Research on the Construction of Intelligent Learning System Based on Big Data

Zheng an Yang*
Department of Electronics and Information Engineering, Lanzhou Vocational Technical College, Lanzhou, Gansu, China

*Corresponding author’ e-mail: 30227503@qq.com

Abstract. Internet technology and online learning are in a period of high integration. The learner-centered online learning environment is no more restricted by distance. Facing the large amount of learning resources on the Internet, learners often have no choice. The problem cannot be solved in time, so it is imperative to build an intelligent learning system that according to the characteristics of learners. This article analyses the design of an intelligent learning system based on big data, and analyses the key points of the fit between big data and personalized learning features. According to the design principles of intelligent learning system based on big data, it is divided into four categories: learning strategies, learning tools, learning resources, and learning activities. To build an intelligent learning system in three dimensions. Then the recommended mode of resources in the intelligent learning system is introduced in detail. According to the above design ideas of the intelligent learning system based on big data, a personalized learning platform is constructed, and the structure, functions, and development technology of the platform are elaborated.

Keywords. Big Data, Data mining, Resource recommendation

1. Introduction
At present, big data is widely used in various fields. It has matured in the field of e-commerce and provided convenience to people’s lives. People are paying more and more attention to the proper terms such as "personalization" and "data mining". It has also received extensive attention from the education community. Educational researchers are also exploring the application of big data in the field of education. How to use big data to achieve personalized learning is a problem that we education technology researchers must break through. Students have transformed from digesters of content to creators of content, learning from the classroom to the environment, and the development of technology and changes in learning concepts have triggered changes in the learning environment. Online learning led by big data can track, record, and master the characteristics, learning needs, learning behaviours and learning foundations of different learners in an all-round way, build learning models based on learners' characteristics, and create personalities for different types of learners With customized learning strategies, learning tools, learning resources, and learning activities, the learning content of each learner is no longer the same, and will be dynamically presented according to the learning trajectory, creating a personalized learning environment for the learner.
2. Research status of intelligent learning environment based on big data

2.1. The status quo of foreign education big data research

The application of big data in education is relatively mature in China. Data-driven classroom learning began in the United States in the early 1960s, and big data was used in elementary education, higher education, and social education [1]. In 2012, the report "Promoting Teaching and Learning Through Educational Data Mining and Learning Analysis" proposed by the U.S. Department of Education laid a theoretical foundation for the application of big data in education, and the report on the application of American big data in education [2]. The research case is analyzed and the challenges faced are explained in detail. In September 2013, Khan Academy launched a learning analysis dashboard, which formed personalized recommendation results based on a large amount of learning trajectory data of learners. In 2014, the Horizon Report was released, explaining the learning analysis technology [3]. For educators, learning analysis technology can mine the information hidden in education data, propose better teaching methods, and locate various learning groups; For educational researchers, learning analysis technology can reveal the relationship between learners and learning resources. For learners, learning analysis technology can create a personalized learning experience for learners [4]. In 2015, Victor Meyer-Schoenberg's "Walking with Big Data: The Future of Learning and Education" [5]. He cited the Khan Academy, MOOC, Duolingo Language Learning Website and other cases, Victor Meyer-Schoen Berg [6] told us that the booming field of online education has produced big data, and education is no longer just "you speak and I listen" and test scoring. For the first time in history, we have a powerful tool with empirical effects [7]. Davenport believes that learning analysis technology has experienced a total of 3 eras, the analysis 1.0 era, the analysis 2.0 era, and the analysis 3.0 era. The analysis 1.0 era focuses on data collection and preparation. This stage requires the longest time, mainly the statement of the past. However, there is a lack of explanation and prediction. In the era of analysis 2.0, big data is different from traditional small-scale data [8]. The scope of big data is expanded to the entire network, and there are many ways to process data. Data analysis has entered the 3.0 era, and data is not limited to the Internet, but extends to the Internet of Things, and even the endless "data universe" [9].

2.2. Current status of domestic education big data research

Through CNKI’s retrieval and analysis of big data education, on the advanced search page, select "big data" and include "education" as the search criteria for precise retrieval, and a total of 7859 search results were obtained (as of November 16, 2020). See Figure 1 for details: Regarding the publication of big data academic paper.

![Figure 1](attachment:figure1.png)

**Figure 1.** Regarding the publication of big data academic paper
From the statistical results, it can be found that although a series of measures taken by the US government in 2012 have attracted worldwide attention, my country's big data research in education started relatively late, and no relevant research was done before 2012. In 2013, a large number of papers were published successively in my country, marking that Chinese scholars have begun to pay attention to the potential value of big data in education, and my country's research on big data in education has entered the initial stage. In the following four years, the number of papers on the application of big data in education has grown substantially, and it is expected to continue to grow in 2021. The application value of big data in education has received extensive attention from Chinese scholars.

