(1) Introduction: Human movements are difficult to predict, especially, when we consider rare behaviors that deviate from normal daily routines. By tracing the behavior of a person over a long period, we can model their daily routines and predict periodical behaviors, whereas rare behaviors, such as participating in the New Year’s Eve countdown, can hardly be predicted readily and thus they have usually been treated as outliers of the daily routines in most existing studies. However, for scenarios such as emergency management or intelligent traffic regulation, we are more interested in rare behaviors than daily routines. Using human mobility Big Data, the rare behavior of each individual in a social crowd is no longer rare and thus it may be predicted when we analyze the crowd behavior at a citywide level.

(2) Methodology: In this study, instead of predicting movement based on daily routines, we make short-term predictions based on the recent movement observations. We propose a novel model called CityMomentum as a predicting-by-clustering framework for sampling future movement using a mixture of multiple random Markov chains, each of which is a Naive Movement Predictive model trained with the movements of the subjects that belong to each cluster.

(3) Data: To address the problem of real-world human mobility and to develop innovative applications that may allow governments to accelerate the promotion of smart cities, we collected an anonymous GPS log dataset from about 1.6 million real mobile-phone users in Japan over a three-year period (August 1, 2010 to July 31, 2013). The dataset stored about 30 billion GPS records with a total file size exceeding 1.5 terabytes using a Hadoop cluster. The data were collected by a mobile operator and a private company under an agreement with the mobile phone users. By default, the positioning function in mobile phones is activated and uploaded about every 5 min. However, data acquisition is not stable and robust due to several factors, such as the loss of signal or battery power. In addition, the positioning function is also turned off automatically when no motion is detected from the mobile phone (e.g. the phone is placed evenly on the table). The characteristics of the dataset also cause handling problems, so we designed the CityMomentum model to address these challenges.

(4) Results: We apply our approach to a big mobile phone GPS log dataset and predict the short-term future movements, especially during the Comiket 80 and New Year’s Eve celebration. We evaluate our prediction by a Earth Mover Distance (EMD) based metric, and show our approach accurately predicts the crowd behavior during the rare crowd events, which makes an early crowd event warning and regulation possible in the emergent situations.

(5) Future steps: We note several limitations of our work. First, a more sophisticated preprocessing will be included in our further research. Moreover, we predict future movements based on the current transition pattern, but it may deviate from the movements in reality if the transition pattern changes suddenly.

Figure 1: Predicted sources of visitors to C80 at 1 h later and predicted 1-h-later locations of the visitors at C80 (in the top row) and the actual sources of and 1-h-later locations of visitors in reality (in the bottom row)