Chapter 57
R4 Model for Case-Based Reasoning and Its Application for Software Fault Prediction

Ekbal Rashid
Aurora's Technological and Research Institute, India

ABSTRACT

Making R^4 model effective and efficient I have introduced some new features, i.e., renovation of knowledgebase (KBS) and reducing the maintenance cost by removing the duplicate record from the KBS. Renovation of knowledgebase is the process of removing duplicate record stored in knowledgebase and adding world new problems along with world new solutions. This paper explores case-based reasoning and its applications for software quality improvement through early prediction of error patterns. It summarizes a variety of techniques for software quality prediction in the domain of software engineering. The system predicts the error level with respect to LOC and with respect to development time, and both affects the quality level. This paper also reviews four existing models of case-based reasoning (CBR). The paper presents a work in which I have expanded our previous work (Rashid et al., 2012). I have used different similarity measures to find the best method that increases reliability. The present work is also credited through introduction of some new terms like coefficient of efficiency, i.e., developer's ability.

1. INTRODUCTION

Nowadays various machine learning techniques available for software fault prediction. Artificial Intelligence (AI) based techniques like case-based reasoning (CBR), genetic algorithms; neural networks have been used for fault prediction. But, the biggest challenge is how to really apply these techniques in fault prediction, and which technique is more effective in respect to time. Case-based reasoning is a most popular machine learning technique (Rashid et al., 2012). Currently the major application areas of case-based reasoning (CBR) are in the field of health sciences, multiagent systems as well as in Web-based planning. Research on CBR in these areas is growing, but most of the systems are still prototypes

DOI: 10.4018/978-1-7998-2460-2.ch057

and not available in the market as commercial products. However, many of the systems are intended to be commercialized (Begum et al., 2011). The novel idea of CBR is that 'similar problems have similar solutions.' There have been several models for CBR that attempt to offer better understanding of CBR (G.R finnie et al., 2003). The client requires a system that can predict the number of faults encountered during run time on the basis of size metrics i.e., lines of code (LOC). The prediction is based upon machine learning technique like case-based reasoning. Software quality estimation models in software engineering are used to predict important attributes such as software development effort, software reliability, and productivity of programmers (Bhattacherjee et al., 2004). Accurate and timely prediction of the development or maintenance cost of a software system is a critical activity in managing a software project. Software reliability provides measurement of software dependability in which probability of failure is generally time dependent. Software quality prediction is a complex mix of characteristics and varies from application to application and users who requests for it. A software quality prediction model can be used to identify the defective modules. Unlike other software quality estimation models, this model uses Case-Based Reasoning (CBR) technique. CBR utilizes the specific knowledge of previously experienced, concrete problem situation (cases), finding a similar past case, and reusing it in the new problem situation: this is the technique of CBR to solve a new problem. CBR is also an approach to incremental, sustained learning since a new experience is retained each time a problem has been solved, making it available for future problems. The results predicted by this system will be more accurate and can be used further to increase the knowledge base, so that further predictions can be less erroneous. Furthermore, different similarity methods like Euclidean, Canberra, Clark, Exponential and Manhattan are used along with weights to improve the results. Moreover, this model can serve as a gateway to further quality estimation and prediction mechanisms for other systems. Rashid et. al., emphasized on the importance of machine learning and software quality prediction: as an expert system (Rashid et al., 2014). The appendix provides the detail information about hardware and software used in this research paper as well as source code used for building the KBS and removing the duplicate record from KBS.

The rest of the paper is organized as follows: section 2 gives the overview of machine learning, section 3 describes the case-based reasoning approach and program logic, section 4 describes the literature survey. In section 5 I present Methodology Overview and similarity function. Section 6 discussed about the normalizer. Section 7 presents the development of model and results and discussion has been presented in section 8.

2. OVERVIEW OF MACHINE LEARNING

Machine learning deals with the problem of building computer programs that improve their performance at some task through experience (Michalski et al., 1998). Machine learning has been utilized in various problem domains. Some typical applications of machine learning are (Rashid et al., 2012):

- Optical character recognition;
- Face detection
- Spam filtering
- Fraud detection
- Medical diagnosis
- Weather prediction