Spammer Detection Based on Task Completion Time Variation in Crowdsourcing

Ayato Watanabe
Graduate School of Informatics, Kyoto University

Keishi Tajima
Graduate School of Informatics, Kyoto University

Research Background

- Spammer detection in crowdsourcing is an important research issue to guarantee the quality of results.
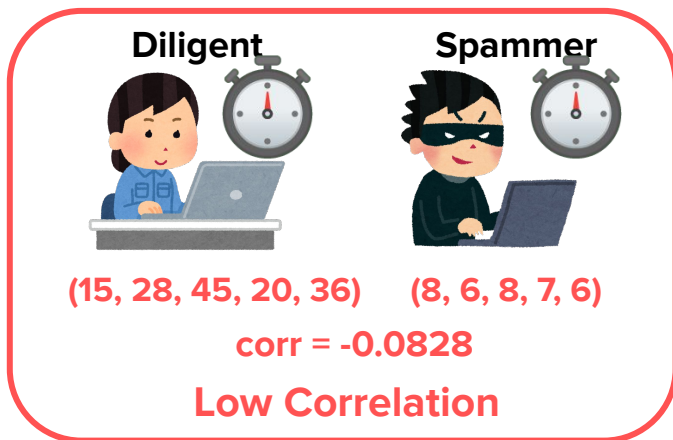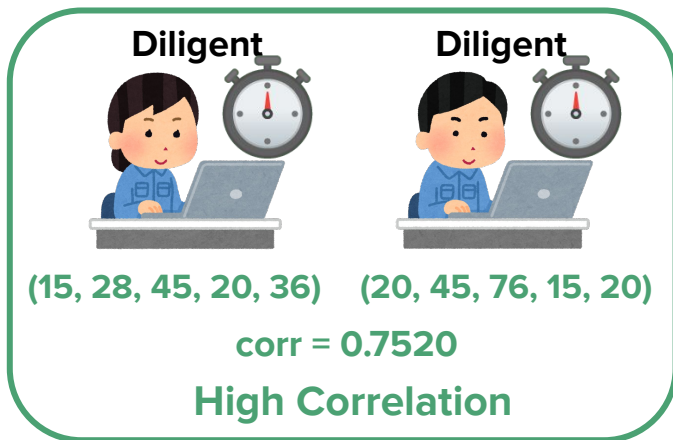
Diligent Worker **Spammer**

Proposed Method

- Assumption:
  - diligent workers take shorter time at easy tasks and longer time at difficult task
  - spammers take short constant time for every task
- We calculate Pearson correlation coefficient between task completion time of a worker and that of the other workers.
- We expect the correlation coefficient value is high between diligent workers and low between a diligent worker and a spammer.

Diligent   **Diligent**

(15, 28, 45, 20, 36)   (20, 45, 76, 15, 20)

corr = 0.7520

High Correlation

Diligent   **Spammer**

(15, 28, 45, 20, 36)   (8, 6, 8, 7, 6)

corr = -0.0828

Low Correlation

Related Work for Spammer Detection

- Rayker and Yu [1] proposed a method of calculating the spammer scores of workers based on how much their answers depend on the true answers.
- Kazai et al. [2] and Chen et al. [3] have used the workers' average task completion time for spammer detection.
- However, there is no spammer detection method using the correlation between the difficulty of task and the task completion time.

[1] V. C. Raykar and S. Yu, "Eliminating spammers and ranking annotators for crowdsourced labeling tasks," JMLR, vol.13, no.1, pp. 491–518, 2012.

[2] G. Kazai, J. Kamps, and N. Milic-Frayling, "Worker types and personality traits in crowdsourcing relevance labels," in CIKM, 2011, p. 1941–1944.

[3] X. Chen, "A real time anti-spamming system in crowdsourcing platform," in ICSESS, 2016, pp. 981–984.

Experiment

- We posted a image classification task on Amazon Mechanical Turk
- Workers were asked to classify 70 images into the following seven categories:
    - Samoyed, German Shepherd, Siberian Husky, Alaskan Malamute, Gray wolf, Coyote, Dhole
    - 10 images in each category
- We recorded the task completion time and the label for each image.
- We collected data of 199 workers.

