A Method of Opportunity Prediction in Mobile Ad Hoc Network

Zhou Chen¹ and Chao Sun²

ABSTRACT

The mobility of nodes in mobile ad hoc network (MANET) makes great changes in network topology. And limited knowledge of the future encounter opportunity leads to a blind and unpredictable packet forwarding behavior in routing decisions. Finding the regularity in topology changes and makes efficient opportunity prediction is the key to routing in mobile ad hoc network. To this end, this thesis proposed KROP, a kernel regression opportunity prediction method in MANET. In KROP, we first extract Adamic-Adar metric and contact frequency metric to form features of node pairs, and use these features to capture the evolution of the local network topology over time. Then, we use kernel regression estimation method to model the historical evolution of the topology and output the probability of a future encounter. According to the given sequence of the probability, we make opportunity prediction of the future. In the comparison experiments on world datasets, we eventually proved KROP outperforms on prediction accuracy.

Keywords—mobile ad hoc network, opportunity prediction, similarity metric; kernel regression; prediction accuracy

I. INTRODUCTION

In recent years, with the development of wireless communication technology and the emergence of a large number of intelligent mobile terminals, the mobile ad hoc network (MANET) [1] research has received extensive attention in academia. This paper conducts research on the opportunity prediction technology in mobile ad hoc network. The difficulty of routing in MANET lies in the uncertainty of future encounter opportunity knowledge, leading to blind and unpredictable message forwarding behavior. Most of the existing research work is based on the assumption of static or historical data coverage, instead of a dynamic topology. Therefore, this paper aims to solve the problem of learning and mining the topology change rules between nodes and giving a more accurate instantaneous probability value to make opportunity prediction in the dynamically changing scene of MANET. To this end, this paper proposed KROP, a kernel regression opportunity prediction method in MANET. In KROP, a feature vector of the node pair is extracted from the network topology to represent the evolution of the topology

¹Email: chenzhou@nrec.com, ²Email: sunc@nrec.com
over time. Then, we use the kernel regression estimation method to model the historical change of the topology and calculate the probability of a future encounter opportunity. Finally, the experimental results verified that KROP has a better performance in predicting accuracy and can provide routing knowledge guidance in MANET.

This paper first introduces the topology characteristics of the MANET, and compares the advantages and disadvantages of existing related work. Then we will introduce the kernel regression opportunity prediction method proposed in this paper, and verify the effectiveness of the prediction method through the actual data set experiment.

II. RELATED WORK

In MANET, nodes use the store-carry-forward routing method to deliver the messages while they are moving and encountering each other. The mobility of nodes and the time-varying links lead to delays in delivery and difficulty in routing. Opportunity prediction refers to predicting the possibility of a node establishing a link with another one at the next moment by using known network topology information or historical evolution information of network topology. If we can predict the network topology at the next moment, the blindness of the routing behavior will be reduced greatly. Similarity based prediction is a typical link prediction method, indicators include: CN [2] (Common Neighbors), AA [3] (Adamic-Adar), Jaccard [4], RA [5] (Resource Allocation). These indicators represent the closeness of a pair of nodes, the higher the indicator values are, the higher the encounter probability between nodes will be. MaxProp [6] routing method defines the encounter probability associated with the number of encounters in last interval and selects the best path with the highest probability. Prophet [7] uses a probabilistic prediction mechanism to evaluate the effectiveness of the forwarding, and the probability of encountering between nodes is constantly updated as the topology changes. The experiment found that the direct use of these indicators for prediction does not performing well, or can’t capture the instantaneous changes in encounter probability.

This paper aims to solve the problem of learning and mining the dynamic change rules between nodes and predict the future encounter opportunity by calculating a more accurate instantaneous probability value. So, this paper proposed a kernel regression opportunity prediction method, KROP.
III. KERNEL REGRESSION ALGORITHM

A. Topology Snapshot

In order to be closer to the actual scene, this paper selects the measured Trace data set to describe the changes of the network topology. This data set contains appearing and disappearing records of links between nodes at a certain time. To capture the evolution of the network topology, we divide network topology changes into a series of topology snapshots \( \{ G_1, G_2, G_3, \ldots, G_t \} \).

Each snapshot \( G_t \) has a weighted symmetric adjacency matrix \( A \) on the time slot \( t \), and the element \( A[i][j] \) in the matrix represents the number of times node \( i \) encountered node \( j \) in this time slot. In Figure 1, the circular dots represents moving nodes, the edge represents encounter times that appears in the time slot.

