Behavioral Profiling for Employees using Social Media: A Case Study based on WeChat*

Xiaochun Ni, Shuai Zeng, Rui Qin, Juanjuan Li, Yong Yuan, *Senior member, IEEE*
The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China
Qingdao Academy of Intelligent Industries, Qingdao, China
Beijing Engineering Research Center of Intelligent Systems and Technology, Chinese Academy of Sciences, Beijing, China
{xiaochun.ni, shuai.zeng, rui.qin, juanjuan.li}@ia.ac.cn, yong.yuan@ia.ac.cn (Corresponding author)

Fei-Yue Wang, *Fellow, IEEE*
The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China
Qingdao Academy of Intelligent Industries, Qingdao, China
Research Center of Military, Computational Experiments and Parallel System, National University of Defense Technology, Changsha, China
feiyue.wang@ia.ac.cn

***Abstract*—With the continuing development and innovation of modern information and communication technologies in recent years, social media platforms, such as WeChat and Microblog, have been witnessed to play a key role in employee management for enterprises or organizations, enabling individual or groups of employees to express their viewpoints or report their works in a real-time fashion. The resulting Cyber-workspace in social media, which exists in parallel with employees' physical workspace, has the potential of greatly changing the forms and functions of the organizations, as well as the knowledge structures and behavioral patterns of employees, thus bringing unprecedented challenges to the existing research efforts on the management of organizational behavior. In this paper, we propose an approach of behavioral profiling for analyzing and better understanding employees, with the input of large amounts of real-time collected data generated by employees' daily reported works on social media. Our aim is to characterize the diversified behavior patterns of employees in an automatic and accurate fashion for organization management. We also validate our proposed method using a real-world dataset collected from WeChat, and the experimental results prove our analysis, as well as the effectiveness of our approach.**

Keywords—employee behavioral profiling; behavior analysis; social media

I. INTRODUCTION

With the continuing development and innovation of modern information and communication technologies, social networks have emerged to serve as a critical medium for information dissemination, diffusion of epidemics, and spread of specific behavioral patterns, typically by shared activities or similarities between individuals [1]. As such, social media platforms, e.g., WeChat and Microblog, have been witnessed to play a key role in employee management for enterprises or organizations, enabling individual or groups of employees to express their viewpoints or report their works in a real-time fashion. In literatures, social media played significant roles in social changes [2], and now has become a useful source of real-time user-generated contents [3]. Many enterprises and organizations have begun to use various forms of social media as key communication channels for a variety of organizational activities [4]. Social media has profoundly changed the forms and functions of the organizations, as well as the knowledge structures and behavioral patterns of employees, thus bringing both valuable opportunity and unprecedented challenges to the existing research efforts on the management of organizational behavior.

On one hand, information asymmetry always exists between supervisors and employees within the organizations. Traditional ways of organizational management prove to have insufficient capability in providing a convenient and effective channel for managers to keep informed about their employees' working state, thus possibly leading to a delay in tracking a task or piece of work. Employees, therefore, might "hitchhike" during work without supervision. In practice, employees typically communicate with each other, forming a social group, in which a social activity can be an action or behavior addressing (directly or indirectly) other people and soliciting a response from them [5]. In recent years, social media platforms prove to be able to provide good supports to facilitate continuously relational and dialogic models of such communication, as well as record voluminous data generated in communications [6]. As such, these online platforms have become the carrier of Web office, undertaking new supports for organizational management. Generally speaking, social media provides a real-time communication channel which is well-designed so that it is convenient and fast for employees to communicate with their colleagues or supervisors and report their work. Typically, working groups or teams can be easily established on social networking platforms. A working group creates a space in which its members communicate and collaborate with each other, and the information within the group is publicly known for each member. It is an

This work is partially supported by National Natural Science Foundation of China (71472174, 71702182, 61533019, 71232006, 61233001, 71402178, 61603381, 61702519) and the Early Career Development Award of SKLMCCS (Y6S9011F52).

environment of mutual supervision and authenticity. A social networking platform provides a real-time and trusted medium for organizational management, and the data from working groups is the basis for quantifying organizational behavior.

