A Novel Variable Order Markov Model Based Wireless Service Prediction Algorithm with User Similarity*

Juan Chen, Hui Zhang*, Longxiang Yang, Hongbo Zhu

Jiangsu Key Lab of Wireless Communications, Nanjing University of Posts and Telecommunications
Nanjing 210003, China

Abstract

More and more researches focus on wireless service prediction problems with the rapid development of wireless network technology. It is necessary to investigate user service behaviors to improve user QoE. Therefore, in this paper, a wireless service prediction algorithm based on variable order Markov model is proposed to predict user service behaviors in wireless networks. In this algorithm, firstly, a user clustering method with user similarity is presented to categorize users. Then, a variable order Markov model based on Probabilistic Suffix Tree (PST) is introduced to predict the future service states of users. Finally, an adaptive fusion method of prediction values is given to improve the prediction accuracy. Simulation results show the effectiveness of the proposed algorithm.

Keywords: User Service Behaviors; Prediction Algorithm; User Similarity; Variable Order Markov Model

1 Introduction

Recently, with the rapid development of wireless network technology, the quality and amount of wireless services have increased significantly. How to provide services for users from wireless services efficiently and guarantee high user QoE attract more and more attention. Network heterogeneity, ubiquitous computing and personalized services have become main characteristics of wireless networks. By analyzing hobbies, behavioral characteristics, and service requirements of a user, the personalized services can be provided to this user.

Therefore, it is significant to do research on wireless service prediction problems. By studying user service behaviors, we can predict the occurrence time, occurrence location, type, content of each service [1, 2]. Heo et al. [3] proposed a novel user demand prediction method, which was

*This work is supported by the National Natural Science Foundation of China (No. 61471203, 61101105), Specialized Research Fund for the Doctoral Program of Higher Education (No. 20113223120001), National Science & Technology Major Projects (No. 2012ZX03001008-003), Jiangsu 973 Project (No. BK2011027), the Research Innovation Program for College Graduates of Jiangsu Province (No. SJJX.0370), and the Priority Academic Program Development of Jiangsu Higher Education Institutions - Information and Communication Engineering.

*Corresponding author.
Email address: zhhjoice@126.com (Hui Zhang).
applied in resource allocation. Considering that users have similar social background or personal preference, Hsu et al. and Eagle et al. [4, 5] presented a group behavior prediction method. Moreover, according to the impact of “Circle of Friends” on user service behaviors, Kubo et al. [6] proposed a social network environment oriented prediction algorithm. By studying and predicting the user view behavior, Chen et al. [7] developed a smart streaming strategy that can significantly improve streaming service quality.

Several researchers have proposed methods for user behavior prediction problems. Ahmed et al. [8] proposed a novel method based on Hidden Markov Models (HMM) to predict web service behaviors, which further suggests an optimal path for the execution of user requests. Vasquez et al. [9] proposed a growing hidden Markov model where motion patterns can be learned incrementally and in parallel with prediction. Awad et al. [10] proposed a new modified Markov model to alleviate scalability issues, and presented a new two-tier prediction framework to predict web browsing behaviors. Haitham Abu-Ghazaleh et al. [11] presented a Markov renewal process for mobility modeling, so as to predict the probability of the next-cell transition for an arbitrary user. However, recently, there are few wireless service prediction references considering the concept of user cluster.

Therefore, a wireless service prediction algorithm is proposed in this paper. Compared with the traditional algorithms, this proposed algorithm performs better in accuracy. This paper is organized as follows (see Fig. 1). Section 2 describes the concept of user similarity and presents a user clustering method. Section 3 builds a variable order Markov model and an adaptive fusion method. Section 4 describes simulation results and performance analyses. Section 5 concludes this paper.

![Fig. 1: The algorithm architecture](image)

## 2 A User Clustering Method Based on User Similarity

### 2.1 Definition of User Similarity

Users who have similar hobbies, behaviors and requirements may choose similar services in wireless networks. Therefore, when predicting wireless service behaviors of a user, we can refer to prediction results of the other users. In order to improve the prediction accuracy, this paper presents a clustering method based on user similarity. The definition of user similarity is given as below.
User similarity is defined as the similar degree of self-characteristics and service behavior characteristics between users. \( D = \{d_1, d_2, \ldots d_k\} \) describes self-characteristics of a user, including age, gender, occupation, etc. \( S = \{s_1, s_2 \ldots s_n\} \) describes service behavior characteristics. Based on these characteristics, users can be categorized effectively.