Given a series of network snapshots \( \{ G_1, G_2, \ldots, G_t \} \), it is desirable to be able to predict the snapshot of the next time slot network topology \( G_{t+1} \). In fact, the snapshot matrix is a simple accumulation of the number of links in a given time period, related to the length of slot. According to the prediction mode of complex networks, this paper assumes that the possibility of establishing links between nodes in the future is related to the instantaneous local network topology. That is to say, whether a node establishes a link with another in \( G_{t+1} \) is related to the local network topology in \( G_t \). When the topology evolution of the node pair is captured, the next snapshot of the node pair can be predicted, or the forwarding opportunity in future can be predicted.

B. Feature Vector

Feature vector of node pair \( <i,j> \) in time slot \( t \):

\[
f_t(i, j) = \{ AA_t(i, j), CF_t(i, j) \}
\]

(3.1)

\[
AA_t(i, j) = \sum_{n \in \text{link}(i)} \frac{1}{\lg k(n)}
\]

(3.2)

\[ CF_t(i, j) = A[i][j] \]
AA\( (i,j) \) is similarity indicator based on common neighbor, AA (Adamic-Adar) indicator, in which \( \Gamma(i) \) and \( \Gamma(j) \) represent the neighbor set of node \( i \) and node \( j \), \( k(n) \) indicating the degree of the common neighbor \( n \), or the number of nodes that establish a link with node \( n \). This indicator assigns a degree-related weight to the common neighbor, and the weight is equal to the reciprocal of the logarithm of the degree of the neighbor node. Neighbors with a greater degree will be more active and have limited contribution to the similarity. It is similar to finding a friend relationship in the real world scenario, interests of friends should be highly relevant. However, if they have an intersection due to a wide interest and wide attention, we could not say they are friends. The meaning of choosing the feature vector \( AA\( (i,j) \) \) is that the similarity between node \( i \) and node \( j \) can be well characterized, which can be used to predict the possibility of encounter opportunity.

\( CF_t(i,j) \) is a commonly used indicator in route metrics, indicating the encounter frequency between nodes. If a pair of nodes has been in frequent contact in the past, we believe that they will be more likely to meet in the future. In this paper, the feature quantity \( CF_t(i,j) \) is used to describe the degree of closeness between nodes.

C. Kernel Regression Prediction Model

The opportunity prediction model is based on the attribute information of the link endpoint and its neighbors, aims to estimate the probability of link occurrence and make predictions. Assume that a set of snapshots \( \{G_1,G_2,...,G_t\} \) was observed in consecutive time slots. If link \( <i,j> \) exists in snapshot \( G_t \) in time slot \( t \), \( Y_t(i,j) = 1 \), otherwise \( Y_t(i,j) = 0 \). And the prediction model is defined:

\[
Y_t+1(i,j) | G \sim Bernoulli\left(p\left(\xi_t(i,j)\right)\right)\]  
\[
\xi_t(i,j) = \{f_t(i,j), E_t(i)\} \]  

(3.3)

The probability \( 0 \leq p(\cdot) \leq 1 \) of a link established in time slot \( t+1 \) obeys the Bernoulli distribution. \( p(\cdot) \) is a function about \( \xi_t(i,j) \). \( E_t(i) \) is a local topology evolution data set of node \( i \), defined as:

\[
E_t(i) = \{\beta_{\xi_t} (f), \beta_{\xi_t'} (f) \forall f \in F\} \]  

(3.4)

In which \( \beta_{\xi_t} (f) \) is the number of node pairs with feature vector \( f \) in the neighbor nodes of node \( i \) in time slot \( t-1 \). \( \beta_{\xi_t'} (f) \) is the number of node pairs which have established links at the next time slot \( t \). This data set indicates a topology evolution from time slot \( t-1 \) to time slot \( t \).

