Data Foraging: Spatiotemporal Data Collection Decisions in Disciplinary Field Science

Cristina Wilson
Temple University, Philadelphia, Pennsylvania, United States

Feifei Qian
University of Pennsylvania, Philadelphia, Pennsylvania, United States

Doug Jerolmack
University of Pennsylvania, Philadelphia, Pennsylvania, United States

Thomas Shipley
Temple University, Philadelphia, Pennsylvania, United States

Sonia Roberts
University of Pennsylvania, Philadelphia, Pennsylvania, United States

Jonathan Ham
Temple University, Philadelphia, Pennsylvania, United States

Daniel Koditschek
University of Pennsylvania, Philadelphia, Pennsylvania, United States

Abstract

Field scientists collect data in a noisy heterogeneous environment, where the value of additional data for characterizing the natural system is weighed against the time and money involved in data collection. This is analogous to foraging for food data is the resource and its collection can be optimized based on energy costs. Here we conduct a novel simulated data foraging study to elucidate how spatiotemporal data collection decisions are made in field sciences, and how search is adapted in response to in-situ data. Expert geoscientists were asked to evaluate a hypothesis by collecting environmental data using a mobile robot. At any point, participants were able to stop the robot and change their search strategy or make a conclusion about the hypothesis. We identified previously unrecognized spatiotemporal reasoning heuristics, to which scientists strongly anchored, displaying limited adaptation in response to new data. We analyzed two key decision factors: variable-space coverage, and fitting error to a given hypothesis. We found that, despite varied search strategies, the majority of scientists made a conclusion as the fitting error converged. Scientists who made premature conclusions, either due to insufficient variable-space coverage or before the fitting error stabilized, were more prone to incorrect conclusions. We believe the findings from this study could be used to improve field science training in data foraging, and aid in the development of technologies to support data collection decisions.