A Standard Measuring Procedure for Software Engineering

L. Benedicenti§, W. Curry¥, M. Mintchev¥, M. Smith¥, G. Succi¥

§Dipartimento di Informatica, Sistemistica e Telematica - Università di Genova, Italy
¥Department of Electrical and Computer Engineering - University of Calgary, Canada

Extended Abstract

This paper describes a standard measuring procedure for software engineering. It is an account of our experience in trying to obtain software engineering experiments that are repeatable and comparable.

Measures in software engineering often lack a solid measurement theory basis. Even more, there is the need for a basic standard procedure to extract meanings from these measures.

Real-life projects often encounter difficulties either in obtaining the required measures (metrics) or in organizing the data collected in order to assess the result of the project. This makes it impossible to compare different projects.

Our approach is based on the following:

- The representational theory of measurement, as proposed by Fenton (1994 and 1997) to identify and measure attributes of the project
- A statistical approach to the measures obtained in order to organize them, present them, and extract a behavior from them, based on the work of Berenson et al (1996), and Winer (1971)
- The use of neural networks to identify relations that defy statistical analysis, based on the works by Kadirkamanathan and Niranjan (1993), and Haykin (1994)

In this paper we detail how to develop and classify a set of measures and how to analyze the data collected.

Then, we present a sample application of our approach on a set of industrial data to determine the correlation between internal properties of code, such as number of lines of code, cyclomatic complexity, amount of reuse, with external properties of the production process, such as effort spent, number of defects found. The data were collected in an Italian software company.

Our approach is innovative for software engineering because it involves the standardization of measuring procedures and of the statistical tools to draw conclusions on the data, so it follows the entire life cycle of the experiment. Moreover, our approach maximizes the experience gained from each project, thus enabling the collection of an experience base.

The usage of Neural Networks generalizes non-linear relations from a set of data. This is useful for model estimation where the modeling function is unknown. Therefore, we employ Neural Networks only when no simpler relationships can be found in the data.

§Address for correspondence: Luigi Benedicenti, DIST, Via Opera Pia 13, 16145 Genova, Italy
The advantage of this approach is that it is repeatable, clear, easily transferable among firms, and can be tailored to suit academic or corporate requirements. It is also a general approach that can cope with several types of measures, representation models, and experimental goals.
Future extensions of this procedure include its automation in process enactment tools: automatic data collection is possible since metrics are univocally defined and automatic model construction is possible by means of statistics tools.