Privacy-Preserving Crowdsourcing

論文の内容の要旨

Abstract

- 論文題目:
  Privacy-Preserving Crowdsourcing
  （プライバシ保護クラウドソーシング）
- 氏名: 梶野 洸

Crowdsourcing is an idea in which requesters outsource tasks to unspecified workers via the Web. A basic procedure can be described using the following three steps. In the assignment step, a crowdsourcing platform matches tasks and workers, employing either a push-type or pull-type assignment strategy. The push-type assignment strategy uses the platform to assign tasks to appropriate workers based on the features of the workers and tasks (*e.g.*, skills, preferences of tasks, and minimum wages), while the pull-type assignment requires workers to choose tasks they like. In the request step, each requester sends a job instruction and task instances (*e.g.*, audio files in case of an audio transcription task) to the allocated workers. Finally, in the delivery step, each worker, having processed the assigned task, sends the results back to the requester. Crowdsourcing provides requesters with easy access to a huge pool of workers and enables workers to work much more flexibly than in the traditional labor market. These unique advantages have led to a number of real applications and businesses as well as new research opportunities such as human computation.

Despite its revolutionary power, it is often pointed out that using crowdsourcing entails several risks including the risk of poor quality task results. Among others, this thesis focuses on the privacy risks. Although

the privacy risks in crowdsourcing have been pointed out in diverse domains, little has been investigated until now. Toward establishing a research basis for privacy-preserving crowdsourcing, this thesis addresses the following two research questions:

- *What types of privacy risks are present in crowdsourcing?*
- *How can we measure and control the privacy risks in crowdsourcing?*

To answer the first research question, we carefully examine the three steps of crowdsourcing and discover that four types of data can lead to privacy issues: features (the assignment step), job instruction and task instances (the request step), and task results (the delivery step). Further, by analyzing the applicability of existing privacy preservation strategies, we find that some types of data cannot be handled by the existing approaches, highlighting the novelty of privacy-preserving crowdsourcing research.

Given this finding, we develop the following three solutions to answer the second research question.

(1) Privacy Preservation in the Assignment Step

We present a privacy-preserving task assignment (PTA) protocol, which computes an optimal task assignment, keeping the features of the workers and tasks private. Observing that our task assignment problem can be reduced to the maximum flow problem, the PTA protocol constructs an instance of the maximum flow problem and solves it by harnessing the push-relabel algorithm, both in a privacy-preserving way. Because the PTA protocol significantly decreases the number of workers who receive instructions and instances compared to the standard pull-type task assignment, our protocol also reduces the privacy risks associated with them. We evaluate the computation overhead induced by cryptography and discuss relaxation methods to reduce it.

(2) Privacy Preservation in the Request Step

We present the utility-privacy trade-off analyzer (UPTA), which enables us to evaluate the trade-off between the utility and privacy

of an instance-privacy-preserving (IPP) protocol. Because an instance is used to perform a task as well as to extract the sensitive information contained within it, an IPP protocol in general has to sacrifice utility for privacy. Therefore, it is essential to quantify the trade-off in order to research instance-privacy preservation. The idea of UPTA is to model the task execution and privacy invasion as sampling of a task result and sensitive value from probability distributions. We estimate the models using crowdsourcing and apply divergence-based measures to the estimated models in order to quantify utility and privacy. As a case study of UPTA, we develop an instance-clipping (IC) protocol and analyze its properties. The IC protocol submits a task with clipped instances of a fixed size. We discuss the performance of the IC protocol as well as the validity of UPTA in the experiments.

(3) Privacy Preservation in the Delivery Step

We present a worker-private latent class (WPLC) protocol, which allows a requester to receive task results without compromising the privacy of the workers associated with the task results. The key observation is that a requester often aggregates results to produce quality-controlled results, which are no longer associated with any worker. The WPLC protocol simulates the aggregation procedure using cryptography to output quality-controlled results without disclosing the results of each worker. We discuss the validity of WPLC by evaluating the disadvantages induced by cryptography, including its computation time.