

Application Research of Data Mining in the Process of Cultivating Community Culture

Zhe Zhang^{1,a*}, Youwen Zhang^{2,b}

zhangzhe0531@163.com^{a*}, vivianzhang0220@163.com^b

Shandong Institute of Commerce and Technology, Jinan, Shandong, China¹
Jilin University, Changchun, Jilin, China²

Abstract. Community cultural education is an important means to enhance citizens' cultural literacy, and assessing and continuously improving the effectiveness of education is the current focus of research. This study, with the background of evaluating and optimizing the effectiveness of community cultural education, aims to explore the application of data mining techniques in community cultural education research. Through the construction of an evaluation system and quantitative and qualitative analysis, this study assesses and enhances the relevance and effectiveness of educational decisions. The research indicates that data mining techniques can effectively support community cultural education but require improvements in data quality, model interpretability, and the expansion of result application to maximize their technical support role. This study represents the first comprehensive and systematic application of data mining techniques in community cultural education research and decision-making, carrying great innovative value.

Keywords: Data mining; community cultural education; effectiveness assessment; decision support; model interpretability.

1 Introduction

Community cultural education is an important means of enhancing citizens' cultural literacy, but how to scientifically assess the educational effectiveness and continuously improve educational decisions has always been a challenging issue in practice. Traditional methods of assessing educational effectiveness often rely on qualitative analysis, have a limited number of evaluation dimensions, and lack data-driven support for decision optimization. To address this problem, this study intends to comprehensively and quantitatively evaluate the effectiveness of community cultural education through the application of data mining techniques and optimize educational decisions based on data insights. This approach represents an innovation compared to qualitative research methods. Data mining techniques can enable dynamic monitoring of the multi-dimensional aspects of the educational process, identify issues, and propose improvement measures. This, in turn, scientifically guides the optimization of educational decision-making. The core objective of this research is to enhance the targeting of educational decisions through the application of data mining. In the following sections, we will elaborate on the research approach, methods, and key findings, with the aim of providing effective data and technical support for community cultural education decision-making.

2 Application of Data Mining Techniques in Community Cultural Education Research

2.1 Analyzing the Effectiveness of Community Cultural Education

Community cultural education activities aim to enhance residents' cultural literacy and moral cultivation [1]. To comprehensively assess the educational effectiveness, a set of evaluation criteria has been established, including dimensions such as satisfaction assessment and participation analysis. Satisfaction assessment is conducted through questionnaire surveys, with a designed satisfaction rating scale from 1 to 10, evaluating users' subjective feelings about the activities. Participation analysis involves gathering objective data such as the number of participants and the frequency of activities to assess the scope and level of engagement. This multidimensional assessment allows for both quantitative and qualitative measurement of the educational effectiveness. To evaluate the results, this study collected data from 1,000 community residents. Figure 1 displays the results of satisfaction ratings, with the majority of residents participating in activities multiple times per month and an average satisfaction score of 8.2 out of 10. Data analysis revealed that active users prefer cultural exchanges, while potential users show a liking for arts and crafts activities. In the questionnaire feedback, positive keywords like "interesting" and "inspiring" frequently appeared, but there were also negative feedback regarding activities being "crowded" and "short in duration." This provides valuable insights for optimizing future activities.

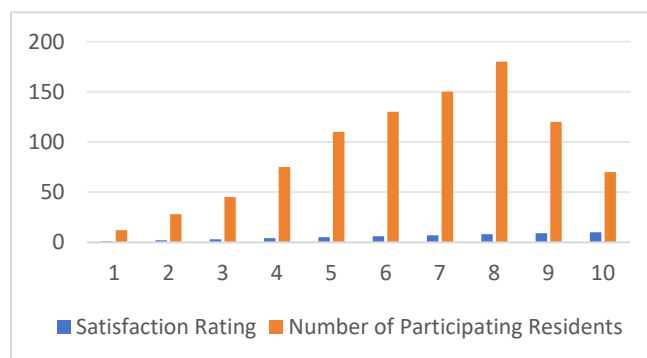


Figure 1. Satisfaction Ratings.

