International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
Impact Factor: 5.22 (SJIF-2017), e-ISSN: 2455-2585
Volume 4, Issue 6, June-2018

Various Techniques of Data Mining- for Software Development Process

Nishu Sethi[1], Neha Bhateja[2]
[1,2]*Department of Computer Science, Amity University Haryana,*

Abstract--A typical software development process has several stages; each with its own significance and dependency on the other. Each stage is often complex and generates a wide variety of data. Using data mining techniques, we can uncover hidden patterns from this data, measure the impact of each stage on the other and gather useful information to improve the software development process. The insights gained from the extracted knowledge patterns can help software engineers to predict, plan and comprehend the various intricacies of the project, allowing them to optimize future software development activities. As every stage in the development process entails a certain outcome or goal, it becomes crucial to select the best data mining techniques to achieve these goals efficiently. In this paper, we survey the available data mining techniques and propose the most appropriate techniques for each stage of the development process. We also discuss how data mining improves the software development process in terms of time, cost, resources, reliability and maintainability.

Keywords— Data Mining, Software Engineering, Software Testing, Classification, Clustering.

I. INTRODUCTION

Data mining can dig out the hidden mineral resources — knowledge from magnanimity data. The demand to calculate and data increases continuously with the rapid development of computer technology. The size and complexity of software systems has increased dramatically. It is becoming more and more difficult to control and supervise the develop activity[2,3]. In software engineering area, traditional qualitative methods and simple statistical techniques is hardly to solve problems bring by explosive growth of data and information. It seems particularly important to find rules from testing data to steer the software development activity. Software testing data mining is to use existing technology or develop new algorithms to distill, analyze and express data. It is a process to dig out useful information and knowledge for software developers and managers.

II. SOFTWARE TESTING

Software testing is a process which provides information about the quality of the software. The objective of Software testing is to determine how well the evaluated software adapts to its specifications, in addition to its regaining reliability after amendments of the software[5].

Testing Methods:

Most common methods of Software Testing are

1. Static and Dynamic Testing
2. White- Box and Black- Box Testing

1.Static Testing: This review or inspect the source code itself, also called Verification Process.

Dynamic Testing: This inspects the actual execution of the programmed code with test cases, also called Validation.

2. White Box Testing: This uses the actual code of the tested program to perform analysis

Black Box Testing: This compares the program input against the output without taking into account internal workings.

III. DATA MINING

Data mining is an interdisciplinary field which involves methods of machine learning, statistics, database systems, and many other fields. Data Mining is defined as a process of knowledge discovery, which finds hidden information. The knowledge Discovery in Databases (KDD), is the overall process of translating raw data into useful information[4,7].

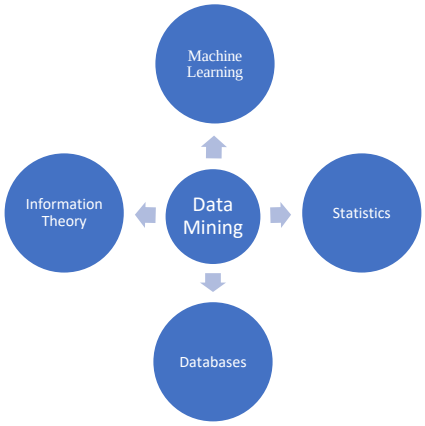


Fig1: Interdisciplinary fields of Data Mining

IV.DATA MINING TECHNIQUES

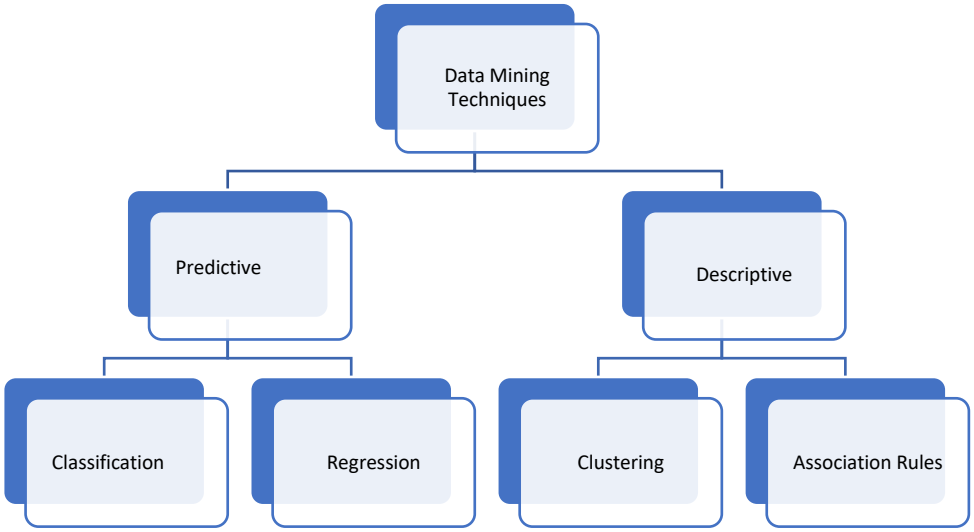
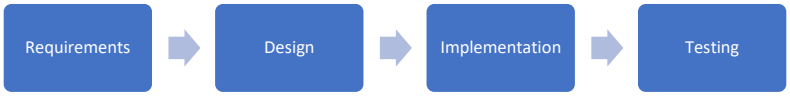