Experiment

- We sort the 199 workers by the spammer score by Raykar and Yu [1],
  and examine the Pearson correlation coefficient
  between task completion time of low/high score workers.
- We replace outlier values exceeding 60 seconds with 60 seconds.

[1] V. C. Raykar and S. Yu, "Eliminating spammers and ranking annotators for crowdsourced labeling tasks," JMLR, vol.13, no.1, pp. 491–518, 2012.

Result

- Among seven classes, Samoyed is the easiest to distinguish, and German Shepherd is the second.
- The heatmap (a) shows the correlation between workers' task completion time for 10 images in Samoyed class, where we can find no clear pattern.
- The heat map (b) shows the correlation for 10 images in German Sheperd class, where we can see more blue or green dots in the top-left quarter.
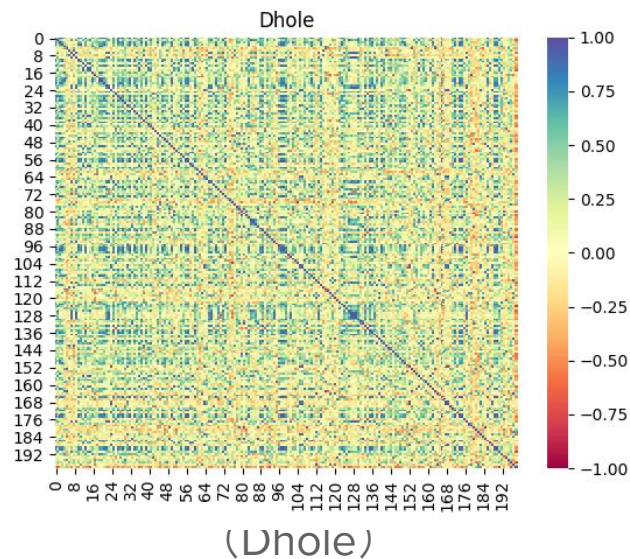

(a) Samoyed

(b) German Shepherd
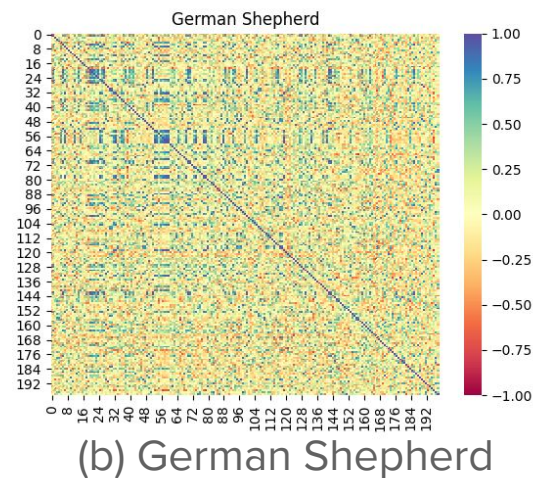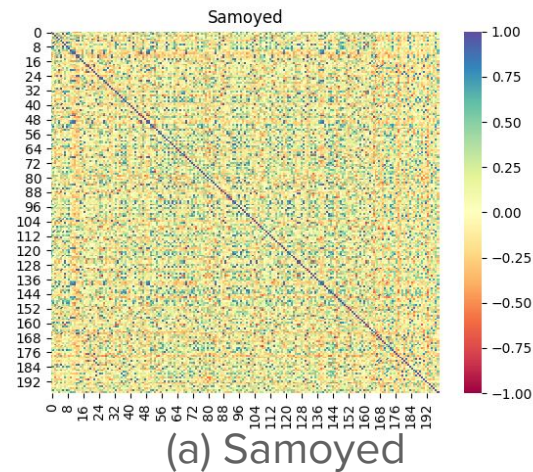
Result

- The heat map for Dhole shows higher correlation
- This is because there was an image with huge
  data size (6.60MB) in Dhole category,
  and it took time for every worker to load it through
  the network.