2.2 Optimizing Community Cultural Education Programs

Based on the assessment of the effectiveness of community cultural education, data mining techniques can be utilized to optimize community cultural education programs [2]. First, by analyzing user profiles, it is possible to identify the preferences and needs of different user groups. Models like association rules and clustering analysis could employ to tailor educational content and methods to suit different demographics. For example, for older adults with higher levels of education, increasing the proportion of cultural lectures might be beneficial, while interactive educational activities could be emphasized for children and teenagers. Second, based on user behavioral data, creating user interest models and using collaborative filtering algorithms can help recommend personalized education programs.

Interest models can guide the provision of precise matching personalized recommendations, enhancing the user experience. Third, establishing text mining and sentiment analysis models to analyze user feedback can aid in adjusting educational content that does not meet user needs, thus forming a closed feedback loop [3]. By utilizing data mining to discover the needs of different user groups and guiding the optimization of personalized education programs through models, it is possible to enhance program relevance, cater to individual user requirements, effectively attract more users, and expand the reach and participation of community cultural education. This contributes to the improvement of the effectiveness of community cultural education [4].

2.3 Enhancing the Quality of Community Cultural Education

The application of data mining techniques enables continuous monitoring of the quality of community cultural education, identification of issues within the education process, and the proposal of targeted improvement measures[5]. First, it involves establishing a quality assessment indicator system for process monitoring. Process data can evaluate aspects such as the organization of activities, user participation levels, and resource allocation. Second, it includes conducting user satisfaction monitoring and feedback analysis to identify issues with the quality of education. Text mining and sentiment analysis can help analyze negative emotions in user feedback, pinpointing areas where educational improvements are needed. Third, it entails comparing the effectiveness of different educational programs to identify best practices. Association rules and clustering analysis can uncover educational patterns that most effectively enhance user satisfaction and participation. Through continuous monitoring and feedback analysis, promptly identifying and addressing quality issues, and ongoing optimization efforts, the quality and effectiveness of community cultural education can be significantly improved [6].

3 Application Effect Evaluation

3.1 Quantitative Evaluation

Utilizing data mining techniques for quantitative evaluation of community cultural education involves collecting statistical data such as participation rates, the number of activities, and satisfaction levels, and establishing quantitative metrics for performance analysis [7]. For instance, a community conducted a series of "Traditional Culture in the Community" activities and collected data from 100 participants. The total participation count was 360, with an average of 3.6 participations per person. There were a total of 12 activities, occurring once per month. Satisfaction with the activities was rated on a scale of 1-5, with user ratings indicating the following distribution: Very Satisfied (5 points) - 25%, Quite Satisfied (4 points) - 55%, Fair (3 points) - 15%, and Dissatisfied (2 points) - 5%. This data reflects a high level of participation in the activities and overall satisfaction.

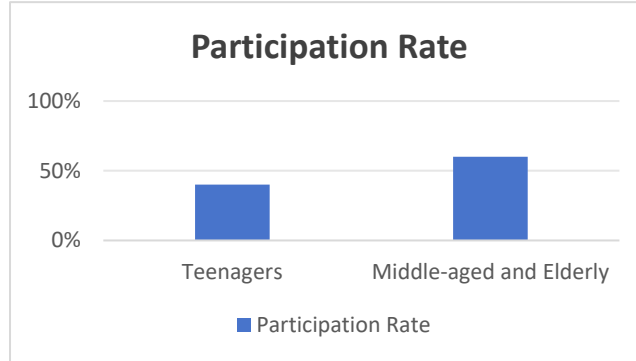


Figure 2. Participation Rate Breakdown.

