Crowdsourcing intelligence for improving disaster forecasts

Xiaohui Wang,[1,6] Fuhai Dao,[2,6] Yatai Ji,[3,6] Sihang Qiu,[3,6] Xian Zhu,[2] Wenjie Dong,[2] Huizan Wang,[1,*] Weimin Zhang,[1,*] and Xiaolong Zheng[4,5,*]

[1]College of Meteorology and Oceanography, National University of Defense Technology, Changsha 410073, China

[2]School of Atmospheric Science, Sun Yat-sen University, Zhuhai 519082, China

[3]College of Systems Engineering, National University of Defense Technology, Changsha 410073, China

[4]The School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 101408, China

[5]The State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

[6]These authors contributed equally

*Correspondence: wanghuizan17@nudt.edu.cn (H.W.); weiminzhang@nudt.edu.cn (W.Z.); xiaolong.zheng@ia.ac.cn (X.Z.)

Received: January 3, 2024; Accepted: May 19, 2024; Published Online: May 22, 2024; https://doi.org/10.1016/j.xinn.2024.100647

Citation: Wang X., Dao F., Ji Y., et al., (2024). Crowdsourcing intelligence for improving disaster forecasts. The Innovation **5(4)**, 100647.

INTRODUCTION

Natural disasters, including floods, storms, and tsunamis, pose a great threat to human societies. A recent study highlighted this concern, revealing that billions of people globally were exposed to flood hazards.[1] In 2023, Super Typhoon Doksuri caused devasting floods in Beijing and Hebei areas, resulting in massive casualties and huge economic losses. Therefore, there is a need for a precise understanding of disaster processes, reliable forecasting of disaster effects, and timely warning of risks to prevent and mitigate major disasters.[2] Numerical modeling stands as the predominant approach to meet these demands. However, the predictive accuracy of such numerical models could be degraded because of various factors: oversimplification of real processes, computational errors, fluctuations of complex environments (e.g., terrains, precipitations, buildings, and plants), and the influence of human activities (e.g., evacuation and rescue) during disasters.

Improving models with a large number of observational data through methods like machine learning or data assimilation is widely recognized as an effective approach. Nevertheless, the lack of disaster-related data and the practical difficulty encountered in data collection, particularly during emergencies, have become major challenges in disaster forecasting. In addition, even after data collection, there is also a research gap in terms of identifying model drawbacks and acquiring improvement solutions based on limited datasets. Crowdsourcing is a rising approach employing crowds to complete various tasks.[3] By soliciting insights from people both physically located in disaster zones and virtually connected online, it holds potential to obtain disaster data and pave the way for model improvements. Yet, the question of how to effectively apply crowdsourcing to improve disaster forecasts remains unexplored.

To bridge existing research gaps, in this work, we propose a concept named "crowdsourcing intelligence." It leverages the wisdom of crowds and artificial intelligence through crowdsourcing activities, to show unprecedented advantages when juxtaposed against traditional computing methods solely relying on either human intelligence or machine algorithms. Crowdsourcing intelligence shows certain advantages and unique features for disaster forecasts, in comparison with typical hybrid intelligence methods. First, crowdsourcing intelligence recruits a larger number of participants and is underpinned by a hierarchical structure involving various human roles to optimize disaster management. Furthermore, in pursuit of wider participation, crowdsourcing intelligence aims to set a lower participation barrier and to minimize task load, which is particularly suitable for non-expert and untrained residents in disaster areas. In addition, crowdsourcing intelligence can solicit voluntary contributions from complex social networks, to reflect social reactions promptly, precisely, and comprehensively. In this work, we designed a universal pipeline for improving state-of-the-art numerical disaster models using crowdsourcing intelligence.

IMPROVING STATE-OF-THE-ART NUMERICAL MODELS FURTHER

Numerical models are so far the most commonly used method to mathematically describe and explain natural disasters. They use fundamental physical laws with various theoretical hypotheses and environmental elements (e.g., meteorological and oceanic elements) to simulate disasters and predict their effects. However, models always produce errors, because modeling disasters essentially entails approximating natural phenomena using mathematics and computer programs under the principle of similarity, which inevitably leads to errors. Such errors could come from simplification of real disaster processes, inaccurate input elements, numerical computations, and factors that are difficult to incorporate during the modeling phase (such as complex environmental factors and human activities). Acknowledging these unavoidable errors, we design a universal pipeline aiming at improving predictive models in disaster forecasts by leveraging crowdsourcing intelligence. The pipeline consists of problem identification, causal analysis, and forecast improvement, as shown in Figure 1.

Crowdsensing for problem identification

The initial step in improving forecasts lies in identifying problems that a disaster model has. The step aims to collect necessary disaster observations comprehensively and, by comparing observations against model outputs, to find and identify model problems. This step is achieved through crowdsensing, which is a domain application of crowdsourcing intelligence. Crowdsensing uses crowd participants (including humans, vehicles, robots, and virtual agents) to execute sensing (data-acquiring) tasks by visiting the targeted regions and collecting on-demand observations physically.[4] We can obtain multi-dimensional error distributions/maps representing where or when the model performs problematically by comparing crowdsensed observations with model outputs.

