IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 27, 2019, accepted September 15, 2019, date of publication September 27, 2019, date of current version October 17, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2944243

A Survey of User Profiling: State-of-the-Art, Challenges, and Solutions

CHRISTOPHER IFEANYI EKE[1,2], AZAH ANIR NORMAN[1], LIYANA SHUIB[1], AND HENRY FRIDAY NWEKE[1,3]

[1]Department of Information System, Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur 50603, Malaysia
[2]Department of Computer Science, Faculty of Science, Federal University, Lafia P.M.B 146, Nigeria
[3]Computer Science Department, Ebonyi State University, Abakaliki P.M.B 053, Nigeria

Corresponding author: Azah Anir Norman (azahnorman@um.edu.my)

This research was supported by the University of Malaya Research Grant (UMRG-RP044C-17HNE) and Grant "Bantuan Kecil" BKS080-2017.

ABSTRACT Advancements in information and communication technology, and online web users have given attention to the virtual representation of each user, which is crucial for effective service personalization. Meeting users need and preferences is an ongoing challenge in service personalization. This issue can be addressed through the building of a comprehensive user profile. A user profile is the summary of the user's interests, characteristics, behaviours, and preferences, while user profiling is the system of collecting, organizing and inferring the user profile information. Many reviews on user profiling have been conducted but none focused on the effective profile modeling process. Hence, this article aims to provide a review of the recent state-of-the-art approach to user profiling. These include methods, description, characteristics, and taxonomy of the user profile. The study of the existing user profiling modeling in the aspect of data acquisition, feature extraction, profiling techniques, and profiling approaches (with the identification of their strengths and weaknesses) and the performance measures are also provided. In addition, the research challenges were also discussed with a focus on privacy, datasets, cold start issues, trust issues, and computational complexity. Moreover, the article identified an open research direction that serves as solutions to the identified challenges and motivation for further researchers in advancing user profiling. The findings showed that an effective modeling process enhances the construction of accurate user profile for service personalization.

INDEX TERMS User profiling, user interest, profiling modeling, personalized service, service recommendation.

I. INTRODUCTION

Advancements in information and communication technology has brought an obvious need to have personalized information systems, whose goal is to adapt information-exchange functionality to the specific interest of their users. User personalization research is a current field of study that dispersed among various domains such as artificial intelligence, data mining, and information science [1]. One of the notable applications of the user personalization is the recommender system [2]. The existing search engine according to Alaoui *et al.* [3] is inefficient and cannot satisfy the user's needs because of the exponential number of services (services and information in digital format remain problematic and uncontrollable), the range of the user's goals (each user has a definite goal, context, and target when searching for information), and bad query formulation (the imprecision of queries leads to an inadequate descriptions of the user's information need). It was also revealed that the existing search engine system did not consider user profiling. Moreover, the current system provides access to a vast amount of distributed and heterogeneous information, giving rise to information overloading, which makes it difficult for users to distinguish relevant information from the irrelevant ones. Furthermore, the assessing of the user's query does not rely on the preference of the user who issued the query and the query contents. For instance, the same query issued by two different users produces similar results despite that they specified different preferences in their query contents. To tackle these issues, there is a need to build a system that offers users information

The associate editor coordinating the review of this manuscript and approving it for publication was Xiao Liu.

appropriate for their needs. The system that is capable of taking various users characteristics and situations in a different context, which may affect their response to the system. This enhanced system depends on the idea of user profile and personalization system. User profile represents the user model instance, useful for the interactive adaptive system whereas personalization system relies on various forms of the user profile for the construction of an efficient recommendation engine.

The user profiling system addresses challenges outlined above by constructing, handling and demonstrating modified information about each user. A good user profile plan is essential in web-based systems and search engine personalization. Moreover, a user profile is the main component of information systems such as adaptive systems. It has played an active role in various domains such as healthcare sectors, banking sectors, social media, e-commerce, security, access control and social networking [4], [5]. For instance, Behavioral-based profiling is essential in healthcare delivery services. In this case, physicians and healthcare assistants are interested in how patients carry out their daily routine to support them and intervene in their situations as the need arises. Likewise, in the smart home, people have different special needs such as their living preferences, in order to ascertain their well-being. In the healthcare domain, user profiling is vital in meeting a wide range of users' need. Moreover, adapting services to specific user requirements cannot be efficiently attained without the proper review of a user profile [6]. The advent of social media has also given rise to user profiling application in the field of advertising, recruiting, marketing and law-enforcement [7].

Numerous descriptions of User-profiles have been identified in the previous studies. Ouaftouh *et al.* [8] defined user profile as a set of information that describes a user. It consists of demographic information such as the user's name, age, country, level of education, etc., which represents user preferences or interests in either a single or group of users. El Allioui and El Beqqali [9], in their study, explained user profile as a set of data structure that describe the environment for human-computer interaction. In a web search engine, the user profile is described as the application of ontology for the systematic representation of the user's interest [10]. It enables the conceptual representation of the knowledge that constitutes user preference and context. Godoy and Amandi [11] described user profile as the narration of a user's behavior, interests, characteristics and preferences obtained through interviews and questionnaires, or dynamically with the aid of machine learning algorithms and data mining techniques. In another study, Kanoje *et al.* [2] defined user profile as the procedure for gathering information of the user's interest. The system utilises such information to tailor services and improve the user's satisfaction. Furthermore, Alaoui *et al.* [3] defined a user profile as the information that offer insight to a user's need and predicts his future intention. They noted that this information depends on three major factors, which include similarities, trace handling, and prediction through machine learning algorithms. In a recent study, Chen and Ghorbani [12] described a user profile as a user pattern that consists of user behavioral tendency and preferences. In their description, they maintained that the user profile knowledge acquired provides an idea of the user's behavior knowledge and can predict his/her intentions. Consequently, it is simple to determine users with similar behavior as long as they have the same user profile. Thus, the prediction of user behavior trends is practically feasible because of the current behavioral model.

The main benefit of building user profiling through behavioural-based approach is due to its ability to provide an efficient mechanism in solving the problem of information overloading inherent in current information systems. However, finding and managing the right information becomes difficult without an effective user profiling systems. These issues can be addressed by using more information about the user's need and objectives through information resources [13]. In addition, effective user profiling technology provides a high measure of personalization and user convenience. This can be seen in most of the current financial institutions such as banks, insurance and credit rating providers that collect a large amount of user information and uses financial metrics to determine the monetary status of the individuals [14].

Few reviews and surveys have been presented on user profiling. For instance, Stamatatos [15] surveyed an automated method for authorship attribution. In their survey, they examined the user profiles in the aspects of text classification and text representation. However, they focused more on the computational settings and requirements for profiling rather than linguistic issues. In another study, Mezghani *et al.* [16] worked on user profile survey using social annotation. The study investigated the social user characteristics and techniques for modeling and updating a tag-based user profile. Thus, the study considered only the tag-based profile modeling on social annotation. Abdel-Hafez and Xu [17] described and carried out a comparative analysis of user modeling technique for a social media site. In their study, the author described the modeling procedure for user profile construction. In another study, Peng *et al.* [18] presented a survey on user profiling that focused mainly on the intrusion detection system. In their study, they carried out a survey on the preventive measure in the context of exploiting the user behavior based on their user profile for the acceptance or denial of a legitimate user on the system. Recently, Chen and Ghorbani [12] presented a user profiling model survey on anomaly detection in cyberspace. The study concentrated more on the anomaly detection model based on user profile by examining the profile modeling, the data source and the commonly used features for profile modeling in the domain of cyber-security behavior. The aforementioned surveys on user profiling have concentrated on authorship attribution, tagged-based social annotation, and cyber-security (instruction and anomaly detection). However, none of the studies have considered state-of-the-art user

TABLE 1. List of acronyms and their meaning.

Acronym	Meaning
ACM	Association for Computing Machinery
CUO	Construction User Ontology
DUWE	Dynamic User and Word Embedding
IEEE	Institute of Electrical and Electronics Engineers
K-NN	K- Nearest Neighbours
LCRF	Linear Conditional Random Field
LIWC	Linguistic Inquiry and Word Count
ML	Machine Learning
MUO	Maintenance of user Ontology
NBC	Naive Bayesian Classifier
NIL	Not in the List
NLP	Natural Language Processing
OUPA	Ontology-Based User Profile Acquisition
SFTG	Social Tie Factor Graph
SKDM	Streaming Keyword Diversification Model
TCRF	Tree-structured Conditional Random Fields
TREC	Text Retrieval Conference
URL	Uniform Resource Locator

profile modeling in the aspect of data acquisition, feature extraction, modeling techniques, and performance metrics. This study seeks to address this gap. Therefore, the scope of this study is to carry out an extensive review of user profile, its processes, and user modeling (with the focus on data source, feature used, profiling approach, strengths, weaknesses, and the performance metrics employed in various approach). This study also provides open challenges found in user profile research and the solutions to the identified challenges.

The contributions of this paper are outlined below:

- *It provides a comprehensive review of user profile methods, profiling types (Static and Dynamic), models and profiling processes.*
- *It provides an extensive review of state-of-the-art user profiling models based on data acquisition, feature extraction, modeling techniques (by identifying their strengths and weaknesses) and performance evaluation*
- *The study identified various research challenges inherent in the current user profiling in relation with service recommendation.*
- *The study also provides an open research direction that could serve as a solution to identified challenges.*

The remainder of this paper is organized as follows. Table1 gives the acronyms used in the articles and their meanings. Section 2 describes the background of user profiling. Section 3 examines the state-of-the-art of user profiling. Section 4 discusses the research challenges in user profiling while section 5 provides the open research direction that could serve as a solution to the identified challenges. The paper concludes with a summary of the findings and offers suggestions for future research.

II. USER PROFILING

The first thing to consider when dealing with personalization is to produce an accurate user information representation (user preferences and interest), usually stored in the user profile. A better retrieval of user result depends solely on the accuracy of the user information representation. Thus, an accurate representation of the user profile is crucial to appropriately obtain better retrieval results [19]. User profiling research started getting consideration since the introduction of expert finding, the task at TREC 20015-enterprise track. [20], [21]. User profiling tasks began by Balog *et al.* [2], when they used the procreative language modeling method to model users by selecting a set of suitable keywords for user profile representation.

User profiling is a virtual representation of individual data related to a particular user in a customized desktop settings. Kanoje *et al.* [2] described user profiling as a means of determining the user's interest data that is built upon the knowledge of the user and the accurate system's retrieval of user satisfaction. While Jang *et al.* [23] noted that the user profile comprises user and service information. User service stores the user's information such as the user's name, ID, personal inclination, and hobbies. On the other hand, service information stores the service name, service provider, service context, service frequency, and value. etc. Yang [24] in her study maintained that user profile assists in the summarization of a vast amount of user information in order to attain personalized information retrieval and product recommendation goal. Generally, the main objective of user profiling is to acquire data about the user interest or subject, and the range of time in which they have shown interest, in such a way to enhance the quality of user information access and ascertain the intention of the user. User profiling performs a substantial role in any field of application such as event analysis, service recommendation and attributes inference. The taxonomy of user profiling is provided in Figure 1.

