Internet behavioral models for improving internet quality of service or user profiling: a systematic literature review

Zhang Lei[1,2], Nor Shahniza Kamal Bashah[1]
[1]College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, Shah Alam, Malaysia
[2]Faculty of Information Engineering, Anhui Business and Technology College, Hefei, China

Article Info

Article history:

Received Dec 13, 2023
Revised Mar 28, 2024
Accepted Apr 16, 2024

Keywords:

Behavior analysis
Behavioral characteristics
Internet behavior model
Internet quality of service
User profiling

ABSTRACT

Internet behavior models have found applications across diverse domains, notably in internet addiction, customer satisfaction analysis, user purchasing behavior prediction, and optimizing internet of things (IoT) sensor performance. However, a notable gap exists in exploring these models in enhancing internet quality of service (QoS), specifically in campus settings, intricately linked to the nuances of students' online behavior. This study elucidates the strategic utilization of internet behavioral models for augmenting internet QoS and facilitating user behavior analysis. Creating datasets grounded in internet users' access behavior represents a pivotal phase, with explicit, implicit, and mixed methods emerging as the prevailing approaches. In this comprehensive literature review, we systematically scrutinized the methods, techniques, and inherent characteristics of constructing internet behavior models according to a systematic literature review process. The qualitative findings extracted from the systematic review encapsulated 1,046 articles, meticulously classified according to predefined inclusion and exclusion criteria. Subsequently, 35 articles were judiciously selected for in-depth analysis. This study culminated in identifying the most pertinent methodologies and salient features pivotal to construct robust internet behavior model for improving internet QoS and user experience.

Corresponding Author:

Nor Shahniza Kamal Bashah
College of Computing, Informatics and Mathematics, Universiti Teknologi MARA
40450 Shah Alam, Selangor, Malaysia
Email: shahniza@uitm.edu.my

1. INTRODUCTION

Internet behavioral models allow the creation of user analytics in virtual environments based on their behavior and characteristics when accessing the internet. Internet behavior models are obtained from multiple data sources (such as logs, interviews, and questionnaires) through different processes (such as machine learning, deep learning, and big data). The main methods for generating the datasets needed to build internet behavior models are explicit, implicit, and mixed methods.

Explicit methods work by collecting results from questionnaires, interviews, and business records, obtained from a population, which can be used in manually created models. Implicit methods use machine learning or deep learning to extract features from data sets such as logs, sessions, and traffic. These features are later used to create or improve these models. In a mixed approach, the dataset is achieved by combining the other two methods mentioned earlier.

Due to the efficiency and ease of use of computing devices, many studies improve user satisfaction, analyze the causes of internet addiction, or predict user purchasing behavior by analyzing users' internet behavior. However, through a systematic review of these articles, this study found that although there have been previous studies on different topics of internet behavior modeling, there is still insufficient research has been made on internet behavior models specifically aimed at improving the internet quality of service (QoS) of campus and user experience. Understandably, there is an extensive need of an internet usage behavioral model based on the relevant attributes discovered from literature review (LR) that dedicated for the internet QoS campus to ensure the issue on insufficient bandwidth and maximum excess on multimedia files are attainable. This will establish user profiling for improving the QoS of the campus network, which is the final artefact expected by this study in response to the problems described and discussed earlier. While at the same time, the behavioral model could provide critical information to guide the choice of internet bandwidth allocation for specific users.

In the previous research, the author analyzed the factors affecting the internet QoS in campus and found that the internet QoS in campus was significantly related to the App's type, access time, location and dormitory type of users when accessing the internet, but was not significantly related to gender, grade, major, user account group and user bandwidth. Next, the author needs to construct internet behavior models to find out which of these factors are highly correlated with campus internet QoS. Two problems need to be solved. First, the internet behavior model needs to be built to prove the correctness of the influencing factors; Second, use the internet behavior model to find out the size of each factor. It must be mentioned that the importance of conducting a systematic review of the literature is to fully understand all the different methods and techniques used by researchers in the field to build models of internet behavior and the main features of internet behaviors used in building models.

The structure of this paper is as follows: section 2 provides an overview of the theoretical background. Next, section 3 shows the methods used in conducting this review. The results from the systematic literature review are then analyzed and described thoroughly in section 4. Finally, section 5 concludes the paper with some insights of future works.

2. THEORETICAL BACKGROUND

This section is structured in two parts. The first subsection introduces the concept of internet behavior model, and the second subsection examines the concepts of user behavior model. Based on the literature then, will explore on how these relate to the myriad of model developed to measure the impact user profiling in improving the internet QoS.

2.1. Internet behavior model

Internet behavior model is a model that utilizes technology and tools to analyze and summarize internet traffic behaviors [1]. It employs information theory methods to characterize the stability, temporal dynamics of behaviors on the internet, as well as traffic patterns of network systems and internet applications. Many studies have utilized the internet behavior model to achieve various objectives, such as predicting peak periods of network traffic [2], enhancing information collection efficiency, reducing time costs [3], and forecasting user lost rates [4]. In the creation of internet behavior models, researchers utilize various data sources including network traffic data, interviews, surveys, and business records. Kovacevic and Kascelan [5] examined gender differences in internet usage patterns using interview data. Novela *et al.* [6] constructed models using survey data to explain the impact of hedonic and utilitarian motivations on attitudes towards online purchasing. Wu *et al.* [4] developed efficient user churn prediction models by profiling users using business record data.

2.2. User behavior model

User behavior model (UBM) is an exploration and utilization of user interests' representation from behavioral history [7], which has been extensively used in recommender systems. Existing UBM research can be categorized into four different directions, including traditional UBM, long sequence UBM, multi-type UBM, and UBM with auxiliary information [8]. Existing studies have demonstrated the power of using user behavior models to solve various problems. For instance, Bahra and Pierre [9] proposed a novel bidirectional trajectory prediction model to simulate user mobility behaviors. They leveraged the potential advantages of bidirectional gated recurrent units (GRU) for accurate predictions. Certainly, modeling user behavior can also be achieved through smartphone power consumption to explore the relationship between energy consumption and user behavior [10]. In the field of architecture, Zhou *et al.* [11] introduced an approach for modeling occupant behaviors in office buildings, providing a theoretical basis for predicting occupant behaviors and simulating building energy consumption. These studies model user behavior through data to extract user behavior features from pre-processed data. However, Vassio and Mellia [12] argue that these features change with people's intrinsic habits and interactive behaviors, and the effectiveness of personalized models can be improved through time optimization.

3. METHOD

This article systematically reviews the literature on behavioral models in the field of the internet, encompassing the definition of research questions, the search process, inclusion and exclusion criteria, article quality evaluation, and result analysis. This approach is preferred as a systematic literature review serves as "A way of mapping areas of uncertainty and identifying where little or no relevant research has been conducted, but where new research is needed." This literature review aims to pinpoint areas offering opportunities for foundational research in internet QoS.