In order to capture the historical evolution process and describe the influence of different evolution processes on the probability of the topology, this paper uses a kernel regression method, assign different weights to the historical evolution process of different adjacent time slots, according to the
similarity between the current node local structure and the historical one. For the estimation of probability \( p(\xi(i,j)) \), this paper chooses the Nadaraya-Watson kernel regression estimation method [10]:

\[
m_h(x) = \frac{\sum_{i=1}^{n} K_h(x - X_i) Y_i}{\sum_{i=1}^{n} K_h(x - X_i)}
\] (3.5)

And \( p(\xi(i,j)) \) is defined as:

\[
p(\xi(i,j)) = \frac{\sum_{i,j,j'} Sim(\xi(i,j),\xi(i',j')) \cdot Y_{\xi_{ij}}(i',j')}{\sum_{i,j,j'} Sim(\xi(i,j),\xi(i',j'))}
\] (3.6)

\( Sim(\cdot,\cdot) \) characterizes the similarity between two variables, defined as:

\[
Sim(\xi(i,j),\xi(i',j')) = K(E(i),E(i')) \cdot I\{f_{ij} = f_{i'j'}\}
\] (3.7)

In which \( I \) equals 0 or 1, indicates whether the feature vector \( f \) matches the target one. Equation (3.6) will be:

\[
p(\xi(i,j)) = \frac{\sum_{i,j,j'} K(E(i),E(i')) \cdot \sum_{i,j,j'} I\{f_{ij} = f_{i'j'}\} \cdot Y_{\xi_{ij}}(i',j')}{\sum_{i,j,j'} K(E(i),E(i')) \cdot \sum_{i,j,j'} I\{f_{ij} = f_{i'j'}\}}
\]

\[
= \frac{\sum_{i,j,j'} K(E(i),E(i')) \cdot \beta_{ij} \cdot f_{ij}}{\sum_{i,j,j'} K(E(i),E(i')) \cdot \beta_{ij} \cdot f_{ij}}
\] (3.8)

\( K \) is a kernel function, which captures the degree of similarity between two parameters. This paper chooses Gaussian kernel function [11] as the function \( K \), defined as:

\[
K(E(i),E(i')) = \exp\left(-\frac{D^2(E(i),E(i'))}{2h^2}\right)
\] (3.9)

\[
D(E(i),E(i')) = \sum_{j,j'} \delta(B_1,B_2)
\]

\[
B_1 = Be(\beta_1^+(f),\beta_1^-(f))
\]

\[
B_2 = Be(\beta_2^+(f),\beta_2^-(f))
\] (3.10)

The function \( D \) captures the distance between the data set \( E_1 \) and \( E_2 \), \( \delta(\cdot) \) is a function of total variation distance. The probability \( p(f) \) of a node pair with the same feature \( f \) establishes a link obeys the beta distribution, the distance \( D \) can be defined as the total variation distance of the beta distribution. \( h \) is a width parameter of the Gaussian kernel function, which can be optimized by cross-validation in the calculation.

Then we get the probability \( p(\xi(i,j)) \) of the node \( i \) encounter node \( j \) in time slot \( t \).
IV. EXPERIMENT RESULTS

A. Experiment Settings

In the experiment of this paper, two measured data sets were selected as the research objects: SIGCOMM [12] and UCSD [13]. These two data sets are the user contact data of a conference and a large communication scene in campus. The basic situation of the data set is as follows:

<table>
<thead>
<tr>
<th>Data set</th>
<th>SIGCOMM</th>
<th>UCSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Record date</td>
<td>2009</td>
<td>2002</td>
</tr>
<tr>
<td>Device</td>
<td>smart phone</td>
<td>PDA</td>
</tr>
<tr>
<td>Contact type</td>
<td>Bluetooth</td>
<td>WiFi</td>
</tr>
<tr>
<td>Device number</td>
<td>76</td>
<td>275</td>
</tr>
<tr>
<td>Days</td>
<td>4</td>
<td>77</td>
</tr>
<tr>
<td>Record interval (min)</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Scene</td>
<td>conference</td>
<td>campus</td>
</tr>
</tbody>
</table>

The SIGCOMM dataset records the communication of 76 mobile smartphone devices in a conference building. The recording lasts for 4 days and the communication method is Bluetooth [14]. In this experiment, the 42-hour record in the first 2 days was selected as the research object.

The UCSD dataset records the contact of 276 students with handheld PDA devices on campus for 77 days. The link between the devices is established via the WiFi [15] hotspot. In late midnight, students usually do not go out for activities. Therefore, in this specific experiment, in order to select the record with predictive value, this paper extracts the daily record of 8:00-24:00 as the research object.

In order to characterize the accuracy of the prediction algorithm, the main accuracy indicators are AUC [16], Precision [2] [17] and Ranking Score [18]. The difference lies in the different focus of the accuracy. The AUC (area under the receiver operating characteristic curve) uses the ROC curve to describe the whole accuracy; the Precision considers the accuracy of the predicted probability in top L; the Ranking Score focuses on the edge sorting.

