Received November 18, 2019, accepted December 18, 2019, date of publication December 23, 2019,
date of current version January 2, 2020.

Digital Object Identifier 10.1109/ACCESS.2019.2961769

User Behavior Analysis by Cross-Domain Log Data Fusion

YE TAO[1], SHUAITONG GUO[1], CAO SHI[1], AND DIANHUI CHU[2]
[1]School of Information Science and Technology, Qingdao University of Science and Technology, Qingdao 266061, China
[2]College of Computer Science and Technology, Harbin Institute of Technology, Weihai 264209, China

Corresponding author: Cao Shi (caoshi@yeah.net)

This work was supported in part by the National Key R&D Program of China under Grant 2018YFB1702902, and in part by the Shandong Province Colleges and Universities Young Talents Initiation Program under Grant 2019KJN047.

ABSTRACT Modern mass customization production allows user interaction activities to be distributed in the full product life cycle via multiple information systems. Investigating user behaviors across the boundaries of different domains helps to deeply integrate isolated fragmental profiles into a comprehensive one, and therefore can provide multi-dimensional, high-quality and valuable services. However, traditional user behavior analysis models are based on individual user profile information derived from separate domains, such as requirement analysis, design, supply chain, logistics, marketing, *etc.*, which have not considered the whole complexity of mass customization manufacturing. In this paper, we introduce the concept of multi-dimensional semantic activity space, where user behavior features are merged and represented as combined vectors. User behavior patterns are discovered by mining action data extracted from log files in different subsystems in the corresponding domains. We also identify distinct categories of user behaviors in various modules and subsystems in the context of an intelligent manufacturing environment. Experiment results show a strong indication that the proposed approach can be applied to reveal variations in typical behavioral aspects of cross-domain participants, in terms of patterns in resource access, operation tasks, performance assessment, *etc.*

INDEX TERMS Mass customization, manufacturing, semantic activity space, log analysis.

I. INTRODUCTION

In the modern manufacturing ecosystem, user behavior analysis plays an essential role in capturing user-specific information, expanding cognition and enhancing the ability to provide customized user experiences. Mass customization [1], referring to the capability to produce customized goods for individual end-users or a mass-market, has become one of the most crucial factors in business competition. To fulfill user requirements with high diversity and heterogeneity, it is necessary to classify customers into different categories and create a group or an individual user profile. Accurate and comprehensive profiles help to allocate the resources, make strategic choices, and provide personalized services. In contrast, a single-faceted, incomplete and ambiguous user profile, which does not reflect overall aspects, often becomes the cause of inefficient resource allocation and ambiguous confusion of user requirements.

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

Typical manufacturing platform consists of a variety of information management systems, *e.g.* manufacturing execution, supply chain management, resource planning customer relationship management, *etc.*, running in an open, dynamic, heterogeneous and cross-domain environment, as shown in **Fig.1**. These systems consist of multiple modalities, each of which has a different representation of the models, the scale of datasets and the types of operations. Multiple participants access these cross-border applications/services in different behavior patterns and have diverse information interests related to their roles and responsibilities. Each system depicts the user profile from specific topics, which can be potentially connected to a broader perception of user behavior.

The dramatic growth of information available to the manufacturing industry has made it easier to gather user data but also leads to information overload at the same time. However, due to the heterogeneous of disparate data sources, precisely profiling user behavior characteristics from multiple isolated systems has become difficult and time-consuming [2]. Therefore, it has brought opportunities and

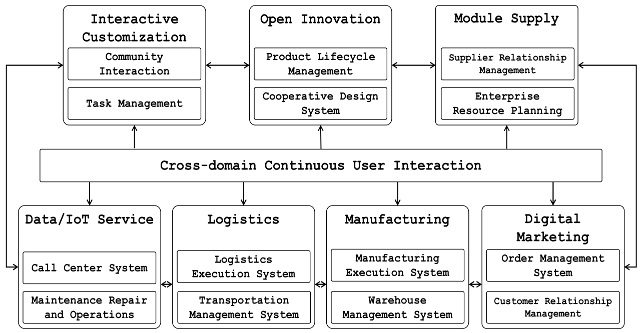


FIGURE 1. User interaction with cross-domain manufacturing platform.

challenges to dynamically discover and identify the interrelated linkage knowledge from multi-dimensional and heterogeneous datasets to figure out a full range of user behavior profiles.

