Extended User Profiling Approaches for Recommendation Systems

Fayzi Bekkamov
Library Information Systems
Tashkent University of Information Technologies named after Muhammad al-Khwarizmi
Tashkent, Uzbekistan
bfayzi93@gmail.com

Yashin Sharifov
Optical Communications Systems and Network Security
Karshi branch of the Tashkent University of Information Technologies named after Muhammad al-Khorezmi
Tashkent, Uzbekistan
sharifovyashin@gmail.com

Abstract. **Currently, the volume of data in information systems and the information needs of users are increasing. This causes information overload and a number of difficulties in finding the necessary information. Therefore, individual approaches, including personalization of user profiles, are important in solving this problem. Creating a user profile is relevant for obtaining information from systems in accordance with the needs of the user and for personalizing the services provided by the system. The lack of direct user profiling in information systems creates a number of problems in providing personalized services to users. Around the world, recent research has focused on developing systems that personalize user profiles based on their data sets. The work done so far on a global scale to create and model user profiles is analyzed. In this article, mathematical algorithms such as TF-IDF, Cosine similarity, Word2Vec are used to model and personalize user data. It also provides a classification scheme for user profiling, modeling and personalization. This classification scheme is based on three components. These are user data collection, user profiling, modeling and personalization components. Additionally, the article also mentions the benefits of user identification.**

Keywords: user profile, user profile modeling, personalization, information need, semantic connection, similarity, ontology.

I. INTRODUCTION

Currently, as a result of the increase in the volume of data in information systems, users experience information overload. This forces users to spend a lot of time searching for information that matches their information needs and creates demands for personalization approaches that match the user's information search profile. Personalization requires first creating and modeling user profiles. Typically used for user profiles, user data modeling, and personalization. User profiles reflect an individual's personal information, demographics (name, age, country, gender, language, profession, academic background, etc.), as well as their interests and preferences.

User profiling is widely used in various search engines, user identification, personalization, recommendations, intelligent learning systems, data filtering. User profile information varies depending on the information content in the information systems field. That is, the user's profile information is updated dynamically depending on his interests. When communicating with users, information systems should include the study of their behavior (topics of interest, social relationships, user evaluations, and goals) [1, 3].

As a result of taking into account the interests of users, user modeling becomes possible. User modeling is the process of collecting user browsing history data from open systems, creating user profiles, storing and testing systems. For example, recommendation-oriented information systems capture the characteristics of registered users, find similar users based on input data, and offer personalized information and services that ultimately satisfy users' information needs. There are two main approaches to user modeling [5, 12]:

- create an initial profile for a new user;
- keep profile information up to date in accordance with the constantly changing interests, preferences and information needs of users.

These two approaches must be fully implemented since user modeling is done based on profile data (based on the above approaches).

Based on user data, personalization is carried out as a result of creating a profile and modeling the user. Personalization mainly relies on two types of user profiles: static and dynamic profiles. Static personalization does not take into account the initial registration of users from information systems and the creation of a profile, and then the needs and interests of users. As a result, systems recommend irrelevant information and services to users. Because the information needs and interests of users change over time.

Print ISSN 1691-5402
Online ISSN 2256-070X
https://doi.org/10.17770/etr2024vol2.8052
© 2024 Fayzi Bekkamov, Yashin Sharifov. Published by Rezekne Academy of Technologies.
This is an open access article under the Creative Commons Attribution 4.0 International License.

Dynamic personalization overcomes the shortcomings of static personalization, that is, it constantly learns the interests and information needs of users and provides services to users based on their characteristics. Dynamic personalization uses clustering and classification algorithms to create user profiles.

II. MATERIALS AND METHODS

For profiling, modeling and personalization of users, a generalized classification scheme is created (Figure 1). This classification scheme is based on three components:

- Data collection;
- Creating and modeling user profiles;
- Personalization.