By gathering essential information, this study aims to enhance our understanding of the functionality of users' internet behavioral models. Armed with this information, the study seeks to formulate an internet behavior model specifically tailored for school users, ultimately contributing to the improvement of internet QoS. The overarching goal is to provide users within campus networks with a faster and more stable internet access experience.

3.1. Research questions

Research questions serve as crucial guideposts, delineating the scope of our study and steering clear of extraneous information. These questions were carefully formulated to ensure a focused investigation. Defining the boundaries of our research helps maintain relevance and coherence throughout the study. The following questions were employed in this review:

a. Q1: What is the purpose of creating an internet behavior model?
b. Q2: What are the steps to create an internet behavioral model?
c. Q3: What techniques are used for the construction of internet behavior models?
d. Q4: What characteristics are used to define the internet behavior model?

Objective interpretation:

a. O1: Aims to uncover the primary applications and domains where internet behavior models find predominant utility through a comprehensive literature review.
b. O2: Serves as a preliminary exploration to grasp the fundamental methodology employed in constructing internet behavior models.
c. O3: Emerges from the necessity to comprehend the spectrum of technologies available for building these models, identifying the most effective ones.
d. O4: Explores the diverse characteristics influencing user behavior in these models, encompassing terminal specifications, temporal considerations, application types, geographical aspects, and other factors shaping user interaction with the internet.

3.2. Search string

The goal of executing the search string is to easily capture in a structured way all results related to internet behavior modeling and internet QoS in scientific databases. The basis and selection of the terminology used are based on the research topic. i) internet behavior model as a subject of research and similar terms; ii) User profiling or user analysis, construct the models by user profiling or user analysis; iii) Quality of service, as a research focus, seeks to use these models to improve internet service quality; and iv) Behavior analysis, analyzing user behavior by behavior model.

This research performed search strings in four scientific databases: IEEE Xplore, ACM Digital Library, Science Direct, and Scopus. Each database yielded different result terms based on the specified same criteria. This variation underscores the importance of using multiple sources to ensure comprehensive coverage of the research topic. The search strings are ("Internet behavior model" or "user profiling" or "user analysis") and ("Quality of service" or "behavior analysis").

3.3. Inclusion and exclusion criteria

Inclusion criteria were developed to classify articles found in scientific databases with the aim of obtaining answers to the posed research questions in relevant publications. This results analysis only includes documents that meet the following criteria:

a. IC1: This document demonstrates different techniques for building internet behavior models.
b. IC2: This document contains learning techniques for modeling internet behavior through validation in a campus environment.
c. IC3: This document describes characteristics relevant to user use of the internet.
d. IC4: This document focuses on internet QoS.

The exclusion criteria were designed to eliminate any articles that were not relevant to our study. Papers meeting at least one of the specified criteria were excluded to maintain focus on pertinent research. This approach helped to refine the dataset, ultimately leading to a more accurate and meaningful analysis. Papers that met at least one of the following criteria were excluded:

a. EC1: This document is for research on IoT optimization, cybercrime, personal privacy, cyberattacks, cloud computing resource scheduling, or 5G protocol optimization.
b. EC2: This document mentions the basic characteristics used in behavioral models, but in terms of real-life behavior rather than internet access behavior.
c. EC3: The article describes the improvement of internet QoS but does not mention the characteristics of users' internet behavior.
d. EC4: Documents come from other internet platforms (such as GitHub or BBS) rather than journals or published conferences.

4. RESULTS AND DISCUSSION

The searching process involves different databases using previously defined search strings. This comprehensive approach ensured that our study broadly collected relevant articles. Among the search results of these four databases, the Science Direct database has the most search results, while the IEEE Explore database has the fewest. Detailed search results are shown in Table 1.

Table 1. Without filter or time period established

Database	Number of papers
ACM Digital Library	148
IEEE Explore	22
Science Direct	802
Scopus	74
All	1,046

4.1. Results

A period of no more than five years was established to take advantage of more recent results based on research and consider more useful information for the study. It also needs to be confirmed that the research is conducted in the field of computational science and not in other scientific fields such as medicine, environmental science, or IoT. For example, research presented behavioral modeling and formal verification of a hybrid machine learning-based fault prediction model in the field of IoT [13]. Another research provides a general four-layer framework for modeling these new trends that novel services for freight transportation and city logistics improve the efficiency of goods transportation [14]. Zhou *et al.* [11] proposed a modeling method of occupant behavior for building energy consumption simulation. The articles as this research were excluded from this study. Likewise, there are requirements that the article must meet the questions in Part A of section 2:
a. REQ1: Does this article describe the creation of an internet behavioral model?
b. REQ2: Does the research report involve internet behavior model building technology?
c. REQ3: Is the study based on reliable sources to define concepts relevant to the topic?
According to the previous requirements, each article found in the database is organized in two steps: i) Through the scientific database, a total of 1,046 articles met the inclusion and exclusion criteria of the systematic review as shown in Table 2; and ii) Sort them into folders so that they are sorted according to the previous requirements shown above, like this: "Yes, Partial and No".

To ensure that the references can focus on internet behavior models, each article found in the database is organized in two steps according to the previous requirements. First, 1,046 articles retrieved from four databases were streamlined according to the inclusion and exclusion criteria. In the end, 854 articles were excluded, and 192 were eligible, as shown in Table 2. Secondly, these articles are classified according to the requirements of REQ1-REQ3. Documents classified as "Yes" mention the requirements of both parts. Those marked "Partial" fulfil at least one of the requirements. Finally, documents classified as "No" are suitable for another type of study. This determines whether the total number of articles found is meaningful for the research effort. The total number of documents that set applicable criteria for inclusion and exclusion for this study is shown in Table 2. The number of documents related to the study is shown in Table 3.

Table 2. Filtered total paper results

Database	Number of files	Exclude	Include
ACM Digital Library	148	84	64
IEEE Explore	22	14	8
Science Direct	802	703	99
Scopus	74	53	21
All	1,046	854	192

Table 3. Results after paper classification

Database	Related results
ACM Digital Library	14 out of 64
IEEE Explore	1 out of 8
Science Direct	17 out of 99
Scopus	3 out of 21
All	35

4.2. Analysis result

This study analyzes results obtained previously to answer the four research questions mentioned earlier. First, the research topics of all research articles were clustered and summarized into eight topics. Then, the methods used in each article were classified at different stages, and quantitative statistics were made. The number of research methods used for each research topic is shown in Figure 1.