On the other hand, data basis is generally static with long updating cycle in traditional human resource management. Therefore, it is difficult for managers to collect information and take control of employee management in time. Also, it is difficult for an employee to prove how much work he/she accomplished for the past recent period. Secondly, due to the lack of quantitative analysis, managers have to rely on subjective judgments on employees' performances. For example, to evaluate employees' performances, a manager should determine a number of different evaluation variables and their respective weights and score for each employee by himself/herself. Even when the manager knows each employee well, which rarely happens in practice, it is still hard for him/her to avoid subjective judgments, as the factors, weights and scores are all determined based on the subjective perception.

With the prevalent use of social media within an organization, a lot of practical and dynamic employee behavior data has been generated, which provides a data base for the assessment of employees, so that we now can "let the data speak". The data generated by social media is real-time, unstructured, and relatively hard to deal with using the traditional human resource analysis method, and it needs to combine the data features to construct the corresponding matching model. That is, based on large amounts of real-time collected data generated by employees' daily reported works on social media, it automatically depicts the dynamic, real-time and accurate employee behavior label, and provides an objective, practical and scientific measurement tool for organization management, which is helpful to realize the effective allocation and rational utilization of the talent in the organization.

Person-organization (PO) fit is a topic having attracted many attentions from both scholars and managers during recent years. Kristof [7] presented a comprehensive definition and conceptual model of PO fit that incorporates supplementary as well as complementary perspectives on fit. These definitional and measurement issues frame a review of the existing literature, as well as provide the basis for specific research propositions and suggestions for managerial applications. Hoffman and Woehr [8] provided a meta-analytic review of the relationship between PO fit and behavioral criteria (job performance, organizational citizenship behaviors, and turnover). Moynihan and Pandey [9] proposed that PO fit shape turnover intention. Cho et al. [10] revealed that millennials perceived higher PO fit for a company with organizational policies supporting employees' social media use. Yu [11] investigated the mechanisms that explain why PO fit impacts organizational attraction. Venkatesh et al. [12] introduced three work outcomes (namely, extrinsic, social, and intrinsic) as determinants of PO and person–job (PJ) fit perceptions of new IT employees from a total rewards perspective. Zhao and Rosson [13] aimed at gaining an in-depth understanding of how and why people use Twitter - a popular micro-blogging tool and exploring microblog's potential impacts on informal communication at work. Huang et al. [14] developed and estimated a dynamic structural framework to analyze social media content creation and consumption behavior by employees within an enterprise. They focused in particular on blogging behavior by employees. However, there is limited research on PO fit based on the work-related information generated on social media. This motivates our research.

Internet use and cyberspace activities have created an overwhelming demand for the rapid development and application of cyber-physical-social systems (CPSS) [15]. Social computing can facilitate the evolution of social administration from traditional paradigms to an intensive, informational, intelligent and innovative model with real-time and complete responses [16]. Conforming this trend, we should make full use of large amounts of real-time collected data generated by employees' daily reported works on social media, transform the data into knowledge, integrate the methods and technologies such as machine learning and human-computer interaction, and indirectly affect the person's consciousness, and indirectly change behavior patterns, to achieve the desired control or management objectives [17]. WeChat is one of the most popular china-based online office platforms, and has been favored by more and more organizations. In this paper, we propose an approach of behavioral profiling for analyzing and better understanding employees, with the input of large amounts of real-time collected data generated by employees' daily reported works on social media. Our aim is to characterize the diversified behavior patterns of employees in an automatic and accurate fashion for organization management. Additionally, we use real-world dataset to validate our proposed method and the experimental results prove our analysis, as well as the effectiveness of our approach. It provides an objective, practical and scientific measurement tool for organizations, and brings new ideas of social-media-based organization management and innovation in the big data era.

The remainder of this paper is organized as follows. Section II introduces the WeChat robot platform. Section III presents the construction of skill knowledge graph. Section IV formulates PO fit model. Empirical studies can provide critical insight into understanding the characteristics of employee behavior and PO model, so we will discuss the practical application of our research in section V. Section VI concludes this paper.