Logs are excellent sources for building user behavior models [3], as it contains valuable information allowing us to observe and understand user behaviors when they are interacting with the target applications, systems or services. Log entries can be seen as time-ordered sequences of events that correspond to certain user activities. It can assist to discover the access patterns and form the individual or group user profiles to improve performance. Besides, logs of routine activities are also beneficial for identifying operational trends, detecting long-term system problems.

In this work, we present a fusion mechanism for determining a unified user behavior representation, by learning user action patterns and characteristics from heterogeneous information gathered by independent systems in multiple domains. After extracting and analyzing the log entries from various data sources in an intelligent manufacturing environment, user behavior features are mapped into a unified, multi-dimensional semantic activity space. Non-parametric clustering methods are applied for user classification.

The rest of this paper is structured as follows. Section II briefly introduces the related work in the fields of user behavior analysis and emerging research directions of log analysis. Section III details the concept of multi-dimensional semantic activity space used to profile users from various data sources. Section IV illustrates the possibility of applying typical non-parameter user classification algorithms to complete the classification by multi-dimensional vectors in the mentioned semantic space. Section V figures out the experiments and discusses the results. Lastly, Section VI presents the conclusions of the paper.

II. RELATED WORK

User behavior data fusion in this paper refers to identifying multiple representations of an individual user or a group of users, and mapping source data into target representation [4]. A user behavior model or profile is constructed to describe user characteristics by extracting diverse information interests and accumulated behavior preferences. Typical user profiles can be modeled via tags or categories [5]. Traditional classification and information retrieval approaches are helpful in the user profiling process [6], [7]. In recent years, ontological methods [8] have gained much popularity and importance to capture, represent and track user interests, and reported better profiling results compared with other methods [9]. Various user profiling studies have been successfully applied in the areas of e-commerce [10], social networks [11], tutorials [12] and healthcare [13]. However, few have concentrated on user behavior modeling techniques on multiple information management systems in the domain of intelligent manufacturing.

Furthermore, most of the researches analysis approaches restrict their scope in a single data source and therefore are not capable of obtaining comparatively complete descriptions of user interests in a comprehensive and distributed environment [14], [15]. Instead, considering the complexity of manufacturing value chains, we focus on an associated, correlated and refined user behavior profile by combining diversified representation from multiple sources of log datasets.

Log files are analyzed for different purposes, such as troubleshooting [16], usage statistics [17], security monitoring [18] and performance profiling [19]. Quite a few methodologies have been applied to deduce what the corresponding person is interested in, by analyzing user log file records [20]. For instance, log entries can be used to identify the typical browsing behavior of a user and subsequently to predict desired pages [21]. User potential interests can also be profiled by tracking logs [22], [23], where the activity information of user requests and page accesses are recorded. However, traditional log analysis methods have inevitable limitations in multiple sources of heterogeneous data fusion. In this paper, logs are recorded and generated throughout various domains in and out of an enterprise scope, distributed over different data sources and stored in various systems.

III. DATA FUSION MODEL

An approach of the data fusion model is described in this section. The modeling procedure consists of two stages: 1) extract features from log items generated by different systems; 2) map the features to a multi-dimensional semantic space. This task can be seen as a typical problem of combining evidence or data fusion.

Fig. 2 presents N different systems of product line in a classic manufacture enterprise environment, where f_{ij} denotes the j^{th} feature extracted from the log file in the i^{th} system, which is represented by S_i. The definition of f_{ij} depends on the specific application and the computation methods varies in S_i. For instance, some typical values may include the login count during a period, the view/submit times of an assignment, the scores/levels of some evaluations, *etc*.

Each f_{ij} contributes to some specific attributes that can be used to portrait a user in a high dimensional space, meanwhile, different f_{ij} may share similar semantic aspects of the user profile. For example, almost every system requires a login, and the frequency of browsing content or leaving comments in different systems often indicate the positivity of a certain user. Therefore, extracted features can be projected to a multi-dimensional activity space, as a result of capturing the implicit semantic relationship among them.

