

RESEARCH

Mobility-aware Personalized Service Recommendation in Mobile Edge Computing

Hongxia Zhang^{*}, Yanghui Dong and Yongjin Yang

^{*}Correspondence:

zhanghx@upc.edu.cn
Department of Computer Science
and Technology, China University
of Petroleum, Qingdao, China
Full list of author information is
available at the end of the article

Abstract

With the proliferation of smartphones and an increasing number of services provisioned by clouds, mobile edge computing (MEC) is emerging as a complementary technology of cloud computing. It could provide cloud resources and services by local mobile edge servers, which are normally nearby users. However, a significant challenge is aroused in MEC because of the mobility of users. User trajectory prediction technologies could be used to cope with this issue, which has already played important roles in service recommendation systems with MEC. Unfortunately, little attention and work have been given in service recommendation systems considering users' mobility. Thus, in this paper, we propose a mobility-aware personalized service recommendation approach based on user trajectory and QoS predictions. In the proposed method, users' trajectory is firstly discovered by hybrid long-short memory networks. Then, given users' trajectories, service QoSs are predicted, considering the similarity of different users and different edge servers. Finally, services are recommended by a center trajectory strategy based on the aforementioned information. Experimental results based on the real base station dataset show that our proposed approach can outperform the traditional recommendation approaches in terms of the accuracy in mobile edge computing.

Keywords: mobile edge computing; mobility; edge service; service recommendation; QoS

Introduction

In mobile edge computing (MEC), a number of edge servers with computation and storage capabilities are deployed to construct a network, termed as the mobile edge network (MEN) [1]. Services can be provided within close proximity of service subscribers for satisfying the high-workload and low-latency requirements through MEN. Application developers and content providers are able to deploy services on edge servers according to context information and mobile edge network information [2]. However, due to the increasingly growing popularity of mobile devices, a large number of mobile services have been developed for mobile devices. For high efficiency and excellent service qualities, users should invoke their desired services through accessing edge servers in mobile edge computing (MEC) rather than through core networks or Internets [3, 4]. It is critical to recommend proper mobile services to users, taking both quality of services (QoS) and network optimizations into consideration.

Undoubtedly, the mobility of users may have detrimental impacts on the QoS-based service recommendation [5]. To be specific, mobile users request services from different edge servers when they move from service areas of one edge server to others. On the other hand, QoS data of a mobile server may prominently

volatile due to the mobility of users, as services have to switch among edge servers frequently according to user requirements, resulting on higher network overhead. Although some approaches [6] on service recommendation have already been proposed in MEC, these approaches directly overlook the impact of users mobility, or merely consider users mobility with unrealistic models, which cannot be used to evaluate the impact on users' mobility, yielding inappropriate service recommendation. Considering practical scenarios, two main problems should be taken into account in service recommendation while neglected in existing literature, which are summarized as follows:

- *Frequent handoffs among edge servers*: The handoffs among different edge servers may occur frequently, especially for high-mobility users, such as autonomous vehicles. In such a case, users request desired services from different edge servers at different moments because of their dynamical location changes [7]. However, each edge server is only able to provide the service for a moment due to the high QoS performance in a small area. The network loading and server loading can be increased by data transmission among servers. Besides, the QoS may be changed or even absent on servers. Therefore, the mobility of users may significantly degrade the service performance in MEC [8].
- *Volatility of QoS data*: In the light of extensive network changes, the QoS data are unstable. One active user invokes the service many times, and QoS data is different each time. Different users who invoke the same service are provided by different QoS data at the same time. This phenomenon is common in real life. Especially in MEC, the QoS data is different each time, because of different performance, surroundings and the number of users using services at that time. So the QoS data is highly volatile, it is more complicated when recommending better services for users.

To solve the aforementioned problems, a mobility-aware personalized service recommendation approach is proposed in this paper. To address the negative impacts caused by user mobility and QoS data volatility in MEC, user trajectories are predicted to obtain candidate servers leveraging the long short-term memory (LSTM) network firstly. Next, we predict QoS data on candidate servers based on the trajectory by reducing the influence of the above two factors. Then, QoS predictions are carried out based on location-based collaborative filtering approach and service recommendations are performed by the center trajectory strategy. The major contributions of this paper are summarized as follows:

- A hybrid LSTM-based mobility prediction approach is proposed to predict user trajectories from long-range and sparse position data, in which server information as well as user characteristics and distances information are considered to improve the prediction efficiency.
- We put forward a mobility-aware personalized service recommendation approach, focusing on predicting the QoS data of servers along trajectories, where user and server information are applied to reduce the influence of the volatility of QoS data. In the proposed approach, servers around predicted places on the trajectory will provide global optimal QoS values and reduce the network overload.

The rest of this paper is organized as follows: Section 2 summarizes the related work. Section 3 introduces a motivation scenario. Section 4 presents our mobility-aware service recommendation approach based on trajectory prediction and edge server similarity. Section 5 describes the implementation of our experiments and performance comparisons. Section 6 draws a conclusion.

Related work

Many literature over the past decade focused on developing service recommendation systems. Most of these works focus on collaborative filtering based recommendation [6,9,10], content-based recommendation [11] and model-based recommendation [12,13].

Shao et al. [14] proposed a user-based collaborative filtering method for predicting the web service QoS value and conducted experiments on 20 web services. Zheng et al. developed a web service recommendation system called WSRec [15–17] for collecting a large-scale real-world web service QoS information through a user-collaborative mechanism. They proposed a hybrid collaborative filtering approach to predict web service QoS value by combining the user-based Pearson correlation coefficient(PCC) approach and the item-based PCC approach [16]. Sun et al. [9] presented a normal recovery collaborative filtering approach to solve the personalized web service recommendation problem. They proposed a new similarity measure for web service similarity computation and proposed a novel collaborative filtering approach. Better prediction accuracy is achieved through fusing the information of similar users and similar web services. Wang et al. [6] performed the prediction via a set of multi-dimensional QoS measures through exploiting the structural relationships among multi-dimensional QoS data. They proposed an integrated QoS prediction approach to recommend efficient Web service for mobile clients. These approaches made a lot of effort on improving the accuracy of service recommendation and achieved great success in the traditional Web service recommendation. Due to users' mobility and QoS volatility, these approaches will lead to large deviations and inappropriate results when recommending services in mobile edge computing.