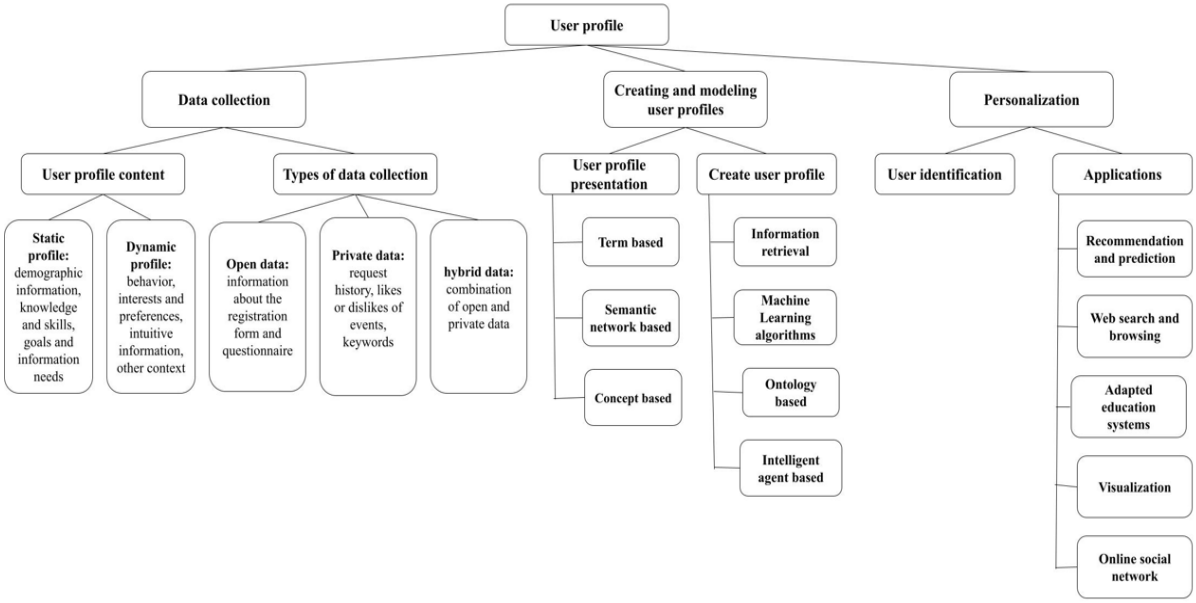

Figure 1. Classification scheme for user profiling, modeling and personalization

III. DATA COLLECTION

The main step in creating a user profile is collecting information about the user. Collecting information from users of information systems, studying their characteristics, information needs, and knowing their interests is an important task in creating personalized systems. Depending on the nature of the data, personalization is carried out. User profile data and data properties must be mutually exclusive [1, 4].

User profile content

A user profile contains various information, including the user's personal information, knowledge and skills, information needs, user goals, behavior, interests, preferences, etc. The following are several types of information for a user profile:

Demographic information. User demographic information includes first name, last name, middle name, country, language, age, gender, education, family members, social background. Demographic information is very important when creating a user profile.

Knowledge and skills. User knowledge and skills are important for student modeling in online learning systems. Students' knowledge is the main feature that determines the appropriate teaching system, and it is important to support them in the subject being taught, give appropriate instructions and adapt the lesson content. Today, some organizational systems take into account the knowledge and skills possessed by users and determine whether a suitable vacancy is available [2, 5].

Goals and information needs. User intent can be determined by their search history in the information systems and applications they use. Information needs are a situation when a user makes a request to search for information in systems (issues a request for an advanced search), with the goal of obtaining the necessary information from systems. Keywords are entered into the search by users, and ratings are recorded by the systems on the server. This is important to provide information and recommendations that match the user's profile and exactly what is needed.

Behavior. Behavior is one of the important pieces of information when creating a user profile. Because many artificial intelligence systems use three types of behavior to determine user behavior: reviews (opinions), ratings and psychological studies (tests, questions, questionnaires). Feedback (opinions) left by the user is one of the main indicators that determines his behavior. Popular social networking systems develop personalized recommendation content based on user feedback.

Interests and preferences. People's interests and preferences are important data when personalizing user profiles. Users' interests may be related to social connections, site names, page topics, sections of information systems, topics in social networks, groups. Also, people's preferences can be obtained from the history of e-commerce systems and from the history of viewing video content. Users' preferences for profile data are modeled based on their short-term and long-term activities. Modeling of long-term activity is mainly carried out in relation to the Google system, that is, created using Google feed themes. Short-term activity

models are created using the history of caches that users have clicked on in information systems.