(Dhole)

careful selection of tasks may be a key to success

Second Experiment

- We calculated the correlation of workers' task completion time
  in these 4 cases:
    - （a）5 images with the highest accuracy and 5 images with the lowest accuracy
    - （b）5 images with the largest variance of completion time and 5 with the smallest
    - （c）worker's average task completion time for 10 images in each category
    - （d）worker's average task completion time for images classified into each category
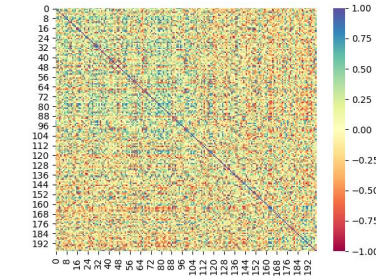
Result

- We expected the top-left area of heat maps has more blue or green dots than other areas.
- In the heat map (d),
  - the average correlation between 120 good workers is **0.1391**, and
  - that between one of the 120 good workers and one of the worst 20 workers is **-0.004**.
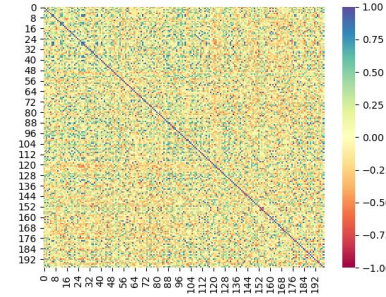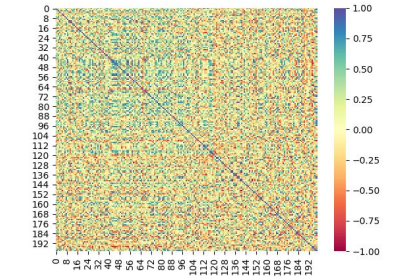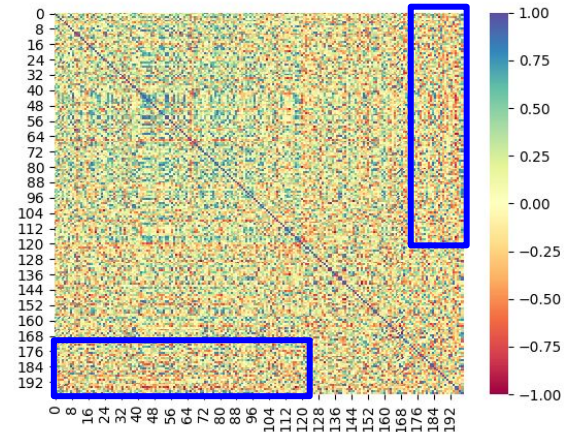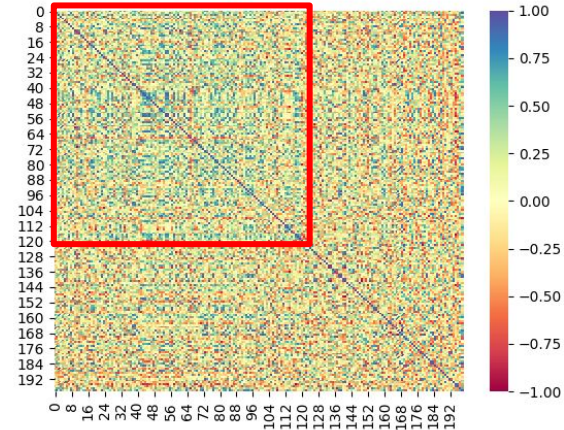

(a)
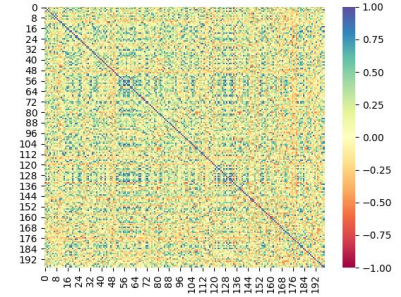
(b)

(c)

(d)

Result

- We expected the top-left area of heat maps has more blue or green dots than other areas.

- In the heat map (d),
  - the average correlation between 120 good workers is **0.1391**, and
  - that between one of the 120 good workers and one of the worst 20 workers is **-0.004**.

Discussion

- We need to choose tasks the completion time of which is more clearly different between a diligent worker and a spammer in order to distinguish spammers.

Conclusion

- In this research, we proposed to use the correlation between workers' task completion time to detect spammers.
- Our experimental result suggests that this approach is potentially useful, but the task design and the selection of tasks seem keys for success.