Analysis of the Functions of Database Information Management System: Taking the User Health Record Platform as an Example

Ying Xiang[a]
Aberdeen Institute of Data Science and Artificial Intelligence, South China Normal University, Taoyuan East Road, Foshan City, China

Keywords: Database, Data Security, Data Prediction.

Abstract: Mental illness is recognized as an important public health problem globally. The disease is accompanied by a huge number of complications and often masks the real psychological problems, causing a lot of losses in manpower, material and financial resources. And currently, medical data often has biases and imbalances. Certain types of case data are sparse, which can easily lead to errors in prediction and diagnosis by algorithms. At the same time, it is important to ensure that patients provide personal data and protect privacy rights. This article concludes that the use of databases can assist doctors to accurately diagnose the corresponding mental illness from a large number of complications faster and provide timely treatment. This article mainly uses a relational database based on the basic functions of the database to conduct research on security and prediction accuracy. Through discussion, this article combines access control and Bayesian networks with databases to make the prediction results in practical applications safer and more accurate and provides relevant ideas for combining databases with disease research.

1 INTRODUCTION

Mental illness is considered an important public health problem globally, with a huge number of accompanying complications that often mask the real psychological problems, causing a lot of losses in human, material and financial resources. As an efficient data management system, the database can store a large amount of relevant data and predict new situations based on past data. Therefore, database management can provide researchers with comprehensive and accurate case records and treatment data, helping to reveal the characteristics and evolution of the disease.

The database not only provides efficient data management but can also be used to store patients' personal information, past medical records, treatment plans, treatment results and other related information. Through the database, medical staff can further analyze the patient by sorting the data to ensure accurate treatment and predict the possibility of recurrence of the disease, thereby making targeted recommendations. At the same time, medical staff can also search for similar cases based on keywords in order to use past successful cases and experience to treat new patients accordingly, reducing the possibility of incorrect diagnosis and improving diagnosis and treatment efficiency. Healthcare professionals also need to take data security seriously, especially when handle personal patient information, to ensure the accuracy and trustworthiness of research on mental illness. However, in the era of big data, ensuring the security of patients' personal data and protecting the privacy of users is particularly important. Moreover, current medical data often have biases and imbalances, and there are less data on certain types of cases, which can easily lead to errors in prediction and diagnosis by algorithms, resulting in lower credibility of prediction results. Therefore, it is particularly important to build a database with high security performance and accurate prediction capabilities.

In order to predict mental illness more accurately, the international cooperation organization Cochrane Collaboration published a medical literature database called Cochrane Library, which contains a large number of systematic reviews of clinical trials of mental illness; the American Psychological

[a] https://orcid.org/0009-0000-6036-7064

Xiang, Y.
Analysis of the Functions of Database Information Management System: Taking the User Health Record Platform as an Example.
DOI: 10.5220/0012916100004508

In *Proceedings of the 1st International Conference on Engineering Management, Information Technology and Intelligence (EMITI 2024)*, pages 159-163
ISBN: 978-989-758-713-9

Association (APA) operates a database called PsycINFO The psychology literature search database contains a large number of studies on mental illness. These databases provide valuable support and reference for mental illness research. At the same time, some hospitals have also adopted related decision-making systems, such as decision trees, logistic regression support, etc., to establish platforms to screen conditions. However, due to the characteristics of mental illness and the current low popularity of mental illness-related knowledge, there is less reliable data available. Therefore, the research on the combination of this field with databases is relatively blank and has high research value.

This article explores how databases can be designed and used to address this issue and ensure the protection of patient privacy. This article focuses on the functions and functions of the database. Then, this article mainly analyzes how to improve the security of the database and the accuracy of diagnosis and treatment of mental illness through reasonable algorithms and methods.

2 BASIC FUNCTION INTRODUCTION

A database is an important tool for storing data. It has a variety of functions that can adapt to the needs of different application scenarios. The main functions of the database include data storage, data retrieval, data security, data analysis and reporting.

Data storage is the basic function of the database. Organize corresponding information by establishing tables and inserting data to store structured and unstructured data orderly. Databases can also store a variety of data, such as numbers, text, dates, images, audio and video. At the same time, database administrators can complete data insertion, update, and deletion operations, either individually or in batches, to meet different business needs. In addition, the database can also ensure data consistency through transaction management and ensure data integrity in multiple operations.

