Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 06:2023
DOI 10.17605/OSF.IO/RM3UJ

AN EFFICIENT APPROACH FOR DATA TRANSMISSION USING THE ENCOUNTER PREDICTION

AMBREEN MEMON*

Information Technology" at Western Institute of Technology at Taranaki, New Zealand. *Corresponding Author Email: A.memon@witt.ac.nz

M. ASHRAF NAZIR

Ph.D. Scholar, Department of Computer Science, Superior University, Lahore, Pakistan and Lecturer at GC University, Lahore. Email: Ashnaz084@gmail.com

KHALID HAMID

Ph.D. Scholar, Department of Computer Science, Superior University, Lahore, Pakistan and Assistant Professor at NCBA & E University East Canal Campus Lahore. Email: Khalid6140@gmail.com

MUHAMMAD WASEEM IQBAL

Ph.D., Associate Professor Department of Software Engineering, Superior University, Lahore, Pakistan. Email: Waseem.iqbal@superior.edu.pk

Abstract

Mobility has gathered researchers' interest as mobile users change their locations, and the need for quality of service and big data transmission has increased. This increase in wireless mobile users and their activities in mobile ad-hoc networks (MANET) has also resulted in the growth of bandwidth allocation. Efficient planning and management of the resources can be done through mobility prediction. Determining mobility is about finding the maximum probability of the next location where a mobile user could travel among different mobile networks. Mobility probability determination increases the authenticity of daily paths a mobile user adopts, making data transmission and resource management efficient. After mobility probability determination, data transmission starts after an encounter among source and destinations nodes. If an encounter doesn't occur at the predicted future location, the resources are consumed for the continuous location next location prediction. The resources include computational resources and energy consumed by these resources for mobility probability determination. Therefore, there is a need for encounter prediction to ensure data transmission to reduce this energy consumption. Different mobility prediction models exist, e.g., the Markov model uses random data to predict the next state/location of the user. Still, current research lacks work on encounter prediction in MANET. In this paper, we have proposed a new method of data distribution via encounter prediction (DDVEP) for the encounter prediction consisting of data mining techniques and sequential algorithms. This algorithm stores the locations sequentially visited in the daily routine and predicts the next encounter of multiple users. The case studies performed using the mobility traces from the individuals on a university campus showed that the accuracy of the encounters predicted by the proposed model was 5% higher than the existing Markov model for the encounters and future location prediction.

Keywords: Similarity Analysis, Mobility, Device-To-Device Communications, Random Forest Model, Opportunistic Connectivity

1. INTRODUCTION

In the current revolutionised digital world, the increasing number of smartphone users is causing mobile traffic growth, reaching 79% in 2019. [1]. These mobile devices can form a

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 06:2023
DOI 10.17605/OSF.IO/RM3UJ

wireless mobile ad-hoc network (MANET), which provides different scalability and flexibility advantages. Different users can join or leave the network without prior notification [2]. These mobile devices can change their location and connect to other devices dynamically and arbitrarily. This node mobility makes the MANET environment unpredictable for real-time data transfer [1-3].

Mobile ad-hoc networks don't require any fixed infrastructure. Due to the absence of any central infrastructure and limited transmission distances, the nodes in MANET are responsible for discovering each other for data transmission [5]. Discovering other nodes is also essential for reliable communication and ensuring data transmission without interruptions. The forecasting of a mobile user's next location is called mobility prediction. For mobility prediction of a specific mobile user, where millions of devices are in the mobile network, determining the behavioural pattern of users is a key step. Individuals usually follow a daily routine in their activities; for example, a user can go to work from 9 A.M – 5 P.M, go home from 5 P.M – 6 P.M, and go to a restaurant from 7 P.M – 9 P.M. These behavioural patterns show that the certain user goes to certain places at specific time intervals. Using these mobility patterns, the future locations of the mobile users can be determined using the probability method. Existing mobility prediction research usually employs the Markov model or its variants [4-6].

After discovering the future location, the sender node also ensures whether the encounter at the location occurs. These operations of mobility prediction, encounter checking, and data transfer across multiple nodes result in huge resource consumption, such as bandwidth, energy, battery power, CPU, and buffer space [3]. To reduce resource consumption and ensure the encounter among devices, there is a need to predict encounters.

This paper uses data mining techniques and a sequential algorithm for mobility and encounter prediction. The proposed algorithm employs a user's daily routine stored as a "sequence." The sequences are formed orderly, and we have divided the data set into different chunks to predict the future location. The experiments showed that the proposed model determined the probability of where users can be at the next locations and encounters with higher accuracy than the low-order Markov model. The accuracy is calculated by training mobility traces in the data set with the Markov model and the proposed sequential-based prediction model on the traces used by Burbey and Martin (2008). The rest of the paper is structured as follows. In Section 2, we present related work. Section 3 explains the proposed algorithm and the implementation details, including the data set used. In Section 4, we detail the experimental setup and analyse the results. Finally, Section 5 highlights the study's contributions and limitations and outlines further work [7-9].

2. LITERATURE REVIEW

One of the Massachusetts Institute of Technology themes is anticipating user behaviour. The time duration is recorded when mobile phones are close to Bluetooth and cell tower IDs. Depending on how they interact with one another, Bluetooth devices exhibit various

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 06:2023
DOI 10.17605/OSF.IO/RM3UJ

behaviours. Presence of business students conducting the same activity at a nearby location[7] Bluetooth signals were constructed in individual houses to check the accuracy of the three techniques of translating data into locations by cell towers [8]. Estimating the object's next location using the dynamic. Bayesian Network reached a successful percentage of 93 % to 99 %. In prediction, the next cell is a sequence of locations that they investigated in networking and communication areas, using these needs to improve those resources for reservation and QoS predict that driving destinations are given a partially travelled route by calculating the probabilities of different possible routes [9-16].

This study aims to predict a similar location so that even researchers can revise the location update from a movement trajectory. Mobile phones or other electronic devices will sometimes be switched off, unremembered, or entering urban canyons or other places with such low GPS coverage. User updates, i.e. significant time lapses between location updates, are crucial facts of those applications such as Facebook and Twitter are dependent. Additionally, users shall access record deletion or prevention recording as per choice to confirm full privacy policies. The author suggest predicting future destinations and previous histories [10-12].

Let 'xyzzxxxy' be the mobility history of a particular user 'u', where x, y, and z are visited locations. These locations are visited in the same order as in the string. After visiting a location, that location is chained in a sequence and shows the mobility history of the user. These mobility patterns help predict a mobile user's future location [12].

