Supplementary materials

In this section, to further explore the effect of different strategies in the three sessions on the quality of result, we conduct more experiments concerning more treatments (as listed in Table 2) beyond the treatments mentioned in Section 3.

Experimental setting

We use the WeChat applet to inject users into the proposed platform and design our experiments in a way that each treatment consists of the same number of tasks (5 in each treatments), so neither treatment appears to be obviously more time-consuming or enjoyable. In the pre-hoc handling of the same subtasks in a phase, if the *synchronous* coordination strategy is implemented for coordinating workers, the redundancy of workers is set to 5. If the *asynchronous* coordination strategy is implemented for coordinating workers, the redundancy of workers is set according to specific quality requirements. In general, in the common crowdsourcing platforms such as CrowdFlower, the salary for each hired worker is set according to the number of tasks a worker has completed and the reward to each hired worker for completing per subtask is fixed. In the experiments, the reward for each hired worker is 0.5 yuan per person for completing per subtask, and the more tasks they complete the more reward they can obtain. In order to protect the privacy of the workers, the workers are anonymous, as the common crowdsourcing platforms do. Importantly, we did not tell workers the strategies in different treatments, which helps us to minimize the biases as the two treatments look the same to workers.

Table 2 Eight experimental treatments.

Methods	Session 1		Session 2		Session 3	
	w/o CCS	With CCS	*Syn*	*Asy*	w/o CCS	With CCS
NSN	☑	⊠	☑	⊠	☑	⊠
CSN	⊠	☑	☑	⊠	☑	⊠
NAN	☑	⊠	⊠	☑	☑	⊠
CAN	⊠	☑	⊠	☑	☑	⊠
NSC	☑	⊠	☑	⊠	⊠	☑
CSC	⊠	☑	☑	⊠	⊠	☑
NAC	⊠	☑	⊠	☑	⊠	☑
CAC	⊠	☑	⊠	☑	⊠	☑

☑ denotes the factor is considered in model, while ⊠ denotes not, CCS is shorted for Considering the Correlation among Subtasks, *Syn* is shorted for synchronous and *Asyn* is shorted for asynchronous as defined previously. In task design, most work [29,17] does not consider the correlation among tasks. In the process of pre-hoc coordination of workers, most work [77,25] adopt the *Asyn* mode. In the process of post-hoc coordination of crowd work, most work [17,2] is without CCS.
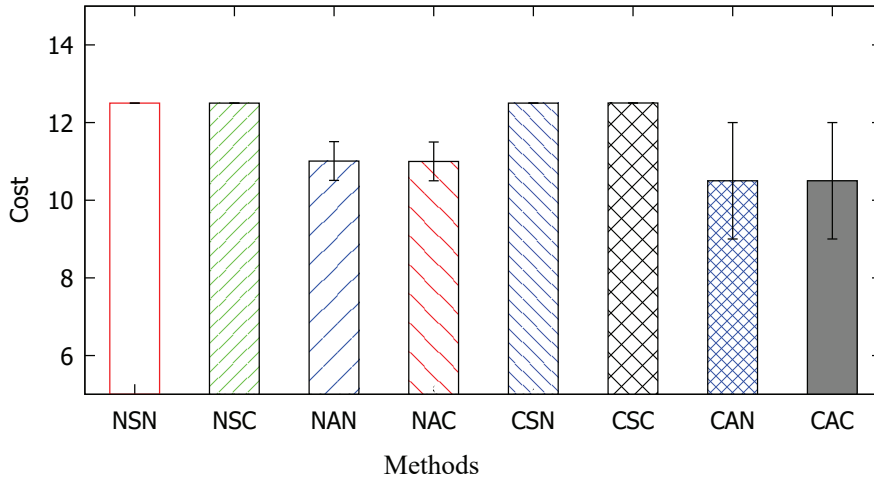
(a) The costs of different methods.
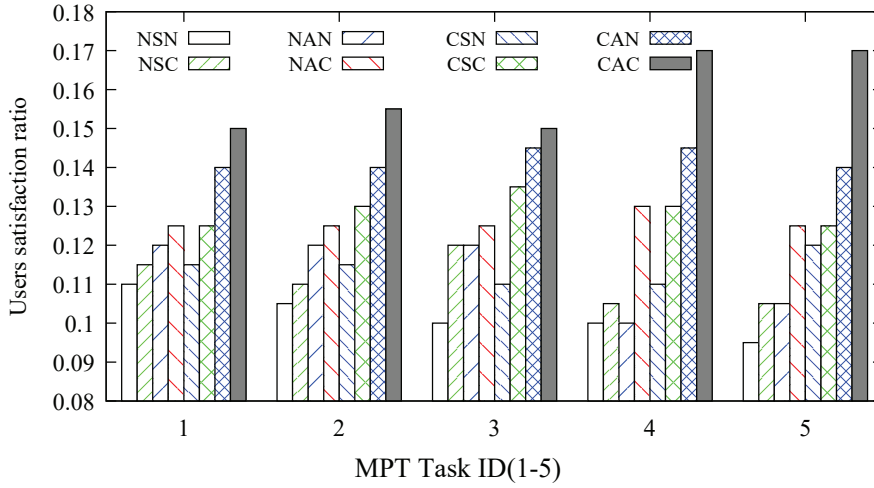
(b) The result quality of different methods.