Secondly, the database has powerful data retrieval capabilities. Users can use query languages such as SQL to filter, sort, group, and other data based on specific conditions to quickly obtain the required information and facilitate data use and extraction.

The database also provides a variety of mechanisms to ensure data security, such as access control, user authentication, and data encryption. By setting permissions, users can restrict their access to and operations on data, prevent unauthorized access and data leakage, and facilitate data management. At the same time, databases can use advanced encryption techniques such as secure hashing algorithms (SHA) to generate unique identifiers in a set of variables, and by including a reasonable number of variables to identify unique patterns, it is possible to compare different data sets. Generate an anonymous unique identifier and merge them (Marco et al., 2023). In addition, the database can back up data regularly to prevent data loss, and in the event of data damage or loss, the data can be restored to its most recent state through the recovery function to prevent data loss to the greatest extent.

Moreover, using the easy-to-understand OLAP paradigm for multi-dimensional big data analysis, and through its flexibility and expressive capabilities, helps people discover the relationship of data from different angles and analyze the data to help users better understand the data and make better decisions. decision-making (Alfredo, 2023). At the same time, the database can also predict future data based on other algorithms, allowing people to take better measures for the future or improve current functions.

To sum up, the database, as a powerful data management tool, has a variety of functions to meet the needs of different application scenarios. With the development of technology, the functions of the database are constantly enhanced and improved, providing better support for various application scenarios.

3 SPECIFIC FUNCTION INTRODUCTION

3.1 Data Security

With the rapid development of Internet technology and the popularity of network applications, the number of websites has shown explosive growth. These websites cover various fields, including but not limited to government departments, commercial enterprises, educational institutions, medical institutions and various social groups. They have built information exchange platforms by establishing websites to provide users with various services and information. Database systems play a vital role behind the scenes of these websites. They are responsible for storing and managing large amounts of user data, including personal information, transaction records, social interactions, etc. However, with the surge in the number of websites and the

expansion of data volume, ensuring that these websites can run safely and stably has become a major challenge for website administrators and database administrators.

The complexity and changeability of the network environment, the numerous loopholes in information systems, and the increasingly sophisticated hacker attack methods have put database security protection under tremendous pressure. In the design of databases, most databases use commercial database systems such as DB2 and SQL Server. Although these database systems have mature technical conditions, problems such as SQL injection may still occur. However, considering the stability of the database, the update speed of more complete patches is usually delayed, which ultimately makes it difficult to deal with database vulnerabilities the first time, and eventually becomes a security risk (Peng et al., 2022). Hackers often use system vulnerabilities to steal confidential data or destroy important data by invading databases to achieve their illegal purposes. This not only violates users' privacy rights, but also brings huge economic losses and reputational damage to website operators.

In order to deal with these security threats, database administrators have taken a variety of measures to protect data security, such as setting access control, data encryption, auditing and monitoring functions, SQL injection protection, etc., so that even if the data is illegally obtained, It also cannot be easily interpreted, preventing hackers from exploiting these vulnerabilities. At the same time, detection technology based on machine learning can also be used to train a machine learning model, so that the machine can distinguish between normal SQL statements and malicious SQL statements by learning algorithms such as decision trees, support vector machines, and neural networks, so that it can automatically Identify patterns of malicious behavior. Although this detection method requires a large amount of training data and computing resources, it can detect unknown attack methods and has high detection accuracy (Zhu et al., 2024).

In addition to the above measures, the database server is also protected by physical security. Such as installing surveillance cameras, using access control systems, establishing security barriers, etc., to protect the physical location of the database server and prevent unauthorized physical access.

In short, with the continuous development of network applications, database security has become an important issue that cannot be ignored. Through the comprehensive use of technology and management methods, the security of the database can be greatly improved, user data is protected from infringement, and the normal operation of the database is maintained.

3.2 Data Prediction

The prediction function of the database is based on existing data and uses data analysis and machine learning algorithms to predict future trends. Its steps mainly include data collection and cleaning, data analysis and modeling, model evaluation and optimization, and prediction and application. Human society generates a large amount of different types of data every day with people's social activities. If these large data sets are to be useful, new tools are needed to analyze them, which has led to the application of machine learning (ML). Liu et al. (2020) define ML methods as the practice of generating predictive models based on learning feature patterns from data, and predicting new data or results through the constructed models (Flores et al., 2023).