However, indicating the future location alone isn't enough when data transmission from a source to a destination is required. The data transmission process may take huge time intervals depending upon the data size.

Moreover, an individual's changed location or mobility preferences may interrupt the data transmission process. For data transmission, we also need to predict user encounters along with the locations where the encounter will occur and the duration of the encounter.

A partial match algorithm can predict the future location of mobile users, whereas sequential patterns can help maintain the user's mobility history and maintain the location's semantics. In this paper, we have proposed a method for encounter prediction and location id where encounter will occur and encounter duration, through mobile user's mobility history [18-20].

3. DETAIL DESCRIPTION OF DATA

To validate the proposed approach, we used the same data set used in the WTD (The Wireless Topology Discovery) experiment as part of a project at UCSD (University of California at San Diego***)*** The data-set consists of traces of WiFi connection from PDAs carried by 3000 students. Whenever a user is connected to an AP, the time of connection, location id, and connection duration are logged.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 06:2023
DOI 10.17605/OSF.IO/RM3UJ

The mobility traces of a user "u" are represented as a sequence of user id, sample time, AP id, Sig_strength, AC_power, and associated flag. Where the entries are explained as follows [10-17]:

USER-ID: is an identifier reserved for a mobile device user

SAMPLE_TIME: is the time at which WTD software collected the data.

AP_ID: is a unique identifier assigned to the detected AP.

SIG_STRENGTH: Represents the strength of AP signals received by the device.

AC_POWER: Stores the value 1 if the device uses AC power and 0 if in case of battery.
ASSOCIATED: shows if the device is related to this AP (1) or not (0).

An instance of a data set is shown in figure 1. While a student's device was powered on, WTD sampled and recorded the information shown in the table every 20 seconds for each access point (AP).

USER_ID		SAMPLE_TIME	AP_ID	SIG_STRENGTH	AC_POWER	ASSOCIATED
123	Sep-22	0:00:00	359	8	0	0
123	Sep-22	0:00:00	363	5	0	0
123	Sep-22	0:00:00	365	11	0	1
191	Sep-22	0:00:00	355	31	0	1
101	Sep-22	0:00:00	353	8	1	0
101	Sep-22	0:00:00	362	30	1	1
129	Sep-22	0:00:00	369	31	0	1
156	Sep-22	0:00:00	360	19	1	1
184	Sep-22	0:00:02	352	29	0	1
184	Sep-22	0:00:02	364	12	0	0
211	Sep-22	0:00:02	366	31	1	1
211	Sep-22	0:00:02	369	18	1	0
35	Sep-22	0:00:05	353	31	1	1
154	Sep-22	0:00:05	362	23	0	1
66	Sep-22	0:00:06	363	25	1	1
66	Sep-22	0:00:06	364	4	1	0
149	Sep-22	0:00:07	354	7	0	0
149	Sep-22	0:00:07	355	10	0	0
149	Sep-22	0:00:07	360	29	0	1

Figure 1: Format of the original data-set

3.1 Pre-processing of Training Data

[12] used the Partial match algorithm to predict the future location. The authors employed the log loss function to calculate the amount of training data for future location prediction. The average log loss function of a test sequence x_0^T = x_0, x_1, x_2……………………., x_T is defined as.

$$l(\hat{p}, x_0^T) = \frac{1}{T}\sum_{i=0}^{T} \hat{p}\ (x_i | x_0\, x_1 \dots \dots x_{i-1}\ ) + log2\ \hat{p}(\ x_0\ ) \qquad (1)$$

For an alphabet'∑' the benchmarked algorithm was trained on a sequence q_n^1 = q_1, q_2, q_3, …., q_n. Minimising the average log-loss corresponds to maximising the probability assignment for the test sequence. In other words, a lower value for average log-loss indicates a better predictor.

3.2 Data set for proposed sequential pattern approach

For training the proposed approach with sequential patterns, the data set we have used is UCSD traces. The data set had many entries and sizes (over 371 Megabytes), so it was reduced.The first step of pre-processing lessened the data amount for pertaining the

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 06:2023
DOI 10.17605/OSF.IO/RM3UJ

important features and noise removal. For this, the following steps were performed on the UCSD data set. For reduction, the data files were parsed into other sub-files, each containing data about 275 users. For additional data reduction, the records with mobile devices connected with an access point were used while completely ignoring the records in which a device didn't connect with an access point despite sensing it. Figure 2. shows an instance of the resulting data file.

Then the data set was first trained with the proposed approach. Since WTD recorded the above information for all APs sensed during a sample, if a device detects three

1	354	2002-09-22 13:07:09.00	00:04:50.00	2002-09-22 13:11:59.00
1	354	2002-09-22 13:12:30.00	00:02:22.00	2002-09-22 13:14:52.00
1	363	2002-09-22 13:15:13.00	00:03:23.00	2002-09-22 13:18:36.00
1	367	2002-09-22 13:31:47.00	00:09:12.00	2002-09-22 13:40:59.00
1	367	2002-09-22 15:51:44.00	00:11:39.00	2002-09-22 16:03:23.00
1	367	2002-09-22 19:19:48.00	00:03:18.00	2002-09-22 19:23:06.00
1	367	2002-09-22 19:25:35.00	00:04:34.00	2002-09-22 19:30:09.00
1	367	2002-09-22 20:13:09.00	00:15:29.00	2002-09-22 20:28:38.00
1	355	2002-09-22 20:28:59.00	00:01:01.00	2002-09-22 20:30:00.00
1	367	2002-09-22 20:30:20.00	00:11:55.00	2002-09-22 20:42:15.00
1	355	2002-09-22 20:42:35.00	00:03:24.00	2002-09-22 20:45:59.00

Figure 2: An instance of the data-set

APs in one sample, three entries were recorded for that sample (which differ only in the AP detected, signal strength, and associated flag).

3.3 Movement Patterns

People visit different locations at some specific timestamps in their daily routines. The locations visited by most of the users are in a particular order or pattern for each day. These mobility patterns help us to predict the future location of a user. These mobility patterns for different users can intersect at different points each day based on their spatial-temporal coordinates.