Domestic research on the application of big data in education is still at the stage of theoretical discussion and model establishment, and there are few practical application cases. For example, in 2009, Zhang Hongbo [10] conducted research on the development status of educational resource sharing environment and sharing mechanism construction, and proposed the construction of a shared learning environment for advanced learning resources. In 2013, Qian Dongming [11] proposed a system analysis framework for the co-construction and sharing of digital educational resources, expounding how to promote the co-construction and sharing of educational resources. In 2014, Gu Xiaoqing conducted research on education big data and provided theoretical guidance for obtaining learning process data through the Experience AP specification released by the United States. In 2015, Yang Xianmin and others conducted research on the application model and policy recommendations of education big data, and put forward six policy recommendations based on the challenges encountered in the application of my country’s education big data. In 2016, Xing Beibei and others conducted research on the sources and collection techniques of education big data, and introduced the sources of education big data and several common data collection techniques. In 2017, Zhu Zhiting and others proposed that although the application of big data in other fields has been effective, it is still in the exploratory stage in the education field. In 2018, Wu Xiaorong pointed out that going with big data is the future of education. These papers discuss how to build an education platform for big data applications, how to realize resource sharing, and how to collect useful data. These theoretical studies are just the preliminary preparations for realizing big data analysis.

3. Design of Intelligent Learning Platform
The intelligent learning system can perform self-paced learning according to a series of embedded learning brackets and learning templates. At the same time, we will collect learners’ basic data and data generated during the learning process, analyze them, and find For the learner’s current problems, adjust his current learning pace, and recommend suitable learning resources and learning paths to the learner, as shown in Figure 2: Resource integration based on big data.

![Figure 2. Resource integration process](image)

3.1. Process data mining
The establishment of a learning process database is a key step in the integration of big data resources. The combination of systematic Web log mining and platform database query is used to mine the learner’s learning process data. The benefits of combining these two methods are to compensate for each other. Inadequate, the use of Web log mining can understand the overall use of the platform,
including the number of page views, various clicks, the number of visitors, and the duration of the visit. The usage of platform resources, including the most common access paths, entry pages, exit pages, web pages, dynamic pages, most downloaded files, etc., helps to dig out the most popular learning resources and the most popular learning modes. Understand the distribution and geographical distribution of visitors' clicks. Knowing the number of visits and clicks in each time period, and finding out the most active time period for learners to learn, helps to rationally arrange learners' learning time. Using Web log mining can mine the data generated by learners in the learning process, which provides favourable conditions for learning analysis.

- **Web log mining**
  Web log mining is to collect learners' browsing information on the Internet, understand learners' daily learning behaviours, dig out learners' browsing habits, and analyze learners' learning interests. Mining data mainly includes learner registration, login information, resource search query information, learner click information, log files, etc. Web log mining is mainly to collect, analyze, and integrate the browsing data generated by learners during the process of platform access.

- **Web data pre-processing**
  Collecting the original log files of the learners' access to the platform cannot be used directly for data mining. The original log files need to be processed to filter out useless browsing records and delete data information that has nothing to do with mining data algorithms, in order to be able to see clearly Log records after data processing, build a learner behavior database during the development of the personalized learning platform, store browsing information tables in the database, and then put the processed data into this browsing information table. See Table 1 for details. User browsing web pages.

| Field name     | type of data | length | Field meaning                        | Remarks       |
|----------------|--------------|--------|--------------------------------------|---------------|
| ip             | varchar      | 29     | The ip requested by the user          | Main chain    |
| url            | varchar      | 20     | Page requested                       |               |
| Start time     | datetime     | 20     | Start time                           |               |
| Terminal time  | datetime     |        | End of visit                         |               |

3.2. Personalized learning resource recommendation algorithm

3.2.1 User-based collaborative filtering recommendation algorithm
The user-based collaborative filtering algorithm in the personalized recommendation of learning resources means that if a user in a group with the same interest characteristics is interested in a thing, the system will consider that this group is interested in this thing, and then recommend it. For similar users, "N Nearest Neighbours" refers to the groups that have the same interest as the current user calculated, sorted according to the degree of similarity of the user’s interest, and the top N users rate the item. The scoring data is used as the basis for personalized recommendation of learning resources. See Figure 3 for details: Recommended process.