A. USER PROFILE TYPES

User profile can be represented in the form of rich semantic-based structure (occasionally improved using ontologies) [25] and a set of weighted keywords [26]. However, the weighted keyword representation is commonly used because the extraction of this profile from a document or other sources is carried out automatically [19]. User profile can be broadly grouped into two main types: static profile and dynamic profile [27]. These categories are explained in details below.

1) STATIC PROFILE

Generally, the user-profiling task is considered as a supervised learning approach. In this type of profile, the data representation relies on the static position through the creation of aggregated representations across the whole datasets [27]. According to Poo *et al.* [28], ''Static profiling approach is a process of analysing user's predictable and static characteristics''. They noted in their methodology that the information provided by user via the static profile is employed in identifying the kind of information that the user is showing interest. A static profile is a type of profile that maintains user information for a long period of time. In other words, the user

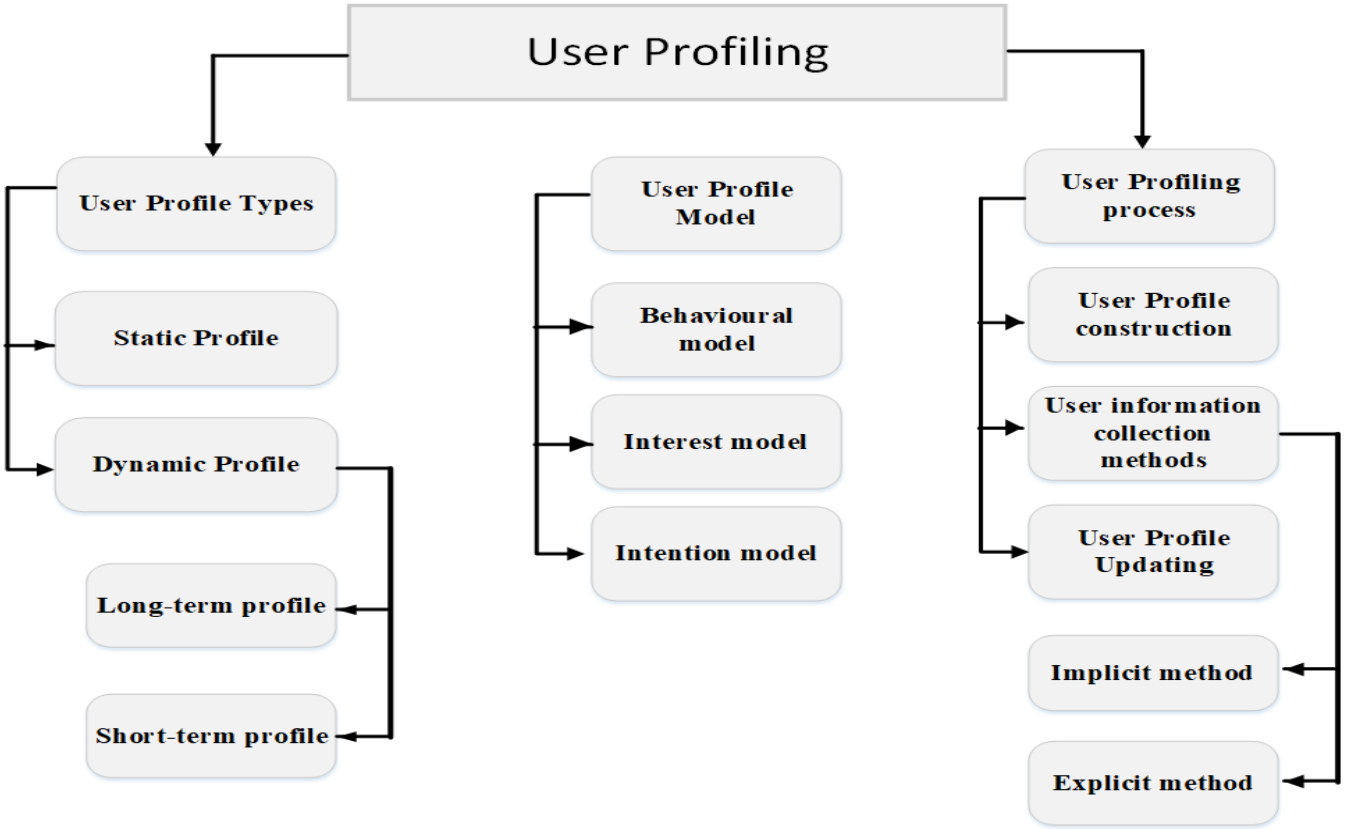

FIGURE 1. User profiling taxonomy.

information does not undergo any change or modification. For example, user's age and sex. Moreover, the information supplied by the users are employed for the creation of static profile. Therefore, the user's attribute and available contents are static in nature, that is, it remains unchanged within a period. In a recent study, Song *et al.* [29] proposed a profile learning method for predicting volunteerism for various social network users. They defined the problem as a binary classification problem. In addition, the study seeks to apply the non-negative matrix factorization approach to infer the missing data in the proposed technique. Thus, the proposed model was able to address a single task problem. However, the proposed method failed to perform multi-task learning. In a related approach, a study conducted by Farseev *et al.* [30] demonstrated the effectiveness of ensemble learning method that combines multiple sources and multi-modal data for demographic user profiling. The experimental analysis showed that the proposed model outperformed the state-of-the-art method and indicated the important of multi-source learning approach for user profile performance improvement. The issue with this form of profiling is that the users rarely provide all their information accurately as they consider their privacy so much important and as such, makes the static profile unreliable

2) DYNAMIC PROFILE

In contrast to the static profile, the dynamic profile is auto-generated by the system and consequently, the user attribute and contents undergo changes over time. In dynamic profiling, the profile information about user's behavior seeks to determine future information of the user more than the present information [2]. In other words, it is referred to as a behavioral or adaptive profile. The dynamic profile is always accurate in a situation where there is a high velocity of data delivery. In addition, the existing user ontology is employed to direct the profile extraction, define the set of relations in question and to provide the entity dictionary.

Some current attempts have been made on the analysis of user temporal behavior to learn dynamic user profile. Akbari *et al.* [31] in their study on profiling of user's wellness employed optimization technique to learn the user's wellness profile on the class of events wellness. Their proposed technique used content information of the tweets together with event category relation for obtaining users wellness event on user's timeline such as eating and exercising. The researchers specifically utilized graph Laplacian as a regularizer in the process of learning to analyse the interrelatedness that exists between different events. Similarly, Liang *et al.* [32] researched the problem of dynamic user

profiling on Twitter. In their study, they employed dynamic user and word embedding (DUWE) model and streaming keyword diversification model (SKDM) in order to address the problem. However, DUWE was employed to track the semantic representation of the words and users dynamically over time, and modeling of their embedding in similar space to effectively measure their similarities. Dynamic profile that considers time may distinguish between long-term and short-term interests. While the short-term profile depicts the current interest of the user, the long-term represents the interest that does not change always.

B. USER MODELING

A user model is a unique characteristic of the adaptive system. It represents user's information for effective adaptation in an adaptive system. For instance, it facilitates in prioritizing an adaptive selection of important/relevant item for user when searching for relevant data. However, the acquisition of data through the adaptive system for user modeling via different sources, (which may consist of implicit observation, the interaction of user or direct capturing of user data explicitly) is essential for the creation and updating of a user model. In order words, the process is referred to as user modeling. Gao *et al.* [33] in their work on user modeling classified user modeling into the following three classes.

- *Behavioral modeling:* The behavioral model is based on human behavioral patterns. In this model, the data obtained during the interaction between the systems and the user is stored, and the estimation of the intended action is obtained after the analysis of the previous action
- *Interest modeling:* The interest model is formed by describing the techniques used in calculating the interest degree of the new item or venue, etc. Interest modeling can be in the form of direct or indirect modeling. The direct approach of interest modeling explicitly request from the users on what they like. On the order hand, the indirect form of interest modeling captures users' preferences based on the previous browsing history such as the period allocated in reading a book or clicking on hyperlinks.
- *Intention modeling:* The intention (conveyed in the form of goal, aim, and purpose) denotes what a user plan to accomplish or the purpose why the user searches for the information. For instance, customers can be grouped into two categories based on their intentions such as the group without buying intention and the group with the buying intention. This form of user modeling attempt to find the ultimate reason why the user started interacting with the system. It is developed upon the foundation of behavioral and interest modeling.

C. USER PROFILE PROCESS

This section describes the user profiling process. It provides the entire processes that are involved in user profiling such as profile construction, user information collection methods (implicit and explicit methods) and user profile updating. The detailed description of each process is given in subsections below.

1) USER PROFILE CONSTRUCTION

User profile can be constructed for an individual by obtaining the user information through direct interaction with the user, or automatically by the system that monitor the activities of the user. In user profile construction, various learning algorithms/information retrieval systems are employed based on the choice of representation. Profile construction can be categorized into the semantic network, keyword and concept profile. Users or experts can manually construct a user profile. However, this approach is hard and time consuming for most users, which hinders the expansion of personalized service adoption. In contracts, the technique that automatically employ user's feedback for profile creation is most popular. Other approaches such as neural networks/genetic algorithms that rely on probabilities or vector space model are generally used and have been found to be more efficient in many domains.

Despite the fact that user profile is normally constructed based on user's topic interest, various studies have extended the profile construction by considering other profile topics that are not of interest to the user [34], [35]. In that regard, the application of both methods are made available for the system to identify the critical documents and eliminate the irrelevant ones concurrently.

2) USER INFORMATION COLLECTION METHODS

Information gathering about a particular user is the starting point for user profiling techniques. However, it is expected that the system exclusively identifies the users, which serves as the essential requirement for the system. The information about the users might be obtained in the form of the user's input or automatically collected by an intelligent agent. In order to obtain user identification, five standard approaches such as login, software agents, enhanced proxy servers, cookies and session ids are applicable. Nonetheless, cookies are more efficient and widely employed among the techniques due to its transparency to the users and the provision of cross-section tracking ability. Furthermore, in terms of consistency and enhancement of accuracy, login approach is preferred since it monitors user's action over a session between computers provided that users are willing to register and login in every visit. In order to create a user profile, user information can be obtained explicitly or implicitly [1].

Explicit Method: The explicit method for user information collection, also known as acquiring user feedback, depends on capturing the user's personal information. In such an instance, the user profile can be obtained explicitly by a means of the direct intervention of the users. A user may be required to express their opinion while filling the form. The explicit information is supplied by users through a survey and registration process [36]. The obtained data may consist

of demographic attributes such as the user's name, user's address, his telephone number, marriage status, job status, birthday, personal interest and hobbies. Other pieces of evidence about the user such as user's online transaction or web activity can also be classified as explicit information. Such pieces of evidence about the user, for instance, may consist of the user's average amount consumed on the item purchased, the categories of the most purchased item and the most frequent web visit by the user [24]. This method of information collection is not efficient and has a low usability rate because users are often reluctant to make their profile information public as they consider their privacy more important. Conversely, filling of the forms is burdensome, and users always tend to avoid it. In addition to the above limitations, this method of data acquisition has inherent time consumption and it is subject to the willingness of the user to participate. Thus, the inability of the user to provide personal information will halt the construction of the profile.

