ISSN 2047-3338

Improving Mobility Prediction Using Data Mining Techniques

Velmurugan. L.[1] and P. Thangaraj[2]
[1]MCA, EBET Group of Institutions, Kangayam, 638108, India
[2]Department of CSE, Bannariamman Institute of Tech., Sathyamangalam, 638401, India
[1]velu.phd@gmail.com

***Abstract*– With the increase in the popularity of mobile devices, WLAN infrastructure planning has become significant for maintaining desired quality of service. Movement pattern of the user plays a key role in maintaining quality of service. Mobility prediction involves predicting the mobile device's next access point as it moves through the wireless network. In this paper, we propose a new method for feature extraction and use Neural network classifier to evaluate the classification accuracy. We evaluate our hypotheses using one month syslog data of Darthmouth college mobility traces available online, to extract mobility features. Our proposed methodology was implemented and achieved a precision of 0.86 and recall of 0.82.**

***Index Terms*– Mobility Prediction, Neural Network, Wireless Infrastructure, Quality of Service and Location Based Service**

I. INTRODUCTION

MOBILITY predictions of wireless devices [1] helps the users in smart access and useful to the service provider in planning of the infrastructure and to provide better quality of service (QoS). Mobility prediction is applicable to GSM mobile networks as well as the wireless networks containing many access points (APs) and routers. Mobility prediction essentially tries to predict the movement of mobile users based on prior mobility models. The user's next location as the user is traveling in the network is helpful for infrastructure planning, resource allocation and future network requirement prediction. This knowledge also allows proactive measures to reduce the impact of handovers and thus, improve the network QoS. Location-based services (LBS) are an emerging application of mobility prediction, wherein information specific to a location is made available to mobile users. LBS provide many value-added services like customized advertising, resource tracking with dynamic distribution like taxis, service people, finding people, finding places based in that particular area [2]. Most users exhibit some kind of regularity in their daily movement. As the user moves through the network, the mobile devices access the network through different APs. It is possible to monitor the movement of a user by identifying when the mobile device connects itself with a particular AP and using the unique id of the AP to find the location of the mobile device. The movement trajectories of a user is generally logged in the form of where is the time at which the mobile device connect to an AP, represents specific AP of the nth location the user has moved to from the defined time. The actual movement of the mobile device is called the user actual path (UAP) which has the form:

$$u_i =< ap_1, ap_2, ..., ap_n >$$

User mobility pattern (UMP) is the frequently used path of the user and can be obtained using the logs from all the APs with the associated time stamp for each unique mobile device. User mobility patterns are extensively used to generate mobility rules. Mobility rules can be generated from the UAP in the form:

$$l_1 \rightarrow l_2$$
$$l_1, l_2, l_3 \rightarrow l_4$$

with support 's' and confidence 'c'. Given an relation of X on Y and the mobility pattern of interest

$$Y' \subseteq Y$$

A multidimensional association rule on Y' is a number of associations whose union is an association of Y'. The above statement can also be represented as a rule by:

$$\forall Y' \subseteq Y, A \rightarrow B$$

is an association rule on Y' iff $A \cup B$ is an association of Y'.

For a given set of mobility patterns in the body, the next location prediction can be determined using the head as the class label and can be given by $a_1, a_2, ..., a_n \rightarrow b_i 10$ Where

$$a_i \in A \text{ and } b_i \in B$$

This paper presents a neural network classification algorithm for the prediction of user movements in a wireless campus environment. The proposed algorithm uses the mobility rules formed from the mobility pattern of users and finally predicting the mobile user's next movement. We use one month trace data of Dartmouth College available in public domain.

II. PREVIOUS RESEARCH

Liu et al, [3] proposed a predictive mobility management algorithm. The proposed method modeled user's movement as elementary paths, which are either circular or straight patterns. The future location of the user is found using the mobile motion prediction (MMP) algorithm. Simulations showed that the proposed algorithm had a prediction efficiency of 95%. However, the efficiency starts decreasing as the randomness of the movement increases.

Yavas et al, [4] proposed a three-phase algorithm for mobility prediction. In the first phase, user mobility traces were extracted from the historical data. Mobility rules were derived in the second phase and finally mobility prediction was done in the third phase. The efficiency of the proposed scheme was tested and compared against two other prediction methods. The proposed method performed better than the other methods.