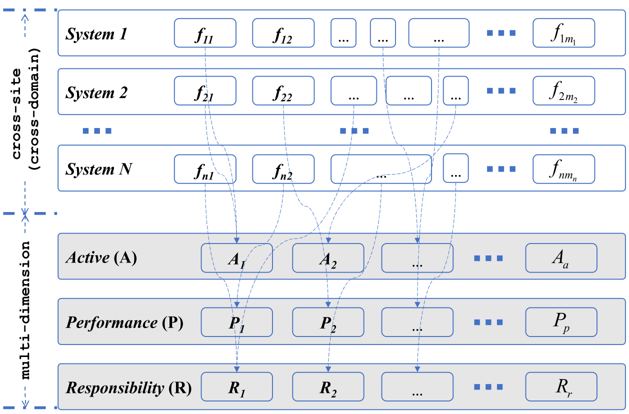


FIGURE 2. Features extracted from cross-domain systems.

The semantic user activity space is composed of two layers. In the first layer, the selected features from different systems form the dimension vector. And then, in the second layer, all dimension vectors are combined to represent the user characteristics. This is defined by the following equation:

$$U_i = (\xi_1 d_1, \cdots, \xi_n d_n), \quad d_i = \sum_{k=1}^{m} \varphi_{ik} f_{ik},$$

where k is the number of the involved cross-domain systems, ξ and φ are used to weight the importance of each component, which can be set based on experience in a simplified model.

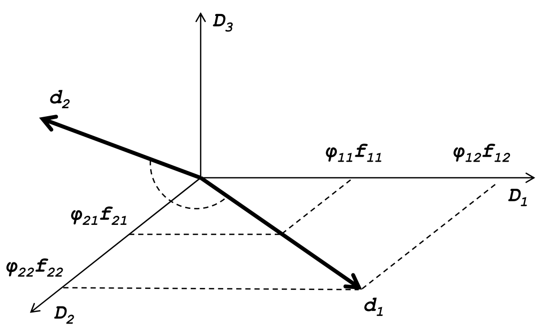


FIGURE 3. Mapping features to a multi-dimension semantic activity space.

Fig. 3 depicts a 3-dimensional semantic activity space model with 2 vectors (d_1 and d_2) representing for individual user profiles. Different from the traditional vector space model, in this paper, vectors are constructed by integrating selected isolated computed features from system log items. For example, if the values of f_{1i} $(i = 1, 2, 3, \ldots, K)$ refer to the login counts from K subsystems, then the combination of all components can be considered as a vector representation for activeness dimension (*e.g.* D_1). Using this approach, any user behavior can be computed and represented in a high-dimensional semantic activity space.

When projection from features to activity space is fulfilled, the comparison between users and user classification can be figured out using vector operations. User similarities can be calculated by comparing the deviation of angles between vectors in the user activity model. Given two sets of vectors which represent different user profiles, the following formula returns a value between zero and one. In other words, if two vector sets representing two users' behavior profiles, the results of the following formula fall into [0, 1], where 0 and 1 indicate the lowest and the highest degree of user similarity, respectively.

$$s\left(U_i, U_j\right) = U_i \cdot U_j / \|U_i\| \left\|U_j\right\|.$$

To exploit the proposed activity space for the benefit of the value chains in a classic manufacture enterprise, three dimensions are occupied for feature classification.

A. ACTIVENESS DIMENSION

The activeness dimension describes the contribution of user participation to all involved systems that the user can access. In general, the value of this degree reveals the stickiness of a certain user and can be derived by measuring how many times the user visits or interacts with specific systems via either web applications or mobile apps. For example, if a user visits the dashboard page of the system, or leaving comments on products, the corresponding log items can be included in the following calculations. Although each system could have its unique approach for user access control, most systems will keep track of the session data for further audit.

B. PERFORMANCE DIMENSION

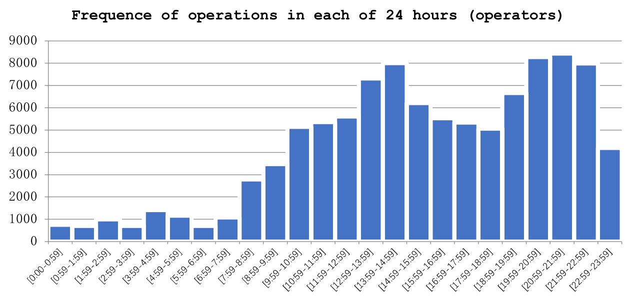
When it comes to intelligent manufacturing systems, complicated assessment methods should be examined to evaluate to report employee's effectiveness about their own goals or administration tasks. The procedure of this kind of measurement is involved in various parameters, such as KPI metrics, critical factors, and indicators, *etc*. There are enough reasons to utilize more efficient algorithms for processing information in log files. And it will result in a normalized range of scores for the performance dimension.