Implicit method: This method of the profile collection relies on implicit user feedback for its creation. In this method, implicit information can be gathered through intelligent agent or data mining techniques that analyse user activity like obtaining user's rule [37]. Alaoui *et al.* [3] explained the implicit profile collection method as a means of acquiring information in order to build a profile that monitors the user's actions. For instance, a user is more anxious about the future usage of the file he created and saved on his system. However, this method has an advantage over the explicit since it does not require any extra-role to be performed by the users during the construction process. This method was demonstrated by Kelly and Teevan [38] during their study in providing the most useful approach employed in acquiring implicit feedback and the type of information that can be obtained from the user's behavior. in addition, the implicit profile method has an advantage of automatic updates as it mostly depends on machine learning application. Nevertheless, it requires a large amount of interaction between the user and the content in the initial stage before the creation of an accurate user profile. Implicit profiling is synonymous with ontology-based user profiling.

3) USER PROFILE UPDATING

The user profile updating usually occur after the successful creation of the user profile. Updating means submission of a particular query to the system. Accordingly, the system retrieves the specific element (searched for) and keywords from the query and then authenticates the occurrence of the target in the profile. The query is strengthened based on the user's selection criteria if the verification is positive. Besides, the system provides useful services to the user based on the hybrid solution or user content matches and estimation [3].

III. MODELING PHASE OF USER PROFILING

This section gives a review of user profile modeling phases. It discusses the sequence of steps that are involved in modeling user profiling. In user profiling, the analysis phase also known as modeling phase is a crucial part of user profile implementation. Considering user profiling as a data-mining task, the idea is to construct a user profile that can represent the activities and interest of all users, which can be analyzed to predict the user's future needs [39]. The modeling phase consists of Data collection, Pre-processing, Feature extraction and Analysis (using various profiling techniques). The subsection below provides a review of the aforementioned phases. The taxonomy of the user profiling techniques is depicted in Figure 2 whereas the summary of the modeling is shown in Table 2.

A. DATA COLLECTION

In this phase, the datasets required to support the comprehensive implementation of the user profile are collected. In addition, the acquired data can be in the form of physical context data represented in terms of time, weather, and temperature also provide extensive support to user profiling [40]. In a recent study, Lakiotaki *et al.* [41] collected different user data to improve user profiling in multi-criteria user modeling for a product recommendation. The authors collected data about users, which include user preference data and preference statement. User preference data define the numeric rating of different items evaluated by users based on their preference order while preference statements depict the ranking of alternative items that belong to the same group (also referred to as a weak-preference order). This preference of user information is then used to build a model that accurately detects user profile interests. Moreover, user profile data are acquired by contextual information such as location, company's name, emotion status or personal digital devices like network connectivity or bandwidth. Early profiling systems focused on acquiring data from users directly. However, this method is ineffective as users are not concerned in providing their information publicly. Consequently, the investigation has now shifted to the user profiling based on the user's behavior. This is referred to as behavioral user profiling. Furthermore, most profile data are sourced from the social platform. In a recent study, Chen *et al.* [42] developed a user profile based on data collected on a social network site and personal information shared by users. These social networks include Twitter, Facebook, Myspace and Instagram. Twitter bio-data, for instance, consists of the user's full name, education status, occupation, location, short biography and number of the tweet, which gives more information about the users such as their interest, what they engage in, where they live (location) and their self-conception.

B. PRE-PROCESSING OF DATA

Most of the profile data collected from social media contain many flaws. In that regard, there is need to clean the obtained datasets so that the actual features that will be extracted for the profile modeling would produce a better performance result. In addition, some extracted data sometimes appear as a duplicate. These duplicate data are removed during the pre-processing stage. Most researchers employed different

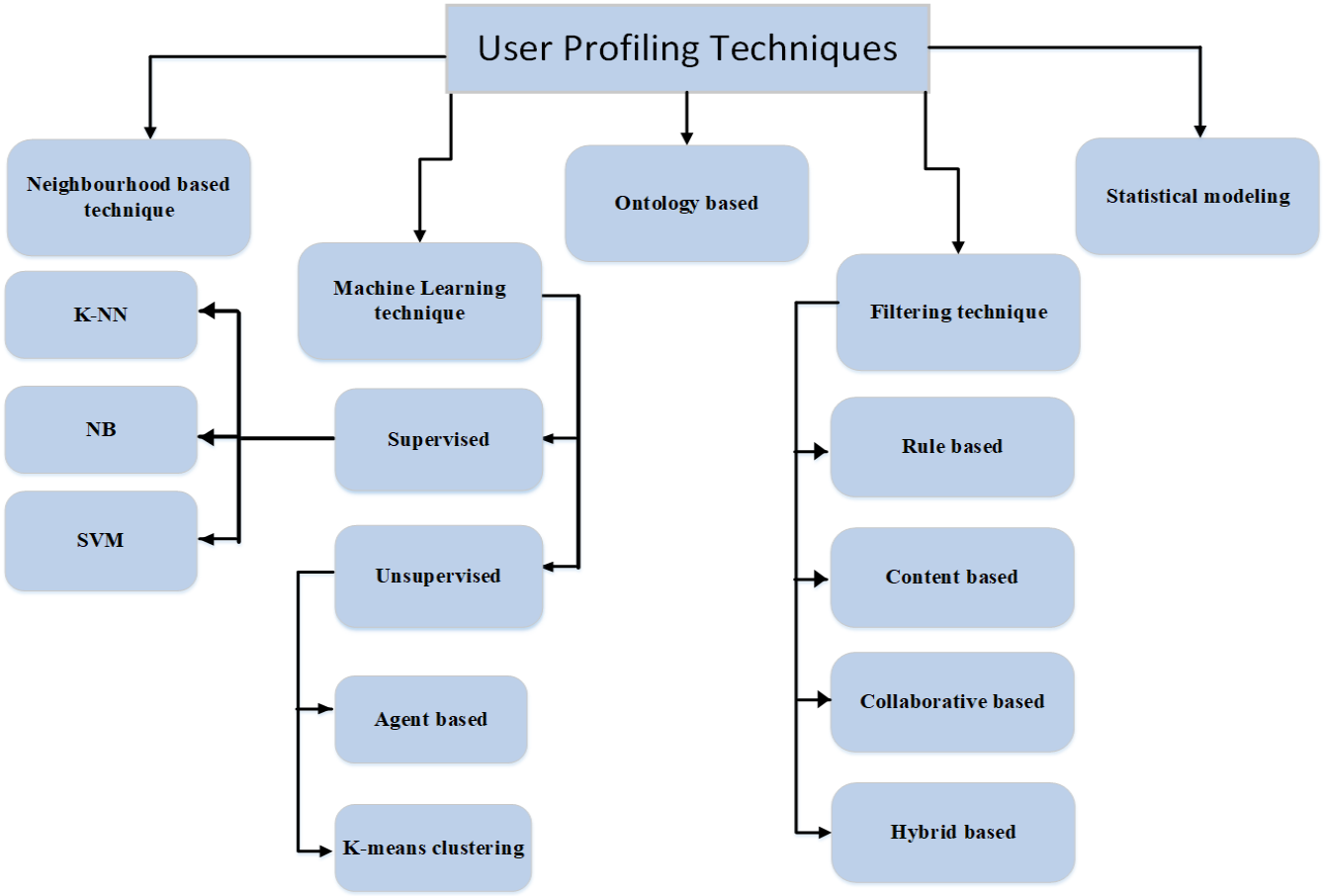

FIGURE 2. User profiling techniques.

pre-processing techniques in their studies on user profiling in order to prepare their data for the analysis phase. For instance, Tang *et al.* [43], implemented an effective and efficient Hidden Markova model approach for data purification to enhance user profile system. The proposed method aided in eliminating the problem of disambiguation in collecting user profile information. Some basic pre-processing techniques are tokenization, stop word removal and tagging. Tokenization is a process of segmenting the textual data into tokens using tokenizer and allocation of the tag to each token.

C. FEATURE EXTRACTION

Feature extraction is one of the crucial aspects that is required in profile modeling. Feature extraction involves the extraction of the user profile features from diverse domains. Different researchers in this domain have employed different procedures in order to extract these features. The feature extraction stage is required in order to extract the needed feature that will improve modeling performance. The most commonly used features in the recently proposed approaches found in literature includes content features, pattern feature, profile features, term feature, and user behavioral features, etc. For instance, Tang *et al.* [43] in their study on authorship profiling evaluated a set of profile feature that consists of six sets of attributes in article publication data. It include: Publication title, Abstract, Publication venue, Abstract authors, Publication year, and References. These features were extracted from the digital library (ACM, Springer, and IEEE) using heuristics. Thus, the attributes of the paper were represented as a feature vector and the number of occurrence as the values. These profile features extracted were trained using support vector machine learning algorithm to find the important page from the web. Moreover, the authors used tree-structured and linear conditional random field (LCRF) to model the relevant features. In addition, the study established that the tree-structure model performs better than the LCRF in comparison.

Li *et al.* [44] in their study also worked on profile extraction. In their work on the social networking site, they extracted data from Twitter using a supervised learning classifier to train the model. Consequently, this method considered network information only, which is not enough to infer the attributes of users. In another study, Mislove *et al.* [45] proposed a method to infer user

TABLE 2. Summary of the modeling phases of profiling.

