

Research on the Application of Data Mining Technology in Software Engineering

Ruize Luo

Beijing University of Technology Beijing 100124

Abstract: With the development of computer science and software engineering, software systems are becoming larger and more complex in scale and function. How to effectively manage and utilize data during development, testing, and maintenance, improve software quality, reduce development costs, and increase productivity has become an important research topic in the field of software engineering. As an effective data analysis method, data mining technology has been widely used in the field of software engineering. Data mining technology can help software engineers mine useful information and knowledge from data, improve the quality and performance of software systems, reduce development costs, and accelerate the software development process. This article introduces the research status and development trend of applying data mining technology in software engineering. Firstly, it introduces the application scenarios and objectives of data mining in the field of software engineering, including defect prediction, demand analysis, and software quality evaluation. It discusses the research hotspots and future development trends of data mining technology in software engineering, including deep learning, interpretable data mining, and cross domain data mining.

Keywords: Software engineering; Data mining; Technology application

Introduction

With the constantly changing demands of new technologies in society, people have increasingly high requirements for software. Real software development requires both quantitative results and sufficient accuracy to meet the requirements of software development. Due to the advantages of wide application fields, mature application technology systems and processes, data mining technology has become a mainstream technology applicable to all aspects of software engineering development. However, in the specific process of software engineering development and application, it is necessary to clarify the mining objects and mining technologies, so as to promote the sustainable development of China's software engineering industry.

1. Overview of application of data mining technology in software engineering

In the initial stage of the emergence of data technology, people referred to it as knowledge discovery from a data repository, which is a database based on information technology for data mining. A new concept of software engineering emerged in the 1960s, with the core idea of using engineering methods to complete a series of software construction and maintenance tasks, and according to user requirements, within a specified time, produce high-quality software with practical application value. Over time, data mining technology has been increasingly applied to the software field, and has made significant progress. Data mining technology is to extract core data with important application value, higher application efficiency, and more new application forms from massive data to achieve unusual processing of software engineering data patterns. Based on the above concepts, the application characteristics of data mining technology in practice are summarized: ensuring the correctness and correctness of data sources; Ensure sufficient data capacity and avoid issues such as incompleteness and noise. The knowledge gained from data should be useful and possibly valuable. The application of technology does not require a lot of professional knowledge, as long as it can be used to analyze and solve specific problems; Used to obtain knowledge that users need and have not yet been understood by users.

2. Data mining applications in software engineering

2.1. Defect prediction

Defect prediction is one of the important issues in software engineering. Through data mining technology, it is possible to discover the rules and patterns of defects from historical data, predict possible future defects in software systems, and take appropriate measures in advance to avoid the occurrence of defects. Defect prediction is an important task in software engineering, which aims to identify potential defects in advance in order to repair them before software release to ensure software quality. Data mining technology has been widely used in defect prediction, which can effectively mine the rules and characteristics of defects, thereby improving the accuracy of prediction.

The process of defect prediction generally includes steps such as data collection, feature extraction, model construction, and evaluation. Data collection refers to the collection of various data generated during the software development process, such as the number of lines of

code, the number of modifications, and the number of submissions. Feature extraction refers to extracting defects related features from collected data, such as code complexity, developer experience, and so on. Model building refers to building prediction models based on extracted features. Common models include decision trees, support vector machines, and so on. Evaluation refers to the performance evaluation of the built model, such as accuracy, recall, etc. The application of data mining technology in defect prediction includes feature selection, classification, clustering, and so on. Feature selection refers to selecting features from all features that have a significant impact on the prediction results, in order to reduce the complexity of the model and improve the accuracy of the prediction. Common feature selection methods include correlation coefficient method, information gain method, chi-square test method, and so on. Classification refers to the classification of software instances into defective and non defective categories. Common classification algorithms include decision trees, naive Bayesian, support vector machines, and so on. Clustering refers to dividing software instances into multiple clusters. Instances within the same cluster have similar characteristics, and instances between different clusters have significant differences. Common clustering algorithms include K-Means, DBSCAN, and so on.

