

# Predicting future location in mobile cache based on variable order of prediction-by-partial-matching algorithm

Lincan Li, Chiew-Foong Kwong, Fenghu Chen, Qianyu Liu, Jing Wang



**University of  
Nottingham**

UK | CHINA | MALAYSIA

Department of Electrical and Electronic Engineering, University of Nottingham Ningbo China, 199 Taikang East Road, Ningbo, China  
International Doctoral Innovation Centre, University of Nottingham Ningbo China, 199 Taikang East Road, Ningbo, China

First published 2018

Copyright 2018 Lincan Li, Chiew-Foong Kwong, Fenghu Chen, Qianyu Liu, Jing Wang

This work is made available under the terms of the Creative Commons Attribution 4.0 International License:

<http://creativecommons.org/licenses/by/4.0>

The work is licenced to the University of Nottingham Ningbo China under the Global University Publication Licence:

<https://www.nottingham.edu.cn/en/library/documents/research-support/global-university-publications-licence.pdf>



**University of  
Nottingham**

UK | CHINA | MALAYSIA

# *Predicting future location in mobile cache based on variable order of prediction-by-partial-matching algorithm*

*Lincan Li<sup>1</sup>, Chiew-Foong Kwong<sup>1\*</sup>, Fenghu Chen<sup>1</sup>, Qianyu Liu<sup>2</sup>, Jing Wang<sup>1</sup>.*

<sup>1</sup>*Department of Electrical and Electronic Engineering, University of Nottingham Ningbo China, 199 Taikang East Road, Ningbo, China.*

<sup>2</sup>*International Doctoral Innovation Centre, University of Nottingham Ningbo China, 199 Taikang East Road, Ningbo, China*

\* *chiew-foong.kwong@nottingham.edu.cn*

**Keywords:** MOBILE CACHING, LOCATION PREDICITON, PREDICTION-BY-PARTIAL-MATCHING.

## **Abstract**

Mobile caching at the edge of the wireless network has been regarded as an ideal approach that can alleviate the user access latencies. While there is a problem that the user terminal (UT) is moving too fast when enter a serving cache area, it may not have enough time to acquire the required data from the cache. One solution is to predict the UT's future location and pre-prepare the requested content at the cache devices that will appear in the UT's future path. Once the UT arrive at the serving cache area, they can directly acquire the data since it is already at the location, rather than send in a request to update the cache. The key point to achieve this reliably is the accuracy of the location prediction. This paper presents a location prediction model based on prediction-by-partial-matching (PPM) algorithm in the mobile cache design. The performance of this model will be compared by using one-order context, two-order context and three-order context, respectively. All the models will be evaluated in a real world data.

## **1 Introduction**

Due to the fast development of the fifth generation of mobile communication (5G) and the rapid expansion of the smart mobile devices, a massive mobile data traffic has been generated within the wireless network [1]. As such, the demand of the wireless network capacity has been increasing intensively. To do this, a cost efficient approach is needed to meet huge demand of the traffic of the content centric work [2]. Advanced caching technique is seen to be an ideal approach to achieve this.

The caching techniques uses mobile cache devices to cache the content at the edge of the wireless networks, which are base stations (BSs) and user terminals (UTs). Figure 1 shows an example of a cache-enabled network. In the cache-enabled network, UTs can acquire the content from these cache devices directly rather than from the core network via the backhaul links. Compared with the traditional method, the cache-enabled networks do not need to

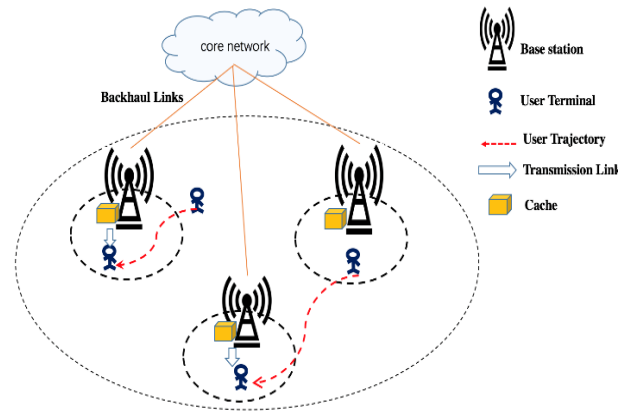


Figure 1 The cache-enabled network.

upgrade the massive high-capacity backhaul links. This is also from the fact that the price of the cache-enabled devices such as solid state drives (SSDs) have been significantly decreased over the past few years. Consequently, the costs can be managed more effectively. Besides that, the delay can be reduced since the requested content are directly from the cache devices at the wireless edge, rather than from the core network via the backhaul links. Therefore, cache technique is a promising approach which can both save the cost and reduce the delays [2] [3].