II. THE WECHAT ROBOT PLATFORM

A. The Application background

With the recent rapid development of mobile Internet, WeChat, as an unimpeded information communication channel and convenient office collaborative platform, gets the favor of more and more organizations, which causes the new form and mode of organization operation, also makes the employee's knowledge structures and behavior patterns change greately. Employee generates a large number of daily work behavior data in the WeChat. Although the generated data is typically short, unstructured and fast-updating, it can accurately reflect the real-time work content of employees, as

well as personal skills development. The major difficulties in processing the data might be effective feature extraction from large amounts of unstructured data, as well as reduction from large amounts of data into knowledge in quantitative analysis of organizational behavior.

B. Analysis and Framework

Based on a framework that incorporates robotics and software-defined surrogates using the ACP-based parallel systems theory [18], this paper proposes the construction of WeChat robot platform, to realize employees' behavior data management. WeChat robot platform uses the focus crawler technology to realize the efficient collection of work related information generated on the WeChat platform. The availability and structured storage of data is realized based on domain related data correction, cleaning and extraction technologies; Skill knowledge database is constructed based on Synonym Clustering, Entity Extraction and other knowledge acquisition technologies. Comprehensive and effective data is the basis for quantifying the employees' behaviors. The WeChat robot platform provides a powerful support for employee behavioral profiling analysis based on open source knowledge and a large amount of real-time collected data generated by employees' daily reported works on social media. The platform architecture is shown in Fig. 1.

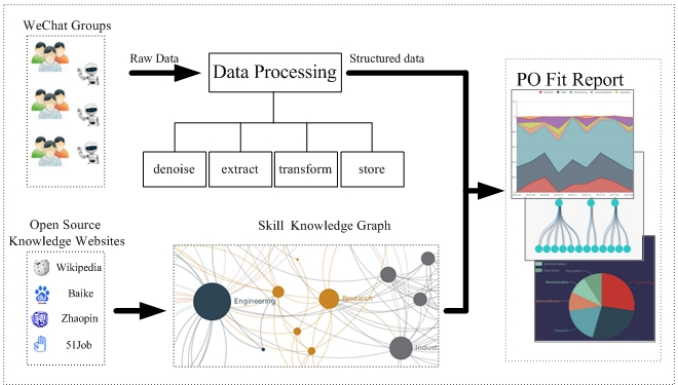

Fig. 1. WeChat robot platform architecture

III. CONSTRUCTION OF SKILL KNOWLEDGE GRAPH

Generally, a skill knowledge graph is a structure of nodes that represents skill-related entities and the relationships between them. Therefore, a social network is modeled as a directed graph $G\{V,E,\mathrm{P}\}$, where V represents the set of n nodes, E represents the set of m directed edges, and $\mathrm{P}: E(v_i,v_j)\rightarrow(0,1)$ is a probability function which associates a probability $p_{i,j}$ with each edge $e_{i,j}$ connecting two nodes v_i and v_j.

The architecture of construction is depicted in Fig. 2.

First, we use crawlers to collect skill-related data from open-source knowledge sites and store them within file systems. Then, we extract information from the data. The extracted results are used for generating skill knowledge graph as well as improving crawlers. To generate skill knowledge graph, we integrate multi-source information to gain entities and extract relationships between them. We also integrate words from the extracted results to obtain a skill-related lexicon. The details are described as follows.