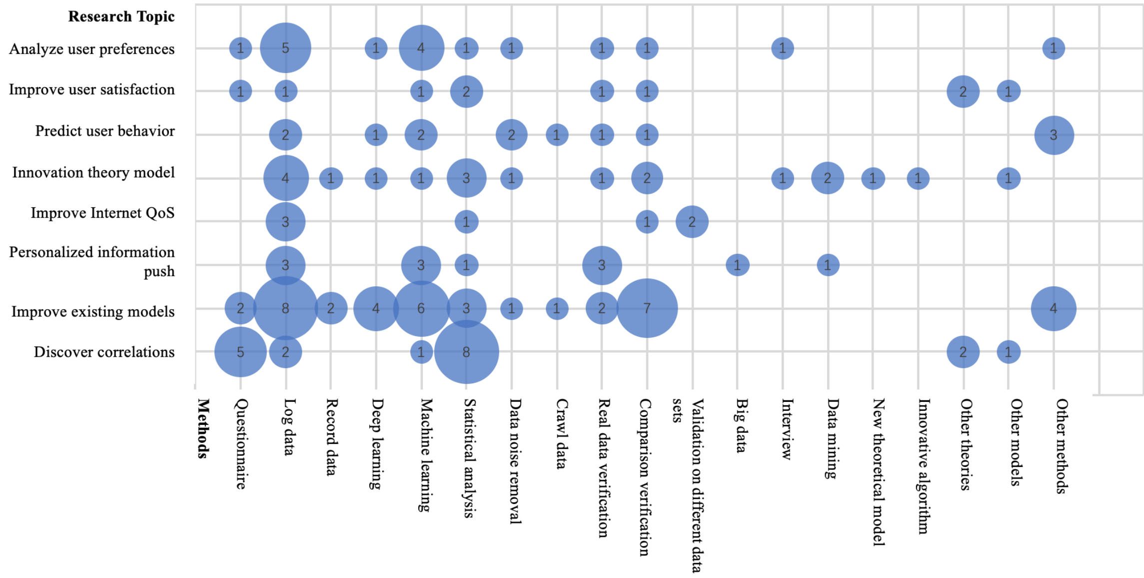


Figure 1. Bubble chart of research topics and methods used

Q1: What is the purpose of creating an internet behavior model?

By scrutinizing the classification results from prior studies, we have structured the objectives behind creating an internet behavior model, encompassing eight key aspects: improving existing models, exploring behavioral correlations, personalized information push, enhancing QoS, theoretical model innovation, user behavior prediction, user satisfaction improvement, and user preference analysis. Notably, the top three categories with the highest paper count are improving existing models (10 papers), exploring behavioral correlations (5 papers), and theoretical model innovations (5 papers); user preference analysis ranks fourth with 4 papers. The research purposes of improving internet QoS, personalized information push, and user behavior prediction each have 3 papers, while user satisfaction improvement holds the lowest count with 2 papers.

In the context of studies aimed at improving internet QoS, Yuan *et al.* [15] conducted an analysis of video data from users and programs. They subsequently developed a visual and user-friendly model to address the challenge of network congestion resulting from a large user base and substantial data traffic, thereby impacting the user's online experience. Soo *et al.* [16] focused on enhancing wireless local area network (WLAN) performance and quality through the strategic load balancing of users (i.e., stations and/or devices) among access points (APs) in an extended service set (ESS). Sen *et al.* [17] managed congestion by offering time-varying discounts to incentivize users to shift some data traffic temporally. Another innovative approach was taken by Kwon *et al.* [18] who tackled network congestion problems by providing time-varying discounts to incentivize users to temporarily shift some data traffic.

Q2: What are the steps to create an internet behavioral model?

The creation process of the Internet behavioral model based on improving internet QoS covers four basic stages: creating data sets, data analysis, creating models, and validating models. Each of these stages is crucial for developing a comprehensive and effective model. However, it is important to note that not all articles contain these four stages, as their inclusion mainly depends on the specific research goals of the article. For instance, some articles focus on improving existing models by refining their parameters or algorithms, while others might concentrate solely on creating new models without proceeding to the verification stage. This variability reflects researchers' diverse approaches to address different aspects of internet service quality.

4.2.1. Create dataset stage

Creating a dataset is an important part of creating a model of internet behavior. An analysis of previous research papers found that there are explicit methods of obtaining data, which are methods of directly contributing personal data in different forms, such as questionnaires, interviews, and service registration. There are also implicit methods, which collect personal data from the system or server. These data are historical data obtained through web logs, transaction records, system logs, and traffic logs, or real-time data obtained through crawlers, and traffic crawling. There are also mixed methods that mix explicit data and implicit data into a dataset by establishing a mixed dataset.

The difference between implicit and explicit methods is the speed at which personal data is collected. Implicit methods have one major advantage over explicit methods is that the researchers only must collect data from the servers or systems. In explicit methods, data must be captured manually, which makes the process slow. Sometimes users are unwilling to cooperate in filling out forms. The mixed method requires more effort than the other two methods, but it gives better predictions or recommendations content when using a server or system, since combining both methods gives you more information on behavioral preferences. Zhu *et al.* [19] integrated the heterogeneous information of the target users such as the user's app, time, and location, into users' behavior trajectories to model the users' app usage preferences. However, some studies used two or three different datasets, but these datasets are all implicit datasets, or all explicit datasets, so it does not belong to the mixed method. For example, Li *et al.* [20] utilized network footprint data (NFP) which consists of deep packet inspection (DPI) data from internet service providers (ISPs) and crawler data from web for analyzing the users' behavior. Liu and Wu [21] used a combination of interviews and observations to understand the user's media product preferences, habits, and the experience of several innovations in the product.

In previous research, 24 articles used implicit data acquisition methods and 7 articles used explicit methods, but only 4 articles used mixed methods. Many articles use datasets from more than one source. For example, 24 articles used logs, 8 articles using traffic logs, 5 articles using web logs, and included other log types such as access logs, operating system logs, session logs, server logs, and transaction or usage records.

Among all the studies using traffic logs as datasets, Adeyemi *et al.* [22] used the machine learning and statistical toolbox in MATLAB 2016a software to statistically calculate the trend of data download and upload traffic in the smart campus during a 12-month period to construct internet data traffic prediction models to guarantee efficient QoS for computer networks and applications. Oliveira *et al.* [2] proposed an adaptive network prediction model called ANP based on the observed network traffic behavior to allow appropriate network resource allocation and strategic planning of network infrastructure. Zhao *et al.* [23] used mobile network traffic obtained from nearly 11,500 users over a 24-hour period at a large university to build a customizable framework, named HIPHARNESS, that can generate smaller models without sacrificing model accuracy. Wu *et al.* [4] investigated IC user usage patterns by analyzing large-scale cellular data sets. Based on the characteristics of internet card (IC) user portraits and usage patterns, they proposed and implemented a deep learning prediction model IC user churn prediction (ICCP) to predict user churn rate.

There are also some studies that use web logs as datasets. Prakash *et al.* [24] analyzed web page prediction methods from web logs containing fields such as IP address, date and time, status code, and number of transmitted bits, and used classification and clustering techniques to predict user navigation patterns. Vassio and Mellia [12] collected 3 years of web data from about 25,000 households to understand how people interact with the network, capture its characteristics and changes, model people's intrinsic habits and interactions, and analyze how users and web services change over time. Ali *et al.* [25] use clickstream, transaction data, and user profile data to represent various sources of web usage data for understanding the user's life within a session on a website. Hong [26] collected comprehensive user behavior logs and geographical dimension information and built a mobile user behavior pattern analysis model based on the analysis of mobile user big data characteristics and Hadoop systems. Luckose *et al.* [27] grouped web server log files and used context-aware cohesive Markov models and Apriori clustering to propose a network recommendation system, which successfully improved web page prediction accuracy and prediction performance.

