Research on User Behavior Patterns under the Background of Big Data

Jieli Sun*, Shuangsi Li², Naishi Yan³, Yanxia Zhao¹, jianke Li¹

¹ School of Information Technology, Hebei University of Economics and Business, Shijiazhuang, Hebei, 050061, China
² School of Humanities, Beijing University of Posts and Telecommunications, Beijing, Beijing, 100876, China
³ School of Computer Science, Beijing University of Posts and Telecommunications, Beijing, Beijing, 100876, China

Corresponding author’s e-mail: 1255844882@qq.com

Abstract. We discuss related researches on Information Quality Control and main CBR process models. We also take a deep study of the user behavior pattern, the formation mechanism of user behavior pattern. The design idea of user behavior pattern construction based on CBR and information quality control is discussed in detail.

1. Introduction
With the improvement of the Internet penetration, more and more people obtain information online. According to The 45th China Statistical Report on internet Development, as of March 2020, the Internet population in China has reached 904 million. The mobile Internet population in China has reached 897 million, increasing by 79.92 million than the end of 2018 [1].

At the same time, in this big data era[2], users get a large quantity of information from rich sources, moreover, it's difficult to ensure the quality of information. So, the need to provide personalized and high-quality information services for users is pressing[3,4]. Whether big data can provide active services for users in an appropriate way or not, the key part is the automatic construction mechanism of user behavior pattern and situation-driven reasoning mechanism. User behavior pattern is widely used in the fields of e-commerce and information service.

2. Related Researches on CBR and Information Quality Control

2.1. Overview of information quality
Information quality is a cross-discipline, which involves many disciplines, such as management, engineering, computer science, psychology. It is generally believed in the existing literature that it is based on Data Quality. Data is usually regarded as a simple fact; and information is put in a certain context and structure.

As early as 1958, some scholars had already studied the data quality issues. Early data quality studies believed that data quality is the accuracy of data. Redman defines data quality in terms of concept level, data value level and form level. The research on data quality mainly focuses on the data
quality issues, the evaluation of data quality and the quality research of information product. Information quality goes with data, and the research on information quality focuses on the user requirements, it doesn’t have a uniform definition so far.

The concept of information quality is put forward by Ballou et al [5]. Richard et al. [6] believe that information quality belongs to a multi-dimensional and hierarchical concept. From the process of users accessing information, the dimension of information quality can be divided to the following parts: accessible, interpretable, useful and believable.

2.2. Related Researches on Information Quality Control
As to information quality architecture, the most representative research is put forward[6]. Richard et al. Strong proposed a groundbreaking framework for comprehensive description of data quality. The framework includes four dimensions: internal data quality, situational data quality, expression quality and accessible quality, respectively. The research on the tools and methods to measure, model and improve the quality of data and information is still in its infancy.

The representative design method based on information quality methodology is an information quality methodology designed, which includes three parts: information quality model, intra-organizational information quality questionnaire and analysis techniques to explain information quality testing methods. Ballou et al. [5] introduced the IP map to model the process of information products for the first time by using symbol modules which are used in the similar data flowcharts.

At present, existing researches on information quality assessment mainly focus on two aspects: evaluation methods, evaluation standards and tools. As of the study on evaluation methods, Richard et al. [6-8] separately put forward the evaluation methods of information quality. Among them, Yang et al. [8] propose an information quality evaluation method called AIMQ to help organizations evaluate their information quality and monitor the process of information quality improvement at any time. In the research of evaluation standards and tools, Yang [8] puts forward two specific standards for information quality evaluation: the stability and comprehensiveness of information sources. He also puts forward an automatic evaluation method based on information source sampling. As of the information quality management, Richard [6] put forward the methods of TDQM (Total Data Quality Management). In addition, other scholars and institutions also work a lot and achieve many results.

2.3. Main CBR Process Models
After analyzing and summarizing the related research literature of CBR, a variety of CBR reasoning process models are proposed. There are mainly four influential models: CBR process model proposed by Hunt[9], CBR cycle R4 reasoning process model proposed by Aamodt and Plaza[10], and R5 reasoning model of CBR proposed by Gavin Finnie et al[11]. The CBR's R4 reasoning process model and CBR's R5 reasoning model will be introduced as follows.

The R4 reasoning model[10] of CBR divides the basic steps of the CBR problem solving process into four steps---that is R4 model: Case Retrieve, Case Reuse, Case Revise and Case Retain. Because the R4 model proposed by Aamodt and Plaza intuitively and abstractly reflects the essential characteristics of CBR reasoning process, it has been widely accepted since its publication in 1994. At present, most of the systems based on CBR are studied on the basis of R4 reasoning process model. However, because the R4 reasoning process model ignores the fact that the construction of the case base of the CBR system is also the main part of CBR. Gavin Finnie et al propose a R5 reasoning model which can overcome the defects of the above reasoning process model.

Gavin Finnie et al extend the R4 reasoning process model of CBR and propose the R5 reasoning model[11,12], which brought the case base generated by CBR system into the unified reasoning process of CBR, so it is possible to generate the case base automatically. The R5 in the CBR's R5 reasoning model means Repartition, Retrieval, Reuse, Revise and Retain, in which Repartition provides the theoretical basis of case retrieval, thus providing a mathematical basis for the construction of case base and case retrieval to a certain extent. The R5 reasoning process model of
CBR overcomes the difficulties of R^4 model and other CBR reasoning models in the description of CBR reasoning process. It’s an ideal CBR reasoning process model.

