Scaffolded Training with Machines: An Experiment on Participant Learning in an On-line Production Community

Abstract

Training users in online communities is important for making high performing contributors. However, several conundrums exist in choosing the most effective approaches to training users. For example, if it takes time to learn to do the task correctly, then the initial contributions may not be of high enough quality to be useful.

We conducted an online field experiment where we recruited users (N = 386) in a web-based citizen-science project to evaluate the two training approaches. The project uses machine learning to teach users how to identify anomalies in a classification task. In one training approach, users received one-time training and were asked to apply their knowledge to the classification task. In the other approach, users were gradually exposed to classes of data that were selected by trained machine learning algorithms as being members of particular classes.

The results of our analysis revealed that the gradual training produced high performing contributors. In our comparison of the treatment and control groups we found users who experienced gradual training performed significantly better on the task (an average accuracy of 90% vs. 54%), contributed more work (an average of 228 vs. 121 classifications), and were retained in the project for a longer period of time (an average of 2.5 vs. 2 sessions). The results suggest online production communities seeking to train newcomers would benefit from training regimes that gradually introduce them to the work of the project using real tasks.