Data source	Feature extracted	Profiling approach	Strengths	Weaknesses	Performance metrics	References
System log	Terms in user interest content	Neighbourhood approach	An adaptive search of user profile	User profile interest relies only on the viewed history of the webpage	Accuracy, Recall, and Precision	[26]
Twitter	User interest, number of friends, tie strength between the user and her friend	Neighbourhood approach	Better accuracy since the increase in tie strength leads to a decrease in average error distance	Difficult to apply in a situation where there is a frequent change in user location	Accuracy, Error distance, and Average error distance	[42]
Sina microblog	Content and semantic interest	Neighbourhood approach	Better performance in terms of recommendation accuracy and coverage	Time complexity in the selection of Neighbourhood and repetition of user-relevant interest	Precision and recall	[80]
Human resources	Key skill and job type	Supervised K-NN	Client-side and server-side of user profiling	Reliance on only demographic data for learning the preference of the user	Precision, Recall and F-measure	[54]
System logs	Content feature, pattern feature and term feature	Supervised SVM	Easily scalable	Limited to academic data and cannot be applied to general web user profiling	Precision, Recall and F-measure	[43]
Mobile logs	My data, preference, my space, rating	K-means clustering	Simplicity and fast coverage	Overloading and less scalable due to the existence of a centralized server	Precision, Recall, F-measure and Coverage	[40]
Mobile Wi-Fi log	Number of tokens, Average token length, Delimiters number, Number of Digits, Number of uppercase letters, and Number of lowercase letters	Supervised learning	Effectiveness of the proposed technique in cleaning noisy Wi-Fi data for user preference profiling	Many access points named by default settings or without any semantics	Utility measure, Normalized Discounted cumulative Gain	[81]
Twitter data set	Relevant keyword with their embedding	k-means clustering	Effectiveness in building user profiles by embedding words and users in a common semantic space	Able to generate only keywords but unable to generate phrases for building user profiling	Precision, Normalized Discounted cumulative Gain, mean reciprocal, and mean average precision	[32]
Event log data from Yahoo Recommends	Universal relevance, Context relevance and Personal relevance	Supervised learning	The proposed method improved the effectiveness of user profiling and content recommendation	The approach is only effective in building user profile based on implicit feedback and did not consider explicit feedback	Mean Average Precision (MAP), Mean reciprocal rank (MRR), Area under curve (AUC)	[82]
Facebook profile	Status update (LIWC), profile picture, Page like	Deep learning	Effectiveness in inferring age, gender and personal traits of social media users.	Profile picture is not an appropriate data source for predicting users personality traits	Area under curve (AUC)	[7]
Twitter	User tweets, user followees tweets, URLs posted by user followees and URLs posted by followees of followees.	Statistical based approach	Profiling of user interest for URLs recommendation	Profiling of user interest only rely on user interaction in twitter	NIL	[73]
System logs	The day the application commenced, the hour the application commenced, and if the application has previously been executed by the user or not	Statistical approach	User profiling is applicable for anomaly detection	Profiling of user only depends on the system log.	NIL	[74]

profile using the social network. They collected profile data of alumni and students of Rice University in the online social network. The attributes of the users such as the academic major(s) of their study, dormitory information and matriculation year, that depends on the network information were extracted as the feature for modeling. Another exciting work on text-extraction based on user mouse behavior features on web text was done by Hijikata [46]. Thus, the implicit importance feedback was fed to the system in order to generate the appropriate recommendations [47].

D. MODELING TECHNIQUES

User profile modeling is a process of building a computational model using the extracted features that can predict user needs or preferences. modeling techniques consist of neighbourhood-based approach, machine-learning approach, ontology-based approach, filtering approach, and statistical modeling approach, User profiles can be analyzed using machine learning algorithms. The machine learning could be in the form of supervised or unsupervised learning. The machine-learning algorithm is used for training and testing of the data. In recent years, various classification algorithms have been employed in the construction of the user profile. For instance, a recent study was carried out by Ying *et al.* [48], which concentrated on relating authorship verification and detection of a compromised account employed K-NN classifier as an instance learning that help in dynamic profile updating. The frequent updating of the baseline learning classifier improves the accuracy of the model and other profile attributes. On the other hand, the filtering approach consists of rule-based, content-based, collaborative and hybrid filtering. The description of the modeling techniques is given in the subsection below.

1) NEIGHBOURHOOD BASED TECHNIQUE

It is feasible that ''follow friends'' have a group of friend's interests referred to as a neighbourhood. However, a neighbourhood with adequate knowledge can assist each user to build the neighbourhood user profile in order to address the inherent shortage of information in personal interest representation. The neighbourhood creation process is a model construction process for the collaborative recommender. The main objective of the neighbourhood construction process is to determine for each user I, an ordered list of j users

$$M_b = (M_1, M_2, \ldots M_j) \tag{1}$$

in such that b $\in M_b$ where (b, M_1) = maximum and Sim (b, M_2) = next max, etc.

Sugiyama *et al.* [26], in their study on user profiling, investigated the adaptive search techniques that can obtain a search result based on the preference of the user. In their study, they employed the user's browsing history to create a user profile. As a result, the system can obtain relevant information for user adaptability, which consists of various pieces of information needed in the absence of human intervention. However, an update on the profile is usually executed whenever there is any modification in the user's web page. In another study, Chen *et al.* [42] utilized STFG model to find the Twitter user's location. In STFG, the factors are analyzed using attributes and correlations whereas the nodes are analyzed using labeled relationships in a social network. The study built a location profile for users specifically by estimating their city-level location. The provision of a ''user's followers'' of friend locations in their profile can propagate their location, and even the venue in their message can be used as a reference to infer their location. Moreover, Jurgens [49] proved that the nearest neighbor could be used to predict the individual location using a cumulative distribution function. By considering Twitter's network. the author indicated that half of the individuals have neighbors who disclosed their location within close proximity. The experimental result of this method showed that the model can improve performance for inferring the home location compared with many state-of-the-art methods. However, the method is limited to the extraction of location features in English, without considering users that use different languages to communicate on Twitter. Also, the home location for users who often move in different locations is another issue with this method, as it cannot estimate the home location for such users.

2) MACHINE LEARNING

According to Portugal et al. [50], ''Machine is an algorithm that uses a computer to simulate human learning and allows computers to identify and acquire knowledge from the real-world, and improve the performance of some tasks based on this new knowledge''. In computational science, machine learning studies the algorithm that it can learn from and establishes a prediction on data. This is achieved by using the input data to build a model in order to create a data-driven decision in place of twitting with many static program instructions [51]. Learning is knowledge acquisition. Computers learn from the algorithm, unlike people who learn from experience because of their ability to reason. The machine learning algorithm consists of two broad forms, namely supervised learning (input mapped to desired output) and unsupervised learning (auto-detection of data disregarding pattern to class assignment). The machine learning approach is a standard method of profiling in the recommendation systems [52]. In this review, both forms of machine learning (the supervised and unsupervised learning), are mostly used as profiling techniques as will be described briefly in the subsection below.

Supervised Learning: This form of learning utilises training data and testing data. In a supervised learning approach, the systems learn how to perform a task of new observation classification from the input data. The algorithm learns from the available training data and uses its application on real data [53]. The most useful supervised learning for profiling is K-Nearest Neighbour, Naive Bayes and Support Vector Machine. K-nearest neighbor is one of the supervised learning algorithms that is employed in both classification and regression problems. The algorithm depends on the measurement of similarities for data classification by using majority neighbor voting. Thus, the assignment of data to the class is determined by the highest nearest neighbor and the accuracy of the classification can be enhanced by increasing the number of the nearest neighbor. KNN algorithm application was demonstrated in a study conducted by Bradley *et al.* [54], on user profiling based on personalization. The study seeks to monitor user activity in order to create a user profile [55]. The task of the personalization technique employed here is regarded as a classification problem. For example, this approach uses the availability of user profile for individual

Job Seeker (which contains the job's history and the user's likes or dislikes) for classification. Moreover, after conducting the search activity, the retrieved result can be classified as either relevant or irrelevant for each job retrieved. Thus, the K-nearest neighbor algorithm was employed in the classification phase [54].

On the other hand, Naive Bayesian classifier is a learning algorithm, which is also referred to as the state-of-the-art of Bayesian classifier. It assumes that there is no relationship between the features in a class, even if the features depend on each other. It is computationally efficient and simple algorithm used in data mining and machine learning applications [56], [57]. The effectiveness of Naive Bayes classifier is found useful in the domain of interactive applications. In addition, Naive conditional independence assumption prevents it from achieving optimal accuracy in profiling. Support vector machine is another supervised learning utilized on user profiling. It is applicable in text classification of genomic data and difficult data types than feature vectors [58]. The algorithm was utilized in a study carried out by Tang et al. [43] to identify the relevant documents on the web by sorting out the profile information using Tree-structured Conditional Random Fields (TCRF).

Unsupervised Learning: In unsupervised learning, the machine learns real-world data on its own since the provision of the labels are not available in this case [59]. The unsupervised machine learning techniques commonly employed in user profiling are multi-agent system [60] and K-means clustering. A multi-agent system is used to improve the retrieval result and evaluation criteria by creating multiple agents that handle different personalization issues and phases. The agent uses the web search model to retrieve the best result that meets user preferences. This approach has an advantage over the traditional search engine in such that the building of the profile starts from the scratch with the basic information and is maintained until the end by utilizing user's feedback [61], [62]. Nonetheless, the profile input is subject to bias as the user description is carried out by the users and the profile degrades over time due to the static nature of the profile [60]. On the other hand, K-means clustering is an unsupervised approach that uses K-means algorithm for user clustering. User clustering is an algorithm that partitions the distinctive datasets into individual group behavior in order to determine user profiles.

Therefore, clustering is deployed to group user data objects depending on the information contained in the data that specifies the objects and their association. The contents of the group behavior determines the grouping of users into separate classes using an algorithm. In addition, the separation of the two groups behavior can be used as the input data, whereby users are divided into different groups depending on the rating of the applications. It also deals with the assignment of the set of observations into clusters in such a way that the observation in a similar cluster looks alike in some sense. Furthermore, it uses a classifier called K-means to classify objects based on attributes into k numbers of the groups. Various researchers have proposed clustering algorithms for user profiles.

For instance, Han and Chen [63] proposed a fuzzy clustering approach to build an ontology-based user profile that represents a sophisticated user's needs. This method is used for an effective information retrieval system. It has a unique feature when compared with other approaches. One of the notable characteristics of fuzzy clustering is a parallel allocation of information to more than one user profile that consists of diverse states of accuracy. However, initial parameter settings is one of the weaknesses commonly found in most user clustering technique like the cluster number and initial position of the centroids. This weakness can be constrained by deploying demographic profiles for individual behaviours, preference profiles for group behavior, and application of the global k-means algorithm on both categories. The result of the two components and their corresponding users are stored in the database. However, this global k-means does not depend on the fundamental parameter values but on its k-means algorithm application as a local search procedure. Thus, it acts incrementally by optimally computing of one new cluster center at every step that minimizes the clustering criterion rather than choosing an initial value basically for every cluster centers [40], [41].

3) USER ONTOLOGY

An ontology is a ''conceptualization of a domain into a human-understandable, but the machine-readable format, which consist of entities, attributes, relationships, and axioms'' [64]. User ontology has more probability of improving user profiling because of the evolution of the semantic web. Ontology technique does not suffer from the difficulties in the interest sharing that is most common in the representation of the user's interest since it can share relevant information in other systems with the effective representation of the user's interests. Han *et al.* [65] proposed an ontology-based user profile acquisition (OUPA) method grouped into construction user ontology (CUO) and maintenance of user ontology (MUO). The MUO profile method uses a K-nearest neighbor algorithm while OUPA acquires the user profile via the automatic construction of user ontology to maintain the representation of personal interests. The automated approach solves the time-consuming problem inherent in the manual approach. In addition, it makes the user profiles stronger and more expressive.

4) FILTERING

The filtering recommendation approach to profiling is a method of filtering information that meets the user's specific need in different situations and removes the irrelevant information about the user. This approach consists of rule-based, content-based, collaborative-based and hybrid methods. The descriptions of these approaches are given below.

Rule-based Approach: In this approach, rules are specified by the information system based on the demographic or a static profile of users obtained via the registration process by

asking users a set of questions. The application of the pre-specified ''if this then that'' rule is used to select the useful information for a recommendation [66]. Its effectiveness usually relies on the knowledge quality of the rules. However, it has poor maintenance issues and is prone to bias since the input is the subject of the user's self-description or their interests.

