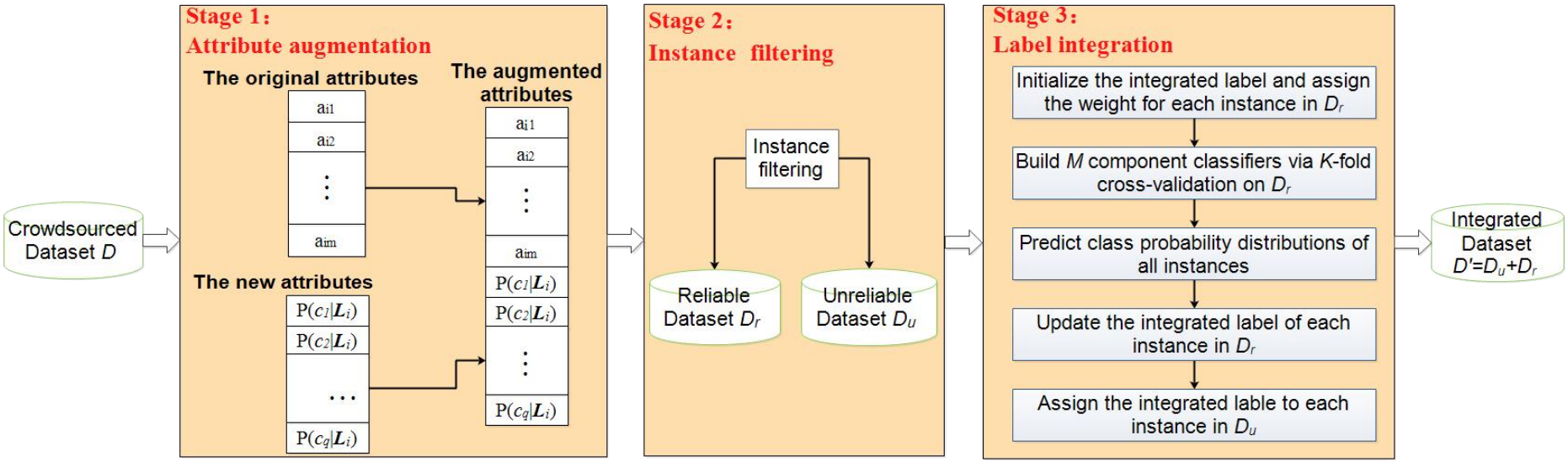
Augmentation-based Label Integration for Crowdsourcing

Yao ZHANG, Liangxiao JIANG, Chaoqun LI

Frontiers of Computer Science, DOI: 10.1007/s11704-022-2225-z

Problems & Ideas

- Problems of label integration for crowdsourcing:
  - Almost all existing label integration methods merely make use of the original attribute information.
  - Almost all existing label integration methods do not pay attention to the quality of the multiple noisy label set of each instance.
- Ideas: a novel three-stage label integration method is proposed which is called attribute augmentation-based label integration (AALI).



Overview framework of AALI.

Main Contributions

- Contributions:
  - An attribute augmentation method is designed to enrich the attribute space;
  - A filter is developed to single out reliable instances with high-quality multiple noisy label sets from a crowdsourced dataset;
  - The cross-validation is used to build multiple component classifiers on reliable instances to predict all instances.

Dataset	DS	ZC	KOS	GTIC	IWMV	AALI
Income	72.67	72.03	71.33	71.83	71.77	**73.4**
Leaves	63.8	64.61	63.54	62.24	64.35	**66.69**
LabelMe	74.7	77.2	76.5	76.7	77.2	**78.7**
NER	**89.64**	88.64	86.83	83.11	87.95	84.11

Dataset	DS	ZC	KOS	GTIC	IWMV	AALI
Income	69.83	69	71	70.33	69.17	**76.67**
Leaves	54.43	52.34	53.39	50.52	52.6	**78.39**
LabelMe	48.8	49	47.5	47.2	47	**52.8**
NER	85.03	84.86	84.06	82.09	84.48	**87.18**

The label quality and model quality comparison results on four real-world datasets. Up: the label quality (%) comparisons of six methods; Down: the model quality (%) comparisons of six methods.