Anomaly detection in crowdsourced work with interval-valued labels

Background

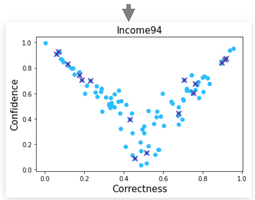
Our published work at SSCI 2021 proved that **worker quality can't be assured** in crowdsourcing, so efficiently detecting anomalies is crucial to getting inferences that are both correct and trustworthy!

Method

N=100
Generate workers with random reliability indexes

Select Attackers
Select random workers to give bad IVLs

Attacker Detection
Gold and regular IVLs are checked for the same means and variances statistically

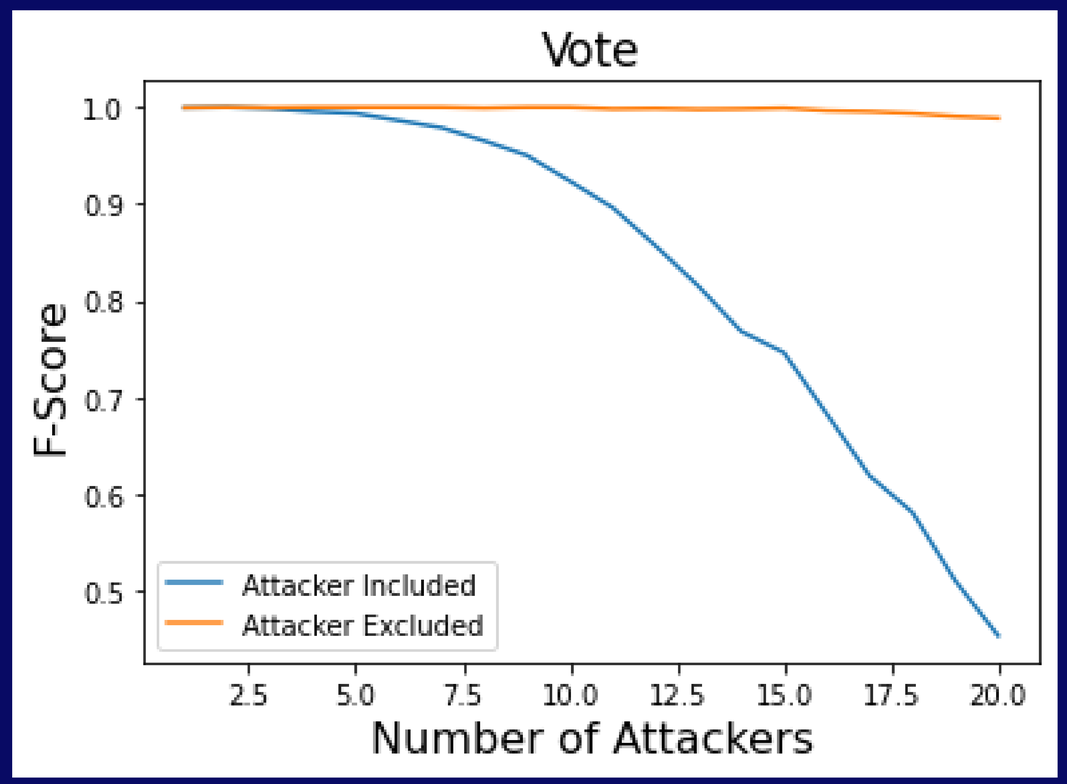


Inference Making
Attacker-excluded and included tests are run using Weighted Interval Majority Voting (WIMV)

Confusion Matrices
Evaluate performance of including and excluding attackers through F_1-score

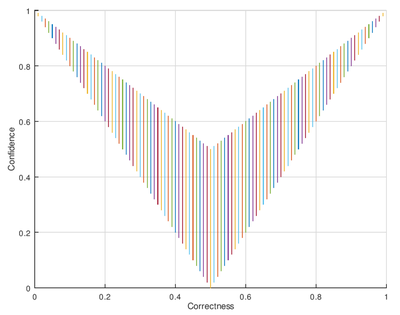
This material is based upon work supported by the National Science Foundation under Award No. OIA-1946391.

Excluding attackers gets reliable results.

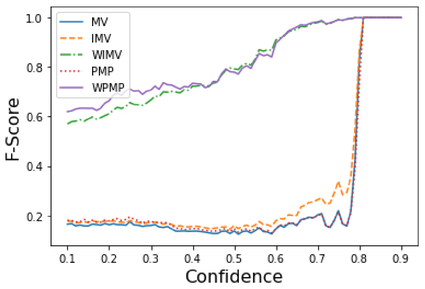


Without excluding, resulting quality falls drastically.
With excluding, quality is almost perfect.

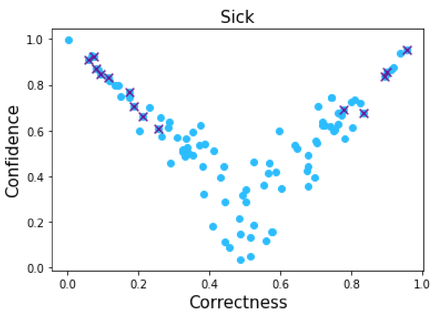
Paper to be published by Springer in the Communications in Computer and Information Science (CCIS) series in the IPMU 2022 Conference in Milan, Italy.



Correctness-Confidence Graph



F_1-score value vs. confidence threshold



Statistically inconsistent workers in Sick dataset

$$t = \frac{dist(\mu(\mathbf{L}_G^j), \mu(\mathbf{L}_X^j))}{\sqrt{\frac{\sigma^2(L_G^j)}{|L_G^j|} + \frac{\sigma^2(L_X^j)}{|L_X^j|}}}$$

$$F = \frac{Var(\mathbf{L}_X^j)}{Var(\mathbf{L}_G^j)}$$

Makenzie Spurling, Chenyi Hu, Huixin Zhan, and Victor S. Sheng