ABSTRACT

Readability metric is considered to be one of the most important factors that may affect games business in terms of evaluating games’ quality in general and usability in particular. As games may go through many evolutions and developed by many developers, code readability can significantly impact the time and resources required to build, update or maintain such games. This paper introduces a new approach to detect readability for games built in Java or C++ for desktop and mobile environments. Based on data mining techniques, an approach for predicting the type of the game is proposed based on readability and some other software metrics or attributes. Another classifier is built to predict software readability in games applications based on several collected features. These classifiers are built using machine learning algorithms (J48 decision tree, support vector machine, SVM and Naïve Bayes, NB) that are available in WEKA data mining tool.

Keywords: Data Mining, Decision Tree, Games Classifiers, Readability Metrics, Vector Machine

DOI: 10.4018/ijeji.2013100104
INTRODUCTION

Software companies are interested in covering the fundamental stages of software development process, by stressing all the practical principles that are applicable during software project development stages.

It’s essential for both developers and customers to get comprehensive knowledge about their project. Many research papers proposed ideas and methods to predict software cost estimation or defect prediction based on collecting software attributes of the software product, process or any other relevant parts.

In addition, the primary concern for software engineers is to identify a variety of factors that may have an impact on software development process. Software readability is considered as a critical factor especially in software maintenance and evolution stages. This is since the more readable code reduces the significant effect on software maintainability. This will eventually reduce overall expenses required for future project extensions for updates, enhancements, support, etc.

Readability of code is considered as one of the essential issues for software developers. Therefore, the level of code readability can clearly affect maintainability and reuse of code. Code readability has a global effect on the code development process. Since, development cost for some projects depends on the level of readability. Many researchers investigated the relation between code readability and the quality of software by proposing readability measures to predict the degree of code readability that may eventually help improve such aspects in future.

In this work, a new approach is proposed to estimate the readability metric using multiple attributes such as: number of statements, number of lines of codes, number of control statements, percentage of comments and other attributes that may affect the process of code readability. The developed readability metric and other metrics are used to predict the type of game source code given a set of games assembled for the evaluation purposes.

Data mining is the process of extracting the information from datasets (Basili & Perricone 1984). Data mining techniques motivate researchers to use their scalable and efficient data analysis methods to analyze software engineering data in diverse areas such as business, entertainment, financial, scientific data...etc. (Zhang & Zhou 2004).

Data mining techniques are used to discover the most valuable information in the organized data by utilizing data analysis methods under adequate computational restrictions in order to find particular patterns over the selected data.

Many research projects were conducted in software engineering filed using data mining methods, where data mining classification techniques such as Support Vector Machines (SVM) (Karatzoglou & Meyer, 2006), (J48) decision tree method, Naïve Bayes (NB) and Artificial Neural Network (ANN), are used in order to come up with effective prediction and identification methods for software modules defects prediction.

This work utilizes data mining techniques to analyze software metrics especially in games domain. We have built a dataset from open source games projects written in C++ and Java, important software features are extracted from them using Source monitor tool (Liu et al. 2008).

SVM is a supervised learning machine algorithm that is used for binary dataset classification (Hastie et al., 2004). Support vector machine aims to build a vector that separates dataset into two categories by applying a simple linear method in high-dimensional space. Moreover, it can show the optimal value for that dataset according to the decision function \( f(x) = \text{sign}(hw \cdot (x) + b) \) (Alsmadi & Magel, 2006; Campwood Software, 2013). A decision or classification tree represents a sequence of policy that can be converted easily into a set of rules. Decision tree helps in recognizing a strategy to achieve a specific target and predict software complexity (Sivaprakasam & Sangeetha, 2012).

In this work, decision tree method (J48) is used as a classification method for the software
Innovative Behaviour of Rural Small Enterprises
www.igi-global.com/article/innovative-behaviour-rural-small-enterprises/58354?camid=4v1a