Toward Worker-Centric Crowdsourcing

Sihem Amer-Yahia
Univ. Grenoble Alpes, CNRS, LIG
F-38000 Grenoble, France
sihem.amer-yahia@imag.fr

Senjuti Basu Roy
New Jersey Institute of Technology
Newark, NJ, USA
senjutib@njit.edu

Abstract

Today, crowdsourcing is used to "taskify" any job ranging from simple receipt transcription to collaborative editing, fan-subbing, and citizen science. Existing work has mainly focused on improving the processes of task assignment and task completion in a requester-centric *way by optimizing for outcome quality under budget constraints. In this paper, we advocate that accounting for workers' characteristics, i.e.,* human factors *in task assignment and task completion benefits* both *workers and requesters, and discuss new opportunities raised by* worker-centric *crowdsourcing. This survey is based on a tutorial that was given recently at PVLDB [2].*

1 A Case for Worker-Centric Crowdsourcing

As more jobs are being "taskified" and executed on crowdsourcing platforms, the role of human workers online is gaining importance. On virtual marketplaces such as Amazon Mechanical Turk, PyBossa and Crowd4U, the crowd is volatile, its arrival and departure asynchronous, and its levels of attention and accuracy diverse. Tasks differ in complexity and necessitate the participation of workers with varying degrees of expertise. As workers continue to get involved in crowdsourcing, a legitimate question is how to improve both their performance and their experience. Existing proposals have been mostly concerned with the development of requester-centric algorithms to match tasks and workers and with preemptive approaches to improve task completion. We believe that new opportunities in developing models and algorithms are yet to be explored in bridging the gap between Social Science studies and Computer Science. Naturally, *understanding the characteristics of workers, here referred to as human factors*, that directly impact their performance and their experience on the platform, is a necessary step toward achieving that goal. We advocate a re-focus of research in crowdsourcing on how to best leverage human factors at all stages that will widen the scope and impact of crowdsourcing and make it beneficial to both requesters and workers. Several other complementary surveys could be found in the literature [3, 6].

Common tasks such as labeling images or determining the sentiment of a piece of text, can be completed by each worker independently. These types of crowdsourcing tasks are known as *micro-tasks*. An emerging area of interest is *collaborative crowdsourcing* where workers complete a task together. Examples include fan-subbing, where workers with complementary skills collaborate to generate movie subtitles in various languages, just hours after movies are made available. Disaster reporting is another example where geographically close people with diverse and complementary skills work together to report the aftermath of an earthquake. Section 2

Bulletin of the IEEE Computer Society Technical Committee on Data Engineering

Worker-specific	Micro-tasks: *Skill, Reputation/Trust, Expected Pay, Acceptance Ratio* Collaborative tasks: *Affinity, Critical Mass, Interaction model*
Task-specific	*Feedback, Incentives, Skill Variety, Task Identity, Task Autonomy, Expected Quality, Budget, Desired Expertise*
Workers and Tasks	*Motivation*

Table 1: A rough characterization of human factors

reviews human factors for micro-tasks and collaborative tasks in conjunction with psychology studies conducted in the 70's in physical workplaces. It then reports on more recent empirical evaluation of human factors in virtual marketplaces. The outcome of that section is a review of existing approaches to model, acquire and learn human factors.

The two main processes that leverage human factors in virtual marketplaces are task assignment and task completion. Section 3 reviews algorithms and approaches for assigning tasks to workers and various approaches to intervene during task completion and improve overall performance. We review task assignment for both micro-tasks and collaborative tasks and draw a connection with findings in psychology. This connection brings an understanding of which factors are most likely to affect workers' choice of tasks and their performance during task completion.

The review we provide in Section 3 naturally leads to the second half of this paper. Section 4 is dedicated to new opportunities raised by leveraging human factors in crowdsourcing. This section exposes a number of promising directions that contribute to worker-centric crowdsourcing, a paradigm shift that we believe will lead to sustainable crowdsourcing.

2 Human Factors at Work

A variety of human factors characterize workers and their environment at work. Their genesis goes back to the 70's when "organization studies" and "work theory" were developing models to understand motivation in physical workplaces. A flagship study is that of Hackman and Oldham in 1976 whose goal was to determine which psychological states are stimulated by which job characteristics. The authors ran experiments on 658 employees in 62 heterogeneous jobs (white collar, blue collar, industry, services, urban and rural settings) in 7 organizations. The study showed that modeling extrinsic motivation such as how much a job pays, and intrinsic motivation such as whether a job provides feedback to workers, are critical for measuring workers' psychological state and hence their satisfaction and performance in the workplace.

