scientific reports

Check for updates

OPEN

Web log mining techniques to optimize Apriori association rule algorithm in sports data information management

Tiantian Li, Fang Liu✉, Xiaobin Chen & Chao Ma

To optimize the current college sports data information management system, this study combines the Apriori association rule algorithm with web application development technology to upgrade the management system. Firstly, this study explores novel log mining techniques in genetic algorithms and web application development technology. Secondly, by integrating log mining techniques to optimize the Apriori algorithm, associations between sports data and information are discovered. Through the optimized algorithm, this study identifies key association rules of sports data information and validates the optimized system's reliability and effectiveness through experiments. The experimental results show that the running time of the traditional Apriori algorithm exponentially grows with the increase in information volume, while the optimized execution efficiency is improved by approximately 10–15%. Additionally, the average retrieval accuracy of this optimized system can reach 98.3%, although the retrieval time also increased by 23%. Therefore, the technology and algorithms proposed in this study have certain application value in the sports information management system and contribute to the optimization of data information management in this field.

Keywords Web log mining technology, Apriori algorithm, Sports data, Information management, Retrieval time

With the proliferation of information technologies, institutions of higher education have increasingly adopted the sports data information management system. A college sports data information management system encompasses the collection, analysis, and administration of sports-related data within academic institutions, facilitated by computer technology, network technology, and other information technology tools[1]. The system can collect and store all data related to college sports, encompassing athlete and coach profiles, competition results, and training conditions[2]. Leveraging web log mining technology in conjunction with the Apriori association rule algorithm can collect user behavior data through web log mining technology. Then, the Apriori algorithm is utilized to conduct frequent itemset mining and association rule mining, thereby elucidating relationships between different information and enhancing the retrieval accuracy and efficiency of the system[3].

In the research on information management systems, Svacina et al. (2020) combined university laboratory safety management systems with natural ecosystems to construct a laboratory ecological safety management system covering ecological subjects, ecological objects, and ecological environments. They used fault tree analysis methods to identify risk factors in the system. Additionally, they constructed a fault tree model for ecological safety accidents in university laboratories[4]. Rak and Żyła (2022) first introduced key technologies and related concepts, analyzed the system's requirements from the perspectives of system functionality and performance, and defined the main functional modules of the system. Then, they designed the system's technical architecture and provided corresponding database designs[5]. In the study of the Apriori association rule algorithm, Su and Wu (2021) proposed the enhanced frequent pattern-Apriori algorithm based on closed item sets. Experimental results showed that compared to the classical Apriori algorithm, the improved algorithm achieved an average performance improvement of 290% on smaller data scales with diverse increments and support levels, without reducing the number of correctly mined frequent item sets[6]. Wang et al. (2023) modeled project features, user behavior characteristics, and content features using neural network algorithms. Additionally, based on the neural network algorithm, they constructed a relevant entrepreneurial project recommendation and resource optimization model and conducted evaluation and analysis[7].

Department of Sports and Arts, Bengbu Medical University, Bengbu 233000, Anhui, China. ✉email: 0200116@bbmc.edu.cn

Although there have been studies exploring the application of association rule algorithms in various fields, there is still insufficient in-depth research specifically in domains like sports data information management. This leads to existing methods potentially not being fully applicable to the characteristics of sports data, such as timeliness and diversity. Additionally, the traditional Apriori algorithm, especially when dealing with large-scale data or significant data increments, may exhibit unclear performance improvements and may encounter inefficiency issues. Furthermore, many studies have not fully considered the scalability of the algorithm, making it difficult to cope with rapidly changing data environments. This study extends conventional research methodologies by integrating innovative log mining technology into web application development technology, thereby enhancing system performance and sports data processing capabilities while gaining insights into students' browsing habits and preferences. Additionally, through the utilization of an optimized Apriori algorithm, correlations within sports data information are identified, leading to improvements in retrieval accuracy and efficiency, and facilitating administrators' access to detailed system information. Ultimately, the reliability and efficacy of the optimized system are empirically validated through experimentation. The contribution of this study lies in the first-time integration of the Apriori association rule algorithm with web application development technology to optimize the sports data information management system. Particularly, by introducing log mining techniques from the genetic algorithm (GA), this study achieves effective optimization of the Apriori algorithm. This integration offers a new solution for sports data management, and through the optimization of the Apriori algorithm, this study successfully improves the algorithm's execution efficiency and its capability to handle large-scale data.