C. RESPONSIBILITY DIMENSION

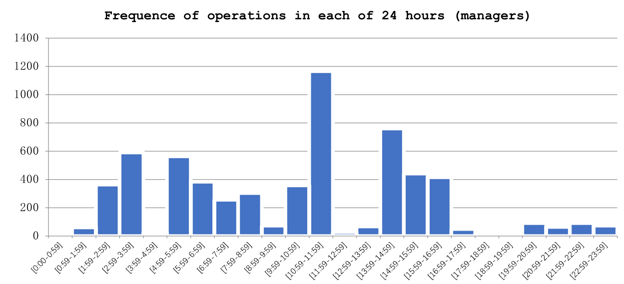
Since the objective of this research is to profile a user behavior model from individual sub-modules or systems crossing different value chains, hence the responsibilities of the target user will have a crucial influence on the final results. More specifically, some employees are responsible for the design (Product Lifecycle Management Systems), product manufacturing (Manufacturing Execution Systems), while others may focus their jobs on sales, financing or accounting (Enterprise Resource Planning).

IV. SAMPLE FEATURES AND CLUSTERING

In this section, we use the page views as a sample feature to illustrate the non-parametric profile clustering approach. By analyzing the log items, it is found that the user action patterns, in most modules or subsystems, have apparent diversity, especially for users with different responsibilities, *e.g.* operators and managers. As shown in **Fig. 4**, for operators, most page visits occur between 12h00 and 18h00, while for managers, there is a significant number of views in the late hours of the day. This is reasonable since most routine tasks are required to be completed during the working hours, while archiving jobs, statistics reports are regularly generated later in the day.



(a) count of page views for operators



(b) count of page views for managers

FIGURE 4. Distribution of page views against time for users with different responsibilities.

Without loss of generality, the Gaussian Mixture Model can be applied here as a non-parametric clustering method, to classify users into different categories according to some of the aforementioned statistical features. For example, given an m dimensional feature vector x the mixture density, the weighted likelihood function can be defined as:

$$p\left(x\,|\lambda\right) = \sum_{i=1}^{m} \omega_i p_i\left(x\right),$$

where every $p_i\left(x\right)$ represents a unimodal Gaussian density, and ω_i are the mixture weights for linear combination. Collectively, the parameters of the density model are denoted as:

$$\lambda = \{\omega_i, \mu_i, \sigma_i\},$$

where μ_i and σ_i are means and covariances.

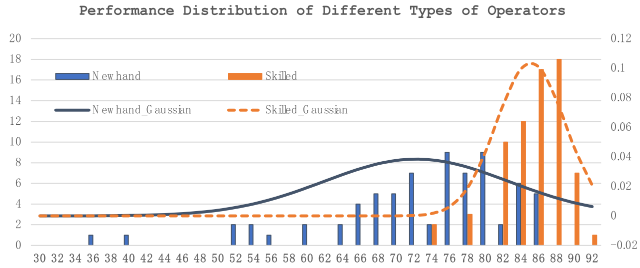


FIGURE 5. Performance distribution of skilled and new hand operators.

Similarly, obvious behavior pattern differences can also be found in the dimension of performance. As shown in **Fig. 5**, blue bars and yellow bars represent the performance distribution of novice operators and skilled users respectively. It indicates that skilled workers generally have higher work efficiency, on the contrary, the efficiency of novices is generally low. Therefore, using non-parametric clustering methods are expected to obtain higher accuracy of behavior pattern recognition.

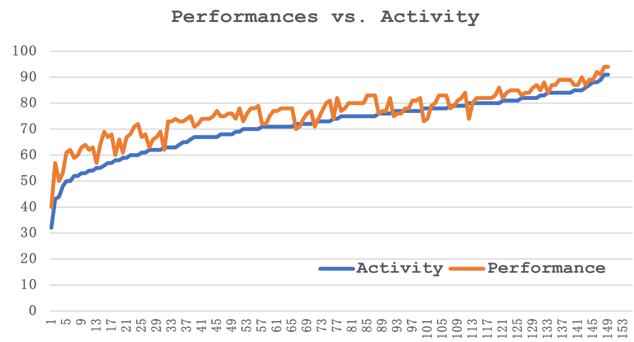


FIGURE 6. Correlation between page views and performance.

