simulation results and real-world data. Two simulations were conducted to measure the effectiveness of adding new stores and operating digital signage. Both simulations identified changes in human mobility for each scenario. Counterintuitive results were also observed, such as a decreased total number of visitors to each building.

Keywords. Foot traffic data, smart cities, simulation of human mobility, classification of behavioral characteristics

Introduction

In recent years, the smart city market has been expanding and data utilization to solve issues related to Sustainable Development Goals (SDGs) is proposed in many studies. For instance, Aziz et al. proposed a system to monitor a person's health condition in real-time using Google Mobile Service (GMS) and Global Positioning System (GPS) technologies [1]. Catlett et al. contributed to solving an issue of community safety with a crime prediction method [2].

Foot traffic data is one of the promising tools to be applied to solving regional issues. A simulator of human mobility provides a way for evaluating the effectiveness of multiple measures to vitalize a city without implementing them in the real world. If a measure of the optimal advertisement or the optimal store layout is identified through simulation, city developers could add value to their city without a large redevelopment project.

[1] Corresponding Author, Mail: kmaru-eeen@g.ecc.u-tokyo.ac.jp.

However, foot traffic data are fragmented temporally and spatially due to the way they are accumulated and their format. Such incompleteness of foot traffic data may be an obstacle to the utilization of such data in smart city initiatives.

1. Previous studies

As a method to supplement the incompleteness of traffic data, many studies have applied data assimilation techniques developed in the field of geophysics. Zhou et al. proposed a method to predict walkers' trajectories in narrow areas from video data using the Kalman filter which is one of the data assimilation techniques [3]. Barrios et al. conducted a prediction of vehicles' trajectory with the Kalman filter [4]. However, it is difficult to obtain wide-area human movement trajectories from foot traffic data with the existing techniques.

In the logistics field, the model based on Wardrop's principle is often used to determine the equilibrium state of traffic flow and guide traffic management and planning decisions. This approach estimates the costs or tradeoffs that individuals make in transportation networks. Rieser expanded an agent-based traffic simulation by incorporating a behavioral model that made agents choose transportation other than cars [5]. Bergkvist et al. proposed a multi-agent simulation (MAS) to simulate the activities of production, storage, and transportation in a supply chain and the decision-making process of the parties involved in each activity [6]. These approaches have the potential to be applied to human mobility forecasting by implementing the decision-making model of humans to quantify the measure of their utilities and costs.

There are many studies on human behavior characteristics utilizing a variety of data. For instance, Chen et al. proposed a method to find out changes in customers' behavior using purchase data from retail stores and e-commerce sites [7]. There are also some studies using trajectory data. Kharrat et al. proposed a clustering algorithm using a vehicle's or a person's trajectory data to group similar trajectories [8]. Li et al. used GPS data in the real world to mine user similarity of trajectories [9]. These approaches to applying trajectory data to find out tendencies in movement path can be expanded by taking features that facilities in a city have into consideration because they are strongly related to a person's behavior. It means that trajectory data might be used to find out the tendency of human behavior in a certain city. The analysis of human behavior characteristics based on foot traffic data is not only useful for the revitalization of the entire city and measures to induce human mobility but also for reproducing human behavior more precisely in a simulation.

In this study, we propose a method for designing a human mobility forecasting simulator that implements a behavioral characteristics classification method using foot traffic data and a decision-making model that complements the incompleteness of the foot traffic data.

2. Proposed method

2.1. Overview of the proposed method

In this study, we propose a method to construct a simulator of human mobility with historical foot traffic data. The method is divided into two phases. In the first phase, the classification of behavioral characteristics of people in a city is conducted with historical foot traffic data accumulated in an actual city. People classified into the same class are regarded as having similar behavior, and the classes made in this phase are used as an attribution of humans in the next phase.

Then, in the second phase, an agent-based simulator is constructed. An agent in the simulator represents a person existing in the city, and it moves around the city, making decisions about its next action. The decision-making model in this simulation is based on attributes assigned to each agent, such as gender, age, class of behavior tendency, and Indicators such as hunger and fatigue levels, etc. Multiple agents are generated at every time step, and they move around the city and leave there after their purpose is achieved. By simulating the behavior of each agent, it is possible to simulate human mobility in an entire city.

The proposed method is novel in two points. First, it combines a decision-making model that incorporates human behavioral tendencies with actual human flow data, allowing for the reproduction of human movement in a spatially and temporally continuous manner. This enables not only macro-level human flow predictions but also detailed analysis.

