Crowd-sourcing labeling

Pros: cheap and fast to obtain large quantity of label data.
Cons: the obtained labels can be very noisy.

Previous work

- Majority voting based confidence. [Donez et al. 2009-2010]
- Incremental relabeling mechanism. [Zhao et al. 2011]

Disadvantage

- Cannot handle label noise during the labeling process.
- The label quality will be heavily affected if the malicious labelers occur at the early stage.
- Only investigate the case where a single copy of labels is engaged.

Motivation

- We introduce the active learning strategy into the framework.
- We want to enable the collaborative work among the multiple labelers.
- We want to handle the label noise during the labeling process.
- We want to detect and even kick out the irresponsible labelers at the early stage.
- We also want to make full use of multiple copies of labels.

Datasets

- ImageNet dataset (10 categories, LLC features)
- Gender face dataset (9441 face images)

Comparisons

- CAL: collaborative active learning (ours).
- CRL: collaborative random learning (ours).
- MIAL: multiple independent active learning (remove cross term from CAL).
- MIRL: multiple independent random learning (remove cross term from CAL).
- SVM-MIAL: multiple independent active learning SVM.
- SVM-MIRL: multiple independent random learning SVM.
- MVAL: single classifier with majority voted labels using logistic loss.
- SVM-MVAL: single classifier with majority voted labels using hinge loss.
- ML-Bernoulli: active learning with multiple labelers (Bernoulli version) proposed by Yan Yan et al. [ICML 2011]
- ML-Gaussian: active learning with multiple labelers (Gaussian version) proposed by Yan Yan et al. [ICML 2011]