Crowd-reasoning for causal analysis

In this step, we leverage human intelligence from crowdsourcing to reason the causes of model problems. We introduce a novel concept—crowd-reasoning—considered as a promising approach to achieve causal analysis with small samples. This step initiates with human participants finding potential problem causes (for instance, using an Ishikawa diagram). Afterward, it analyzes causal relationships by employing AI algorithms or crowdsourcing. We recommend two task designs for crowd-reasoning. Crowd-mapping enables crowds to label potential environmental and geographical causes, with visual assistance by using maps and street-view imagery. Causal relationships between model problems (identified in step one) and potential causes could be consequently analyzed using AI/statistical tools. Another task design is error-matching. It requires the model to generate plenty of hypothetical error distributions in advance, by intentionally introducing variations into model parameters and model structures. The crowd is asked to match actual error distributions (acquired from step one) with those artificially created error distributions. Through error-matching, the most plausible causes can be found.

Improving disaster forecasts

After identifying model problems and reasoning their causes through crowdsourcing, the third step is to improve disaster forecasts. We provide two main strategies. *Model correction* improves disaster forecasts by updating the model structure, modifying the model parameters, or fixing training data biases. The goal is to fix model problems that should be (but were not) considered during modeling phases, such as oversimplification of physical processes, systematic biases, and unreasonable parameters. It improves forecasts by updating the model parameters/structure/data through white-box, gray-box, or black-box system identification methods. *State correction* is another strategy to improve disaster forecasts, which directly fixes the system state, rather than making changes to models. The goal is to overcome the unexpected effects of complex, dynamic, or stochastic factors that exist in the system state, which are challenging to incorporate during modeling phases. Based on the problem causes found in step two, we can predict the state errors they cause and, accordingly, eliminate them in disaster forecasts.

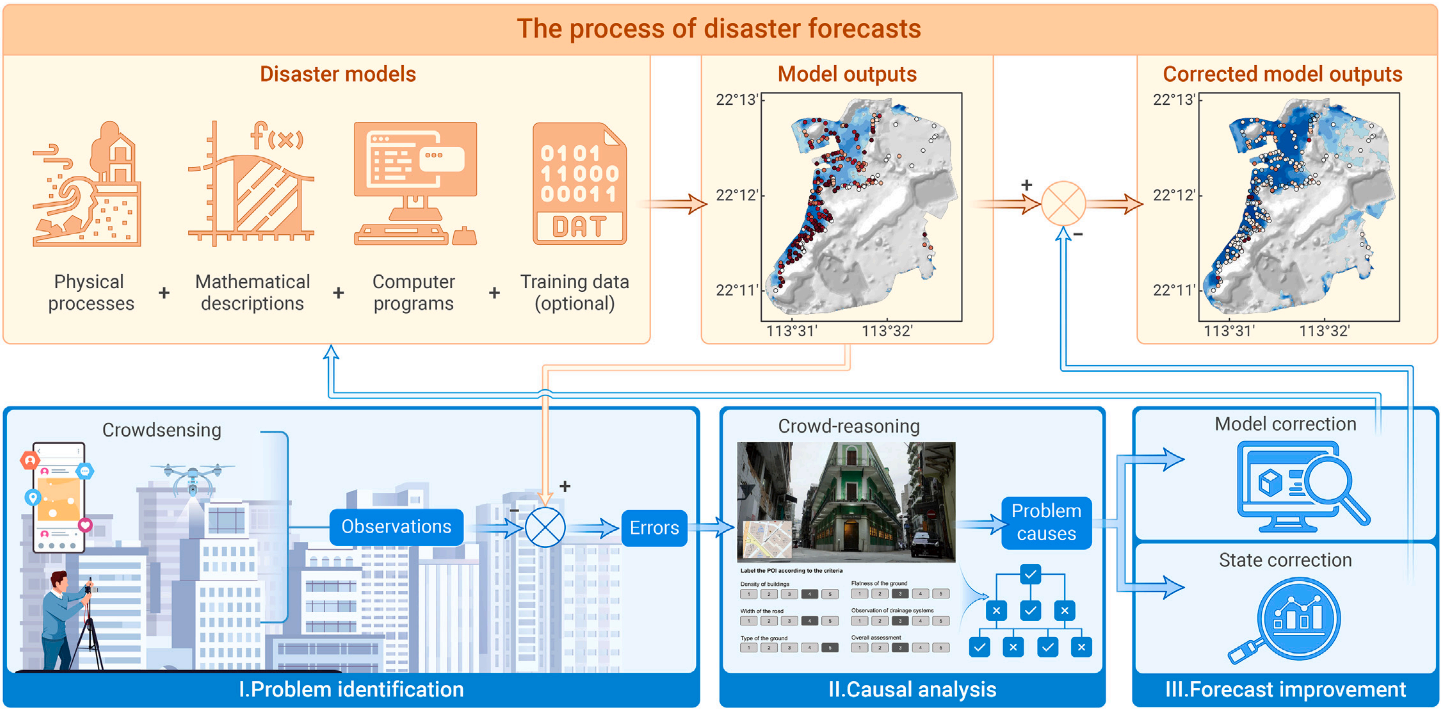

Figure 1. A universal pipeline for improving disaster forecasts integrating two intelligent crowdsourcing approaches: crowdsensing and crowd-reasoning Crowdsensing in disaster regions identifies model errors, and crowd-reasoning analyzes the causes of the errors. Afterward, disaster models can be improved through two error correction approaches.

