future internet

MDPI

Article

Selecting Workers Wisely for Crowdsourcing When Copiers and Domain Experts Co-exist

Xiu Fang [1,†], Suxin Si [1,†], Guohao Sun [1,*], Quan Z. Sheng [2], Wenjun Wu [1], Kang Wang [1] and Hang Lv [1]

[1] School of Computer Science and Technology, Donghua University, Shanghai 201600, China; xiu.fang@dhu.edu.cn (X.F.); 2202491@mail.dhu.edu.cn (S.S.); 180730223@mail.dhu.edu.cn (W.W.); 2202425@mail.dhu.edu.cn (K.W.); 2202516@mail.dhu.edu.cn (H.L.)

[2] School of Computing, Macquaire University, Sydney, NSW 2109, Australia; michael.sheng@mq.edu.au

* Correspondence: ghsun@dhu.edu.cn

† These authors contributed equally to this work.

Abstract: Crowdsourcing integrates human wisdom to solve problems. Tremendous research efforts have been made in this area. However, most of them assume that workers have the same credibility in different domains and workers complete tasks independently. This leads to an inaccurate evaluation of worker credibility, hampering crowdsourcing results. To consider the impact of worker domain expertise, we adopted a vector to more accurately measure the credibility of each worker. Based on this measurement and prior task domain knowledge, we calculated fine-grained worker credibility on each given task. To avoid tasks being assigned to dependent workers who copy answers from others, we conducted copier detection via Bayesian analysis. We designed a crowdsourcing system called SWWC composed of a task assignment stage and a truth discovery stage. In the task assignment stage, we assigned tasks wisely to workers based on worker domain expertise calculation and copier removal. In the truth discovery stage, we computed the estimated truth and worker credibility by an iterative method. Then, we updated the domain expertise of workers to facilitate the upcoming task assignment. We also designed initialization algorithms to better initialize the accuracy of new workers. Theoretical analysis and experimental results showed that our method had a prominent advantage, especially under a copying situation.

Keywords: crowdsourcing; task assignment; truth discovery; domain; copier

Citation: Fang, X.; Si, S.; Sun, G.; Sheng, Q.Z.; Wu, W.; Wang K.; Lv, H. Selecting Workers Wisely for Crowdsourcing When Copiers and Domain Experts Co-exist. *Future Internet* **2022**, *14*, 37. https://doi.org/10.3390/fi14020037

Academic Editors: Vijayakumar Varadarajan, Rajanikanth Aluvalu and Ketan Kotecha

Received: 20 December 2021
Accepted: 19 January 2022
Published: 24 January 2022

1. Introduction

With the rapid development of mobile Internet and the popularity of intelligent terminal devices, the wide use of crowdsourcing is gradually being integrated into people's lives [1]. Crowdsourcing platforms, such as Amazon Mechanical Turk (AMT (https://www.mturk.com (accessed on 19 December 2021))), have been widely applied in many fields, including video analysis [2], knowledge discovery [3], Smart Citizen (http://www.smartcitizen.me (accessed on 19 December 2021)), and human–robot interaction studies [4]. On the platforms, requesters publish questions to be answered. The participating public obtains material rewards or meets their interests by completing these tasks.

After the answers are collected from all workers, the crowdsourcing platform comprehensively integrates the data to identify the truth. However, different workers may provide conflicting answers. The quality of their answers is also uneven. Some answers may be close to the truth, while some answers may be far from the truth. To resolve the conflicts, many methods have been proposed in the crowdsourcing area, which can be divided into two categories:

- The methods that adopt a redundancy-based strategy [3,5]: Majority voting chooses the answer given by the majority of workers as the estimated truth. Gold-injected methods [6] use a small number of tasks with basic facts to evaluate the credibility of workers. The expectation–maximization-based method [7] evaluates worker credibility and forecasts the truth at the same time;

- The methods [8] that improve accuracy by eliminating bad workers: These methods believe that aggregating answers from a small amount of high-credibility workers may achieve better accuracy than blindly pursuing more workers. A typical method [9] is to use qualification tests to distinguish bad workers and stop assigning tasks to them.

Most current methods quantify worker credibility as a single value [10–12] or a confusion matrix [7,13]. They assume each worker has the same professional level on different tasks. However, each worker has different domain expertise. For example, a basketball fan has higher credibility in answering basketball-related tasks than volleyball-related tasks. Assigning tasks to workers who are not good at the related domain may lead to obtaining a low-quality collection of workers' answers. Therefore, it is very important to consider the credibility of workers in different domains. ARE [14] considers the domain in task assignment to select one expert to complete each task. Actually, in crowdsourcing, it often needs not a worker, but a large number of high-quality workers. MDC [15] aggregates truth by considering domains of workers in the truth discovery stage instead of the task assignment stage, which may assign tasks to low-quality workers. When considering the domain expertise of workers, initializing the accuracy of workers is also a problem that needs to be studied. Most current methods in crowdsourcing and truth discovery initialize the accuracy of workers to a fixed value, hoping that workers can complete a large number of tasks to approach their real accuracy. We propose an initialization method to make the initialization accuracy closer to the real accuracy of workers.

The existing methods also make another assumption that all workers are independent of each other. However, due to the convenience of the Internet and the accessibility of information, it is easy for workers to copy, crawl, and aggregate information published by others. Therefore, copying is common among workers [16]. Answers provided by one worker, no matter whether true or false, can be copied by many other workers. In the case that wrong answers spread among workers via copying, if we do not distinguish copiers, we may identify the wrong answers as the truth. When submitting answers, copiers will not state that answers are obtained by copying others. Therefore, taking into consideration the possible dependence between workers and avoiding assigning tasks to copiers can often lead to more precise crowdsourcing results.

In this paper, to relax the workers' unified credibility assumption, we quantified the credibility of each worker as a vector, with each element demonstrating the worker's expertise on a specific domain. Intuitively, tasks should be assigned to domain experts to obtain more accurate answers. A task may belong to multiple domains. Therefore, we calculated the fine-grained worker credibility for each task based on worker domain expertise and prior task domain classification. To relax the worker independence assumption, we calculated the probability and direction of copying among workers and removed copiers. We propose a system, selecting workers wisely for crowdsourcing (**SWWC**), as an overall solution, which consists of two stages, i.e., the task assignment stage and the truth discovery stage. The former stage determines how to assign tasks to domain experts wisely and efficiently. The latter stage adopts an iterative method to calculate worker credibility and the truth from each other.

To summarize, this work makes the following contributions:

1. To the best of our knowledge, we are the first to propose a crowdsourcing system that comprehensively considers the worker domain expertise and copier detection;
2. We used a greedy strategy to select experts in task assignment and updated worker domain expertise vectors in truth discovery for more precise quantification. Copier removal was then conducted to facilitate task assignment;
3. We conducted extensive experiments to demonstrate the effectiveness of our approach via comparison with baseline methods on two real-world datasets and one synthetic dataset.

The remainder of this paper is organized as follows. Section 2 discusses the related work. Section 3 defines the problem and gives an overview of the entire process. In Section 4, we introduce our task assignment algorithm, and in Section 5, we describe our truth discovery algorithm. Section 6 introduces our method to update the domain expertise of workers and the algorithm to initialize new workers. Section 7 shows the performance results and analysis of different models. Finally, we conclude our paper and point out the future work in Section 8.

2. Related Work

2.1. Task Assignment in Crowdsourcing

In early crowdsourcing systems, such as CDAS [6], the candidate tasks are randomly assigned to workers. AskIt! [17] is yet another crowdsourcing platform, which assigns the tasks that have the highest uncertainty, again disregarding the quality (or expertise) of the incoming worker. Recently, such as in OptKG [18] and CrowdDQS [19], task assignment is modeled by a Markov decision process or solved by using maximum potential gain. Some works about task assignment in crowdsourcing focus on different perspectives. Xi et al. [18] addressed the budget allocation problem using an extended Markov decision framework. Parameswaran et al. [20] proposed optimal and heuristic algorithms to efficiently find assignment strategies. Gao et al. [21] introduced a cost-sensitive method to determine whether questions can be better solved by crowdsourcing or machine-based methods. Mo et al. [22] explored how to optimize the plurality of an HIT. Sheng et al. [23] studied the extent to which repeated labeling helps to achieve a better result. Mo et al. [24] explored how to assign heterogeneous tasks (tasks of multiple types) to workers. CrowdSelect [25] increases the accuracy of crowdsourcing tasks through behavior prediction and user selection. There are also some works that have studied worker models and discussed how to infer the parameters [26,27].

Some works in crowdsourcing have studied subjective perceptions concerning worker engagement [28], moods [29,30], and satisfaction [31]. Reference [28] quantified worker engagement and worker retention and showed that conversational interfaces could significantly better retain crowd workers. References [29,30] also showed that worker moods could affect quality-related crowdsourced outcomes. Some researchers attempted to apply conversational interfaces in crowdsourcing. Reference [32] designed an HTML-based conversational interface for microtasking, which can be directly embedded on crowdsourcing platforms, saving the inconvenience of redirecting to other messaging applications. Some studies [33] focused on the pricing of tasks in crowdsourcing to make better rules. In recent years, mobile crowdsourcing has emerged as a method to harness human power to perform spatial tasks. Reference [34] investigated the quality-aware online task assignment (QAOTA) problem in mobile crowdsourcing.