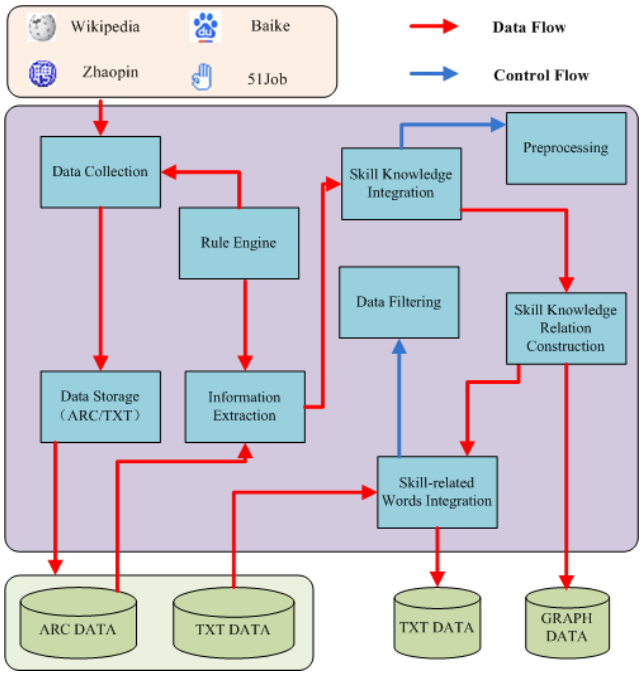

Fig. 2. The architecture of skill knowledge graph construction

A. Data source

We construct our skill knowledge graph from four websites: Wikipedia, Baike, Zhaopin, and 51Job.

Wikipedia and Baike are two typical online free knowledge bases, providing abundant structural knowledge items involving almost all fields. Note that each item consists of a name, a description and a list of names of related items. Zhaopin and 51Job are two famous recruiting websites in China. Each of them organizes jobs into a three-layer hierarchical architecture, which can be considered as multiple tree-like nets, which are called job trees in our paper. Every recruitment page, which consists of the name and the descriptions for a single job, is connected to a leaf node.

B. Data collection

We design and implement several crawlers to collect information from these websites. As the ranges of knowledge items on Wikipedia and Baike are too wide, we carefully select the skill-related pages and put them into the to-be-crawled list. Specifically, we firstly crawl recruitment pages on Zhaopin and 51Job, and extract keywords from all job names with a tool for word segmentation in Chinese. Then we search Wikipedia and Baike with these keywords to find domain-constrained knowledge, which are crawled finally.

C. Knowledge Integration

As a result of heterogeneity of the four knowledge resources, we have to integrate and combine the multi-source information and reduce the ambiguity. We assume that items with same knowledge/job name are the same entity. We use

the job trees of Zhaopin as the initial graph. Then we integrate items from the other three sites into the graph. An entity whose name is equal to an existing node is simply integrated with the node by appending descriptions and relationships for it. Those who are not existed in the graph yet are added as new nodes. Finally we gain a knowledge graph, including all the items and their relationships from the four sites.

D. Relation Construction

In the knowledge graph, each relationship between knowledge/jobs is assigned with a weight based on its type. If there are multiple relationships between two nodes, the average value will be chosen as the final integrated weight. Therefore we have a value between 0 and 1 for each edge, which is the associate probability $p_{i,j}$.

For any two node v_i and v_j in the knowledge graph, the associate probability is

$$p_{i,j} = p_{i,j_1} * p_{j_1,j_2} * \ldots p_{j_l,j} \tag{1}$$

where $v_{j_1}, v_{j_2} \ldots v_{j_l}$ are nodes along the shortest path from v_i to v_j.

IV. Person-Organization Fit Model

A. Relationships between words and knowledge

We split entities' names and descriptions using word segmentation and count word frequency. Let C be the set of words. Note that we defined that a word $c_k \in C$ is directly related to an entity $v_i \in V$ if and only if c_k is contained in the name or the description of v_i. Supposed that a word $c_k \in C$ is related to an entity $v_i \in V$, $w_{i,k}^1$ and $w_{i,k}^2$ are the corresponding word frequency of c_k in the name and the description of v_i, the integrated weight of c_i is computed as in the Equation (2).

$$w_{i,k} = R_1 * w_{i,k}^1 + R_2 * w_{i,k}^2 \tag{2}$$

Note that R_1 and R_2 are all constants, and $R_1 >> R_2$.