Intuition data. Intuition plays a key role in user profiling, and data management is carried out without mathematical algorithms. Users' intuitions often shape initial perceptions of resources and services. When creating a user profile, their intuitive characteristics are determined. These signs are the choice of language, a history of continuous visions, the type of mental activity and the ability to think logically. Intuition helps users decide whether to trust a website or brand. It's like an internal compass that helps users navigate cyberspace. Intuition allows users to quickly make decisions, reason, and quickly sort and process information. Gut feelings help users connect emotionally with a resource and service. User profiling is not just about data, but also about understanding the person behind the profile.

Other contexts. It is important to describe the content of resources in information systems. The general and summary content of the resource is considered user profile information. Also, the location of users is important as contextual information in social networks. Because the location information of users can be accurately provided. In this case, source addresses and user locations are compared, and relatively similar information is presented in the form of a user profile feed. Social networks collect information about the psychological state of users, what they like or dislike (like/dislike), as well as emotional data. Some social networks have found that users' emotions directly influence how they use the system. All of this is important information for modeling the user profile in the context of the type. This type of information is then used to provide users with the information they need [6, 8].

Types of data collection

Data collection is carried out in normal, automatic and semi-automatic modes, respectively. The accuracy of user profile data depends on the amount of data generated as a result of user interaction with the system. In addition, the personalization stage reflects how, where and when to use the collected data. There are three types of data collection methods [4, 6]:

- open data;
- private data;
- hybrid data.

Open data. Data collection using open data is carried out directly in connection with the user himself. Registration of users from the information system, requests entering the system, pages divided into categories, assessments of system resources, storage of request history - data collection using an open method. For example, Amazon's system asks users to leave reviews and ratings for resources and services. Amazon then uses the information collected in this open manner to provide personalized recommendations to users. Open collected data is also used to determine user preferences in recommender systems. Some information systems use query keyword extraction, user reviews, and messages saved on social networking pages to express user interests. Data collected in an open manner is also static in that it remains valid until the user explicitly changes his or her interests and preferences. Additionally, this method is limited to asking users to indicate their interests using keywords.

Private data. Regarding private data, information about the user's interests is usually learned automatically. Online information systems automatically collect user data. Automatic data collection differs from static mode, in which system usage history is collected from system to system on a regular basis. By creating a user profile based on the information collected in private data, personalized search and recommendations are successfully implemented. Information is generated from personal data by tracking the user's profile browsing history, time spent on each page, and user history. Private data collection is currently supported by internet browsers and collects all user actions on the system server.

Hybrid data. Hybrid data is collected semi-automatically with limited user intervention. Information systems collect information from users based on a hybrid approach, using open and hidden methods. This approach is necessary for effective user profiling and timely and accurate data collection. A user profile is created by combining users' personal information, opinions, interests, history of queries entered into systems, and users' history in the system log. Data collection based on this method is carried out by combining open and private methods.

IV. CREATING AND MODELING USER PROFILES

User profile presentation. The second stage of the user profile is the creation and modeling of the user profile based on the collected data. The collected data is processed using mathematical algorithms and plays an important role in modeling the user profile. A user profile view is a collection of pre-collected data that reflects a user's interests. Modeling the user profile representation is carried out in three ways: term-based modeling, semantic network based and concept based modeling methods [8].

Term based. A user model based on keywords and terms generates digital vectors using the TF-IDF algorithm for texts in information systems. User input and saved query data is converted into a vector representation (Word2Vec) and then similarity algorithms (Cosinus, Jacquard, etc.) are discovered. Texts in information systems based on the user's request, that is, keywords, can be sorted by user profile information using TF-IDF algorithms. In this case, the TF-IDF algorithm evaluates to the value range [0,1].

$$TF_{td} = \frac{n_{td}}{\sum_k n_{td}} \quad (1)$$

Here, TF – number of frequencies of the word in the document, t – text, d – document.