To address these problems, many studies have been studied on the location-based recommendation [18–20]. Zheng et al. [21] presented a user-centered collaborative location and activity filtering approach to cluster many users data, applied the collaborative filtering to find like-minded users and like-patterned activities at different locations. They modeled the user location-activity relations with a tensor representation and proposed a regularized tensor and matrix decomposition solution to address the sparse data problem in mobile circumstances. This approach only considered the users locations, not service locations. Chen et al. [22] proposed a location-aware Web service recommend system, which employed both location information and QoS values to cluster users and services. They made personalized QoS prediction for users based on the clustering results according to user locations and QoS values. Location-Aware and personalized collaborative filtering approach [23] leveraged locations of both users and Web services when selecting similar neighbors for the target user or service. The method presented an enhanced similarity measurement for both users and Web services, by taking into account the personalized influence of them. Wang et al. [6] indicated that user mobility often makes

service QoS prediction values deviate from actual values in mobile edge computing networks. They proposed a service recommendation approach based on collaborative filtering and make QoS prediction based on user mobility. They divided service invocation into two cases and calculated user or edge server similarity separately. But the frequency of mobility did not take into consideration, which may bring immigration among different edge servers frequently.

Although the location is vital to service recommendation at the current time, it cannot explain the mobile trend of users, and it is not sufficient to predict QoS data only based on the next location when recommending services. Therefore, these approaches will bring inaccurate even bad results in MEC. To address these problems, we propose an approach that both consider the user trajectory and the data volatility to provide high quality services.

Scenario and Problem Formulation

Scenario

A mobile edge network consists of multiple mobile edge servers (*MESs*), which provide network access services for mobile users anytime in a certain range, such as airports, large shopping malls, museums, etc. Suppose that a mobile user named Jenny often uses services to obtain surrounding information in an airport, such as AR/VR, location, discount, etc.

Considering Jenny is invoking service s by accessing the nearby mobile edge server *MES1*. As Figure 1 depicts, she uses service s continuously on her traveling from A to B along the trajectory as the black dotted line shown. In this situation, how to predict the next accessing edge server and the corresponding QoS value of the service becomes an important issue. If service s is always provided by the nearest edge server, that is to say, the green-line option: $MES1 \rightarrow MES2 \rightarrow MES5 \rightarrow MES3 \rightarrow \dots$ will provide the service successively. However, service s in *MES2* may have bad quality (e.g., the response time is 100ms) for accessing by many people at that time. After that, Jenny will access *MES5* with the response time is 95ms for the shortest distance. Meanwhile, the response time of service s is 95ms when accessing *MES4* during the entire moving process. Therefore, the red-line option: $MES1 \rightarrow MES4 \rightarrow MES6 \rightarrow \dots$ will obtain the best experience globally, which has the same path direction as Jenny's trajectory. Therefore, the users experience can be improved and the network delays can be reduced compared to the case where only the nearest edge server is recommended. In the above scenario, the trajectory defines user mobility more accurate than the current location.

Problem Formulation

In this section, we first define the mobility prediction model, then the mobility-aware personalized service recommendation problem is formalized. Trajectory data are the foundation of the mobility prediction, which characterize locations and times of moving objects. A trajectory is a chronologically ordered sequence of spatio points. A spatio point q_i is defined as a tuple of its coordinate $\langle x_i, y_i \rangle$. A sequence sq is a set of spatio point q_i , i.e. $sq = q_1 q_2 \dots q_n, (n \in D)$. A subsequence of sq is defined as $sub(sq)_l = q_{r-l} \dots q_{r-1} q_r$, where $r-l$ is the start position on the subsequence, and l is the number of spatiotemporal points.

Formally, given a sequence set SQ and an active user u , the trajectory set of user u is a chronologically ordered subsequence. $Tr(u) = sub(sq_r)_l$ contains all the trajectories generated by user u with the sequence length l . A trajectory of user u which is recorded at the t time window can be defined as $Tr(u)_t = sub(sq_t)_{l=q_t-l \dots q_t-1 q_t} \subseteq Tr(u)$.

If the current time window is t , then the trajectory history can be denoted as $Tr(u)_1, Tr(u)_2, \dots, Tr(u)_{t-1}$. For the convenience of calculation, we discretize continuous time interval t to hour of day and day of week.

The mobility prediction problem can be formalized as:

Problem 1 Given a user's history trajectory set $Tr(u)$ and the current trajectory $Tr(u)_t$, the mobility prediction is to predict the user's trajectory $\widehat{Tr}(u)_{t+1}$ at the next time interval, the prediction length $l = len(Tr(u)_{t+1})$, that is, construct the predicting model $f(Tr(u), Tr(u)_t) \mapsto \widehat{Tr}(u)_{t+1}$.

Analysis results [24] show that people's trajectory with high probability, which confirms the existence of mobility patterns for users. If the mobility pattern can be obtained through the history trajectory, then the Problem 1 will be solved based on the related mobility pattern information.

Definition (Mobility pattern) Given a trajectory set Tr of users U , the users mobility pattern mp is a set of typical trajectories that can indicate the characteristics of users movement and its uniqueness.

Apparently, QoS properties, including response time, throughput, failure-probability, etc., are heavily depend on network environments and location of users [22]. As a result, in service recommendations, the QoS properties of MEC servers along trajectories should be considered rather than only selecting nearest ones. Therefore, our motivating problem is to make more accurate QoS estimations based on the trajectory for personalized service recommendations in MEC to improve users quality of experience and reduce the overload of networks. Therefore, the user mobility-aware personalized service recommendation problem in MEC can be formulated as follows:

Problem 2 Let U be a set of users, S be a set of services, MES be a set of MEC servers, an active user $u \in U$ is using a service $s \in S$ providing by $mes \in MES$ when roaming at m , then u leaves the service area of mes or the QoS of s gets worse gradually, the $mes_i \in MES$ around the predicted trajectory $\widehat{Tr}(u)_{t+1}$ will be recommended to provide service s for user u for better QoS. The problem is predicting the QoS data of service s on the MEC servers along the predicted trajectory $\widehat{Tr}(u)_{t+1}$, then the server $mes_j \in MES$ with the best QoS data will recommend to u .

To reach this goal, several challenges must be addressed, including: i) How to discover the user's mobility pattern and predict the trajectory. ii) How to redefine the collaborative filtering algorithm to predict QoS data and perform service recommendation when considering edge servers information.

Mobility-aware personalized service recommendation in MEC

In this section, we propose a two-stage mobility-aware personalized service recommendation approach. Firstly, a hybrid LSTM-based model for mobility prediction approach is presented to predict the user trajectory by combining the user information, historical trajectory information and distances between servers. Then a

personalized server recommendation approach is presented to predict the QoS data on the mobile edge servers around the trajectory by weakening the volatility of QoS data.