3.4 Determining Future Location

The human mobility prediction helps estimate the next location a person will visit. For example, people generally go to offices or workplaces every day on weekdays and go to malls for shopping on weekends or after work. In people's daily life, temporal regularities are evident. There is a certain probability of returning to some highly frequently visited places, including houses or offices. Many urban human mobility predictors try to capture spatial-temporal regularities to help estimate traffic congestion, air pollution prediction, and mobile ad-hoc network generation. However, In this research, we determine the future locations through the time-space matrix of a specific user, where each time entry may be a time of any day or week. Whereas spatial entry captures the positions of a user at a specific time. These matrices log the location and time at which a user with a specific id visits a particular location.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 06:2023
DOI 10.17605/OSF.IO/RM3UJ

3.5 Encounter Patterns

The mobility of mobile users is quite homogenous and usually shows similarities. Multiple users meet at different locations. If users meet simultaneously at the exact location, we say an encounter has occurred. In this research, encounter patterns are all instances from the data-set where two users encountered each other. These encounter patterns consist of the time and location at which the encounter occurred. For a particular encounter, the time t_1 of u_1 at location L_1 matches the time t_2 of u_2. Where t_n refers to the time of presence of user u_n at the location L. For instance, Figure 3 shows two friends, user1 and user2, who study at the same university. The X-axis shows the places both friends have visited, and the y-axis shows when both users visit. The intersections of both lines in the graph show user 1 with user 4 at the same place and time. Figure 3. shows a sample encounter among user 1 and user 4.

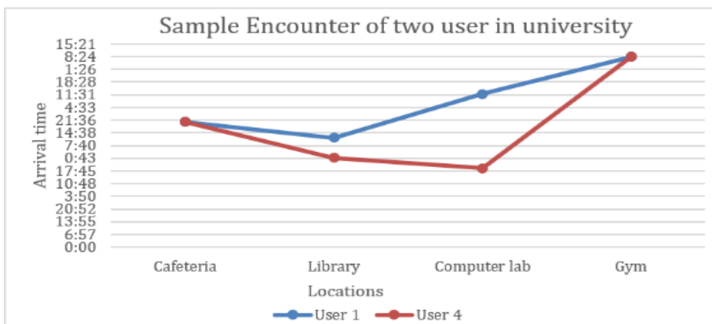

Figure 3: Example of user encounter

An encountered pair is represented in the following sequence in our proposed system:

4. ENCOUNTER PREDICTION METHODOLOGY

To predict users' encounters, we first determine the probability of the next location of a user through his mobility traces. We have proposed an algorithm that stores the user's mobility history data in sequences to predict future locations (see Figure 4). Each new visited location appears as a new entry in the sequence. The future locations of different users, in the form of spatial-temporal associations, help to predict whether users will have an encounter at a specific location or not [18-25].

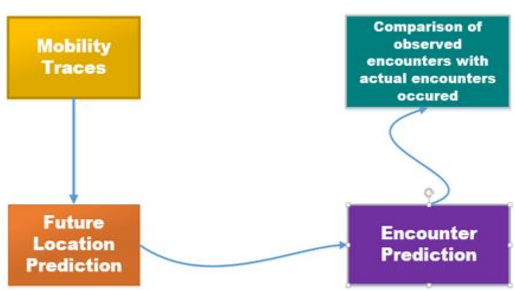

Figure 4: General steps of the proposed algorithm

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 06:2023
DOI 10.17605/OSF.IO/RM3UJ

4.1 Predictive Model

For big data transfer to distant locations and optimization of resource consumption, we proposed an algorithm for mobility prediction and predicting different user encounters. The proposed technique uses data mining with the sequential algorithm. It stores the mobility history of a mobile user "node" in a data structure called "sequence". These mobility traces are collected and used for future location and encounter prediction.

4.1.1 Algorithm for Predicting Sequential Patterns

The proposed algorithm takes values of transaction sequences as input and generates the frequently visited places in the form of sequential patterns. These sequential patterns are used to predict the mobile user's next location. The algorithm works in the way that if a person visits specific places in a certain pattern for a particular time period, then that sequential pattern is regarded as a candidate sequence pattern and added to the previously available set of sequential patterns. These sequential patterns represent the history of movement of the mobile user, which is the order in which certain places are visited, and the time span for which the user stayed in a place. Algorithm 1 shows the method for predicting sequential patterns [26].

Input: D_s $\quad$ *Mobility Traces*
Output: FP_u $\quad$ *frequently visited patterns of user*

let S be the set of sequences
for $k = 1$; $\mathrm{FP}_{uk}! = 0$ $k++$ **do**
  $\mathrm{C}_{NCS}(FP_{uk})$ ▷ new candidate sequence
  for *Mobility sequences* ; $t_s \in D_s$ **do**
    C_{total} = Subsequence (C_{NCS}, t_s)
for all candidate sequences; C $\in C_{NCS}$ **do**
    $C_{count}++$
  end for
  $FP_u = \{C \in C_{NCS}\}$
end for

Algorithm 1: Predicting sequential patterns

The proposed algorithm uses mobility traces of mobile users (logged during each movement), i.e., to form sequential patterns of frequently visited places by the user. Let (FP_{uk}) the frequent pattern of the user, S be the existing sequences, showing the user's mobility. In case of a new movement by the user, a new sequence, i.e., candidate sequence (CNCS), is generated. This new candidate sequence is added to existing mobility sequences, where all candidate sequences combine to form a set, F_K of frequently visited places in the form of patterns. Frequently visited user patterns are employed further to predict his future locations.

Let's take the example of the user having id 2 from the traces in the data-set. On September 22, at time 1:36:05, the user is at 353. At time 1:36: 25, when the user visited location 357, a new candidate sequence was generated and added to user 2. After this, the user 2 visits locations 362, 357, and 353. So, the frequently visited patterns of user 2 are in the form <353, 357, 362>

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 06:2023
DOI 10.17605/OSF.IO/RM3UJ

4.1.2 Proposed Computational Model for Encounter Prediction

This section describes a computational model for encounter prediction step by step.

