IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 4, May 2023

User Behavior Charecterization using Mobile Internet Usage

Prof Nandini S, Devaraj Patil, G Pawan Nikil, H. Hari Raj, Madhava K V
Department of Information science and Engineering
S J C Institute of Technology, Chickballapura, Karnataka, India

Abstract*: Smart gadgets provide ubiquitous mobile Internet connection, which is hastening the expansion of mobile Internet. This article examines the mobile user behavior from three perspectives using mobile traffic statistics gathered at China's central urban 2G and 3G networks over a week: Data use, mobility, and application use are the first three categories. To analyze the resource usage on mobile Internet, it divides mobile users into various groups. It notices that users with high levels of mobility and traffic tend to burn up a lot of radio and data resources at once. Users' application access behavior is intimately tied to their data usage and mobility patterns. Users can be grouped based on how they use an application, and different application categories can be found by the methods they use.*

Keywords: K-means Clustering , KNN , K Nearest Neighbor..

I. INTRODUCTION

Smart gadgets provide ubiquitous mobile Internet connection, which is hastening the expansion of mobile Internet. This article examines the mobile user behavior from three perspectives using mobile traffic statistics gathered at China's central urban 2G and 3G networks over a week: Data use, mobility, and application use are the first three categories. To analyze the resource usage on mobile Internet, it divides mobile users into various groups. It notices that users with high levels of mobility and traffic tend to burn up a lot of radio and data resources at once. Users' application access behavior is intimately tied to their data usage and mobility patterns. Users can be grouped based on how they use an application, and different application categories can be found by the methods they use.

II. LITERATURE SURVEY

A novel framework was put forth by Matias Callara and Patrice Wira in 2016 to analyze user behavior in a distributed computing environment using machine learning algorithms. Discriminating between close user groups is the goal. Users in these groups all exhibit similar behaviors. The user's behavior-related events are recorded and added to a database. To ascertain the user groups, a method is created. For each user, a personalized prediction of application launches and session opens is made using a non-parametric method of generating a probability density. Within a complete virtualization environment for workstations and applications in actual hospital conditions, these algorithms have been applied and proven effective. In order to analyze user behaviors in a distributed computing environment, Matias Callara and Patrice Wira (2016) suggested a new framework employing machine learning methods. The goal is to distinguish between similar user groups. Users who exhibit similar behaviours make up these categories. Events pertaining to the user's actions are noted and added to a database. To identify the user groups, a strategy is devised. Application launches and session openings are forecasted individually for each user using a non-parametric method of predicting a probability density. These algorithms have been put into practice and shown to operate well in a complete virtualized environment for workstations and applications in a hospital.

Researchers Alyssa Pea, Ehsanul Haque Nirjhar, and Andrew Pachuilo (2018) [28] have shown that "Explainable AI" enables analysts to make more educated conclusions. AI models are frequently understood, diagnosed, and improved through visualizations. However, it is not yet apparent what kinds of interactions and how the visualizations are interpreted are acceptable for a specific model. It may be helpful to understand how users engage with AI visualizations and to construct a naturalistic model of explanation if research into sense making for visual analytics is continued.

The rapid development of electric vehicles (EVs) has presented issues to power grids and transportation networks, according to Ziqi Zhang and Bo Liu's research from 2019 [31]. It is necessary to accurately capture EV consumers' usage patterns in order for EVs to engage with electrified transportation networks. As a result, this study suggests a model that is driven by the complete data chain (FDC) to mine all of the features of EVs. A driving portrait and origin-destination (OD) distribution are used to first mine the driving characteristics of 150 private electric cars (PREVs), 100 commercial electric vehicles (CEVs), and 50 official electric vehicles (OEVs) in Chongqing, China. Last but not least, the outcomes of the analysis of EV user characteristics are examined, and regional charging load analysis and a comparison of urban road traffic flow. The research results offer a data source and a user behavior model for the design, management, and control of transportation and power grids.

The goal of Luca Vassio and Marco Mellia's study from 2020 [32] was to determine whether the need to comprehend how people interact with the web motivated the research I undertook for my PhD. ISPs and network administrators may better see and understand how users and online services evolve over time thanks to this information. My study focuses on two complementing areas thanks to user traffic traces and logs: This work demonstrates how to model users' behaviour and recreate users' online activity from passive measures using (i) data analytics and (ii) user modelling. It introduces machine learning techniques to recognise the websites and pages that users voluntarily visit. It emphasizes the development of gadget usage, navigational organization, and interactions with social networks and search engines.