Since this paper is a prediction for the probability of future encounter opportunity, the Precision indicator in [2] is selected:

\[
\text{Precision} = \frac{l}{m} \quad (4.12)
\]
in which \( m \) is the total predicted number of links at next time slot, and \( l \) is the actual number of links established in these predicted links, that is, the number of links be correctly predicted. For simplicity, let \( m \) be the total number of links actually present at the next time slot in the dataset.

B. Prediction accuracy

In order to verify the reasonable validity of the proposed kernel regression opportunity prediction method KROP, we compared the method to some classical prediction methods in the experiments with the SIGCOMM and UCSD dataset scenarios respectively. Methods used for comparison include some route indicators based prediction methods: Prophet indicator, MaxProp indicator and OPF [19] indicator. And also include simple similarity indicators AA and contact frequency indicators CF. Figure 2 shows the performance comparison of various prediction methods:

![Graph showing prediction accuracy comparison](image)

(a)SIGCOMM

(b)UCSD

Figure 2. Prediction accuracy of different methods.

Figures 2(a) and 2(b) shows the experimental results in the SIGCOMM and UCSD scenarios, respectively. As can be seen from the figure, the KROP prediction method can achieve 19\% and 25\% accuracy improvement compared to the classical method. Prophet's probability value can reflect the topology situation in a period of time, but cannot describe the correlation of continuously time slots, rules from the historical data are ignored; AA considers the similarity of the local topology, failing to reflect the
instantaneous changes of topology, and does not take the periodicity and regularity of human activities into account. The MaxProp indicator performs poorly.

KROP in the SIGCOMM scenario performed litter batter than in the UCSD scenario. It may be that the nodes meet and communicate more frequently in the conference scenario, which is convenient for prediction. The campus scene is sparse and the time span is longer.

C. Prediction accuracy of active nodes

It is found in the experiment that due to the sparsity of the network, most of the node pairs in each time slot are not linked, and there are a small number of active nodes with more neighbors and higher degrees in this network. Since the kernel regression algorithm in this chapter aims to predict encounter opportunity, the experimental of the prediction accuracy of active nodes is made here for comparison. We choose top \( k \) nodes with high degree as research objects in each time slot, and observe the change of the prediction accuracy with the length of the time slot.

![Figure 3. Prediction accuracy of active nodes.](image)

Figure 3(a) and Figure 3(b) show the change of the prediction accuracy of the active nodes with the length of time slot in the data sets SIGCOMM and UCSD, respectively. It can be seen from the figure that nodes with higher
degree, or more active nodes, got higher prediction accuracy, up to more than 90%, which is higher than the average accuracy of the prediction in the previous section.

The practical significance of the experimental results is that in the MANET scenario, the prediction accuracy of these active nodes can reach a very high level. On the one hand, in the process of packet forwarding, the active nodes should be preferred, or active nodes can be used as a relay center; on the other hand, when the active node carries the packet for copying and forwarding, it should minimize the number of copies of the message, select efficient forwarding nodes according to its high prediction accuracy to improve forwarding efficiency.

D. Prediction accuracy of different time

Taking the SIGCOMM scenario as an example, the changes and intensities of links in the network are different at different times of the day. The average prediction accuracy of all time periods is shown in the previous two sections. This section will select the data record of a certain day and analyze the accuracy change at the time slot from 00:00 to 24:00.

![Figure 4. Prediction accuracy of different time.](image)

As shown in Figure 4, the prediction accuracy of each prediction method is different at different time of day. The accuracy is close to 0 during midnight, and up to 70% in the morning or afternoon. As the prediction method aims to predict the link appearance, active network scenario will get higher prediction accuracy compare to an inactive scenario.

V. CONCLUSION

Based on the existing probability prediction research, this paper proposes a kernel regression-based opportunity prediction method KROP, which extracts the high-efficiency features of node pairs in the network to represent the change of network topology and predicts by kernel regression method.
KROP analyzes the changes of the historical data set, mines the evolution rules of the topology, and estimates the probability of future link occurrence. Using the KROP method, it is possible to predict changes in the future topology of the network and provide immediate and targeted guidance for routing in MANET. The prediction method also achieved a high prediction accuracy in the comparative experiments.

Future research work includes predicting the duration of future links and the probability of future link disconnection. In the MANET scenario, this is useful for judging whether the destination node is reachable. For unreachable nodes, the service can be denied, avoiding wasting time and caching.

REFERENCES