Most of the work related to cache design did not consider the effects of user mobility, which is very important to reduce the latency [2]. For example, a UT is moving too fast when enter a serving cache area, it may not have enough time to acquire the required data from the cache [4]. To mitigate this, the cache data needs to be sent prior to the arrival of the UT. By predicting the UT's future location and pre-preparing the requested content at the cache devices. Once the UT arrive at the serving cache area, they can directly acquire the data since it is already at the location, rather than send in a request to update the cache. The key point to achieve this reliably is the accuracy of the location prediction. In this paper, we proposed a location prediction model based on prediction-by-partial-matching (PPM) algorithm which has a high perdition rate.

Table 1 The notification.

Symbol	Description
$X$	The future location
$S$	The UT trajectory set
$C$	The context
$n$	The order of the context
$P$	The probability of prediction
$R$	The region set
$SP$	The special location
$N_{SP C}$	The number of times of one specific location appears in the sequence when given a UT trajectory
$N_{\Sigma SP C}$	The number of times of the total set of the special location appear in the sequence when given a UT trajectory
$P_{PPM}$	The probability of PPM
$P_{reality}$	The probability of reality

## 2. Related work

The PPM algorithm is mostly used in lossless compression [5]. According to [6], each lossless compression algorithm can be used for prediction. Hence, PPM can also be used to predict future location with high accuracy [7]. In [8], a cooperative caching scheme has been proposed for the content delivery in the vehicular content centric network. From the historical path records, the probability of the future location of the vehicles can be computed by PPM algorithm. This allow the algorithm to choose the caching nodes in the future paths to decrease the communication delay and to increase the hit ratio. According to proposal in [9], PPM algorithm can address the problems in the location prediction, such as the zero frequency problem. This algorithm deals with the zero frequency problem by using a special escape mechanism.

Reference [7] discussed a proposal for PPM algorithm for both the geographic location information and the semantic information, e.g. the reason why the terminals are in the specific location, the time when the terminals arrive and leave. The prediction of this proposed PPM model considers both the future destinations and the future routes, which helps the determination of the base nodes in the UTs' future paths. In [10], a PPM algorithm using three-order context is proposed. By given the information of the historical time, location and the future time, the future location can be accurately predicted. Specifically, the accuracy of the prediction is approximately 92%.

In [11], a spatial-temporal prediction method based on PPM algorithm is proposed. This novelty of this approach is using the temporal context information to predict the future leaving/arriving time and the future location.

## 3. Methodology

The notification is shown in Table 1

### Algorithm 1: Data collection algorithm

```

Input:  $R$   $\triangleright$  the longitude and latitude of the location is (x, y)
Output:  $SP$   $\triangleright$  the special location point
1   $SP \leftarrow a, b, c, d, e, f, g$ 
2  for all  $r \in R$  do
3    if  $r_i \in a$  then
4       $a \leftarrow r_i$ 
5    else if  $r_i \in b$  then
6       $b \leftarrow r_i$ 
7    else if  $r_i \in c$  then
8       $c \leftarrow r_i$ 
9    else if  $r_i \in d$  then
10      $d \leftarrow r_i$ 
11   else if  $r_i \in e$  then
12      $e \leftarrow r_i$ 
13   else if  $r_i \in f$  then
14      $f \leftarrow r_i$ 
15   end
16   end
    
```

### 3.1. Data collection

The input of the PPM algorithm is the location data which were collected from volunteers in Ningbo during a six-month period. An application which can be used to record the GPS information (longitude and latitude coordinates) was installed into the volunteer's UT to capture their UT's trajectory. The special locations were designated and when a volunteer arrived the location, it will be recorded automatically including other useful information such as the specific landmark of the location such as building name, which floor, etc. There are six special locations were chosen to represent the UT trajectory. These six special locations are designated as location 'a', location 'b', location 'c', location 'd', location 'e' and location 'f' respectively. Algorithm 1 summarizes the procedure of the data collection.