$$IDF(w) = log\left(\frac{N}{DF_t}\right) \quad (2)$$

Here, $IDF(w)$ – number of documents containing words.

$$W_{t,d} = TF_{t,d} * log\left(\frac{N}{DF_t}\right) \quad (3)$$

After the query data entered by the user is converted to Word2Vec, the user profile is modeled using similarity algorithms. Cosine similarity is expressed in the following order.

$$Sim(u,v) = \frac{\sum_{i\in I}(R_{u,i})(R_{v,i})}{\sqrt[2]{\sum_{i\in I}(R_{u,i})^2}\sqrt{\sum_{i\in I}(R_{v,i})^2}} \quad (4)$$

The similarity of Jaccard is expressed in the following:

$$Sim(u,v) = \frac{|R_u|\cap|R_v|}{|R_u|\cup|R_v|} \quad (5)$$

Similarities are measured using these similarity algorithms. The TF-IDF algorithm analyzes requests to information system user profiles and provides customized data sequences.

Semantic network based. Users face problems of polysemy and synonyms when entering queries using keywords in information systems. To overcome these problems, user profiles are linked to the semantic web and each node in the network is weighted. An ontology must be built between the semantic web, user interests, and query history. For example, the WordNet system created an ontology-based user profile model based on user interests and queries. An ontology-based knowledge base of user profiles greatly simplifies working with the semantic web model.

Concept based. Concept-based user profile models are similar to the semantic web model, in which no ontology is built. In this model, semantic network nodes and relationships between nodes are important to the user's profile. Conceptual models are used to determine the user's level of interest in a topic. This model can also use TF-IDF algorithms to weight user profile queries.

Create user profile

There are two important concepts to consider when creating a user profile: first, what data sources users use for what purposes, and second, how to create a user profile using the selected data sources. Research shows that four concepts are important when creating user profiles today. These are data search, approaches based on machine learning algorithms, creation of ontologies, approaches based on intelligent agents. User profiles are created based on interests and preferences. Additionally, the four concepts above may not be fully utilized when creating user profiles. Profile information must be updated to ensure that the chosen construction method accurately reflects users' preferences.

Information retrieval. Information retrieval plays an important role in user profiling as it allows information systems to collect relevant information about user preferences, behavior, interests and demographics. Information retrieval helps collect information about user interactions, such as search queries, browsing history, and content preferences. By analyzing user interaction data, information systems can build a database of their goals and interests. This database allows for more specific recommendations and suggestions, tailored to users' preferences and goals after creating a profile. With user profiles, information systems can operate more efficiently by consistently providing relevant notifications, updates, and alerts based on users' interests and past interactions. In addition, information retrieval allows you to match resources based on user profiles. This improves the recommendations of relevant resources for individual users, which improves user satisfaction.

Machine Learning algorithms. Machine learning algorithms are widely used to classify documents in information systems into topics to create user profiles. The resulting topics are used as keywords and concepts to create user profiles. In information systems, decision trees, Bayesian classifiers, and K-nearest neighbors (KNN) machine learning algorithms are used to identify document topics by learning classifications from a set of documents matching a user's query. Machine learning algorithms KNN and Bayesian classifier play an important role in creating user profiles through efficient analysis and classification of data. KNN is a simple algorithm used for classification task. This algorithm is useful for creating user profiles and is based on the similarity of data points. When creating user profiles, KNN can be used to find similar users based on various characteristics and attributes. For example, given a dataset of user goals and behavior, KNN identifies users with similar preferences by measuring the distance between their feature vectors. When creating user profiles, a Bayesian classifier is used to classify users into different categories and segments based on their information and behavior. For example, in e-commerce platforms, Bayesian classifiers are used to segment users based on their purchasing history, browsing behavior, such as frequent shoppers, occasional shoppers, and non-buyers.