As shown in Figure 2, our proposed approach consists of three steps as follows:

- 1) Trajectory prediction. In this stage, a hybrid LSTM-based model(LSTMM) is presented to predict trajectories. User mobility patterns can be discovered by the historical trajectories and user information based on LSTM, then a fully connected classifier is built to predict the future trajectory through combining user mobility pattern and the current position. The servers along the future trajectory consist of the candidate servers set CS , which considering the frequency mobility of users in the recommendation process.
- 2) QoS prediction: The QoS data provided by the candidate server set CS will be predicted in this stage. We propose a personalized service recommendation approach which considers the similarity of both users and servers to weaken the volatility of QoS data. The QoS values on edge servers along the trajectory are predicted based on similar users and similar servers.
- 3) Service recommendation: Based on predicted QoS values along the trajectory, we make service recommendations to meet user requirements.

a Hybrid LSTM-based Mobility Prediction

In this section, we present a hybrid LSTM-based mobility prediction approach to discovery user mobility patterns and predict the user trajectory, then the edge servers around the trajectory will be the candidate servers recommended.

1. Users mobility patterns discovery

Each trajectory has different spatio points and variable temporal distances. To process these long-term variable length trajectories, we employ Long Short-Term Memory (LSTM) networks to discover user mobility patterns. LSTM is a particular implementation of recurrent neural networks (RNN), which can learn long-term dependency information and avoid the problem of gradient disappearance. It is generally considered much suitable for time series. The network architecture of the LSTM predictor is schematically presented in Figure 3.

Specially, we take the user information and the historical trajectories as the input to model the characteristic of mobility patterns. Before training the model, the spare representation of user and location information are converted into a dense vector. Position vector pos_i is embedded to $\mathbf{p}_i$ by

$$p_i = V_p \cdot pos_i, \quad (1)$$

where V_p is an trainable embedding matrix for all positions, and pos_i is an one-hot vector of a spatio point. User information vector u_j is embedded to $\mathbf{u}_j$ by

$$u_j = V_u \cdot us_j, \quad (2)$$

where V_u is an trainable embedding matrix for all positions, and us_j is an one-hot vector of a user.

For the trajectory vector $\mathbf{Tr}(\mathbf{u}_j)_t = \{\mathbf{p}_s, \mathbf{p}_{s+1}, \dots, \mathbf{p}_t\}$, let h_t and h_{t+1} denote the previous and current embedding state respectively. The basic LSTM model, which takes $\mathbf{p}_i$ and h_t as inputs, the implementation is as follows:

$$\begin{aligned} f_{t+1} &= \sigma(W_f[h_t, p_{t+1}] + b_f) \\ i_{t+1} &= \sigma(W_i[h_t, p_{t+1}] + b_i) \\ o_{t+1} &= \sigma(W_o[h_t, p_{t+1}] + b_o), \end{aligned} \quad (3)$$

where f_{t+1} , i_{t+1} , o_{t+1} are the forget gate, input gate and output gate respectively, σ stands for the standard sigmoid function, W are different gate parameters, and b is the bias. The memory cell c_{t+1} is updated by partially replacing the existing memory unit c_t with a new cell as:

$$c_{t+1} = f_{t+1}c_t + i_{t+1}\sigma(W_c[h_t, p_{t+1}] + b_c). \quad (4)$$

The embedding of the trajectory $\mathbf{Tr}(\mathbf{u}_j)_{t+1}$ which expresses the mobility pattern information is then output by,

$$h_{t+1} = o_{t+1} \tanh(c_{t+1}). \quad (5)$$

2. Users trajectory Prediction

To predict the following trajectory of an active user, a fully connected encoder takes the encoded $\mathbf{u}_j$ and $\mathbf{Tr}(\mathbf{u}_j)_t$ as input and predicts the probability Pr_{t+l} of user u_j belonging to each place P_i in the future $t+l$ time by

$$Pr_{t+l} = \text{softmax}(fc_{t+l}), \quad (6)$$

where fc_{t+l} is the output of feature data after fully connection operation in the l_{th} branch, σ stands for the standard ReLU function [25]. Then, a softmax layer is used to normalizes the output value fc_{t+l} whose formula is given as,

$$fc_{t+l} = W_{i2}\sigma(W_{i1}[u_j, tr_j] + b_{i1}) + b_{i2}, \quad 1 \leq i \leq seq \quad (7)$$

Besides, due to the limitation of users movement, the distance among the current server and the future servers is also an important factor during the trajectory prediction. $dist(pos_i) = [d_1, d_2, \dots, d_j, \dots, d_{|p|}]$ is defined as a $|P|_{th}$ vector represents the distance between the current server position pos_i and the positions of other edge servers. A typical geographic distance algorithm [26] is adopted to calculate the distance $d_j (1 \leq j \leq |P|)$, which means the distance between pos_i and pos_j , where pos_i and pos_j specified by the coordinates $\langle x_i, y_i \rangle$ and $\langle x_j, y_j \rangle$,

$$dist(pos_i)_j = 2R \arctan\left(\frac{\sqrt{hav(\theta)}}{1 - \sqrt{hav(\theta)}}\right), \quad (8)$$

with $hav(\theta) = \sin^2\left(\frac{y_1 - y_2}{2}\right) + \cos y_1 \cos y_2 \sin^2\left(\frac{x_1 - x_2}{2}\right)$, where θ is a central angle between the two edge servers and R is the radius of the Earth, which we assume to be 6371 km.

Considering the distance effect on the future position result, a parameter $\alpha(0 \leq \alpha \leq 1)$ is employed to adjust the final results. The probability of position on $t + l$ time can be formula by,

$$P_{t+l} = \text{softmax}((1-\alpha) \cdot \text{Pr}_{t+l} + \alpha \cdot \text{dist}(\text{pos}_i)). \quad (9)$$

3. Implementation

The model is trained by maximizing the likelihood of the true position being generated from the predicted value. Hence, we jointly learn all the parameters by minimizing the negative cross entropy loss $L = - \sum_{l=1}^m y_l \log(P_l | W, b)$ for the future trajectory. An important aspect of the training phase is all parameters should be learned jointly, since the hybrid LSTM layers of the encoder and decoder units are trained in an end to end fashion. Thus, an adaptive gradient descent algorithm is used to backpropagate the loss for each trajectory at every time slot t . The pseudo-code of the trajectory prediction is shown as Algorithm 1.