![Figure 3. Recommendation process](image-url)
Extract the user rating data for all items to form a M*N two-dimensional matrix, which represents the number of users, N represents the number of items, users (User), the set of users is expressed as \( \text{Un} = \{U_1, U_2, \ldots, U_n\} \), M represents the number of items (Item) \( \text{Im} = \{I_1, I_2, \ldots, I_m\} \), the i-th row and j-th column in the formed matrix represent \( U_i \)'s rating of \( I_j \), \( V_{ij} \). As shown in the following table, the value is usually set in the range 0-5. It shows in Table 2 for details.

| Item | Use  | \( U_1 \) | \( U_2 \) | \( U_1 \) | \( U_n \) |
|------|------|-----------|-----------|-----------|-----------|
| I_1  | V_{11} | V_{12} | \ldots | V_{1i} | \ldots | V_{1a} |
| I_2  | V_{21} | V_{22} | \ldots | V_{2i} | \ldots | V_{2a} |
| \ldots | \ldots | \ldots | \ldots | \ldots | \ldots | \ldots |
| I_j  | V_{ji} | V_{j2} | \ldots | V_{ji} | \ldots | V_{ja} |
| \ldots | \ldots | \ldots | \ldots | \ldots | \ldots | \ldots |
| I_m  | V_{mi} | V_{m2} | \ldots | V_{mi} | \ldots | I_{mn} |

Then, similar users are obtained, and the items with high predicted scores are recommended to the target users. First calculate the similarity between users and find all the neighbours of similar users. Then determine the interest similarity between neighbouring users. Finally, according to the predicted value, items with higher scores are pushed to users. Specific steps are as follows:

- According to the user-item rating matrix, calculate the similarity value \( U_{\text{Similarity}} \) between ordinary users in the platform and target users.
- The similarity value is arranged in order from large to small, and the user with the large similarity value is taken as the neighbour user.
- Calculate all the items evaluated by the neighbours, filter out the items evaluated by the target user, and get the recommendation list \( V \).
- Find out users who have commented on list \( V \) among neighbour users, and calculate the similarity \( I_{\text{Similarity}} \) between these users and the target user.
- Calculate the predicted value of the project, and recommend the project with a higher predicted value to the target user.

3.2.2 Item-based collaborative filtering recommendation algorithm

As the number of users using the platform gradually increases, the system takes more time to calculate the similarity between users and the calculation speed is slower. Therefore, the project-based collaborative filtering method is cited. The project-based collaborative filtering method is faster in calculation. More directly, this algorithm mainly recommends based on the degree of similarity between the rated items and the unrated items by the target user. The specific steps are as follows:

- Count the set \( V \) of all unrated items of the target user.
- Count the set \( I \) of all the rated items of the target user.
- Calculate the similarity \( S \) between \( V \) and \( I \).
- Calculate the predicted value of set \( I \) according to the weight of preference, and recommend items with high scores in \( I \) to target users.

3.3. Intelligent learning platform design framework

3.3.1 The platform system framework

The platform of the framework is shown in Figure 4 Functional structure.
3.3.2 System database model
The main data sheet is designed as follows:

- Table 3 for Learner user table:

  | Field name | Data type | length | Field meaning       | Remarks      |
  |------------|-----------|--------|---------------------|--------------|
  | ID         | int       | 20     | Learner ID          | Primary key  |
  | username   | Varchar   | 40     | username            |              |
  | password   | Varchar   | 32     | password            |              |
  | Realname   | Varchar   | 40     | actual name         |              |
  | User-idcard| char      | 20     | User ID number      |              |
  | User-mobile| Varchar   | 11     | User phone number   |              |
  | User-email | Varchar   | 32     | User mailbox        |              |
  | Random-code| Varchar   | 32     | Email activation    |              |
  | gender     | Varchar   | 1      | Male (0), female (1)|              |
  | Gmt-creat  | datetime  |        | Creation time       |              |
  | Gmt-modify | datetime  |        | Change the time     |              |

  - The resource id is the primary key and is used as the unique identifier. See Table 4 Learning Resource Table for details.

  | Field name    | Data type  | length | Field meaning    | Remarks      |
  |---------------|------------|--------|------------------|--------------|
  | ID            | int        | 20     | Resource ID      | Primary key  |
  | Resource-name | Varchar    | 40     | Resource Name    |              |
  | Res-description| Text      |        | Resource description |        |
  | Resource-tag  | Varchar    | 200    | Resource tag     |              |
  | Resource-ip   | char       | 20     | Upload IP        |              |
  | Resource-type | int        | 4      | Resource Type    |              |
  | Uploadtime    | datetime   |        | Upload time      |              |
  | Download-count| int        |        | Downloads         |              |
  | Reading-count | int        |        | Reading          |              |
Verify-type int Approval Status
Initial-score float Initial rating

- See Table 5 for details: Resource comment form.

