Modeling Transition and Mobility Patterns

Adele Hedrick[[1]](#footnote-1),\*, Ying Zhu2, and Ken Pu1

1 Faculty of Science, 2 Facult of Business and IT

University of Ontario Institute of Technology  
2000 Simcoe Street North

Oshawa, Ontario, Canada

{adele.hedrick, ying.zhu, ken.pu}@uoit.ca

**Abstract.** We present a solution to model user transitions and mobility patterns without the need of accessing any cloud services, and thus completely preserving user privacy. Our algorithm relies solely on the sensor inputs of the mobile device to gather environmental fingerprints. A real-time hierarchical clustering algorithm efficiently organizes the individual signatures into a hierarchy of meaningful significant locations at various time scale. By applying (normalized) information measure and neural network based learning, we are able to identify the most salient transition patterns that best characterize the mobility data of the user. Our algorithms are completely online, and do not rely on any networked resources. Thus, user can gain insight to their own mobility activities, and has total control of how this information is to be shared with other applications. We will demonstrate using real-life data the effectiveness and efficiency of our approach. Several appealing visualizations will be showcased. The resulting transition models can be utilized towards better user experience for a variety of mobile applications such as activity scheduling and travel route planning.

**Keywords:** human mobility · modeling · privacy

1. Introduction

Smart devices and mobile computing have become an integral part of the modern everyday life. Most of us carry a relatively powerful computing device that is constantly collecting information using an array of different sensors including the wireless network connectivity. While the smartphones are capable of providing us geolocation and mobility history, much of the algorithms and methods to achieve this require additional cloud services. For instance, Google Maps require access to the cloud Web services and other cellular network based approaches [4] to generate the geolocations and the mobility history of the device. Depending on the situation, user may be concerned with the leakage of privacy while still wish to retain the ability to access and analyze one’s own mobility history and patterns. We argue that this can be achieved by leveraging off of the device sensor inputs. Most of the collected sensor information is thrown away and never used in any meaningful way. However, there is a great deal of value that we can derive from the on-device data by applying intelligent data process [6,7].

Our motivation is to design efficient data analytic processing pipelines that are suitable for the mobile platform in order to achieve the following goals:

1. Summarize the real-time wireless data observations collected by the mobile device in a space-efficient manner.
2. The summarization allows us to construct transition models of the mobile user, and the models can be used to gain insight of the user mobility patterns.
3. The analysis should solely rely on the data collected by the mobile device at the user’s discretion. No external cloud services and network exchange are needed to ensure that all the mobility analysis does not lead to any privacy leakages.
4. Background and Problem Definition

Mobile devices have a number of continuous sensors which are sources of data. The most notable one is the on-device wireless network adapter. It naturally gathers the ambient wireless hotspots. Each wireless hotspot has a unique identifier known as BSSID. Modern mobile devices can very efficiently scan the available BSSID’s at fixed time intervals. Each scan reports a set of BSSIDs and their relative strengths. We assume that the device continuously gathers a stream of BSSID, and we call this the *raw readings* of the mobile device.

**Definition:** *Raw Reading*

*The raw readings are a sequence of BSSID sets and their signal strength, denoted as:*

*where are the timestamps and is a collection of BSSIDs and their strength.*

**Definition:** *Transition Model*

*A transition model consists of two parts.*

1. *A set of significant locations . Each location* ***L*** *is a set of BSSIDs.*
2. *An family of prediction functions: that predicts the likelihood that we will be at location L at some time in the future t.*
3. *A transition model: that predicts the likelihood that we will go to location if the current location is .*

This paper describes an end-to-end pipeline that process the data in the raw reading to a transition model.

Given the density of Wi-Fi hotspots, each distinct physical location (e.g. home or office) may yield dozens if not more BSSIDs in the reading. To further complicate the situation, Wi-Fi signals can fluctuate and intermittently be disrupted due to hardware glitches. Moreover, the distribution of BSSIDs is highly non-uniform. In an urban environment, each physical location may correspond to several dozens of distinct BSSIDs while on the road or in rural areas, each physical location may only have one or two BSSIDs.

The first problem toward our objective of transition modeling is to learn a mapping of BSSID readings to physically significant locations. Our previous work CITATION address this specific problem using an unsupervised algorithm. The algorithm constructs a hierarchy of locations with increasing geographical scale. The hierarchy consist of nodes which are physical locations. At successively higher levels, the locations correspond to larger physical scales. The root corresponds to the entire physical range that the mobile phone has covered, while the leaf nodes corresponds to specific locations. For completeness, we will briefly outline our solution for location identification. The physical location hierarchy is used as the basis of the contribution of this paper.

