



Improved Vote Aggregation Techniques for the Geo-Wiki Cropland Capture Crowdsourcing Game

Artem Baklanov (1,2), Steffen Fritz (1), Michael Khachay (2), Oleg Nurmukhametov (2), Carl Salk (1), Linda See (1), and Dmitry Shchepashchenko (1)

(1) International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria, (2) Krasovsky Institute of Mathematics and Mechanics, Ekaterinburg, Russian Federation (This research was supported by Russian Science Foundation, grant no. 14-11-00109)

Crowdsourcing is a new approach for solving data processing problems for which conventional methods appear to be inaccurate, expensive, or time-consuming. Nowadays, the development of new crowdsourcing techniques is mostly motivated by so called Big Data problems, including problems of assessment and clustering for large datasets obtained in aerospace imaging, remote sensing, and even in social network analysis. By involving volunteers from all over the world, the Geo-Wiki project tackles problems of environmental monitoring with applications to flood resilience, biomass data analysis and classification of land cover. For example, the Cropland Capture Game, which is a gamified version of Geo-Wiki, was developed to aid in the mapping of cultivated land, and was used to gather 4.5 million image classifications from the Earth's surface. More recently, the Picture Pile game, which is a more generalized version of Cropland Capture, aims to identify tree loss over time from pairs of very high resolution satellite images. Despite recent progress in image analysis, the solution to these problems is hard to automate since human experts still outperform the majority of machine learning algorithms and artificial systems in this field on certain image recognition tasks. The replacement of rare and expensive experts by a team of distributed volunteers seems to be promising, but this approach leads to challenging questions such as: how can individual opinions be aggregated optimally, how can confidence bounds be obtained, and how can the unreliability of volunteers be dealt with? In this paper, on the basis of several known machine learning techniques, we propose a technical approach to improve the overall performance of the majority voting decision rule used in the Cropland Capture Game. The proposed approach increases the estimated consistency with expert opinion from 77% to 86%.