Ontology based. In user profiling, an ontology is a formal representation of knowledge in a specific domain of data representation. Ontologies provide a common framework for describing concepts and relationships that facilitate interaction between different systems and applications. Creating an ontology is useful in situations where user profiles need to be shared and integrated across multiple information systems. Creating an ontology to represent user preferences, interests, and behavior is essential for personalized user profiling recommendations. Ontology-based user profiles enable information systems to understand the context of user actions and provide tailored recommendations and services. Ontology creation is also useful for tasks such as identifying inconsistencies in user profiles, identifying latent user preferences, and predicting future behavior based on historical data. Ontology-based approaches play an important role in creating user profiles by providing a structural and semantic representation of user data, facilitating interaction, enabling personalization, supporting data integration, and enhancing reasoning and inference capabilities [8, 13, 15].

Intelligent agent based. In the context of user profiling, agent-based approaches play an important role. Intelligent agents collect information about users and create user profiles. These profiles capture users' preferences, behavior, and context. Using data mining and ontologies, profiles can be created with context updates. These updated profiles provide a personalized experience and a better understanding of user needs. By learning relevant contexts based on user behavior, intelligent agents dynamically adapt. Ontology segments represent these relevant contexts. The integration of intelligent agents, data mining and ontologies improves

overall user profiling performance. Intelligent agent approaches enable users to create more informed and context-sensitive user profiles, leading to more personalized services and interactions.

V. PERSONALIZATION

The third step will be to use the information in the user's profile to provide personalized services. Once a user profile is created, personalized recommendation systems are used to provide personalized services in search, social media, and various service delivery processes [10].

User identification. Accurate user identification is critical to any information system that creates profiles representing individual users. With personalization, usernames, email addresses, and passwords form the basis of user identification. This information provides access to personalized services and content. Age, gender, location, nationality and preferences help identify the user. Demographic data helps create initial user profiles. A user identification allows you to personalize behavior by associating behavior with specific people. Personalization, when information systems recognize a returning user, can tailor content and recommendations. For example, an e-commerce system greets the user by name when he logs in, suggests products based on past purchase history, meaning the system remembers the user's preferences. This process is performed using the user identification. User identification is the foundation of effective personalization.

Applications

Recommendation and prediction. Recommender systems and predictions are essential for personalization. Recommender systems provide users with personalized resources and services based on their preferences, behavior, and historical data. Users feel understood (observed) when information systems offer relevant resources (movies, books, products). For example, the Netflix system suggests shows and movies based on viewing history and user profile ratings. And Amazon recommends resources based on your browsing and purchase history. Predictive systems allow information systems to adapt to user profiles in real time. For example, predicting share prices based on stock market trends [7].

Web search and browsing. Searching and browsing play an important role in personalizing and shaping users' online experiences. When you enter a query into a search engine, it returns results tailored to the users' profile rather than generic results. However, remembering previous searches and visited websites helps the user understand your interests. It takes into account social interactions to better understand user preferences. In addition to search, website personalization customizes the entire site experience for each user. Personalized search is the process of searching a collection of documents and web pages based on the user's interests and preferences. User interest is generated by analyzing the websites that users search for.

Adapted education systems. Adapted education systems play an important role in improving personalized learning experiences. Adapted education systems organize individual lessons for students, taking into account their different levels of knowledge. They dynamically adjust the learning period based on each student's mastery of concepts. Adapted education systems do not follow a curriculum, these systems determine areas of learning based on individual needs, abilities and preferences [9, 14].

Visualization. In modern information systems, decision making is often based on data. Visualization allows us to understand complex information by representing it visually. Effective user profiling requires the creation of accurate user models. Visualization helps to visualize these patterns. Graphs and charts can display user attributes, preferences, and interactions with other people. The user profile reports personalization activities. And visualization helps to understand which features are important to users. Real-time visualization enables dynamic adaptation. As user behavior changes, customizing profiles and personalizing experiences becomes important. Visualization allows users to create meaningful user profiles, improve personalization, and deliver personalized experiences.

Online social network. Online social networks generate large amounts of data through user interactions, messages and connections. This information provides relevant information about the user's behavior, preferences and interests. Online social networks allow you to create complete user profiles. These profiles include individual characteristics such as demographic information, interests, and social connections. Analysis of social connections (friends, groups) helps improve user profiles [11].