### 3.2. Prediction by partial matching algorithm (PPM)

The output of the PPM algorithm is a value between 0 and 1, which can be regarded as the probability of the next location. By given an input sequence  $S$ , an  $n$ -order context  $C$  and the future location  $X$ , the model of the PPM algorithm is shown like Figure 2, and the predictive probability of the future location based on the context is represented as  $P_n(X|C)$ . In addition, in order to deal with the zero frequency problem, an escape mechanism is used [8]. The definition of the zero frequency problem is that  $X$  cannot be found after the context. The escape mechanism is used to generate the escape probability  $P_n(escape|C)$ , which displays the probability that  $X$  cannot be found after the context  $C$ . The computation is start from the maximum given order  $n$ . If the future location can be found after the context, the probability  $P_n(X|C)$  will be output and the process is finished. If the location cannot be found after the context, the escape probability  $P_n(escape|C)=1-P_n(X|C)$  will be output. When occurs escape, the order is replaced by the order-1. Then the new order is used to restart the last

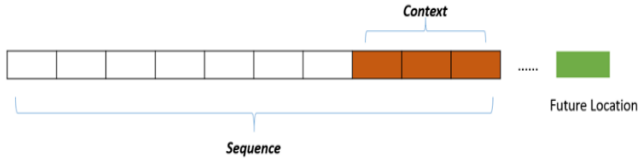


Figure 2 The model of PPM algorithm.

Table 2 The number of times of each corresponding sequence in the input sequence.

Order=2				Order=1			
C	X	N	P	C	X	N	P
ac	a	1	1/2	c	a	1	1/5
	esc	1	1/2		d	2	2/5
					esc	2	1/5

process and output the probability. The process will be finished when the future location X is found after the context or the order=-1. And the final predictive probability is multiplied by the sub probabilities of each processes. The complete expression of the probability is like:

$$P(X|C) = \begin{cases} P_n(X|C), & \text{if } X \in \sum C \\ P_n(\text{escape}|C) \times P_n(X|C'), & \text{otherwise} \end{cases} \quad (1)$$

Where  $P_n(X|C)$  and  $P_n(\text{escape}|C)$  mean the probability and escape probability corresponding to the related order.  $C'$  and  $n'$  represent the suffix of C and the new order in an escape scenario.  $\sum C$  represents X appearing after the context in the training sequence. In addition, when order is -1, the probability  $P_{-1} = \frac{1}{N}$ , where N is the number of the sequence. At the same time,  $N_{CX}$  stands for the counts of the appearance of the XC in the sequence. We can acquire the following equations,

$$P_n(X|C) = \frac{N_{CX}}{N + \sum_{X \in \sum C} N_{CX'}} \quad \text{if } X \in \sum C \quad (2)$$

$$P_n(\text{escape}|C) = \frac{N}{N + \sum_{X \in \sum C} N_{CX'}} \quad (3)$$

Algorithm 2 summarizes the procedure of the PPM algorithm.

Given an example to help understand the computation of PPM algorithm, where the input sequence is  $\{a, b, c, d, e, a, c, a, d, a, b, d, c, d, a\}$  the UT trajectory is 'ac' and the future location is 'd'. We obtain the probability  $P(d|ac)$  with Eq. (1). Firstly, the context is 'ac' and the sequence 'acd' cannot be found in the input sequence, the probability  $P(d|ac) = P_n(\text{escape}|c) \times P_n(d|C')$ . Based on Eq. (3), we output the escape probability  $P_n(\text{escape}|ac)$ . Then, the new order is replaced by the original order -1,  $C'$  is the suffix of context and  $C'='c'$  the order is replaced by order-1. The new sequence 'cd' can be found in the input sequence. Then the probability  $P_n(d|c)$  can be obtain according to Eq. (2). Table 2 summarizes the number of times of each corresponding sequence appears in the input sequence.

---

**Algorithm 2 PPM algorithm**


---

Input: S

Output: P.

▷ the probability of prediction

```

1:  P ← 0
2:  for all X ∈ S do
3:    If CX ∈ S then
4:      P ← P(X|C)
5:    else if C'X ∈ S then
6:      P ← P(escape|C) × P(X|C')
7:    end
8:  end

```

---

$$P_n(\text{escape}|ac) = \frac{1}{1+1} = 1/2,$$

$$P_n(d|c) = \frac{2}{2+(1+2)} = 2/5,$$

$$P(d|ac) = P_n(\text{escape}|ac) \times P_n(d|c) = 1/5.$$

### 3.3. Predicting the future location by prediction by partial matching algorithm

For the location prediction model, the UT trajectory corresponds to the input sequence of the PPM algorithm and each special location corresponds to the symbol of the input sequence. The probability of the UT's future location can be acquired from the above formula in 3.1. To acquire the probabilities of each location corresponding to the given sub-trajectory and rank the probabilities from top to bottom. The higher value is, the location has more probability that will appear in the UT's future path. The location which has a very high probability can be think as the predicting future location and pre-prepare the requested data for the cache devices in the future location.