2.2. Demand analysis

The application of data mining technology in demand analysis can not only help developers better understand user needs, but also help optimize the process of demand analysis and improve the efficiency and quality of software development. Data mining technology can help software engineers mine user needs and behaviors from user feedback data, thereby better performing demand analysis and ensuring that software systems meet user needs. In the early stages of software development, requirements analysis is a crucial step. It involves multiple aspects of software system requirements collection, analysis, clarification, management, and verification. If the requirements analysis is not sufficient and accurate enough, it will lead to significant problems in later software design and development. Therefore, data mining technology can also be applied to requirements analysis.

Data mining techniques can be used in requirements analysis to collect and analyze requirements. By mining a large amount of user feedback, market research data, competitor information, etc., you can discover the potential needs and behavior patterns of users, helping to better understand their needs. At the same time, data mining technology can also discover the similarities and differences between different user needs, thereby helping to determine the needs of various types of users and providing guidance for software design and development. In requirements analysis, data mining techniques can also be applied to the classification and prioritization of requirements. By analyzing the correlation and weight between requirements and other requirements, requirements can be divided into different categories and the priority of each requirement can be determined.

2.3. Software quality assessment

Software quality assessment is one of the important issues in software engineering. Data mining technology can be used to evaluate the quality of software systems, evaluate their reliability, efficiency, security, and other indicators, help software engineers identify potential quality issues, and take timely measures to improve them.

Software quality assessment is a very important step in the software development process. It is to ensure that the final results of software development meet the needs and expectations of users, and also ensure the availability, stability, reliability, and other requirements of the software. Data mining technology can play an important role in software quality assessment. Firstly, data mining technology can be used for software defect analysis. In the process of software development, there are many defects that can have a significant impact on software quality if they cannot be detected and repaired in a timely manner. Using data mining technology, you can mine a large amount of defect data, analyze the causes and laws of defects, and propose corresponding defect prevention and repair plans based on these data and laws, thereby improving the quality of software.

Secondly, data mining technology can be used for software performance evaluation. Software performance is one of the important indicators for software quality evaluation, which reflects the operational efficiency and resource utilization of software. Using data mining technology, you can mine a large amount of performance data, analyze the impact of different factors on software performance, and provide guidance for software performance optimization. In addition, data mining technology can also be used for software user satisfaction evaluation. Software user satisfaction reflects the degree of user satisfaction with software, and is one of the important indicators for evaluating software quality. By using data mining technology, it is possible to analyze user behavior data, user feedback data, market research data, etc., thereby mining the key factors and laws of user satisfaction and providing guidance for improving user satisfaction.

Finally, software testing is one of the important means to ensure software quality. Through the use of data mining technology, it is possible to mine defect information in test data, analyze the causes and laws of defects, and provide reference and guidance for the formulation of test strategies and test plans. The application of data mining technology in software quality assessment can discover the root causes of software quality problems, improve software quality, reduce software development costs, and improve development efficiency.

3. Conclusion

In summary, due to the continuous development of data analysis technology, data analysis is increasingly widely used in software engineering. The development of software engineering is a relatively complex process, including many programs and development data that require real-time updates. Introducing data mining technology into software projects can efficiently integrate and process the data involved in the project, thereby providing technical support and data assurance for the implementation of the project. Through in-depth analysis of data mining technology, the application of data mining technology in software projects is revealed, providing an efficient and low-cost method for the construction of software projects, and making contributions to the sustainable development of software projects.

References

- [1] Desheng Dong. Application of Data Mining Technology in Software Engineering [J]. Information Recording Materials, 2021,22 (12): 165-167.
- [2] Zhiyun Liu. Analysis of the Application of Data Mining Technology in Software Engineering [J]. Computer Knowledge and Technology, 2021,17 (24): 27-28.
- [3] Binbin Wang. Research on the Application of Data Mining Technology in Software Engineering in the Era of “Internet +” [J]. Journal of Beijing Institute of Graphic Technology, 2021,29 (04): 148-151.

Author profile:Luo Ruize [March 31, 2001], male, native place: Zunyi City, Guizhou Province, ethnic group: Han Nationality, research direction: software engineering.