Some task assignment methods take the domain expertise of workers into consideration. These methods mainly focus on three aspects: expert search, expertise modeling, and expertise representation. The purpose of expert search is to find experts who have knowledge of a specific domain and can solve a specific task [35–37]. Fang et al. [35] proposed a discriminative learning framework to model the correlation conditional probability between work and a task. In many application scenarios, various data in modeling expertise are required. Guan et al. [38] mined the fine-grained knowledge of web users, by analyzing their web surfing data, to facilitate expertise modeling. For expertise representation, some early methods built a knowledge base that contains the descriptions of workers' skills [39] or uses labels to represent the expertise of every worker. More advanced methods are based on topic modeling [40,41]. Latent Dirichlet allocation (LDA) [42] learns the document–topic and topic–word distributions by analyzing documents. LDA and TwitterLDA [43] exploit diverse domains in each task using topic models. However, they require a user to input the number of latent domains and cannot capture the related domain(s) of each task explicitly and correctly, without considering the semantics in texts. DOCS [44] judges the domain of task according to the knowledge base. Our method is to select high-quality

workers based on task domain classification. Gagan Goel et al. [45] allocated tasks to workers with matching constraints. ARE [14] selects one expert to complete each task according to the completion of historical tasks, which is contrary to the original intention of the crowdsourcing platform. If no worker meets the requirements of the expert, ARE cannot give a solution. Our method selects multiple high-quality workers for each task and updates their domain expertise vector iteratively in truth discovery, achieving more precise crowdsourcing results. MDC [15] considered domain expertise in the subsequent truth discovery step instead of the task assignment step. Therefore, it may assign tasks to low-quality workers and consume unnecessary costs. Our paper selected domain experts in the task assignment stage to save costs.

Unfortunately, we have not found an algorithm that considers the dependence between copiers in the task assignment step of crowdsourcing. In the truth discovery step, many methods consider copying. However, the truth discovery step is after the task assignment step, and many tasks are still assigned to the copier (which consumes unnecessary budget). We considered copying in the task assignment step, which can directly avoid assigning tasks to copiers.

2.2. Truth Discovery in Crowdsourcing

The most basic truth discovery method is majority voting. This approach regards all workers as equally trustworthy. However, in reality, different worker may have different accuracies [46]. To differentiate worker credibility, early methods, such as D&S [7], use a confusion matrix to model the credibility of workers. After that, more advanced methods, such as TruthFinder [47], LTM [48], and PrecRec [49] have been proposed. They measure worker credibility by additionally applying worker answering models or incorporating more considerations, such as task difficulty. However, these methods do not take worker domain expertise into consideration, assuming that workers have the same credibility for all tasks. Based on this assumption, they measure the accuracy of each work as a unified value or matrix.

To relax this assumption, there are some current efforts in the truth discovery step in crowdsourcing [50]. They try to utilize the fine-grained credibility of sources. FaitCrowd [51] uses a probabilistic graphical method to divide tasks into topic-level clusters and estimate every source's topical credibility accordingly. However, this method needs to determine the number of topics in advance, and the semantics of topic clusters is lacking. IniCrowd [52] is a similar method. It uses similarity metrics and topic models to obtain the similarity and topic distribution of each task. Tasks with high text similarity are assigned to the same domain. Nevertheless, this method may simply divide tasks with a similar text description syntax together. In fact, these tasks are in different domains. Lin et al. [53] proposed a method that considers domains and multi-truth. They inferred the domain expertise of a data source based on its data richness in different domains. However, this method is inapplicable to the crowdsourcing scenario, where we cannot expect all workers to provide answers to a large number of tasks.

In crowdsourcing, a large number of truth discovery algorithms for many aspects have been proposed. Miao et al. [54] propose a privacy-preserving truth discovery framework called PPTD. This framework uses the threshold homomorphic cryptosystem to guarantee the confidentiality of workers' values and weights. Tang et al. [55] proposed a non-interactive privacy-preserving truth discovery framework, which protects workers' data while obtaining truth. Wu et al. [56] proposed an unsupervised learning method to quantify the workers' credibility and long-term reputations. This method uses an outlier detection technique to filter out anomalous data items. Xiao et al. [57] proposed a protocol called BUR, which can employ nearly the least number of workers while ensuring that the overall accuracy of each task meets the requirements (given threshold). Jin et al. [58] proposed a framework for multi-requester mobile crowdsourcing, called CENTURION, which consists of a truth discovery step and an incentive step. However, none of these studies have considered the factors of the domain and copying.

In addition, copying is ubiquitous in crowdsourcing. The credibility of copiers is not in line with their real level. Some truth discovery works have taken source correlation into consideration. Dong et al. [16] proposed a method to consider the relationship between sources in truth discovery. Source correlations were inferred based on the intuition that "if two sources provide the same false values, it is very likely that one copies from the other". However, this model does not precisely demonstrate how a potential correlation can impact the estimation of sources' trustworthiness. MBM [59] takes into account both the copier and multi-truth problems. IMC^2 [60] takes into comprehensive consideration the copying and incentive mechanism in crowdsourcing, but it still calculates the probability of copiers among all workers, which is very heavy in the actual crowdsourcing application. How to deal with copiers in crowdsourcing scenario is different from that in truth discovery. In a crowdsourcing system, we need to avoid assigning tasks to copiers, while when conducting truth discovery, we need to penalize the copiers by assigning them a lower reliability.

3. Problem Definition

In this section, we formally define the problem of selecting workers wisely for crowdsourcing and provide the details of our crowdsourcing model. The system selecting workers wisely for crowdsourcing generally involves four components in its life cycle:

Inputs include:

- `A set of domains, D.` $D = \{1, 2, \ldots |D|\}$. This contains all the possible domains involved in the tasks in the system. All domain IDs are named from 1 to $|D|$;
- `A set of tasks, T.` Each $t_j \in T$ is a numerical selection task. $|l_j|$ indicates the number of options for the task t_j. For example, if the task is to give Obama's age, then the $|l|$ is 130. In addition, we use D_{t_j} to represent the domain set involved in task t_j;
- `A set of workers, W.` Each $w_i \in W$ applies to complete tasks in the crowdsourcing system;
- `A set of labels, L.` Each label $L_i \in L$ is the information voluntarily provided by the worker at the time of registration, indicating some characteristics of the worker, such as age, occupation, etc. These labels are used to better initialize the domain expertise of workers in the initialization algorithm;
- `M.` The upper limit of the number of tasks each worker can complete. We set M as a constant;
- `K.` The number of workers required for each task. We set K to another constant.

Intermediate variables are generated and updated during the SWWC procedure:

- `Worker domain expertise vector.` Each w_i is modeled as a vector $\left[v_i^1, v_i^2, \ldots v_i^{|D|}\right]$, where each $v_i^k \in [0, 1]$ indicates the expertise of worker w_i in answering tasks in domain $k, 1 \leq k \leq |D|$. A higher value v_i^k means that worker w_i is better at domain k. The system updates it after w_i completes tasks;
- `Fine-grained worker credibility,` q_i^j. This reflects the capability of worker w_i providing true value to task t_j. It is calculated based on the worker domain expertise vector;
- `Selected workers set,` $\hat{W}_{t_j}$. This collects the workers selected to complete task t_j. $|\hat{W}_{t_j}| \leq K$ (Section 4.2);
- m_{w_i}. This depicts the number of tasks that has already been assigned to worker w_i. m_{w_i} should never exceed M;
- `A set of answers provided by workers` after task assignment, A. Each $a_i^j \in A$ depicts the answer provided by worker w_i on task t_j;
- `Historical task completion records, H.` This collection contains all tasks that have been previously completed by workers. Each $h \in H$ is modeled as a vector $[w_i, t_j, a_i^j, \bar{a}_j]$, where $\bar{a}_j$ is the estimated truth of task t_j. We denote by ϕ the observation of H.

Output is:

- `Estimated truth,` $\bar{a}_j$. This is obtained by integrating the answers of all workers on task t_j.

Ground truth is the factual truth for each task $t_j \in T$, denoted as $\hat{a}_j$, which is used to measure the effectiveness of methods.

Different from most existing methods, we took the dependence of workers into consideration in order to reduce the impact of copiers. Based on the given historical task completion records, we calculated the probability of copying between each pair of workers. The basis for detecting plagiarism among workers is that if two workers w_i and $w_{i'}$ always share the same wrong answers, there may be a dependence between them. We use $w_i \perp w_{i'}$ to represent that workers w_i and $w_{i'}$ are independent and use $w_i \sim w_{i'}$ to describe that there is a dependence relation between w_i and $w_{i'}$.

To make the computation tractable, we assumed that the dependence of workers satisfies the following assumptions:

- `Independent copying:` The dependence of any pair of workers is independent of the dependence of any other pair of workers;
- `No loop dependence:` The dependence relationship between workers is non-transitive;
- `Uniform false value distribution:` For each task, there are multiple false values, and an independent worker has the same probability of providing each of them.

We formally define the problem of selecting workers wisely for crowdsourcing as follows:

The problem of selecting workers wisely for crowdsourcing: Given a set of tasks (T) and a set of workers (W), assign tasks to appropriate independent workers by considering the domain. Then, calculate the estimated truth ($\bar{a}_j$) from the answers for each task t_j, satisfying that $\bar{a}_j$ is as close to the ground truth $\hat{a}_j$ as possible.

Figure 1 shows an overview of the process of SWWC. SWWC contains two stages, i.e., the task assignment stage and the truth discovery stage. Requesters submit the task and task-related domain information to the crowdsourcing platform. When there are some workers applying to complete tasks, the crowdsourcing platform starts the task assignment step. For each new task, it first conducts fine-grained worker credibility calculation. Then, worker selection, as well as copier detection and removal are iteratively operated to wisely assign tasks to independent domain experts. After workers complete the task, they submit their answers to the system. Then, the truth discovery stage launches. In this stage, the truth computation and fine-grained worker credibility estimation are conducted in an iterative manner. Finally, the crowdsourcing platform returns the estimated truth to the requesters as the result of the tasks. At the same time, the crowdsourcing platform updates the work domain expertise vectors according to the estimated truth, which helps to better measure the real level of workers.