Akoush et al, [5] proposed a mobility path prediction model and hybrid Bayesian neural network model for predicting locations on cellular networks. The proposed model is based on the probability model to represent uncertainty in the relationships learned. Markov chain Monte Carlo method is applied to N sample values of posterior weights distribution obtained from the Bayesian training. These samples vote for the best prediction. Realistic mobility patterns were used for simulation studies. Results showed that the proposed algorithm achieved higher prediction accuracy when compared to standard neural network techniques.

Sakthi et al, [6] proposed a mobility prediction algorithm based on infrequent mobility patterns. The proposed algorithm computes new mobility patterns faster and avoids scanning of full database. The computation time for datasets at various support counts were computed using the proposed method. The results proved that the mobility patterns are generated in less time using proposed method than the re-computing approach. The work proved that live prediction can be done in a grid environment based on live data captured.

Liu et al, [7] proposed a two-level scheme for mobility prediction with a global mobility model (GMM) and local mobility model (LMM). GMM handled the domain at handover from one cell to another and included the inter cell trajectory whereas the LMM looked into the micro level including speed, distance, etc at which the mobile device moves over the network.

III. DATASET

Dartmouth College provides mobility traces as a community service to researchers. Mobility trace collected over a period of three years within the Dartmouth college campus is available. We use one month syslog data in this work. Initially 476 APs were available and was increased to 566 over the period of time. The SSID is the same for all APs; 115 subnets covered the 188 buildings and hence devices were forced to obtain new IP addresses some times. Mobility trace of 5500 students and 1200 faculty were collected during the three year period. The log recorded in a syslog server included the timestamp to each message with each message containing the AP name, MAC address of the card and the type of message. The various messages used are authenticated, associated, reassociated, roamed and disassociated. Whenever a mobile device connects to the network it is first authenticated, after authentication, the device must associate with one of the AP enabling all traffic between device and network. Reassociation occurs when another AP with better signal strength is available. Roaming is used when a device re-associates with an new access point. Disassociated message is sent when the device no longer needs the network. Sample data from the syslog is given in table 1.

TABLE 1: THE SYSLOG CAPTURED FOR A SPECIFIC USER

1034100785	AcadBldg35AP1
1034100842	AcadBldg18AP3
1034100851	AcadBldg35AP1
1034100908	AcadBldg18AP3
1034100963	AcadBldg35AP1
1034101020	AcadBldg18AP3
1034101022	AcadBldg35AP1
1034101080	AcadBldg18AP10
1034101082	AcadBldg35AP1
1034101139	AcadBldg18AP3

The first column represents the unix time stamp and the second column represents the specific access point the user has associated with.

Generation of mobility rules: Let us say the user mobility pattern is:

$$l = < l_1, l_2, l_3, ...l_n >$$

Mobility rules generated from this pattern are

$$l_1, l_2, l_3, l_4 \to l_{c5}$$
$$l_2, l_3, l_4, l_5 \to l_{c6}$$
$$l_3, l_4, l_5, l_6 \to l_{c7}$$
$$....$$
$$l_{k-4}, l_{k-3}, l_{k-2}, l_{k-1} \to l_{ck}$$

where l_{ck} represents the clustered value of the access point. The clustered value in the head is represented by all access points that are close to each other in the network.

Fig. 1 shows the frequent item set for minimum support of 35% and confidence of 10%. Fig. 2 represents the distribution of the clustered access points.

