

2023 IEEE MLSP Data Challenge

Call for Participation: Volunteer Retention and Future Collaboration Prediction

In recent years, there has been a rise in online crowdsourcing platforms that facilitate volunteering efforts, enabling individuals to collaborate on collective goals. Such efforts can play an instrumental role in responding to community needs, especially during times of crisis, such as earthquakes and pandemics. To optimize these efforts, it is crucial to model volunteers' task participation and collaboration behavior. In this challenge, we aim to learn such a model from data using machine learning and signal processing techniques. We present a volunteer participation dataset with two tasks.

Volunteer Dataset: The dataset is collected from a mobile crowdsourcing app called “Anti-Pandemic Pioneer” used for organizing volunteers during the COVID-19 pandemics in Shenzhen, China. It contains records of each volunteer participating in a group activity (task). Each record includes features such as volunteer ID, task ID, timestamp, task location, and etc. Task names and descriptions will also be provided.

Task 1: Retention Prediction: Predict the number of future participations of volunteers based on their participation history. Groundtruth labels will be provided as a list of “Volunteer ID - # of future participation” pairs. This task focuses on modeling volunteer retention, an important research area in social and management science. Improving volunteer retention enhances the overall volunteering experience and encourages sustained engagement.

Task 2: Future Collaboration Prediction: Predict whether two volunteers will collaborate in future tasks.. The participation records of volunteers as well as the ground truth edge list of their participation graph will be provided. Future collaboration prediction has diverse applications, such as understanding volunteer group behaviors and developing social recommendation tools for volunteers.

Key Points & Competition Details:

- Participants can choose to participate in either one or both tasks.
- The datasets for both tasks have identical structure but sampled from distinct districts, ensuring no overlap.
- The evaluation metrics are Mean Squared Error for Task 1 and Area Under Precision-Recall Curves (AUPR) for Task 2.
- Both datasets are split into a train set (5,000 volunteers) and a test set (5,000 volunteers). The test sets are further split into private test set (2/3 for final evaluation) and public test set (1/3 for getting feedbacks during the competition).
- Participants can access all datasets but only labels for the train sets.
- Submissions can be made for the public test set twice a day, and immediate scores

(MSE and AUPR) will be provided.

- Submission deadline will be two months after the start of the competition.