Another finding is the correlated relationship between the page views (activity dimension) and the performance indicators (performance dimension). **Fig. 6** clearly shows that with the continuous growth of the page view frequency, the corresponding performance shows a significant increasing trend. The normalized curve of performance generally fluctuates around the curve of page view in a small range.

V. RESULTS AND DISCUSSION

A. LOG PREPROCESSING

Log preprocess is a critical step in effective user behavior mining, including access log cleaning, session identification, and transformation of access log data to an appropriate format. Generally, data preparation needs to meet the requirements of the particular mining task. In this part, we have collected logs items from different modules deployed in a discrete manufacturing enterprise, where the user logs in with a single user ID and password to gain access to all related sub-systems. Once a user is defined with a unique identifier, it is possible to associate the visitor's activities across different systems to that unique ID. After log cleaning, session identification, and low support page filtering, we have collected more than 60000 log items generated by more than 130 users within a certain period.

Depending on the type of system an event can be composed of different information attributes. Usually, they contain timestamps, activity identifiers, and some key parameters. Specifically, Table 1 lists ten fields of the action that are recorded by data recording modules embedded in all server applications crossing the intelligent manufacturing platform. The log report allows us to learn which resources and activities in a module have been accessed, when, and by whom.

TABLE 1. Fields in raw log files.

Field	Value
request_session	The identification string of each session.
request_time	The timestamp of the log item.
remote_user	The user who invokes this activity and creates this log item.
server_module	The module that creates the log item.
server_url	The page that the user visited.
operation	The type of operation that the user has taken.
user_role	The responsibility of the target user.
server_target	The influenced users. (optional)
request_protocol	The initialization source of the action. (Web/App/Client/*etc.*)

An obvious problem of the raw logs is that the items are usually provided in a tabular format with the poor logical organization. And it may also contain missing/redundant lines, illegal values, and other noise data. Thus, preprocessing such as filtering and cleaning raw log files is critical.

Furthermore, as described in the above section, the collected raw user log is at a relatively fine granularity, which means that it is hard to discover the high-level semantic relationship. Therefore, some computational and statistical methods should be applied to get meaningful features. For example, login counts from all sub-systems in a certain period are calculated as an essential feature that contributes to the Activeness dimension. The ratio of qualified products, indicating the working efficiency, acts as an effective feature in the performance dimension.

Fig. 7 shows the trends of the log size with the increase of the users. The growth of the number of users also implies the passage of time. The main purpose of this experiment is to study how features and characteristics of log data change with time in heterogeneous systems. In a modern mass customization production environment, user behaviors often show a wide diversity of patterns in systems in different domains.

In domain A, as time goes on, it shows an approximately linear relationship between the number of users and the size of logs, indicating that the number of logs generated by each user is basically the same. For example, the logs generated by some proxy modules show similar characteristics like this. In domain B, in the early stage of the deployment, the number of logs surges sharply, but with the increase of users,

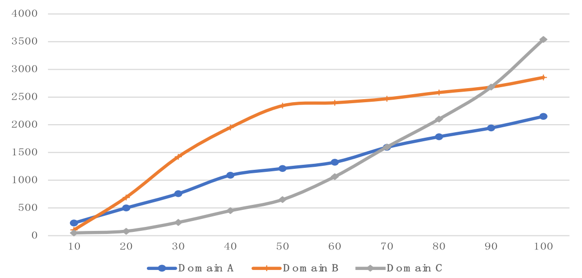


FIGURE 7. The scale of logs in different systems.

generated log item tends to be stable, and the growth rate slows down obviously. It suggests that there is an extreme increase in operations as soon as the system started, such as data migration or synchronization. In domain C, it has been found that with the increase of users, the number of logs shows a rapid growth, which indicates that the interaction activities among users made major contributions. Most modules with social network communication functionalities work like this.