Secondly, in addition to factors such as hunger levels and behavioral tendencies based on gender and age, the proposed method incorporates the behavioral tendencies of each cluster identified through clustering analysis using actual human flow data into the decision-making model. This results in a more precise reproduction of human behavior, allowing for the replication of specific behavioral tendencies unique to the users of a particular area due to the features and layout of buildings in that area.

2.2. Classification of behavioral characteristics of urban users

The flow of the behavioral characteristics' classification method is shown in Fig. 1. First, preprocessing is performed using urban foot traffic data as input, and if anomalous data are found by analyzing the clustering results, the preprocessing method is improved.

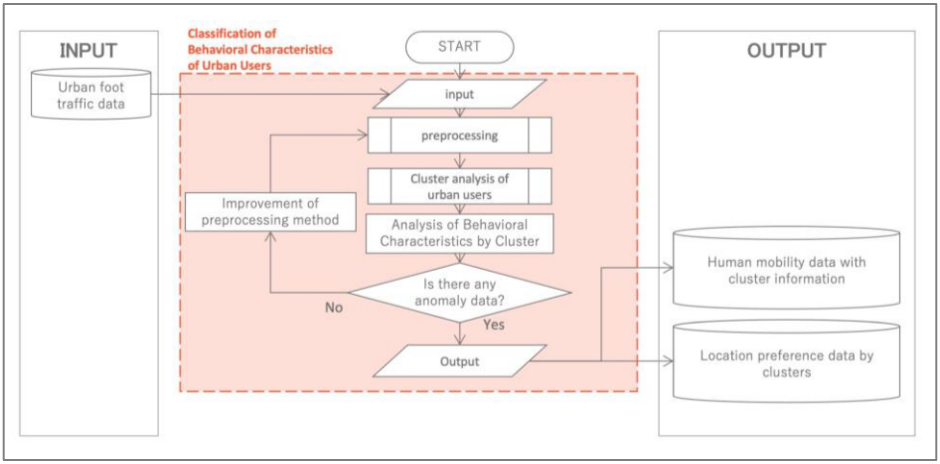

Figure 1. Flowchart for classification of behavioral characteristics based on foot traffic data.

2.2.1. Data and preprocessing

The foot traffic data to be processed should include (1) information that can identify individual users; (2) information that can be linked to city map information; and (3) an index to evaluate the accuracy of the user's location at a certain time.

There are four main preprocessing tasks to be performed. The first is to filter the data of users who cannot be identified. The second is to link the data with the city map data that contains information on the location of facilities and aisles in the target city. The third is the elimination of highly uncertain data. The fourth is the elimination of anomalous data. If anomalous data are found during the analysis of the clustering results, it is necessary to improve the fourth preprocessing step sequentially. After the above four preprocessing steps, the input data for clustering is prepared.

2.2.2. Cluster analysis of city users

In this part, we compare the results of two clustering methods: non-hierarchical cluster analysis by k-means++ and hierarchical cluster analysis by Ward's method. In the former, the number of clusters is determined by the elbow method. The two methods are evaluated by a quantitative comparison, in which the variation of the number of users belonging to a cluster is evaluated in terms of standard deviation.

2.2.3. Analysis of Behavioral Characteristics by Cluster and Identification of Anomalous Data

To verify that there are differences in behavioral characteristics among the clusters, we analyze the tendency of each facility by cluster. If there is a difference in the tendency of the facilities used by each cluster, an analysis of the tendency of the time spent at each facility is conducted. The two main purposes of the above two analyses are to obtain knowledge on the characteristics of behavioral trends by cluster and to generate input data for the simulator to be constructed in the next phase. Finally, to analyze the characteristics of behavioral tendencies within each cluster in detail, we analyze the behavioral flow of representative users belonging to each cluster.

If any anomalous data are identified during the analysis of the output results, the preprocessing method is improved, and the cluster analysis is performed again.

2.3. Construction of a simulator for predicting human mobility

In this study, a multi-agent simulation (MAS) is used for human mobility forecasting. The simulator consists of four types of components: Agent, Store, Building, and Departure point. Agent represents humans in the virtual city space.