### 3.4. The likelihood of the location in reality & the difference value between PPM and reality

In this section, the probability of the future location when given a known UT sub-trajectory are defined as the following Eq. (4). In Eq. (4),  $N_{SP}$  represents the number of times of one specific location appears in the sequence when given a UT trajectory.  $N_{\sum SP|C}$  represents the number of the times for the total set of location appears in the sequence when given a UT trajectory.

$$P = \frac{N_{SP}}{N_{\sum SP|C}} \quad (4)$$

The difference value is defined as the following Eq. (5). In Eq. (5),  $P_{PPM}$  represents the probability of PPM and  $P_{reality}$  represents the probability of reality.

$$\text{Difference value} = |P_{PPM} - P_{reality}| \quad (5)$$

The prediction accuracy is defined as Eq. (6).

$$\text{Prediction accuracy} = 1 - \frac{\text{Difference value}}{P_{reality}} \quad (6)$$



## 4. Results

### 4.1. The Comparison between the simulation result and the reality

In this section, the objective is to investigate the probability in PPM and reality considering the difference types of order. We randomly select ten UT trajectories with the future location, the probability of the one order PPM, two order PPM and three order PPM with their corresponding reality are shown in Figure 3, Figure 4, Figure 5, respectively. From Figure 3, with Eq. (5), the minimum difference value is 0.7%, the maximum difference value is 2.9% and the average difference value is 1%. From Figure 4, the minimum difference value is 0.6%, the maximum difference value is 3.6% and the average difference value is 1.7%. From Figure 5, the minimum difference value is 0.6%, the maximum difference value is 3.5% and the average difference value is 1.9%. The difference value indicates that the probability of PPM is almost the same with the reality. In addition, in this three types of PPM model, the minimum prediction accuracy is 86.6% and the average prediction accuracy is 95.7%, which can be regarded that PPM algorithm has a high prediction accuracy

### 4.2. The influence of different order for the location prediction

After the validation that PPM has a high location prediction accuracy, we investigate the influence of different order for the location prediction. Table 3 displays the number of the future location that has a probability more than 0.9 based on 1-order, 2-order and 3-order PPM model, respectively. The UT trajectory has a probability larger than 0.9 means that the future location can be well predicted for its corresponding historical UT trajectory. It can be seen that 1-order PPM has 1, 2-order PPM has 6 and 3-order PPM has 13 UT trajectory whose prediction probability is larger than 0.9. It can be concluded that 1 location can be well predicted in 1-order PPM model, 6 locations in 2-order PPM model and 13 in 3-order PPM model when given a corresponding UT trajectory. Therefore, in this model, 3-order PPM algorithm has the best location prediction performance compared with the other two PPM algorithms.

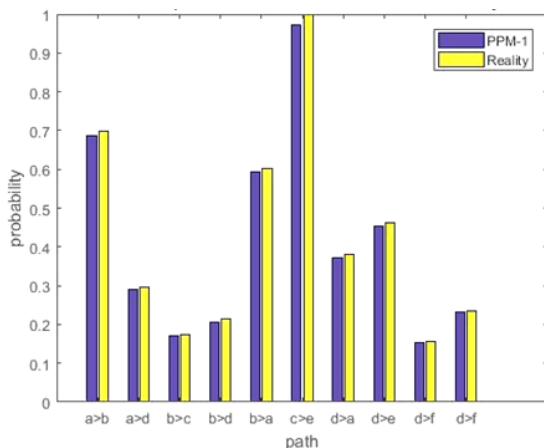


Figure 3 The probability of 1-order PPM and reality. Blue line represents 1-order PPM and yellow line represents reality.

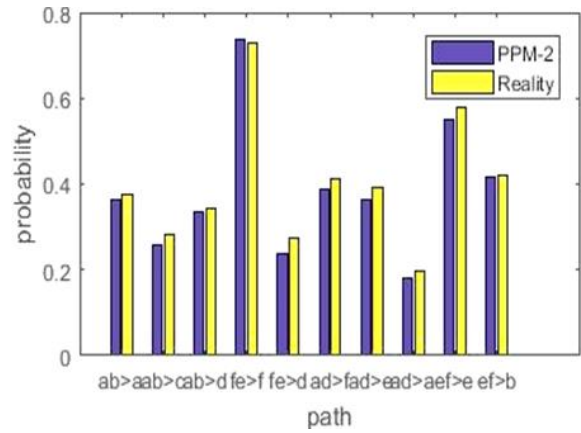


Figure 4 The probability of 2-order PPM and reality. Blue line represents 2-order PPM and yellow line represents reality.