B. RESULTS AND ANALYSIS

By observing the user behaviors in sub-systems, we classified all operations into 17 basic categories, which are indicated by the operation field in Table 1. To improve the readability, we illustrated typical 3 out of 17 operations for a specific user in **Fig. 8**.

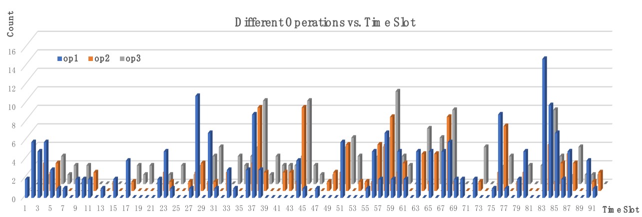


FIGURE 8. Different operation categories of a typical user in a certain period.

As shown, the *op1* (red bars) represents the basic routine activities such as login to the system and view pages. *op2* (yellow bars) denotes some submit operations, *e.g.* adding lines of supplier information, changing the status of an open task into resolved, *etc*. *op3* (green bars) shows statistical activities, *e.g.* generating daily/weekly/monthly reports, *etc*. It is reasonable that *op2* and *op3* are less frequently detected, as these types of operations do not occur every time the user accesses the system.

We plot users into the proposed semantic activity space to evaluate the effectiveness of classification, as shown in **Fig. 9**. The axes indicate the three dimensions defined as activity (A), responsibility (R), and performance (P). According to the corpus used in this experiment, 67% of the users are operators (red dots), 23% of them are managers (green blocks), and the rest are supervisors (purple triangles). It is found that those users with management responsibility have higher values in P and A.

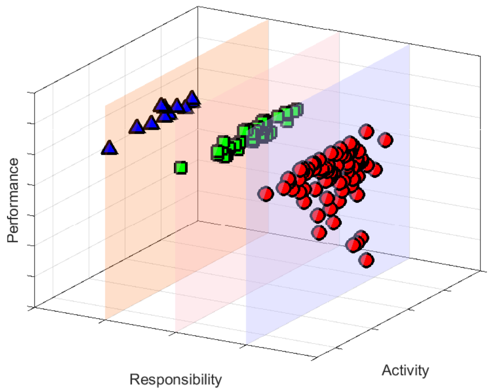


FIGURE 9. User classification via features from semantic activity space.

To further investigate the variance of user behaviors in different subsystems/modules, we studied the COSMOPlat [24] developed by Haier Corp. It is an internet-based platform where customers can create personalized orders for products. The full cycle of mass customization is divided into 7 main stages, which are implemented and supported by the corresponding subsystems, *i.e.* user interaction (D_1), R&D (D_2), supply chain management (D_3), intelligent manufacturing (D_4), e-commerce (D_5), logistics (D_6) and service (D_7).

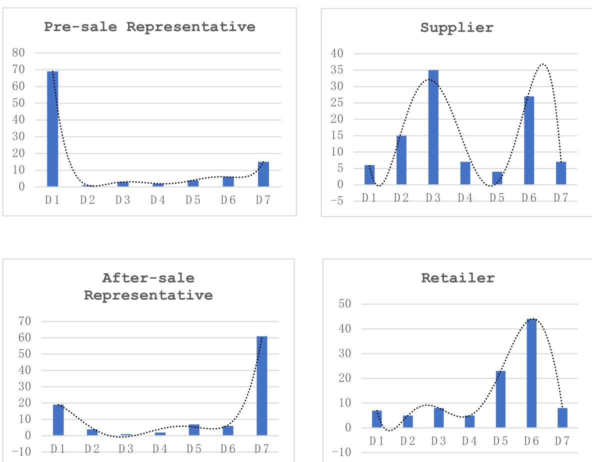


FIGURE 10. Typical user behavior logs statistics in different systems.

As shown in **Fig. 10**, four classes of typical users are chosen to investigate the number of logs they generated in different systems. Pre-sale and after-sale employees are responsible for collecting user requirements and feedbacks from social community modules (*e.g.* blog, group chats, *etc.*), therefore, their operations are focused on D_1 and D_7, respectively. Suppliers concentrate on querying demand and current logistics situation of products, parts, and assemblies, thus their interaction logs are mainly generated by D_3 and D_6. For distributors and online retailers, they are mainly responsible for operations and maintenance of e-commerce and logistics systems, so the major part of their interactions take place in D_5 and D_6.