Optimization of sports data information management system based on web log mining and Apriori algorithm

Application of web log mining technology in data informatization

When a user interacts with the server, various logs are generated, capturing essential information such as the user's IP address and the nature of their request execution. These logs are critical in analyzing user behavior and optimizing websites[8,9]. Central to this process is web log mining, a fundamental task involving the preprocessing of log documents, encompassing data cleaning, integration, transformation, and specification[10]. This preparatory phase yields a refined, simplified, and more precise dataset, conducive to further analysis and mining. In Web log mining, the primary task is to analyze logs to obtain users' page access patterns, such as access time, frequency, and source[11]. This information enables a comprehensive understanding of user behavior, including their preferred pages and functions, and may even unveil latent needs and desires. The analysis can optimize the website, meet user requirements, and improve the user experience.

Web log mining technology finds crucial applications in security analysis and performance optimization[12]. By analyzing the log files, anomalous login attempts and potential security breaches can be detected, safeguarding website integrity. Concurrently, log mining facilitates the identification of performance bottlenecks within websites, enabling timely configuration adjustments to enhance access speed and responsiveness[13–15]. This technology plays a pivotal role in understanding users' needs, optimizing website experience, ensuring security, and improving performance. By scrutinizing log documents, insights into user behavior can be gleaned, leading to website enhancements and improved user satisfaction[16–18]. Before data mining, the original data is often disorganized and unsuitable for direct analysis. Data mining is divided into three stages: data preparation, data preprocessing, and data mining and result display. Data preparation should follow the principles of result-oriented, focusing on the intrinsic meaning of the data, and improving algorithm efficiency[19]. This requires selecting appropriate data according to the research purpose and excluding anomalous data to mitigate misleading outcomes and reduce processing time. Algorithmic efficiency is further bolstered through strategies such as distributed processing or increased concurrency, ensuring shortened processing cycles without compromising accuracy. There are often problems in actual data, such as attribute mismatches, null values, out-of-range values, or type mismatches, which affect data integrity and accuracy[20]. Hence, data preprocessing is a necessary step. Once the initial two stages are completed, refined and valuable data is obtained, paving the way for the delineation of analysis objectives based on the data or identified issues, thereby guiding the pursuit of desired results and knowledge.

The application of the Apriori association rule algorithm and GA in sports data management systems

Association rules play a pivotal role in uncovering implicit connections among multiple entities, akin to the butterfly effect, thereby elucidating correlations between various elements[21–23]. Through association rule analysis, insights such as "the presence of certain attributes coincides with the presence of others" or "the occurrence of specific events triggers subsequent events" can be derived. The Apriori algorithm employs an iterative approach to achieve this goal. It begins by exploring all candidate 1-item sets and obtains frequent -item sets through pruning. Subsequently, it iteratively connects and prunes to uncover frequent 2-item sets, and so forth, until larger item sets cannot be formed. Finally, the frequent k-item sets derived from the Apriori algorithm constitute the analysis's conclusion sets. Table 1 presents the sports scores of four college students.

Mining association rules constitutes a crucial data mining task aimed at extracting valuable information from extensive datasets[24–27]. It encompasses two primary subtasks: firstly, the frequent item sets are filtered from the cleaned dataset, and the memory consumption and algorithm complexity are reduced by mining the low-frequency items and their relationships[28–30]. Secondly, it involves the analysis of these frequent item sets to unearth potential relationships among the data, thus revealing association rules. Through statistical, machine learning, and artificial intelligence techniques, the data are analyzed, and the laws and trends behind them are mined. This is helpful to deeply understand the data and find its practical applications. Association rule mining

Student ID	Projects with individual scores above 90
1	High jump, Long jump, Basketball
2	High jump, Soccer, Tennis
3	High jump, Long jump, Basketball, Baseball
4	Long jump, Basketball

Table 1. Sports scores of 4 college students.

is widely used in e-commerce, medical care, social network analysis, and other domains, albeit encountering challenges such as managing large datasets, algorithmic efficiency, data quality, and privacy preservation.

In the sports data management system, the adoption of GA demonstrates notable efficacy in optimizing the computational overhead, especially in the key task of generating frequent item sets. GA, mimicking natural selection and genetic mechanisms, has been successfully deployed to tackle various complex problems, including the generation of frequent item sets in data mining. The sports data management system involves a lot of data, ranging from athlete performance metrics to competition results and team statistics. In such a system, the generation of frequent item sets emerges as a fundamental and imperative task, entailing the identification of recurrent data patterns within vast datasets. While traditional methods like the Apriori algorithm prove effective in certain scenarios, they often falter when confronted with colossal and intricate datasets, leading to computational challenges.