INSIGHTS FOR DISASTER PREVENTION

We validated our proposed pipeline using two flood events, yielding valuable insights, particularly pertinent to flood prevention. Typhoon Hato and Typhoon Mangkhut hit China's Pearl River Delta in 2017 and 2018, respectively. These two record-breaking extreme events caused Macau's two largest flood disasters since 1925. We replicated the physical process of these disasters using ADCIRC (https://adcirc.org). Crowdsourcing campaigns were carried out, to first detect errors and then find error causes. Afterward, we built machine learning models and parameter estimation methods, using crowd-labeled error causes as inputs, to predict and then eliminate state errors and model errors. To this end, the error of corrected predictions could be reduced to a sufficiently low level. The details of this preliminary experiment are available online publicly at https://osf.io/wk965/.

Our crowdsourced causal analysis based on crowdsensing observations highlighted crucial considerations for flood prevention, summarized as follows: (1) *narrow or steep urban canyons*: These types of streets can dramatically increase the risk of local inundation and threaten the reliability of prediction. (2) *Low-altitude areas distant from water*: the prediction accuracy decreases sharply as the distance from water increases. Low-altitude places away from water are at risk of inundation but are usually neglected. (3) *Huge artificial structures* (e.g., nearshore seawalls): they could bring up severe systematic biases to flood predictions, even though the model was well calibrated by meteorological and hydrological data.

DISCUSSION

In this work, we have shown the feasibility of using the power of crowds to identify model problems and the effectiveness of leveraging crowd intelligence to analyze their causes, with the aim of improving disaster forecasts. In terms of method and design, this work shows its novelty in integrating multiple crowdsourcing approaches and aggregating heterogeneous data in disaster forecasts. In terms of implementations and applications, this work has demonstrated that crowdsourcing intelligence could be an important approach to further improve the performance of a state-of-the-art numerical model and to provide more reliable disaster predictions. Furthermore, we outline the future integration of large-scale AI models and the importance of devising incentive mechanisms for future crowdsourced disaster prevention.

Integrating large-scale AI models

We emphasize the great potential of our pipeline in terms of facilitating advancements in disaster modeling and management by integrating large-scale AI models.[5] We also highlight research opportunities for using large language models (LLMs) in our pipeline. Through a suitable LLM reasoning-acting framework, it is possible to involve AI agents in every stage of the crowdsourcing pipeline, consequently alleviating human workload and making the crowdsourcing process autonomous. Moreover, we envision that our work will serve as an indispensable data source for AI training or for modeling human/environmental impacts.

Crowdsourcing incentives

We acknowledge the necessity of actively and effectively involving crowds in disaster areas to harness the strength of crowdsourcing intelligence in disaster forecasts. It directs us toward exploring crowdsourcing incentives. Apart from traditional monetary incentive designs that are prevalent in online crowdsourcing marketplaces, we anticipate that voluntary incentives will play an important role in the future. Through a reasonable incentive design, we can motivate social contributors (particularly residents in the disaster area) to participate actively, by sharing valuable on-the-ground data/information through crowdsourcing activities. Reciprocally, these contributors would gain access to high-quality public services (such as monitoring real-time disaster situations and receiving reliable risk alerts) supported by their contributions to enhance their safety during crises.

REFERENCES

1. Tellman, B., Sullivan, J.A., Kuhn, C., et al. (2021). Satellite imaging reveals increased proportion of population exposed to floods. Nature ***596***(7870): 80–86. https://doi.org/10.1038/s41586-021-03695-w.
2. Yang, W., Zhang, J., Hua, P., et al. (2023). Global framework for flood risk management under climate change and urbanization. The Innovation Geoscience ***1***(1): 100009. https://doi.org/10.59717/j.xinn-geo.2023.100009.
3. Howe, J. (2006). The rise of crowdsourcing. Wired magazine ***14***(6): 176–183.
4. Guo, B., Wang, Z., Yu, Z., et al. (2015). Mobile Crowd Sensing and Computing: The Review of an Emerging Human-Powered Sensing Paradigm. ACM Comput. Surv. ***48***(1): 1–31. Article 7. https://doi.org/10.1145/2794400.
5. Xu, Y., Liu, X., Cao, X., et al. (2021). Artificial intelligence: A powerful paradigm for scientific research. Innovation ***2***(4): 100179. https://doi.org/10.1016/j.xinn.2021.100179.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (nos. 42306217, 62202477, 72225011, and 42276205), Hunan Provincial Natural Science Foundation of China (2023JJ10053), and the National Key Research and Development Program of China (2021YFC3101500).

DECLARATION OF INTERESTS

The authors declare no competing interests.