Algorithm 1 Mobility Prediction

Input: The trajectory set Tr ; The user set U ; The trajectory length lag ; The prediction trajectory length seq ;
Output: Trained model

- 1: // Stage 1. Data preprocessing.
- 2: Obtain all spatio points in Tr , calculate the distance matrix D between each points, and construct an empty train set TS ;
- 3: Construct train set TS ;
- 4: **for each** $u \in U$ **do**
- 5:
- 6: Extract the sub-trajectory $St(u)$ of length is $(lag+seq)$ from $Tr(u)$
- 7: Split $St(u)$ into $tr=(p_1, \dots, p_{lag})$, $tp=(pre_1, \dots, pre_{seq})$
- 8: Get distance vector $dist_p$ between the current location p_{lag} and all locations from the distance matrix D
- 9: Add $(u, tr, tp, dist_p)$ to TS
- 10: **end for**
- 11: // Stage 2. Model parameter training.
- 12: Initialize model parameters
- 13: **repeat**
- 14: **for each** $ts \in TS$ **do**
- 15:
- 16: Get $(u, tr, tp, dist_p)$ from ts
- 17: $emb_u = \text{emb}(u)$ //According to Eq.(1)
- 18: $emb(p_i) = \text{emb}(p_i)$, $p_i \in tr$ //According to Eq.(2)
- 19: Get trajectory vector representation v_{tr} of the trajectory tr by enter $emb(p_i)$ to LSTM
- 20: $join = \text{concat}(emb_u, v_{tr})$
- 21: **for each** $1 \leq l \leq seq$ **do**
- 22: Calculate predict probability Pr_{t+l} by Formula (6)
- 23: Calculate per time probability of belonging to each known location P_l by Formula (9)
- 24: Calculate $lossEntropy(P_l, \text{onehot}(pre_l))$
- 25: **end for**
- 26: Update model parameters by adaptive gradient descent
- 27: **end for**
- 28: **until convergence**

Algorithm 1 takes the trajectory set Tr , the user set U , trajectory length lag and prediction length seq as input, and output is a trajectory predicted model with stable parameters. Line 1-9 is the data pre-processing stage, where an empty training set is constructed in line 2, and the distance matrix D between each spatio point is calculated in Line 5-7. Lines 5 and 6 indicate that the sub-sequence $St(u)$ of length $(lag + seq)$ is extracted from the user trajectory $Tr(u)$, then the front lag locations of each sub-sequence are used as training data, and the seq locations

are used as training labels, and line 7 get the distance $dist_p$ between the current location p_{lag} and all locations from the distance matrix D . Line 11-26 is the stage of model training, line 11 initializes all trainable parameters of the model. The line 15 and 16 convert the sparse representation of the user and the location into a dense vector emb_u and $emb(p_i)$ through the trainable embedding matrix, and line 17 inputs the sequence location vector into the LSTM-cell to obtain the vector representation encoder v_{tr} of the trajectory. Vector **join** is to connect the user's representation and the trajectory representation as a user trajectory pattern vector and as an input to the location prediction classifier in line 18. The line 20 and 21 calculate the output of the location prediction classifier and output the final prediction probability P_{t+l} in conjunction with the location distance according to the Formula (6) and (9). Line 22 calculates the training loss and line 24 update all trainable parameters by adaptive gradient descent.

Mobility-aware Personalized Service Recommendation

The number of QoS data on MEC servers is not large in MEC, because of the user mobility and the service quantity. When services are provided, it can be described by a user-service-server matrix, the majority of empty data in the matrix donates that the service has never been invoked by any user on the server before, which leads to the matrix density. Besides, the QoS data vary greatly on different servers, due to the variety of edge servers. The traditional similarity calculation methods do not properly handle the QoS data in different vector spaces. Therefore, the existing well-know similarity measurement methods for the service recommendation is not suitable. To overcome these shortcomings, a location-based collaborative filtering (LCF) method is proposed to recommend services.

1. Normalized similarity calculation

In our approach, we first measure the similarity between servers and users to predict QoS data on certain servers. The user-service-server matrix is decomposed into user-service matrix US and server-service matrix SS firstly. Then we normalized each row of the original US and SS to eliminate data variety.

Assume the server s and t co-provide services J , $sdist(s, t)$ calculates the Euclidean distance between s and t in the $|J|$ -dimensional vector space, while $sdis_{max}$ calculates the maximal Euclidean distance in the $|J|$ -dimensional vector space. Therefore the formula to measure the similarity between two servers s and t is as follows:

$$\begin{aligned} Sim(s, t) &= 1 - \frac{sdis(s, t)}{sdis_{max}} \\ &= 1 - \frac{\sqrt{\sum_{j \in J} \left(\frac{r_{s,j} - r_{smin}}{r_{smax} - r_{smin}} - \frac{r_{t,j} - r_{tmin}}{r_{tmax} - r_{tmin}} \right)^2}}{\sqrt{|J|}}, \end{aligned} \quad (10)$$

where $J = J_s \cap J_t$ is the set of services co-provided by servers s and t ; $r_{s,t}$ is the value of service j provided by server u in the server-service matrix SS ; r_{smin} and r_{smax} denote the lowest and the highest values from server s in SS , respectively; r_{tmin} and r_{tmax} denote the lowest and the highest values from server t in SS , respectively. $Sim(s, t) = 0$ represents that two servers are dissimilar and $Sim(s, t) = 1$ indicates that two servers are the same.

To calculate the similarity between users, similarly, based on normalizing service QoS data, the original user-service matrix US is also mapped into the row-normal user-service matrix where each row is in the range of $[0,1]$. The formula to measure the similarity between two users u and v is as follows:

$$Sim(u, v) = 1 - \frac{\sqrt{\sum_{i \in I} \left(\frac{r_{u,i} - r_{umin}}{r_{umax} - r_{umin}} - \frac{r_{v,i} - r_{vmin}}{r_{umax} - r_{umin}} \right)^2}}{\sqrt{|I|}}, \quad (11)$$

where $I = I_u \cap I_v$ is the set of services co-invoked by users u and v ; $|I|$ is the number of I ; $r_{u,i}$ is the QoS value of service i from user u in the user-service matrix P ; r_{umin} and r_{umax} denote the lowest and the highest values from user u in US , respectively; r_{vmin} and r_{vmax} denote the lowest and the highest values from user v in US , respectively. $Sim(u, v) = 0$ represents that two users are dissimilar and $Sim(u, v) = 1$ indicates that two users are the same.

2. Location-based Collaborative Filtering

Based on the normalized similarity measurement approach, we propose a Location-based Personalized QoS Prediction method. In our approach, the prediction value is calculated by the value of similar users and similar servers.

When predicting the unknown QoS value $\hat{r}_{u,i}$ of user u on service i , our approach recover the original scale of user u or service i . In the user-based QoS value prediction, we define our user-based CF as follows:

$$\hat{r}_{u,i} = r_{umin} + (r_{umax} - r_{umin}) \frac{\sum_{u' \in U} Sim(u, u') * nr_{u',i}}{\sum_{u' \in U} Sim(u, u')}, \quad (12)$$

where U denotes the set of similar users to user u , who have invoked service i ; $nr_{u',i}$ is the value of service i from user u' in the row-normal matrix US_{nu} ; r_{umin} and r_{umax} are the lowest and the highest values from user u in the original matrix US , respectively; and $Sim(u, u')$ can be computed by formula(11).