In addition, transfer learning and adding explainable artificial intelligence can also help database predictions become more accurate. Transfer learning can fully use the feature representation capabilities of existing models in new disease prediction by utilizing existing data and models, such as artificial learning network models, thereby improving the accuracy and generalization capabilities of the new model. Migration learning has a high data utilization rate. It can extract and use existing models' feature extraction capabilities to greatly reduce data analysis and annotation and improve data utilization. At the same time, it can avoid building a model from scratch. However, suppose there is a large difference between the original rules and the new rules. In that case, the effect of transfer learning may decrease, and some irrelevant features may be introduced, which may reduce the model's performance.

In actual use, while ensuring the reliability of predicted data, it is also necessary to provide users with a clear decision-making process, so explainable artificial intelligence needs to be added. Designers can use the SHAP model in game theory: Shapley value. In a game, participants cooperate to achieve certain benefits. The purpose of the Shapley value is to allocate each player's contribution to different ways of cooperation. For example, the Shapley value defines a reasonable distribution mechanism so that the benefits obtained by each participant are proportional to their contribution to the game (Lundberg et al., 2017). Database administrators can use this method to determine the data type that has a

greater impact on the final result and quickly find the corresponding relationship.

4 DATABASE DESIGN: TAKING THE MENTAL ILLNESS RECORDING PLATFORM AS AN EXAMPLE

4.1 Method of Data Storage

Usually, data storage methods are divided into two types of databases: relational and non-relational. Since different complications of patients have different characteristics, this article chose a relational database as the data storage method for this database. Relational databases can use relational algebra and powerful Structured Query Language (SQL) capabilities to filter information within the database, allowing doctors to solve complex queries through combined operations (Córdoba-Hidalgo et al., 2022). Therefore, doctors can use a patient's symptoms to search the database to see if other patients also have the same symptoms and find the best medical advice and treatment while identifying possible complications. Through this method, doctors can quickly locate the patient's symptoms, screen out other hidden diseases, strive for the best treatment time, and help the patient recover quickly.

4.2 How to Ensure Data Security

SQL injection attack is a relatively common database attack method. It is mainly based on the SQL language database system and uses vulnerabilities in web applications to inject attacks on the database to create, read, modify or delete sensitive data (Kasim, 2021). Since vulnerabilities in Web systems are unavoidable, Text Plan adopts access control to formulate policies based on the principle of least privilege to prevent relevant information from being directly exposed to users in order to deal with security threats such as SQL injection attacks and ensure the security of the database. This article uses role-based access control to classify different users, grant different permissions, and manage permissions. For example, doctors can query patient information and apply it to submit new case information, and only database administrators can add, modify, or delete cases. At the same time, when users log in, they also need to pass user authentication, such as entering user name, password, using multi-factor authentication, etc. to ensure that the identity of the login is authentic and reliable, so as to ensure the security of the database.

4.3 Data Prediction

By using Bayesian models for probabilistic inference, the impact of other diseases can be reduced and the condition can be determined more accurately.

Bayesian network is a graphical model used to represent dependencies between variables and provide explainable results through probabilistic inference (Koller et al., 2009). Bayesian networks can use the patient's symptoms, signs and examination results as nodes, various possible diseases as output nodes, and the dependencies between nodes are expressed as directed edges. The probability of each disease is provided by calculating the posterior probability, thereby providing prediction results for the patient's disease.

Patients often have more than one symptom, so a multi-feature fusion method based on Bayesian theory is applied. The performance and generalization ability of the model can be improved through feature combination, feature selection, etc., and the results can be further accurately predicted (Li, 2019).

Feature combination is to combine different features to generate new features. For example, a patient's two physical conditions, A and B, can be multiplied to obtain a new feature, AB. This captures the interaction between features and provides more information for analysis on the Bayesian network.

Feature selection selects the most representative and effective features among multiple features and assigns an important value or weight to the features for feature ranking (Kanti et al., 2021). People can use methods such as correlation, information gain, and chi-square tests to evaluate the degree of association between features and target variables, and select features with higher correlations for analysis to reduce negative impacts. For example, by assessing the correlation between the patient's various physical indicators and the current discomfort and by selecting highly relevant indicators, people can more quickly find the potential cause and provide rapid diagnosis and treatment.