1. Two nodes "i" and "j" will have an encounter $e_{i,j}$ only when both are connected to the same AP, a_x at the same time t_z .
2. The variable $c_i, a_x(t_z)$ states that the node "i" is connected to AP a_x at the time t_z
    a) The variable has value 1 when such a connection exists
    b) The variable has a value of 0 when no such connection exists.
3. If a node "i" has only one radio, then at any given time, t_z there can be only one AP a_x for which variable $c_i, a_x(t_z)$ has value one.
    a) This means "The sum of all $c_i, a_x(t_z)$ for all AP in set A is less than or equal to 1"
4. The probability of a node "i" being connected to AP a_x at a time t_z can be defined as $P_i, a_x(t_z)$ and can be derived by calculating the ensemble average at the time t_z taken over all experiments (days)
    a) We can assign weights to all experiments, i.e., less weight to earlier and more weight to later experiments, if we want to fine-tune the outcome.
5. The probability of encounter $e_{i,j}$ between node "i" and "j" at AP a_x can be as $e_i,_j(a_x,t_z) = P_i,a_x(t_z) \times P_j,a_x(t_z)$
6. Also, a system constraint must be imposed. It states that an encounter $e_{i,j}(a_x,t_z)$ can only occur if no encounter is present at $t-1$ the time on the same AP a_x

We use a network among two mobile nodes, "i" and "j", selected from the traces of 275 users where U= (u_1, u_2, u_3… u_{275}). Also, $e_{x,y}(a_m, t_n)$ is the set of all encounters that may occur among nodes x,y, at the same time t_n, where $t_n \epsilon$ t, which is a period of 7 days starting from Monday to Sunday. When an encounter occurs among any two nodes, the encounter will occur if both these nodes are connected to the same access point a_m. In this case, signal strength will be 1; otherwise, 0. The access points are at specific locations having id from AP = (AP_{350}, AP_{351}, AP_{352}……..AP_{370}). The probability of encounter among two nodes is the product of user 1 being at an access point at a specific location and the probability of user 2 at the same access point.

5. PROPOSED TECHNIQUE EVALUATION

In this section, we evaluated the proposed data mining technique used with the sequential algorithm. The evaluation is done step by step, first for the model for predicting future locations, then listing the observed encountered patterns of users and evaluating the performance of the encounter prediction model [27-33].

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 06:2023
DOI 10.17605/OSF.IO/RM3UJ

5.1 Evaluation of Prediction Model

To evaluate the prediction model, first, we performed some simulations. The code for the prediction algorithm was developed in Java programming language, where the prediction model generated users' locations at different time intervals. These locations are predicted based on the mobility history data of a mobile user. The output was in the form :<user_id, time, location_id>. Figure 5. Shows an instance of actual prediction output from the simulations.

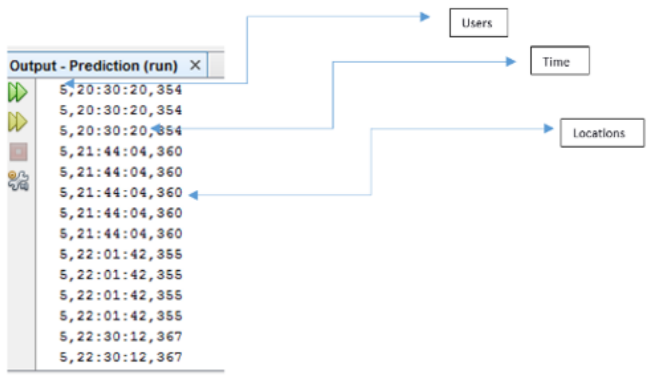

Figure 5: Prediction of location

5.1.1 Results for Probability of Different Users at Specific Locations

We used four traces of different users from the UCSD data-set to calculate the probability of a user being at a specific location in the future. We applied data mining techniques with the sequential algorithm. The graphs in Figure 6 show the probabilities computed through the proposed model for users 1, 2, 3, and 4 to be available at specific locations. The X-axis shows the location's id, and the Y-axis shows the probability of the user being present at a location. The most common locations of users with the highest possibilities are from 350 to 360.

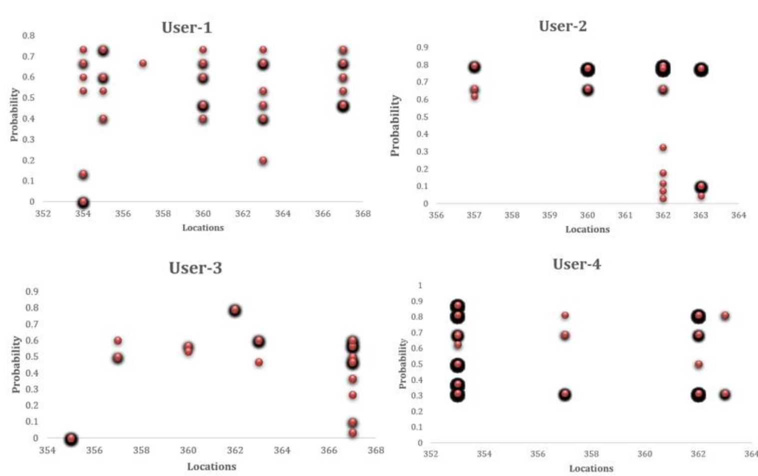

Figure 6: Probability results for user1, user2, user3, user4

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 06:2023
DOI 10.17605/OSF.IO/RM3UJ

5.2. Encounter Patterns Observed

The proliferation of mobile phones has created paradigms of communication and interaction for these devices. Mobile users tend to follow patterns in their movement observed by the devices' encounters. The knowledge and understanding of behavioural patterns can help predict future users' locations in the networks comprising mobile devices owned by different mobile users. This subsection lists down all the encounters that were observed through the traces. We inspected the data set and determined the users with the same location simultaneously. Figure 7. Shows the observed data about the encounter of five randomly chosen users from the data-set with respect to their different locations.

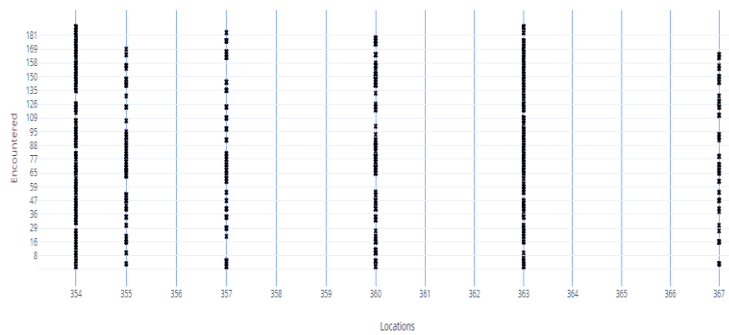

Figure 7: The encounter of 5 users with each other in 7 days

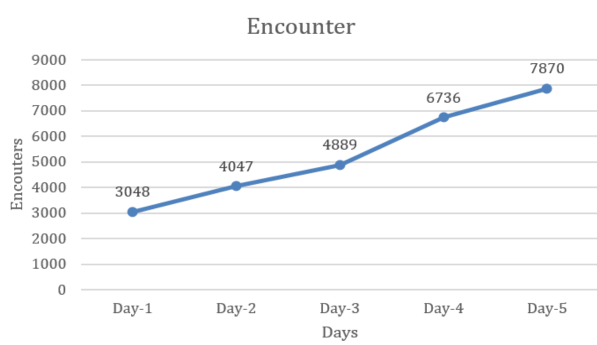

Figure 8: Shows 5 days encounter

Figure 8. Shows all encounters of the user having id 2 in five days. Although the number of encounters was high on day 5 compared to day 1 from the analysis, we have found that the reason for high encounters on a specific delay is directly related to the delay in data transfer. The greater the delay for data transmission, the greater the encounters. The low number of encounters is due to the low delay for data transfer to another node.