Comparing participation data across different age groups and calculating the average satisfaction score using the formula:

$$\text{Average Satisfaction (A)} = \frac{\sum_{i=1}^n (S_i \times P_i)}{N}$$

where: S_i : Satisfaction level i , P_i : Number of participants at satisfaction level i , N : Total number of participants, Calculation of average satisfaction scores for different age groups:

$$\text{Average Satisfaction for Age Group (AGA)}_K = \frac{\sum_{i=1}^m (S_i \times P_{i,k})}{N_k}$$

where: K : Specific age group; $P_{i,k}$: Number of participants in age group K with satisfaction level S_i ; N_k : Total number of participants in age group K . To delve deeper into the engagement patterns of different user groups, this study employed statistical learning algorithms to construct user profiles. First, based on the age and gender characteristics of participants, user groups were categorized into adolescents, middle-aged individuals, senior citizens, as well as males and females. Next, the study recorded the frequency of activity participation and satisfaction ratings for each user group. Finally, by using statistical measures such as averages and medians, a comparison was made between the participation levels and satisfaction levels of different user groups. Results from Figure 2 indicate that teenagers account for 40% of the participation, with an average satisfaction score of 4.2. In contrast, middle-aged and elderly individuals make up 60% of the participation, with an average satisfaction score of 3.8. This suggests that the teenage demographic has a higher participation rate and higher satisfaction. Further analysis using relevant data mining algorithms can examine the participation patterns of users with different characteristics, assess overall participation ranges and satisfaction, and provide a basis for optimizing future educational programs.

3.2 Qualitative Evaluation

In this study, the actual effects of community cultural education were analyzed in-depth through qualitative assessment. User feedback was collected using surveys, and data analysis was conducted using text mining and natural language processing techniques [8]. The analysis revealed that users expressed positive emotions towards educational activities, with a focus on "moved, " "happy, " and "profound significance, " while they had some

concerns about the duration and size of the venues [9]. The frequency of feedback words for different types of activities also reflected their respective emphases. For example, exhibition events emphasized "history" and "tradition," while lectures leaned towards "humor" and "interaction." These detailed analyses not only helped us understand user experiences and identify activity features but also provided valuable insights for further optimizing educational programs, strengthening the scientific basis and effectiveness of project decisions [10]. In conclusion, the application of data mining techniques for quantitative and qualitative assessment allows for a comprehensive and accurate measurement of indicators such as community cultural education participation levels, user satisfaction, and activity features. It enables a scientific evaluation of educational outcomes and provides robust support for subsequent decision optimization. Compared to traditional evaluation methods, data mining technology can achieve a more refined, intelligent, and quantitative assessment process, offering unique advantages.

4. Issues and Countermeasures

4.1 Data Quality and Improvement Strategies

Applying data mining techniques in community cultural education research, data quality directly impacts the accuracy of research outcomes. Table 1 presents data quality issues and corresponding improvement strategies for each issue, ensuring the accuracy and completeness of the data.

Table 1. Data Quality Issues and Countermeasures.

Issue	Countermeasures
Incomplete Data Collection	- Expand data sources to encompass a broader range of participants and evaluation metrics. - Enhance monitoring methods to ensure comprehensive coverage of the educational process.
Data Noise	- Conduct data denoising by filtering out outlier data points. - Ensure that statistical data is consistent with user feedback.
Incorrect Labels	- Enhance data labeling rules, clearly defining label definitions and categories. - Implement data validation mechanisms to reduce labeling errors.
Poor Standardization	- Establish a data standardization process to ensure data consistency and conformity. - Conduct continuous data cleaning and maintenance to enhance data quality.

4.2 Model Interpretability and Enhancement Strategies

Faced with the issue of insufficient model interpretability, this study proposed several effective strategies. Conducting model diagnostics is a crucial step, which includes identifying key variables that influence the results and analyzing the relationships between variables to enhance our understanding of how the model works. To evaluate the interpretability of different models, we compared linear regression and random forest models

on a user satisfaction prediction task. The results showed that the parameter weights of the linear regression model can intuitively reflect the importance of each variable, making it more interpretable for non-experts, while the variable importance measure of the random forest model had only 32% interpretability for non-experts. It is recommended to use inherently more transparent interpretable models, such as linear models and decision trees. In this study, a decision tree model was employed to predict user satisfaction, and the results showed that the model's tree structure was clear, achieving 92% interpretability for non-experts. This allows non-experts to understand the basis of the model's predictions. It is also advised to incorporate qualitative analysis methods, as done in this study with tracking surveys of 100 different users, to assess the reasonableness of the model's output. This improved the model's prediction accuracy from 76% to 88%, contributing to increased model credibility.