Content-based Filtering: In this user profile filtering approach, the user's interests depends solely on the matching of items when its contents are compared with the user profile. However, the best matches after the comparison are taken as user interest. Godoy and Amandi [11], on their survey study on user profile for personal information agent noted that users exhibit related behaviours under similar circumstances. It is also referred to as cognitive filtering [67]. Thus, this approach has a content dependence issue due to the difficulties in analysing the limited contents of the items and as such, reduces the performance results [2], [68].

Collaborative based filtering: In collaborative based filtering, user's interest in an item is established based on the user's previous interests on the same item. It usually depends on the knowledge that a user who decided on items in the past will possibly go for it in the near future. Thus, this method matches users with related interests into groups of peers, thus allowing the aforementioned idea of recommending the item within a similar group of users [40]. This type of filtering works with an algorithm that aggregates the feedback provided by different users and recommends items for users by considering the similarities between users in order to offer recommendations to the target users [69]. The algorithm is classified into a model-based and memory-based. The memory-based algorithm such as vector similarities and collaborative analysis searches the user's database for user profiles that are similar to the active user profile that contains recommendation [70]. This method is widely used in e-commerce and social media among others because it accommodates all types of items and aids in finding the likely user's interest. In contrast to memory-based, the model-based approach assumes that users of the same group (e.g. age, sex, social group) have the same profile as a result of their similar behavior [68]. The success of this method deeply relies on how well the clustering of profiles is associated with the users. For example, amazon.com broadly employs collaborative filtering recommendation algorithms to personalize its web page on individual customers based on their interests. They built their filtering algorithm due to the fact that the existing filtering recommendation algorithm is negligible and unsalable in nature compared to the number of amazon products and a large number of customers that visits amazon.com on a daily basis. Their collaborative filtering algorithm is highly scalable and produces high - quality recommendations with large datasets [71]. These filtering approaches cannot help in a cold-start situation with the absence of the user's initial ratings.

Hybrid based Filtering: This method combines the features of the content-based filtering and collaborative filtering method to enhance their performance. This is to prevent the drawbacks inherent in the collaborative and content-based method. The approach has been proven effective in several application areas such as web search, electronic commerce, sensing, monitoring, and financial-based systems. The implementation is determined by joining the prediction result obtained from content-based and collaborative-based method mixed with their characteristics. The study conducted by Park and Chang [67] on modeling the customer profile for both individual and group behavior, demonstrated the effectiveness of hybrid filtering approach in product recommendation by considering both user and group interest.

5) STATISTICAL MODEL

A statistical model is a technique used for building a user profile by employing a list of keywords or user logs. In web system, this technique may consist of a highly frequent word obtained from the visited web page by the user [43]. Recently Chen *et al.* [72] presented URL recommendation model for recommending Twitter's user with the URL contents stream they are interested in. The constructed model gives a recommendation established on Twitter's user profile whereas their user profiles are dependent on user's tweet, favorite URL, users' followees' and social voting within users' neighborhoods. In another study on the user profile, Corney *et al.* [73] concentrated on building a user profile using computer system logs. The constructed profile extracted the usage pattern of the program and processes running on the system. In such an instance, the activities that vary from the user profile can serve as a signal to the administrator and subsequent action can be taken. Consequently, this study only concentrates on user logs, in which the behavior of the user is limited.

E. PERFORMANCE EVALUATION

The evaluation phase is very crucial in any modeling problem in order to test the performance of the model. In user profiling, the following standard evaluation measures indicated in equation (2) - (6) are used to evaluate the performance of the model. They can be calculated by estimating the True Positive (the accurate representation of the user interest or preferences), False Positive (the false representation of user's need or preference), True Negative (the accurate representation of user's need by a different user), and False Negative (false identification of the user's preference of genuine users by an imposter).

Accuracy (ACC): The accuracy provides the percentage ratio of the overall instance found to the overall instance. It is computed as

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

where TP represents a true positive number, TN denotes a true negative number, FN represents a false negative number and FP represents false positive number

Precision (PRE): Precision is a performance metrics computed as the ratio of true positive to the summation of the true

positive and false positive

$$PRE = \frac{TP}{TP + FP} \tag{3}$$

Recall (REC): Recall is the incorrect representation of the user's interest or preferences. It computationally represents the ratio of true positive to the summation of the true positive and the false negative.

$$REC = \frac{TP}{TP + FN} \tag{4}$$

F-measure (F-M): F-measure represents the combination of precision and recalls particularly when there is severe equality of false positive and false negative. It computes the harmonic mean of precision and recall and assumes 0 and 1 values.

$$F - M = \frac{2 * (Precision * Recall)}{Precision + Recall} \tag{5}$$

Mean reciprocal rank (MRR): This is a measure of the average of a multiplicative inverse of the rank of a target of the testing set across the number of target tags in the testing set. In information retrieval, MRR metrics is often employed in the application domain such as search engine and recommender system. It is calculated as

$$MMR = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{r_i} \tag{6}$$

where r_i is the rank of the target tag.

Different performance parameters such as Accuracy, Precision, Recall, and F-score have been used in a recent study carried out by Kaur *et al.* [74] on authorship verification by using textual features to test the authorship tweet posted by users on Twitter. When the required feature is well selected, the performance parameter of the accuracy, recall, precision, and F-score will be as high as possible, which indicates good performance.

IV. OPEN RESEARCH CHALLENGES

This section highlights the challenges encountered by researchers in the creation of user profile studies. These research challenges are currently limiting the real-world application of user profiling. Figure 3 shows the pictorial representation of major challenges inherent in the creation of a user profile while the details are explained below.

A. MANUAL PROFILE CREATION PROBLEM

Profile creation method that uses some social network services such as YouTube and LinkedIn requires individuals or experts to manually create the user profiles by supplying their personal data themselves [75]. The profile obtained through this method is termed an incomplete profile, as the individuals are not willing to provide some of their profile information while creating profile themselves [43]. In addition, it is difficult to implement, time-consuming and cannot be extensively used for service personalization. However, the automatic approach for building a behavioral profile using machine-learning techniques that can learn user behavior is more efficient and reliable.

B. PRIVACY

Privacy is a challenge that is peculiar to a behavioral profile and calls for further investigation. In a recent study on a behavioral profile in the intrusion detection system, it was revealed that an attacker that gains access to user profile information could use it to compromise some aspect of behavioral profile that is difficult to hide. For example, a user that often posts information on the social network site may be vulnerable to attackers if the attacker studies the profile writing style of the user. An intruder can use it to create the profile of that user as he reads it. In addition, the keystroke of a user as he enters the text into the public site can also be vulnerable to the attackers for creating their profile from the users [18].

C. INSTABILITY AND LIMITATIONS

Studies carried out by Grcar *et al.* [76] and Li *et al.* [77] considered user visitation models that consist of long-term and short-term interest. In their work, they approved all the visited web pages to be interesting to the user irrespective of the time spent in reading the page. However, there is a potential ignorance on the strength of the interest at the side of the user as the equal opportunity is opened to all the pages. This calls for an identification of pages that are not relevant to the construction of the user profile. For example, pages like webmail, portal entry pages, and search engine result are not suitable for user profile construction. However, the extension of the stop-word collection with some common internet word that allows the user to specify their irrelevant URL expression in the profiling process will make a negative impact on such pages ineffective during the profile construction phase. Another challenge found in profile creation is the limitation of the long-term and short-term folder to a predefined size in handling a page view. For instance, if the short-term folder was set to keep most frequently viewed pages (n = 5) whereas the long-term folder is set to also keep the viewed pages for a longer time (n = 300), then the initial visit on any page it kept in both folders. Eventually, it gets pushed out by the other pages that viewed afterward [76]. However, the most recent page is saved in the short-term folder whereas the last viewed page is saved in the long-term folder. This limits user interest and can lead to instability due to the frequent change of user interest.

D. INACCURATE MULTIDIMENSIONALITY REPRESENTATION OF USER PROFILE

The representation of a user profile in more than one dimension poses a significant challenge in the recommendation service. In a recommendation service, for example, the user profile that considers the rating given to video music might not be utilized to the rating of the restaurant for the same user. This constraint calls for further investigation in order to create an adequate profile that captures different users'

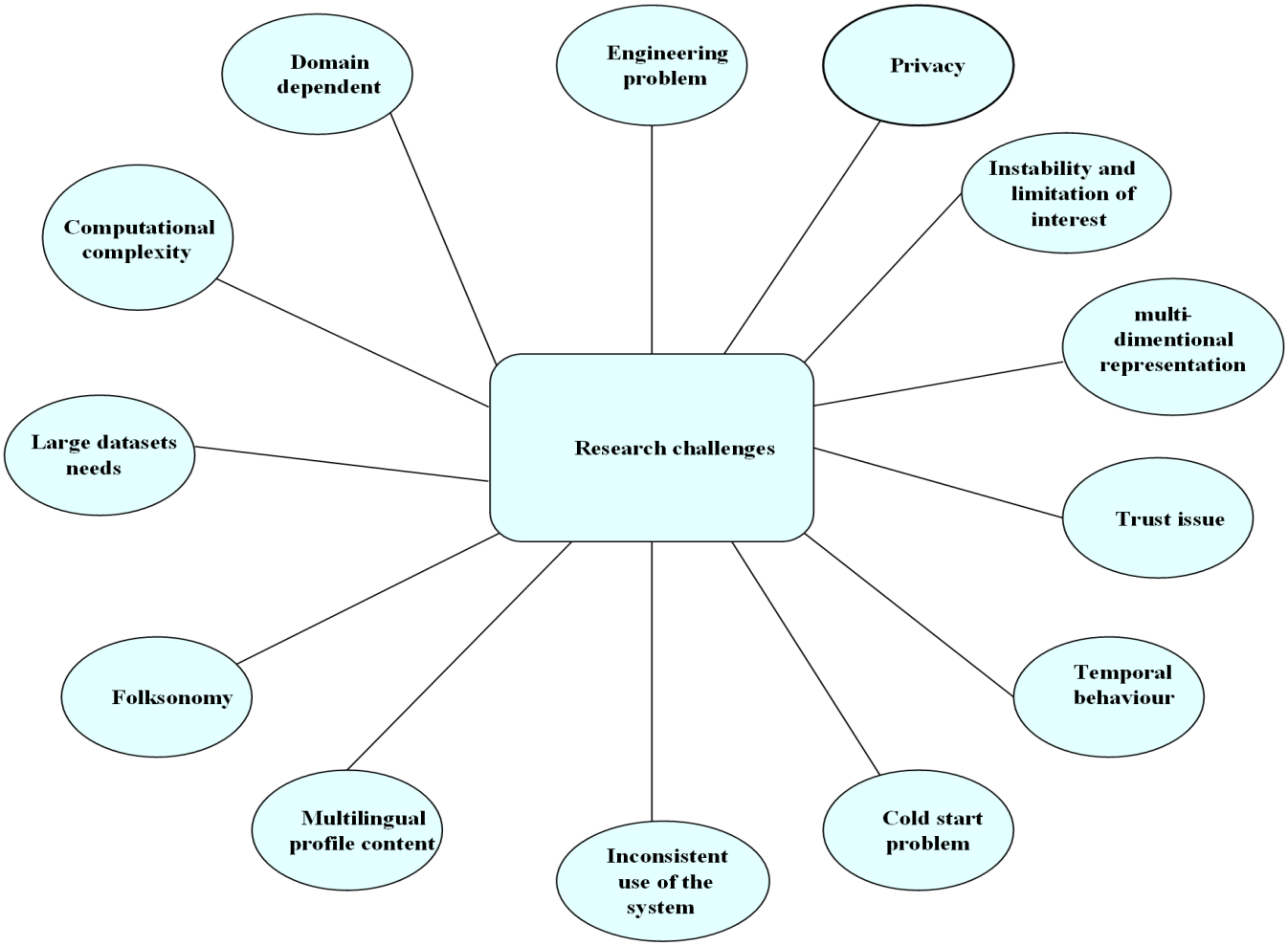

FIGURE 3. User profiling research challenges.

information such as user preferences, interests, and demographics [68]. This profiling approach employs the available user's related information to personalize different services from any third-party service provider. In addition, the feature weight is to be given due consideration for the accurate use of multi-dimension profiles because of different preferences and service personalization. For example, the book interest information of a user may not be as relevant as the user's income information for a personalized restaurant recommendation service.

