Human Mobility Prediction using Day of the Week probability

Hitoshi Ishikawa
PEZY Computing, K.K.
Tokyo,Japan
ishikawa@pezy.co.jp

ABSTRACT

In this paper, we describe our method used at the prediction for the Human Mobility Prediction Challenge (HuMob Challenge) 2024 [1]. Here, we utilized the fact that human behavior patterns seem to depend on cyclical rhythms. We improved a simple weekly model using information about which days of the week a particular date resembles.

CCS CONCEPTS

• Human-centered computing

KEYWORDS

human mobility prediction, day of the week probability

1 Basic Idea

It is natural to think that human behavior patterns depend on cyclical rhythms, such as daily, weekly, monthly, and annual, which we tried to take advantage of in this challenge.

The reference data provided here is for 75 days, which seems too short to capture monthly or annual rhythms, so we decided to consider only daily and weekly rhythms.

Our approach is simple and intuitive.

If a person was at a place at 9 a.m. last Sunday, then the same person is likely to be the same place at 9 a.m. this Sunday. On the other hand, the probability of being in the same place on different days of the week might be low.

In the simplest way of thinking, the probability that a user in a city stays at x, y is given as follows.

$$P_{city,user}(x, y|d, t) = P_{city,user}(x, y|d\%7, t)$$

This assumes that the behavior of this user exactly depends on the 7 days rhythm, i.e. weekly rhythm.

Then, we can predict the location x, y of this user as follows.

$$XY_{city,user}(d, t) = arg_{x,y}max\ P_{city,user}(x, y|d, t)$$

Clearly, this is an overly simple model and needs to be a bit more improved.

For example, people may behave differently on a regular Monday from on a holiday Monday. On the other hand, the behavior of people on a holiday Monday might be similar to that of Saturday and/or Sunday. In this sense, applying a strictly weekly rhythm may be oversimplifying the model.

Here, we improve the weekly rhythm model by introducing probabilities to capture such cases.

2 Method

From now on, Monday to Sunday is referred to as DOW (Day of the Week). Since the reference data does not include DOW information, it remains unclear what DOW a specific d is. We use an abstract integer variable dow whose value ranges from 0 to 6.

In the simple weekly model, a specific day d corresponded to one $dow(= d\%7)$. Now, we improve this model letting each day d have some probabilities to each dow.

In this model, the probability that a user in a city stays at x, y is modified as follows.

$$P_{city,user}(x, y|d, t) = \sum_{dow=0}^{6} P_{city,user}(x, y|dow, t)\ P_{city,user}(dow|d)$$

HuMob'24, October 29-November 1 2024, Atlanta, GA, USA

ACM 979-8-4007-1150-3/24/10…
https://doi.org/10.1145/3681771.3699911

However, it is difficult to obtain $P_{city,user}(dow|d)$, since usually the behavior at the day d for this user is hidden as a query and not shown in the reference data.

To calculate this, we assumed that the probability of dow under d is independent of $city$ and $user$. Assuming this, the above formula comes to use $P(dow|d)$ calculated based on all cities and user information.

$$P_{city,user}(x,y|d,t) = \sum_{dow=0}^{6} P_{city,user}(x,y|dow,t)\ P(dow|d)$$

Note that when performing the above calculation, there are cases where reference data does not exist for a specific t and therefore $P_{city,user}(x,y|d,t)$ cannot be determined. In such case, we shift t to a nearby point t^* where the value can be calculated and treat $P_{city,user}(x,y|d,t^*)$ as a substitute for $P_{city,user}(x,y|d,t)$.

.

The calculation proceeds as follows.

2.1 Count the number of occurrences

First, from the reference data, the number of times each user stayed at position x,y at time t on day d is counted as an occurrence count $Occ\_d_{city,user}(x,y,d,t)$.

Also, by calculating dow as $dow = d\%7$, we also get the occurrence count for each dow as $Occ\_dow_{city,user}(x,y,dow,t)$.

And then, each occurrence count is normalized as follows.

$$\overline{Occ\_d_{city,user}(x,y,d,t)} = \frac{Occ\_d_{city,user}(x,y,d,t)}{\sum_{x,y,t}(Occ\_d_{city,user}(x,y,d,t))}$$

$$\overline{Occ\_dow_{city,user}(x,y,dow,t)} = \frac{Occ\_dow_{city,user}(x,y,dow,t)}{\sum_{x,y,t}(Occ\_dow_{city,user}(x,y,dow,t))}$$

Following figure illustrates an image of normalized occurrence count for specific dow,t and d,t. They can be treated as vectors based on x,y.