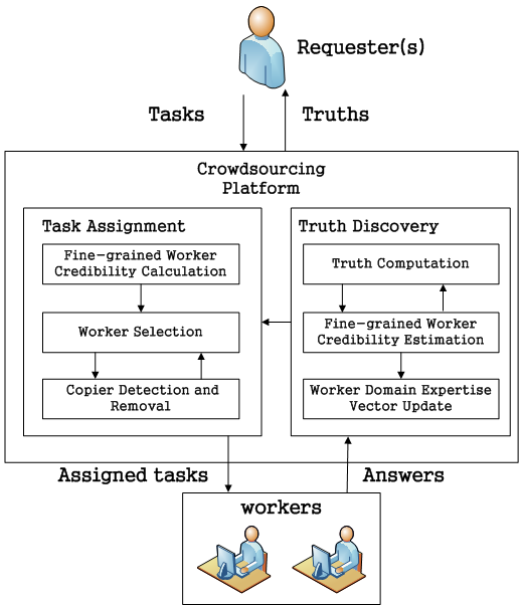

Figure 1. The process of SWWC.

4. Task Assignment

This section presents our approach for task assignment. All tasks run in the order of first come, first execution. Each task follows these three steps: fine-grained worker credibility calculation (Section 4.1), worker selection (Section 4.2), and copier detection and removal (Section 4.3). After Step 3 is executed, if there are no copiers, tasks are assigned according to the assignment strategy obtained in Step 2. If there are copiers, remove the copiers and then perform Step 2 until there are no copiers in the assignment strategy. The input, output, and execution flow of the task assignment algorithm are shown in Algorithm 1.

Algorithm 1 Task assignment.

Input: The set of tasks T, the set of workers W, historical task completion of all workers H, M, K

Output: Selected worker set for all tasks $\left\{\hat{W}_{t_j}\right\}$

Initialize: $\{m_{w_i}\} = 0$, iscopier = true
for $j = 1$ to $|\mathrm{T}|$ **do**
    //calculation of worker fine-grained credibility
    for $i = 1$ to $|\mathrm{W}|$ **do** compute q_i^j by (1)
    end for
    sort workers in descending order of credibility
    while iscopier **do**
        //worker selection
        workerselection (t_j, $\bar{W}_{t_j}$, K, $\{m_{w_i}\}$)
        //copier detection and removal
        copierdisposal (iscopier, $\left\{\hat{W}_{t_j}\right\}$, $\{q\}$)
    end while
end for
return $\left\{\hat{W}_{t_j}\right\}$

4.1. Fine-Grained Worker Credibility Calculation

In reality, tasks may contain multiple domains, and workers may be good at different domains. In order to assign tasks to appropriate workers, we need to calculate the credibility of workers for each task. We define it as the worker fine-grained credibility, e.g., q_i^j. For single-domain tasks, we can regard worker's domain quality in that domain as the fine-grained credibility for these tasks. However, it is unreasonable to consider only one domain of the task for multi-domain tasks. We should consider all the domains covered by the task. For multi-domain tasks, we take the average value of the worker's domain quality in all the domains involved in the task as the fine-grained credibility of the worker. Here, fine-grained credibility is the worker's credibility for this specific task. The calculation of worker fine-grained credibility is shown in Formula (1).

$$q_i^j = \frac{1}{|D_{t_j}|} \sum_{d \in D_{t_j}} v_i^d \tag{1}$$

After obtaining the fine-grained credibility of all workers. For each task, we rank the workers in descending order of fine-grained credibility to obtain a candidate worker set $\bar{W}_{t_j}$. The total time complexity of Step 1 is $O(|T|(|W||D| + |W|^2))$.

4.2. Worker Selection

In this section, we formulate the worker selection problem as an optimization problem [45,61,62] and propose a greedy strategy to solve it more effectively. The basic idea is to find an assignment scheme to make all tasks be globally completed best. This can be achieved by selecting the first K workers from $\bar{W}_{t_j}$. However, note that each worker has an

upper limit (M) on the number of tasks that can be completed. For example, we assign M tasks to worker w_1. In the following tasks, w_1 is still the best worker. Which K tasks will the worker w_1 be assigned? We need to take this situation into account. We use set $\{m_{w_i}\}$ to record the number of tasks that worker w_i has been assigned.

For the set of tasks, Step 2 finds an optimal assignment scheme S to maximize the summation of the overall worker fine-grained credibility of all tasks in the scheme, i.e.,

$$S = argmax \sum_{t_j \in T} \left(\sum_{w_i \in \hat{W_{t_j}}} q_i^j \right) \tag{2}$$

subject to:

$$max(m_{w_i}) \leqslant M \tag{3}$$

where $\sum_{w_i \in \hat{W_{t_j}}} q_i^j$ is the overall credibility of workers that globally complete t_j. Summing such overall credibility of each task, $\sum_{t_j \in T} \left(\sum_{w_i \in \hat{W_{t_j}}} q_i^j \right)$ can measure the credibility of all globally completed tasks.

Next, we prove that the optimal problem of worker selection is NP-hard.

Proof. Given a set of workers W, each of which corresponds to their credibility, then the system has a set of tasks T, each of which needs K workers to complete. In order to meet scheme S, we may need to detect $(C_{|W|}^K)^{|T|}$ possible cases to find the optimal solution. There is no doubt that this has an exponential time complexity, which cannot be obtained in polynomial time. Therefore, the optimal problem of worker selection is a typical NP-hard problem. □

A Greedy approximation algorithm:

As the optimal problem of worker selection is NP-hard, we developed a greedy-based approximation algorithm to efficiently solve it. The algorithm takes task t_j, candidate worker set $\bar{W_{t_j}}$, number of workers needed K, and number of tasks completed by workers $\{m_{w_i}\}$ as the input. The algorithm obtains an allocation set $\hat{W_{t_j}}$ for each task t_j. For each task t_j, it selects the worker with the highest credibility from the candidate worker set $\bar{W_{t_j}}$ to assign the task. When t_j selects worker w_i, if m_{w_i} has reached the upper limit, it skips the worker and removes w_i from $\bar{W_{t_j}}$. Otherwise, t_j increases m_{w_i} by one, inserts w_i into $\hat{W_{t_j}}$, and removes w_i from $\bar{W_{t_j}}$. Perform this operation K times until the number of workers required for the task is reached, and output $\hat{W_{t_j}}$. The algorithm pseudo code of the worker selection part is shown in Algorithm 2. The time complexity for all tasks to complete worker selection is O(K | T |).

Algorithm 2 Worker selection.

```
function WORKERSELECTION(t_j, W̄_{t_j}, K, {m_{w_i}})
    for r = 1 to K do
        choose the first w_i ∈ W̄_{t_j}
        if m_{w_i} < M then
            insert w_i into Ŵ_{t_j}
            remove w_i from W̄_{t_j}
            m_{w_i} + +
        else
            remove w_i from W̄_{t_j}
        end if
    end for
end function
```

4.3. Copier Detection and Removal

Existing task assignment methods mostly assume that all workers complete their tasks independently. However, in reality, copying is ubiquitous. We classified workers into two groups, i.e., independent workers and copiers. Given a task, it may have multiple distinct false answers, but only one single true answer. It is common to observe two workers sharing the same true answer, but it is a rare event to see two workers providing the same false answer. Therefore, if two workers share many false answers, they are highly likely to be dependent, i.e., $w_i \sim w_{i'}$.

We applied the Bayesian method to compute the dependent probability of two workers, w_i and $w_{i'}$. The purpose of the algorithm is to avoid assigning tasks to copiers in the task assignment step. Therefore, we not only need to detect whether there is a copying relationship between a pair of workers, but we also need to identify the direction of dependence. Intuitively, copiers normally tend to copy information from highly reliable workers. Therefore, given $w_i \sim w_{i'}$, we regard the worker with lower reliability as the copier. We denote w_i a copy from $w_{i'}$ as $w_i \to w_{i'}$. In this paper, we focused on the direct copying relationship between any pair of workers, and leave the complex global copying detection as future work.

In order to calculate the dependence probability, we first partitioned the tasks completed by w_i and $w_{i'}$ in H into three sets. T^t is the set of tasks on which w_i and $w_{i'}$ provide the same true values; T^f is the set of tasks on which they provide the same false values; T^d is the set of tasks on which they provide different values.

