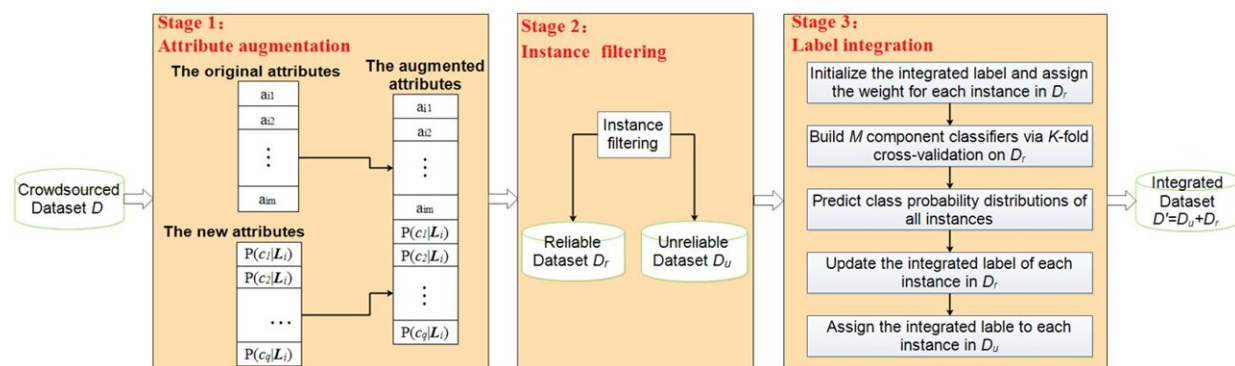


Attribute augmentation-based label integration for crowdsourcing

October 30 2023



Credit: *Frontiers of Computer Science* (2022). DOI: 10.1007/s11704-022-2225-z

Crowdsourcing provides an effective and low-cost way to collect labels from crowd workers. Due to the lack of professional knowledge, the quality of crowdsourced labels is relatively low. A common approach to addressing this issue is to collect multiple labels for each instance from different crowd workers and then a label integration method is used to infer its true label. However, almost all existing label integration methods merely make use of the original attribute information and do not pay attention to the quality of the multiple noisy label set of each instance.

To solve these issues, a research team led by Liangxiao JIANG published their new research in [Frontiers of Computer Science](https://doi.org/10.1007/s11704-022-2225-z).

The team proposed a novel three-stage label [integration](#) method called attribute augmentation-based label integration (AALI). AALI enhances the performance of label integration by improving the discriminative ability of the original attribute space and identifying the quality of each instance's multiple noisy label set. Experimental results on simulated and real-world crowdsourced datasets demonstrate that AALI outperforms all the other state-of-the-art competitors in terms of label quality and model quality.

In the research, they design an attribute augmentation method to enrich the attribute space, and then develop a filter is to single out reliable instances with high-quality multiple noisy label sets from a crowdsourced dataset. Finally, they use the cross-validation to build multiple component classifiers on reliable instances to predict all instances.

In the [first stage](#), AALI defines class membership probabilities generated from a multiple noisy label set as new attributes and constructs the augmented attributes by concatenating the original attributes with the new attributes. In the second stage, AALI develops a filter to single out reliable instances with high-quality multiple noisy label sets. As a result, the original dataset is divided into a reliable dataset and an unreliable dataset. In the third stage, AALI uses majority voting to initialize integrated [labels](#) of all instances in reliable dataset while estimating the certainty of each integrated label and assigning it to the weight of each instance.

Next, AALI uses K-fold cross-validation to build M component classifiers on reliable dataset to predict class probability distributions of all instances. At last, AALI updates the integrated label of each instance in reliable dataset and assigns the integrated label to each instance in unreliable dataset. The extensive experimental results on both simulated and real-world crowdsourced [datasets](#) validate the superiority of AALI.

Future work can focus on finding the optimal value of the developed filter's threshold using an optimization method.

More information: Yao Zhang et al, Attribute augmentation-based label integration for crowdsourcing, *Frontiers of Computer Science* (2022). [DOI: 10.1007/s11704-022-2225-z](https://doi.org/10.1007/s11704-022-2225-z)

Provided by Frontiers Journals

Citation: Attribute augmentation-based label integration for crowdsourcing (2023, October 30) retrieved 9 May 2024 from <https://techxplore.com/news/2023-10-attribute-augmentation-based-crowdsourcing.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--