論文の内容の要旨

A Study on Human Mobility Using Cell Phone Traces

（携帯電話トレースを用いた人の移動に関する研究）

ダンスタン　マテケニャ

In today's urban spaces, people generate data everyday either consciously (e.g., by participating in participatory sensing projects such as OpenStreetMap) or inadvertently (e.g., by using metro systems or making phone calls). The result is that we have huge collections of digital traces telling *when* and *where* people go (this is what is considered as human mobility data). These huge collections of data often hide interesting information with high potential for decision making in many domains such as urban planning. However, there exists a gap between the raw data itself and the benefits which can be reaped from it. In other words, in order to leverage this kind of data in urban planning or disaster management, we need to come up with robust tools to mine useful insights from the data. The work in this thesis contributes to the bridging of this gap by providing a set of techniques for mining mobility data. We develop these techniques by studying key aspects of human mobility using data generated from call detail records (CDRs).

Although human mobility research has attracted huge attention, it is still open to further investigation. This is partly due to the fact that human behaviors are inherently fuzzy and dynamic. Also, most human mobility studies which leverage massive datasets make conclusions which are inevitably specific to characteristics of the datasets used in the study and therefore cannot be easily generalized. For instance, suppose a study uses GPS location data coupled with social network (SN) data to develop some algorithm for location prediction. This type of algorithm cannot be adopted wholesale in a scenario where the available location data is sparse (such as that from cellular networks) and also with no access to SN data. It is against this background that we tackle three problems related to human mobility as follows: *visualization of mobility*, *residence change discovery* and *location prediction*. In the following sections, we provide more details about these works.

We first develop a web-based framework for interactive visualization of mobility patterns in order to allow easy and quick interpretation of trends. The user interface of our system (see Figure 1) consists of two main components: the left part is the dashboard while the right part is the map area. The dashboard has four sub-panels: the top most part (shown in red in the screenshot) is the timer which gets activated during animation of user movement. The statistics panel, is where we display mobility statistics for current user. In what we call the navigation panel, the user of our system can choose to load new user data and also whether they want to see animation of current user movement.

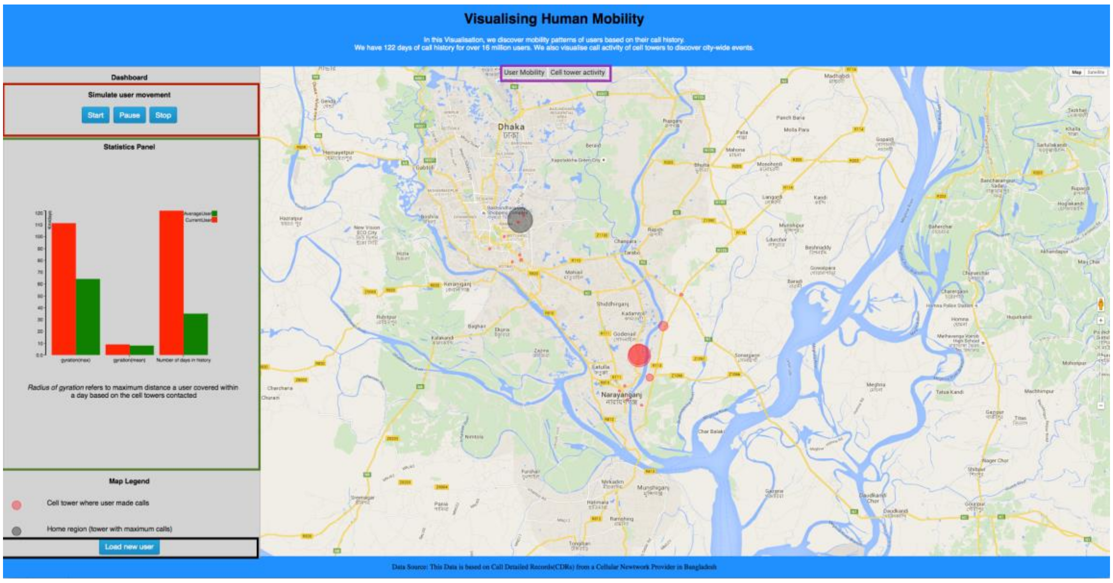

Figure 1 : Screenshot of visualization platform

In our work on residence change discovery, we explore the use of CDRs as surrogate residence history to infer change of place of residence (home). In particular, we investigate two research questions: first, whether we can reliably discover a person's residence change from unlabeled CDR data. Second, if we can develop an algorithm that can autamatically carry out this task. To this end, we first formulate the *residence change discovery problem* and then propose a sequential spatio-temporal clustering technique-*MoveSense* to solve this problem. We use a large scale CDR dataset with over 3.5 billion call records and 16 million unique users to conduct experiments to validate our technique. In the experiment, we use an *unsupervised classification approach-anomaly detection* to classify users as either having changed place of residence or not and find that across the three categories of test datasets, the technique performed well (see Figure 2) with average *detection rate* of *71 percent*, *68 percent* and *72 percent*.

