

MISQ Archivist

Crowd-Squared: Amplifying the Predictive Power of Search Trend Data

Erik Brynjolfsson, Tomer Geva, and Shachar Reichman

Abstract

Big data generated by crowds provides a myriad of opportunities for monitoring and modeling people's intentions, preferences, and opinions. A crucial step in analyzing such big data is selecting the relevant part of the data that should be provided as input to the modeling process. In this paper, we offer a novel, structured, crowd-based method to address the data selection problem in a widely used and challenging context: selecting search trend data. We label the method "crowd-squared," as it leverages crowds to identify the most relevant terms in search volume data that were generated by a larger crowd. We empirically test this method in two domains and find that our method yields predictions that are equivalent or superior to those obtained in previous studies (using alternative data selection methods) and to predictions obtained using various benchmark data selection methods. These results emphasize the importance of a structured data selection method in the prediction process, and demonstrate the utility of the crowd-squared approach for addressing this problem in the context of prediction using search trend data.

Keywords: Prediction, big data, search trends, data selection, crowdsourcing