E. TRUST ISSUES

Users often spread false information on social media. However, using this information in building user profile will lead to a false profile creation, which might be misleading in service personalization due to the fact that user profile is expected to be a true information about an individual. Consequently, there is a need for more investigation on the profiling approach that will be able to distinguish fake profiles from authentic profiles. Trust issue was examined in a study carried out by Vassileva *et al.* [78] but their approach was not able to provide a substantial fake profile detection method to efficiently estimate the trust in the profile information.

F. TEMPORAL BEHAVIOUR

In a behavioral profile, there are often changes in users' interests. The current user behavior might change within a period and this calls for further investigation on a method that monitors the changes in behavior and performs profile updating based on the state of the behavior and the change in interests. However, this can be achieved by considering the current interests of a user as a function of her interest in the previous time interval.

G. COLD START PROBLEM

user profile creation process requires sufficient information gathering. However, when the information is not sufficient, it will lead to a cold-start problem. The cold start problem is common in learning and adapting dynamic user profiles for personalization, where the system is not capable of providing an effective personalization service in order to learn the user profile. This is very common in content-based personalization systems [82].

H. INCONSISTENT USE OF THE SYSTEM

User profile personalization for a user recommendation might be left unused for a very long period. The user might decide to use the system after a change in user interest had taken place. In such a situation, the system will continue to offer recommendation services based on the previous interest of the user by employing the knowledge of the previous information without considering the change in the user interest. Consequently, when the user profile utilizes the previous information, the profile information will be more dedicated, and the system will likely find it difficult to offer a new recommendation to the users [82]. Thus, there is a need to consider an automatic update of user interest even when the system is abandoned for a period.

I. MONOLINGUAL PROFILE CONTENT

By monolingual profile content, we mean having the contents of user profile represented in just one language. Most profile contents are written in the multilingual language. For instance, an ACM computing survey on user profiling was carried out by Barforoush *et al.* [75] using several content languages other than English. The systems to analyse such contents automatically are not available as most techniques for profile analysis are carried out in English. Consequently, there is a need to pay more attention to language independence study on user profile in order to solve the monolingual content extraction problem of user profile. However, a recent work conducted on multilingual by Nagy and Farkas [83] did not provide sufficient solution in solving the problem. Therefore, more work is needed in that area.

J. FOLKSONOMY

Despite the easy implementation and evolution inherent in the folksonomy profiling approach, there is still a tag polysemy problem in this approach. Tag polysemy is a situation where tags of the same word come in diverse forms. For example, blogging, blog, and blogs are polysemy. This approach might not yield good accuracy in the analysis phase since it cannot handle synonyms and homonyms [16]. A technique that will be able to distinguish different polysemy tags is another research area in folksonomy profile that requires more investigation.

K. THE NEED FOR THE LARGE DATASETS

A collection of large datasets is the most crucial issue on a direct application of a machine-learning algorithm for profile modeling. In most cases, the algorithm needs a relatively large dataset to build a model that can produce acceptable accuracy. The learning algorithm in most cases needs several training examples with datasets for better accuracy. This is possible since there are many classification model option that can be applied. Moreover, the annotated data is another issue that is most common in supervised machine learning. The availability of the volume of the annotated training data will determine the efficiency of the supervised learning approach of user profiling [5]

L. COMPUTATIONAL COMPLEXITY

The advancement in information and communication technology, particularly the internet is bringing new opportunities in assisting users through user profiling. However, the amount of information in existence and the number of online users has generated many challenges in user profiling. Some of the popular sites such as Yahoo and Google that receive over one million visits on a daily basis are liable to computational complexity in user modeling if every user's profile is to be generated automatically [5]. Therefore, method that can handle computational overhead is required in a user profile.

M. DOMAIN DEPENDENT

Domain-dependent is one of the shortcomings commonly found in rule-based profiling techniques. It requires a design of profile extraction rules that need human experts. However, the creation of rules is labor demanding and consumes a lot of time. However, it is one of the most straightforward approaches in user profiling that produces better accuracy on a given domain, yet the handcrafted rules are highly domain-dependent and cannot perform well on a different domain [75]. Hence, the rule-based approach are not appropriate for the large-scale domain, but highly effective on small-scale domains.

V. SOLUTIONS TO RESEARCH CHALLENGES

This section provides a promising direction for further researches on user profiling. Based on the review of this study, the following open research directions that serve as the solutions to the challenges above challenges are proposed: They include ontology representation of the user profile, general-purpose profile, dynamic profile, fake profile detection, secured user profile creation, distributed system, and language independence. These solutions to open research challenges is presented in Figure 4 and described in the subsection below.

A. ONTOLOGY REPRESENTATION OF USER PROFILE

The ontology representation of the user profile especially in e-learning has a feature of defining the terms employed in explanation and representation of knowledge in the learner's profile. It makes the sharing of understanding among the users and reuse of the domain knowledge possible. Besides, it analyses and extracts the domain knowledge from the operational knowledge, explicitly shares, and exchanges profiles within the system upon an agreed model. However, using a specified weight to distinguish between the relevant and irrelevant information for effective user profiling can provide a solution to the problem like wrong multi-dimensional representation of the user profile. This can be carried out by investigating a means of changing different profile representation into unified knowledge. A mechanism that can map profile information that has different schemas into a unified system and can divide into different profile schema representation is

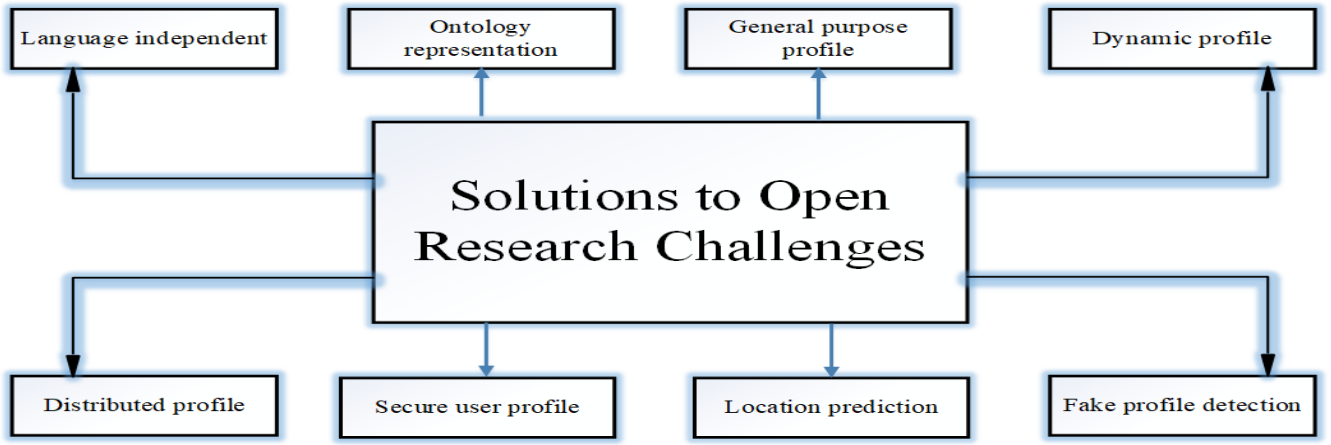

FIGURE 4. Solutions to research challenges.

the best way to achieve profile transformation into a unified format.

B. GENERAL PURPOSE PROFILE

Most user profiling systems are domain-dependent. In contrast, the general-purpose profile is independent of any domain and is applicable in different applications. One of the methods of implementing the general-purpose profile is by employing the building blocks of the profile system that do not rely on the domain.

For instance, in an entity profile, the profile data extraction requires training by semi-supervised learning algorithms. Likewise, the open information extraction technology is used for profile components extraction in order to facilitate the domain independent and to scale up different quantities especially in the web domain.

C. DYNAMIC PROFILE

User profiles may exist in the form of a static profile or dynamic profile. Dynamic profile, for instance, is a form of profile where the system automatically generates the profile of the users without requiring users to supply any information. In the contrast, the static profile requires users to supply their information before the profile can be created. In such case, users seldom provide all their information accurately due to some privacy issue, which makes the static profile unreliable. In addition, there is an entity relationship and limitation inherent in populating static profile, which requires the application of the existing user ontology to direct the profile extraction, define the set of relations in question and provide the entity dictionary. However, the emerging of dynamic profile now serves as a bypass to an entity and relationship limitation inherent in populating a static profile. Therefore, there is a need to focus more on dynamic profiling.

D. FAKE PROFILE DETECTION

The increase in social network usage influences social interaction and entity profile sharing among users. Protecting the privacy of individual users has become a serious challenge due to the amount of personal information sharing among friends in the online social network. This has given rise to the exploitation of the personal profile of users to create profiles. This problem is demanding because there is no reliable mechanism in most cases that can detect and differentiate the fake profile from the authentic profile. An in-depth study of fake profile detection method would help and detect the fake information in the profile and will conversely develop confidence in the profile information.

E. SECURED USER PROFILE

As discussed from the previous section that the insecurity of the writing style and the keystroke of the behavioral profile are vulnerable to the attackers and they employ it to create the profile of the user. It should be noted that one of the solutions to this security challenge is to keep the user information secret from the attacker in order to make profile falsification that corresponds to the genuine user very difficult. However, the possibility of this solution will immensely depend on the reliability of the information source and keystroke since it is very easy to manipulate the input data. Thus, there is a need for an additional security mechanism to strengthen the user profile so that the intruder will find it difficult to falsify.

F. DISTRIBUTED SYSTEM

Most dynamic user profiles are sourced from a large-scale system like social network site. Mining profile information from such a system poses a great challenge to the user because such an entity profile is subject to noise, scale, uncertainty, ambiguity, and trustworthiness and is most often out of date. This inherent nature of the profile has shifted the state-of-the-art research to focus on a new approach that extracts user-centric information from social media and the web. This can be achieved by using a distributed system like the multi-agent system in a large-scale and complex environment. Further research on the extraction of profile information from the

web or social media using a distributed artificial intelligence method is required in this research area.