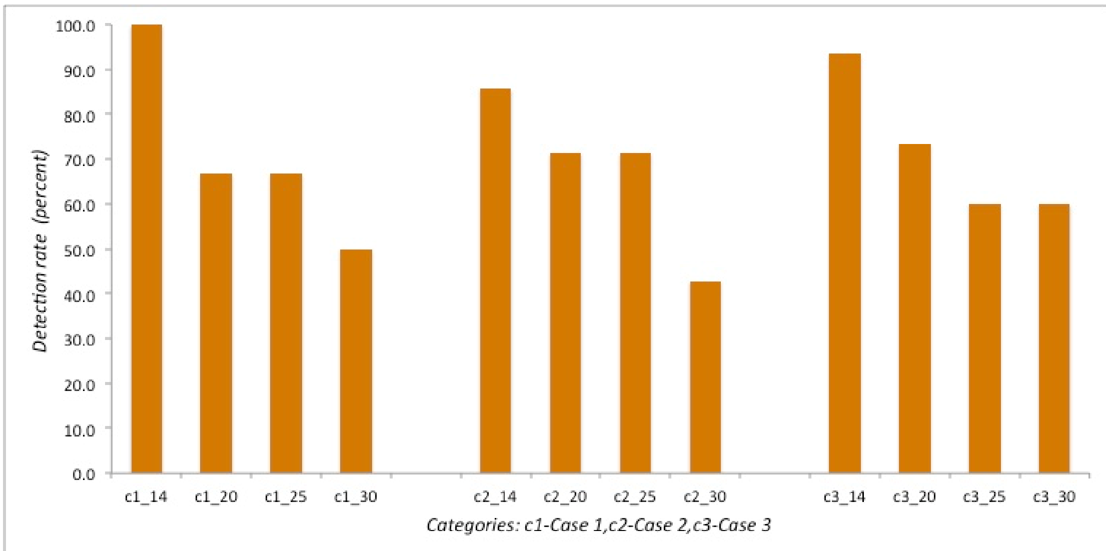

Figure 2 : Detection rates for user classification using MoveSense

In the location prediction task, the broad research question we address is: can we leverage big data to enhance performance of location predictors without relying on external data sources? In order to answer this question, we first carry out spatio-temporal analysis of user call behavior and call activity and use the insights to propose an *enhanced bayes predictor* which leverages large scale data. We conducted an experiment by using both the regular Naïve Bayes and our enhanced bayes predictor to make predictions for selected users in our CDR dataset and compared performance. Results reveal that overall, the enhancements we propose improve the predictors' (i.e., when compared to regular Naïve Bayes without our enhancements) performance by 17 percentage points (see Figure 3 below).

Secondly, we investigate the potential to improve performance (accuracy and training time) of location prediction models by again leveraging large scale data. Given that users closer in space would exhibit similar mobility behaviors, our idea is to create what we are calling a *community model* for a group of users in a given geographic area and then use parameters from this model to enhance performance for individual users in the same community. We choose to experiment with logistic regression classifier. The results from our experiments show that our idea to use *community-wide learned model parameters* in individuals works very well and reduces training time for individual models by nearly one order of magnitude as shown in Figure 4.