5 CONCLUSIONS

This article conducts research and analysis on the functions of the database, protecting its security and increasing the accuracy of data prediction. Taking the database design of the mental illness recording platform as an example, a relational database is used

to facilitate users to quickly search for relevant information. At the same time, access control is used to prevent SQL injection attacks and reduce the exposure of key information to users to ensure database security. In terms of prediction, Bayesian models are used to conduct correlation analysis on various complications, and explainable artificial intelligence is added to facilitate doctors to confirm patients' symptoms faster. By calculating the probability to find the main cause of the disease, the accuracy and reliability of the prediction results are ensured while excluding complications and other erroneous data.

Although the Bayesian network assumes that each condition is independent, the model is relatively simple and computationally efficient and can avoid erroneous data through probabilistic analysis. However because of this assumption, the model struggles to make complete predictions. Moreover, this model requires a large number of training samples for data training, and there is still a lack of samples in this field, so it is still difficult for this model to be accurately applied in disease diagnosis.

Due to the small number of relevant references, it is difficult for this article to form an accurate feasibility analysis conclusion. At the same time, due to the lack of detailed understanding of the human nervous system and convolutional neural networks, it is still at a highly theoretical stage and is difficult to apply in specific practice. However, the Bayesian network proposed in this article can be used to analyze new cases in the database, provide ideas for subsequent research, and reduce the severe psychological burden on patients due to the impact of long-term complications and further aggravate mental illness.

REFERENCES

Alfredo, C,2023, Compressing Big OLAP Data Cubes in Big Data Analytics Systems: New Paradigms, a Reference Architecture, and Future Research Perspectives. Communications in Computer and Information Science, 1849, CCIS.

Córdoba-Hidalgo, P., Marín, N., & Sánchez, D,2022, RL-instances: An alternative to conjunctive fuzzy sets of tuples for flexible querying in relational databases. Fuzzy Sets and Systems, 445, 184‑206.

Flores, K. R., de Carvalho, L. V. F. M., Reading, B. J., Fahrenholz, A., Ferket, P. R., & Grimes, J. L,2023, Machine learning and data mining methodology to predict nominal and numeric performance body weight values using Large White male turkey data sets. Journal of Applied Poultry Research, 32(4), 100366.

Kanti Ghosh, K., Begum, S., Sardar, A., Adhikary, S., Ghosh, M., Kumar, M., & Sarkar, R,2021, Theoretical and empirical analysis of filter ranking methods: Experimental study on benchmark DNA microarray data. Expert Systems with Applications, 169, 114485.

Kasim, Ö,2021, An ensemble classification-based approach to detect attack level of SQL injections. Journal of Information Security and Applications, 59, 102852.

Koller, D., & Friedman, N,2009, Probabilistic graphical models: Principles and techniques. The MIT Press eBooks.

Li, Z. H,2019, To explore the underlying molecular mechanism of mental illness based on the comorbidity and multi-feature fusion Bayesian model [Doctoral dissertation, East China Normal University]. Retrieved from https://kns.cnki.net/kcms2/article/abstract?v=0Vs2Vpqj5wcperIKNbCrv7xQoET3WbTi_jdtg7TSc0X5bbDguOSggsDT8Mw9AJR_LOcfcdmEkWc2io7BTO4vA_dSqxFIsWj05h5KGpIAb2f9JX1cGunKAR85nPorN_OwG_1uX5Apb5U=&uniplatform=NZKPT&language=CHS

Lundberg, S., & Lee, S.-I,2017, A Unified Approach to Interpreting Model Predictions. arXiv: Artificial Intelligence.

Peng, X. M., & Hu, J,2022, Database Security System under the background of big data. Computer knowledge and technology, 09, 11-12.

Tomietto, M., McGill, A., & Kiernan, M. D,2023, Implementing an electronic public health record for policy planning in the UK military sector: Validation of a secure hashing algorithm. Heliyon, e16116.

Wang, F,2019, Digital Communications World. Journal of Digital Communication, 25(4), 256-268.

Wang, Z,2023, Modern industrial economy and informatization. Journal of Modern Industrial Economy, 35(1), 32-45.

Zhu, Z. N., & Jin, J. Q,2024, Research on attack technology based on SQL Injection vulnerability. Computer knowledge and technology, 01, 98-100 + 103.