### Table 5. Resource Comment Form

| Field name   | Data type | length | Field meaning                  | Remarks         |
|--------------|-----------|--------|--------------------------------|-----------------|
| ID           | Decimal   | 20     | Comment ID comments            | Primary key     |
| Comment-content | text      |        | Commenter id                   |                 |
| User-id      | decimal   | 20     | Comment vocabulary             |                 |
| Comment-vocabulary | text      |        | Comment id                     |                 |
| Comment-time | datetime  |        | Comment time                   |                 |
| Resource-id  | decimal   | 20     | Comment resource id            |                 |

- See Table 6 for details: Learning resource download form.

### Table 6. Download table of learning resources

| Field name    | Data type | length | Field meaning       | Remarks         |
|---------------|-----------|--------|---------------------|-----------------|
| Resource-ID   | decimal   | 20     | Resource ID         | Primary key     |
| Download-time | datetime  |        | Download time       |                 |
| User-id       | decimal   | 20     | Downloader id       |                 |

- See Table 7 for details: Resource score sheet.

### Table 7. Resource Scoring

| Field name        | type of data | length | Field meaning                  | Remarks         |
|-------------------|--------------|--------|--------------------------------|-----------------|
| User-id           | decimal      | 20     | Rater id                       | Primary key     |
| Resource-ID       | decimal      | 20     | Resource ID                    |                 |
| Rating            | float        |        | fraction                       |                 |
| Resource-type     | varchar      | 10     | Types of learning resources    |                 |
| Rating-time       | Datetime     |        | Scoring time                   |                 |

- For details, see Table 8: Resource browsing table.

### Table 8. Resource browse

| Field name     | Data type | length | Field meaning       | Remarks         |
|----------------|-----------|--------|---------------------|-----------------|
| ID             | Decimal   | 20     | Comment ID          | Primary key     |
| Resource-ID    | decimal   | 20     | Resource ID         |                 |
| Start-time     | datetime  |        | Starting time       |                 |
| finish-time    | datetime  |        | Complete time       |                 |

### 4. Conclusion

The intelligent learning system based on big data is proposed to make education researchers pay more attention to the research and development of new learning environment; let the builders of learning environment change the learning concept from the intelligent learning system based on big data is the development trend of the Internet education era. Internet learning has played an important role and is also the direction of future education development. The application prospects of big data in intelligent learning systems are very broad, but there are still some challenges. In terms of technology, it is mainly in the four links of data mining: Data collection, storage, processing, and visualization of data.
results. Second, computer hardware is facing challenges in processing data. In addition, big data faces other issues, such as people's lack of awareness of big data, the credibility of data, and the legality of data disclosure. The application of big data in personalized learning is also in continuous exploration and development. All problems will be solved accordingly, and related research will be more systematic and comprehensive.

Acknowledgments
Thanks for the following fund for supporting this article: Research Project of Lanzhou Vocational and Technical College_"Research on Grant Evaluation System Based on Big Data" [2018xy-16].

References
[1] Zhang Y N. 2016 Research on the application of big data in education (Shanghai: East China Normal University).
[2] Jiang Q, Zhao W, Li Y F, Li S. 2017 Research on the dashboard of learning analysis based on big data vol 1 (China Audio-visual Education) pp 112-120.
[3] Zhu Z T, Sun Y Y, Peng H C. 2017 Interpretation of the cultural implication of education big data vol 1 (Audio-visual Education Research) pp 28-36.
[4] Xie Z P, Jin C, Liu Y. 2018 Personalized knowledge recommendation model based on constructivist learning theory vol 1 (China Audio-visual Education) pp 112-120.
[5] Lam X N, Vu T, Le T D, et al. 2018 Addressing cold-start problem in recommendation systems. (Proceedings of the 2nd international conference on Ubiquitious information management and communication) pp 208-211.
[6] Zhang W. 2017 Design and implementation of a personalized recommendation system for learning resources based on collaborative transition algorithm (Tianjin: Tianjin Normal University).
[7] Sun Y H, Tang Z W. 2017 Research on the construction of personalized learning environment based on big data vol 2 (Audio-visual Education for Primary and Secondary Schools) pp 51-54.
[8] Guo L T. 2017 Research on personalized learning mode under smart learning environment. (Changchun: Northeast Normal University).
[9] Zhao H, Chen L, Zhao Y T. 2015 Design of a personalized learning strategy guidance system based on learning style vol 5 (China Audio-visual Education) pp 67-72.
[10] Sun Y H. 2017 Discussion on personalized learning tools based on big data vol 3 no 19 (Quality Education in West China) pp 127-128.
[11] Li D Y. 2017 Research on resource push scheme in intelligent learning platform (Changchun: Northeast Normal University).