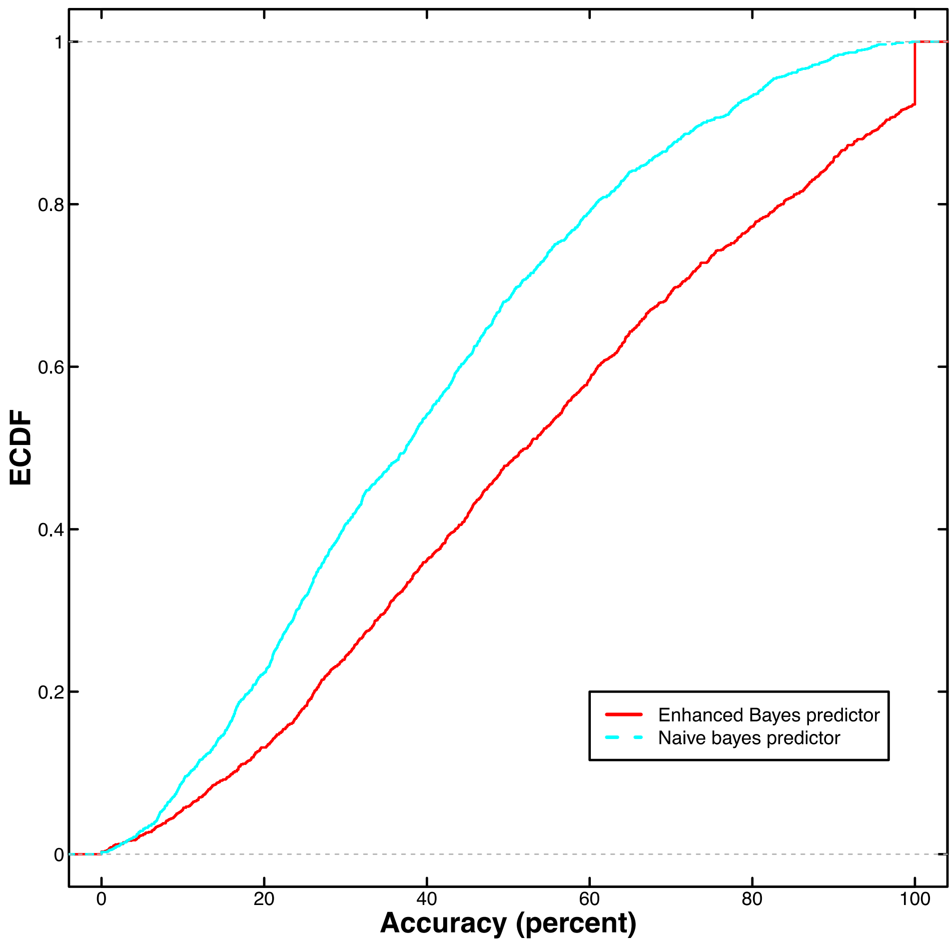

Figure 3 : Accuracy comparison for Naive Bayes vs. enhanced version

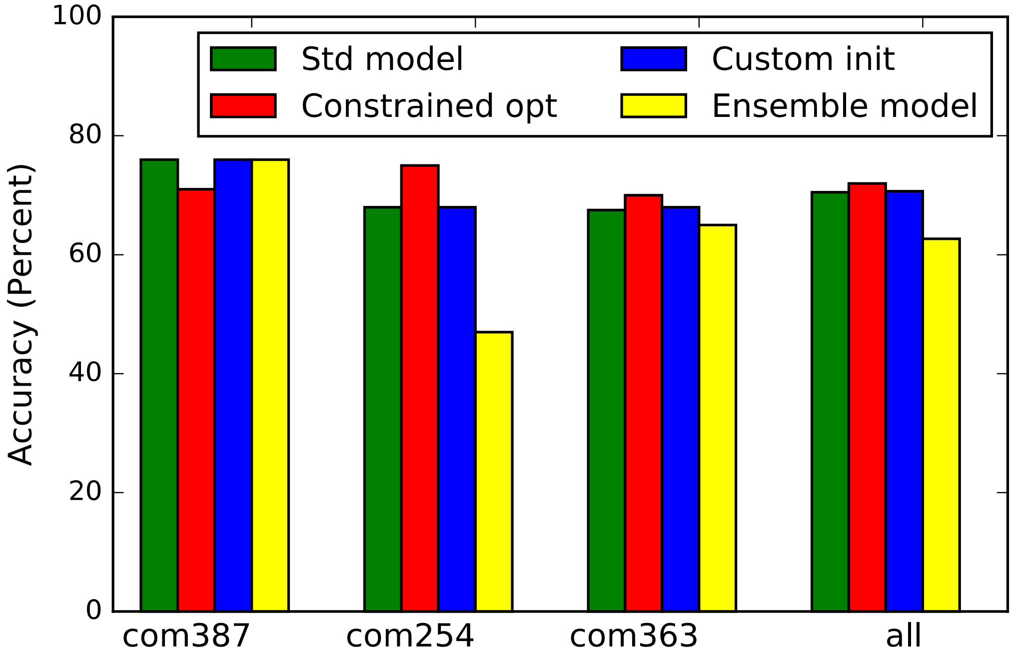

Figure 4 : Comparison of training time of model across 3 communities

In this thesis, we conducted a study of human mobility using data generated from call detailed records. We undertook this study because thorough understanding of laws that govern human movements has useful applications in public health, urban planning, disaster management, traffic engineering and marketing. We created a visualization platform to enable easy interpretation of data. We then tackled the residence change discovery problem because it has potential application in profiling internal migration rates. Finally, our work in location prediction also has benefits. In any prediction system, it is natural to aim for higher accuracy because it will in turn lead to improved service delivery while the idea of reducing training time of models is useful for scalability of models.