Optimization of sports data information management system

In the traditional algorithm, despite the utilization of connection and pruning steps for filtering, it remains necessary to traverse every transaction in the database and tally every candidate item sets containing the item when handling frequent item sets[31]. Given the typically substantial scale of databases, frequent database operations tend to consume considerable memory and performance. In addition, the algorithm produces numerous invalid candidate sets in the loop traversal, failing to exclude the elements that should not participate in the combination. Optimization measures primarily focus on three aspects. Firstly, database compression is employed[32]. By deleting or marking entries during database traversal, redundant inspections are circumvented, and unnecessary operations are minimized. Secondly, dynamic reduction of candidate item sets is implemented, whereby candidate item sets meeting the minimum support threshold are directly incorporated into the frequent item sets. Finally, item sets are pre-filtered, and the unqualified item sets are deleted before the connection step to mitigate the invalid calculations. The introduction of GA can further enhance the efficiency of this process. By simulating natural selection and genetic mechanisms, GA can adopt a global search strategy when searching for the optimal solution. When dealing with frequent item sets, GA iteratively improves the solution through crossover and mutation operations, thereby effectively reducing computational and search costs while enhancing mining accuracy. In summary, the existing sports data information management system is optimized, and the optimized system structure is displayed in Fig. 1.

The purpose of database compaction is to reduce duplicate checks and unnecessary operations when traversing the database, which can be achieved by deleting or marking entries. The compressed database is calculated as shown in Eq. (1):

$$T' = \{t' \mid t' = \mathrm{compress}(t), \forall\, t \in T\} \tag{1}$$

T' and T represent the compressed and original database; t refers to the transaction; $\mathrm{compress}$ denotes the calculation function; t' means the compressed transaction format. When generating a frequent itemset, the performance of the algorithm can be optimized by dynamically resizing the candidate set. The process for adding a frequent itemset is as follows:

$$F_k = F_{k-1} \cup \{c \mid c \in C_k, sup(c) \geq min_sup\} \tag{2}$$

F_k means the frequent itemset; c indicates the candidate set; sup represents the support degree function. The unqualified item set is deleted through the pre-filtering operation before the connection step, to reduce invalid calculations. A filter function is defined as:

$$C'_k = \mathrm{filter}\,(C_k) = \{c \mid c \in C_k, \mathrm{some_condition}(c)\} \tag{3}$$

C'_k indicates the filter function; filter stands for the calculation function; $\mathrm{some_condition}$ refers to the filter condition. In connection with the operation process of web log mining, the designed system function structure is classified into five parts. The specific content is given in Fig. 2.

Figure 2 illustrates that the user management module is responsible for distinguishing between ordinary users and administrators in the system, and for managing data access and operations differently based on their respective permissions. Subsequently, the data import module imports log data from multiple websites through task management functionality, ensuring the initial integration and availability of data. After data importation, the data preprocessing module cleans and organizes the data as necessary, including user identification, session identification, and transaction tracking, preparing the data for mining and analysis. The core management rule mining module utilizes association rule algorithms to discover valuable information and patterns from the