Fig 2: Various Data Mining Techniques

Descriptive miningtasks characterise properties of the data in a target data set. The scope of these tasks is to derive patterns , correlations, trends, anomalies, that summarise the underlying relationships in data [6]. Descriptive data mining tasks 7 are often exploratory in nature and frequently require post-processing techniques to validate and explain the results.

Predictive mining tasks perform induction on the current data in order to make predictions. The objective here is to predict the value of a particular attribute based on the values of other attributes[8,9]. The attribute to be predicted is commonly known as the target or dependent variable, while the attributes used for making the prediction are known as the explanatory or independent variables.

V. STAGES OF SOFTWARE DEVELOPMENT



Software Development Stage	Data Mining Techniques	Input Data	Data Analysis
Requirement	Text Mining	Mailing lists	Data Summarization
	Classification	Documentation	Classification of requirements
Design	Clustering Design	document Data gathering	Labelling
Implementation	Clustering	Source code	Software processes
	Classification	SCM	Bug tracking
	Text Mining	Source code	Bug tracking
	Frequent Pattern Mining & Association rule	Program dependence graph	Neglected conditions, Defect correction
Testing	Classification	Program execution	Software behavior classifiers
	Clustering	Execution Profiles	Cluster of Execution Profiles

Table 1: Data Mining Techniques used in different stages of software development [1].

VI. CONCLUSION

In this paper, we have established the need and importance of using data mining techniques to aid software engineering, especially to tackle problems such as the occurrence of bugs, rise in the cost of software maintenance, etc. that can affect software productivity and quality. We have also listed the sources of software engineering data that can be mined, most common stages in the development process as well as the data mining techniques that can be applied in these stages. However, the major contribution of our work lies in the specification of the data mining technique most suited for a particular stage in the development process. We have observed the advantages of using such powerful data mining techniques, especially in terms of time, cost, resources, reliability and maintainability.

VII. REFERENCES

[1] NidhinThomas , Atharva Joshi, Rishikesh Misal and Dr Manjula R,(2016). 'Data Mining Techniques used in Software Engineering: A Survey', International Journal of Computer Sciences and Engineering. Vol 4 No 3, pp.28-34.

[2] Shusi Yu, Shuigeng Zhou, Jihong Guan, "Software Engineering Data Mining: A Survey", Journal of Frontiers of Computer Science and Technology, vol. 6, no. 1, pp. 1-31, June 2012.

[3] Zhang Liangjun, Yang Tan, Xiao Gang, Xu Shengbing, 'Data analysis and mining', Machinery Industry Press, 2016.

[4] NishuSethi,(2016), 'A Comprehensive Study on Data Science and Its Practical Applications', International Journal on Recent and Innovation Trends in Computing and Communication, Vol 5, No 5, pp. 573-575.

[5] Neha Bhateja, NishuSethi (2016), 'Comprehensive Study on Aspects of Software Testing for Quality Assurance', International Journal of Computer Science and Mobile Computing, Vol 5, No 5, pp 703-707.

[6] J. Dong,Y. Zhao andT. Peng, A review of designpattern miningtechniques, InternationalJournalof SoftwareEngineeringand Knowledge Engineering 19(6) (2009), 823–855.

[7] Neha Bhateja,(2017), 'A Business Approach to Data Driven Science - A Study', International Journal of Advanced Research in Computer Science, Volume 8, No. 5, pp. 2483-2485.

[8] Shivangi, (2017). 'Customer Relationship Management and Data Mining' International Journal of Technical Innovation in Modern Engineering & Science. Volume 4 No 5, pp. 1251-1254.

[9] Juhi Singh,(2018), 'A comprehensive study on Open Source Tools and Techniques of Data Mining' , International Journal of Technical Innovation in Modern Engineering & Science. Volume 4 No 5, pp. 1482-1487.