We first considered the situation where the two workers w_i and $w_{i'}$ are independent. Since there is only one true value, the probability that w_i and $w_{i'}$ provide the same true value for task t_j, denoted by P_t^j for convenience, is:

$$P_t^j = P\left(t_j \in T^t | w_i \perp w_{i'}\right) = q_i^j \cdot q_{i'}^j \tag{4}$$

Based on the assumption of a uniform false value distribution in Section 2, we think any independent worker has the same probability of providing each false value for task t_j. Thus, the probability that worker w_i provides a false value for task t_j is $\frac{1-q_i^j}{|l_j|-1}$. The probability that w_i and $w_{i'}$ provide the same false value for task t_j, denoted by P_f^j, is:

$$P_f^j = P(t_j \in T^f | w_i \perp w_{i'}) = \frac{(1-q_i^j)\cdot(1-q_{i'}^j)}{|a|^j} \tag{5}$$

Then, the probability that w_i and $w_{i'}$ provide different values on task t_j, denoted by P_d^j, is:

$$P_d^j = P(t_j \in T^d | w_i \perp w_{i'}) = 1 - P_t^j - P_f^j \tag{6}$$

Thus, the conditional probability of observing ϕ is:

$$P(\phi | w_i \perp w_{i'}) = \prod_{t_j \in T^t} P_t^j \cdot \prod_{t_j \in T^f} P_f^j \cdot \prod_{t_j \in T^d} P_d^j \tag{7}$$

We next considered the situation where $w_i \to w_{i'}$ (similar for $w_{i'} \to w_i$). There are two cases where w_i and $w_{i'}$ provide the same value for the task t_j. First, with probability r, one copies the value from the other, and so, the value is true with probability $q_{i'}^j$ and false with probability $1 - q_{i'}^j$. Second, with probability $1 - r$, the two workers provide the value independently, and so, the probability of being true or false is the same as that in the situation where w_i and $w_{i'}$ are independent. Thus, we have:

$$P(t_j \in T^t | w_i \to w_{i'}) = q_{i'}^j \cdot r + P_t^j \cdot (1-r) \tag{8}$$

$$P(t_j \in T^f | w_i \to w_{i'}) = (1 - q_{i'}^j) \cdot r + P_f^j \cdot (1 - r) \tag{9}$$

Finally, the probability that w_i and $w_{i'}$ provide different values on task t_j is the probability that w_i provides a value independently, and the value differs from that provided by $w_{i'}$:

$$P(t_j \in T^d | w_i \to w_{i'}) = P_d^j \cdot (1 - r) \tag{10}$$

Thus, the conditional probability of ϕ is:

$$P(\phi | w_i \to w_{i'}) = \prod_{t_j \in T^t} \left[q_{i'}^j \cdot r + P_t^j \cdot (1 - r) \right]$$

$$\cdot \prod_{t_j \in T^f} \left[(1 - q_{i'}^j) \cdot r + P_f^j \cdot (1 - r) \right] \cdot \prod_{t_j \in T^d} \left[P_d^j \cdot (1 - r) \right] \tag{11}$$

We compute $P(w_i \to w_{i'} | \phi)$ accordingly:

$$P(w_i \to w_{i'} | \phi) = [1 + \frac{1 - \alpha}{\alpha} \cdot \prod_{t_j \in T^t} \frac{P_t^j}{q_{i'}^j \cdot r + P_t^j \cdot (1 - r)}$$

$$\cdot \prod_{t_j \in T^f} \frac{P_f^j}{(1 - q_{i'}^j) \cdot r + P_f^j \cdot (1 - r)} \cdot (\frac{1}{1 - r})^{|T^d|}]^{-1} \tag{12}$$

Here, $\alpha = P(w_i \sim w_{i'})$ $(0 < \alpha < 1)$ is the a priori probability that two workers are dependent.

By applying the Bayesian method, we can compute the dependent probability for every pair of workers w_i and $w_{i'}$. If the probability $P(w_i \to w_{i'})$ is larger than threshold β, then we say there is a dependency between w_i and $w_{i'}$. If both $P(w_i \to w_{i'})$ and $P(w_i \to w_{i'})$ are larger than threshold β, then we should identify the direction of copying. We regard the worker with lower accuracy as the copier between them. To punish the copiers, we not only avoid assigning the current task to them, but also halve their credibility to facilitate copying detection for the upcoming tasks.

The algorithm pseudo code of the copier detection and removal part is shown in Algorithm 3. The time complexity for all tasks to complete copier detection and removal is $O(C_K^2|H|)$. If Step 2 and Step 3 cycle v times, the time complexity of task assignment is $O(|T|(|W||D| + |W|^2 + v(K|T| + C_K^2|H|)))$. In practice, $v \leq 5$.

Algorithm 3 Copier disposal.

```
function COPIERDISPOSAL(iscopier, {Ŵ_{t_j}}, {q})
    iscopier = false
    copiernum = 0
    for w_i ∈ Ŵ_{t_j}, w_{i'} ∈ Ŵ_{t_j}; w_i ≠ w_{i'} do
        compute P(w_i ~ w_{i'} | φ) by (4–12)
        if P(w_i ~ w_{i'} | φ) > β then
            if q_i^j < q_{i'}^j then copier = w_i
            elsecopier = w_{i'}
            end if
            iscopier = true, copiernum++
            remove copier from Ŵ_{t_j}
            halve the credibility of the copier
        end if
        K = copiernum
    end for
    return iscopier, K
end function
```

5. Truth Discovery

This section presents our truth discovery method. We observe that there are two relations between workers' credibility and tasks' truth: (i) given a task t_j, if the worker's credibility for t_j is high, then his/her answer is likely to be the truth for t_j; (ii) given a worker w_i, if w_i often answers tasks correctly, then w_i has a high credibility. Therefore, truth and workers' credibility depend on each other. Based on these intuitions, we developed our approach, in which the truth computation step and the worker credibility estimation step are iteratively conducted until convergence. Next, we introduce these two steps in detail and analyze the time complexity of the truth discovery step.

Truth computation step: In this step, workers' credibility is assumed to be fixed. Then, the estimated truth $\bar{a_j}$ can be inferred through weighted aggregation. In general, for each task, we obtain the estimated truth by the following formula.

$$\bar{a_j} = \frac{\sum_{w_i \in \hat{W_{t_j}}} a_i^j \cdot q_i^j}{\sum_{w_i \in \hat{W_{t_j}}} q_i^j} \tag{13}$$

This follows the principle that answers from reliable workers are considered more in the aggregation.

Worker credibility estimation step: In this step, workers' credibility is identified based on the current estimated truth. As q_i^j denotes worker w_i's credibility for task t_j, we decide whether to improve or reduce the credibility of w_i according to the difference e_i^j between a_i^j and $\bar{a_j}$. We used the following principles to update the worker's credibility: (i) If e_i^j is less than the threshold θ, we think worker's answer is good and improve his/her credibility. (ii) If e_i^j is larger than θ, we believe that worker's answer is poor and his/her accuracy is reduced. We use the indicator function to represent this:

$$F = \begin{cases} 1, \ e_i^j \leqslant \theta \\ 0, \ e_i^j > \theta \end{cases} \tag{14}$$

Then, we calculate the accuracy of the worker q_i^j:

$$q_i^j = (q_i^j + 0.5 \cdot (1 - q_i^j))^F \cdot (q_i^j - 0.5 \cdot \frac{e_i^j}{|l|} \cdot q_i^j)^{|F-1|} \tag{15}$$

The above formula shows how to improve the credibility of reliable workers and reduce the credibility of unreliable workers.

After we obtain the new credibility of workers, we can calculate the estimated truth again. We iteratively calculate these two values until they converge. The estimated truth obtained after convergence is the final estimated truth we need, that is the last round of $\bar{a_j}$. In the truth discovery, there is also the step of updating worker domain expertise, which we analyze in Section 6.1. The pseudo code of the whole truth discovery algorithm is shown in Algorithm 4.

Time complexity: For each task, suppose it takes u iterations to converge, then the time complexity of obtaining the final $\bar{T_j}$ is $O(uk)$. The time complexity of updating worker domain expertise is $O(|W||D|)$. Thus, the total time complexity is $O(|T|(uk + |W||D|))$. In practice, $u \leq 20$, the time complexity is linear.

Algorithm 4 Truth discovery.

Input: The set of answers $\{< t, w, a >\}$
Output: truth $\bar{a}$, updated worker domain quality

```
for j = 1 to |T| do
    repeat
        compute ā_j by (13)
        for each w_i ∈ Ŵ_{t_j} do
            compute q_i^j by (14,15)
        end for
    until convergence
    for each w_i ∈ Ŵ_{t_j} do
        compute v_i^k by (16)
    end for
end for
return ā, updated worker domain quality
```

6. Worker Domain Expertise Renewal and Initialization

The domain expertise for workers involves two operations: one is updating the domain expertise of workers after workers complete their tasks; the other is initializing the domain expertise for new workers. Next, we introduce them in Sections 6.1 and 6.2, respectively.

6.1. Worker Domain Expertise Renewal

After obtaining the truth for each task, we need to update the domain expertise vector of the participating workers who complete the task according to $\bar{a}_j$. According to this, the domain expertise vector can better reflect the real level of workers. The method of updating the domain expertise vector is similar to the credibility update method, the formula is as follows:

$$v_i^k = (v_i^k + 0.5 \cdot \frac{1}{\left|D_{t_j}\right|} \cdot (1 - v_i^k))^F \cdot (v_i^k - 0.5 \cdot \frac{1}{\left|D_{t_j}\right|} \cdot \frac{e_i^k}{|l|} \cdot v_i^k)^{|F-1|} \tag{16}$$

6.2. Worker Domain Expertise Initialization

For a new worker, we need to set her/his initial domain expertise vector. This is a cold-start problem. Most of today's crowdsourcing and truth discovery methods set it to a fixed value of 0.5. The accuracy of workers is updated through multiple rounds of iterations to make it close to the real level of workers.

We initialize workers' domain expertise vector through the information provided by workers during registration. Workers need to fill in some personal information when registering, including gender, age, occupation, etc. Different workers have different accuracy in different domains, which is related to workers' age, occupation, and so on. For example, programmers tend to know more about computer problems, and the older they are, the deeper their knowledge may be. Therefore, we explored how to use the worker registration information to better optimize the accuracy initialization of the worker domain.

We use each registration information option as a label (such as men, college students, teachers). For each label, we calculate a domain expertise vector according to the accuracy of existing workers in the crowdsourcing system, which represents the accuracy of the person containing the label in each domain. Each element L_a^k in the label vector L_a is calculated as follows:

$$L_a^k = \frac{1}{|W|} \cdot \sum_{w_i \ contains \ L_a} v_i^k$$

where $|W|$ refers to the number of workers with the L_a label.