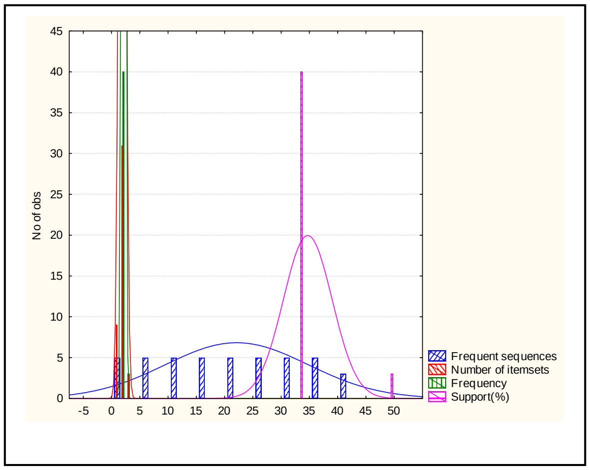

Fig. 1: Frequent Item set for one month data

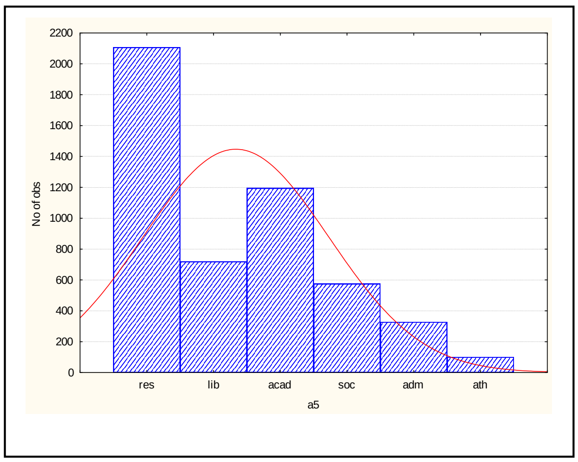

Fig. 2: shows the distribution of the clustered access points in the head

IV. RESEARCH METHOD

A. Neural Network

Neural networks are made up of multiple layers of computational units or artificial neurons, usually interconnected with each other. The artificial neurons use a mathematical model for information processing based on connectionist approach to computation [8]. Neuron consists of weights, thresholds and a activation function. The block diagram of a simple neural network is shown in Figure 3. Outputs are obtained by propagating the inputs from input layer through the hidden layer. Output signal is computed using weights, bias and activation function. The neural network is trained by propagation rule by backpropagating the errors and changing the weights of nodes. The major advantages of neural network are:

- Complex non-linear functions are created from simple linear functions
- Neural network process learns from data
- The network modifies itself as the data set changes

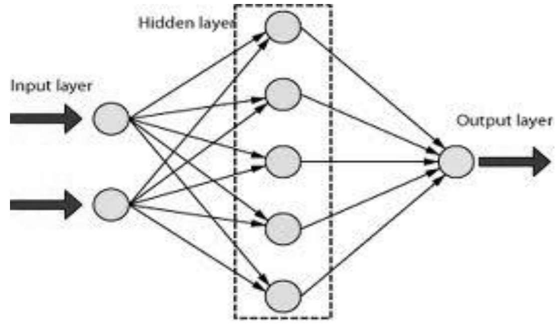

Fig. 3: A Simple Neural Network Model

The total input for a given neuron is given by:

$$s_k = \sum_j w_{jk} y_j + \theta_k$$

where s_k is the total or effective input for unit k, w_{jk} is the weight of the connection, y_k is the current activation and θ_k is the bias.

Activation function A_f takes the input and the current activation and gets the new activation value during learning by:

$$y_k(t) = A_f(y_k(t-1).s_k(t-1))$$

The activation function controls the value of each neuron and limits it generally to 0, 1 using a threshold function and the most commonly used is the sigmoid function:

$$A(s_k) = \frac{1}{1+e^{-s_k}}$$

V. RESULTS

The experiments were conducted using mobility traces from Dartmouth college campus dataset. The classification accuracy of the mobility predictions is evaluated using Naive Bayes, CART, Multilayer Perceptron and Feed Forward Neural Network. Table 2 shows the classification accuracy and Root mean squared error (RMSE) for the various methods. Fig. 5 shows the graph of the classification accuracy.

TABLE 2: CLASSIFICATION ACCURACY AND ROOT MEAN SQUARED ERROR OF VARIOUS METHODS

Classification Method	Classification Accuracy in %	RMSE
Naïve Bayes	72.0052	0.3271
CART	70.9635	0.3473
MLP	75.26	0.2712
FFNN	78.52	0.2437

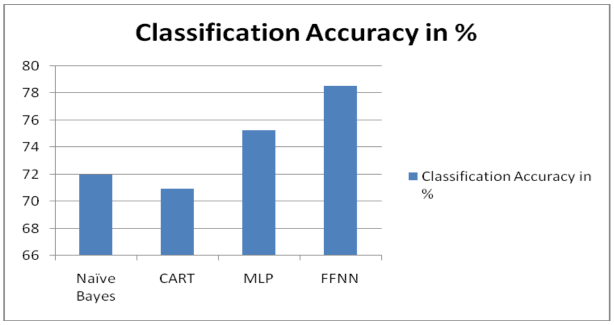

Fig. 5: Classification accuracy for various classification methods

Table 3 gives the weighted average precision and recall of all classification methods and Fig. 6 depicts the graph of the same.