Through a review of the above studies, it can be found that implicit datasets such as logs are widely used in research on internet behavioral model building. On the other hand, explicit datasets are also used for the construction of internet behavioral models, although their acquisition cycle is relatively long. Hua [28] took female students from higher vocational colleges in Jiangsu Province as the research object, and used surveys and principal component analysis methods to find out the factors that affect their financial behavior, in order to achieve the goal of proposing strategies to help them manage their finances rationally. Qiu *et al.* [29] collected 395 questionnaires, used SPSS 23.0 and AMOS 24.0 software to analyze the data, and constructed an unsustainable usage behavior model of knowledge paying users. Chemnad *et al.* [30] conducted a problematic internet use questionnaire among 334 mobile phone users with the aim of identifying typical interactions between the use of social media applications on smartphones and problematic internet use (PIU). Kovacevic and Kascelan [5] using face-to-face interview data from 20 internet users and questionnaire data completed by 1,147 respondents, found out the differences in internet usage patterns between different genders, especially related to usage intensity and types of online activities.

It can be seen from the construction of internet behavior models and user behavior analysis, the implicit datasets became the main data source for many researchers, because it can objectively display users' internet behavior. Explicit datasets such as interviews and questionnaires are mainly used to conduct research on gender differences and problematic usage behaviors. Mixed data requires the use of some aggregation techniques to integrate the two types of data into one dataset for more complex data modeling and data analysis.

4.2.2. Data analysis stage

The data analysis phase groups different tasks, allowing the construction of models from the previously created datasets. Different data analysis methods are used when creating the model, depending on the content of the data analysis and the role of the data analysis results in the model. The most relevant methods are shown in Table 4. In Lopez-Fernandez's research, all statistical analyses were performed using IBM SPSS software and a significance level of $p < 0.05$ [31]. Abubakar and Al-Zyoud [32] analyzed the data using partial least squares structural equation modeling (PLS-SEM) and found that problematic internet use reduced safety behaviors, while temporal autonomy enhanced safety behaviors, and Novela *et al.* [6] used the same data analysis technology. Some researchers do not indicate which analysis methods were used in their research. Oztoprak's research used users' internet usage history to classify users but did not explain the specific classification method [33]. Other researchers used multiple data analysis methods in their research, depending on the complexity of the dataset and whether the behavioral model was constructed as a theoretical innovation model or to improve the existing model. In this phase the elements required to create the model are identified.

4.2.3. Create model stage

At this stage, computer systems or services are typically used to build behavioral models that are used to achieve the research goals of examining, customizing, predicting, or performing any specific task. Gan [34] established a precise marketing model according to user behavior characteristics and attributes, thereby improving the marketing effect of the enterprise. Paramaeswari and Sarno [35] conducted a study to find out what factors could influence the user interest in using e-commerce. Shaikh *et al.* [36] contributed to the research on cyberbullying behavior by offering a conceptual validated model that predicts Malaysian university students' cyberbullying behavior. As in the data collection stage, user behavior can be tracked through explicit or implicit methods. In this way, when including surveys and forms, researchers need to set in the monitoring points for building the models. So, in the questionnaire of Velki *et al.* [37] research, the first part consisted of 4 items measuring computer users' potentially risky behavior and the second part consisted of 6 questions measuring the level of users' information security awareness.

4.2.4. Validate model stage

This is the final stage of the internet behavioral model creation process, but some studies do not go through this stage. In this stage, researchers compare the models they created with models previously created by other researchers, or with models they created using other technologies, to illustrate advantage on the prediction accuracy, time efficiency, and Compatibility. Bünder and Kuchen [38] reported on a controlled AB/BA crossover experiment comparing the efficiency of our approach to the industry proven JBehave tooling and took 63% less time to specify automated tests. Ding *et al.* [10] proven that their usage pattern models (UPM) cover 94% of user behaviors and achieve up to 20% improvement in accuracy of energy representation. In addition, there are also some studies that replace the original model with the created model in real-life usage environments (such as advertising push and video push) and compare the results before and after to verify the model's contribution to specific fields. Erdem *et al.* [39] analyzed the similarity between the actual browsing behaviors and the newly generated browsing behaviors to investigate the quality of newly generated test scripts.

Q3: What techniques are used for the construction of internet behavior model?

There are various technologies employed in constructing internet behavior models. Table 4 displays the most frequently used technologies in the four stages, sorted by their frequency of use. The table also outlines the advantages and disadvantages of these technologies.

The techniques used in the dataset creation stage have been described in detail in Q2 and will not be elaborated here. The data analysis stage uses the most technology, because this stage involves a variety of different datasets, so, the most effective method is needed to carry out data analysis. Yin *et al.* [40] extracted application identifiers from the dataset, divided the applications into 16 categories, and extracted

128-dimensional feature vectors, and then clustered them by bisecting the K-means method, and finally obtained 7 different MAUP provides effective feature vectors for the next step of model construction. Kovacevic and Kascelan [5] used the interview records and questionnaire results of internet users to generate a decision tree (DT) using the CART method, and then conducted K-means cluster analysis to investigate the differences between different genders in internet usage patterns. Bahra and Pierre [9] used the Douglas-Peucker (DP) algorithm to eliminate unnecessary raw data. This operation reduced the data noise level and significantly reduced the time complexity, resulting in a significant improvement in prediction accuracy and in a scenario closer to the real world and more practical. Bhati *et al.* [41] collected 449 samples of online buyers and conducted confirmatory factor analysis using SPSS 24.0 and AMOS 24.0 software to test reliability and verify the construction of the measurement model.

Compared with the data analysis stage, the methods of creating model stage mainly focus on machine learning and deep learning. The use of machine learning to create behavioral models mainly uses genetic algorithms (GAs), support vector machines (SVMs), decision trees (DTs), genetic algorithms, Markov models and other algorithms, while deep learning mainly uses neural network algorithms. In machine learning, Putri *et al.* [42] trained and combined two SVMs through stacked generalization to produce a driver behavioral model. Abououf *et al.* [43] uses machine learning to predict the probability of the workers performing a given task, based on their learned behavioral models. Aung and Thein [44] used a bagging decision tree classifier with hyperparameter optimization to achieve a prediction accuracy of 98.97% for user behavior, significantly improving the network management method. Rajapaksha and Asanka [45] found that the random forest classification model had better results than other models. It has an accuracy of about 98% and a precision of 0.98, and they believe that the model can correctly classify e-commerce event log data as purchased or not purchased. With the advance of neural networks, many neural-based models have been designed to predict a single users' behavior, i.e., social link behavior or consumption behavior [46]. Han *et al.* [47] created two embeddings: i) action and ii) interaction features and used them to build a neural network to learn a low-dimensional vector representation of behavior over time that can identify when good learning occurs, or greatest success is achieved. In addition to machine learning and deep learning, Hong [26] designed a user behavior analysis system based on the Hadoop distributed platform to study mobile user behavior mining models based on big data.

