Cornell University Library

We gratefully acknowledge support from the Simons Foundation and member institutions

arXiv.org > stat > arXiv:1305.0015

Search or Article-id

(Help | Advanced search)

All papers | Go!

Statistics > Machine Learning

Inferring ground truth from multi-annotator ordinal data: a probabilistic approach

Balaji Lakshminarayanan, Yee Whye Teh

(Submitted on 30 Apr 2013)

A popular approach for large scale data annotation tasks is crowdsourcing, wherein each data point is labeled by multiple noisy annotators. We consider the problem of inferring ground truth from noisy ordinal labels obtained from multiple annotators of varying and unknown expertise levels. Annotation models for ordinal data have been proposed mostly as extensions of their binary/categorical counterparts and have received little attention in the crowdsourcing literature. We propose a new model for crowdsourced ordinal data that accounts for instance difficulty as well as annotator expertise, and derive a variational Bayesian inference algorithm for parameter estimation. We analyze the ordinal extensions of several state-of-the-art annotator models for binary/categorical labels and evaluate the performance of all the models on two real world datasets containing ordinal query-URL relevance scores, collected through Amazon's Mechanical Turk. Our results indicate that the proposed model performs better or as well as existing state-of-the-art methods and is more resistant to `spammy' annotators (i.e., annotators who assign labels randomly without actually looking at the instance) than popular baselines such as mean, median, and majority vote which do not account for annotator expertise.

Submission history

Which authors of this paper are endorsers?

Link back to: arXiv, form interface, contact.