Towards Instantaneous Recovery from Autonomous System Failures via Predictive Crowdsourcing

John Joon Young Chung, Fuhu Xiao, Nikola Banovic, Walter S. Lasecki
Computer Science & Engineering | University of Michigan, Ann Arbor
{jjyc, fuhu, nbanovic, wlasecki}@umich.edu

ABSTRACT
Autonomous systems (e.g., long-distance driverless trucks) aim to reduce the need for people to complete tedious tasks. In many domains, automation is challenging because systems may fail to recognize or comprehend all relevant aspects of its current state. When an unknown or uncertain state is encountered in a mission-critical setting, recovery often requires human intervention or hand-off. However, human intervention is associated with decision (and communication, if remote) delays that prevent recovery in low-latency settings. Instantaneous crowdsourcing approaches that leverage predictive techniques reduce this latency by preparing human responses for possible near future states *before* they occur. Unfortunately, the number of possible future states can be vast and considering all of them is intractable in all but the simplest of settings. Instead, to reduce the number of states that must later be explored, we propose the approach that uses the crowd to first predict the most relevant or likely future states. We examine the latency and accuracy of crowd workers in a simple future state prediction task, and find that more than half of crowd workers were able to provide accurate answers within one second. Our results show that crowd predictions can filter out critical future states in tasks where decisions are required in less than three seconds.

Author Keywords
Real-time Crowdsourcing; Human Computation; Prediction.

CCS Concepts
•**Human-centered computing** → *Collaborative interaction;*

INTRODUCTION
Autonomous systems hold the potential to take over tedious or dangerous tasks from humans. However, these systems are not perfect and may fail in mission-critical settings [3, 4]. For example, when an autonomous vehicle fails to recognize an object in its environment, this failure can lead to a collision within a few seconds [2]. In such cases, it is common practice for humans to take over the control from the system [8] — but this typically takes several (3 or more) seconds, even in the best case [8], which may not be quick enough to prevent accidents. Remote human intervention similarly may be of limited use due to delays introduce by recruitment [1] and network latency [11].

Instantaneous crowdsourcing [9] is an accelerated form of real-time crowdsourcing [7, 5, 1, 10] that has the potential to address this latency. Predictive crowdsourcing prefetches human responses for all possible near-future states, and can be used to enable instantaneous crowdsourcing. However, with a large number of near-future states to be considered, prefetching responses for all future states would make the approach inefficient because it would require a large number of workers.

To reduce the number of future states that a system needs to consider, we propose an approach in which we ask crowd workers to predict *possible* near-future states. We explore the feasibility of this approach by studying crowd workers' ability to make accurate, low-latency predictions. We focused on the task of predicting whether or not an unidentified object *could* be dangerous in an autonomous driving scenario, as it will serve as the first signal on whether the vehicle should consider future states of the object or not. Our results show that crowd workers can accurately predict future states with a latency of ∼1 second, with precision and recall both at or above 0.87. Our results are a step towards using human prediction to quickly and accurately reduce the size of the reachable state space for predictive crowdsourcing approaches.

EXPERIMENTAL RESULTS
Our proposed workflow leverages remote crowd workers when an autonomous system fails to recognize part of its environment. For example, when an autonomous vehicle fails to recognize an object, that object could be many possible things and cause many possible future states. To reduce the number of possible futures that a predictive crowdsourcing system must consider, the vehicle can first ask crowd workers if the object is dangerous or not. This can significantly reduce the number of scenarios that require direct human oversight.

We hypothesize that crowd workers can quickly and accurately decide if an object is dangerous. To test their ability to predict relevant futures, we asked workers to watch videos from the perspective of the driver that contain a highlighted object that could not be recognized. The object could cause future danger (e.g., near-by humans or vehicles that could move into the

UIST '19 Adjunct, October 20-23, 2019, New Orleans, LA, USA.

ACM ISBN 978-1-4503-6817-9/19/10 ...$15.00.
http://dx.doi.org/10.1145/3332167.3357100

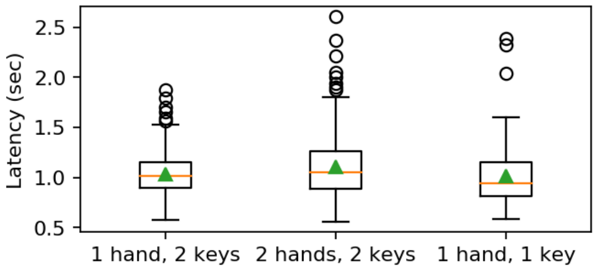

Figure 1. Latency in getting human responses for three input conditions.

vehicle's path) or not (e.g., trees or fire hydrants, which cannot move). We asked workers to answer the question "Will the object possibly collide with our vehicle and cause an accident in the near future?" as quickly as they could accurately.

Some response latency is due to the workers' motor speed when answering. Thus, we examined three input methods for answering yes-or-no questions: 1) using one hand with left and right arrow keys (`1H2K`), 2) using two hands with left and right shift keys (`2H2K`), and 3) using only one key, the space bar, for answering "yes" and using no-input as "no" (`1H1K`).

We collected data from 60 US-based crowd workers who had task acceptance rate above 97% from Amazon Mechanical Turk using LegionTools [6]. We randomly assigned the workers into one of the three conditions above (20 per condition).

We introduced the task to the crowd workers and provided them with two-part training. First, we asked them to make a simple binary decision by pressing on keys on their keyboard in response to a prompt within a short period of time. The prompt was a short text saying either "Movable" or "Won't Move". The crowd workers had to press on the key corresponding to the prompt within 0.8 second. Each crowd worker completed 20 training prompts. Second, we asked them to watch and answer the object question in two tutorial videos.

The crowd workers then moved on to the main task and answered the question about objects in 10 videos each. Five videos contained an object that can be dangerous and five videos contained an object that is not dangerous to the vehicle. The order of the videos was randomized.

We measured the prediction latency by recording how long after the start of the video the workers gave inputs. We also measured the correctness of each prediction.

Our results showed that workers could quickly predict whether an object can be dangerous or not (Figure 1). The ratio of workers who could answer before the end of the video was 100%, 99.5%, and 89% for `1H2K`, `2H2K`, and `1H1K`, respectively. The average latency was 1.03, 1.11, and 1.01 seconds for `1H2K`, `2H2K`, and `1H1K`, respectively. For calculating the average latency, we did not include the latency of negative inputs of `1H1K`, or consider workers who could not provide an answer before the end of the video.

More than half of workers were capable of giving inputs within 1.06 seconds, regardless of the condition. There was a significant difference in latency between conditions (Kruskal-Wallis, $Chi-Square = 8.80$, $p < 0.05$), and `1H1K` significantly outperformed `2H2K` (Mann-Whitney $U = 7042$, $p < 0.05/3$ with Bonferroni correction). Accuracy, precision, and recall were also high, all being above or equal to 0.87. With Cochran's Q test for accuracy, there was no significant difference between the conditions ($N = 200$, $Q = 3.59$, $p > 0.1$).

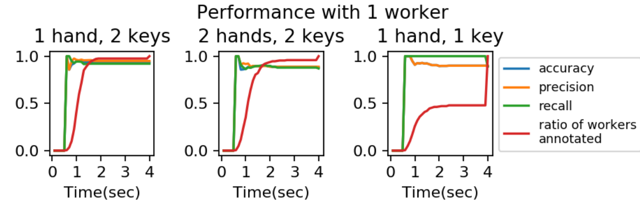

Figure 2. Performance of individual workers, according to time.

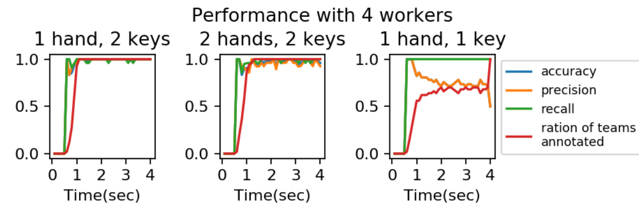

Figure 3. Performance of teams with four workers, according to time.