### 2.2 User Similarity Calculation

User self-characteristic is considered as an important factor in calculating user similarity. The similarity with self characteristics between user \( m_1 \) and user \( m_2 \) is given in Eq. (1, 2), \( D = \{d_{m_1}, d_{m_2}, \ldots d_{m_j}\} \), where \( d_{m_j} \) is the \( j \)th characteristic value of the user \( m \), and \( k \) is the number of characteristics types.

\[
Self(m_1, m_2) = \frac{\sum_{j=1}^{k} cha(d_{m_1j}, d_{m_2j})}{k} \tag{1}
\]

\[
cha(d_{m_1j}, d_{m_2j}) = \begin{cases} 
1 & d_{m_1j} = d_{m_2j} \\
0 & d_{m_1j} \neq d_{m_2j}
\end{cases} \tag{2}
\]

For user \( m \), the service support degree \( u_{m;i} \) is defined as the user’s preference for service \( i \), seen in Eq. (3, 4). The influence factors include using time \( t \) and using frequency \( f \), where \( t_{m;i} \) is the proportion of using time to total time for service \( i \), and \( f_{m;i} \) is the proportion of the using times to total times for service \( i \).

\[
u_{m;i} = \sqrt{t_{m;i}^2 + f_{m;i}^2} \tag{3}
\]

\[
t_{m;i} = \frac{\sum_{t_i \in T_i} t_i}{\sum_{T_i \in T} \sum_{t_i \in T_i} t_i} \quad f_{m;i} = \frac{\sum_{f_i \in F_i} f_i}{\sum_{F_i \in F} \sum_{f_i \in F_i} f_i} \tag{4}
\]

The similarity with service behavior characteristics between use \( m_1 \) and user \( m_2 \) can be shown in Eq. (5), where \( n \) is the number of service types and \( u_{m_1;i} \) is the service support degree.

\[
Beh(m_1, m_2) = \frac{\sum_{i=1}^{n} u_{m_1;i} u_{m_2;i}}{\sqrt{\sum_{i=1}^{n} u_{m_1;i}^2} \sqrt{\sum_{i=1}^{n} u_{m_2;i}^2}} \tag{5}
\]

Therefore, the user similarity between use \( m_1 \) and user \( m_2 \) can be derived by analyzing self-characteristics and service behavior characteristics. The improved user similarity \( Sim(m_1, m_2) \) is given in Eq. (6), where \( u_{m;i} \) is the service support degree, and \( Self(m_1, m_2) \) is the similarity with self-characteristics.

\[
Sim(m_1, m_2) = \frac{\sum_{i=1}^{n} u_{m_1;i} u_{m_2;i} Self(m_1, m_2)}{\sqrt{\sum_{i=1}^{n} u_{m_1;i}^2 Self(m_1, m_2)} \sqrt{\sum_{i=1}^{n} u_{m_2;i}^2 Self(m_1, m_2)}} \tag{6}
\]
2.3 The User Clustering Method Flow

According to user similarity, users can be categorized into different clusters. Therefore, a user clustering method based on fuzzy clustering is proposed. In this clustering method, every user has a membership degree to each cluster. The user clustering method flow is shown as follows (see Fig. 2):

1. Data standardization: there are \( m \) indexes to express a user’s characteristics containing self-characteristics and service behavior characteristics. These initial data should be normalized to \([0, 1]\).

2. Building fuzzy similar matrix: based on self-characteristics and service behavior characteristics, we calculate user similarity \( Sim_{ij} = R(m_i, m_j) \) between each two users. Then the fuzzy similar matrix can be built in Eq. (7).

\[
Sim = \begin{bmatrix}
Sim(1,1) & Sim(1,2) & \ldots & Sim(1,m) \\
Sim(2,1) & Sim(2,2) & \ldots & Sim(2,m) \\
& \vdots & \ddots & \vdots \\
Sim(m,1) & Sim(m,2) & \ldots & Sim(m,m)
\end{bmatrix}
\]  

(7)

3. Calculating the transitive closure of matrix \( Sim \): when matrix \( Sim \) is reflexive, symmetric and transitive, the transitive closure \( R = t(Sim) \) can be computed by using the square method, which is also called fuzzy equivalent matrix.

4. Dynamic clustering: the \( \lambda \) cut-matrix \( R_\lambda \) can be obtained by confidence level \( \lambda \), \( \lambda \in [0, 1] \), then users can be categorized according to the \( \lambda \) cut-matrix.

Fig. 2: Flow chart of the user cluster method
3 A Wireless Service Prediction Algorithm

A wireless service prediction algorithm based on variable order Markov model is proposed in this paper. Variable Order Markov (VOM) models are an important class of models that extend the well known Markov chain models. Its prediction accuracy is approximate to high-order Markov models, while its complexity is similar to low-order Markov models. Furthermore, a fusion prediction method is presented in this paper. Fusing with the other user models which have high similarities, this method improves the prediction accuracy effectively.