3. Construction of User behavior pattern
The user behavior pattern based on case-based reasoning and information quality control in this paper is designed and implemented based on the R5 reasoning process model of CBR. During the research, we not only study the classical link of CBR but study the new link of R5 reasoning model of case-based reasoning. We study the automatic acquisition, the generation and the case base construction of user behavior case and propose corresponding solutions.

3.1. Design goal of user behavior pattern based on CBR and information quality control
The user behavior pattern based on CBR and information quality control reflects the interaction between the user and the information environment. It places the user behavior pattern between the user and the personalized service. On the one hand, the data information is extracted from the information resources provided by the information service system. On the other hand, the relevant information of the user who use the system is processed, which forms the continuous interaction process between the user information flow and the information service system data flow.

The problems of user behavior pattern based on CBR and information quality control is the demand of information users. The second part is the user group with these demands. The next part is how user demand information affects user behavior and how to affect similar user demands and behavior. Based on the demand analysis of the above user behavior patterns, the specific design objectives are listed as follows:

• enhance the function of case data analysis and processing of user behavior pattern to meet the requirements of different users. In order to meet the different data analysis and processing requirements in the operation of the personalized service system, all kinds of user behavior pattern case base data must be planned and designed in the analysis and design of system.
  • The accuracy and standardization of user behavior pattern will directly affect the subsequent recommendation steps of personalized service system. It is related to the result of personalized recommendation. So it’s important to analyze and process the user behavior data. Obtaining and storing correct user behavior data plays an vital role in the construction of user behavior pattern.
  • The user behavior pattern should be flexible to adapt to the changes of personalized user demands and personalized information services. And it should be simple to maintain and easy to use.

3.2. User behavior pattern construction
User behavior pattern needs to go through three stages: user behavior case representation and user behavior case base construction, user behavior case similarity judgment, user behavior case retrieval and user behavior pattern reasoning to realize the function of user behavior pattern construction. The result will produce a user behavior case base that represents the users’ background knowledge or interests, needs and related recommendation results. The user behavior case similarity judgment and case retrieval phase will be based on the user behavior case base, use a variety of case retrieval and personalized recommendation techniques to find user behavior pattern cases, and then present the comprehensive calculation results of these cases to users and update the user behavior case base. The user behavior pattern construction process figure is shown in Figure 1.

• User behavior pattern data collection and preprocessing. After logging in to the system, the system users first register, and after becoming registered users, they can get fine-grained personalized recommendation service.
  • The users behavior data using the system is first saved in the system log. After the relevant user behavior data are cleaned, analyzed and processed by the personalized recommendation system, it is stored in the user temporary database and relevant knowledge base of the system.
  • Analyze and establish user behavior cases and pattern cases, establish and update the corresponding case base.
The user behavior cases base includes user information, user personalized recommendation results, and generates user behavior pattern cases as the basic data of recommendation reasoning in the system.

Figure 1 User Behavior Pattern Construction Process Figure

4. conclusion.
With the operation of personalized system, its user behavior case base is growing. The user behavior case base may be redundant, contradictory and bloated, which will affect the efficiency and accuracy of the solution, so it is necessary to further study the update and revision of the user behavior case base.

Acknowledgments
This paper is supported by Science and technology plan project of Hebei Province (No.18960122D, No.16960314D), and Technology Fund Project of Hebei University of Economics and Business (No.2020PY10).

References
[1] China Internet Network Information Center. (2020) http://www.cnnic.net.cn/.
[2] Wu, J. (2016) Smart era. CITIC Press Group, Beijing.
[3] Chen, L. C., Kuo, P. J., Liao, I. E. (2015) Ontology-based library recommender system using MapReduce. Cluster Computing, 18(1):113-121.
[4] Jin, Z. F. Design and Implementation of Big Data-based Educational Resources Personalized Recommendation System. Beijing: University of Chinese Academy of Sciences.
[5] Ballou, D. P., Pazer, H. L. (1985) Modelling data and process quality in multi-input, multi-output information system. Management Science, 31(2):150-162.
[6] Richard, Y. W., Reddy, M. P., Henry, B. K. (1995) Toward quality data: An attribute-based approach. Decision Support Systems, 13:349-372.
[7] Yang, W. L., Diane, M. S.,(2002) AIMQ:a methodolog for information quality assessment. Information & Management, 40:133-146.
[8] Yang, W. L., (2003) Context-reflective data quality problem solving. Journal of Management Information Systems, 20(3):93-119.

[9] Hunt, J. (1995) Evolutionary case based design, In: I.D. Watson (Ed.), Progress in Case-based Reasoning, Springer, Berlin LNAI 1020:17-31.

[10] Aamodt, A., Plaza, E. (1994) Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches[J]. AI Communications, 7(1):39-59.

[11] Finnie, G., Sun, Z. (2003) R5 model for Case-based reasoning[J]. Knowledge-Based Systems, 16(1):59-65.

[12] Sun Z., Finnie G., Webe K. r. (2004) Case base building with similarity relations[J]. Information Sciences, 165(1-2):21-43.