Similarly, Figure 9 shows the total number of encounters in seven days. The period of seven days was chosen randomly from the data set. Both figures show that increasing the days or time affects encounters (means encounters will increase). The increased number of days is directly related to the delay required to transmit data to the other node.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 06:2023
DOI 10.17605/OSF.IO/RM3UJ

So, by increasing the number of days, the encounters will increase. The gradual increase in the encounters is because the encounters that occurred on a day were added to the encounters occurring on the next day.

Figure 10. Shows the total number of encounters that occurred among all users. The results are different from those in figure 6. The latter shows the encounters happening in seven days while figuring 10. Shows all encounters that occurred in the data set. The encounters were relatively less on Saturday and Sunday as these are weekend days. The traces used for the experiment are of the UCSD campus, which may have only a few numbers of mobile users present on the campus these days.

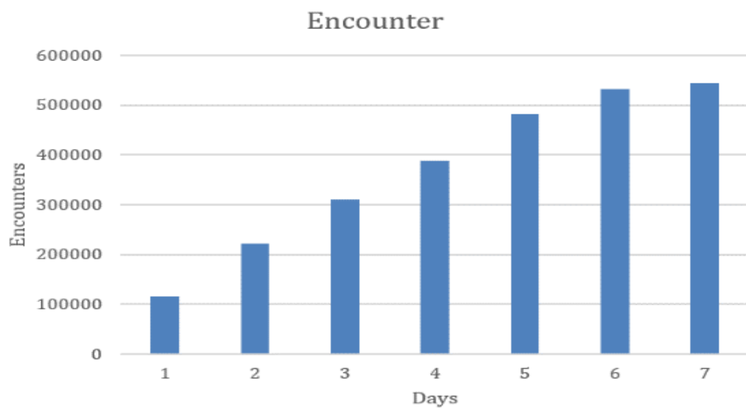

Figure 9: The total number of encounters among all users

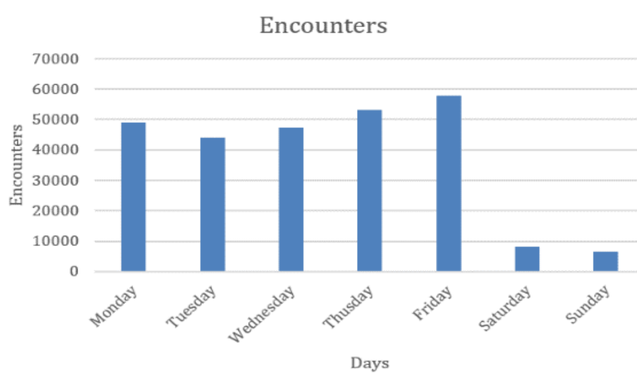

Figure 10: Shows the total number of encounters that occurred among all users

5.3. Encounter Prediction Results

A novel approach for encounter prediction is proposed in this research. However, work exists for prediction, e.g., prediction of the next arrival time to a specific location, time for encounters, and future place prediction; however, the literature doesn't address the encounter prediction. Encounter prediction based on spatial-temporal preferences can help to exchange data at a lower cost of time. Encounter prediction majorly addresses the time for the encounter of different users (based on their future location) and the time for the encounter. The predicted location determined through encounters in the proposed

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 06:2023
DOI 10.17605/OSF.IO/RM3UJ

work estimates the time interval when data among two or more nodes could be exchanged.

For predicting the future locations, we used small time intervals. The small interval was chosen because of the variance in a person's movement. A small interval was highly affected by a slight change in the daily schedule. The data at different locations were recorded at different time intervals. Each dot represents a mobile user in our case. n encounter of more than one user shows that their respective mobile nodes are at the same location. The predictions for encounters for different users are shown based on the sequential history movements. The prediction for the encounter of other users also consists of the specific locations.

The encounter prediction is based on a specific user's time in a particular position. Stay time at a location is also logged in the sequence and helps to predict the encounter of different nodes at a location. Figure 11. shows the stay time of other users at multiple locations. Each circle represents a user. The figure shows how long multiple users with different IDs stayed at a particular location.

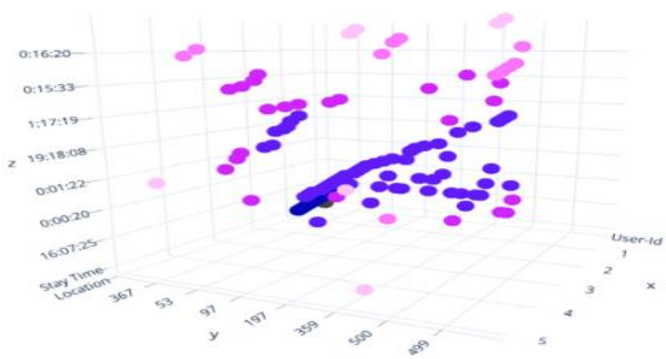

Figure 11: Users stay time location

Figure 12 presents encounter probability of user 2 with different users along with the location id where encounter is predicted.

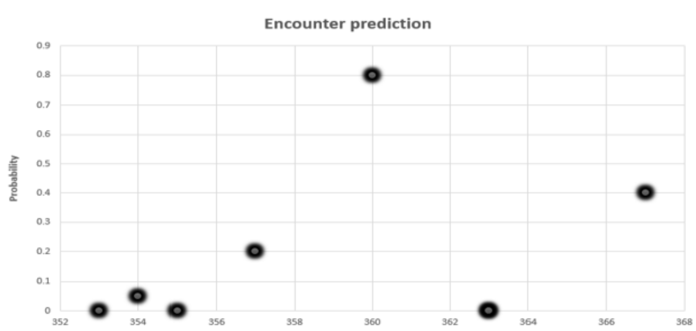

Figure 12: Results for encounter prediction of user 2

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 06:2023
DOI 10.17605/OSF.IO/RM3UJ

Figure 13. Shows encounter prediction for different users. Encounter prediction results consist of the time when an encounter will occur at a specific place with a user id.