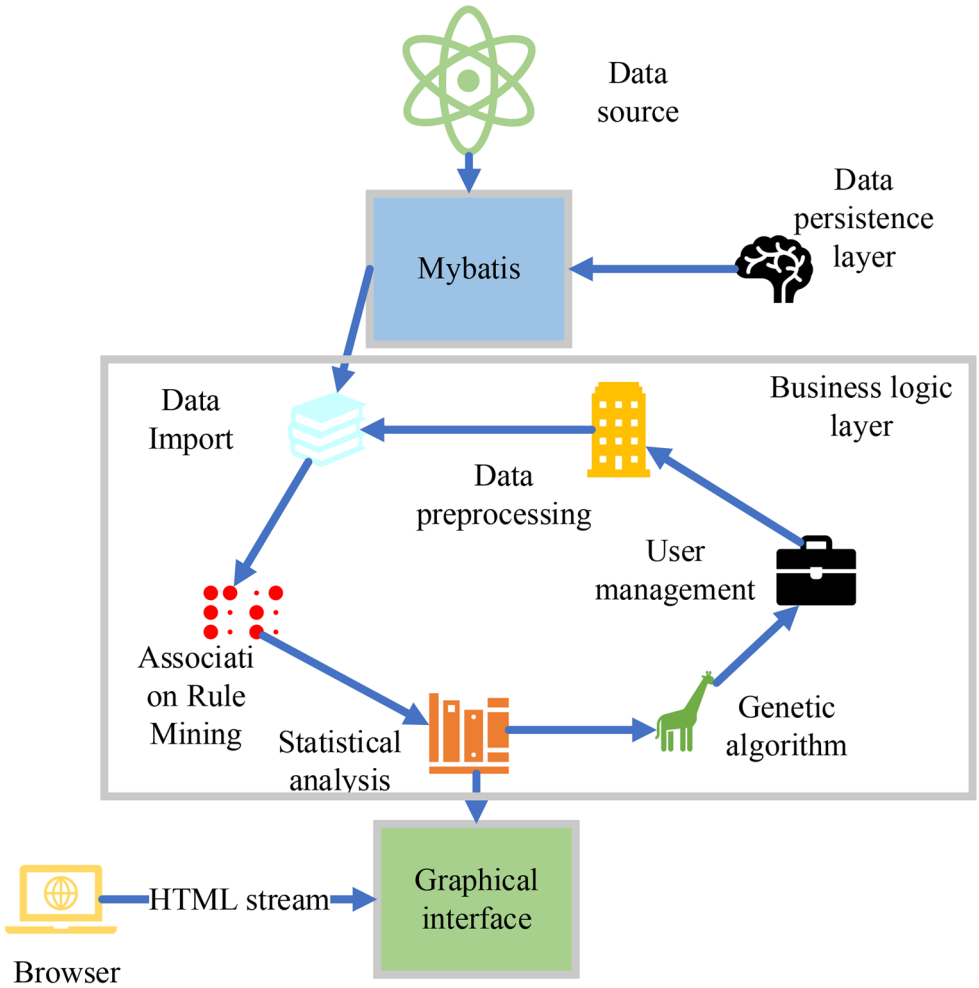


Figure 1. Optimized sports data information management system.

preprocessed data, which is a critical step in system implementation. Finally, the statistical analysis module is responsible for statistically analyzing key performance indicators such as daily website traffic and user page visits to assess the operational status of the website and user behavior trends. The collaborative work of these modules enables the entire system to efficiently process and analyze large amounts of sports data, thereby enhancing the quality and efficiency of information management. Data preprocessing stands as a pivotal step preceding data mining, given that data sourced from diverse origins often harbor inaccuracies, inconsistencies, or missing values, termed "dirty data." Utilizing such data directly for mining purposes risks yielding inaccurate and misleading outcomes. Hence, it is necessary to convert these dirty data into usable transactional data to meet the requirements of mining algorithms and remove data with no reference value. Preprocessing procedures encompass data cleaning and transformation, with methods tailored to specific application contexts and data formats. Especially in association rule mining, transitioning original data into a relational database constitutes a critical juncture. Association rule mining necessitates transactional data format, wherein item occurrences are recorded in transaction units, whereas relational databases usually store data in tabular form. In the process of data transfer, the application of data cleaning is essential to mitigate noise, redundancy, and missing values, thereby mitigating the influence of dirty data. Subsequently, the cleaned data requires conversion into transactional data and storage within the relational database, while preserving data integrity and consistency. To sum up, preprocessing is the basis of effective data mining, notably in association rule mining, wherein correct data preprocessing and transformation are paramount to facilitating the extraction of valuable insights from the data.

System performance analysis under the optimization of web log mining technology and Apriori algorithm

Comparison and analysis of performance before and after optimization of the Apriori association rule algorithm

The dataset used in this study is sourced from Kaggle, covering multiple domains including education and sports. This allows researchers to access data related to students' athletic performance, health tests, or other sports-related information, which is essential for researching the sports data information management system.