Sentiment analysis advanced with the introduction of the concept paper Understanding the behaviour of individuals or a specific user using his comments or tweets in various social media.Nirmal Varghese Babu and Fabeela Ali Rawther in 2008 [33]. To comprehend the general feelings present in the data gathered from various social media, sentiment analysis or opinion mining is used. Because of the Internet and many social media platforms like Twitter, Facebook, Instagram, and others where people may share their opinions, people are more exposed to the outside world.

III. METHODOLOGY

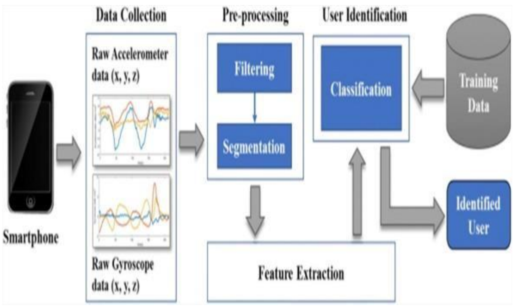

Fig 3.1 System architecture

The system architecture shows how semi- supervised data collecting works. Up until the element's creation date, public information is saved in descending chronological order. In order to move on to the next instruction, the Agent program runs a script to manage the node list and check for the node's existence. An access token is necessary for servers to connect with each other for security concerns. The program agent modifies replies and related filtering procedures in addition to renewing the access token in order to store the data in the dataset.

K-means Clustering

The unsupervised learning algorithm K- Means Clustering is used to solve the clustering problems in machine learning or data science. This chapter explains the K- means clustering algorithm and shows us how to use Python to implement it.

K-Means algorithm functions:

Step-1: Select K to get the number of clusters.
Step 2: Pick K locations or centroids at random. It might not be the input dataset.
Step 3: Assign each data point to its nearest centroid, which will create the K clusters that have been predetermined.
Step 4: Determine the variance and relocate each cluster's centroid.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 4, May 2023

Step 5: Re-assign each data point to the new nearest centroid of each cluster by repeating the third step.
Step 6: move to step 4 if there is a reassignment; otherwise, move to step 7.
Step 7: The model is prepared for completion.

KNN:
K-NN algorithm:
Step-1: Choose the Kth neighbor's number.
Step-2:Calculate the Euclidean distance between K neighbours in step two.
Step 3: Choose the K nearest neighbours based on the Euclidean distance that was calculated.
Step 4: Count the number of data points in each category among these k neighbours.
Step 5: Assign the fresh data points to the category where the neighbor count is highest.
Step 6: Our model is complete.
Classify a new data point in order to use it. Take into account fig. 6.2 below.

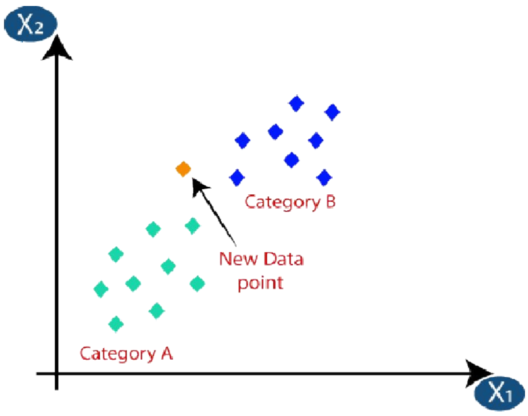

Fig 3.2. K-Nearest Neighbor

First, we'll decide on the number of neighbors; we'll go with k=5.
Then, we will calculate the Euclidean distance between the data points. The distance between two points, which we have already examined in geometry, is known as the Euclidean distance. It is calculable as follows:

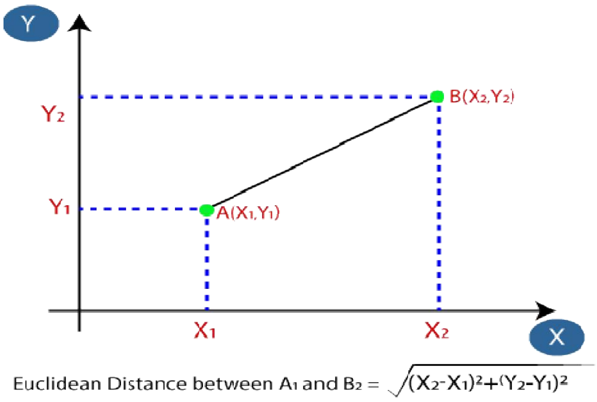

Fig 3.3. Euclidean Distance

By computing the Euclidean distance, we were able to determine the distances between the nearest neighbours, which were determined to be three in category A and two in category B. Think on the photo below:

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 4, May 2023

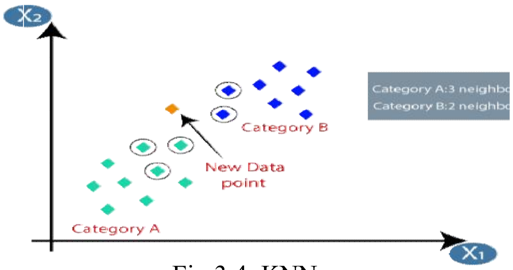

Fig 3.4. KNN

The three closest neighbors are from category A, as can be shown in fig. 6.4, hence this new data point must also be from group A.