And the associate probability between c_k and v_i is

$$p'_{i,k} = \frac{w_{i,k}}{\sum_{c_i \in C} w_{i,k}} \tag{3}$$

B. Relationships between knowledge and positions

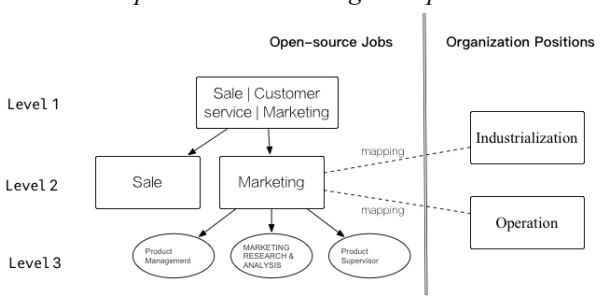

Fig. 3. Mapping skills to positions

In practice, there are different position systems for different organizations. Therefore we should map each intermediate node in skill-trees to at least one position in a specific organization based on expert knowledge as shown in Fig. 3.

Let $v(d)$ be the set of nodes mapping to a single position d, we assumed that each node $v_f \in v(d)$ is strongly connected to d, so that the associate probability $p''_{f,d}$ between v_f and d is 1. As for other nodes in the graph, the probability is

$$p''_{i,d} = \frac{\sum_{v_f \in v(d)} p_{i,f}}{|v(d)|}, v_i \in V, v_i \notin v(d) \tag{4}$$

C. Relationships between job logs and positions

Supposed that an employee's job logs are composed by a set of words $C = \{c_1, c_2, c_3, \ldots\}$. For each entity v_i that is directly related to c_k, we compute the associate probability $p'_{i,k}$ between c_k and v_i according to the Equation (3). So the associate probability between c_k and a position d is

$$p'''_{k,d} = \sum p'_{i,k} * p''_{i,d} \tag{5}$$

Therefore the associate probability between the job logs C and the position d is

$$p'''_{C,d} = p'''_{1,d} * p'''_{2,d} * p'''_{3,d} * \ldots \tag{6}$$

Choose the position with the highest associate probability, we can fit an engagement to a position of a specific organization.

V. An Empirical Study

Next, we conduct an empirical study to validate our proposed method using a real-world dataset collected by WeChat robot platform. A set of rules or policies are formulated to encourage and regulate the postings of group members, especially the daily work reporting [19].

A. Experimental Scenario and Dataset

The organization we chosen for the empirical study has more than 200 employees and each employee is assigned to one specific division. For each division, only one working group is allowed to create in the WeChat platform. An employee submits work related contents in his or her own group. Each employee should have an career anchor, and all those positions can be classified into seven types:

- Administration, focused on administrative affairs.
- Engineering, focused on project development.
- Research, focused on scientific research.
- R&D, focused on research and development.
- Industrialization, focused on marketing.
- Operation, focused on incubation and operation of enterprises.

- Management, focused on high-level management of the organization.

All positions are not strictly distinguishable, there might be overlaps between each other. The dataset contains more than 70 thousands data records collected from July 2016 to August 2017.

B. Analysis

In order to validate our model, we conduct an empirical study from three levels:

1) Analyzing behavioral profiling from a randomly-selected employee .

First, we choose one employee, analyze his/her work related data generated in WeChat of a week, and extract the key features as shown in Fig. 4.