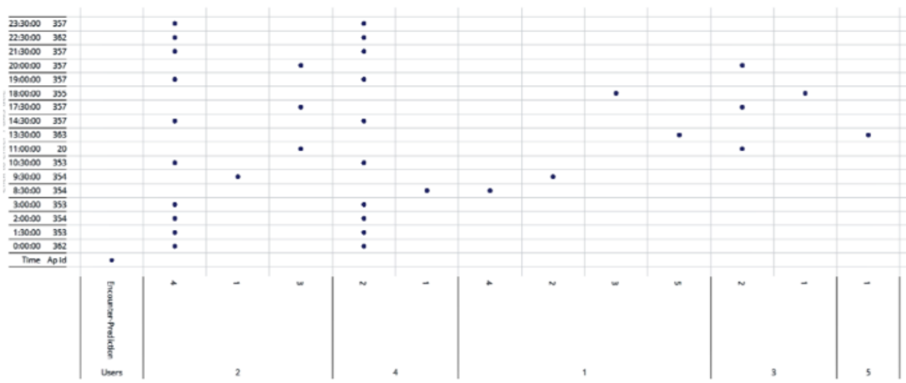

Figure 13: Results for encounter prediction of different users

6. COMPARISON OF PROPOSED APPROACH WITH BENCHMARK MARKOV MODEL

We have compared the proposed approach with the work done by [12], who employed the 'Prediction by Partial Match' model for future location prediction based on mobility patterns. The algorithm used data of varying lengths of historical data for constructing the prediction model. Each string is processed through each character, and for each substring, a table is built, and the corresponding character is stored along with it. The number of times a particular character occurs is stored in the table. For prediction, the table is traversed along with the provided context. The traversal returns the character with the highest occurrence, representing the state.

6.1 Prediction Accuracy of Proposed Approach vs Markov Model

To predict future locations or human encounters, author [12] has used "Partial Match", a variant of the Markov model.

Partial Match algorithm

The partial match algorithm is a variable order Markov Model that relies on generating a predictive model using variable length sequences of the previous context. The context stored in the sequences is the characters that precede a specific sequence. These variable length sequences store the historical data in a

$$p(\varphi) = \frac{C(\varphi)}{1+C}, \mathrm{c}(\varphi) > 0 \qquad (2)$$

A table that is built for each order from 0 to maximum. The purpose of creating the table is to store all the substrings, characters, and the count of the characters that appear after a specific substring. The region from the set of high probability sequences is selected as a result of prediction. All the generated candidate sequences are chained together through the k-order Markov chain model. A counter in each created table stores how

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 06:2023
DOI 10.17605/OSF.IO/RM3UJ

many times a string in the sequence has been encountered in a particular context. The partial match model was trained on different timestamps that show the location of individuals at multiple time periods. The authors used arithmetic coding from [10] to build the tables and calculate the probabilities. For arithmetic coding, at first different probabilities are calculated. Then, the first probability that a symbol will occur in the context again is calculated in the following way:

Where $c(\varphi)$ is the frequency of occurring of the symbol φ in any context "#i". Here, i is the alphabet whose occurrence we are looking for, and "#" represents the space. It is written only to increase the readability of space. C is the number of times the context "#i" occurs. C is defined in equation (3).

$$C = \sum \varphi E A^{C}(\varphi) \qquad (3)$$

Where A is the coding alphabet (say in ASCII) in which coding of the occurring alphabet is performed.

The escape probability for a character φ that hasn't appeared in the context before is

$$1 - \sum\nolimits_{\varphi \varepsilon A_{,} c(\varphi) > 0} p(\varphi) = \frac{1}{1+c} \qquad (4)$$

Suppose "q" be the number of characters that have occurred in a context and "a" is the size of the coding alphabet "A" then, the characters that haven't happened yet in the context are "a - q ". For simplicity, we assign an overall coding probability to these no-occurring characters. The overall coding probability is

$$p(\varphi) = \frac{1}{1+c} * \frac{1}{a-1}, c(\varphi) = 0 \qquad (5)$$

Here q denotes the total number of characters that occur in the context.

To compare the proposed approach with the benchmark, we have determined the accuracy of encounter prediction by both techniques. We have used the formula for calculating the accuracy of both approaches: Accuracy = (correctly predicted locations/total observed locations) * 100.

The accuracy was calculated by training the Markov model and proposing a sequential-based prediction model on the traces used (I. Burbey, 2011). The data presented in the proposed method was divided into chunks and was presented during simulation. These models were executed, and predictions made by both models were logged. These predictions were compared to the observed mobility of the users, which further helped calculate the accuracy.

For calculating the accuracy of the proposed model for encounter prediction, we have regarding the data in original traces as correct for the locations and temporal id's from UCSD traces. Therefore, for each user u_1, u_2, u_3, ………….., u_{275}, the predicted encounters were compared with each original trace in the data-set.

Figure 14. And Figure 15 shows the comparison between both approaches.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 06:2023
DOI 10.17605/OSF.IO/RM3UJ

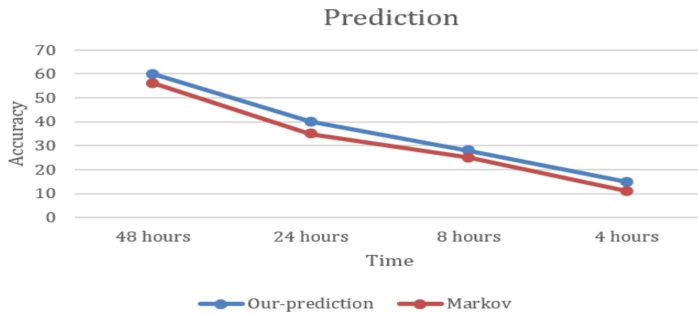

Figure 14: Prediction accuracy comparison

The prediction accuracy of the Markov model is relatively less than the proposed approach. During the time interval of 8 hours, the accuracy of the Markov model is 25%, and the proposed approach shows an accuracy of 28%. With the increased time, although the accuracy improves, the proposed approach's accuracy is higher than the benchmark.