In validate model stage, previous researchers usually used comparative verification between different models, verified the authenticity with historical data with real results, or used different datasets to verify the model. To verify the effectiveness of the model, Wang *et al.* [3] conducted a series of experiments on six web applications according to three types of user behavior traces. Zheng *et al.* [48] applied their analysis to a real-world usage trace from 13 mobile users of a U.S. ISP for one billing cycle. The most used method is comparative verification. Yu *et al.* [49] verified the superiority of the model that predicts which applications are most popular by comparing it to existing state-of-the-art methods. This method can intuitively see the advantages of the constructed model in terms of accuracy, time complexity, and number of feature factors.

Q4: What characteristics are used to define internet user behavior?

When constructing internet behavior models, the characteristic factors used are quite different due to their different research objectives. However, according to the different objects that researchers focus on, it can be divided into the extraction of user features and the extraction of behavioral features. Lin *et al.* [50] defined students' life through attributes such as time of the record, location ID (L.ID), location coordinates (Latitude, Longitude), location coordinates provider (L.P.), set of Wi-Fi APs scanned, phone lock status (Lock), the app in foreground (A.ID). Rapp and Boldi [51] explored the users' characteristics such as gender, age, profession, education level, target behavior(s), instrument(s) used, period of use, and defined a behavior change model to identify obstacles that may hinder successful implementation. Tranos and Stich [52] use internet usage, gender, income, age, qualification, population density, house price, LIC, distance to urban center, distance to London, distance to exchange, yearly trend, year features to build a logistic regression model to test whether web content of local interest attracts individuals to the internet. Chemnad *et al.* [30] use the features such as age, gender, country, app name, start use time, end use time, to identify the relationship between the use of social media applications on smartphones and problematic internet usage (PIU).

In addition to these user characteristics, more models adopt behavioral characteristics as considerations. Yuan *et al.* [15] used the characteristic factors of time, APP, and number of visits to study the user behavior characteristics of video on demand programs. Prakash *et al.* [24] predict user navigation patterns through features such as homepage, city, mobile operating system, mobile phone model, access status, and payment method. Ding *et al.* [53] investigated the pattern of users' fine-grained document reading behavior by analyzing the characteristics of exposure, viewport, skimming behavior, reading behavior, attractiveness, and perceived satisfaction. Maliki *et al.* [54] analyze the IP address, timestamp, domain, category ID features in browsing history data and develop users' security behavior profiles. However, not only the features listed can produce significant results when building models, as some models, despite supplementing them with other features, can produce the same results as without them.

Table 4. The most relevant approach to creating models of internet behavior

Stage	Method	Specific methods or algorithms	Advantage	Disadvantage
Create a dataset	Explicit	Questionnaires, interviews, business records, semi-structured interviews, transaction records, experimental records	High authenticity, allowing targeted selection of research objects to collect data	The data collection cycle is long, and the data results are greatly affected by the subject group's subjective opinions
	Implicit	Traffic logs, session logs, operating system logs, web logs, event monitoring logs, crawler data	The amount of data is large, the data collection speed is fast, and the data is relatively objective	Unable to specify specific collection objects, there is a lot of noise data
	Mixed	Mixes explicit and implicit data	Get more information	The data set is complex and requires specialized tools to integrate the data
Data analysis	Cluster analysis	K-means, Apriori, Python, Box and whisker plot method, K-Prototype clustering	Data clustering algorithms are more mature and have a lot of software support	Clustering quality is difficult to assess
	Machine learning	Decision trees, support vector machines, Bayesian classifiers, neural networks	Fast computing speed and many algorithms	Relies on large sample sizes and cannot identify multiple options
	Noise data removal	binning, return	Reduce erroneous data and reduce the number of influencing factors	leading to deviations between results and actual
	Structural equation	SPSS, AMOS, SEM	Suitable for discontinuous data and long-term data	Larger sample size required
	Correlation Analysis	Apriori	The algorithm is fast and can handle incomplete data	High computational complexity
Create model	Machine learning	GA, SVM, DT, genetic algorithm, Markov model, cohesive Markov model	Improve precision, improve accuracy, shorten time,	The model needs to include more influencing factors
	Deep learning	Neural networks	Improve prediction success rate, wide model coverage, and lower data set requirements	Factors such as time information and user's fine-grained location are not considered
	Big Data	Hadoop, Map reduce	Can form real-time data model	The data set needs to be large enough, and it is difficult to extract effective factors and requires a lot of calculations
	Technology acceptance model (TAM)	Theory of planned behavior (TPB), the theory of reasoned action (TRA)	Distinguish the behavioral outcomes of different groups	There are few choices of data sets from different social platforms
Validate model	Comparison verification	Comparison of model results with other benchmark models	Can be verified using simulated data sets, and the model effect is obvious	Different models have different behavioral characteristics
	Real data verification	feature ranking-selection methods, case studies, historical data comparison	Model results tend to be more realistic	Only historical data can be compared
	Validation on different data sets	Run the same model on different test data sets and observe the difference in results	Visualize different outcomes for different groups	Differences in influencing factors in different data sets lead to model mismatch

4.3. Discussion

Leveraging internet behavioral models to enhance internet QoS represents a multifaceted and evolving approach aimed at optimizing user experience and network performance. Through an extensive review of previous studies, the goal of creating an internet behavior model is constructed around eight key aspects: improving the existing model, exploring behavioral correlation, personalized information push, enhancing internet service quality, theoretical model innovation, user behavior prediction, user satisfaction improvement, and user preference analysis. Although a great deal of research has focused on the development of internet behavioral models in recent years, there is still little research devoted to improving internet QoS. Out of 35 research documents describing behavioral models or behavioral analysis of the internet, only two explicitly mention improvements in internet QoS. Yuan *et al.* [15] developed a visual model based on user and program video data analysis to alleviate network congestion caused by large data traffic and enhance online experience. Soo *et al.* [16] strategically balanced WLAN performance by redistributing user load among access points (APs), incentivizing users to temporarily divert data traffic to effectively manage congestion issues.

Building a model of internet behavior involves several key stages, starting with the creation of datasets that contain explicit, implicit, or hybrid methods of data acquisition. Implicit datasets such as logs are widely used for their ability to objectively show user behavior, while explicit methods such as surveys and interviews provide insight into gender differences and problematic usage behaviors. Of the four articles using a hybrid dataset, two used both traffic log data and questionnaire data, one analyzing the relationship between individual internet behavior and local internet content of interest, and the other investigating the impact of different social media applications on problematic internet use. After acquiring the dataset, the subsequent data analysis phase uses various techniques such as statistical analysis, machine learning, and deep learning to extract meaningful insights from the dataset.

In the creation phase, researchers utilize different techniques to build behavioral models tailored to specific research goals. The choice of technique often depends on the particular objectives and data characteristics involved in the study. Commonly employed modelling techniques encompass regression analysis, time series modelling, clustering analysis, and collaborative filtering. Regression analysis is frequently used to predict user behavior based on historical data, while time series modelling helps understand and forecast trends over time. Clustering analysis groups users with similar behaviors, enabling more targeted service improvements. Collaborative filtering, often used in recommendation systems, predicts user preferences by leveraging similar users' preferences. These diverse techniques collectively enhance the robustness and applicability of the behavioral models in addressing various aspects of internet service quality.

