What is Wrong with our Estimates?

Using Interactive Views to Uncover Patterns in Problem Report Estimation.
Patrick Knab, Martin Pinzger, Harald Gall

Issue tracking repositories contain a wealth of information for reasoning about various aspects of software development processes. Our approach uses a combination of graphical views to investigate details of individual problem reports while maintaining the context provided by the surrounding data population, which enables the detection and detailed analysis of hidden patterns.

Multiple Views to Discover Important Connections

For the visualization of effort measures, we use Polymetric Views, the width of the boxes is determined by the value of the estimatedEffort and the height by the value of the actualEffort. The effort measure is the sum of all efforts that were exerted to resolve the issue described by the problem report. The color of the box is mapped to the identity of the analyzer (other mappings are also possible).

To visualize the duration of process steps we use a pie chart visualization. The size (i.e., the area) of the pie is mapped to the total time from the creation of the PR until it was closed. Finally with a PR state transition view we provide a new angle and starting point for an analysis. In this view configuration, all occurring state transitions are displayed in an aggregated form. The width of the arrows between the states is mapped to the number of problem reports that exhibit the corresponding transition.

Distinct Patterns For Different Analyzers

This view groups and colorizes PRs according to the analyzer that did the estimates. The width of rectangles represents the estimated effort, the height the actual effort. Looking for patterns we can see, that there is a mix of estimation errors as well as some fairly well estimated PRs. We can also see, that there are about six analyzers that do the most analyses (more than 10). There are some differences in the performance of the analyzers. Considering the selections for analyzer A and analyzer B, we can see that analyzer A mainly underestimates the effort, whereas analyzer B mainly overestimates. Looking at the differences in the width of the boxes of analyzer A we get the impression that he uses mostly one standard estimate for all his PRs where analyzer B has more variability in his estimates.

Skipping the Analysis Phase: Bad or Good?

To answer this question we start with our state transition visualization were the transition submitted to in_resolution is selected. Only a small amount of all PRs take this shortcut while most of them go first into the in_analysis phase.

To assess the effects on the resolution process, we use our pie chart visualization were we can explore additional properties of the problem reports that skipped the initial analysis. The view shows that there are a lot of almost totally green pies, that there are small as well as bigger circles indicating shorter and longer processing time, and that there are PRs that although they skipped the initial analysis phase have nonetheless spent time in analysis. The pies labeled a), b), and c) have a blue wedge in their pie chart. Selecting the PR labeled b) and displaying it in a state transition view, we can examine its life-cycle in more detail.