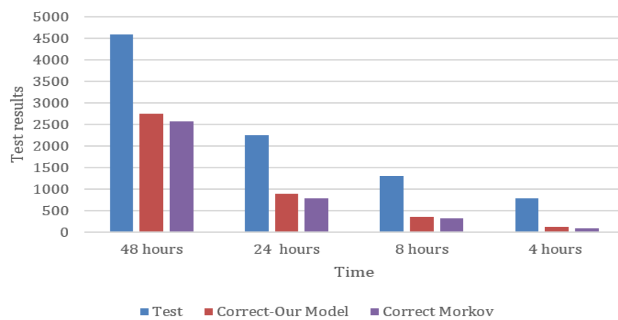

Figure 15: Prediction accuracy comparison

The proposed model A better approach than the existing one. Results of prediction through [12] partial match algorithms are relatively nontrivial. For location prediction, the results through this algorithm are not a single location but multiple locations with different probabilities. Moreover, this algorithm doesn't consider temporal values for prediction, whereas temporal values play quite an essential role in predicting an individual's future location. A comparison between the proposed models based on a sequential algorithm was performed with Markov model-based prediction algorithm. Figure 15. Shows the result for accurate comparison of both models as mentioned above. The results showed that the proposed model predicted locations and encounters with higher accuracy than the Markov model. The highest accuracy was achieved for the predictions made in the time interval of 48 hours.

The proposed model's higher accuracy than the Markov model requires demand-based prediction at a specific time interval. As the Markov model assumes the user to be in the same state, the sequential patterns that a human follows in his daily routine are completely ignored. These sequential mobility patterns determine a particular user's routine and are almost identical daily. Markov model makes the prediction based on randomly chosen data at any timestamp without keeping in view the mobility history of a user and the order in which specific locations were visited. The sequential model

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 06:2023
DOI 10.17605/OSF.IO/RM3UJ

considers the positions, the specific time stamp at which they were visited, and the visiting order. One of the drawbacks of using a sequential algorithm is the processing time is higher in this case, as this model has to process each sequence for accessing the location at a specific time interval. The proposed algorithm's novelty lies in that the data set is divided into chunks, and the time required to process a chunk is comparatively lower than the time needed to process the whole data set. As during simulations, in the case of the Markov model, where the whole data set was used for training the algorithm, the total time for obtaining results was 8 hours. Whereas, for the sequential algorithm with data mining, where the data set was divided into chunks and then trained, the time for getting results was 4 hours.

7. CONCLUSION

In this paper, we have discussed the importance of understanding human mobility patterns. These patterns help define user behaviour in the form of movement. Furthermore, these mobility patterns help to predict the future location. In this work, we have employed data mining techniques and a sequential algorithm for predicting encounters among multiple users. Predicting the encounters reduces resource consumption, i.e., energy and bandwidth during data transmission in daily movements. At first, provided a detailed description of the data set we employed to train the proposed technique. The data set was then pre-processed. The proposed technique first used mobility traces of different users for future location prediction and then employed these future locations for encounter prediction. We have proposed an algorithm for predicting future locations. This algorithm used a data structure called "sequence" for adding visited locations in a chain form. We have also provided a mathematical computational model for the encounter prediction of different users. We have performed simulations and simulated the proposed prediction algorithm for experimentation and benchmarked PPM. The data set was trained on both techniques, and respective prediction accuracies were computed. We have compared users' observed encounters to the proposed system's predicted encounters, to calculate the prediction accuracy results as shown in figure 14 and figure 15. In the future work we will show if the prediction is correct then the data will be transferred through D2D "Device to Device" otherwise the data will be transferred through the internet and we will work with live traces from the cloud.

References

1. Siris VA and Dimopoulos D. Multi-source mobile video streaming with proactive caching and D2D communication. In: *2015 IEEE 16th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)* 2015, pp.1-6. IEEE.
2. Baby B and Pramila RS. Survey on analysis of energy optimization in MANET routing. *International Journal of Engineering & Technology* 2018; 7: 1951-1955.
3. Chaudhari SS and Biradar RC. Traffic and mobility aware resource prediction using cognitive agent in mobile ad hoc networks. *Journal of Network and Computer Applications* 2016; 72: 87-103.

4. Chellapa R, Jennings A and Shenoy N. A comparative study of mobility prediction in fixed wireless networks and mobile ad hoc networks. In: *IEEE International Conference on Communications, 2003 ICC'03* 2003, pp.891-895. IEEE.
5. Vu TK and Kwon S. On-demand routing algorithm with mobility prediction in the mobile ad-hoc networks. *arXiv preprint arXiv:160908141* 2016.
6. Munaye YY, Juang R-T, Lin H-P, et al. Radio Resource Allocation for 5G Networks Using Deep Reinforcement Learning. In: *2021 30th Wireless and Optical Communications Conference (WOCC)* 2021, pp.66-69. IEEE.
7. Ziebart BD, Maas AL, Dey AK, et al. Navigate like a cabbie: Probabilistic reasoning from observed context-aware behavior. In: *Proceedings of the 10th international conference on Ubiquitous computing* 2008, pp.322-331.
8. Elnekave S, Last M and Maimon O. Predicting future locations using clusters' centroids. In: *Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems* 2007, pp.1-4.
9. Roddick JF and Spiliopoulou M. A survey of temporal knowledge discovery paradigms and methods. *IEEE Transactions on Knowledge and data engineering* 2002; 14: 750-767.
10. Eagle N, Clauset A and Quinn JA. Location Segmentation, Inference and Prediction for Anticipatory Computing. In: *AAAI Spring Symposium: Technosocial Predictive Analytics* 2009, pp.20-25.
11. Holbach H and Bourassa MA. The effects of gap wind induced vorticity, the ITCZ, and monsoon trough on tropical cyclogenesis. In: *2012 International Ocean Vector Winds Science Team Meeting* 2012.
12. Theerthagiri P. Mobility prediction for random walk mobility model using ARIMA in mobile ad hoc networks. *The Journal of Supercomputing* 2022: 1-32.
13. UCSD"https://crawdad.org/ucsd/cse/20080825/". 2008-08-25.
14. A. Memon, W. Liu and A. Al-Anbuky, "CatchMe If You Can: Enable Sustainable Communications Using Internet of Movable Things," 2016 IEEE 14th Intl Conf on Dependable, Autonomic and Secure Computing, 14th Intl Conf on Pervasive Intelligence and Computing, 2nd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech), Auckland, New Zealand, 2016, pp. 947-952, doi: 10.1109/DASC-PICom-DataCom-CyberSciTec.2016.163.
15. Memon, A.; Kilby, J.; Breñosa, J.; Espinosa, J.C.M.; Ashraf, I. Analysis and Implementation of Human Mobility Behavior Using Similarity Analysis Based on Co-Occurrence Matrix. *Sensors* 2022, *22*, 9898. https://doi.org/10.3390/s22249898
16. Memon, A.; Kilby, J.; Breñosa, J.; Espinosa, J.C.M.; Ashraf, I. Analysis and Implementation of Human Mobility Behavior Using Similarity Analysis Based on Co-Occurrence Matrix. *Sensors* 2022, *22*, 9898. https://doi.org/10.3390/s22249898
17. Rasool, N.; Khan, S.; Haseeb, U.; Zubair, S.; Iqbal, M. waseem; Hamid, K. Scrum And The Agile Procedure’s Impact On Software Project Management. *Jilin Daxue Xuebao Gongxueban Journal Jilin Univ. Eng. Technol. Ed.* 2023, *42*, 380–392, doi:10.17605/OSF.IO/MQW9P.
18. Hamid, K.; Iqbal, M. waseem; Niazi, Q.; Arif, M.; Brezulianu, A.; Geman, O. Cloud Computing Network Empowered by Modern Topological Invariants. *Appl. Sci.* 2023, *13*, 18, doi: 10.3390/app13031399.
19. Hamid, K.; Iqbal, M. waseem; Fuzail, Z.; Muhammad, H.; Basit, M.; Nazir, Z.; Ghafoor, Z. Detection of Brain Tumor from Brain MRI Images with the Help of Machine Learning & Deep Learning. 2022, doi:10.22937/IJCSNS.2022.22.5.98.