In the server-based value prediction, we first cluster servers based on K-means to reduce the influence of the servers sparsity. Then we create server-based CF to predict the unknown QoS value $\hat{r}_{s,i}$ of server s when provides service i , whose formula is given as:

$$\hat{r}_{s,i} = r_{smin} + (r_{smax} - r_{smin}) \frac{\sum_{s' \in S} Sim(s, s') * nr_{s',i}}{\sum_{s' \in S} Sim(s, s')}, \quad (13)$$

where S denotes the set of similar servers to server s , who have provided service j ; $nr_{s',j}$ is the value of service j provided by server s' in the row-normal matrix SS_{nu} ; r_{smin} and r_{smax} are the lowest and the highest values from server s in the original matrix SS , respectively; and $Sim(u, u')$ can be computed by formula(10).

To make use of the information from both similar users and similar servers, a parameter $\lambda (0 \leq \lambda \leq 1)$ is employed to determine how much does the prediction rely on user-based CF or server-based CF. Our approach makes prediction by employing the following equation:

$$\hat{r}_{s,u,i} = \lambda \times \hat{r}_{u,i} + (1 - \lambda) \times \hat{r}_{s,i}, \quad (14)$$

when $\lambda = 1$, the approach only uses information of similar users to make prediction. When $\lambda = 0$, it uses information of similar server for predicting the QoS data. When $0 < \lambda < 1$, it systematically combines the user-based CF approach and the item-based CF approach to fully utilize the information of both similar users and similar servers.

3. Service Recommendation

Based on the above trajectory prediction and location-based collaborative filtering approach(LCF) for an active user. The length of the predicted trajectory is not long in terms of the prediction accuracy. [27] shows that with the length of the predicted trajectory increases, the prediction accuracy will gradually decrease. After predicting an active user trajectory, the mobile computing platform will choose servers around the trajectory and predict the QoS data on these servers. We assume that the best position is the center of the predicted trajectory which is the best distance to provide services for the user. servers around the best (within b km radius) will be considered to provide good QoS data for the active user. The mobile edge computing platform will calculate the QoS data of the service on these servers based on LCF, and recommend the active user to use that on the new edge server with the best QoS data or still use the service provided by the current edge server. In this way, personalized service recommendations can be achieved without conducting the expensive and time-consuming real-world service invocations on the edge server. The process of mobility-aware personalized service recommendation(MPSR) for an active user is shown in Algorithm 2.

Algorithm 2 MPSR Algorithm

Input: an active user $u \in U$; History trajectories htr of u ; The service s invoked by u ; Trajectory length lag ; Prediction Trajectory seq ; Neighbor's search radius r

Output: Server with the best QoS

```

1:  $p\_servers = MobilityPrediction(u, htr, lag, seq)$  //Using the trained predicted locations model
   to obtain a prediction sequence of length  $seq$ 
2:  $center\_p = get\_center(p\_servers)$  // find servers center.
3:  $nb\_servers = get\_neighbors(center\_p, r)$  // Find neighbors near the center point within radius  $r$ .
4:  $QoS = dictionary()$  //Initialize a candidate set and save the QoS of each server.
5: for each  $s \in nb\_servers$  do
6:    $QoS[s] = Get\_QoS(u, b, s)$  //According to Eq.(14)
7: end for
8:  $best\_server = get\_best\_server(QoS)$  // find the best  $QoS$  server
9: return  $best\_server$ 

```

Results and Discussion

In this section, we present experiments to verify the performance of our approach and compare the results with other recommendation methods. Our experiments are intended to (1) validate the rationality of our proposed approach; (2) compare our approach with other trajectory prediction methods; (3) compare our approach with other recommendation methods without considering the mobility; and (4) analyze parameters of our approach to achieve optimum performance.

Dataset Description

In our experiments, we merge two real-world dataset on data fusion [28]: the WSDream dataset [16] and the Shanghai Telecom user connecting base station dataset [6]. The WSDream dataset contains the real-world QoS evaluation obtains

from 142 users on 4500 web services in 64 time intervals. The QoS evaluation results are the user-service matrices on response time and throughput which can be used as the user-service matrix in our approach. The Shanghai Telecom dataset contains the information about 6357 users invoked on 3233 base station. The base station is considered as the edge server in our experiment. We merge two datasets as [6] into a hybrid dataset, which contains both QoS data and edge server locations.

Evaluation Metric

Since our approach tries to predict top-K candidate positions for each user trajectory, we adopt the accuracy of top-K (ACC@K) to evaluate the performance. ACC@K is the percentage of accurate predictions, i.e., correctly predicts the following positions for users, which is calculated as:

$$ACC@K = \frac{|y \text{ in top-K}|}{N}, \quad (15)$$

To evaluate the QoS value prediction accuracy, we adopt the commonly used Mean Absolute Error(MAE) and Root Mean Squared Error(RMSE) as the accuracy metrics. MAE is defined as follows:

$$MAE = \frac{\sum_{u,i} |r_{u,i} - \hat{r}_{u,i}|}{N}, \quad (16)$$

where $r_{u,i}$ is the actual QoS value of user u using service i , $\hat{r}_{u,i}$ is the corresponding predicted QoS value, N is the total number of test samples. The RMSE is defined as follows:

$$RMSE = \sqrt{\frac{\sum_{u,i} (r_{u,i} - \hat{r}_{u,i})^2}{N}}, \quad (17)$$

the meaning of the $r_{u,i}$, $\hat{r}_{u,i}$ and N are the same as the formula (16). For the position samples, the smaller the values of RMSE are, the more accurate the prediction.

Performance Comparison on QoS Prediction

For the sake of demonstrating the performance and effectiveness of our location-based collaborative filtering recommendation approach(LCF). We compare the MAE and RMSE with other approaches. The other approaches are as follows:

- **User-Based Collaborative Filtering Recommendation Approach (UCF):** This approach is to predict QoS values by user-based prediction algorithm using Pearson correlation coefficient, employs similar users for service recommendation;
- **Service-Based Collaborative Filtering Recommendation Approach(SCF):** This approach is to predict QoS values by item-based prediction algorithm using Pearson correlation coefficient, employs similar web services for service recommendation;
- **Hybrid Collaborative Filtering Recommendation Approach (HCF):** This approach is to predict QoS values by combining the user-based and service-based prediction approach for service recommendation;

We make service recommendations in terms of qualities of service(response time and throughput) for randomly active users and compare our approach with others. The experimental results are shown in Table.1.