3.1 Variable Order Markov Prediction Model (VOMM)

In contrast to the Markov chain models, where each variable in a sequence depends on a fixed number of variables, in VOM models this number of variables is unfixed. The VOM models are also called context trees, in this paper a model called Probabilistic Suffix Tree (PST) has been proposed. Probabilistic Suffix Tree (PST) is a model of probability distribution for the event occurring in sequences. And it is used to predict the service state by calculating probability distribution vector. Each node of PST has probability distribution vector which denotes the probability of next node appearance.

A PST is a non-empty tree, where each edge in the tree is labeled by a single state of the service sequence. Therefore the service sequence is labeled on the PST by going up the tree from the node to the root. The flow of Build-PST \((p_{\text{min}}, \gamma, r, L)\) is as follows:

1. The initial PST only includes one root node, where the probability vector denotes the frequency of each symbol occurring in total symbols. If the frequency of symbols exceeds threshold value \(p_{\text{min}}\), these symbols become candidate child nodes.

2. The recursive expansion of candidate child nodes.
   (a) Calculate the probability vector of the following candidate nodes.
   (b) Mark one candidate node as string \(s\). If there exists \(\sigma \in \sum, p(\sigma | s)/p(\sigma | \text{suffix}(s)) \geq r\) or \(p(\sigma | s)/p(\sigma | \text{suffix}(s)) \geq 1/r\), this node can be added to the tree. The parent node is the marked as string \(\text{suffix}(s)\).
   (c) If the depth of this node is less than the maximum depth \(L\) of PST, and the occurring probability of \(\sigma s\) exceeds \(p_{\text{min}}\), for every \(\sigma \in \sum\), the node marked as \(\sigma s\) is an candidate node of this node.

According to the history service states of users, PST could predict the future service states. In Fig. 3, \(P = \{p_1, p_2, \ldots, p_n\}\) describes the probability vector of each node, and \(S = \{s_1, s_2, \ldots, s_n\}\) describes types of service states. With the given PST tree \(T\) and sequence of history services, one node can traverse the whole PST from the root node \(R\). Firstly, at time \(t\), the root node \(R\) chooses the child tree according to the current state. Then, it traverses the child tree according to the state at time \(t-1\). When the traversing of history states is finished, the probability vector on the leaf is the probability of each service state in next time. Therefore, the future service states can be predicted by judging the threshold value \(I_{\text{min}}\).
After building PST, VOM model is used to predict service states. In D order Markov model, when the length of service sequence $q$ is $m$ and the length of service sequence $s$ is less than $L$, the probability of service sequence $s$ occurring in the sequence $q$ is given in Eq. (8, 9).

$$P(s) = \frac{1}{m - D + 1} \sum_{j=L}^{m} x_j(s) \quad (8)$$

$$x_j(s) = \begin{cases} 1, & \sigma_{j-|s|+1} \ldots \sigma_j = s \\ 0, & \sigma_{j-|s|+1} \ldots \sigma_j \neq s \end{cases} \quad (9)$$

Then, the probability of service state $\sigma$ occurring behind service sequence $s$ is given in Eq. (10)

$$P(\sigma|s) = \frac{\sum_{j=L}^{m-1} x_{j+1}(s\sigma)}{\sum_{j=L}^{m} x_j(s)} \quad (10)$$

For one user, we can predict the probability of each service state at time $t+1$, according to the history service states before time $t$. In all probabilities of service states at time $t+1$, the maximum probability is $P_{t+1}^{\text{max}}$. If it exceeds the threshold value $P_{t+1}^{\text{min}}$ ($P_{t+1}^{t+1} > P_{t+1}^{\text{min}}$), the corresponding service state is the predicted state.

### 3.2 An Adaptive Fusion Method

In a fusion method of prediction values, $M_p$ is VOM model of the predicted user, and $\{x_i\}_{i=1}^{t}$ is learning set. Users are chosen from learning set, and corresponding models are $\{M_j\}$, $j = 1, 2, \ldots, n$. Therefore, the fusion model can be built in Eq. (11).

$$M_c = \lambda_0 \cdot M_p + \lambda_1 \cdot M_1 + \lambda_2 \cdot M_2 + \cdots + \lambda_n \cdot M_n$$

$$\sum_{i=0}^{n} \lambda_i = 1 \quad (11)$$
Assumed that \( m_c(x_i) \) is the prediction value and \( Y_i \) is the observed value, \( E \) should be minimized:

\[
E = \sum_{i=1}^{t} |m_c(x_i) - Y_i|^2
\]

(12)

Therefore, the adaptive fusion method is built as follows:

1. From learning set, we choose \( n \) users who are similar to the predicted user, their \( \text{Sim}(m, m_j) \) should exceeds threshold value \( v(\text{Sim}(m, m_j) \geq v) \);
2. Random values from \([0, 1]\) are given to \( \{\lambda_j\} \) and then to be normalized; \( x_i \) is the service state, \( m_{ji} \) is each branch prediction value, \( m_c(x_i) \) is the fusion prediction value, and \( Y_i \) is the observed value.
3. Calculating the distance between \( m_{ji} \) and \( Y_i \)

\[
d_{ji} = |m_{ji} - Y_i|^2, j = 1, 2, \cdots, N
\]

(13)

4. \( d_{ji} \) is sorted from small to large; choose the largest \( d_{ji} \) and smallest \( d_{ji} \), and put corresponding \( \lambda_j \) into a same group; Then choose the \( j_{th} \) largest \( d_{ji} \) and \( j_{th} \) smallest \( d_{ji} \), and put corresponding \( \lambda_j \) into a same group. To the end, all \( \lambda_j \) are put in groups.
5. For \( \lambda_j \) in the same group, the bigger \( \lambda_j \) minus \( \eta(t)\lambda_j \) and the smaller \( \lambda_j \) plus \( \eta(t)\lambda_j \), all \( \lambda_j \) can be modified accordingly. \( \eta(t) \) is given in Eq. (14), where \( t \) is learning times, \( T \) is the total learning times and \( \eta(0) \) is the initial value.

\[
\eta(t) = \eta(0)(1 - t/T)
\]

(14)

6. Studying all service states and setting the learning time \( t = t + 1 \), till \( t = T \), we can get \( \lambda_j \).
7. Calculating the difference value between \( \lambda_j \) and normalized \( \text{Sim}(m, m_j) \) in each model, if the value exceeds \( u \), return to step 5 and continue to study.

To the end, we get accurate \( \{\lambda_j\} \), and the fusion method has been built to predict user behavior services. This fusion prediction model would improve the prediction accuracy.

4 Simulation Analysis

In order to verify the performance of the wireless service prediction algorithm, we use the real user data to analyze and predict user service behaviors. We collect wireless service behaviors of 20 users in 30 days. Wireless service can be divided into 4 categories: streaming media services, download services, burst services, and session services. We set parameters for \( p_{\text{min}} = 0.001, \gamma = 0, r = 1.05 \). Precision and recall rate are shown in Eq. (17, 18), where \( R \) is the number of correct prediction, \( W \) is the number of incorrect prediction, \( Q \) is the number of requested prediction, and \( U \) is the number of correct requested prediction.

\[
\text{precision} = \frac{R}{R + W} \times 100\%
\]

(15)

\[
\text{recall rate} = \frac{U}{Q}
\]

(16)
As shown in Fig. 4, the prediction precision of three models is improved and reaches a maximum value when $order = 2$. Then the precision of traditional Markov model is decreased, while VOM model and fusion VOM model always keep steadily with the increase of order. And the precision of fusion VOM model is higher than VOM model. The precision of Variable Order Markov (VOM) model is decreased with the increase of recall rate. And the precision of fusion VOM mode is better than the other two model’s, reaching 81%. Experiments show that this fusion prediction model would improve the prediction accuracy.

![Fig. 4: Prediction precision with order](image1)

![Fig. 5: Prediction precision with recall rate](image2)

Compared with non-cluster Markov model, cluster Markov model shows the effectiveness on prediction (see Fig. 6). With the increase of sequence length, the prediction of two models is increased, and no-cluster model shows the higher precision than cluster model. But with the sequence becoming longer, the precision of the cluster model is increased obviously. Therefore it can be verified that clustering Markov model based on user similarity will improve prediction precision effectively. As shown in Fig. 7, traditional Markov model keeps lower stable precision,
Fig. 6: Prediction accuracy with sequence length

Fig. 7: Prediction accuracy with sequence length

because this model is not affected by sequence length of services. And the precision of VOM and fusion VOM model is increased with the growth of sequence length. In the fusion VOM model, the prediction precision is better than VOM model, which shows that this algorithm obviously improve the prediction accuracy.

5 Conclusion

A wireless service prediction algorithm based on variable order Markov model is proposed in this paper. Compared with the traditional prediction algorithms, the proposed algorithm improves the accuracy of prediction, and the simulation results show the effectiveness. Space complexity of this algorithm is \(o(n)\) and time complexity is \(o(n^2)\). Therefore, how to reduce time complexity of this algorithm will be a future research direction.

References


[6] H. Kubo, R. Shinkuma, T. Takahashi, Demand prediction based on social context for mobile content services, IEEE International Conference on Communications Workshops (ICC), 2011, 1-5


