Anomaly Detection in Crowdsourcing: Why Midpoints in Interval-Valued Approach

Alejandra De La Peña, Damian L. Gallegos Espinoza, and Vladik Kreinovich
Department of Computer Science, University of Texas at El Paso
El Paso, Texas 79968, USA
adelapena5@miners.utep.edu, dlgallegose@miners.utep.edu
vladik@utep.edu

What is crowdsourcing: a brief reminder. In many practical situations, we need to perform a large number of reasonably simple tasks, tasks that do not require high qualifications. For example, deep learning requires that a large number of labeled examples by available. In many cases, we do not have that many labeled examples. So we need someone to label a large number of photos, or a large number of speech recordings. One way to perform these tasks is *crowdsourcing*, when people all over the world are paid to solve the corresponding tasks. Example: label pictures for training a machine learning algorithm.

Need to detect anomalies. Most crowd-workers work conscientiously; however, since the payment is proportional to the number of answers, there are also many cases when crowd-workers do a sloppy job, not spending enough time on analyzing the corresponding picture and therefore producing answers that are often wrong. Such wrong answers prevent machine learning algorithms from getting high quality results. It is therefore important to be able to detect such anomalous crowd-workers and dismiss their answers.

A natural way to do it is to include examples with known labels into the list of tasks. Then, we can gauge the quality of a crowd-worker by the number of wrong answers that he/she has on these examples. If this number is unusually high, then all the answers provided by this crowd-worker should be dismissed.

Need to take into account degrees of confidence. Crowd-workers are often not 100% confident in their answers. To help machine learning, it is therefore desirable to collect not only the answers, but also the degrees indicating how confident is the crowd-worker in each answer. This way, the neural network will be able to weigh these answers with different weights: if its answer is different from the confident answer of a crowd-worker, then the algorithm should continue training, but if the difference is only with not very confident crowd-workers, then maybe there is no need to adjust.

These degrees of confidence are used to detect anomalies. If the answer is wrong but the crowd-worker is not very confident about it, this may be an honest mistake. However, if there are many wrong answers with high degrees of confidence, this indicates an anomaly. Sometimes, these degrees also affect the amount of payment: the higher degree of confidence, the higher the pay – since one way to gain more confidence is to spend more time analyzing the corresponding picture or recording.

Interval-valued degrees of confidence. Crowd-workers are usually unable to describe their degree of confidence by a single number. So, it makes sense to allow the crowd-workers to mark their confidence by selecting an interval $[\underline{x}, \overline{x}]$ of possible degrees. For example, an interval $[0.7, 0.8]$.

How to detect anomalies based on interval-valued degrees: formulation of the problem. A natural idea is to utilize formulas that have been successful in detecting anomalies based on numerical degrees. To apply these formulas, we need to select a single value x from the corresponding interval $[\underline{x}, \overline{x}]$. In other words, we need an algorithm $x = f(\underline{x}, \overline{x})$ that generates a number based on the bounds of the worker-generated interval. Which algorithm $f(\underline{x}, \overline{x})$ should we select? An empirical analysis has shown that the more accurate anomaly detection happens when we use arithmetic average $\dfrac{\underline{x} + \overline{x}}{2}$. In this talk, we use natural invariances to explain this empirical result.