Reliable Aggregation Method for Vector Regression Tasks in Crowdsourcing

Joonyoung Kim, Donghyeon Lee, Kyomin Jung

Machine Intelligence Laboratory

13 May 2020
What is Crowdsourcing?

- Crowdsourcing
  - Process for structuring unstructured data using human resources
What is Crowdsourcing?

- **Crowdsourcing**
  - Process for structuring unstructured data using human resources

- **Three Elements**
  - Requester
  - Workers
  - Tasks
Challenges in Crowdsourcing

- Noisy responses
  - Lack of expertise, low payment
  - Free money collector
  - Spammer, adversarial workers
Challenges in Crowdsourcing

- **Noisy responses**
  - Lack of expertise, low payment
  - Free money collector
  - **Spammer**, adversarial workers

- **Redundancy on querying tasks**
  - Helpful but need more cost
Previous works on Crowdsourcing

- **Goal**: To infer the ground truth from responses with minimum cost
  - Existing methods focus on only binary or multiclass classification tasks (discrete choices)
Previous works on Crowdsourcing

- **Goal**: To infer the ground truth from responses with minimum cost
  - Existing methods focus on only binary or multiclass classification tasks (discrete choices)

- Our work targets **real-valued vector regression**. This type of task is more difficult to find true answer from noisy responses.
Vector Regression Task

(a) Movie rating 
(b) Image Object 
(c) Pose estimation

- **Goal**: Find the exact size and direction of a vector!
  - More **difficult** to find tight float values than classification
Image Object Localization

- **Find the object’s coordinates in the image**
  - Ground Truth (rectangular shape)
    > 4 coordinates \((x_{tl}, y_{tl}), (x_{br}, y_{br})\)
  - For simplicity, focus on x-axis only
  - Worker Response: **Tight** bounding box
    > Normalized with its width \((x_{max})\)
      \[
      A_x = (x_{tl}, x_{br} - x_{tl}, x_{max} - x_{br}) / x_{max} \tag{1}
      \]
    > Mapping a single response to a point in 2D-simplex
Task Assignment

- **Task-Worker Mapping**
  - Bipartite Graph $G = \{[m], [n], E\}$.
  - Each task $i \in [m]$ is assigned to $l_i$ workers
  - Each worker $j \in [n]$ solves $r_i$ tasks
  - For $(i, j)$ pair, there is worker $j$’s response $A_{ij} \in E$. 

![Diagram showing task-worker mapping](image-url)
Naive Aggregation Method

\begin{itemize}
  \item \textbf{Majority Voting (MV)}
  \[ \hat{t}_i^{(MV)} = \sum_{j \in \delta_i} \frac{1}{l_i} A_{ij}. \]
\end{itemize}
Naive Aggregation Method

- **Majority Voting (MV)**
  \[ \hat{t}_i^{(MV)} = \sum_{j \in \delta_i} \frac{1}{l_i} A_{ij}. \]

- **Outlier Rejection (Top-k, \(|\Delta_i| = k < l_i|)**
  \[ \hat{t}_i^{(Top-k)} = \sum_{j \in \Delta_i} \frac{1}{k} A_{ij}. \]
Proposed Method - 1

- **Iterative Update**

\[
A_{ij} \left\| A_{ij} - x_{i\rightarrow j}^{(k)} \right\|_2
\]

- **Task Message (x-message)**

\[
x_{i\rightarrow j} = \sum_{j' \in \delta_i \setminus j} \left( \frac{y_{j'\rightarrow i}}{y_{\delta_i \setminus j}} \right) A_{ij'}
\]  

(2)

- **Worker Message (y-message)**

\[
y_{j\rightarrow i} = \left( \frac{1}{\hat{r}_j} \sum_{i' \in \delta_j \setminus i} \left( \left\| A_{i'j} - x_{i'\rightarrow j} \right\|_2 \right) \right)^{-1}
\]

(3)
Proposed Method - 2

\[
x_{i \rightarrow j} = \sum_{j' \in \delta_i \setminus j} \left( \frac{y_{j' \rightarrow i}}{y_{\delta_i \setminus j}} \right) A_{i,j'}
\]
Proposed Method - 2

- Task Message Update

\[
x_{i \rightarrow j} = \sum_{j' \in \delta_i \setminus j} \left( \frac{y_{j' \rightarrow i}}{y_{\delta_i \setminus j}} \right) A_{ij'}
\]  

- Relative reliability of the worker \( j' \in \delta_i \setminus j \)
Proposed Method - 3

Other Tasks \quad Worker j \quad Task i

\begin{equation}
\delta_j \setminus \{ j \} \quad x_{i' \rightarrow j} \quad \hat{r}_i \quad y_{j \rightarrow i} \quad x_{i' \rightarrow j} \end{equation}