Fig. 2 shows a flowchart of the simulator construction. First, virtual city space is designed by inputting city map data, Store data, Visitor data, and configuration parameters. Then, the components of Store, Building, and Departure are created from the input data, mainly city map data and store data, to reproduce an actual city. Before generating Agents in a virtual city space, a behavioral model to decide how each Agent behaves in the virtual space including a decision-making model is implemented. As the simulator starts to run, Agents are generated every time step from Departure points and move around the virtual city space. Each Agent is created with attributes such as gender, age, and class of behavioral tendency assigned based on statistics from the input data, mainly visitor data. Each agent behaves based on the behavioral model that has both of common rule and dependency on attributions. The actual simulation is performed to verify the consistency with the real data, and the parameters are adjusted by Bayesian optimization. The output is the attribute information and behavioral history of each Agent. Agent mobility in the virtual city space is observed with these data.

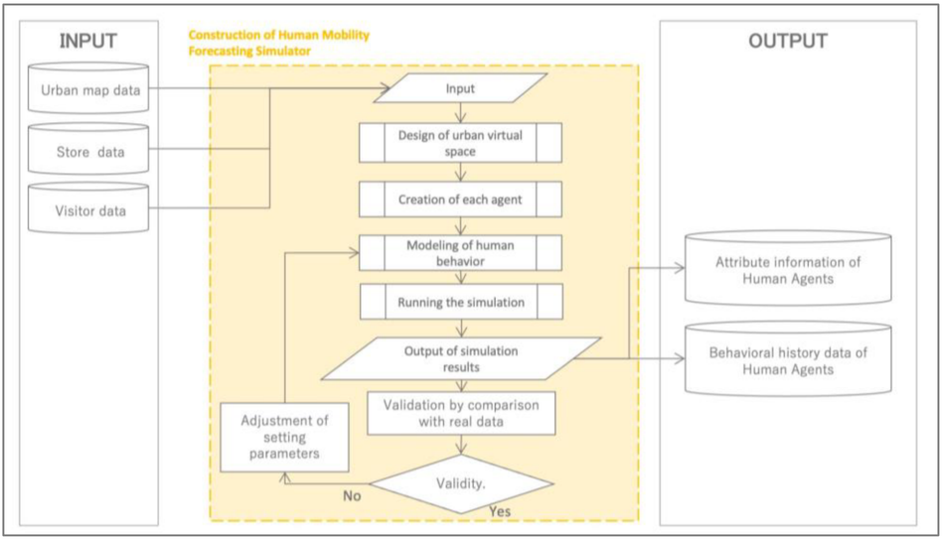

Figure 2. Flowchart for constructing a human mobility simulator.

The decision-making model implemented in human behavior modeling for determining and updating destinations and making stopover decisions follows the principle of minimizing loss and maximizing utility, referring to the model based on Wardrop's principle often used in travel route selection in the field of logistics. The evaluation function used is shown below.

$$U_{i,j} = \frac{\alpha_i CP_{i,j} + \alpha_i GP_{i,j} + \alpha_i AP_{i,j} + \alpha_i HS_{i,j} + \alpha_i IS_{i,j}}{5} - \frac{\alpha_i CC_{i,j} + \alpha_i VC_{i,j} + \alpha_i MC_{i,j}}{3} \quad (1)$$

$U_{i,j}$: Utility of store j by agent i

$CP_{i,j}$, $GP_{i,j}$, $AP_{i,j}$: Preference for store j dependent on agent i's cluster, gender, and age

$HS_{i,j}$: Hunger score obtained if store j was a restaurant according to the hunger level of agent i.

$IS_{i,j}$: Information score for store j held by agent i

$CC_{i,j}$: Cost due to the amount of money that agent i spends in store j

$VC_{i,j}$: Revisit cost if agent i has already visited store j or the building where store j is located

$MC_{i,j}$: Travel costs incurred by agent i going to store j

α_i: The weight that agent i has for each term

Since the evaluation function has a simple structure consisting of five Benefit terms and three Cost terms, terms can be added to each of them as needed. To find optimal parameters that cannot be estimated from real data, we use the Bayesian optimization approach due to its efficiency and accuracy. This approach is widely used in the machine learning fields and applied in MAS. For instance, a method proposed by Kiyotake to estimate OD information (departure time, place of Origin, and Destination) by using Bayesian optimization to estimate parameters in MAS realized efficient and highly accurate estimations [10].

If the simulator is validated drawing a comparison between the simulation result and real data under the same conditions, we will use it to simulate scenarios corresponding to measures to solve urban problems such as city decline and measure the effectiveness of the measures.

3. Case study and simulation

A case study was conducted using data on human mobility in Ebina City of Japan collected by beacon devices that detect the signals of smartphones of city users.