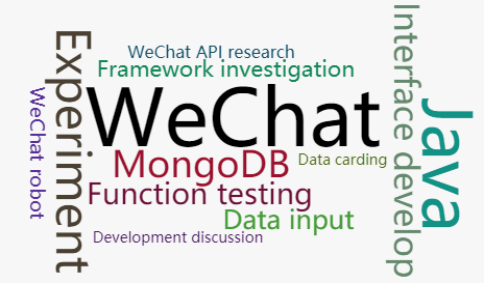

Fig. 4. The key features of one employee's work report in a week

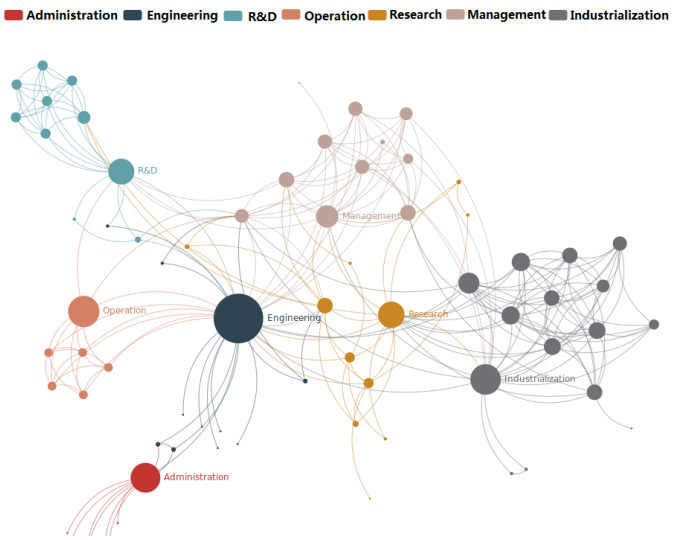

Fig. 5. Demonstration of skills mapping to positions. There are seven positions in our empirical study: Administration, Engineering, R&D, Operation, Research, Management and Industrialization

Combining the skill knowledge graph (Fig. 5) and PO fit model, we can formulate an employee behavioral profiling in a week (Fig. 6). It is intuitive to see that the skill used by this employee in this week's work tends to be consistent with the "Engineering". Managers can fully understand what work the employee focuses on within this week and which skills are mainly used by him/her.

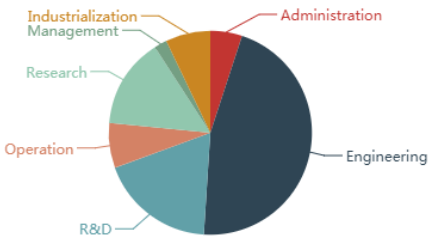

Fig. 6. Example of employee behavioral profiling in a week

From the other point of view, we can get a new behavioral profiling of the employee by analyzing the data in the past one year as shown in Fig. 7. This can clearly describe the change of the employee's skills and work focus.

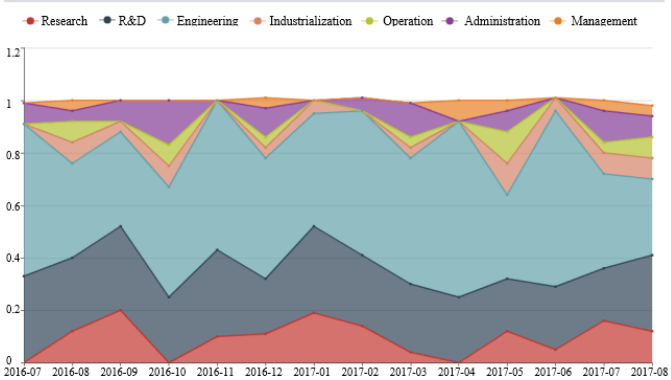

Fig. 7. Example of employee behavioral profiling in a year

2) Comparative analysis of employees from different divisions.

Then we choose another two employees from different divisions for comparison, one employee is from HR division, the other is from senior management division. The comparative analysis is shown in Fig. 8. We can see that the skills used in personnel management are well differentiated, "Administration" takes a large part of his/her work. However, The ability to perform managerial positions is equally balanced.

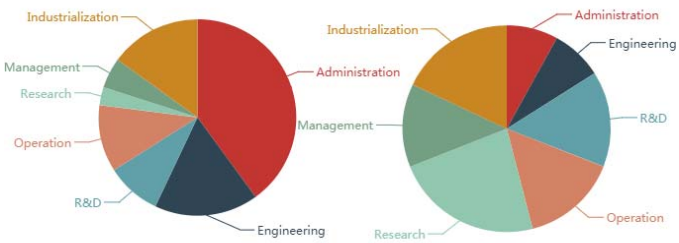

Fig. 8. Comparative analysis of employee behavioral profilings

3) Group profiling for a division.

Once we can formulate a profile of an individual employee, we can then realize group profiling for a division, as shown in Figure 9. It's useful for managers to obtain the change of team's work focus in a long period. And it is helpful to realize the effective allocation and rational utilization of the talent in the organization.