The second sub-problem of transition modeling is to distinguish the significance of the locations on in the hierarchy. Each location corresponds to some node in the hierarchy (to be described later). Since the locations of different granularity and geographical scale, they have unequal degree of *interestingness* for the purpose of constructing the transition model as defined above. In this paper, we will present a scoring function that assigns the degree of *interestingness* of a location. The novelty of the ranking function is that it uses a combination information theory and unsupervised machine learning to *learn* the score of the locations.

1. Hierarchical localization

In this section, we will describe the method for constructing a multi-scale locations from the raw readings. Reads can refer to our previous publication [1] for any details omitted. For completeness, we will outline the key elements of the hierarchical localization algorithm which are essential for the contribution of this paper as described in Section 4.



**Fig. 1.** A partially constructed hierarchy of locations. The leaf nodes are the BSSID sets from the raw readings. Each reading is successively processed and added to a partially constructed hierarchy.

We successively process each reading, and merge it incrementally into a hierarchy of location nodes as shown in Figure 1. The location nodes near the root cover larger sets of BSSIDs, and thus spanning greater geographical scale. The algorithm is online, and performs well independent of the length of the raw readings.

Let *H* be the hierarchy constructed by online segmentation. Each node is a location that has a set of BSSIDs. Due to the online nature of the algorithm, it is possible that a single physical location (e.g. home) can appear multiple times in the location hierarchy. For this reason, we utilize a hash function that can map a set of BSSID to a hash value using *min-wise* [2] hashing. The reason we choose min-wise hashing is that the hash value can be used to perform approximate Jaccard similarity [2] between two sets of BSSIDs.

As shown in Section 3.1 and Section 3.2, we are able to construct the following from the raw readings:

1. A hierarchy of locations, with each level having successively broader geographic coverage.
2. A min-wise hash based dictionary of locations where the key is the min-wise hash value of the BSSID set of the location, and the value is the BSSID set.
3. An information theoretic and learning based ranking and modeling

This section presents the main contribution of this paper, namely a scoring function that uses the normalized information measure and a machine learning based score to assess the interestingness of a given location.

At an intuitive level, a location is *uninteresting* if it appears so rarely that it does not contribute to any mobility *pattern*. At the other extreme, a location is equally uninteresting if it spans such large geographical scale that it’s nearly always being detected by the device. So, we need to design a scoring function that captures the following two aspects of interestingness:

1. Location should be part of a pattern.
2. The presence at a location should be informative.

Fortunately, there are well established mathematical models that corresponds to the concepts of *pattern* and *information.* The former can be captured by machine learning algorithms, while the later can be described by statistical measure of entropy and mutual information. In this section, we describe a data analytics pipeline that computes a two-dimensional score for locations that reflect the presence of pattern and information.

* 1. From activity graph to a feature space

Recall that the raw readings *R* consists of a sequence of timestamped BSSID sets, written , where is a set of BSSIDs detected at timestamp .

**Definition:** *Activity sequence*

*Given a location node with BSSIDs , the activity sequence of L, written , is given as .*

The activity sequence of *L* provides the presence of location *L* over time. The timestamps are absolute values, i.e. UNIX system timestamps. Thus, every is unique. We transform the timestamp into *feature vectors*.

(month, day of week, hour of day)

So, the activity sequence in the feature space becomes a feature activity sequence:

* 1. A neural network based pattern score

A location is *interesting* if it exhibits a pattern. Neural networks [3] are excellent at capturing *patterns*. By controlling the number of neurons and layers in the network architecture, we can control the degree of complexity of the pattern we seek. For our application, we utilize a shallow multi-layer perceptron (MLP) architecture to fit a best model for the observation derived from . The input to the MLP is the feature vectors , and the training is done on the output of the activity level .

If there exists a distinct pattern, the trained MLP will exhibit high degree of accuracy from the training. Namely the model fits the data well. We stress that the MLP needs to be appropriately designed so that there is no overfitting to the data.

**Definition:** *Predictability*

*The predictability of a location , written , is given as the training accuracy of MLP with the feature vectors as inputs, and activity level as the output.*

Predictability of a location is a measure of the degree of a pattern in the activity sequence of that location. While this will filter out uninteresting locations that have very rare appearances, unfortunately, predictability by itself is not enough to guarantee interesting locations. Consider locations in the hierarchy that span large geographic scale, such as the root location. It’s always present, so the activity sequence is nearly: . This is exceedingly predictable by any MLP, so it will have a very high predictability score. We will address this issue by introducing a second score that measures the amount of information in the distribution in the feature space.