3.1. Classification of behavioral characteristics of city users

After confirming that the data used met the conditions, four preprocessing steps were performed. As for the method of excluding anomalous data, the method of discriminating by switching counts among beacon devices to identify the devices that no one user carries around, such as the PC in a store, left many anomalous data. Therefore, we improved the preprocessing method by implementing the switching count as the name of the building where the beacon devices are installed. The preprocessed data was divided into weekdays and weekends and converted into matrices that represent how many times each city user was detected by each beacon device. And then, unnecessary columns were removed to eliminate sparsity, and dimensionality was compressed using a Non-negative Matrix (NMF). NMF is an effective method to deal with matrices with sparse properties, for instance, this approach was applied in research of clustering documents [11].

As a result of the clustering, we were able to classify users of Ebina City according to their behavioral characteristics. Since we confirmed that there were differences in the tendency of facilities used and the time spent in each cluster, we created input data including the information about the tendency and population of each cluster for the simulator constructed in the next phase. Finally, the behavioral flow analysis of the representative users belonging to each cluster was conducted to confirm the differences in behavioral tendencies among subgroups within the same cluster and the similarities among the subgroups.

3.2. Construction of human mobility forecasting simulator

The simulator was constructed according to the flowchart shown in Fig. 2. 141 Stores were created in ten Buildings, and a station was set up as a Departure point. Fig. 3 shows the virtual city space of Ebina.

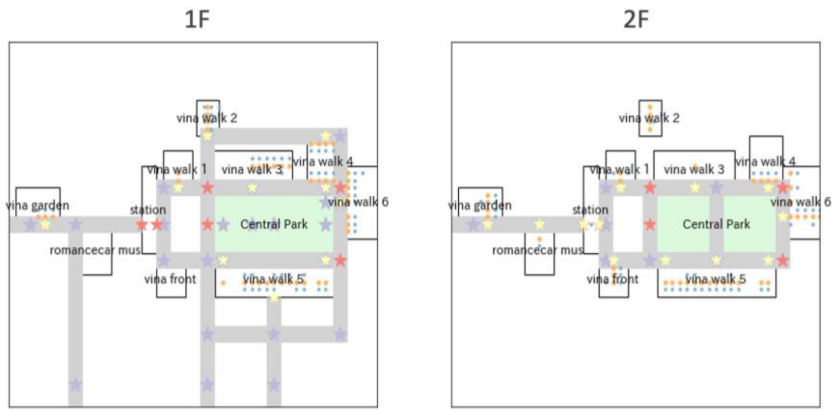

Figure 3. Image of Ebina City simulation.

In the modeling of human behavior, the two processes of stay time and hunger calculation were designed based on the characteristics of the input data of this case study.

After adjusting the parameters by Bayesian optimization, we conducted a validation test of the simulator we constructed. As shown in Table 1, the correlation between the simulation results and the real data is significant for both weekend and weekday data.

Table 1. Results of the uncorrelated test.

		Correlation coefficient	P-value	Result
Weekend	Total number of visitors by building	0.783	0.0126	Significant correlation.
Weekend	Average time of stay by building	0.820	0.0067	Significant correlation.
Weekday	Total number of visitors by building	0.740	0.0225	Significant correlation.
Weekday	Average time of stay by building	0.889	0.0014	Significant correlation.

3.3. Simulation Results

Fig. 4 shows an example of the movement trajectory of a certain Agent generated in the case study under the setting of a Sunny day as a weather parameter on a weekend as a day of week parameter. In this figure, the Agent's footprint is represented by a red point. The Agent is assigned the attributes shown in Table 2 and departs from a station (1). Upon Departure point, the Agent determines its destination using the evaluation function described above and moves toward a store on the first floor of VINAWALK Building 1 which is a commercial facility and one of ten buildings in virtual Ebina City (2). Then, according to the stay time calculation described above, the time spent at the store is determined to be 17 minutes, and after the stay, she started to move again to a retail store on the second floor of VINAWALK Building 5 by updating the destination using the evaluation function (3). While staying at the store, the hunger level of the Agent changes from -6.87 at the time of the arriving at the store to 6.68 at the time of leaving there, as shown in Table 3. It is considered that a change in the Agent's internal state such as hunger level change may be the reason why the Agent chose a restaurant on the same floor as the next destination when updating the destination when she was about to leave a retail store (4). Then, when she left the restaurant, she also updated her destinations and chose to return to the station because there was no store with a positive utility anymore (5). By generating Agents that perform such a series of actions from the station every time step, we reproduced the flow of human mobility in the virtual city space.