G. LANGUAGE INDEPENDENCE

Research on multilingual profile digital content has increased. Language independence in profiling context is referred to as the representation of web-based or online social network profile contents in multi-languages. However, a system to analyse information in a different language and multilingual digital components electronically is uncommon. Most web contents, including the social media site are displayed in multiple languages. As a result, most researchers in the profile domain are now shifting the attention from monolingual profile contents to language-independent. No work has been done on the analysis of multilingual user profile contents. Therefore, research in finding the mechanism for language independence that will be able to analyse profile content in any language in order to produce an accurate profile is needed.

H. LOCATION PREDICTION

Obtaining user's location information via social network is another issue inherent in user profiling due to its unavailability since most users do not usually provide a location in their profiles. This may be due to privacy concern, lack of trust or their mindsets towards it. There is need for effective techniques that can predict the user's location. However, these techniques have been understudied by researchers recently. In this case, user location can be achieved by building a user location prediction technique that can envisage user location by employing different implicit information within the network. Moreover, these techniques will facilitate its application in a diverse domain such as location-based recommendation, disaster management, location-based advertisement, monitoring of disaster eruption, and emergency report.

VI. CONCLUSION

This article has carried out a survey on various approaches for user profiling. A user profile is the representation of the users' need, preferences, interests, and behavior whereas user profiling is the practice of gathering, organizing and interpreting the user profile information. It is also defined as the act of representing the user's need or interest. Valuable information about users can be obtained from diverse sources through profile extraction. User profile extraction can be in the form of manual inputs provided by the user via forms and surveys or automatically by applying various extraction techniques to extract data from the web, mobile network or social networks. These techniques aid user-profiling systems in obtaining interesting information about users. This comprehensive survey investigated the concepts and the state-of-the-art approaches of user profiling. It surveyed the modeling process in the aspect of data collection, pre-processing, features extraction, the modeling approach, and the performance evaluation. It further provided a summary of the strengths and weaknesses inherent in each approach. Finally, it discussed the research challenges encountered in the previous study within the field of user profiling. Based on the survey, the study reveals that there are still open challenge that need to be addressed and proposes open research directions that can help in building an accurate user profile for service personalization.

ACKNOWLEDGMENT

This research was supported by the University of Malaya Research Grant (UMRG-RP044C-17HNE) and Grant "Bantuan Kecil" BKS080-2017.

REFERENCES

[1] S. Gauch, M. Speretta, A. Chandramouli, and A. Micarelli, "User profiles for personalized information access," in *The Adaptive Web*. Berlin, Germany: Springer, 2007, pp. 54–89.
[2] S. Kanoje, S. Girase, and D. Mukhopadhyay, "User profiling trends, techniques and applications," Mar. 2015, *arXiv:1503.07474*. [Online]. Available: https://arxiv.org/abs/1503.07474
[3] S. Alaoui, Y. EL Bouzekri EL Idrissi, and R. Ajhoun, "Building rich user profile based on intentional perspective," *Procedia Comput. Sci.*, vol. 73, pp. 342–349, Jan. 2015.
[4] S. E. Middleton, N. R. Shadbolt, and D. C. De Roure, "Ontological user profiling in recommender systems," *ACM Trans. Inf. Syst.*, vol. 22, no. 1, pp. 54–88, Jan. 2004.
[5] G. I. Webb, M. J. Pazzani, and D. Billsus, "Machine learning for user modeling," *User Model. User-Adapted Interact.*, vol. 11, nos. 1–2, pp. 19–29, Mar. 2001.
[6] B. Chikhaoui, S. Wang, T. Xiong, and H. Pigot, "Pattern-based causal relationships discovery from event sequences for modeling behavioral user profile in ubiquitous environments," *Inf. Sci.*, vol. 285, pp. 204–222, Nov. 2014.
[7] G. Farnadi, J. Tang, M. De Cock, and M.-F. Moens, "User profiling through deep multimodal fusion," in *Proc. 11th ACM Int. Conf. Web Search Data Mining*, Feb. 2018, pp. 171–179.
[8] S. Ouaftouh, A. Zellou, and A. Idri, "User profile model: A user dimension based classification," in *Proc. 10th Int. Conf. Intell. Syst., Theories Appl.*, Oct. 2015, pp. 1–5.
[9] Y. El Allioui and O. El Beqqali, "User profile Ontology for the Personalization approach," *Int. J. Comput. Appl.*, vol. 41, no. 4, pp. 31–40, Jan. 2012.
[10] S. Calegari and G. Pasi, "Ontology-based information behaviour to improve Web search," *Future Internet*, vol. 2, no. 4, pp. 533–558, 2010.
[11] D. Godoy and A. Amandi, "User profiling in personal information agents: A survey," *Knowl. Eng. Rev.*, vol. 20, no. 4, pp. 329–361, 2005.
[12] M. Chen and A. A. Ghorbani, "A survey on user profiling model for anomaly detection in cyberspace," *J. Cyber Secur. Mobility*, vol. 8, no. 1, pp. 75–112, 2019.
[13] I. Dickinson, D. Reynolds, D. Banks, S. Cayzer, and P. Vora, "User profiling with privacy: A framework for adaptive information agents," in *Intelligent Information Agents*. Berlin, Germany: Springer, 2003, pp. 123–151.
[14] H. J. Pandit and D. Lewis, "Ease and ethics of user profiling in black mirror," in *Proc. Companion Web Conf.*, Apr. 2018, pp. 1577–1583.
[15] E. Stamatatos, "A survey of modern authorship attribution methods," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 60, no. 3, pp. 538–556, 2009.
[16] M. Mezghani, C. A. Zayani, I. Amous, and F. Gargouri, "A user profile modelling using social annotations: A survey," in *Proc. 21st Int. Conf. World Wide Web*, Apr. 2012, pp. 969–976.
[17] A. Abdel-Hafez and Y. Xu, "A survey of user modelling in social media websites," *Comput. Inf. Sci.*, vol. 6, no. 4, pp. 59–71, 2013.
[18] J. Peng, K.-K. R. Choo, and H. Ashman, "User profiling in intrusion detection: A review," *J. Netw. Comput. Appl.*, vol. 72, pp. 14–27, Sep. 2016.
[19] L. M. de Campos, J. M. Fernández-Luna, J. F. Huete, and E. Vicente-Lopez, "Using personalization to improve XML retrieval," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 5, pp. 1280–1292, May 2014.
[20] N. Craswell, A. P. de Vries, and I. Soboroff, "Overview of the TREC 2005 enterprise track," in *Proc. Trec*, vol. 5, 2005, pp. 1–7.
[21] S. Liang and M. de Rijke, "Formal language models for finding groups of experts," *Inf. Process. Manage.*, vol. 52, no. 4, pp. 529–549, Jul. 2016.

[22] K. Balog, T. Bogers, L. Azzopardi, M. De Rijke, and A. Van Den Bosch, "Broad expertise retrieval in sparse data environments," in *Proc. 30th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jul. 2007, pp. 551–558.

[23] C. Jang, H. Chang, H. Ahn, Y. Kang, and E. Choi, "Profile for effective service management on mobile cloud computing," in *Advanced Communication and Networking*. Berlin, Germany: Springer, 2011, pp. 139–145.

[24] Y. Yang, "Web user behavioral profiling for user identification," *Decis. Support Syst.*, vol. 49, no. 3, pp. 261–271, Jun. 2010.

[25] X. Tao, Y. Li, and N. Zhong, "A personalized ontology model for Web information gathering," *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 4, pp. 496–511, Apr. 2010.

[26] K. Sugiyama, K. Hatano, and M. Yoshikawa, "Adaptive Web search based on user profile constructed without any effort from users," in *Proc. 13th Int. Conf. World Wide*, May 2004, pp. 675–684.

[27] A. Farseev, M. Akbari, I. Samborskii, and T.-S. Chua, "'360° user profiling: Past, future, and applications' by Aleksandr Farseev, Mohammad Akbari, Ivan Samborskii and Tat-Seng Chua with Martin Vesely as coordinator," *ACM SIGWEB Newsletter*, no. 4, 2016.

[28] D. Poo, B. Chng, and J.-M. Goh, "A hybrid approach for user profiling," in *Proc. 36th Annu. Hawaii Int. Conf. Syst. Sci.*, Jan. 2003, p. 9.

[29] X. Song, L. Nie, L. Zhang, M. Akbari, and T.-S. Chua, "Multiple social network learning and its application in volunteerism tendency prediction," in *Proc. 38th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Aug. 2015, pp. 213–222.

[30] A. Farseev, L. Nie, M. Akbari, and T.-S. Chua, "Harvesting multiple sources for user profile learning: A big data study," in *Proc. 5th ACM Int. Conf. Multimedia Retr.*, Jun. 2015, pp. 235–242.

[31] M. Akbari, X. Hu, N. Liqiang, and T.-S. Chua, "From tweets to wellness: Wellness event detection from Twitter streams," in *Proc. 13th AAAI Conf. Artif. Intell.*, Feb. 2016, pp. 87–93.

[32] S. Liang, X. Zhang, Z. Ren, and E. Kanoulas, "Dynamic embeddings for user profiling in Twitter," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2018, pp. 1764–1773.

[33] M. Gao, K. Liu, and Z. Wu, "Personalisation in Web computing and informatics: Theories, techniques, applications, and future research," *Inf. Syst. Frontiers*, vol. 12, no. 5, pp. 607–629, Nov. 2010.

[34] K. Hoashi, K. Matsumoto, N. Inoue, and K. Hashimoto, "Document filtering method using non-relevant information profile," in *Proc. 23rd Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jul. 2000, pp. 176–183.

[35] D. H. Widyantoro, T. R. Ioerger, and J. Yen, "Learning user interest dynamics with a three-descriptor representation," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 52, no. 3, pp. 212–225, 2001.

[36] T. Raghu, P. Kannan, H. R. Rao, and A. B. Whinston, "Dynamic profiling of consumers for customized offerings over the Internet: A model and analysis," *Decis. Support Syst.*, vol. 32, no. 2, pp. 117–134, Dec. 2001.

[37] G. Adomavicius and A. Tuzhilin, "Expert-driven validation of rule-based user models in personalization applications," *Data Mining Knowl. Discovery*, vol. 5, nos. 1–2, pp. 33–58, Jan. 2001.

[38] D. Kelly and J. Teevan, "Implicit feedback for inferring user preference: A bibliography," *SIGIR Forum*, vol. 37, no. 2, pp. 18–28, 2003.

[39] B. Mobasher, "Data mining for Web personalization," in *The Adaptive Web*. Berlin, Germany: Springer, 2007, pp. 90–135.

[40] M. Amoretti, L. Belli, and F. Zanichelli, "UTravel: Smart mobility with a novel user profiling and recommendation approach," (in English), *Pervasive Mobile Comput.*, vol. 38, pp. 474–489, Jul. 2017.

[41] K. Lakiotaki, N. F. Matsatsinis, and A. Tsoukias, "Multicriteria user modeling in recommender systems," *IEEE Intell. Syst.*, vol. 26, no. 2, pp. 64–76, Mar./Apr. 2011.