VI. CONCLUSION

In this research, we examined the mobile user behaviour with a thorough multi-dimension analysis, focusing on three features: 1. Data usage, 2. Mobility pattern, and 3. Application usage. We did this by using real traffic data gathered from mobile Internet in a significant metropolitan area of countries.

To analyze the resource usage on mobile Internet, we divide mobile users into various groups. Users with high levels of mobility and traffic tend to simultaneously demand a lot of radio and data resources. Users' application access behavior is intimately tied to their data usage and mobility patterns.

ACKNOWLEDGEMENT

We would like to thank our college, SJC Institute of Technology, and our mentor professor. Nandini S ma'am and our project coordinator Aravind Tejas Chandra who have provided us with the opportunity to work on this project and given us support with guidance to make this project a success.We also want to express our gratitude to our coworkers for their support, enthusiasm, and contributions to this effort. Without them, our project wouldn't be a success.

REFERENCES

[1]. The work "User Behaviour Analysis with Machine Learning Techniques in Cloud Computing Architectures," written by Matias Callara and Patrice Wira, was presented at the 2018 International Conference on Applied Smart Systems (ICASS).

[2]. "Detecting Changes in User Behavior to Understand Interaction Provenance during Visual Data Analysis," Alyssa Pena, Ehsanul Haque, and Andrew Pachuilo, ceur- ws.org/Vol-2327/IUI19WS-UIBK-3.pdf.

[3]. Modelling and Analysis of User Behaviour in Electric Vehicles Based on Full Data Chain Driven," by Ruisheng Wang, Qiang Xing *, Zhong Chen, Ziqi Zhang, and Bo Liu Sustainability 2022, 14, 8600, driven. https://doi.org/10.3390/su14148600.

[4]. "Data Analysis and Modelling of Users' Behavior on the Web," by Luca Vassio and Marco Mellia. Dissertation Sessions for the 2019 IFIP/IEEE International Symposium on Integrated Network Management

[5]. Nirmal Varghese Babu and Fabeela Ali Rawther's study, "User Behaviour Analysis on Social Media Data Using Sentiment Analysis or Opinion Mining," The International Research Journal of Technology and Engineering (IRJET) has an e-ISSN of 2395-0056 and a p-ISSN of 2395-0072. Its volume six issue six is from June 2019.

[6]. An empirical examination of user content creation and usage patterns on mobile websites. by A. Ghose and S. P. Han. 2011; "Manage Science," vol. 57, no. 9, p.1671–1691.

[7]. The authors present an unsupervised framework for sensing individual and cluster behaviour patterns from data on human mobility. J. Zheng and L. M. Ni in Proc. ACM Conf. Ubiquitous Comput., 2012, pp. 153–162.

[8]. Characterising data usage trends in a large cellular network was discussed by Y. Jin et al. in Proc. Cellular Networks, Operation, Challenges, and Future Design: An ACM SIGCOMM Workshop, 2012, pp. 7–12.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 4, May 2023

[9]. The study "Analysis of Users and Non- Users of Smartphone Applications," 27th volume of Telematics Inform, no. 3, pp. 242- 255, August 2010. H. Verkasalo, C. López- Nicolás, F. J. Molina-Castillo, and H. Bouwman.

[10]. Characterising web use on smartphones, in Proc. SIGCHI Conf. Human Factors Comput. Syst., 2012, pp. 2769– 2778. C. Tossell, P. Kortum, A. Rahmati, C. Shepard, and L. Zhong.

[11]. The article "Anatomizing application performance differences on smartphones," published in Proc. 8th Int. Conf. Mobile Syst., Appl., Services, 2010, pp. 165-178, by J. Huang, Q. Xu, B. Tiwana, Z. M. Mao, M. Zhang, and P. Bahl.

[12]. ACM Trans. published a paper titled "Mobile information access: A study of emerging search behaviour on the mobile Internet." Web, vol. 1, no. 1, p. 4, May 2007. K. Church, B. Smyth, P. Cotter, and K. Bradley.