VI. CONCLUSION

This article provides some information on the creation, modeling and personalization of the advertiser profile. A classification scheme for the user profile is given. Classification scheme organizers, data collection, user profiling, and personalization steps were analyzed. The first step of the classification scheme includes user profile content and data collection approaches. Information about static and dynamic profile views of user profile content has been provided. Including static profile data, demographic data, knowledge and skills, goals and information needs were analyzed. Dynamic profile data is divided into behavioral data, interests and goals, and other types of contextual data. Data collection approaches are divided into open data, closed data, and hybrid data. Registration form and survey data are open data, query history, event likes and dislikes, keywords are private data and the combination of open data and private data is hybrid data.

The second stage of the classification scheme involves presenting the user profile and creating a user profile. The user profile presentation included information about Term Based Models, TF-IDF Algorithm, Word2Vec Algorithm, Cosine and Jaccard Similarity Algorithms, Semantic Network Based Models, and Concept Based Models. While creating the user profile, detailed information was provided on data mining, machine learning algorithms, KNN algorithm,

Bayesian classifier, ontology based models, intelligent agent based models.

The third level of the classification scheme includes user identification, applications, recommendation and prediction systems, web search and browsing, personalized learning systems, visualization, and personalized private network.

REFERENCES

[1] Jianqiao Hu, Feng Jin, Guigang Zhang, Jian Wang, and Yi Yang. 2017. A User Profile Modeling Method Based on Word2Vec. IEEE International Conference on. IEEE, 410-414

[2] Jyoti Shokeen, Chhavi Rana, A study on features of social recommender systems. © Springer Nature B.V. 2019

[3] Madadipouya K., Chelliah S., A Literature Review on Recommender Systems Algorithms, Techniques and Evaluations. BRAIN: Broad Research in Artificial Intelligence and Neuroscience, ročník 8, č. 2, July 2017

[4] Rustamov A, Bekkamov F, Recommender systems: an overview, Scientific reports of Bukhara State University, 2021/3(85)

[5] Marat Rakhmatullaev, Sherbek Normatov, Fayzi Bekkamov, Fuzzy model for determining the information needs of library users, Environment. Technology. Resources. 14th international scientific and practical conference. June 15-16, 2023, Rezekne Academy Of Technologies, Rezekne, Latvia

[6] Elahi, Mehdi; Ricci, Francesco; Rubens, Neil. A survey of active learning in collaborative filtering recommender systems. Computer Science Review – via Elsevier 2016

[7] Marat Rakhmatullaev, Sherbek Normatov, Fayzi Bekkamov, Tavsiya etish tizimlarining umumiy tahlili, Raqamli Transformatsiya va Sun'iy Intellekt ilmiy jurnali, VOLUME 1, ISSUE 4, DECEMBER 2023.

[8] Susan Gauch, Mirco Speretta, Aravind Chandramouli, Alessandro Micarelli, User profiles for personalized information access. In The adaptive web. Springer, 2007

[9] Deborah L. Taylor, Michelle Yeung, A. Z. Bashet, Personalized and Adaptive Learning, Innovative Learning Environments in STEM Higher Education (pp.17-34), 2021

[10] Matthew Montebello, User profiling and Personalisation, Springer 2017

[11] Justin Gilbert, Suraya Hamid, Ibrahim Abaker Targio Hashem, Norjihan Abdul Ghani, Fatokun Faith Boluwatife, The rise of user profling in social media: review, challenges and future direction, Social Network Analysis and Mining Springer (2023)

[12] Sara Ouaftouh, Ahmed Zellou, Ali Idri, User profile model: a user dimension based classification, 2015 10th International Conference on Intelligent Systems: Theories and Applications (SITA), 10.1109/SITA.2015.7358378

[13] Y. Elallioui and O. El Beqqali, User profile Ontology for the Personalization approach, International Journal of Computer Applications, Volume 41, 2012

[14] Soulef Benhamdi, Abdesselam Babouri, Raja Chiky, Personalized recommender system for e-Learning environment, © Springer Science+Business Media New York 2016

[15] John K. Tarus, Zhendong Niu, Ghulam Mustafa, Knowledge-based recommendation: a review of ontology-based recommender systems for e-learning, © Springer Science+Business Media Dordrecht 2017