Regression analysis is a statistical technique utilized to forecast relationships between variables. By scrutinizing the user's historical behavioral data and other pertinent factors, a regression model can be constructed to anticipate the user's future behavior. For instance, an e-commerce website can predict a user's forthcoming purchase intention and amount by analyzing their purchase history along with factors like age, gender, and geographic location.

Time series modeling is a statistical approach employed to analyze time-dependent data in order to comprehend trends in user behavior over time. For example, a news website can employ a time series model to examine the correlation between users' visits and news events for providing tailored content recommendations during significant news occurrences.

Cluster analysis is a technique used for categorizing users into distinct groups based on their behavior data and other characteristics that exhibit similar patterns of interests. This aids in personalized recommendations and targeted marketing strategies. As an illustration, a travel website can utilize cluster analysis to group users into different categories of travel enthusiasts while offering relevant travel products and services based on each group's specific interests.

Collaborative filtering is an approach that recommends products or content by examining both the historical behaviors of individual users as well as those of other similar users who share comparable behavioral patterns. This method operates under the assumption that if two users display analogous behaviors, they are likely interested in similar products or content. By assessing similarities among users, a recommendation model can be developed for delivering personalized suggestions. The collaborative filtering technology can be employed by a video streaming platform to provide personalized recommendations of movies or TV shows that are highly likely to match users' preferences. At the same time, validation of these models is critical and often involves comparative validation against existing models or real-world use environments to assess accuracy, efficiency, and compatibility.

The characteristics that define internet user behavior cover a variety of factors, from user demographics (e.g., age, gender, income) to behavioral attributes (e.g., time spent, application preferences, browsing patterns). By analyzing these characteristics, researchers can gain a deeper understanding of user behavior and preferences, allowing them to develop more targeted and effective models of internet behavior. Notably, research shows that the success of internet behavior models in enhancing internet QoS or analyzing user behavior depends on the inclusion of characteristic attributes such as: i) age, ii) gender, iii) APP category, iv) access time, v) access location, vi) device model, vii) number of visits, and viii) access status. Therefore, the optimal adaptability of these attributes should be considered in the internet behavior model aimed at improving the quality of internet service.

In summary, the integration of internet behavior models into QoS improvement initiatives holds significant promise for optimizing network resources, enhancing user experience, and mitigating performance challenges. By leveraging insights gleaned from user behavior analysis and employing advanced modeling techniques, researchers can pave the way for a more responsive, adaptable, and user-centric digital ecosystem characterized by improved internet service quality.

5. CONCLUSION

Internet user profiling and behavior modeling technology has been applied in many fields. For example, in the field of e-commerce, by analyzing users' purchasing behaviors, users' needs can be better understood, so as to provide personalized recommendation services. In the field of finance, by analyzing users' trading behaviors, users' risk preferences can be judged, so as to provide users with better investment services. In the field of medicine, by analyzing users' medical data, doctors can better understand patients' conditions, so as to provide patients with better treatment plans. In the field of social media, by analyzing users' social behaviors, users' preferences and interests can be better understood, so as to provide users with more accurate social services.

Through a review of previous literature, it is evident that the primary aim of internet behavior models is to utilize users' past behavioral data to elucidate patterns, subsequently employing these patterns to forecast users' forthcoming actions. This suggests that previous studies have predominantly focused on demonstrating the influence of fixed user group behaviors on the group itself, with only a few models showcasing the impact of fixed user group behaviors on other entities. The study of internet behavior entails several pivotal steps, encompassing dataset creation, meticulous data analysis, model construction, and subsequent model validation. Most prior investigations have relied on singular implicit datasets. Even studies utilizing mixed datasets have typically examined each dataset independently. Given that user internet behavior spans numerous domains and yields diverse datasets, it is regrettable that, thus far, there exists no effective methodology for investigating user behavior across multiple datasets. In preceding research on internet behavior model construction, methodologies have been relatively simplistic, predominantly focusing on machine learning, cluster analysis, and deep learning. Consequently, there remains a necessity for subsequent researchers to explore more efficient research methodologies.

In summary, future research on internet behavior needs to explore how the broader impact of user behavior on external entities can yield valuable insights into network dynamics and user interactions. Developing powerful frameworks capable of integrating diverse datasets from diverse domains enables a more complete understanding of user behavior across platforms and contexts. Adopting an interdisciplinary approach that combines insights from fields such as social sciences, network theory, and cognitive psychology could facilitate the development of more nuanced and comprehensive models.

ACKNOWLEDGEMENTS

This work was supported by the Natural Science Foundation of Anhui Province, China (Grants No.2022AH052795). In addition, the authors would like to thank Anhui Technology and Vocational College for supporting the development of this study.

REFERENCES

[1] K. Xu, "Behavior modeling of network traffic," in *Network Behavior Analysis*, Singapore: Springer Singapore, 2022, pp. 21–37.
[2] D. H. L. Oliveira, F. M. V. Filho, T. P. de Araujo, J. Celestino, and R. L. Gomes, "Adaptive model for network resources prediction in modern internet service providers," in *2020 IEEE Symposium on Computers and Communications (ISCC)*, Jul. 2020, pp. 1–6, doi: 10.1109/ISCC50000.2020.9219550.
[3] W. Wang, J. Guo, Z. Li, and R. Zhao, "Behavior model construction for client side of modern web applications," *Tsinghua Science and Technology*, vol. 26, no. 1, pp. 112–134, Feb. 2021, doi: 10.26599/TST.2019.9010043.
[4] F. Wu *et al.*, "Characterizing internet card user portraits for efficient churn prediction model design," *IEEE Transactions on Mobile Computing*, vol. 23, no. 2, pp. 1735–1752, Feb. 2024, doi: 10.1109/TMC.2023.3241206.
[5] D. Kovacevic and L. Kascelan, "Internet usage patterns and gender differences: a deep learning approach," *IEEE Consumer Electronics Magazine*, vol. 9, no. 6, pp. 105–114, Nov. 2020, doi: 10.1109/MCE.2020.2986817.
[6] S. Novela, Y. O. Sihombing, Novita, E. Caroline, and R. Octavia, "The effects of hedonic and utilitarian motivation toward online purchase intention with attitude as intervening variable," in *2020 International Conference on Information Management and Technology (ICIMTech)*, Aug. 2020, pp. 75–80, doi: 10.1109/ICIMTech50083.2020.9211197.
[7] W. Zhang, J. Qin, W. Guo, R. Tang, and X. He, "Deep learning for click-through rate estimation," *arXiv preprint arXiv:2104.10584*, 2021.
[8] Z. He *et al.*, "A survey on user behavior modeling in recommender systems," *arXiv preprint arXiv:2302.11087*, 2023.
[9] N. Bahra and S. Pierre, "A bidirectional trajectory prediction model for users in mobile networks," *IEEE Access*, vol. 10, pp. 1921–1935, 2022, doi: 10.1109/ACCESS.2021.3139867.
[10] M. Ding, T. Wang, and X. Wang, "Establishing smartphone user behavior model based on energy consumption data," *ACM Transactions on Knowledge Discovery from Data*, vol. 16, no. 2, pp. 1–40, Apr. 2022, doi: 10.1145/3461459.
[11] X. Zhou, Y. Mei, J. Yan, and L. Liang, "A quantitative modeling method of occupant behavior of office building based on human dynamics," in *2020 IEEE 4th Conference on Energy Internet and Energy System Integration (EI2)*, Oct. 2020, pp. 1325–1330, doi: 10.1109/EI250167.2020.9346817.
[12] L. Vassio and M. Mellia, "Data analysis and modelling of users' behavior on the web," *2019 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*, pp. 665–670, 2019.
[13] A. Souri, A. S. Mohammed, M. Yousif Potrus, M. H. Malik, F. Safara, and M. Hosseinzadeh, "Formal verification of a hybrid machine learning-based fault prediction model in internet of things applications," *IEEE Access*, vol. 8, pp. 23863–23874, 2020, doi: 10.1109/ACCESS.2020.2967629.

