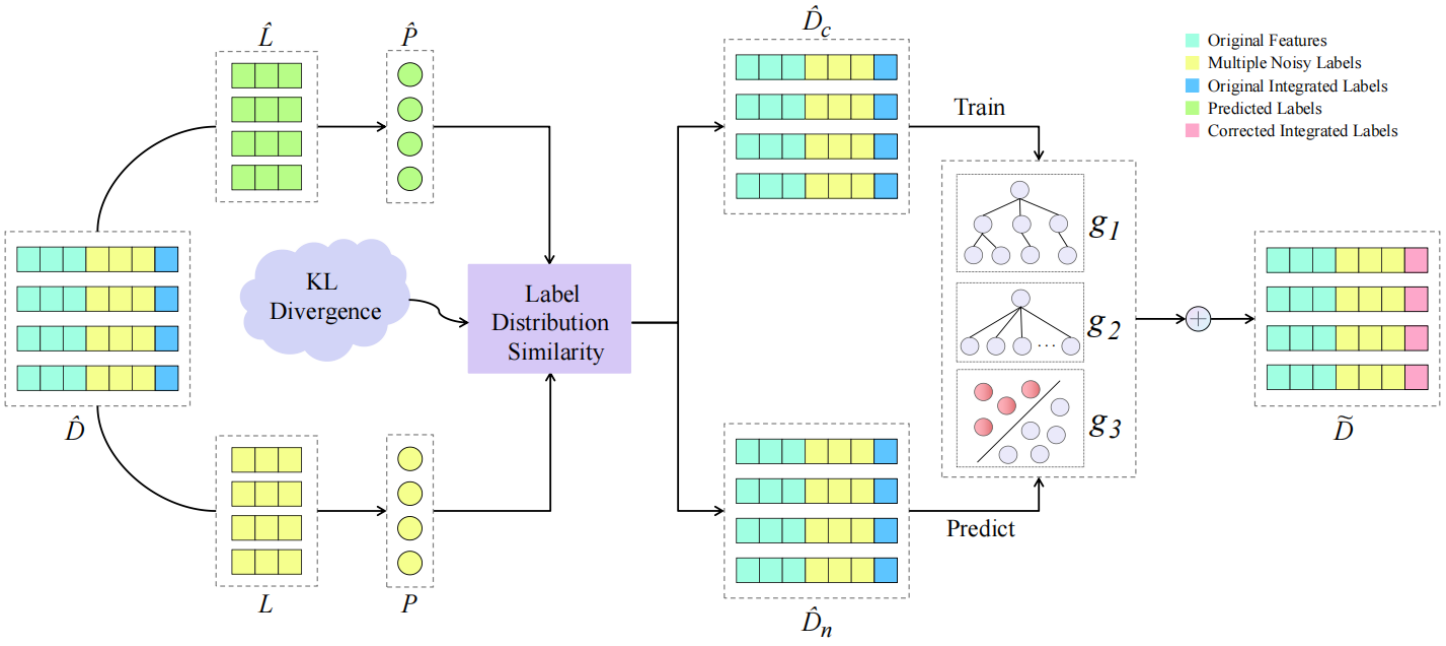
Label Distribution Similarity-based Noise Correction for Crowdsourcing

Lijuan REN, Liangxiao JIANG, Wenjun ZHANG, Chaoqun LI

Frontiers of Computer Science, DOI: 10.1007/s11704-023-2751-3

Problems & Ideas

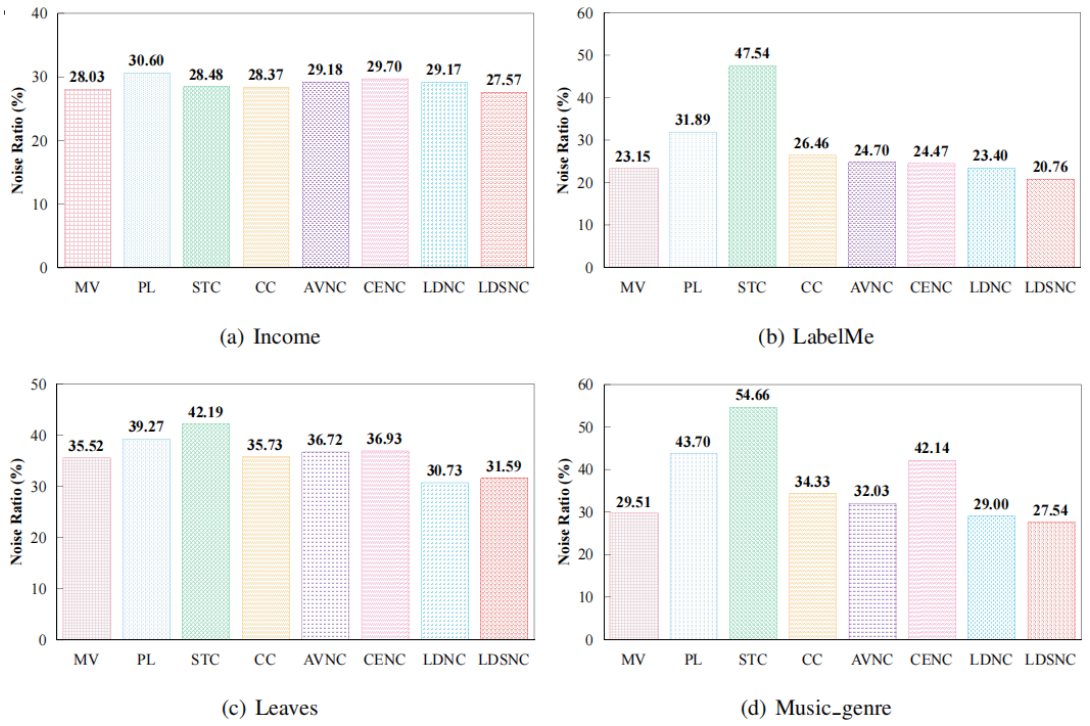
- Problems of noise correction for crowdsourcing:
  - Existing methods rarely consider an instance's information from both its features and multiple noisy labels simultaneously when identifying a noise instance.
- Ideas: A novel noise correction method is proposed which is called as label distribution similarity-based noise correction (LDSNC).



Overall framework of LDSNC.

Main Contributions

- Contributions:
  - A novel noise correction method is designed to reduce the impact of noise in integrated labels by considering each instance's information from both its features and multiple noisy labels simultaneously;
  - The KL divergence is used to measure the similarity between the predicted label distribution and multiple noisy label distribution of an instance.



Noise ratio (%) comparisons for LDSNC versus MV, PL, STC, CC, AVNC, CENC and LDNC on four real-world datasets.