* 1. An information theoretic score

Consider the feature sequence . We can compute a histogram which describes the activity distribution over all possible features.

**Definition:** *Activity Distribution in Feature Space*

*Let be a feature (namely a tuple of month, day of week, hour of day), and a location in the hierarchy. The activity distribution is defined as:*

*Namely, is the sum of all the activities of L during timestamps that match the feature f. We* ***normalize*** *the distribution h(f) such that it is a valid probability mass function over the features.*

**Definition:** *Support*

*The support of a location , written , is the percentage of features that has activity for the location.*

We use normalized entropy to measure the amount of redundancy exhibited by the activity distribution. Let be the number of distinct features seen in the feature sequence. The normalized entropy of is given by

The information gained when detecting the presence of a location is the reduction in the normalized entropy from that of the root location (one with the largest geographical scale). So, the relative information of a location is given by

**Definition:** *Interestingness Score*

*The interestingness of a location L is given by the vector: .*

The score provides a definitive and mathematical ordering of all the locations detected in the hierarchy *H*. Namely, we can say that is more interesting than , written , if we have and . Given that the scoring function is two dimensional, we do not guarantee that all pairs of locations are comparable.

* 1. Transition modeling

We know present the main thesis of the paper, namely transition modeling of the mobile user.

To identify the interesting locations, we filter all locations in the hierarchy by two thresholds: minimal predictability and minimal information . The interesting locations are defined as

The appropriate choice for and depends on the application. If patterns are of greater importance, should be increased, and if local precision is required, than is to be increased. Furthermore, the choices of the thresholds will also determine the number of interesting locations.

As given in Section 2, we are to construct a predictable model for individual interesting locations in , and also the transition probabilities. Both are quite straight-forward.

The location predication is immediately given by

which is to be computed by the MLP neural network. The transition probability can be constructed from the *interesting sequence*. For each reading , we can estimate the location of by the following

That is, is the interesting location that has the greatest similarity to . This induces a sequence of *interesting* locations given by

The transition probability is given by

1. Conclusion and Future Work

We have presented a complete pipeline that maps raw Wi-Fi scans collected by the mobile device to a transition model of the user without resorting to any additional cloud services, thus preserving the user privacy completely.

The contribution of this paper is the design of a two-dimensional score that assesses the interestingness of a location. The score uses a simple machine learning model (MLP) to capture any patterns if present, and an information theoretic measure to capture the amount of information of a location. Together, we are able to automatically sift out the most interesting locations encountered by the mobile device. This allows us to construct a compact transition model that are capable of:

1. Predict future occurrence of locations in the future.
2. Predict the sequence of locations most likely to occur in the future.

As future work, we would like to investigate ways that the transition model can create a more integrated and engaging user experience without any leakage of the user privacy by means of predicative decision making.

References

1. Adele Hedrick, Ken Q Pu, and Ying Zhu, Hierarchical Temporal Mobility Analysis with Semantic Labeling, In Proc. 2016 International Conference on Computational Science and Computational Intelligence (CSCI 2016), Las Vegas, NV. pp. 8 pages, Dec 15-17, 2016

2. Cohen, E., Datar, M., Fujiwara, S., Gionis, A., Indyk, P., Motwani, R., ... & Yang, C. (2001). Finding interesting associations without support pruning. IEEE Transactions on Knowledge and Data Engineering, 13(1), 64-78.

3. Haykin, Simon, and Neural Network. "A comprehensive foundation." Neural Networks 2.2004 (2004): 41.

4. Mayorga, Carlos Leonel Flores, et al. "Cooperative positioning techniques for mobile localization in 4G cellular networks." Pervasive Services, IEEE International Conference on. IEEE, 2007.

5. Cover, Thomas M., and Joy A. Thomas. Elements of information theory. John Wiley & Sons, 2012.

6. Song, C., Qu, Z., Blumm, N., & Barabási, A. L. (2010). Limits of predictability in human mobility. Science, 327(5968), 1018-1021.

7. Jensen, Bjørn Sand, et al. "Estimating human predictability from mobile sensor data." Machine Learning for Signal Processing (MLSP), 2010 IEEE International Workshop on. IEEE, 2010.

1. \* Please note that the LNCS Editorial assumes that all authors have used the western

   naming convention, with given names preceding surnames (first name then last name).

   This determines the structure of the names in the running heads and the author index.

   No academic titles or descriptions of academic positions should be included in the addresses.

   The affiliations should consist of the author’s institution, town, and country. [↑](#footnote-ref-1)