[14] A. Keimer *et al.*, “Information patterns in the modeling and design of mobility management services,” *Proceedings of the IEEE*, vol. 106, no. 4, pp. 554–576, Apr. 2018, doi: 10.1109/JPROC.2018.2800001.

[15] Y. Yuan, X. Wang, and G. Bin, “Analysis of user behavior in a large-scale internet video-on-demand (VoD) system,” in *Proceedings of the 5th International Conference on Multimedia and Image Processing*, Jan. 2020, pp. 153–158, doi: 10.1145/3381271.3381288.

[16] W. K. Soo, T.-C. Ling, A. H. Maw, and S. T. Win, “Survey on load-balancing methods in 802.11 infrastructure mode wireless networks for improving quality of service,” *ACM Computing Surveys*, vol. 51, no. 2, pp. 1–21, Mar. 2019, doi: 10.1145/3172868.

[17] S. Sen, C. Joe-Wong, S. Ha, and M. Chiang, “Time-dependent pricing for multimedia data traffic: analysis, systems, and trials,” *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 7, pp. 1504–1517, Jul. 2019, doi: 10.1109/JSAC.2019.2916490.

[18] H. Kwon, S. Lee, and D. Jeong, “User profiling via application usage pattern on digital devices for digital forensics,” *Expert Systems with Applications*, vol. 168, Apr. 2021, doi: 10.1016/j.eswa.2020.114488.

[19] K. Zhu, Y. Xiao, W. Zheng, X. Jiao, and C.-H. Hsu, “A novel context-aware mobile application recommendation approach based on users behavior trajectories,” *IEEE Access*, vol. 9, pp. 1362–1375, 2021, doi: 10.1109/ACCESS.2020.3046654.

[20] C. Li, K. Yu, and X. Wu, “Co-clustering analysis of mobile users’ usage behavior on apps,” in *Proceedings of the 2nd International Conference on Telecommunications and Communication Engineering*, Nov. 2018, pp. 214–219, doi: 10.1145/3291842.3291856.

[21] X. Liu and K. Wu, “The research of user-based media product innovation in the mobile environment,” in *2018 International Joint Conference on Information, Media and Engineering (ICIME)*, Dec. 2018, pp. 87–91, doi: 10.1109/ICIME.2018.00027.

[22] O. J. Adeyemi, S. I. Popoola, A. A. Atayero, D. G. Afolayan, M. Ariyo, and E. Adetiba, “Exploration of daily Internet data traffic generated in a smart university campus,” *Data in Brief*, vol. 20, pp. 30–52, Oct. 2018, doi: 10.1016/j.dib.2018.07.039.

[23] Z. Zhao, Y. Chen, and S. Li, “Internet use behavior of the elderly: family support, psychological needs, and optimistic personality perspective,” in *2021 5th International Conference on Education and E-Learning*, Nov. 2021, pp. 276–281, doi: 10.1145/3502434.3502454.

[24] O. P. P. G., A. S, S. M, and G. R, “Analyzing the user navigation pattern from web logs using maximum frequent pattern approach,” in *2021 6th International Conference on Inventive Computation Technologies (ICICT)*, Jan. 2021, pp. 877–883, doi: 10.1109/ICICT50816.2021.9358751.

[25] N. M. Ali, A. M. Gadallah, H. A. Hefny, and B. Novikov, “An integrated framework for web data preprocessing towards modeling user behavior,” in *2020 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon)*, Oct. 2020, pp. 1–8, doi: 10.1109/FarEastCon50210.2020.9271467.

[26] R. Hong, “Research on mobile user behavior mining model based on big data,” in *2019 4th International Conference on Mechanical, Control and Computer Engineering (ICMCCE)*, Oct. 2019, pp. 461–4613, doi: 10.1109/ICMCCE48743.2019.00110.

[27] V. Luckose, J. Chembath, J. A. R. Ponnusamy, S. Sharma, P. Kaur, and S. Smiley, “Web usage pattern detection using cohesive markov model with apriori algorithm,” in *2022 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS)*, Jun. 2022, pp. 226–229, doi: 10.1109/I2CACIS54679.2022.9815465.

[28] H. Hua, “Research on the influential factors of female college students’ internet financial behavior in higher vocational colleges in China,” in *Proceedings of the 2023 14th International Conference on E-Education, E-Business, E-Management and E-Learning*, Feb. 2023, pp. 216–221, doi: 10.1145/3588243.3588265.

[29] J. Qiu, K. Wei, and R. Yao, “Model construction and empirical study on unsustainable usage behavior of paid knowledge users,” *IEEE Access*, vol. 11, pp. 46956–46965, 2023, doi: 10.1109/ACCESS.2023.3274866.

[30] K. Chemnad, M. Aziz, S. B. Belhaouari, and R. Ali, “The interplay between social media use and problematic internet usage: Four behavioral patterns,” *Heliyon*, vol. 9, no. 5, May 2023, doi: 10.1016/j.heliyon.2023.e15745.

[31] O. Lopez-Fernandez, “Generalised versus specific internet use-related addiction problems: a mixed methods study on internet, gaming, and social networking behaviours,” *International Journal of Environmental Research and Public Health*, vol. 15, no. 12, Dec. 2018, doi: 10.3390/ijerph15122913.

[32] A. M. Abubakar and M. F. Al-zyoud, “Problematic internet usage and safety behavior: does time autonomy matter?” *Telematics and Informatics*, vol. 56, Jan. 2021, doi: 10.1016/j.tele.2020.101501.

[33] K. Oztoprak, “Mobile subscriber profiling and personal service generation using location awareness,” *Advances in Electrical and Computer Engineering*, vol. 18, no. 3, pp. 105–112, 2018, doi: 10.4316/AECE.2018.03014.