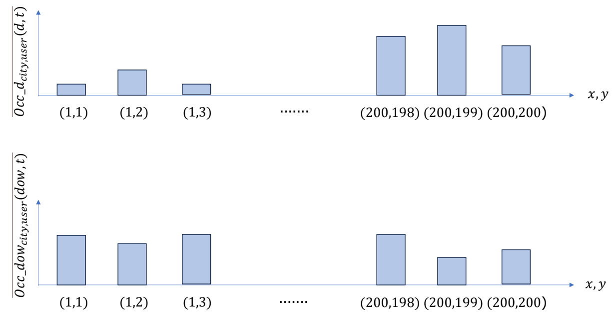

Figure 1 : an image of normalized occurrence count

It is also possible to treat them as vectors based on x,y and t for each dow or d.

2.2 Calculate probability of dow under d

Similarity between d and dow for each $city,user$ is calculated as a dot product of two vectors.

$$Sim_{city,user}(d,dow) = \sum_{x,y,t} \overline{Occ\_d_{city,user}(x,y,d,t)}\ \overline{Occ\_dow_{city,user}(x,y,dow,t)}$$

We previously assumed that the probability of dow under d is independent of $city$ and $user$. Therefore, it is reasonable to define probability as follows.

$$Sim(d,dow) = \sum_{city,user} \frac{Sim_{city,user}(d,dow)}{\sum_{dow=0}^{6} Sim_{city,user}(d,dow)}$$

$$P(dow|d) \equiv \frac{Sim(d,dow)}{\sum_{dow=0}^{6} Sim(d,dow)}$$

2.3 Predict x,y for each user

Probability of x,y under given dow and t is calculated as follows.

$$P_{city,user}(x,y|dow,t) = \frac{Occ\_dow_{city,user}(x,y,dow,t)}{\sum_{x,y}(Occ\_dow_{city,user}(x,y,dow,t))}$$

Then, probability of x,y under given d and t is calculated as follows.

$$P_{city,user}(x,y|d,t) = \sum_{dow=0}^{6} P_{city,user}(x,y|dow,t)\ P(dow|d)$$

The predicted location x, y for each user is as follows.

$$XY_{city,user}(d,t) = arg_{x,y}max\, P_{city,user}(x,y|d,t)$$

3 Result

The following figure shows $P(dow|d)$ for each day d calculated by the above method.

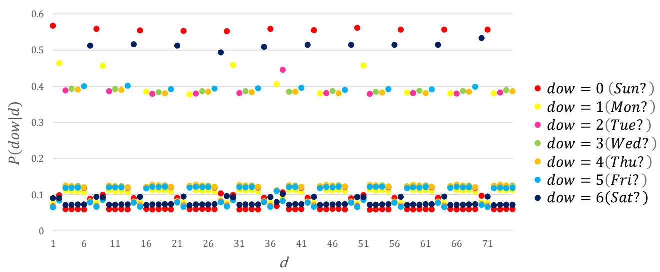

Figure 2 : $P(dow|d)$ for each day d

Although we are not sure which dow corresponds to the exact day of the week, we can see that two colors (purple and red) behave differently from the other five colors. Therefore, we can naturally infer that purple and red represent Saturday and Sunday respectively. Also, we can see that yellow marker sometimes behaves differently from the other markers that are assumed to represent weekdays. This seems to indicate that this model is able to capture phenomena that cannot be captured by a simple weekly model.

4 Discussion

In our method, there are no theoretical reason to treat d (or dow) and t differently. Instead of using $P_{city,user}(x,y|dow,t)$ and $P(dow|d)$, we could introduce another variable tow (Time of the Week) whose value ranges from 0 to 335(=7*48-1) and define the following formulation.

$$P_{city,user}(x,y|d,t) = \sum_{tow=0}^{335} P_{city,user}(x,y|tow)\, P(tow|d,t)$$

Although this formulation has potentially an ability to treat daily variation of human behavior and might lead to more precise prediction, we did not adopt it due to the computation issue. This formulation requires a large number of probabilities to be calculated in intermediate stages, which may increase the computation time and space. Furthermore, the reference data may not include enough information to calculate each probability, which will lead to unpreferable results. However, with future powerful computation resource and fruitful information, we may be able to capture daily variation of human behavior more effectively.

REFERENCES

[1] Yabe, T., Tsubouchi, K., Shimizu, T. et al. YJMob100K: City-scale and longitudinal dataset of anonymized human mobility trajectories. Sci Data 11, 397 (2024). DOI:https://doi.org/10.1038/s41597-024-03237-9