VI. CONCLUSION

In this contribution, we introduced a novel approach for building user behavior profile models from distributed and heterogeneous information systems in a mass customization environment. Different sources of log data provide different behavior aspects of a target group of users, therefore combining and merging features from various domains helps to compensate for the shortcomings of incomplete behavior patterns computed from a single data source, and thus makes the target profile more adequate.

This paper focuses on modeling the semantic activity space, where similar features (*e.g.* login time, performance evaluation, user responsibilities) are mapped into an integrated and comprehensive user profile. We have completed some preliminary validation works of the proposed models and algorithms on a dataset exported from existing isolated systems (*e.g.* user interaction, supply chain management, e-commerce, logistics, *etc.*). And we found that it is possible to classify users and get more valuable new information according to their typical behavior patterns and some of the overall statistical indicators, through non-parameter classification methods.

However, there are still many problems that should be solved in the areas of multi-source heterogeneous user behavior data fusion. One possible further research direction is trying to apply more sophisticated and advanced prediction approaches to achieve better integration effects and improving customized API invocation efficiency based on user profiles.

REFERENCES

[1] S. Guo, T.-M. Choi, B. Shen, and S. Jung, "Inventory management in mass customization operations: A review," *IEEE Trans. Eng. Manag.*, vol. 66, no. 3, pp. 412–428, Aug. 2019, doi: 10.1109/TEM.2018.2839616.

[2] D. K. Srivastava, B. Roychoudhury, and H. V. Samalia, "Importance of user's profile attributes in identity matching across multiple online social networking sites," in *Proc. 8th Int. Conf. Cloud Comput., Data Sci. Eng. (Confluence)*, Jan. 2018, pp. 14–15, doi: 10.1109/CONFLUENCE.2018.8442455.

[3] H. Deng, J. S. Ally, J. Zhao, Q. Yao, C. Wang, and Q. Mat, "Modeling users behavior in cellular networks with mobile big data," in *Proc. 28th Wireless Opt. Commun. Conf. (WOCC)*, May 2019, pp. 1–5, doi: 10.1109/WOCC.2019.8770688.

[4] C. I. Eke, A. A. Norman, L. Shuib, and H. F. Nweke, "A survey of user profiling: State-of-the-art, challenges, and solutions," *IEEE Access*, vol. 7, pp. 144907–144924, 2019, doi: 10.1109/ACCESS.2019.2944243.

[5] M.-F. Moens, J. Li, and T.-S. Chua, *Mining User Generated Content*. London, U.K.: Chapman & Hall, 2014.

[6] M. Pennacchiotti and A.-M. Popescu, "A machine learning approach to Twitter user classification," in *Proc. ICWSM*, 2011, vol. 11, no. 1, pp. 281–288.

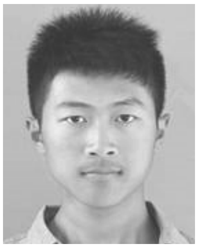
[7] P. Ristoski and H. Paulheim, "Semantic Web in data mining and knowledge discovery: A comprehensive survey," *J. Web Semantics*, vol. 36, pp. 1–22, Jan. 2016.

[8] A. Hawalah and M. Fasli, "Dynamic user profiles for Web personalisation," *Expert Syst. Appl.*, vol. 42, no. 5, pp. 2547–2569, Apr. 2015.

[9] C. Martinez-Cruz, C. Porcel, J. Bernabé-Moreno, and E. Herrera-Viedma, "A model to represent users trust in recommender systems using ontologies and fuzzy linguistic modeling," *Inf. Sci.*, vol. 311, pp. 102–118, Aug. 2015.

