Advance Software Engineering Literature review

Topic: Data mining/machine learning with software Engineering

Group members
Unish Shah
Andrii Melnyk

CSC 8350: Advanced Software Engineering
Department of Computer Science, Georgia State University
Spring, 2018

Abstract:

The main aim of our literature review will be to show how the data mining and machine learning algorithms help in improving the field of software engineering in different aspects. Machine learning is a field of computer science that gives computers the ability to learn without being explicitly programmed. We will discuss about five issues of software engineering field and show how to solve or improve them using machine learning algorithms.

**Software reuse** is the process of using existing software components to develop new software product. Searching and retrieving reusable components for efficient software reuse are the major challenges faced during reuse. To reuse any software component, there should be an efficient retrieval mechanism to retrieve the reusable component from repository. For the of Software reuse, **genetic algorithm** can be used to resolve it. We will show how this process is done in our literature review.

As the development platforms in software technology grow rapidly; the development of software systems has become more and more complex and versatile. It is quite necessary for the software industries to develop software products with good qualities. As the size of the product and development technologies increasing, there is need for **identification of software defects** at the early stage of development phase. Detection of defects are appropriately and accurately measured in the early stage to increase the quality of the product. **Boosting model** would be used in identifying the defects. We will show how this process is done in our review.

**Design patterns** helpful for software development are the reusable abstract documents which provide acceptable solutions for the recurring design problems. But in the process of reverse engineering, it is often desired to identify as well as recognize design pattern from source code, as it improves maintainability and documentation of the source code. To recognize the design patterns from source codes we can use **Neural networks and Decision tree** algorithms of data mining.

**Requirement prioritization** for software products is one of the most important activities in software development. Prioritization is a critical step towards making great choices with respect to product planning for single and multiple releases. There are numerous requirement prioritization techniques and selecting the most appropriate one is a challenging task. **Fuzzy logic and Decision trees** will show us how to select the best technique for requirement prioritization.
Software development and maintenance tasks often need to change the structure of code without changing the functionality of the code. This kind of code changes are often called **Refactoring**, which has long been recognized as an important way to improve the design of existing code. We will discuss about **GEMS**, an algorithm which recommends a ranked list of code fragments with high accuracy and great speed which will improve refactoring process.

**References**


doi: 10.1109/TENCON.2016.7847994