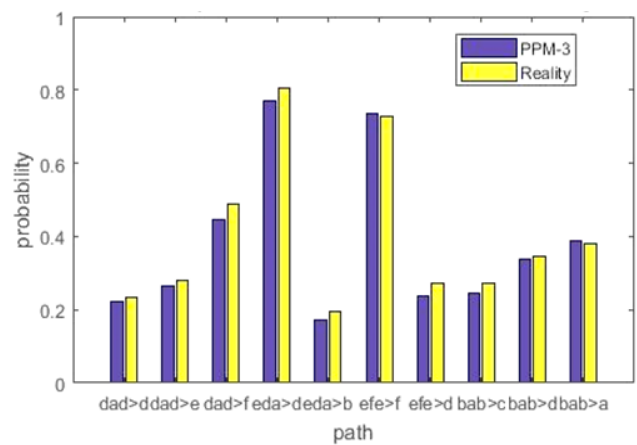


Figure 5 The probability of 3-order PPM and reality. Blue line represents 3-order PPM and yellow line represents reality.

Table 3 The number of the UT trajectory that has a high probability > 0.9.

	1 order	2 order	3 order
P>0.9	1	6	13

## 5 Conclusion

In this paper, we proposed a future location prediction model from the historical UT trajectory. We first proposed an algorithm to discover the special location from the UT historical trajectory (longitude and latitude coordinates) and use these special locations to better represent the UT trajectory. We predict the UT's future location by the optimised UT trajectory based on PPM algorithm. From the simulation result, it has shown that PPM has a high prediction accuracy comparing with the reality. Apart from this, in our experiment, the 3-order PPM has the best location prediction performance compared with the 1-order PPM and 2-order PPM algorithm.

## 6 Acknowledgements

This work is supported by Ningbo Natural Science Programme (NBNSF), project code 2018A610095.

## 7 References

- [1] X. Zhang and Y. Cao, "A cooperation-driven ICN-based caching scheme for mobile content chunk delivery at RAN," *2017 13th Int. Wirel. Commun. Mob. Comput. Conf. IWCMC 2017*, pp. 1437–1442, 2017.
- [2] R. Wang, X. Peng, J. Zhang, and K. B. Letaief, "Mobility-aware caching for content-centric wireless networks: Modeling and methodology," *IEEE Commun. Mag.*, vol. 54, no. 8, pp. 77–83, 2016.
- [3] N. Zhang, J. Guan, C. Xu, and H. Zhang, "A Dynamic Social Content Caching under User Mobility Pattern," pp. 1136–1141, 2014.
- [4] M. Chen, Y. Hao, L. Hu, K. Huang, and V. Lau, "Green and Mobility-aware Caching in 5G Networks," *IEEE Trans. Wirel. Commun.*, vol. 16, no. 12, pp. 1–1, 2017.
- [5] Y. Zhang and D. A. Adjeroh, "Prediction by partial approximate matching for lossless image compression," *IEEE Trans. Image Process.*, vol. 17, no. 6, pp. 924–935, 2008.
- [6] R. Begleiter, R. El-Yaniv, and G. Yona, "On Prediction Using Variable Order Markov Models," *J. Artif. Intell.*, vol. 22, pp. 385–421, 2004.
- [7] C. Paper, F. Dantas, and N. Neto, "Predicting Routes and Destinations of Urban Trips using PPM Method," no. September, 2016.
- [8] L. Yao, A. Chen, J. Deng, J. Wang, and G. Wu, "A Cooperative Caching Scheme Based on Mobility Prediction in Vehicular Content Centric Networks," *IEEE Trans. Veh. Technol.*, vol. 9545, no. c, pp. 1–10, 2017.
- [9] J. Yang, J. Xu, M. Xu, N. Zheng, and Y. Chen, "Predicting next location using a variable order Markov model," *Proc. 5th ACM SIGSPATIAL Int. Work. GeoStreaming - IWGS '14*, pp. 37–42, 2014.
- [10] I. Burbey and T. L. Martin, "Predicting future locations using prediction-by-partial-match," *Proc. first ACM Int. Work. Mob. entity localization Track. GPS-less Environ. - MELT '08*, p. 1, 2008.
- [11] R. Roor, M. Karg, A. Liao, W. Lei, and A. Kirsch, "Predictive ridesharing based on personal mobility patterns," *2017 IEEE Intell. Veh. Symp.*, no. Iv, pp. 1322–1329, 2017.