- **Worker Message Update**

\[ y_{j \rightarrow i} = \left( \frac{1}{\hat{r}_j} \sum_{i' \in \delta_j \setminus i} ( || A_{i'j} - x_{i' \rightarrow j} ||_2 ) \right)^{-1} \] (3)
Proposed Method - 3

**Worker Message Update**

\[
y_{j \rightarrow i} = \left( \frac{1}{\hat{r}_j} \sum_{i' \in \delta_j \setminus i} (\|A_{i'j} - x_{i' \rightarrow j}\|_2) \right)^{-1}
\]  \hspace{1cm} (3)

- **Distance** between worker \( j \)'s response and the average response of workers who solves task \( i' \)
Dirichlet Crowd Model

- Crowd’s response $\sim$ **Dirichlet distribution** $f(x; \alpha)$

$$f(x; \alpha) = \frac{1}{B(\alpha)} \prod_{d=1}^{D+1} x_d^{(\alpha_d - 1)}$$  \hspace{1cm} (4)

- Generalized beta distribution
- Parametric exponential family
- Assume the ground truth is located in the **center** of the simplex

- Three types of crowds standard 2-simplex space

(a) Adversarial  \hspace{1cm} (b) Spammer  \hspace{1cm} (c) Hammer
Theorem 1.

For fixed $l > 1, r > 1$ and dimension $D \geq 1$, assume that $m$ tasks are assigned to $n$ workers according to a random $(l, r)$-regular bipartite graph. If the average quality satisfies $q > (1 + (D + 1)/\hat{l}\hat{r})$, then when $k \to \infty$ the average error of the our algorithm achieves

$$E_{\text{ALG}} \leq \left( \frac{(1 + 1/\hat{l}\hat{r})^2}{(\sqrt{2} + 1)q\hat{r}} \right) \cdot \frac{1}{\hat{l}m} \sum_{i \in [m]} T_i, \quad (5)$$

- where $q$ denotes the average reliability of crowds.
Performance Analysis

- **Oracle Estimator**
  - Who *already know* the reliability of every worker
  - Theoretical *lower* bound under Dirichlet crowd model

- **Verification with Synthetic Dataset**

- Comparison of average errors between different methods varying the reliability of crowds
Experimental Results - Real-world Dataset 1

- **Image Object Localization**
  - 2,000 arbitrary images sampled from MSCOCO Dataset
  - For each image, assign $l = 25$ distinct workers
  - Crowdsourcing Web Platform: *Figure Eight*

(a) Ground truth  (b) Responses  (c) MV  (d) Ours

- **Question:** Find the bat and draw a bbox as tight as possible.
Experimental Results - Real-world Dataset 2

• **Human Pose Estimation**
  - 1,000 arbitrary images sampled from LSPET Dataset
  - Irregular bipartite graph (general task assignment)

• **Question:** Mark distinct dots on the 14 human joints in the image.
Experimental Results

- As task degree $l$ increases, performance gap becomes larger.
- If task degree $l$ becomes more than 15, performance gain is saturated.
## Experimental Results

### Performance Comparison

| Dataset | MSCOCO       | LSPET       |
|---------|--------------|-------------|
| Type    | Box($\ell_2$) | Box(IoU)    | Joints | Angles |
| $\mathcal{WV}$ | 0.22227      | 0.89593     | 0.15877 | 0.10524 |
| $\mathcal{MV}$ | 0.22090      | 0.89666     | 0.15858 | 0.10462 |
| $\mathcal{IP}$ | 0.22026      | 0.89712     | 0.15483 | 0.10462 |
| Welinder | 0.21886      | 0.89821     | N/A     | N/A     |
| DALE    | 0.21834      | 0.89914     | N/A     | N/A     |
| Top-K   | 0.18869      | 0.91250     | 0.12222 | 0.10051 |
| MeanShift | 0.18034      | 0.92150     | 0.11812 | 0.09962 |
| Ours    | **0.14837**  | **0.93445** | **0.09308** | **0.09941** |
Conclusion

• We propose a new aggregation method for vector regression tasks which are generally handled in real crowdsourcing system.

• Our algorithm is robust to Spammer, even adversarial worker distinguishing the reliabilities of workers.

• Through extensive experiments, we observed the considerable gains of our approach with real crowd-sourced dataset.
1. Karger, David R., Sewoong Oh, and Devavrat Shah. “Iterative learning for reliable crowdsourcing systems.” Advances in neural information processing systems. 2011.

2. Lee, Donghyeon, et al. “Reliable multiple-choice iterative algorithm for crowdsourcing systems.” ACM SIGMETRICS Performance Evaluation Review. Vol. 43. No. 1. ACM, 2015.

3. Zhou, Yao, and Jingrui He. “Optimizing the Wisdom of the Crowd: Inference, Learning, and Teaching.” arXiv preprint arXiv:1806.09018 (2018).