The results show that the MAE and RMSE values of our approach are all smaller than other approaches under different matrix density. We set the matrix density from 10% to 90%. The MAE and RMSE values increase gradually when the matrix density is set as 10% to 70%, which shows that prediction accuracy decreases. Because prediction results are in deviation when less information on similar users and services. The MAE and RMSE values increase dramatically when the matrix density increases 90%, because similar users and servers are rare. The MAE and RMSE values of our approach are smaller than other approaches, indicating that the prediction accuracy can be improved by our approach because we consider the volatility and the mobility. We take the similarities of both users and servers into consideration. In MEC environment, QoS data provided by different edge servers and various conditions, the QoS prediction provided by similar servers will eliminate volatility effectively. Besides, we normalize QoS data before calculating similarities.

Performance Comparison on Trajectory Prediction

For the sake of demonstrating the trajectory prediction accuracy of our hybrid LSTM-based mobility prediction(LSTMM) approach. We compare the ACC@K with *persistent*(benchmark), *Seq2Seq* [29] and *CNNM* [30].we randomly select 20% data from our dataset to verify the trajectory prediction accuracy and compare our approach with other methods.

Figure 4 shows the ACC@1 and ACC@5 values in different prediction trajectory length. The row labels $T + 1$, $T + 2$, and $T + 3$ express positions on the $(T + l)_{th}$ time in the predicted trajectory. The difference between prediction accuracy in each position is small because they are predicted at the same time. When the ACC@5 chooses the top 5 candidate position as the correct position, its value is much better than ACC@1, which only chooses the most likely position. Compared with other approaches, the ACC@1 and ACC@5 values of our approach are greater slightly. Compared with the *Seq2Seq* model only uses LSTM coding, the ACC@1 value of our approach increases by about 4% on average due to considering user characteristics and distances among servers. CNNM also results in poor prediction performance, because it is based on the convolutional network, which is unsuitable to solve the long-term dependence problem. Therefore, our approach takes advantage of long-term and short-term dependence and considers more information, which lead to more accurate results than other approaches.

Impact of recommendation strategies

Trajectory prediction plays an important role in service recommendation. We analysis the impact from two aspects, first we compare the recommended QoS data with our approaches to other recommending strategies:

- **Random Location Service Recommendation(RLSR):** Randomly select a server around the current server, it will be the candidate server providing the service at the next time.
- **Nearest Location Service Recommendation(NLSR):** Select the nearest server to the current server as the candidate server providing the service at the next time.

Impact of MES distribution

Figure 5 shows the distribution of historical trajectory points for three typical active users randomly selected from the dataset. The distribution of each active user historical trajectory point and edge servers are very different. For user 58, the distribution of both trajectory points and the edge servers are well-proportioned. The distribution of trajectory points is well-proportioned, and the distribution of edge servers are dis-proportioned for user 68. Both of the historical trajectory points of users 83 and the surrounding servers are distributed dis-proportioned. Then, we analysis the recommendation results for the three active users shown in Figure 6. The results show that the greater the QoS data provided, the greater the range of distances. There are also some slight fluctuations, which are caused by the dis-proportioned distribution of edge servers. The throughput increases obviously when the range of distance is more than 1 km, and the response time reduces when the range of distance is between 0.8 km to 2 km. On the whole, the QoS data changes are less obvious with radius increases after 2 km. Hence, we predict QoS data in the range between 0.8 km and 2 km for active users to examine which distance ranges provided the most accurate service recommendations.

Impact of parameter λ

Different geographical areas may have different data correlation characteristics. The parameter λ makes our prediction method more feasible and adaptable to different data sets. We fuse information from both similar users and similar servers to predict the missing value for active users. To study the impact of the parameter λ to our hybrid collaborative filtering method, we vary the value of λ from 0 to 1 with a step value of 0.1. If $\lambda = 1$, we only consider information from similar users, and if $\lambda = 0$, we only consider information from similar servers. MAE and RMSE are compared to study the effect of the λ for the prediction accuracy. The experimental results are shown in the Table. 2, which shows that MAE gains the best performance when $\lambda = 0.8$ and RMSE gains the best performance when $\lambda = 0.7$. It means that user-based collaborative filtering has a greater impact on the prediction results on our dataset. Because in our experimental data, the number of servers is larger than the number of users, the average number of samples per user is much larger than the number of samples corresponding to each server, which results in the sparsity of the server-service matrix. Therefore, λ will prefer to set 0.7 as the fusion factor in the experiment.

Conclusion

In this paper, we investigate the problem of service recommendation in Mobile Edge computing for frequent mobility scenarios. We propose a two-stage approach for predicting QoS values on mobile edge servers by systematically combining user mobility and location-based Collaborative filtering. We first present a hybrid LSTM-based mobility prediction approach to obtain an active user trajectory by jointly consider server information, user characteristics and distance information. Then a mobility-aware personalized service recommendation approach is proposed to acquire the QoS values based on the mobility results. We conduct simulation experiments to study the performance of the proposed work base on the real dataset. Large-scale

real-world experiments are conducted and the comprehensive experimental results show the effectiveness and feasibility of our approach.

In our future study, we will research on more trajectory prediction method, e.g., social GAN neural networks and ESN, can be employed to predict the future position of users to further improve the QoS values prediction. And other factors should be considered, e.g., the impact of time, when the service is recommended in the mobile edge computing.

Acknowledgements

This work was partially supported by the Major Scientific and Technological Projects of CNPC, China, under Grant ZD2019-183-004, and by the Fundamental Research Funds for the Central Universities, China, under Grant 20CX05019A.

Availability of data and materials

The details of experimental parameters are given in Section Experiment.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

Hongxia Zhang, Yanhui Dong, and Yongjin Yang conceived and designed the study. Yanhui Dong performed the simulations. Hongxia Zhang and Yongjin Yang wrote the paper. All authors reviewed and edited the manuscript. All authors read and approved the final manuscript.