Fig. 11 Evaluation of different treatments.

Additional experiments concerning costs of treatments

Fig 11(a) plots the costs of different treatments. We can observe that among the eight treatments, the costs of NSN, NSC, CSN, and CSC are the same as one another and all equal to 12.5 yuan. This is because, in the strategy for post-coordination of crowd work, the costs are the same in $K = 1$ case of the quality-aware method and the greedy algorithm without considering the correlation among subtasks. In the second session of these treatments, the number of workers' redundancy is determined. So the cost of recruiting

workers is also determined, which is $0.5 \times 5 \times 5 = 12.5$ yuan. Concerning the costs of four groups treatments NAN (NSN), NAC (NSC), CAN (CSN), and CAC (CSC), the former NAN, NAC, CAN, and CAC are lower than the latter NSN, NSC, CSN, and CSC. This is because the latter uses asynchronous subtask processing methods in their second sessions.

We can also observe that concerning the costs of two groups' treatments NAN (CAN) and NAC (CAC), the former is higher than the latter. This is because in the former treatments of each group, in their first sessions, the correlation among subtasks is exhibited to workers. In this way, the quality of workers is improved as discussed previously. Thus, in the second session, there is a need for fewer workers for achieving a certain result quality in asynchronous subtask processing methods.

We can also see that concerning the costs of four groups NSC (NSN), NAN (NAC), CSN (CSC), and CAN (CAC), the former NSC, NAN, CSN, and CAN are the same as the latter NSN, NAC, CSC, and CAC. This is because in the strategy for post-coordination of crowd work, the costs are the same in $K = 1$ case of the quality-aware method and the greedy algorithm without considering the correlation among subtasks.

In summary, considering the correlation not only does not incur more costs in the third session but also can reduce the costs by improving the quality of crowd work in the first session. Moreover, if the asynchronous subtask processing methods are adopted in the second session, the costs can be reduced.

Additional experiments concerning the result quality of treatments

To compare the quality of the outcomes of MPTs obtained in eight treatments, we recruit the other 40 workers and ask each of them to select the best one from the eight treatments. Then we use the approval ratios to quantify the result quality of MPTs obtained in different treatments. The approval ratios of treatment can be computed with the ratio of the number of workers who selected this treatment as the best one she thinks to the number of all the workers who participate in the evaluation tasks. The approval ratios of different treatments is plotted in Fig 11(b).

We can also observe that concerning the approval ratios of four groups treatments NSN(CSN), NAN(CAN), NSC(CSC) and NAC(CAC), the former NSN, NAN, NSC, and NAC are lower than the latter CSN, CAN, CSC and CAC. This is because in the latter groups of treatments, in their first sessions, the correlation among subtasks is exhibited to workers. In this way, the quality of workers is improved as discussed previously. Thus, the crowd work of these workers is better and deserves more votes of approval.

We can also see that concerning the approval ratios of four groups treatments NSN(NAN), CSN(CAN), NSC(NAC) and CSC (CAC), the former NSN, CSN, NSC, and CSC are the same as the latter NAN, CAN, NAC and CAC. This is because even in the second session, they adopt different strategies, dif-

ferent strategies in the second session mainly affect the costs but the quality of crowd work.

Fig 10(b) also plots that concerning the approval ratios of four groups treatments NSN(NSC), NAN(NAC), CSN(CSC), and CAN(CAC), the former NSN, NAN, CSN, and CAN are the lower than the latter NSC, NAC, CSC, and CAC. This is because in the latter groups of treatments, in their third sessions, the correlation among subtasks is considered in the process of post-coordination of crowd work in different phase-level subtasks of the MPTs. In this way, as a whole, a complete MPT can be better handled.

Among all the treatments, the CAC obtains the highest approval ratios, this is because, in these treatments, the correlation among subtasks is considered both in the first and the third sessions.