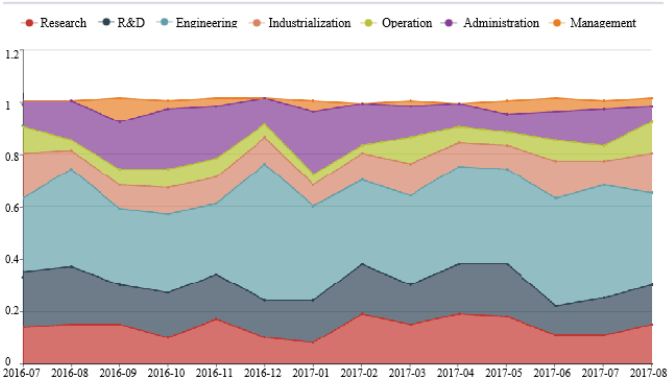

Fig. 9. Group profiling for a division

VI. CONCLUSIONS AND FUTURE WORKS

In this paper, we propose an approach of behavioral profiling for analyzing and better understanding employees, with the input of large amounts of real-time collected data generated by employees' daily reported works on social media. First, we introduce our "WeChat robot platform" which is well designed to collect employees' daily work behavior data, transfer raw data into structured records and obtain key features through data analysis. Then we explain how to construct the skill knowledge graph. Next, we formulate our PO fit model. Finally, we conduct an empirical study to validate our model and analysis using a real-world dataset released by a median-size organization in China.

This paper represents our preliminary study and work on the emerging knowledge automation area [20-23] and our main purpose is to stimulate more in-depth investigation and innovative research in this new direction. In our future work, we are planning to analyze employees' behavior using a Long Short Term Memory (LSTM) model to build a personalized behavior model for each employee, and we are also planning to improve our WeChat robot platform using Blockchain [24-26], which is an emerging and fast-growing decentralized architecture and distributed computing paradigm.

REFERENCES

[1] X. L. Zheng, Y. G. Zhong, D. Zeng, and F. Y. Wang, "Social influence and spread dynamics in social networks", Frontiers of Computer Science, vol. 6, no. 5, pp. 611-620, 2012.

[2] F. Y. Wang, "Social media and the Jasmine revolution", IEEE Intelligent Systems, vol. 26, no. 2, pp. 2-4, 2011.

[3] Y. Lv, Y. Chen, X. Zhang, Y. Duan, and N. Li, "Social media based transportation research: The state of the work and the networking," IEEE/CAA Journal of Automatica Sinica, vol. 4, no. 1, pp. 19-26, 2017.

[4] M. S. Weber, and W. Shi, "Enterprise social media", The International Encyclopedia of Organizational Communication, pp. 1–9, 2016.

[5] S. Zeng, and J. Zhang, "Analyse social influence on student motivation based on social activity network", IEEE International Conference on Service Operations and Logistics, and Informatics, Beijing, China, pp. 133-138, July 10-12, 2016.

[6] J. M. Stanton, "Reactions to employee performance monitoring: Framework, review, and research directions", Human Performance, vol. 13, no. 1, pp. 85-113, 2000.

[7] A. L. Kristof, "Person-organization fit: An integrative review of its conceptualizations, measurement, and implications", Personnel Psychology, vol. 49, no. 1, pp. 1-49, 1996.

[8] B. J. Hoffman, and D. J. Woehr, "A quantitative review of the relationship between person–organization fit and behavioral outcomes", Journal of Vocational Behavior, vol. 68, no. 3, pp. 389-399, 2006.

[9] D. P. Moynihan, and S. K. Pandey, "The ties that bind: Social networks, person-organization value fit, and turnover intention", Journal of Public Administration Research and Theory, vol. 18, no. 2, pp. 205-227, 2007.

[10] J. Cho, D. J. Park, and Z. Ordonez, "Communication-oriented person–organization fit as a key factor of job-seeking behaviors: Millennials' social media use and attitudes toward organizational social media policies", Cyberpsychology, Behavior, and Social Networking, vol. 16, no. 11, pp. 794-799, 2013.