TABLE 3: WEIGHTED AVERAGE PRECISION AND WEIGHTED AVERAGE RECALL FOR VARIOUS METHODS

Classification Method	Weighted Avg. Precision	Weighted Avg. Recall
Naïve Bayes	0.73	0.72
CART	0.714	0.71
MLP	0.71	0.73
FFNN	0.75	0.77

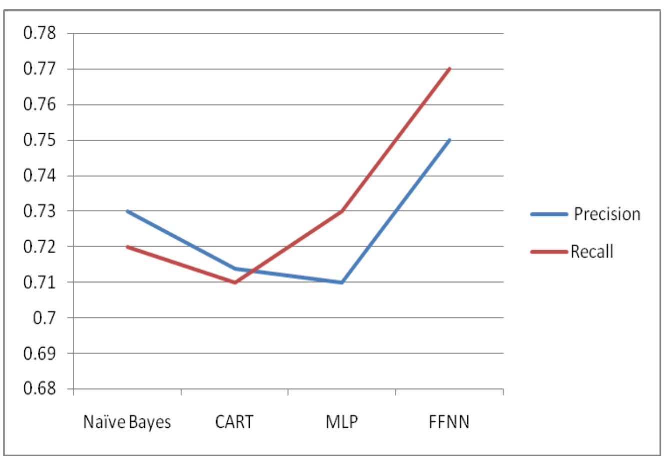

Fig. 6: Weighted Precision and Recall

From Fig. 5 it is seen that neural network based classifiers improves the classification accuracy by 9.05% to 10.65% compared to CART and Naïve Bayes classifier. Similarly from Fig. 6 it can be seen that precision is higher than recall for all the methods under consideration.

VI. SUMMARY AND CONCLUDING REMARKS

In this study, a novel mechanism to extract features from system log was proposed for mobility prediction. A two step method for feature extraction and prediction was proposed using a mathematical model. In the second stage the head of the extracted rules were clustered to become the class label for the prediction model. Classification accuracies obtained from various classifiers were compared. It is seen that classification accuracy of Neural Network classifier are higher than other probability or decision tree based classifiers.

REFERENCES

[1] Aljadhai.A, T. Znati, 2001.Predictive mobility support for QoS provisioning in mobile wireless environments, IEEE J.Select. Area Commun. 19 (10) pp 1915–1930.

[2] Liu.T, P. Bahl, I. Chlamtac, 1998. Mobility modeling, location tracking, and trajectory prediction in wireless ATMnetworks, IEEE J. Select. Area Commun. 16 (6) pp 922–936.

[3] Liu.G.Y, M.Q. Gerald, 1995.A predictive mobility management algorithm for wireless mobile computing and communications, Proceedings of the IEEE International Conference on Universal Personal Communications, pp. 268–272.

[4] Gokhan Yavas, Dimitrios Katsaros, Ozgur Ulssoy, and Yannis Manolopoulos,2005.A Data Mining Approach for Location Prediction in Mobile Web Environments, IEEE Transaction on Data and Knowledge Engineering, pp.121-146.

[5] Sherif Akoush, Ahmed Sameh, 2007. Mobile User Movement Prediction Using Bayesian Learning for Neural Networks, IWCMC'07, August, page 191-6.

[6] Sakthi.U and R.S. Bhuvaneswaran, 2009. Incremental DataGrid Mining Algorithm for Mobility Prediction of Mobile Users, Journal of Computer Science 5 (2): 146-153.

[7] Liu, T.P. and I.B. Chlamtac, 1998. Mobility modeling, location tracking, and trajectory prediction in wireless ATM networks. IEEE J. Select. Area Commun. 16: 922-936.

[8] Abdi, H. “A Neural network primer”. Journal of Biological Systems.2. (1994) pp: 247-281.