[34] W. Gan, “Design of network precision marketing based on big data analysis technology,” in *2020 International Conference on E-Commerce and Internet Technology (ECIT)*, Apr. 2020, pp. 77–81, doi: 10.1109/ECIT50008.2020.00026.

[35] R. P. I. Paramaeswari and R. Sarno, “Analysis of e-commerce (Bukalapak, Shopee, and Tokopedia) acceptance models using TAM2 method,” in *2020 International Seminar on Application for Technology of Information and Communication (iSemantic)*, Sep. 2020, pp. 505–510, doi: 10.1109/iSemantic50169.2020.9234271.

[36] F. B. Shaikh, M. Rehman, A. Amin, A. Shamim, and M. A. Hashmani, “Cyberbullying behaviour: a study of undergraduate university students,” *IEEE Access*, vol. 9, pp. 92715–92734, 2021, doi: 10.1109/ACCESS.2021.3086679.

[37] T. Velki, A. Mayer, and J. Norget, “Development of a new international behavioral-cognitive internet security questionnaire: preliminary results from Croatian and German samples,” in *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, May 2019, pp. 1209–1212, doi: 10.23919/MIPRO.2019.8756765.

[38] H. Bünder and H. Kuchen, “A model-driven approach for behavior-driven GUI testing,” in *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing*, Apr. 2019, pp. 1742–1751, doi: 10.1145/3297280.3297450.

[39] I. Erdem, R. F. Oguz, E. Olmezogullari, and M. S. Aktas, “Test script generation based on hidden Markov models learning from user browsing behaviors,” in *2021 IEEE International Conference on Big Data (Big Data)*, Dec. 2021, pp. 2998–3005, doi: 10.1109/BigData52589.2021.9671312.

[40] J. Yin *et al.*, “Mobile app usage patterns aware smart data pricing,” *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 4, pp. 645–654, Apr. 2020, doi: 10.1109/JSAC.2020.2971896.

[41] N. S. Bhati, L. Vijayvargy, and A. Pandey, “Role of e-service quality (E-SQ) on customers’ online buying intention: an extended theory of planned behavior,” *IEEE Access*, vol. 10, pp. 77337–77350, 2022, doi: 10.1109/ACCESS.2022.3190637.

[42] R. R. Mardi Putri, C.-H. Yang, C.-C. Chang, and D. Liang, “Smartwatch-based open-set driver identification by using GMM-based behavior modeling approach,” *IEEE Sensors Journal*, vol. 21, no. 4, pp. 4918–4926, Feb. 2021, doi: 10.1109/JSEN.2020.3030810.

[43] M. Abououf, S. Singh, H. Otrok, R. Mizouni, and E. Damiani, “Machine learning in mobile crowd sourcing: a behavior-based recruitment model,” *ACM Transactions on Internet Technology*, vol. 22, no. 1, pp. 1–28, Feb. 2022, doi: 10.1145/3451163.

[44] S. T. Aung and T. Thein, “Internet traffic categories demand prediction to support dynamic QoS,” in *2020 5th International Conference on Computer and Communication Systems (ICCCS)*, May 2020, pp. 650–654, doi: 10.1109/ICCCS49078.2020.9118431.

[45] P. Rajapaksha and D. Asanka, "Recommendations to increase the customer interaction of e-commerce applications with web usage mining," in *2023 IEEE International Conference on Contemporary Computing and Communications (InC4)*, Apr. 2023, pp. 1–6, doi: 10.1109/InC457730.2023.10263131.

[46] J. Li, L. Wu, R. Hong, K. Zhang, Y. Ge, and Y. Li, "A joint neural model for user behavior prediction on social networking platforms," *ACM Transactions on Intelligent Systems and Technology*, vol. 11, no. 6, pp. 1–25, Dec. 2020, doi: 10.1145/3406540.

[47] L. Han, A. Checco, D. Difallah, G. Demartini, and S. Sadiq, "Modelling user behavior dynamics with embeddings," in *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, Oct. 2020, pp. 445–454, doi: 10.1145/3340531.3411985.

[48] L. Zheng, C. Joe-Wong, M. Andrews, and M. Chiang, "Optimizing data plans: usage dynamics in mobile data networks," in *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, Apr. 2018, pp. 2474–2482, doi: 10.1109/INFOCOM.2018.8486351.

[49] D. Yu, Y. Li, F. Xu, P. Zhang, and V. Kostakos, "Smartphone app usage prediction using points of interest," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 4, pp. 1–21, Jan. 2018, doi: 10.1145/3161413.

[50] M. Lin, V. W. Zheng, and S. Xiang, "Sequential context modeling for smart devices by collaborative hidden Markov model," in *2018 IEEE 4th World Forum on Internet of Things (WF-IoT)*, Feb. 2018, pp. 771–777, doi: 10.1109/WF-IoT.2018.8355155.

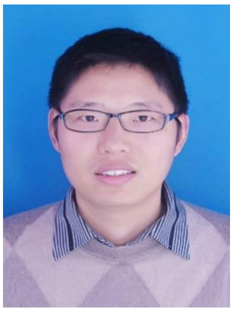
[51] A. Rapp and A. Boldi, "Exploring the lived experience of behavior change technologies: towards an existential model of behavior change for HCI," *ACM Transactions on Computer-Human Interaction*, vol. 30, no. 6, pp. 1–50, Dec. 2023, doi: 10.1145/3603497.

[52] E. Tranos and C. Stich, "Individual internet usage and the availability of online content of local interest: a multilevel approach," *Computers, Environment and Urban Systems*, vol. 79, Jan. 2020, doi: 10.1016/j.compenvurbsys.2019.101371.

[53] Z. Wu, J. Mao, K. Xu, D. Song, and H. Huang, "A passage-level reading behavior model for mobile search," in *Proceedings of the ACM Web Conference 2023*, Apr. 2023, pp. 3236–3246, doi: 10.1145/3543507.3583343.

[54] N. A. Maliki, A. Zainal, F. A. Abdoh Ghaleb, and M. N. Kassim, "User security behavioral profiling using historical browsing website," in *2021 International Conference on Data Science and Its Applications (ICoDSA)*, Oct. 2021, pp. 163–168, doi: 10.1109/ICoDSA53588.2021.9617493.

BIOGRAPHIES OF AUTHORS



Zhang Lei received the B.Eng. degree in computer science and technology from Anhui Normal University, Anhui Province, China in 2003, and the Master's degree in Computer Science from Hefei University of Technology, Anhui Province, China in 2008. He is currently pursuing his PhD at Universiti Teknologi Mara in Malaysia. He is an associate professor at Anhui Business and Technology College in Anhui, China. His current research interests include data mining, interconnection behavior analysis, and local area network operations and management. He can be contacted at email: zhangleiahbvc@gmail.com.

Nor Shahniza Kamal Bashah holds a Ph.D degree in telematics from Norwegian University of Science and Technology, Norway. She is an associate professor at the College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, Malaysia. Her research interest is involved in the field of mobile and wireless communication and semantic web. She can be contacted at shahniza@uitm.edu.my.