[10] S. Hernández, P. Álvarez, J. Fabra, and J. Ezpeleta, "Analysis of users' behavior in structured e-commerce Websites," *IEEE Access*, vol. 5, pp. 11941–11958, 2017.
[11] R. Tinati, L. Carr, W. Hall, and J. Bentwood, "Identifying communicator roles in Twitter," in *Proc. 21st Int. Conf. World Wide Web*, Apr. 2012, pp. 1161–1168.
[12] J. E. Greer and G. I. McCalla, *Student Modelling: The Key to Individualized Knowledge-Based Instruction*, vol. 125. Springer, 2013.
[13] M. Rabbi, M. H. Aung, M. Zhang, and T. Choudhury, "MyBehavior: Automatic personalized health feedback from user behaviors and preferences using smartphones," in *Proc. ACM Int. Joint Conf. Pervas. Ubiquitous Comput.*, 2015, pp. 707–718.
[14] Y. Ma, Y. Zeng, X. Ren, and N. Zhong, "User interests modeling based on multi-source personal information fusion and semantic reasoning," in *Proc. Int. Conf. Act. Media Technol.* Berlin, Germany: Springer, 2011, pp. 195–205.
[15] P. Bhargava, O. Brdiczka, and M. Roberts, "Unsupervised modeling of users' interests from their Facebook profiles and activities," in *Proc. 20th Int. Conf. Intell. User Interfaces*, 2015, pp. 191–201.
[16] A. R. Chen, "An empirical study on leveraging logs for debugging production failures," in *Proc. IEEE/ACM 41st Int. Conf. Softw. Eng., Companion (ICSE-Companion)*, May 2019, pp. 126–128, doi: 10.1109/ICSE-Companion.2019.00055.
[17] M. Kumar and Meenu, "Analysis of visitor's behavior from Web log using Web log expert tool," in *Proc. Int. Conf. Electron., Commun. Aerosp. Technol.(ICECA)*, vol. 2, Apr. 2017, pp. 296–301. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8212820/authors#authors, doi: 10.1109/ICECA.2017.8212820.
[18] V. P. Mishra, J. Dsouza, and L. Elizabeth, "Analysis and comparison of process mining algorithms with application of process mining in intrusion detection system," in *Proc. 7th Int. Conf. Rel., Infocom Technol. Optim. (Trends Future Directions) (ICRITO)*, Aug. 2018, pp. 613–617, doi: 10.1109/ICRITO.2018.8748748.
[19] Y. Zhao, X. Wang, H. Xiao, and X. Chi, "Improvement of the log pattern extracting algorithm using text similarity," in *Proc. IEEE Int. Parallel Distrib. Process. Symp. Workshops (IPDPSW)*, May 2018, pp. 507–514, doi: 10.1109/IPDPSW.2018.00087.
[20] A. Kemmar, Y. Lebbah, and S. Loudni, "A constraint programming approach for Web log mining," *Int. J. Inf. Technol. Web Eng.*, vol. 11, no. 4, pp. 24–42, 2016.
[21] K. R. Suneetha and R. Krishnamoorthi, "Identifying user behavior by analyzing Web server access log file," *Int. J. Comput. Sci. Netw. Secur.*, vol. 9, no. 4, pp. 327–332, 2009.
[22] G. Neelima and S. Rodda, "Predicting user behavior through sessions using the Web log mining," in *Proc. Int. Conf. Adv. Hum. Mach. Interact. (HMI)*, Mar. 2016, pp. 1–5.
[23] I. H. Sarker, A. Colman, M. A. Kabir, and J. Han, "Phone call log as a context source to modeling individual user behavior," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput., Adjunct*, 2016, pp. 630–634.
[24] P. Li, Y. Cheng, and F. Tao, "A manufacturing services collaboration framework toward industrial Internet platforms," in *Proc. IEEE Int. Conf. Intell. Saf. Robot. (ISR)*, Aug. 2018, pp. 166–171, doi: 10.1109/IISR.2018.8535991.

YE TAO received the B.S. and Ph.D. degrees from the Department of Computer Science and Technology, Shandong University, in 2005 and 2009, respectively. He is currently an Associate Professor with the College of Information Science and Technology, Qingdao University of Science and Technology. His research interests include service computing and software engineering.



SHUAITONG GUO received the B.S. degree, in 2019. He is currently the master's degree with the Qingdao University of Science and Technology. His research interests include data fusion and data mining.

CAO SHI received Ph.D. degree from Central South University, in 2011. He is currently with the Qingdao University of Science and Technology. His research interests include natural language processing and image and video processing.

DIANHUI CHU is currently the Director of the College of the Computer Science and Technology, Harbin Institute of Technology. He has undertaken more than 20 projects, such as the National 863 Plan and the National Science and Technology Support Plan. His research interests include service computing, service engineering, manufacturing information engineering, and so on.