Author details

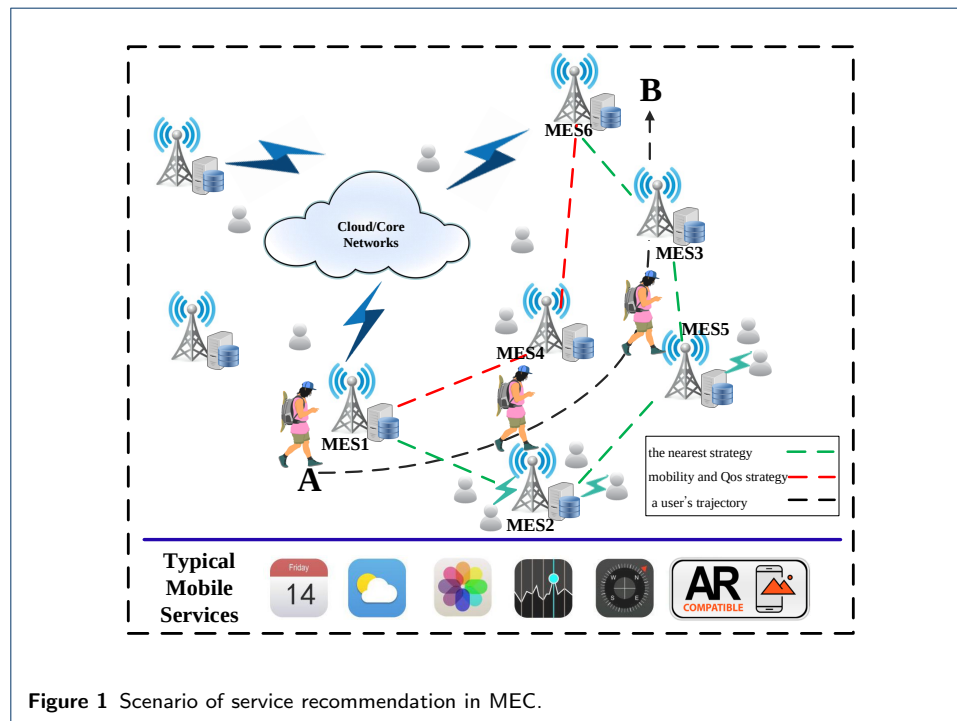
Department of Computer Science and Technology, China University of Petroleum, Qingdao, China.

References

1. Mach, P., Becvar, Z.: Mobile edge computing: A survey on architecture and computation offloading. *IEEE Communications Surveys Tutorials* **19**(3), 1628–1656 (2017). doi:10.1109/COMST.2017.2682318
2. Chiaraviglio, L., Cuomo, F., Gigli, A., Maisto, M., Zhou, Y., Zhifeng Zhao, Honggang Zhang: A reality check of base station spatial distribution in mobile networks. In: 2016 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), pp. 1065–1066 (2016). doi:10.1109/INFCOMW.2016.7562255
3. Shi, W., Cao, J., Zhang, Q., Li, Y., Xu, L.: Edge computing: Vision and challenges. *Internet of Things Journal*, *IEEE* **3**(5), 637–646 (2016)
4. Kai Peng, S.W. Hualong Huang, Leung, V.C.M.: End-edge-cloud collaborative computation offloading for multiple mobile users in heterogeneous edge-server environment. (2020). <https://doi.org/10.1007/s11276-020-02385-1>
5. Wang, Z., Zhao, Z., Min, G., Huang, X., Ni, Q., Wang, R.: User mobility aware task assignment for mobile edge computing. *Future Generation Computer Systems* **85**(AUG.), 1–8 (2018)
6. Wang, S., Zhao, Y., Huang, L., Xu, J., Hsu, C.H.: Qos prediction for service recommendations in mobile edge computing. *Journal of Parallel & Distributed Computing* **127**(MAY), 134–144 (2017)
7. Arshad, R., Elsayy, H., Sorour, S., Al-Naffouri, T., Alouini, M.-S.: Handover management in dense cellular networks: A stochastic geometry approach. (2016). doi:10.1109/ICC.2016.7510709
8. Li, B., Zhang, H., Lu, H.: User mobility prediction based on lagrange's interpolation in ultra-dense networks. In: *IEEE International Symposium on Personal* (2016)
9. Sun, H., Zheng, Z., Chen, J., Lyu, M.R.: Personalized web service recommendation via normal recovery collaborative filtering. *IEEE Transactions on Services Computing* **6**(4), 573–579 (2013)
10. Li, Y., Guo, Y.: Cultural distance-aware service recommendation approach in mobile edge computing. *entific programming* **2018**(PT.1), 2181974–121819748 (2018)
11. Rutkowski, T., Romanowski, J., Woldan, P., Staszewski, P., Rutkowski, L.: A content-based recommendation system using neuro-fuzzy approach. In: 2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) (2018)
12. Yao, L., Sheng, Q.Z., Ngu, A.H.H., Yu, J., Segev, A.: Unified collaborative and content-based web service recommendation. *IEEE Transactions on Services Computing* **8**(3), 453–466 (2015)
13. Nice, N., Koenigstein, N., Ben-Elazar, S., Keren, S., Paquet, U., Finkelstein, Y.: Multilingual content based recommendation system (2018)
14. Shao, L., Zhang, J., Wei, Y., Zhao, J., Mei, H.: Personalized qos prediction forweb services via collaborative filtering. In: *IEEE International Conference on Web Services* (2007)
15. Zheng, Z., Ma, H., Lyu, M.R., King, I.: Wsrec: A collaborative filtering based web service recommender system. In: *IEEE International Conference on Web Services* (2009)
16. Zheng, Z., Ma, H., Lyu, M., R., King, I.: Qos-aware web service recommendation by collaborative filtering. *Services Computing*, *IEEE Transactions on* (2011)
17. Zheng, Zibin, Ma, Hao, Lyu, Michael, R., King, Irwin: Collaborative web service qos prediction via neighborhood integrated matrix factorization. *IEEE Transactions on Services Computing* **6**(3), 289–299 (2013)
18. Logesh, R., Subramaniaswamy, V., Vijayakumar, V.: A personalised travel recommender system utilising social network profile and accurate gps data. *Electronic Government An International Journal* **14**(1) (2018)