We also analyzed accuracy, precision, and recall of teams with different numbers of workers over time (Figure 2 and 3). We set the agreement threshold to 0.5, meaning that more than half of the workers who answered had to agree on the answer. For each team, we considered the inputs of workers as soon as they provided them. This means that we could reach an agreement before all of them answered. With this method, for the `1H2K` and the `2H2K` conditions, we could get reliable performance as quickly as 1 second. The recall was the most robust in the `1H1K` condition across time, but the precision and the accuracy dropped with the addition of workers and the increase in latency.

CONCLUSION AND FUTURE WORK

We found that crowd workers can generate accurate responses with low latency for the task of predicting whether an object can be dangerous. With the aggregation, using 2 keys yields better precision and accuracy than using 1 key. To efficiently prevent autonomous system failures with instantaneous crowdsourcing based on predictive techniques, we will extend this investigation to predict critical future states. For example, to predict more precise future states automatically, we can use crowdsourced information to rapidly build behavioral models of the object. With a predictive crowdsourcing workflow, we can avoid accidents arising from failures of autonomous systems either by informing the system of possible dangerous states or human contributors.

ACKNOWLEDGMENTS

This work was supported in part by the USDOT Center for Connected and Automated Transportation (CCAT) at the University of Michigan, and by a DARPA Young Faculty Award.

REFERENCES

[1] Michael S. Bernstein, Joel Brandt, Robert C. Miller, and David R. Karger. 2011. Crowds in Two Seconds: Enabling Realtime Crowd-powered Interfaces. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology (UIST '11)*. ACM, New York, NY, USA, 33–42. DOI: http://dx.doi.org/10.1145/2047196.2047201

[2] National Transportation Safety Board. 2018. Preliminary report Highway HWY18MH010. https://www.ntsb.gov/investigations/AccidentReports/Reports/HWY18MH010-prelim.pdf. (2018). Accessed: 2019-07-10.

[3] Lee Gomes. 2014. Hidden obstacles for Google's self-driving cars. *MIT Technology Review* (2014).

[4] Philip Koopman and Michael Wagner. 2017. Autonomous Vehicle Safety: An Interdisciplinary Challenge. *IEEE Intelligent Transportation Systems Magazine* 9, 1 (Spring 2017), 90–96. DOI: http://dx.doi.org/10.1109/MITS.2016.2583491

[5] Walter Lasecki, Christopher Miller, Adam Sadilek, Andrew Abumoussa, Donato Borrello, Raja Kushalnagar, and Jeffrey Bigham. 2012. Real-time Captioning by Groups of Non-experts. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology (UIST '12)*. ACM, New York, NY, USA, 23–34. DOI: http://dx.doi.org/10.1145/2380116.2380122

[6] Walter S. Lasecki, Mitchell Gordon, Danai Koutra, Malte F. Jung, Steven P. Dow, and Jeffrey P. Bigham. 2014. Glance: Rapidly Coding Behavioral Video with the Crowd. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology (UIST '14)*. ACM, New York, NY, USA, 551–562. DOI: http://dx.doi.org/10.1145/2642918.2647367

[7] Walter S. Lasecki, Kyle I. Murray, Samuel White, Robert C. Miller, and Jeffrey P. Bigham. 2011. Real-time Crowd Control of Existing Interfaces. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology (UIST '11)*. ACM, New York, NY, USA, 23–32. DOI: http://dx.doi.org/10.1145/2047196.2047200

[8] Lutz Lorenz, Philipp Kerschbaum, and Josef Schumann. 2014. Designing take over scenarios for automated driving: How does augmented reality support the driver to get back into the loop? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 58, 1 (2014), 1681–1685. DOI: http://dx.doi.org/10.1177/1541931214581351

[9] Alan Lundgard, Yiwei Yang, Maya L. Foster, and Walter S. Lasecki. 2018. Bolt: Instantaneous Crowdsourcing via Just-in-Time Training. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 467, 7 pages. DOI: http://dx.doi.org/10.1145/3173574.3174041

[10] Elliot Salisbury, Sebastian Stein, and Sarvapali Ramchurn. 2015. Real-time Opinion Aggregation Methods for Crowd Robotics. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems (AAMAS '15)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 841–849. http://dl.acm.org/citation.cfm?id=2772879.2773261

[11] Steven Eric Vozar. 2013. A Framework for Improving the Speed and Performance of Teleoperated Mobile Manipulators. *PhD thesis* (2013).