We gathered the most common human factors from the literature and characterized them as *worker-specific*, *task-specific*, or specific to *both workers and tasks.* Table 1 contains a summary of the most common factors identified in both physical and virtual marketplaces.

In practice, human factors are mostly acquired via questionnaires and qualification tests. They can also be learned from workers' previous performance in completing tasks. We review the literature on modeling and acquiring human factors.

2.1 Worker-Specific Human Factors

In this subsection, we discuss the human factors that are related to workers. Only Skill and Reputation/Trust are discussed because Expected Pay is acquired directly from workers via a questionnaire, and Acceptance Ratio is computed as a proportion of tasks for which the worker's contribution has been accepted (out of all tasks the worker completed).

2.1.1 Skill and Reputation/Trust

Existing research has investigated the skill and trust estimation problem in several ways, primarily in the context of micro-tasks. For example, for labeling tasks, a probabilistic model was proposed to infer the true label of each image, the expertise of each labeler, and the difficulty level of each task [44]. For annotation tasks, Bayesian solutions were used to iteratively establish and refine a particular golden standard, measure the performance of annotators with respect to that standard, or eliminate spammers [34, 35]. For the same kind of tasks, a more recent work focused on determining worker confidence intervals and worker error rates [20]. A follow-up work [21] designed solutions for the case where not all workers have attempted every task, tasks have non-Boolean responses, and workers have different biases for making false positive and false negative errors for Boolean tasks. Additional work focused on the problem of identifying workers who systematically disagree both with the majority and with the rest of co-workers [41, 42].

In contrast to micro-tasks, there exists only one effort in estimating human factors in team-based tasks [33]. In that work, skill estimation is based on modeling task quality as an aggregation of individual worker skills and their collaboration effectiveness, and on solving an optimization problem under different skill aggregation functions, *sum, max, and min*. The optimization problem reverse-engineers worker skills and collaboration effectiveness from observed outcome quality.

2.2 Task-Specific Human Factors

In this subsection, we describe human factors that are pertinent to tasks, namely Feedback and Incentives. Skill Variety represents the number of different skills a task requires from a worker. Task Identity represents whether a task is part of a bigger task or not. Task Autonomy indicates if a worker depends on others. Expected Quality, Desired Expertise and Budget are used to set a minimum threshold on workers' contributions.

2.2.1 Feedback

The importance of task feedback is studied in CrowdFlower [4]. It was shown that both immediate and long-term feedback helps to improve the quality of completed tasks. This study also indicates that workers expect they should be provided with a meaningful explanation of why their work is rejected, or in case their work is accepted, they expect reasonable turnaround time between submitting the work and receiving payment for it.

2.2.2 Incentives

Incentives have been studied using qualitative and quantitative approaches. In [40], two different types of incentives are studied - social incentives and financial incentives. That work empirically shows that both types can improve worker productivity. A recent work [17] focuses on performance-based payments (PBP) through financial incentives. It empirically tests the effect of varying outcome quality threshold in order for workers to receive a bonus and the effect of varying the bonus amount on task quality. It also recommends running a pilot experiment to determine whether a task is effort-responsive and then design PBP schemes.

A quantitative study [8] presents algorithms for dynamic pricing to meet (a) a user-specified deadline while minimizing total monetary cost, or (b) a user-specified budget constraint while minimizing total elapsed time.

2.3 Worker- and Task-Specific Human Factors

Finally, we describe human factors that are pertinent to both workers and tasks. In this context, studying workers' motivation in completing tasks has been the center of attention. One of the earliest studies of motivation in virtual marketplaces was conducted on Amazon Mechanical Turk [24]. The goal of that study was to empirically verify which of several intrinsic and extrinsic motivation factors were considered important to workers. Figure 1 summarizes the results of an offline evaluation of 13 human factors related to motivation. The results were obtained by asking workers to fill out questionnaires after completing tasks. The goal of the questionnaires was to determine which task-specific factors, Skill Variety, Task Identity, etc., and which other factors, Social Contact, Human Capital Advancement, etc., affect motivation. While Payment remains a highly motivating factor, the study also points out the cumulative importance of "Enjoyment-Based Motivation" factors when compared to Payment. One other highlight is that a large proportion of workers declared that "Human Capital Advancement" was an important motivation in completing tasks.