When a new worker arrives, we initialize his/her domain expertise vector according to the label vector and the new worker registration information. The initialization calculation formula is as follows:

$$v_i^k = \frac{1}{|L|} \cdot \sum_{L_a \ belongs \ to \ w_i} L_a^k$$

where $|L|$ refers to the number of labels belonging to worker w_i.

At this step, we complete the domain expertise vector initialization of workers by using the registration information.

7. Experiments

We present the experiment settings in Section 7.1. In Section 7.2, we compare our method with baseline methods on two real-world datasets in terms of effectiveness and efficiency. In Section 7.3, we compare our method on one synthetic dataset with copiers. In Section 7.4, we validate the expertise of different workers in different domains and analyze the experimental results of our initialization algorithm.

7.1. Experimental Settings

We conducted experiments on two real-world datasets and one synthetic dataset:

- `MovieLens` [63]: This dataset contains 1,000,209 anonymous ratings of approximately 3900 movies made by 6040 MovieLens users. All ratings are in the following format: UserID-MovieID-Rating. UserIDs range between 1 and 6040. MovieIDs range between 1 and 3952. Ratings are made on a 5-star scale. Another file contains the domain to which each movie belongs. Each user has at least 20 ratings;
- `Anime` (https://www.kaggle.com/CooperUnion/anime-recommendations-database (accessed on 19 December 2021)): This dataset contains information on user preference data from 73,516 users on 12,294 anime works. Each user is able to add anime works to his/her completed list and give it a rating. This dataset is a compilation of those ratings. All data are displayed in the following format: userid-animid-rating (ratings range from 0 to 10). The domain information of anime works is in the description document of the anime works;
- `Synthetic dataset`: This dataset was synthesized on the basis of MovieLens dataset by manually adding copiers. We added different proportions of copiers, in which each copier randomly copies a worker in the MovieLens dataset. Due to randomness, we generated 100 synthetic datasets for each proportion, and the experimental results of each method were averaged. We discuss the performance of various algorithms via tuning the proportion of copiers.

Now, there are many crowdsourcing methods to solve the crowdsourcing problem from different angles. We compared our method with the following seven baseline methods:

- `RandomMV`: This method uses a random strategy for task assignment and aggregates workers' answers to generate truth by using majority voting;
- `D&S` [7]: This method also uses the random strategy for task assignment. For truth discovery, it uses the EM algorithm, which calculates worker accuracy and truth;
- `ASKIT!` [17]: This method uses an entropy-like method to define the uncertainty of each task and infers the truth by majority voting. The task with the highest uncertainty is the next one to be assigned to the worker;
- `CDAS` [6]: It provides an estimated accuracy for each result from workers based on the workers' historical performances. Each task we are already confident in is terminated and no longer assigned to workers. At each step, CDAS selects at random a non-terminated task to assign to the incoming worker;
- `ARE` [14]: This method selects one expert for each task based on the professional domain and proficiency of workers' knowledge. In this model, experts accept tasks equal to or lower than their proficiency;

- `MDC` [15]: This method considers the domain factors of tasks and workers to aggregate better results in the truth discovery stage. Calculate the truth, and update the worker's domain credibility by the proportion of the task in each domain, the worker's answer, and domain credibility;
- `SWWC-NoCopier`: This is a variant of our method SWWC, but it assumes all workers are independent.

We adopted accuracy and efficiency to evaluate the performance of the methods. The goal of crowdsourcing is to obtain the best answer for each task. Therefore, the lower the error obtained by the method and the closer the truth estimated by the method is to the ground truth, the better. Thus, we used the MAE and RMSE to measure accuracy. In addition, we used program running time to measure efficiency. Because of the randomness, we ran each method 100 times to evaluate their average performance:

- `MAE`: This quantifies the average error between the estimated truth and the ground truth. The lower the MAE, the better the estimated truth. The formula is as follows:

$$\text{MAE} = \frac{1}{|T|}\left(\sum_{j=1}^{|T|} \bar{a}_j - \hat{a}_j\right) \tag{17}$$

- `RMSE`: This can well measure the deviation between the estimated value and the ground truth. The lower the RMSE, the better the estimated truth. The formula is as follows:

$$\text{RMSE} = \sqrt{\frac{1}{|T|}\left(\sum_{j=1}^{|T|} \bar{a}_j - \hat{a}_j\right)^2} \tag{18}$$

K was set as 10 for all methods. Because of the randomness, we ran each method 100 times to evaluate their average performance. We implemented all compared methods in Python 3.7, on a Window's server with an 8-core Intel(R) Core(TM) i7-3770 CPU @ 3.40 GHz cores and 8 GB memory.

7.2. Comparative Study on Two Real-World Datasets

We compared SWWC with the baseline approaches mentioned in Section 7.1 with the same set of tasks for qualification. Table 1 and Figure 2 show the performance of these approaches on accuracy and efficiency. We can see that our method performed better than others. Since these datasets do not contain copiers, the MAE of our method SWWC-NoCopier was the lowest, followed by our another method, SWWC, which considers copying. Although there are no copiers in these datasets, SWWC may calculate some workers as copiers in the calculation process, which will cause some misjudgment. However, we can see from the results that there were few misjudgments and little impact on the results. Our method SWWC was still superior to all other baseline methods (except SWWC-NoCopier).

Table 1. Comparison of different methods on two real-world datasets (1).

	MovieLens		
	MAE	**RMSE**	**Time (s)**
SWWC	0.3677	0.7583	1855.29
SWWC-NoCopier	**0.3606**	**0.7222**	**14.07**
RandomMV	0.6914	1.0969	56.31
D&S	0.4364	0.7942	54.74
ASKIT!	0.7348	1.1110	1561.88
CDAS	0.4134	0.7801	59.66
ARE	0.8792	1.2247	69.53
MDC	0.3851	0.7243	69.60

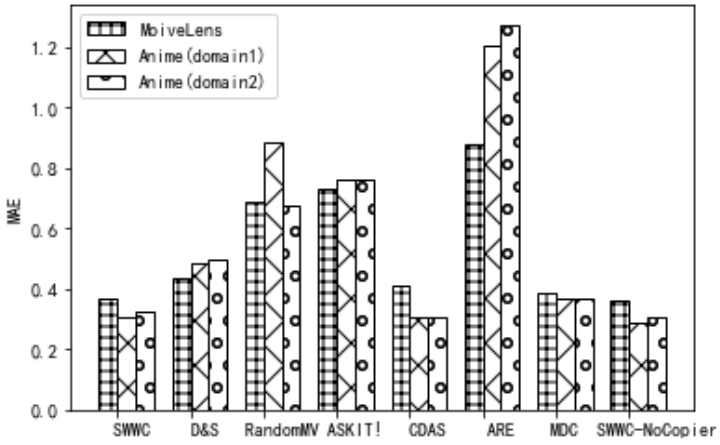

Figure 2. MAE of different methods on two real-world datasets.

RandomMV, ASKIT!, and ARE performed worst on both datasets. The performance of RandomMV largely depends on the quality of the selected worker set: if the set contains more good workers, the accuracy would be high; otherwise, the accuracy would be low. The main problem of method ASKIT! is the sequence of task assignment. This method does not focus on selecting high-quality workers for each task. Although the ARE method considers domain factors to conduct task assignment, it only selects one for each task. When no qualified expert was available in the system, ARE failed to assign the related tasks. Therefore, the experimental results of RandomMV, ASKIT!, and ARE were not ideal. CDAS and MDC were second only to our method. CDAS considers the confidence of the estimated values. This is an effective way to improve the accuracy of the results. MDC considers domain factors in the truth discovery process, which is helpful to improve the accuracy. However, they do not consider the domain in the task assignment stage, which resulted in a lower accuracy compared with our method. D&S had a certain effect on improving the accuracy of the results. The EM algorithm enhanced the accuracy in the truth discovery stage, so the deviation between the estimated value and the truth was slight, e.g., a small RMSE. However, the lack of consideration of the domain factor makes it perform poorly in the MAE.

In order to verify whether only the domain information type is beneficial to the results, we also experimented with the domain information of broadcast mode on the anime dataset. Anime (Domain 1) in Table 2 is the type of anime, and Anime (Domain 2) is the broadcast mode of anime. Anime (Domain 1) includes Action, Adventure, Comedy, Drama, etc. Anime (Domain 2) includes TV, Movie, OVA, ONA, and Special. From the experimental results, we can see that the MAE was the lowest whether considering the domain type or broadcast mode, which proves our idea: considering the domain factor in task assignment is beneficial to the truth. However, considering different domains, the improvement effect of the results was also different. In our experiment, type was better than broadcast.

We measured the efficiency of compared methods by running time. From Tables 1 and 2, we can see that our method SWWC took a relatively long time for conducting copying detection. We iteratively detected copiers while assigning tasks to proper workers, so as to improve the accuracy of crowdsourcing. The variant of our method, i.e., SWWC-NoCopier, that excludes the copier detection module consumed the least execution time. It demonstrated that the efficiency of our method is linear with the number of tasks.

Table 2. Comparison of different methods on two real-world datasets (2).

	Anime (Domain 1)			Anime (Domain 2)		
	MAE	**RMSE**	**Time (s)**	**MAE**	**RMSE**	**Time (s)**
SWWC	0.3043	1.1036	3614.597	0.3243	1.0234	3524.965
SWWC-NoCopier	**0.2875**	1.0368	**5.611**	**0.3075**	0.9521	**5.483**
RandomMV	0.8882	1.1077	48.123	0.6779	0.8865	52.328
D&S	0.4850	0.6799	46.375	0.4971	**0.6459**	50.762
ASKIT!	0.7605	0.9749	1507.95	0.7605	0.9749	1507.95
CDAS	0.3097	1.0935	39.32	0.3097	1.0935	39.32
ARE	1.2089	1.6919	519.29	1.2757	1.6080	502.35
MDC	0.3663	**0.4879**	489.131	0.3662	1.1027	482.58

7.3. Comparative Study on One Synthetic Dataset

To study the impact of copiers in crowdsourcing, we conducted a comparison study on a synthetic dataset with copiers. By tuning the proportion of copiers in the dataset from 10% to 30%, we obtained three datasets. We ran all eight methods on those datasets. All method results are shown in Table 3. From Figure 3, we can see that when the proportion of copiers increased, the MAEs of different methods increased to varying degrees, but our method (SWWC) could still maintain the lowest MAE. The errors of five methods, i.e., SWWC-NoCopier, D&S, ARE, MDC, and CDAS, increased linearly when the proportion of copiers increased, while the error of RandomMV increased exponentially. Among all the methods, ASKIT! showed the worst performance when copiers existed in the dataset. ASKIT! determines which task to assign through uncertainty. It does not involve selecting appropriate workers to complete it, so copiers have a great impact on it. Because SWWC identifies the copiers and removes the copiers from the task assignment, each task can be assigned to highly reliable workers. Thus, the accuracy of the answer is improved. For this reason, our method maintains the error at the lowest level. Experimental results showed that our method can effectively perform task assignment and truth discovery in crowdsourcing with copiers.

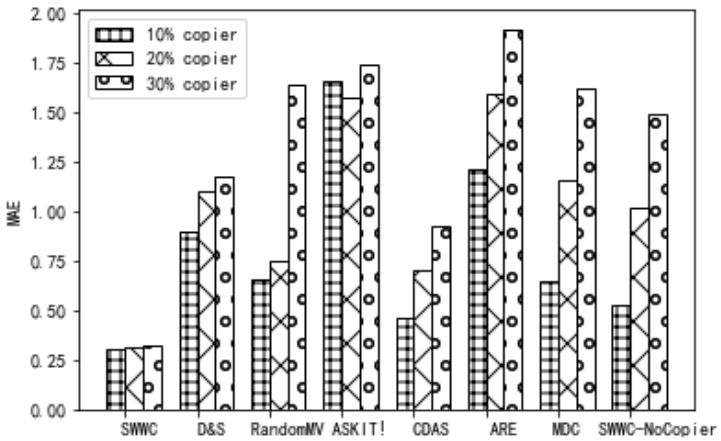

Figure 3. MAEs of different methods on one synthetic dataset.

Table 3. Comparison of different methods on one synthetic dataset.

	10% Copiers		20% Copiers		30% Copiers	
	MAE	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**
SWWC	**0.3677**	0.7583	**0.3685**	**0.7583**	**0.3702**	**0.7583**
SWWC-NoCopier	0.5278	**0.6272**	1.0189	1.0951	1.4917	1.5435
RandomMV	0.6611	1.0908	0.7469	1.1896	1.6419	2.1663
D&S	0.8970	1.0769	1.0974	1.2058	1.1716	1.2551
ASKIT	1.6580	2.2113	1.5712	2.1211	1.7439	2.2494
CDAS	0.4631	0.7967	0.7013	0.9290	0.9266	1.0898
ARE	1.2088	1.6919	1.5936	2.0176	1.9235	2.3544
MDC	0.65	0.7598	1.1594	1.2300	1.6219	1.6697

7.4. Validation of Worker Domain Expertise and Initialization Algorithm

7.4.1. Diverse Accuracies across Domains

To validate the claim that different workers have different domain expertise, we randomly chose five workers from the MoiveLens dataset as representatives and investigated their accuracy diversity on different domains. Figure 4 presents that all five workers showed different accuracies on different domains, with the X-axis representing the IDs of 18 domains and the Y-axis depicting the worker accuracy. With all five workers' accuracy fluctuating through the eighteen domains, we observed the following two diversities on worker domain expertise: (i) Each worker may have different expertise in different domains. For example, as shown in Figure 4, worker5 showed high accuracy in Domain 2 "Adventure" (0.949) and Domain 8 "War" (0.834), but low accuracy in Domain 11 "Horror" (0.263). (ii) Given a collection of workers, each domain may have different experts. The expert in Domain 9 "Drama" is worker1, with an accuracy of 0.963. However, she ranked lowest in Domain 10 "Music" with a low accuracy of 0.452, while the expert in Domain 10 "Music" was worker3. These observations confirmed that it is inaccurate to quantify worker reliability by using one single value. These results also indicate the necessity of considering the domain in task assignment, because workers may be good at one domain, but know nothing about another domain.

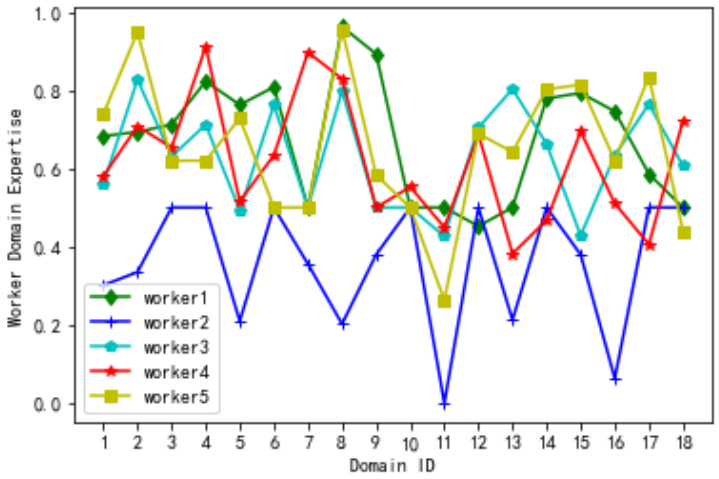

Figure 4. Domain expertise diversity of five workers.

7.4.2. Initialization Algorithm

Firstly, we discuss the accuracy of different labels in different domains through the MoiveLens dataset, as shown in Figure 5. The first subfigure in Figure 5 shows the accuracy in different domains of labels younger than 18 and between 25 and 34. The second figure shows the accuracy in different domains of labels for educators and the unemployed. From the figure, we can observe that the domain accuracy of the label between the ages of 25

and 34 was greater than that of the label less than 18, and the domain accuracy of the label of scholars was also greater than that of the unemployed. These are also in line with the actual situation. We can find that different labels had different effects on different domains, so it is reasonable to consider labels to initialize workers' domain accuracy.

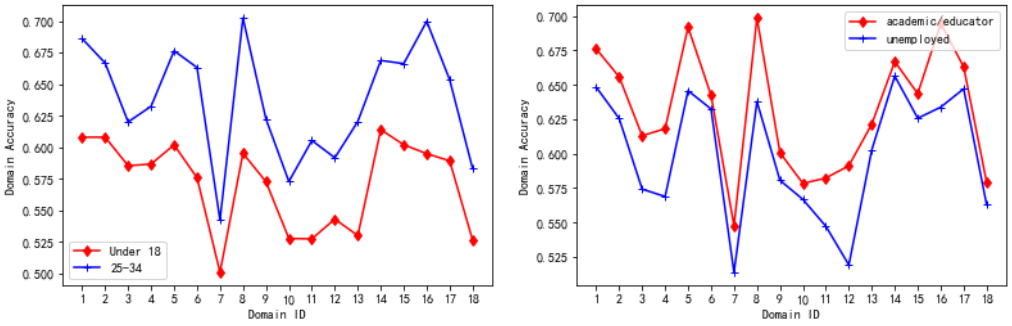

Figure 5. Domain accuracy of different labels.

Next, we compared the difference between our algorithm and the method of setting different fixed values. We used the initialization accuracy and the real accuracy of workers to calculate the MAE for comparison. Figures 6 and 7 show the MAE of our initialization method and different fixed values in various domains. We can clearly see that our method was superior to fixed values in most domains no matter how large the fixed value was. The average MAEs of fixed values of 0.4 to 0.7 in various domains were 0.2378, 0.1727, 0.1569, and 0.1691, respectively. These four values are greater than the average MAE of our initialization algorithm of 0.1481. Therefore, compared with the commonly used method of setting the initial value, our initialization algorithm can more effectively reflect the accuracy of an individual domain, especially when workers answer fewer questions.

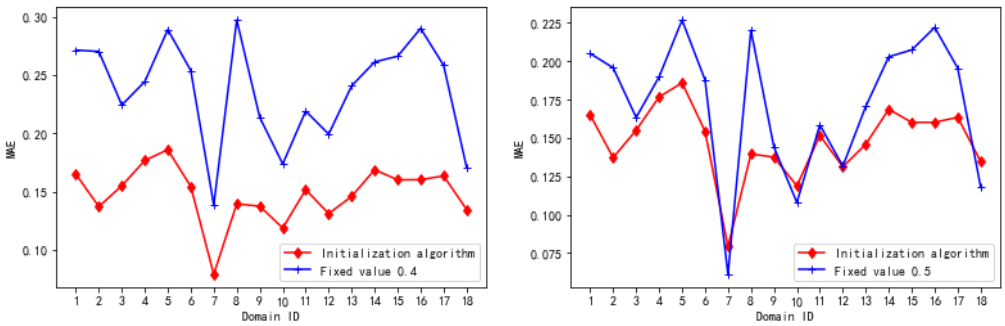

Figure 6. MAEs of different domains (a).

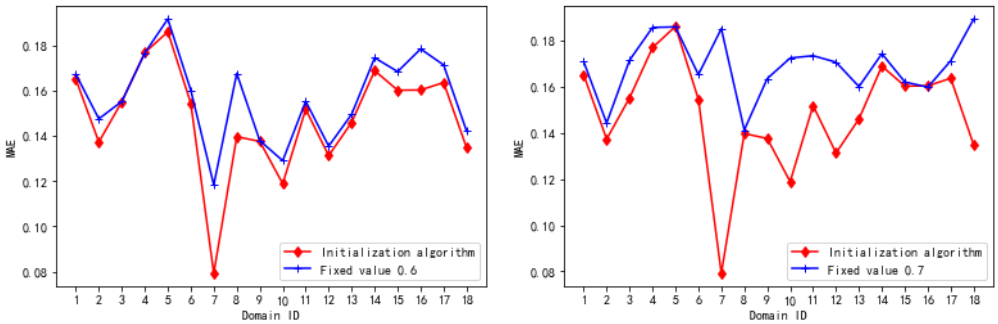

Figure 7. MAEs of different domains (b).

8. Conclusions and Future Work

In this paper, a crowdsourcing system, which consists of a task assignment stage and a truth discovery stage, was designed to comprehensively consider the domain expertise of

workers and copiers. We modeled the credibility of worker at the domain level. By taking the domain classification of tasks as a priori knowledge, we assigned tasks to the domain experts via a greedy algorithm. We also took the multi-domain tasks into consideration. We adopted a Bayesian model to detect the copiers and then removed them. Worker selection and copier detection were conducted in an iterative manner. Compared with previous methods, our method took less time to detect copiers, which is undoubtedly more in line with the application scenario of crowdsourcing. In the truth discovery stage, we used an iterative method to infer the truth and calculate fine-grained worker credibility. Worker domain expertise vectors were then updated based on the estimated truth. We also proposed a new initialization method to better initialize workers in crowdsourcing. The above details our method and its advantages. Our experiments on real-world datasets and synthetic dataset confirmed the superiority of our method, especially when there are copiers.

In the future, we plan to study the following four aspects: Instead of taking the domain classification of tasks as a priori knowledge, we will design an algorithm to automatically classify the tasks into domains based on the text description of the tasks. Given multiple attributes of a collection of tasks, there could be multiple ways of domain classification. We plan to design an algorithm to automatically choose the best way of classification to facilitate task assignment. Different tasks may have different difficulties, and more difficult tasks may require more workers to complete. We plan to conduct task assignment by additionally considering task difficulty. Although our initialization method can initialize workers more accurately than setting a fixed value, we can still use other machine learning methods to find a more appropriate initialization strategy.

Author Contributions: This research work was completed by S.S. as part of his thesis (supervised by X.F.) for the degree of Masters of Computer Science at Donghua University. X.F. and S.S. both contributed to the ideas and methodologies behind the work. The implementation of the algorithms was performed by S.S. and W.W. The data collection was performed by K.W. and H.L. The manuscript was written by X.F. and S.S. Revisions were implemented by G.S. and Q.Z.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Special Funds for Fundamental Scientific Research Operation Fees of Central Universities 21D111208. The authors would like to thank the anonymous Reviewers for their valuable feedback on this work.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets generated and/or analyzed during the current study are not publicly available due to the requirements of the project funding sources, but may be available from the corresponding author upon reasonable request and subject to the approval of the funding sources.

Acknowledgments: We are grateful for the server provided by Donghua University to assist us in running the experiment and obtaining the experimental data and results in a short time. In addition, we would like to thank the team that published the MovieLens dataset online and the kaggle website, which provided the anime recommendations dataset.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Eickhoff, C. Cognitive Biases in Crowdsourcing. In Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, Los Angeles, CA, USA, 5–9 February 2018; pp. 162–170.
2. Saito, S.; Kobayashi, T.; Nakano, T. Towards a Framework for Collaborative Video Surveillance System Using Crowdsourcing. In Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion, San Francisco, CA, USA, 27 February–2 March 2016; pp. 393–396.
3. Fan, J.; Lu, M.; Ooi, B.C.; Tan, W.; Zhang, M. A hybrid machine-crowdsourcing system for matching web tables. In Proceedings of the 2014 IEEE 30th International Conference on Data Engineering, Chicago, IL, USA, 31 March–4 April 2014; pp. 976–987.

4. Tan, J.T.C.; Hagiwara, Y.; Inamura, T. Learning from Human Collaborative Experience: Robot Learning via Crowdsourcing of Human-Robot Interaction. In Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, Vienna, Austria, 6–9 March 2017; pp. 297–298.
5. Wang, J.; Kraska, T.; Franklin, M.J.; Feng, J. CrowdER: Crowdsourcing Entity Resolution. *Proc. VLDB Endow.* **2012**, *5*, 1483–1494. [CrossRef]
6. Liu, X.; Lu, M.; Ooi, B.C.; Shen, Y.; Wu, S.; Zhang, M. CDAS: A Crowdsourcing Data Analytics System. *Proc. VLDB Endow.* **2012**, *5*, 1040–1051. [CrossRef]
7. Dawid, A.P.; Skene, A.M. Maximum likelihood estimation of observer error-rates using the em algorithm. *Appl. Stat.* **1979**, *28*, 20–28. [CrossRef]
8. Xu, L.; Hao, X.; Lane, N.D.; Liu, X.; Moscibroda, T. More with Less: Lowering User Burden in Mobile Crowdsourcing through Compressive Sensing. In Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Osaka, Japan, 7–11 September 2015; pp. 659–670.
9. Kae-Nune, N.; Pesseguier, S. Qualification and Testing Process to Implement Anti-Counterfeiting Technologies into IC Packages. In Proceedings of the 2013 Design, Automation & Test in Europe Conference & Exhibition (DATE), Grenoble, France, 18–22 March 2013; pp. 1131–1136.
10. Demartini, G.; Difallah, D.E.; Cudré-Mauroux, P. ZenCrowd: Leveraging probabilistic reasoning and crowdsourcing techniques for large-scale entity linking. In Proceedings of the 21st International Conference on World Wide Web, Lyon, France, 16–20 April 2012; pp. 469–478.
11. Franklin, M.J.; Kossmann, D.; Kraska, T.; Ramesh, S.; Xin, R. CrowdDB: Answering queries with crowdsourcing. In Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data, Athens, Greece, 12–16 June 2011; pp. 61–72.
12. Zheng, Y.; Cheng, R.; Maniu, S.; Mo, L. On Optimality of Jury Selection in Crowdsourcing. In Proceedings of the 18th International Conference on Extending Database Technology, Brussels, Belgium, 23–27 March 2015; pp. 193–204.
13. Zheng, Y.; Wang, J.; Li, G.; Cheng, R.; Feng, J. QASCA: A Quality-Aware Task Assignment System for Crowdsourcing Applications. In Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, Victoria, Australia, 31 May–4 June 2015; pp. 1031–1046.
14. Han, F.; Tan, S.; Sun, H.; Srivatsa, M.; Cai, D.; Yan, X. Distributed Representations of Expertise. In Proceedings of the 2016 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics, Miami, FL, USA, 5–7 May 2016; pp. 531–539.
15. Liu, X.; He, H.; Baras, J.S. Crowdsourcing with multi-dimensional trust. In Proceedings of the 2015 18th International Conference on Information Fusion, Washington, DC, USA, 6–9 July 2015; pp. 574–581.
16. Dong, X.L.; Berti-Équille, L.; Srivastava, D. Truth Discovery and Copying Detection in a Dynamic World. *Proc. VLDB Endow.* **2009**, *2*, 562–573. [CrossRef]
17. Boim, R.; Greenshpan, O.; Milo, T.; Novgorodov, S.; Polyzotis, N.; Tan, W.C. Asking the Right Questions in Crowd Data Sourcing. In Proceedings of the 2012 IEEE 28th International Conference on Data Engineering, Arlington, VA, USA, 1–5 April 2012; pp. 1261–1264.
18. Chen, X.; Lin, Q.; Zhou, D. Optimistic Knowledge Gradient Policy for Optimal Budget Allocation in Crowdsourcing. In Proceedings of the International Conference on Machine Learning, Atlanta, GA, USA, 16–21 June 2013; Volume 28, pp. 64–72.
19. Khan, A.R.; Garcia-Molina, H. CrowdDQS: Dynamic Question Selection in Crowdsourcing Systems. In Proceedings of the 2017 ACM International Conference on Management of Data, Chicago, CA, USA, 14–19 May 2017; pp. 1447–1462.
20. Parameswaran, A.G.; Garcia-Molina, H.; Park, H.; Polyzotis, N.; Ramesh, A.; Widom, J. CrowdScreen: Algorithms for filtering data with humans. In Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data, Scottsdale, AZ, USA, 20–24 May 2012; pp. 361–372.
21. Gao, J.; Liu, X.; Ooi, B.C.; Wang, H.; Chen, G. An online cost sensitive decision-making method in crowdsourcing systems. In Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data, Indianapolis, IN, USA, 6–10 June 2013; pp. 217–228.
22. Mo, L.; Cheng, R.; Kao, B.; Yang, X.S.; Ren, C.; Lei, S.; Cheung, D.W.; Lo, E. Optimizing plurality for human intelligence tasks. In Proceedings of the 22nd ACM International Conference on Information & Knowledge Management, San Francisco, CA, USA, 27 October–1 November 2013; pp. 1929–1938.
23. Sheng, V.S.; Provost, F.J.; Ipeirotis, P.G. Get another label? Improving data quality and data mining using multiple, noisy labelers. In Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Las Vegas, NA, USA, 24–27 August 2008; pp. 614–622.
24. Mo, K.; Zhong, E.; Yang, Q. Cross-task crowdsourcing. In Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Chicago, CA, USA, 11–14 August 2013; pp. 677–685.
25. Qiu, C.; Squicciarini, A.C.; Carminati, B.; Caverlee, J.; Khare, D.R. CrowdSelect: Increasing Accuracy of Crowdsourcing Tasks through Behavior Prediction and User Selection. In Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, Indianapolis, IN, USA, 24–28 October 2016; pp. 539–548.
26. Dai, P.; Lin, C.H.; Mausam; Weld, D.S. POMDP-based control of workflows for crowdsourcing. *Artif. Intell.* **2013**, *202*, 52–85. [CrossRef]