[11] K. Y. T. Yu, "Person–organization fit effects on organizational attraction: A test of an expectations-based model", Organizational Behavior and Human Decision Processes, vol. 124, no. 1, pp. 75-94, 2014.

[12] V. Venkatesh, J. B. Windeler, K. M. Bartol, and et al., "Person–organization and person–job fit perceptions of new IT employees: Work outcomes and gender differences", Management Information Systems Quarterly, vol. 41, no. 2, pp. 525-558, 2017.

[13] D. Zhao, M. B. Rosson, "How and why people Twitter: The role that micro-blogging plays in informal communication at work", Proceedings of the ACM 2009 International Conference on Supporting Group Work, Sanibel Island, Florida, USA, May 10-13, 2009, pp. 243-252, 2009.

[14] Y. Huang, P. V. Singh, and A. Ghose, "A structural model of employee behavioral dynamics in enterprise social media", Management Science, vol. 61, no. 12, pp. 2825-2844, 2015.

[15] X. Wang, L. Li, Y. Yuan, P. Ye, and F. Y. Wang, "ACP-based social computing and parallel intelligence: societies 5.0 and beyond", CAAI Transactions on Intelligence Technology, vol. 1, no. 4, pp. 377-393, 2016.

[16] F. Y. Wang, X. Wang, Y. Yuan, T. Wang, and Y. Lin, "Social computing and computational societies: the foundation and consequences of smart societies", Chinese Science Bulletin, vol. 60, no. 5-6, pp. 460-469, 2015.

[17] F. Y. Wang, "Software-defined systems and knowledge automation: a parallel paradigm shift from Newton to Merton", Acta Automatica Sinica, vol. 41, no. 1, pp. 1-8, 2015.

[18] T. Bai, S. Wang, Z. Shen, D. Cao, N. Zheng, and F. Y. Wang, "Parallel robotics and parallel unmanned systems: Framework, structure, process, platform and applications", Acta Automatica Sinica, vol. 43, no. 2, pp. 161-175, 2017.

[19] J. Li, S. Wang, X. Ni, S. Zeng, Y. Yuan, X. Wang, F. Wang, "How Reporting Policies Influence Employee Performance: An Empirical Study", IEEE/CAA Journal of Automatica Sinica, under review.

[20] F. Y. Wang, J. J. Zhang, X. Zheng, X. Wang, Y. Yuan, X. Dai, J. Zhang, and L. Yang, "Where does AlphaGo go: From Church-Turing Thesis to AlphaGo Thesis and beyond", IEEE/ CAA Journal of Automatica Sinica, vol. 3, no. 2, pp. 113-120, 2016.

[21] F. Y. Wang, "The destiny: Towards knowledge automation-preface of the special issue for the 50th anniversary of Acta Automatica Sinica", Acta Automatica Sinica, vol. 39, no. 11, pp. 1741-1743, 2013.

[22] S. Zeng, S. Wang, Y. Yuan, X. Ni, Y. Ouyang. "A survey on question answering systems towards knowledge automation", Acta Automatica Sinica, vol.43, no.9, pp. 1491-1508, 2017

[23] R. Qin, S. Zeng, J. Li, and Y. Yuan, "Parallel enterprises resource planning based on deep reinforcement learning", Acta Automatica Sinica, vol. 43, no. 9, pp. 1588-1596, 2017.

[24] Y. Yuan, T. Zhou, A. Zhou, Y. Duan, and F. Y. Wang, "Blockchain: From data intelligence to knowledge automation", Acta Automatica Sinica, vol. 43, no. 9, pp. 1-6, 2017.

[25] Y. Yuan, F. Y. Wang, "Blockchain: The state of the art and future trends", Acta Automatica Sinica, vol. 42, no. 4, pp. 481-494, 2016.

[26] Y. Yuan, F. Y. Wang, "Towards blockchain-based intelligent transportation systems", The 19th IEEE International Conference on Intelligent Transportation Systems (ITSC2016), Rio de Janeiro, Brazil, pp.2663-2668, 2016.