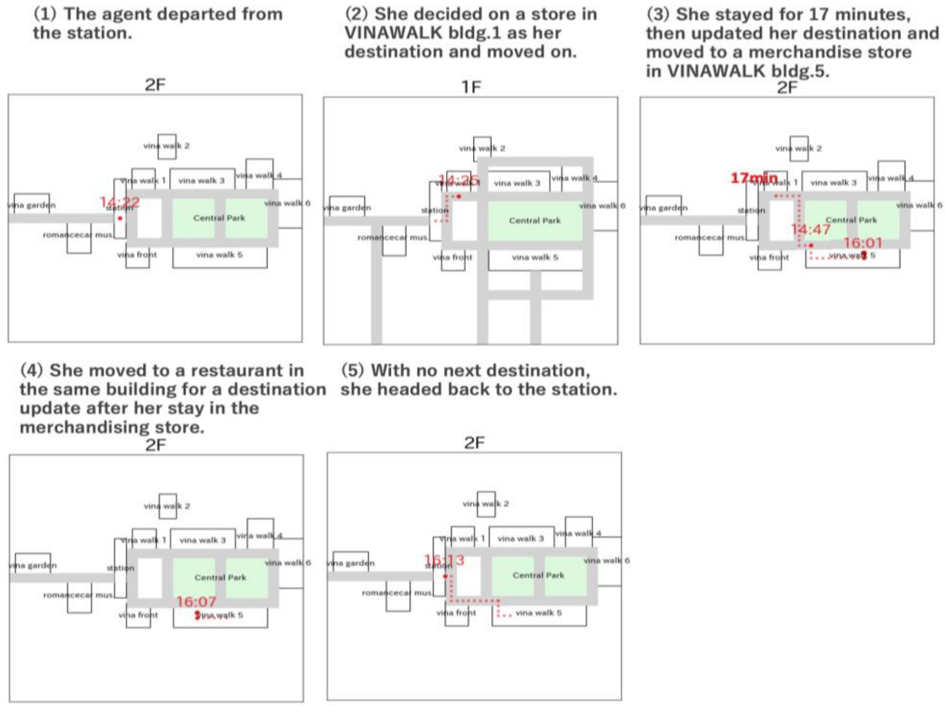

Figure 4. Example of a certain Agent movement trajectory.

Table 2. Results of the uncorrelated test.

id	1831
cluster	9
gender	F
age	27
Initial fatigue level	1.00
Fatigue rate of change	0.0405
Initial hunger level	-12.22
Hunger rate of change	0.048

Table 3. Hunger level at the time of arrival and at the time of leaving the store.

Timestamp	Hunger Level
14:50:30	-6.87
…	…
16:01:30	6.88

Sensitivity analysis was conducted for the four setting parameters: weather parameter, weekday/weekend parameter, gender ratio, and age distribution, and it was confirmed that all of them have an impact on the results. For example, as shown in Fig. 5, the number of visitors to each building decreases when it rains, as indicated by the change in the flow of people when the weather parameter is changed. This result confirms that the decision-making modeling in human behavior modeling works effectively.

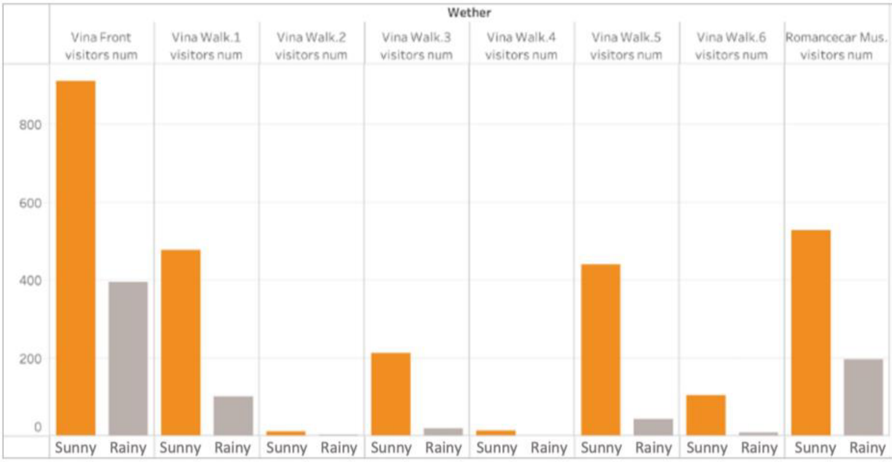

Figure 5. Results of sensitivity analysis of weather parameters.