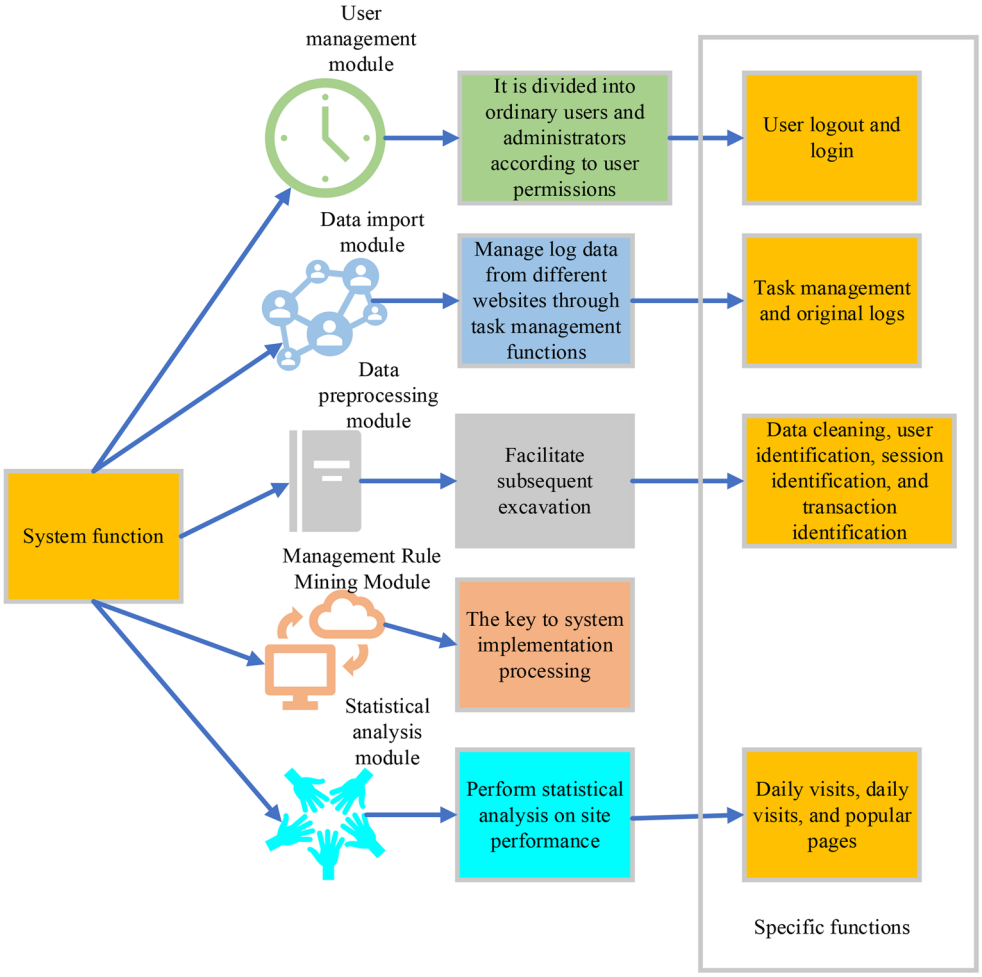


Figure 2. System functional structure diagram.

Datasets on Kaggle are typically provided by community members or collaborating organizations, ensuring data quality and diversity. Researchers can access sports data from diverse regions and backgrounds, which helps increase the breadth and depth of research. The datasets can be downloaded through the official website (https://www.kaggle.com/). The experimental environment is as follows: the operating system is Windows 10, the internal storage is 4GB, the CPU is 2.0 GHz, and the memory is 500G. The algorithms before and after optimization are employed to mine varying dataset sizes, and the time taken to generate frequent item sets and conduct association rule mining is recorded. After running the algorithms ten times, the average runtime is computed, and the experimental findings are revealed in Fig. 3.

Figure 3 denotes that before algorithm optimization, the running time of the Apriori algorithm exhibited an exponential increase with dataset size, leading to suboptimal performance in processing large-scale data and failing to meet practical requirements. After optimization, it can be found that the optimized algorithm markedly enhances the execution efficiency. Moreover, for a data volume of 100, the traditional algorithm requires 0.73s to generate frequent item sets, whereas the optimized algorithm takes 0.15s. However, when the processed data is 2,000, the traditional algorithm necessitates 15.76 s, while the optimized algorithm only takes 3.58s. As data volume increases, the advantages of the optimized model become increasingly apparent. Notably, the optimized algorithm demonstrates a minimum enhancement of 10–15% in execution efficiency. This underscores the effectiveness of the optimized Apriori algorithm in processing large-scale data.

Horizontal comparison of experimental results and analysis of sports data information management system

For the experiments, data from "21 Places to Find Free Datasets for Data Science Projects" is utilized. This dataset, provided by the University of Washington, comprises free datasets for data science projects and is accessible via the university's official website (https://careers.uw.edu/blog/2021/10/05/21-places-to-find-free-datasets-for-data-science-projects-shared-article-from-dataquest/). The sample set is divided into five subsets, each with 2000 students' sports data. It includes Binary Search Algorithm (BSA), Linear Search Algorithm (LSA), Binary Search Tree Algorithm (BSTA), Hash Table Algorithm (HTA), and Breath-First Search Algorithm (BFSA). The results of retrieval accuracy and retrieval time of the system are suggested in Fig. 4:

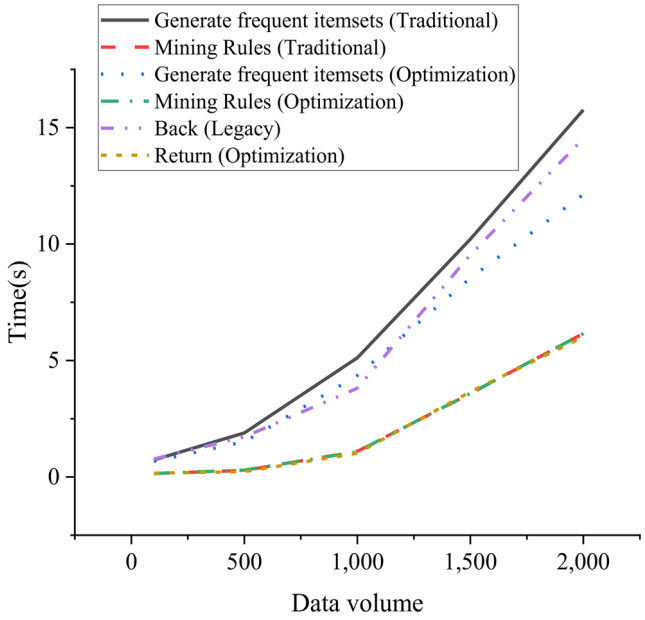


Figure 3. Performance comparison before and after algorithm improvement.

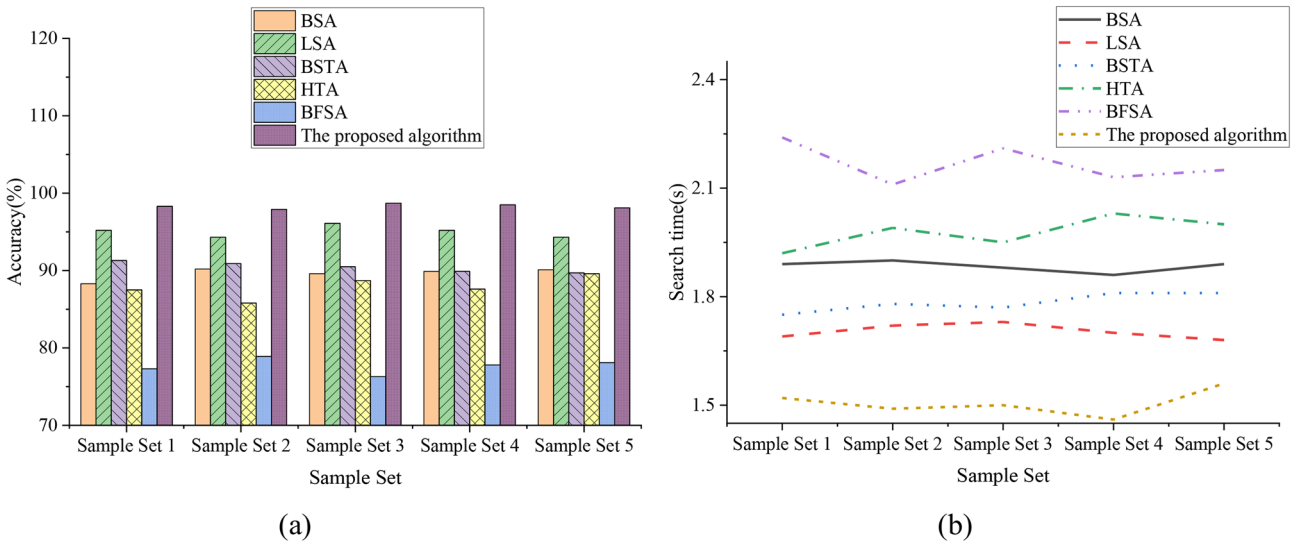


Figure 4. The results of system retrieval accuracy and retrieval time (**a**): Retrieval accuracy; (**b**): Retrieval time.

According to the experimental results in Fig. 4, compared with the traditional model, the proposed algorithm achieves the highest retrieval accuracy of 98.5%, signifying its ability to accurately locate user-required information. Furthermore, after many experimental tests, the average retrieval accuracy of this algorithm is 92.3%. This result underscores the algorithm's efficacy and application value in information retrieval tasks, meeting user's demands for accurate and comprehensive information. Regarding retrieval time, the proposed algorithm outperforms other models, exhibiting the shortest retrieval time and the highest stability. Analysis of experimental data reveals an average retrieval time of 1.51 s for this algorithm, approximately 23% faster than alternative models. The optimized system is deployed on a small scale at University A, where users rate their experience after each use on a scale of 1 to 4, indicating satisfaction levels from low to high. During the usage period, a total of 976 users provided ratings. The results are outlined in Table 2:

According to the scoring results, 76.84% of users are satisfied with the new system. This substantial satisfaction rate underscores the considerable enhancements and advantages of the new system across multiple facets.

Score	Number of people
Very satisfied: 4	219
satisfied: 3	531
General: 2	203
Unsatisfied: 1	23

Table 2. User satisfaction score.

