Mining Massive Data for a More Responsive Software Engineering Process

Type: Research-oriented
Khushboo Neema & Prakruthi Rao

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

Abstract:
Large amount of data like code, written documents, design documents, run-time documents, version-control system (VCS), etc. is produced in software development. These data are collected and stored by the software organizations to obtain useful information and better understanding of the processes and products. The availability of huge amount of data combined with the difficulty of manually browsing data to retrieve knowledge for multiple stages in software engineering, created a need for data mining techniques. To improve productivity and quality, software engineers are increasing applying data mining algorithms to various software engineer tasks.

This review will focus on the different data mining techniques that can be efficiently applied to the different stages of software engineering data. Data that is mined is generated by: requirements elicitation, development analysis, testing, debugging, and maintenance. Based on this classification, the different mining approaches like classification approach, clustering, Apriori and Frequent Pattern Tree Algorithm and text mining is used and categorized according to the corresponding parts of the development process and the task they assist. These data mining techniques enhance the accuracy and comprehensiveness of software development as well as the credibility of the software.

The main objective is to enable a more responsive, interactive, and adaptive software engineering process by mining huge unstructured data to a structured, actionable knowledge. Thus, this review explains how the software engineering tasks benefits from the applied data mining approaches.

Five Main References: