

# GROUP-BASED LIFE PATTERN EXTRACTION FROM LARGE SCALE GPS DATA OF MOBILE PHONE

Apichon Witayangkurn<sup>\*a</sup>, Teerayut Horanont<sup>b</sup>, Mariko Shibasaki<sup>c</sup> and Ryosuke Shibasaki<sup>d</sup>

<sup>a</sup> PhD Candidate, Department of Civil Engineering, the University of Tokyo, 4-6-1, Komaba, Meguro-ku, Tokyo 153-8505, JAPAN; Tel: +81-3-5452-6412 Fax: +81-3-5452-6414;  
E-mail: [apichon@iis.u-tokyo.ac.jp](mailto:apichon@iis.u-tokyo.ac.jp)

<sup>b</sup> Researcher, Institute of Industrial Science, the University of Tokyo, 4-6-1, Komaba, Meguro-ku, Tokyo 153-8505, JAPAN; Tel: +81-3-5452-6412 Fax: +81-3-5452-6414;  
E-mail: [teerayut@iis.u-tokyo.ac.jp](mailto:teerayut@iis.u-tokyo.ac.jp)

<sup>c</sup> Student, Center for Spatial Information Science, the University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa-shi, Chiba 277-8568, JAPAN; Tel: +81-4-7136-4291 Fax: +81-4-7136-4292;  
E-mail: [ms@csis.u-toyko.ac.jp](mailto:ms@csis.u-toyko.ac.jp)

<sup>d</sup> Professor, Center for Spatial Information Science, the University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa-shi, Chiba 277-8568, JAPAN; Tel: +81-4-7136-4291 Fax: +81-4-7136-4292;  
E-mail: [shiba@csis.u-toyko.ac.jp](mailto:shiba@csis.u-toyko.ac.jp)

**ABSTRACT:** Understanding behaviors of people especially daily life pattern is essential for planning and management of urban facilities and services. For instance, people live north side of the city tend to have same activity everyday by going out their home in very early morning and go back at late evening. It may indicate high congestion of the traffic or inadequate transportation in morning time since lot of people have to go out very early so that they can reach the office on time. With the availability of mobile phone technology, people trajectory data can be obtained directly through mobile phones in a format of GPS log. With such trajectory data, it is possible to discovery a life pattern of user. Moreover, by combining trajectories data of many users, the groups of people with the same life pattern can be identified. The main dataset used in this research is GPS-based trajectory data obtained from mobile phone embedded auto-GPS function of about 1.5 million users in a one-year period. With power-consumption function, mobile phone is embedded with accelerometer to detect the movement of mobile phone. It automatically deactivate the position sending function if no movement is detected and results in very low data rate of GPS data regarding its specification which is five minutes sending interval. However, from another perspective, such function can be used to indicate the current activity of the user or the active periods of user. For example, there are less numbers of GPS points if user is stationary and lots of GPS points while commuting. In this paper, we propose a technique to discovery group-based life patterns from the active periods of the multiple users with the aim to find group people with same activity such as ordinary worker (work during day-time) or nighttime worker. The technique involves four steps: active-periods calculation, data grouping, normalization and data clustering. The Iterative K-means clustering with cosine-similarity distance function is applied to extract similar pattern of people. In addition, because it requires processing the whole dataset which are very large, Hadoop, cloud computing platform, is used for data processing to accelerate the processing speed. Finally, we are able to define 24 groups of people based on their activity patterns.

**KEY WORDS:** GPS, Mobile Phone, Mobility Analysis, Life Pattern, Hadoop