[42] J. Chen, Y. Liu, and M. Zou, "Home location profiling for users in social media," *Inf. Manage.*, vol. 53, no. 1, pp. 135–143, Jan. 2016.

[43] J. Tang, L. Yao, D. Zhang, and J. Zhang, "A combination approach to Web user profiling," *ACM Trans. Knowl. Discovery Data*, vol. 5, no. 1, p. 2, Dec. 2010.

[44] J. Li, A. Ritter, and E. H. Hovy, "Weakly supervised user profile extraction from Twitter," in *Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics* vol. 1, 2014, pp. 165–174.

[45] A. Mislove, B. Viswanath, K. P. Gummadi, and P. Druschel, "You are who you know: Inferring user profiles in online social networks," in *Proc. 3rd ACM Int. Conf. Web Search Data Mining*, Feb. 2010, pp. 251–260.

[46] Y. Hijikata, "Implicit user profiling for on demand relevance feedback," in *Proc. 9th Int. Conf. Intell. User Interfaces*, Jan. 2004, pp. 198–205.

[47] G. Jawaheer, P. Weller, and P. Kostkova, "Modeling user preferences in recommender systems: A classification framework for explicit and implicit user feedback," *ACM Trans. Interact. Intell. Syst.*, vol. 4, no. 2, p. 8, Jul. 2014.

[48] Q. F. Ying, D. M. Chiu, S. Venkatramanan, and X. Zhang, "User modeling and usage profiling based on temporal posting behavior in OSNs," *Online Social Netw. Media*, vol. 8, pp. 32–41, Dec. 2018.

[49] D. Jurgens, "That's what friends are for: Inferring location in online social media platforms based on social relationships," in *Proc. 7th Int. AAAI Conf. Weblogs Social Media*, vol. 13, 2013, pp. 273–282.

[50] I. Portugal, P. Alencar, and D. Cowan, "The use of machine learning algorithms in recommender systems: A systematic review," *Expert Syst. Appl.*, vol. 97, pp. 205–227, May 2018.

[51] L. Fuyan, "An attribute selection approach and its application," in *Proc. Int. Conf. Neural Netw. Brain*, vol. 2, Oct. 2005, pp. 636–640.

[52] M. J. Pazzani, "A framework for collaborative, content-based and demographic filtering," *Artif. Intell. Rev.*, vol. 13, nos. 5–6, pp. 393–408, Dec. 1999.

[53] S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, "Supervised machine learning: A review of classification techniques," *Emerg. Artif. Intell. Appl. Comput. Eng.*, vol. 160, pp. 3–24, 2007.

[54] K. Bradley, R. Rafter, and B. Smyth, "Case-based user profiling for content personalisation," in *Proc. Int. Conf. Adapt. Hypermedia Adapt. Web-Based Syst.* Berlin, Germany: Springer, 2000, pp. 62–72.

[55] R. Rafter, K. Bradley, and B. Smyth, "Personalised retrieval for online recruitment services," in *Proc. BCS/IRSG 22nd Annu. Colloq. Inf. Retr.*, Cambridge, U.K., 2000, pp. 5–7.

[56] Z. Shi, Y. Huang, and S. Zhang, "Fisher score based naive Bayesian classifier," in *Proc. Int. Conf. Neural Netw. Brain*, vol. 3, Oct. 2005, pp. 1616–1621.

[57] Z. Xie and Q. Zhang, "A study of selective neighborhood-based naïve Bayes for efficient lazy learning," in *Proc. 16th IEEE Int. Conf. Tools Artif. Intell.*, Nov. 2004, pp. 758–760.

[58] K. Nyberg, T. Raiko, T. Tiinanen, and E. Hyvönen, "Document classification utilising ontologies and relations between documents," in *Proc. 8th Workshop Mining Learn. Graphs*, Jul. 2010, pp. 86–93.

[59] M. E. Celebi and K. Aydin, *Unsupervised Learning Algorithms*. Berlin, Germany: Springer, 2016.

[60] I. F. Moawad, H. Talha, E. Hosny, and M. Hashim, "Agent-based Web search personalization approach using dynamic user profile," *Egyptian Inform. J.*, vol. 13, no. 3, pp. 191–198, Nov. 2012.

[61] J.-W. Ahn, P. Brusilovsky, J. Grady, D. He, and S. Y. Syn, "Open user profiles for adaptive news systems: Help or harm?" in *Proc. 16th Int. Conf. World Wide Web*, May 2007, pp. 11–20.

[62] I.-T. Nebel, B. Smith, and R. Paschke, "A user profiling component with the aid of user ontologies," in *Proc. Workshop Learn.-Teach.-Knowl.-Adaptivity*, Karlsruhe, Germany, 2003, pp. 1–5.

[63] L. Han and G. Chen, "A fuzzy clustering method of construction of ontology-based user profiles," *Adv. Eng. Softw.*, vol. 40, no. 7, pp. 535–540, Jul. 2009.

[64] P. Giaretta and N. Guarino, "Ontologies and knowledge bases towards a terminological clarification," *Towards Very Large Knowl. Bases: Knowl. Building Knowl. Sharing*, vol. 25, no. 32, pp. 307–317, 1995.

[65] L. Han, G. Chen, and M. Li, "A method for the acquisition of ontology-based user profiles," *Adv. Eng. Softw.*, vol. 65, pp. 132–137, Nov. 2013.

[66] O. Choi and S. Y. Han, "Personalization of rule-based Web services," *Sensors*, vol. 8, no. 4, pp. 2424–2435, 2008.

[67] Y.-J. Park and K.-N. Chang, "Individual and group behavior-based customer profile model for personalized product recommendation," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 1932–1939, Mar. 2009.

[68] A. Cufoglu, "User profiling-a short review," *Int. J. Comput. Appl.*, vol. 108, no. 3, p. 8887, 2014.

[69] M. Nilashi, D. Jannach, O. B. Ibrahim, and N. Ithnin, "Clustering- and regression-based multi-criteria collaborative filtering with incremental updates," *Inf. Sci.*, vol. 293, pp. 235–250, Feb. 2015.

[70] M. Nilashi, O. Ibrahi, and K. Bagherifard, "A recommender system based on collaborative filtering using Ontology and dimensionality reduction techniques," *Expert Syst. Appl.*, vol. 92, pp. 507–520, Feb. 2018.

[71] G. Linden, B. Smith, and J. York, "Amazon. com recommendations: Item-to-item collaborative filtering," *IEEE Internet Comput.*, vol. 7, no. 1, pp. 76–80, Jan. 2003.

[72] J. Chen, R. Nairn, L. Nelson, M. Bernstein, and E. Chi, "Short and tweet: Experiments on recommending content from information streams," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, Apr. 2010, pp. 1185–1194.
[73] M. Corney, G. Mohay, and A. Clark, "Detection of anomalies from user profiles generated from system logs," in *Proc. 9th Australas. Inf. Secur. Conf.*, vol. 116, Jan. 2011, pp. 23–32.
[74] R. Kaur, S. Singh, and H. Kumar, "AuthCom: Authorship verification and compromised account detection in online social networks using AHP-TOPSIS embedded profiling based technique," *Expert Syst. Appl.*, vol. 113, pp. 397–414, Dec. 2018.
[75] A. A. Barforoush, H. Shirazi, and H. Emami, "A new classification framework to evaluate the entity profiling on the Web: Past, present and future," *ACM Comput. Surv.*, vol. 50, no. 3, p. 39, Jun. 2017.
[76] M. Grčar, D. Mladenič, and M. Grobelnik, "User profiling for interest-focused browsing history," in *Proc. Workshop End User Aspects Semantic Web*, 2005, pp. 99–109.
[77] L. Li, Z. Yang, B. Wang, and M. Kitsuregawa, "Dynamic adaptation strategies for long-term and short-term user profile to personalize search," in *Advances in Data and Web Management*. Heraklion, Greece: Springer, 2007, pp. 228–240.
[78] J. Vassileva and J. Zhang, "Trust, reputation and user modeling," in *Proc. Int. Conf. User Modeling, Adaptation, Personalization*. Berlin, Germany: Springer, 2011, pp. 225–229.
[79] J. Zheng, B. Zhang, X. Yue, G. Zou, J. Ma, and K. Jiang, "Neighborhood-user profiling based on perception relationship in the micro-blog scenario," *J. Web Semantics*, vol. 34, pp. 13–26, Oct. 2015.
[80] Y.-C. Fan, Y.-C. Chen, K.-C. Tung, K.-C. Wu, and A. L. Chen, "A framework for enabling user preference profiling through Wi-Fi logs," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 3, pp. 592–603, Mar. 2015.
[81] E. Zhong, N. Liu, Y. Shi, and S. Rajan, "Building discriminative user profiles for large-scale content recommendation," in *Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2015, pp. 2277–2286.
[82] A. Hawalah and M. Fasli, "Dynamic user profiles for Web personalisation," *Expert Syst. Appl.*, vol. 42, no. 5, pp. 2547–2569, Apr. 2015.
[83] I. Nagy and R. Farkas, "Person attribute extraction from the textual parts of Web pages," *Acta Cybern.*, vol. 20, no. 3, pp. 419–440, 2012.

CHRISTOPHER IFEANYI EKE received the B.Sc. degree in computer science from Ebonyi State University, Nigeria, and the M.Sc. degree in mobile computing from the University of Bedfordshire, Luton, U.K. He is currently pursuing the Ph.D. degree with the Department of Information Systems, Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia. His research interests include data science, NLP, information systems and security, cloud computing, machine learning, big data, and social media analytics.

AZAH ANIR NORMAN received the Ph.D. degree in information systems security. She was a Security Consultant with MSC Trustgate.com (subsidiary body of MDEC Malaysia) a certification authority in Malaysia for more than four years. She actively supervises many students at all levels of study from the bachelor's up to master's and Ph.D. degrees supervisions. She is currently a Lecturer with the Department of Information Systems, Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur. She was awarded a few research grants which focused on social media security and cybersecurity practices. Her articles in the selected research area have been printed in local and international conferences and ISI/Scopus WOS journals. Her research interests include e-commerce security, information systems security management, security policies, and standards.

LIYANA SHUIB received the B.Sc. degree (Hons.) in computer science from Universiti Teknologi Malaysia, Skudai, Malaysia, the master's degree in information technology from Universiti Kebangsaan Malaysia, and the Ph.D. degree from the University of Malaya. She is currently pursuing the Ph.D. degree. She is also supervising several master's degree students. She is a Senior Lecturer with the Department of Information System, Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur. She has more than 20 articles relevant to her research interests.

HENRY FRIDAY NWEKE received the B.Sc. degree in computer science from Ebonyi State University, Nigeria, the M.Sc. degree in computer science from the University of Bedfordshire, U.K., and the Ph.D. degree from the Department of Information Systems, Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia. His research interests include big data, machine learning, deep learning, biomedical sensor analytics, human activity recognition, multi-sensor fusion, cloud computing, wireless sensor technologies, and emerging technology.