20. Hamid, K.; Iqbal, M. waseem; Niazi, Q. Discovering Irregularities from Computer Networks by Topological Mapping. *Appl. Sci.* 2022, *12*, 1–16, doi:10.3390/app122312051.

21. Hamid, K.; Iqbal, M. waseem; Ashraf, M.U.; Gardezi, A.; Alqahtani, M.; Shafiq, M. Intelligent Systems and Photovoltaic Cells Empowered Topologically by Sudoku Networks. 2022, *74*, 4221–4238.

22. Hamid, K.; Bhatti, S.; Hussain, N.; Fatima, M.; Ramzan, S.; Iqbal, M. waseem Irregularity Investigation Of Certain Computer Networks Empowered Security. 2022, *41*, 75–93, doi:10.17605/OSF.IO/TJ6XN.

23. Hamid, K.; Iqbal, M. waseem; Muhammad, H.; Fuzail, Z.; Nazir, Z. Anova Based Usability Evaluation Of Kid's Mobile Apps Empowered Learning Process. *Qingdao Daxue XuebaoGongcheng JishubanJournal Qingdao Univ. Eng. Technol. Ed.* 2022, *41*, 142–169, doi:10.17605/OSF.IO/7FNZG.

24. iftikhar, tayyaba; Zubair, S.; Iqbal, M. waseem; Hamid, K.; Nazir, A. Cost Effective About Iot-Based Accordingly To Weather Monitoring System Development And Improve Average Of Crops. *Jilin Daxue Xuebao GongxuebanJournal Jilin Univ. Eng. Technol. Ed.* 2023, *42*, 123–152, doi:10.17605/OSF.IO/TKAH8.

25. Riasat, H.; Akram, S.; Aqeel, M.; Iqbal, M. waseem; Hamid, K.; Rafiq, S. Enhancing Software Quality through Usability Experience and Hci Design Principles. 2023, *42*, 46–75, doi:10.17605/OSF.IO/MFE45.

26. Hussain, D.; Rafiq, S.; Haseeb, U.; Hamid, K.; Iqbal, M. waseem; Aqeel, M. Hci Empowered Automobiles Performance By Reducing Carbon-Monoxide. 2022, *41*, 526–539, doi:10.17605/OSF.IO/S5X2D.

27. Hamid, K.; Muhammad, H.; Iqbal, M. waseem; Nazir, A.; shazab; Moneeza, H. Ml-Based Meta Model Evaluation Of Mobile Apps Empowered Usability Of Disables. *Tianjin Daxue Xuebao Ziran Kexue Yu Gongcheng Jishu BanJournal Tianjin Univ. Sci. Technol.* 2023, *56*, 50–68.

28. Hamid, K.; Muhammad, H.; Iqbal, M. waseem; Bukhari, S.; Nazir, A.; Bhatti, S. Ml-Based Usability Evaluation Of Educational Mobile Apps For Grown-Ups And Adults. *Jilin Daxue Xuebao GongxuebanJournal Jilin Univ. Eng. Technol. Ed.* 2022, *41*, 352–370, doi:10.17605/OSF.IO/YJ2E5.

29. Hamid, K.; Iqbal, M. waseem; Nazir, Z.; Muhammad, H.; Fuzail, Z. Usability Empowered By User's Adaptive Features In Smart Phones: The Rsm Approach. *Tianjin Daxue Xuebao Ziran Kexue Yu Gongcheng Jishu BanJournal Tianjin Univ. Sci. Technol.* 2022, *55*, 285–304, doi:10.17605/OSF.IO/6RUZ5.

30. Hamid, K.; Iqbal, M. waseem; Muhammad, H.; Basit, M.; Fuzail, Z.; † Z.; Ahmad, S. Usability Evaluation of Mobile Banking Applications in Digital Business as Emerging Economy. 2022, 250, doi:10.22937/IJCSNS.2022.22.2.32.

31. Muhammad, H.; Basit, M.; Hamid, K.; Iqbal, M. waseem; Shahzad, S.; Muneem, F.; Shaheryar, M. *Usability Impact of Adaptive Culture in Smart Phones*; 2022;

32. Nazir, Z.; Iqbal, M. waseem; Hamid, K.; Muhammad, H.; Nazir, A.; Hussain, N.; ann, Q. Voice Assisted Real-Time Object Detection Using Yolo V4- Tiny Algorithm For Visual Challenged. 2023, *56*, 2023, doi:10.17605/OSF.IO/APQYH.

33. Hamid, K.; Iqbal, M. waseem; Zain, S.; Shairoze, M.; Shaheryar, M.; Zubair, M.; Rehman, M. Wireless Router Forensics: Finding Artifacts of Suspect Traces with a Raspberry Pi and Kali Linux. 2022, *41*, 502–525, doi:10.17605/OSF.IO/RK4YF.