Sociometrics in Software Engineering: Measuring the Contribution of a Software Engineer in a Software Project through GitHub

Sourojit Das, Master of Science in Computer Science,
University of Dublin, Trinity College, 2019

Supervisor: Prof. Stephen Barrett

Contribution is fundamental to the concept of performance analysis. It is integral to judging the worth of a person in a team, an employee in an organization, and participants in any activity in general. With the industry stressing on rightsizing and optimized workforce management, it is of paramount importance that the contribution of employees is comprehensively monitored and adequately rewarded. However, the subjective and bias-prone nature of existing processes has led to widespread employee dissatisfaction, especially in the software industry.

Traditional LOC-based metrics fail to measure contribution as they do not consider the full range of activities performed by a software engineer in a project. Thus, there arises a need for a comprehensive metric to measure contribution. This research seeks to build a data-model to measure the contribution of a software engineer in a software project by using data mined from GitHub and Gitter.

A high-level model for contribution is constructed and then expanded, by a top-down approach, to create a model that considers both the amount of work done, and its quality. This model represents an expert's view of measuring contribution. Using the data mined from GitHub and Gitter a data- model is constructed, from bottom-up, that seeks to identify useful signals to quantify the concepts expressed by the high-level top-down model. This data-model is conceptualized by applying the principles of Measurement Theory in software engineering and is constantly refined using defeasible reasoning to better quantify the concepts of the high-level model. Finally, an expert system is used to measure the contribution of a software engineer in a project.

The proposed solution is then subjected to a sanity test against the manual evaluation methods proposed by the expert in the high-level model for a particular GitHub repository. The results indicate that the data model is a reasonably accurate representation of the model proposed by the expert and succeeds in providing a rank-ordered list of developers on the basis of their contribution to the project. This listing is consistent with that obtained from a manual evaluation of contribution using the expert's model, though some loss in conceptual fidelity is observed.

Keywords: Contribution, Expert System, GitHub, Data Model, Defeasible Reasoning.