Conclusion

With the continuous development of science and technology, Web log mining technology and the Apriori association rule algorithm are integrated into the sports data information management system. This study leverages GA, Apriori association rule algorithm, and Web application development technology to optimize and upgrade the management system. Firstly, a Web application development technology based on log mining technology is introduced. Secondly, the Apriori algorithm is refined through log mining integration to uncover the correlation between sports data and information, thus improving the retrieval accuracy and retrieval time. Finally, experimental verification is conducted, confirming the reliability and effectiveness of the optimization system. The experimental results demonstrate that the proposed algorithm enhances execution efficiency by at least 10%~15%, underscoring its robust performance in handling vast information volumes. Compared with the traditional management system, the proposed algorithm remarkably improves the information retrieval time and accuracy, with an average retrieval accuracy of 98.3% and a 23% reduction in retrieval time. This improvement is attributed to the algorithm's utilization of association techniques to enhance information correlations, thus shortening the retrieval time while improving the accuracy. Therefore, the proposed algorithm offers distinct advantages in accelerating information retrieval speed and holds significant practical application value.

However, several shortcomings exist. Firstly, there is considerable scope for enhancing the efficiency of algorithm selection and optimization within the association rule analysis algorithm. Particularly, efforts to enhance the Apriori algorithm should focus on minimizing loop traversal in frequent itemset discovery and construction. Secondly, the sample size is relatively limited, with only 1,500 data points selected in this study. Given that real-world datasets typically exceed this quantity, future studies will expand the sample set to bolster credibility further.

Data availability

All data generated or analyzed during this study are included in this published article [and its supplementary information files].

Received: 16 August 2023; Accepted: 26 September 2024
Published online: 15 October 2024

References

1. Di Vaio, A., Palladino, R., Pezzi, A. & Kalisz, D. E. The role of digital innovation in knowledge management systems: a systematic literature review. *J. Bus. Res.***123** (2), 220–231 (2021).
2. Dwivedi, Y. K. et al. Impact of COVID-19 pandemic on information management research and practice: transforming education, work and life. *Int. J. Inf. Manag.***55** (7), 102211 (2020).
3. Chayanukro, S., Mahmuddin, M. & Husni, H. Understanding and assembling user behaviours using features of Moodle data for eLearning usage from performance of course-student weblog. *Journal of Physics: Conference Series. IOP Publishing***1869**(1), 012087 (2021).
4. Svacina, J. et al. On vulnerability and security log analysis: a systematic literature review on recent trends. *Proc. Int. Conf. Res. Adapt. Convergent Syst.***5** (2), 175–180 (2020).
5. Rak, T. & Żyła, R. Using Data Mining techniques for detecting dependencies in the Outcoming Data of a web-based system. *Appl. Sci.***12** (12), 6115 (2022).
6. Su, Y. S. & Wu, S. Y. Applying data mining techniques to explore user behaviors and watching video patterns in converged IT environments. *J. Ambient Intell. Humaniz. Comput.***5** (1), 1–8 (2021).
7. Wang, Z., Deng, Y., Zhou, S. & Wu, Z. Achieving sustainable development goal 9: a study of enterprise resource optimization based on artificial intelligence algorithms. *Resour. Policy*. **80**, 103212 (2023).
8. Abd El-Aziz, A. A., Pandian, P. S., Almuayqil, S. N. & Alruwaili, A. S. A. Framework for Clustering & enhanced Approach for frequent patterns in web usage mining. *J. Adv. Sci.* **32** (5), 77–79 (2020).
9. Ibrahim, K. K. & Obaid, A. J. Fraud usage detection in internet users based on log data. *Int. J. Nonlinear Anal. Appl.***12** (2), 2179–2188 (2021).
10. Obaid, A. J., Ibrahim, K. K., Abdulbaqi, A. S. & Nejrs, S. M. An adaptive approach for internet phishing detection based on log data. *Periodicals Eng. Nat. Sci.***9** (4), 622–631 (2021).
11. Naqvi, R. et al. The nexus between big data and decision-making: A study of big data techniques and technologies. Proceedings of the International Conference on Artificial Intelligence and (AICV Cham). *Springer International Publishing***7**(2), 838–853 (2021).
12. Landauer, M., Skopik, F., Wurzenberger, M. & Rauber, A. System log clustering approaches for cyber security applications: a survey. *Computers Secur.***92** (13), 101739 (2020).
13. Manikandan, R. & Saravanan, V. A novel approach on particle Agent Swarm optimization (PASO) in semantic mining for web page recommender system of multimedia data: a health care perspective. *Multimedia Tools Appl.***79** (1), 3807–3829 (2020).
14. Wu, S. Research on the application of spatial partial differential equation in user-oriented information mining. *Alexandria Eng. J.***59** (4), 2193–2199 (2020).
15. Leno, V. et al. Identifying candidate routines for robotic process automation from unsegmented UI logs. 2020 2nd International Conference on Process Mining (ICPM). *IEEE***5**(2), 153–160 (2020).
16. Jalal, A. A. & Ali, B. H. Text documents clustering using data mining. *Int. J. Electr. Comput. Eng.***11** (1), 5–11 (2021). (2088–8708).

17. Bedi, P., Goyal, S. B., Rajawat, A. S., Shaw, R. N. & Ghosh, A. A framework for personalizing atypical web search sessions with concept-based user profiles using selective machine learning techniques. Advanced Computing and Intelligent Technologies: Proceedings of ICACIT 2021. *Springer Singapore***9**(3), 279–291 (2022).
18. Yu, H. Online teaching quality evaluation based on emotion recognition and improved AprioriTid algorithm. *J. Intell. Fuzzy Syst.***40** (4), 7037–7047 (2021).
19. Bach, R. L. & Wenz, A. Studying health-related internet and mobile device use using web logs and smartphone records. *PloS One*. **15** (6), 234663 (2020).
20. Zhang, J. & Dong, L. Image monitoring and management of hot tourism destination based on data mining technology in big data environment. *Microprocess. Microsyst.***80** (6), 103515 (2021).
21. Guo, Y., Wang, N., Xu, Z. Y. & Wu, K. The internet of things-based decision support system for information processing in intelligent manufacturing using data mining technology. *Mech. Syst. Signal Process.***142** (5), 106630 (2020).
22. Wang, H. B. & Gao, Y. J. Research on parallelization of Apriori algorithm in association rule mining. *Procedia Comput. Sci.***183** (27), 641–647 (2021).
23. Cong, Y., Research on data association rules mining method based on improved apriori algorithm. International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE). *IEEE***11**(1), 373–376 (2020).
24. Wang, C. & Zheng, X. Application of improved time series Apriori algorithm by frequent itemsets in association rule data mining based on temporal constraint. *Evol. Intel.***13** (1), 39–49 (2020).
25. Ren, X. & Computing, M. Application of Apriori Association Rules Algorithm to Data Mining Technology to Mining E-commerce Potential Customers. 2021 International Wireless Communications and (IWCMC). *IEEE***9**(5), 1193–1196 (2021).
26. Fauzan, F., Nurjanah, D. & Rismala, R. Apriori association rule for course recommender system. *Indonesia J. Comput. (Indo-JC)*. **5** (2), 1–16 (2020).
27. Mehta, A. & Bura, D. Mining of association rules in R using Apriori algorithm. Advances in Communication and Computational Technology: Select Proceedings of ICACCT 2019. *Springer Singapore***13**(2), 181–188 (2021).
28. Wicaksono, D., Jambak, M. I. & Saputra, D. M. The comparison of apriori algorithm with preprocessing and FP-growth algorithm for finding frequent data pattern in association rule. Sriwijaya International Conference on Information Technology and Its Applications (SICONIAN 2019). *Atlantis Press***3**(3), 315–319 (2020).
29. Sharma, A. & Ganpati, A. Association rule mining algorithms: a comparative review. *Int. Res. J. Eng. Technol.***8** (11), 848–853 (2021).
30. Xie, H. Research and case analysis of apriori algorithm based on mining frequent item-sets. *Open. J. Social Sci.***9** (04), 458 (2021).
31. Mohapatra, D. et al. Interpretation of Optimized Hyper Parameters in Associative Rule Learning using Eclat and Apriori. 5th International Conference on Computing and (ICCMC). *IEEE***4**(1), 879–882 (2021).
32. Fale, P. N., Moundekar, N., RiteshSaudagar, P. K., Rode, M. & Borkar, J. Review on optimization of Apriori Algorithm for finding the Association rules in different business and other datasets for Retrieval of relations between different entities. *Int. J. Sci. Res. Sci. Eng. Technol.* **3** (1), 1096–1098 (2022).

Acknowledgements

1. Fund Project: Key research project of Humanities and Social Sciences in Colleges and universities of Anhui Province, construction and empirical research of competency model of physical education teachers in Colleges and universities of Anhui Province, SK2021A0446;
2. Anhui quality engineering project, sports club teaching team, 2021jxtd183.

Authors' contributions

Tiantian Li, Fang Liu, and Xiaobin Chen contributed to conception and design of the study. Chao Ma organized the database. Tiantian Li performed the statistical analysis. Fang Liu wrote the first draft of the manuscript. Tiantian Li, Fang Liu, and Xiaobin Chen wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-74427-z.

Correspondence and requests for materials should be addressed to F.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2024, corrected publication 2024