A later study [37] found a synergistic interaction between intrinsic and extrinsic motivators and demonstrated that increasing levels of payment increases task throughput regardless of other factors. However, increasing task throughput does not necessarily mean that workers do a good job at completing tasks. It was indeed shown that increasing pay does not increase the quality of workers' contributions [26].

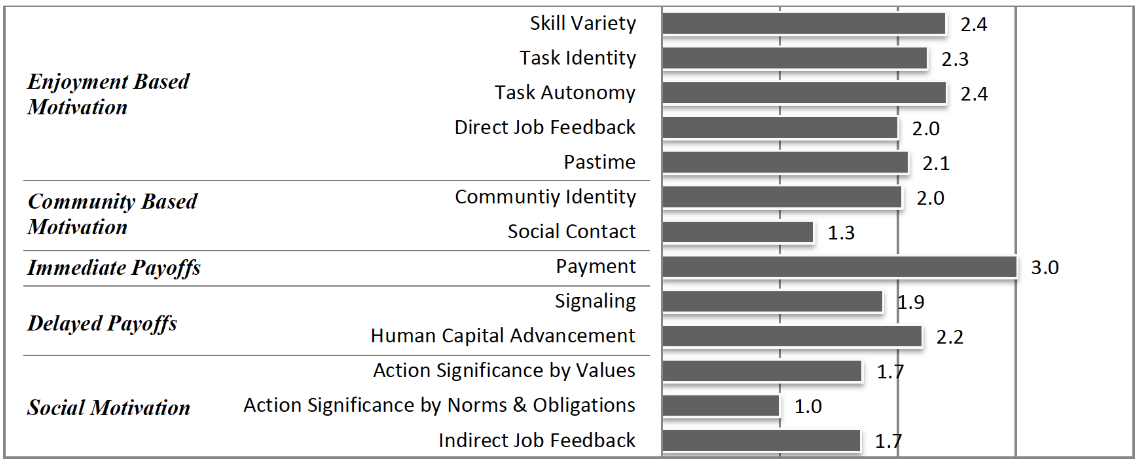


Figure 1: Results of an offline evaluation of human factors [24]

3 Human Factors in Task Assignment and Task Completion

Human factors are mostly recognized in a requester-centric fashion primarily for the purpose of task assignment and in a limited way for task completion. We present existing work in two parts - first, in the context of micro-tasks, then for collaborative tasks.

3.1 Micro-Task Assignment

Amazon Mechanical Turk only allows self-assignment to tasks - although task designers could specify desired worker qualifications for a given set of tasks via system-centric qualifications (e.g., HIT Approval Rate, also referred to as Acceptance Ratio) and platform-centric qualifications (e.g., marital status, education level, political affiliation, etc.).

Related work performs task assignment for micro-tasks, acknowledging primarily worker skill and budget. This body of work belongs to one of the following two kinds.

Researchers in Artificial Intelligence, Machine Learning, and Operations Research primarily assume that workers are distinguishable. Each individual worker has an associated id with a known skill and cost. Given a task with a budget and desired accuracy threshold, the task assignment problem is studied as a *matching* problem between workers and tasks [9, 15, 16, 60].

On the other hand, researchers in Databases have primarily assumed that workers are indistinguishable. A generic skill and cost model is assumed for all workers and a given set of tasks have an associated budget. Different types of datasourcing applications [5] are considered as human-machine intelligence tasks - such as filtering [28], sorting and join [25], deduplication and clustering [43], categorization [36], evaluating order queries [10], or even mining [1] involving human workers. Under this setting, the task assignment problem becomes that of selecting the best subset of tasks for the worker pool that satisfies the budget and is most likely to maximize quality.

3.2 Collaborative Task Assignment

Collaborative tasks need to be performed by a set of workers together as a team (e.g., collaboratively editing an article where each article has a minimum quality and a maximum cost requirement, as well as the need for complementary skills) and it is assumed that the workers' profiles are known (skill per domain, requested wage). In a recent work [38], the objective function is formalized so as to guarantee that each task surpasses its quality threshold, stays below its cost limit, and that workers are not over-utilized or under-utilized. Given the innate uncertainty induced by human involvement, a third human factor, Acceptance Ratio (e.g., computed as the probability that a worker accepts a recommended task) is also used in the problem formulation.