27. Karger, D.R.; Oh, S.; Shah, D. Iterative Learning for Reliable Crowdsourcing Systems. In Proceedings of the Neural Information Processing Systems, Granada, Spain, 12–15 December 2011; pp. 1953–1961.
28. Qiu, S.; Gadiraju, U.; Bozzon, A. Improving Worker Engagement Through Conversational Microtask Crowdsourcing. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 25–30 April 2020; pp. 1–12.
29. Qiu, S.; Gadiraju, U.; Bozzon, A. *Just the Right Mood for HIT!—Analyzing the Role of Worker Moods in Conversational Microtask Crowdsourcing*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 381–396.
30. Zhuang, M.; Gadiraju, U. In What Mood Are You Today?: An Analysis of Crowd Workers' Mood, Performance and Engagement. In Proceedings of the 10th ACM Conference on Web Science, Boston, MA, USA, 30 June–3 July 2019; pp. 373–382.
31. Mavridis, P.; Huang, O.; Qiu, S.; Gadiraju, U.; Bozzon, A. Chatterbox: Conversational Interfaces for Microtask Crowdsourcing. In Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization, Larnaca, Cyprus, 9–12 June 2019; pp. 243–251.
32. Qiu, S.; Gadiraju, U.; Bozzon, A. TickTalkTurk: Conversational Crowdsourcing Made Easy. In Proceedings of the Conference Companion Publication of the 2020 on Computer Supported Cooperative Work and Social Computing, Virtual Event, 17–21 October 2020; pp. 53–57.
33. Shin, S.; Choi, H.; Yi, Y.; Ok, J. Power of Bonus in Pricing for Crowdsourcing. *Proc. ACM Meas. Anal. Comput. Syst.* **2021**, *5*, 36:1–36:25. [CrossRef]
34. Miao, X.; Kang, Y.; Ma, Q.; Liu, K.; Chen, L. Quality-aware Online Task Assignment in Mobile Crowdsourcing. *ACM Trans. Sens. Netw.* **2020**, *16*, 30:1–30:21. [CrossRef]
35. Fang, Y.; Si, L.; Mathur, A.P. Discriminative models of integrating document evidence and document-candidate associations for expert search. In Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, Geneva, Switzerland, 19–23 July 2010; pp. 683–690.
36. Deng, H.; King, I.; Lyu, M.R. Formal Models for Expert Finding on DBLP Bibliography Data. In Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, Pisa, Italy, 15–19 December 2008; pp. 163–172.
37. Serdyukov, P.; Rode, H.; Hiemstra, D. Modeling multi-step relevance propagation for expert finding. In Proceedings of the 17th ACM Conference on Information and Knowledge Management, Napa Valley, CA, USA, 26–30 October 2008; pp. 1133–1142.
38. Guan, Z.; Yang, S.; Sun, H.; Srivatsa, M.; Yan, X. Fine-Grained Knowledge Sharing in Collaborative Environments. *IEEE Trans. Knowl. Data Eng.* **2015**, *27*, 2163–2174. [CrossRef]
39. Davenport, T.H.; Prusak, L. Working knowledge: How organizations manage what they know. *Ubiquity* **2000**, *2000*, 6. [CrossRef]
40. Mimno, D.M.; McCallum, A. Expertise modeling for matching papers with reviewers. In Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, London, UK, 19–23 August 2007; pp. 500–509.
41. Rosen-Zvi, M.; Chemudugunta, C.; Griffiths, T.L.; Smyth, P.; Steyvers, M. Learning author-topic models from text corpora. *ACM Trans. Inf. Syst.* **2010**, *28*, 4:1–4:38. [CrossRef]
42. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
43. Zhao, W.X.; Jiang, J.; Weng, J.; He, J.; Lim, E.; Yan, H.; Li, X. Comparing Twitter and Traditional Media Using Topic Models. In Proceedings of the European Conference on Information Retrieval, Dublin, Ireland, 18–21 April 2011; pp. 338–349.
44. Zheng, Y.; Li, G.; Cheng, R. DOCS: A Domain-Aware Crowdsourcing System Using Knowledge Bases. *Proc. VLDB Endow.* **2016**, *10*, 361–372. [CrossRef]
45. Goel, G.; Nikzad, A.; Singla, A. Allocating tasks to workers with matching constraints: Truthful mechanisms for crowdsourcing markets. In Proceedings of the 23rd International Conference on World Wide Web, Seoul, Korea, 7–11 April 2014; pp. 279–280.
46. Han, L.; Maddalena, E.; Checco, A.; Sarasua, C.; Gadiraju, U.; Roitero, K.; Demartini, G. Crowd Worker Strategies in Relevance Judgment Tasks. In Proceedings of the 13th International Conference on Web Search and Data Mining, Houston, TX, USA, 3–7 February 2020; pp. 241–249.
47. Yin, X.; Han, J.; Yu, P.S. Truth Discovery with Multiple Conflicting Information Providers on the Web. *IEEE Trans. Knowl. Data Eng.* **2008**, *20*, 796–808.
48. Zhao, B.; Rubinstein, B.I.P.; Gemmell, J.; Han, J. A Bayesian Approach to Discovering Truth from Conflicting Sources for Data Integration. *Proc. VLDB Endow.* **2012**, *5*, 550–561. [CrossRef]
49. Pochampally, R.; Sarma, A.D.; Dong, X.L.; Meliou, A.; Srivastava, D. Fusing data with correlations. In Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data, Snowbird, UT, USA, 22–27 June 2014; pp. 433–444.
50. Zheng, Y.; Li, G.; Li, Y.; Shan, C.; Cheng, R. Truth Inference in Crowdsourcing: Is the Problem Solved? *Proc. VLDB Endow.* **2017**, *10*, 541–552. [CrossRef]
51. Ma, F.; Li, Y.; Li, Q.; Qiu, M.; Gao, J.; Zhi, S.; Su, L.; Zhao, B.; Ji, H.; Han, J. FaitCrowd: Fine Grained Truth Discovery for Crowdsourced Data Aggregation. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, Australia, 10–13 August 2015; pp. 745–754.
52. Galland, A.; Abiteboul, S.; Marian, A.; Senellart, P. Corroborating information from disagreeing views. In Proceedings of the Third ACM International Conference on Web Search and Data Mining, New York, NY, USA, 3–6 February 2010; pp. 131–140.
53. Lin, X.; Chen, L. Domain-Aware Multi-Truth Discovery from Conflicting Sources. *Proc. VLDB Endow.* **2018**, *11*, 635–647. [CrossRef]

54. Miao, C.; Jiang, W.; Su, L.; Li, Y.; Guo, S.; Qin, Z.; Xiao, H.; Gao, J.; Ren, K. Cloud-Enabled Privacy-Preserving Truth Discovery in Crowd Sensing Systems. In Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems, Seoul, Korea, 1–4 November 2015; pp. 183–196.
55. Tang, X.; Wang, C.; Yuan, X.; Wang, Q. Non-Interactive Privacy-Preserving Truth Discovery in Crowd Sensing Applications. In Proceedings of the IEEE INFOCOM 2018-IEEE Conference on Computer Communications, Honolulu, HI, USA, 15–19 April 2018; pp. 1988–1996.
56. Yang, S.; Wu, F.; Tang, S.; Gao, X.; Yang, B.; Chen, G. On Designing Data Quality-Aware Truth Estimation and Surplus Sharing Method for Mobile Crowdsensing. *IEEE J. Sel. Areas Commun.* **2017**, *35*, 832–847. [CrossRef]
57. Xiao, M.; Wu, J.; Zhang, S.; Yu, J. Secret-sharing-based secure user recruitment protocol for mobile crowdsensing. In Proceedings of the INFOCOM 2017-IEEE Conference on Computer Communications, Atlanta, GA, USA, 1–4 May 2017; pp. 1–9.
58. Jin, H.; Su, L.; Nahrstedt, K. CENTURION: Incentivizing multi-requester mobile crowd sensing. In Proceedings of the IEEE INFOCOM 2017-IEEE Conference on Computer Communications, Atlanta, GA, USA, 1–4 May 2017.
59. Wang, X.; Sheng, Q.Z.; Fang, X.S.; Yao, L.; Xu, X.; Li, X. An Integrated Bayesian Approach for Effective Multi-Truth Discovery. In Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, Melbourne, Australia, 19–23 October 2015; pp. 493–502.
60. Jiang, L.; Niu, X.; Xu, J.; Yang, D.; Xu, L. Incentivizing the Workers for Truth Discovery in Crowdsourcing with Copiers. In Proceedings of the 2019 IEEE 39th International Conference on Distributed Computing Systems, Dallas, TX, USA, 7–9 July 2019; pp. 1286–1295.
61. Ho, C.; Vaughan, J.W. Online Task Assignment in Crowdsourcing Markets. In Proceedings of the AAAI Conference on Artificial Intelligence, Toronto, ON, Canada, 22–26 July 2012; pp. 45–51.
62. Tran-Thanh, L.; Stein, S.; Rogers, A.; Jennings, N.R. Efficient crowdsourcing of unknown experts using bounded multi-armed bandits. *Artif. Intell.* **2014**, *214*, 89–111. [CrossRef]
63. Harper, F.M.; Konstan, J.A. The MovieLens Datasets: History and Context. *ACM Trans. Interact. Intell. Syst.* **2016**, *5*, 19:1–19:19. [CrossRef]