Empirical Assessment of Software Engineering Research: Pitfalls and Solutions

Massimiliano Di Penta

Department of Engineering,
University of Sannio, Italy

The availability of a wide variety of software repositories, ranging from Questions and Answer forums to mailing lists, forges and issue trackers opens the road for building recommender systems aimed at supporting developers in their activities. Upon evaluating such recommenders, in most cases researchers focus on the underlying approach capability of providing accurate and complete results. This is insufficient because a tool could achieve an almost perfect precision and recall, however, its provided features do not really help software developers in their everyday's tasks.

This keynote will report my personal experience (and those of my collaborators) in evaluating recommenders, showing that an offline evaluation of the approach precision and recall is only a very preliminary starting point. Importantly, different kinds of (qualitative and quantitative) evaluations, having different size and level of control, and above all involving humans, are required to achieve results able to convince practitioners of the actual usefulness and applicability of a tool. Moreover, I will discuss and emphasize the important role played by contextual information in the evaluation of recommender systems.