19. Wang, H., Terrovitis, M., Mamoulis, N.: Location recommendation in location-based social networks using user check-in data. In: Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. SIGSPATIAL'13, pp. 374–383. Association for Computing Machinery, New York, NY, USA (2013). doi:10.1145/2525314.2525357
20. Park, M.-H., Hong, J.-H., Cho, S.-B.: Location-based recommendation system using bayesian user's preference model in mobile devices. In: Proceedings of the 4th International Conference on Ubiquitous Intelligence and Computing. UIC'07, pp. 1130–1139. Springer, Berlin, Heidelberg (2007)
21. Zheng, V.W., Cao, B., Zheng, Y., Xie, X., Yang, Q.: Collaborative filtering meets mobile recommendation: A user-centered approach. In: Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence. AAAI'10, pp. 236–241. AAAI Press, Menlo Park, CA (2010)
22. X.Chen, Z.Zheng, Q.Yu, M.R.Lyu: Web service recommendation via exploiting location and qos information. IEEE Transactions on Parallel and Distributed Systems **25**(7), 1913–1924 (2014). doi:10.1109/TPDS.2013.308
23. Tang, M., Jiang, Y., Liu, J., Liu, X.: Location-aware collaborative filtering for qos-based service recommendation. In: 2012 IEEE 19th International Conference on Web Services, pp. 202–209 (2012). doi:10.1109/ICWS.2012.61
24. E.Ben Zion, B.Lerner: Identifying and predicting social lifestyles in people's trajectories by neural networks. EPJ Data Sci **45**(7) (2018)
25. Schmidt-Hieber, J.: Nonparametric regression using deep neural networks with ReLU activation function. arXiv e-prints, 1708–06633 (2017). 1708.06633
26. Ramachandran, S., Deshpande, O., Roseman, C.C., Rosenberg, N.A., Feldman, M.W., Cavalli-Sforza, L.L.: Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in africa. Proceedings of the National Academy of Sciences of the United States of America **102**(44), 15942–15947 (2005)
27. Arshad, R., ElSawy, H., Sorour, S., Al-Naffouri, T.Y., Alouini, M.-S.: Handover Management in Dense Cellular Networks: A Stochastic Geometry Approach. arXiv e-prints, 1604–08552 (2016). 1604.08552
28. Khaleghi, B., Khamis, A., Karray, F.O., Razavi, S.N.: Multisensor data fusion: A review of the state-of-the-art. Inf. Fusion **14**(1), 28–44 (2013). doi:10.1016/j.inffus.2011.08.001
29. Karatzoglou, A., Jablonski, A., Beigl, M.: A seq2seq learning approach for modeling semantic trajectories and predicting the next location. In: Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. SIGSPATIAL '18, pp. 528–531. Association for Computing Machinery, New York, NY, USA (2018). doi:10.1145/3274895.3274983
30. Lv, J., Li, Q., Wang, X.: T-CONV: A Convolutional Neural Network For Multi-scale Taxi Trajectory Prediction (2017). 1611.07635

Figures



Tables

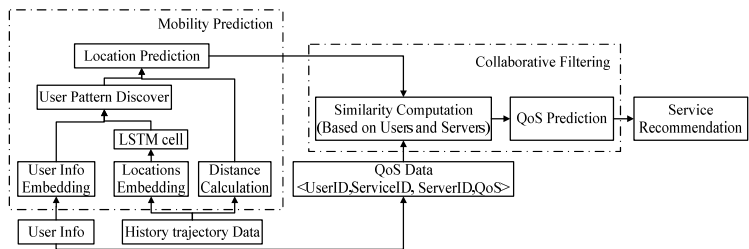


Figure 2 Framework of mobility-aware personalized service recommendation approach.

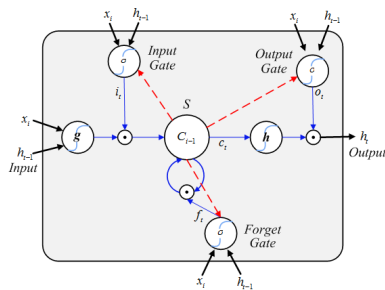


Figure 3 Network architecture of the LSTM predictor.

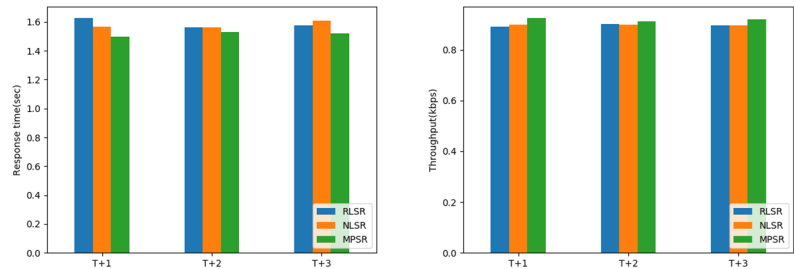


Figure 4 Impact of different recommending strategies on QoS.

Table 1 Impact of data sparsity on prediction accuracy

	methods	10%		30%		50%		70%		90%	
		RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
Response time	UCF	3.9938	1.8100	4.0156	1.8179	4.0232	1.8260	4.0387	1.8337	4.0610	1.8665
	SCF	3.4903	1.7540	3.5075	1.6987	3.5328	1.6317	3.5789	1.5558	3.6934	1.5177
	TCF	3.6131	1.4923	3.6648	1.4838	3.7054	1.4795	3.7644	1.4994	3.8547	1.5740
	LCF	<u>2.5627</u>	<u>1.2895</u>	<u>3.0548</u>	<u>1.4132</u>	<u>3.2744</u>	<u>1.4561</u>	<u>3.2966</u>	<u>1.4820</u>	<u>3.4647</u>	<u>1.7172</u>

Table 2 The effect of lambda taking different values on prediction accuracy

Response time	λ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
	MAE	1.7542	1.6131	1.5177	1.4550	1.4168	1.3853	1.3620	<u>1.3413</u>	1.3462	1.3501
	RMSE	3.8204	3.5451	3.2867	3.0712	2.9005	2.7829	<u>2.7253</u>	2.7314	2.8009	2.9292

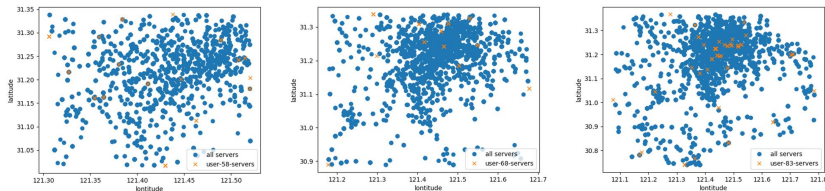


Figure 5 Historical track point distribution and neighbor distribution for different users.

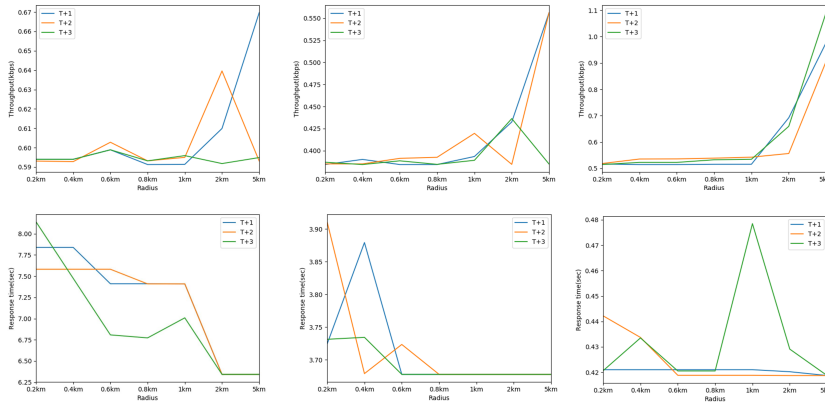


Figure 6 The influence of the choice of neighbor radius on the user level.