In addition, for the purpose of measuring the effectiveness of measures that are expected to be applied to solving various urban problems, two simulations were conducted: (1) the change in human mobility when a store is added to a vacant tenant, and (2) the change in human mobility when a signage advertisement is put into operation.

Not only were changes in human mobility confirmed for each scenario in both cases but also counterintuitive results were obtained for several scenarios, where the addition of stores and signage operation reduced the total number of visitors by building.

4. Conclusion

In this study, we proposed a human mobility simulator design that implements a behavioral characteristics classification method using foot traffic data and a decision-making model which the model based on Wardrop’s first principle used in lots of logistics studies was applied to. The application of the classification method is useful to understand features of human behavior unique to each city and reproduce human behavior more precisely. The decision-making model contributes to solving the problem of spatial and temporal incompleteness in foot traffic data in its utilization which has been an obstacle despite the background of the growing interest in traffic data as the smart city market expands. By solving the problem of spatial and temporal incompleteness in foot traffic data, the proposed method in this study enables us to analyze the result of human mobility simulation in a certain city not only from the macro perspective but also from the micro perspective. Through a case study, the effectiveness of the proposed method was demonstrated, and new knowledge that is expected to be utilized in solving urban problems could be obtained through simulation and analysis.

Our study is helpful for people who devise measures for the vitalization of their city and who conduct studies that human behavior or decision-making could be applied to.

Acknowledgment

We are grateful to Odakyu Electric Railway Co., Ltd., and Softbank Corp. for providing valuable data. This research was carried out with the support of the Institute for Beyond AI of the University of Tokyo.

References

[1] K. Aziz, S. Tarapiah, S. H. Ismail, S. Atalla, Smart real-time healthcare monitoring and tracking system using GSM/GPS technologies, *2016 3rd MEC International Conference on Big Data and Smart City (ICBDSC)*, 2016, pp. 1-7.
[2] C. Catlett, E. Cesario, D. Talia, A. Vinci, Spatio-temporal crime predictions in smart cities: A data-driven approach and experiments, *Pervasive and Mobile Computing*, 2019, Vol. 53, pp. 62-74.
[3] B. Zhou, X. Tang, and X. Wang, Learning Collective Crowd Behaviors with Dynamic Pedestrian-Agents*, International Journal of Computer Vision,* 2015, Vol. 111, pp. 50-68.
[4] C. Barrios, Y. Motai, Improving Estimation of Vehicle's Trajectory Using the Latest Global Positioning System with Kalman Filtering, *IEEE Transactions on Instrumentation and Measurement*, 2011, Vol. 60, no. 12, pp. 3747 – 3755.
[5] M. Rieser, *Adding transit-to-an-agent-based transportation simulation: Concepts and implementation*, PhD thesis, Technical University of Berlin, 2010.
[6] M. Bergkvist et al., A Hybrid Micro-Simulator for Determining the Effects of Governmental Control Policies on Transport Chains, *Davidsson, P., Logan, B., Takadama, K. (eds) Multi-Agent and Multi-Agent-Based Simulation. MABS 2004. Lecture Notes in Computer Science*, 2005, Vol. 3415, pp. 236-47.
[7] Mu-ChenChen et al., Mining changes in customer behavior in retail marketing, *Expert Systems with Applications,* 2005, Vol. 28, Issue 4, pp. 773-781.
[8] A. Kharrat, I. S. Popa, K. Zeitouni, S. Faiz, Clustering Algorithm for Network Constraint Trajectories, *Lecture Notes in Geoinformation and Cartography*, 2008, pp. 631-647.
[9] Q. Li, Y. Zheng, X. Xie, Y. Chen, W. Liu, W.-Y. Ma, Mining User Similarity Based on Location History, *GIS '08: Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems*, 2008, pp. 1-10.
[10] H. Kiyotake, M. Kohjima, T. Matsubayashi, H. Toda, Multi Agent Flow Estimation Based on Bayesian Optimization with Time Delay and Low Dimensional Parameter Conversion, *Principles and Practice of Multi-Agent Systems. PRIMA 2018. Lecture Notes in Computer Science*, 2018, vol 11224, pp. 53-69.
[11] F. Shahnaz, M.W. Berry, V. Pauca and R.J. Plemmons, Document clustering using nonnegative matrix factorization. *Information Processing & Management*, 2006, Vol. 42(2), pp. 373-386.