Discovering Determinants of High Volatility Software

Carolyn Seaman, A. Güneş Koru, and Sreedevi Sampath
University of Maryland Baltimore County
{cseaman, gkoru, sampath @ umbc.edu}

This topic paper presents a line of research that we are proposing that incorporates, in a very explicit and intentional way, human and organizational aspects in the prediction of troublesome (defect-prone or change-prone or volatile, depending on the environment) software modules. Much previous research in this area tries to identify these modules by looking at their *structural characteristics*, so that increased effort can be concentrated on those modules in order to reduce future maintenance costs.

The work proposed here extends this idea. Our approach is to take one step back up the causal chain that leads to software volatility (i.e. unnecessarily high rates of change) by discovering the *characteristics of changes* that result in software that exhibits structural characteristics that in turn indicate high volatility. We are considering a wide variety of change characteristics, including human, organizational, and process factors. In particular, we are concerned with identifying those characteristics that can be eliminated or mitigated by changing programming and management practices during development and maintenance, so that volatility can then be reduced.

We plan to build our empirical models based on data from both open source and closed source systems. The qualitative and quantitative data we will examine on software changes include:

- The names of the changed modules
- The type and date of the change
- The textual changes that were made to each module
- The person or persons who made the change and the organizational relationships between them
- Structural measures, taken both before and after the change, including size, coupling, cohesion, inheritance, and complexity
- How well the change was tested, e.g. through code coverage
- The motivation for the change
- The conditions under which the change was made, e.g., schedule pressure, diffusion of the task, etc.
- The difficulty of the change

It is expected that some of these data will be qualitative in nature. In order to determine which qualitative characteristics are relevant, developers will be interviewed about what characteristics of a change they think will have an effect on structural characteristics. This process will be iterative, and the set of predictor variables will be refined over the course of this research effort.

Some of these factors will be quantifiable, some on an ordinal scale and others categorical. Appropriate analysis techniques will be chosen accordingly. However, we may also have some factors that we cannot code into quantitative variables of any kind. In order to take such factors into consideration in this analysis, we will employ a standard qualitative analysis technique, such as the constant comparison method, to determine the degree to which the factor is relevant to the increases or decreases in the structural measures.

We expect to find a collection of relationships in this analysis, not a single or simple explanation. To explore issues of causality, we will conduct a classroom-based empirical study in which two groups make software changes that vary in the change characteristics from our preliminary models. The values of the structural characteristics of the modified software will then be compared to determine if in fact the change properties determine variations in structural properties.

The outcome of this work will be a set of models that describe the relationships between software change characteristics (in particular human, organizational, and process characteristics), changes in the structural properties of software modules (e.g. complexity), and the future volatility of those modules. The impact of such models is two-fold. First, maintainers will have an improved technique for estimation of maintenance effort based on recent change characteristics. Also, the models will help identify management practices that are associated with high future volatility.

978-1-4244-3712-2/09/$25.00