The experimental results presented in this paper indicate that despite assigning tasks to highly-qualified workers, for some cases, quality of completed tasks was low due to conflicting opinions, edit wars, or proliferation of edits. A follow-up work [33] proposes additional human factors to ensure that *teams are not too large* - in particular, it introduces *Upper Critical Mass* as a human factor that constrains the size of a team. For a given task, if the required number of workers are too many, then the objective is to form sub-teams, where both intra-team and inter-team collaboration is allowed. By leveraging related research in Psychology [14], the authors model collaborative aspects between workers as *Affinity*. Affinity is formalized using socio-demographic attributes, such as region, age, or psychological characteristics [27]. The objective function intends to form a set of teams for a collaborative task, where each sub-team must satisfy the critical mass constraint, their intra- and inter-team affinity is maximized, while satisfying a minimum quality requirement and a maximum cost budget. The problem is proved to be NP-hard and the authors propose a staged solution by designing approximation algorithms, where each stage individually admits an approximation factor.

3.3 Micro-Task Completion

Improving micro-task completion has mainly relied on monitoring worker motivation and taking appropriate actions to improve completion. Some efforts [7, 39] have shown that including a diversion or a break, such as showing an entertaining video, improves workers' motivation. More recently, pre-emption was exercised to interrupt workers who have not completed tasks on time [12].

3.4 Collaborative Task Completion

Failing to complete a micro-task has only "local" impact since it does not prevent other workers to complete that same task in parallel. Failing from completing a collaborative task is however more damaging since it leads to an uncompleted task for all. Despite its importance, monitoring task completion for collaborative tasks is still in its infancy. There exists one piece of work to date that addresses this problem [38]. This work proposes an adaptive task completion scheme that handles scenarios such as the arrival of new workers and tasks, and the departure of workers without finalizing tasks. The proposed solution is based on formulating a marginal IP problem that re-assigns tasks to workers when new workers or tasks arrive, or when workers leave. The assignment is based on the same objective functions described in Section 3.2.

4 Challenges and Opportunities

So far, we have discussed how human factors are modeled and monitored to optimize the main processes of a crowdsourcing platform, namely task assignment and task completion. In this section, we aim to widen the scope and impact of human factors in crowdsourcing and discuss how they can enable worker-centricity. In our opinion, it is essential to shift from requester-centric optimizations to an approach that integrates what workers want from a crowdsourcing platform. This shift will induce a tighter integration between human factors and the processes of a crowdsourcing system. In this section, we discuss some essential elements reailzing this shift. We start with the need for declarative tools that let workers benefit from crowdsourcing. We then examine the evolving nature of human factors and its impact on workers' performance. We move on to discuss evaluation and other factors that affect workforce organization and enabling experimental repeatability in crowdsourcing.

4.1 Declarativity

There have been several efforts in the database community to develop declarative languages that let requesters decompose complex tasks or specify task assignment criteria. Designing a declarative language that helps workers exploit the potential of a crowdcourcing platform appears as a natural goal. Workers should be able to express a number of desiderata such as acquiring or improving a specific skill, or being entertained for a specific period of time. The development of worker-centric primitives such as finding tasks of interest to a given worker or being notified when a particular requester posts tasks, opens new modeling and algorithmic opportunities that complement existing solutions for task assignment and task completion. Such primitives should be designed with an understanding of how worker-centric and task-centric human factors interleave and affect performance. Moreover, they should not hinder requester-centric optimizations. Rather, they should complement them. The overarching goal should be to bring together all the components of a crowdsourcing platform and benefit from an adaptive framework within which workers are observed and provided tasks that serve them and serve requesters. Such an integrative approach would close the loop between different crowdsourcing processes as shown in Figure 2.