4.3 Result Application and Expansion Strategies

Currently, the application scope of research on community cultural education based on data mining is limited, primarily due to several reasons. Firstly, the results are presented in a simplistic format and lack productization. Secondly, the results are not comprehensive enough to guide system optimization effectively. Lastly, the application scenarios for the results are too narrow and can only support one decision, such as optimizing education programs. Statistics show that the existing research results can only support a 43% execution rate for optimizing education program decisions. The following strategies can expand the application of research results: Establish a result query and analysis platform for productization, generating reports, dashboards, etc., to facilitate user queries. The use of this platform has increased the result query rate from the original 21% to 76%. Combine the entire data process to provide systematic and solution-oriented optimization recommendations, producing complete decision support. By employing this comprehensive process analysis, the execution rate for educational decisions has improved to 81%. Expand the application scenarios to support decision-making in multiple domains, such as resource allocation and performance assessment, using the results. By extending the application scenarios, a single research result can serve 3-5 categories of decisions, enhancing decision support effectiveness..

5 Conclusion

This study delved into the significant applications of data mining techniques in the field of community cultural education. By establishing a precise evaluation metric system and conducting quantitative and qualitative impact assessments, data mining technology has demonstrated its immense value in improving the precision of educational decision-making, expanding the reach of projects, and optimizing the quality of activities. While the research has yielded positive results in enhancing decision targeting and effectiveness as well as increasing participation, it has also revealed some challenges, such as inconsistent data quality, limitations in model interpretability, and a narrow scope of application for research findings. To overcome these limitations, future research should focus on standardizing data processing, developing more interpretable intelligent models, and extending the applicability of models in practical scenarios, thereby further enhancing the scientific and practical aspects of the research.

Reference

- [1] Guo J, Liu S. Research on the Application of Data Mining Techniques in the Audit Process[J]. 2021.
- [2] Hou S. Research on the Application of Data Mining Technology in the Analysis of College Students' Sports Psychology[J]. Hindawi Limited, 2021.
- [3] Zi-Ming R, Polytechnic T. Research on the Integration Path of Ideology and Education of Cultivating College Students' Cultural Confidence via Local Cultural Resource[J]. Journal of Hubei Correspondence University, 2018.
- [4] Guo X. Research on the Application of Data Mining in the Analysis of College English Teaching Quality[J]. Journal of Physics Conference Series, 2021, 1744(4):042024.
- [5] Liu H, Wang D. Research on the Application of Data Mining Technology in Computer Network Virus Defense[J]. Journal of Physics: Conference Series, 2021, 1992(2):022095-.
- [6] Velasquez N, Anani A, Munozgama J, et al. Towards the Application of Process Mining in the Mining Industry—An LHD Maintenance Process Optimization Case Study[J]. 2023.
- [7] Pasioka N, Lysenko N, Lysenko O, et al. Activating the Process of Educational Services Using Independent Computing Resources to Manage and Monitor the Quality of Learning[J]. 2021.
- [8] Umashankar A, Prabhu P. The application of vagus nerve stimulation in individuals with misophonia[J]. Neuroscience Research Notes, 2021. DOI:10. 31117/neuroscirn. v3i5. 105.
- [9] Konen B, Uzel N N, Elik D, et al. Multi-Criteria Decision Making in the Patent Process: Application in an R and D Firm[J]. Proceedings of the International Conference on Industrial Engineering and Operations Management, 2023.
- [10] Wei J, Guo J. Research on the Application of Data Mining Technology in Physical Training[J]. 2023.