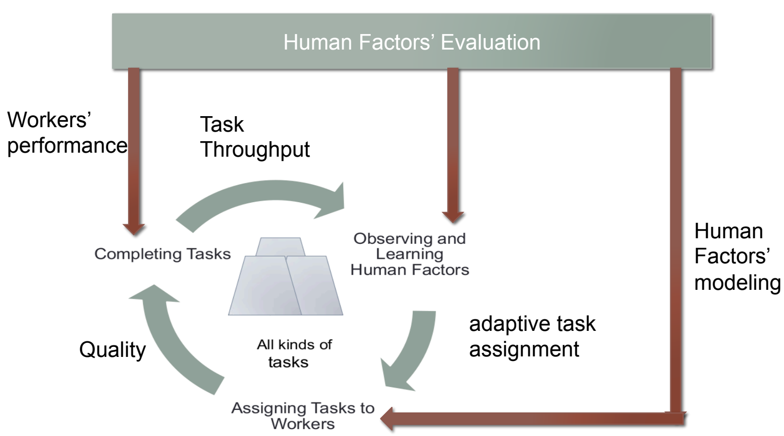
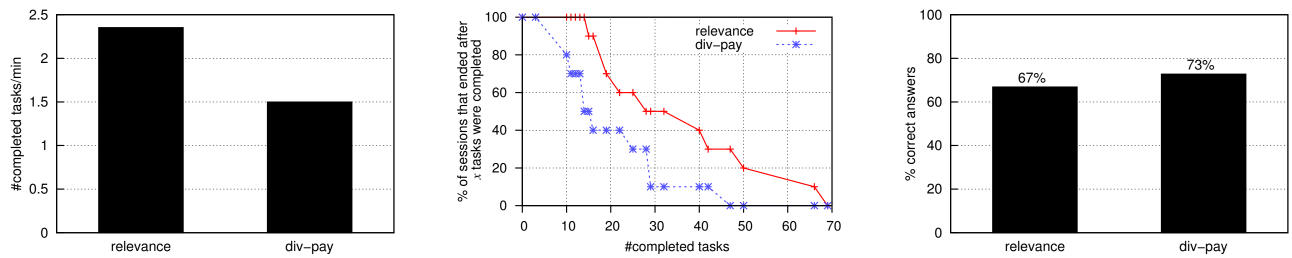


Figure 2: Adaptive task assignment and completion

4.2 Adaptivity

In practice, the evolving nature of human factors requires to re-think how they are integrated into crowdsourcing processes. The accuracy of such factors depends on the strategy used to acquire and refine them. Intuitively, observing workers for a longer period of time should result in more accurate skill values. On the contrary, workers' motivation is task- and context-dependent, i.e., how long a task takes, what other tasks are present, or who else is involved in case of collaborative tasks. In short, motivation is more ephemeral than skills and the length of time required to learn a worker's motivation should be "shorter" than the length of time required to learn a worker's skill. Moreover, while workers' skills increase monotonically as they complete more tasks, motivation varies with time.



(a) Task throughput (b) Worker retention (c) Outcome quality

Figure 3: Workers' performance for relevance-based and diversity/payment-based task assignments

In a recent effort [29], we proposed observing workers as they completed tasks and learning their motivation as a combination between task diversity (intrinsic motivation) and payment (extrinsic motivation). The premise of this work was that workers' motivation evolves over time and that task assignment could be improved from one session to the next by monitoring workers and capturing their motivation as they choose and complete tasks. We compared two strategies: relevance-based, that provides workers tasks matching their skills, and diversity/payment-based, that provides them tasks achieving a combination of diversity and payment. We found that when measuring task throughput, akin to the number of tasks completed per time unit, and worker retention, akin to the likelihood of workers completing many tasks, the relevance-based strategy was superior (see Figures 3a and 3b). However, as shown in Figure 3c, the diversity/payment strategy resulted in contributions of higher quality when compared to a ground-truth. As a result, we could draw two conclusions: workers were faster at picking and completing tasks when no context switching was required, i.e., in the case where tasks were relevant and similar to each other. However, they generated higher quality contributions for tasks that optimized their motivation, i.e., the observed balance between task diversity and payment. Additional experiments can be found in [29].

These preliminary results allow us to argue for the need for adaptive crowdsourcing processes that incorporate human factors as they evolve.

4.3 Evaluation, Deployment Strategies, and Repeatability

Evaluating the performance of a crowdsourcing platform is a major concern that poses a number of challenges. The variety of profiles workers have and the diversity of tasks made available on a crowdsourcing platform raise the need for a careful evaluation. So, what is being evaluated and what are the approaches used for that? Evaluation in crowdsourcing has mostly focused on measuring two kinds of indicators. Requester-centric indicators are task throughput, the number of tasks completed per time unit, worker retention, the likelihood of workers completing many tasks, and payment. Worker-centric indicators are motivation and satisfaction.

Table 2 summarizes the evaluation protocols used in crowdsourcing and the performance indicators that are measured. Evaluation is usually performed in an offline or an online manner. Offline evaluation relies on questionnaires deployed before task completion to measure expected workers' performance, or after task completion, to measure their satisfaction. Online evaluation on the other hand, relies on actually deploying tasks to workers and measuring requester- or worker-centric performance indicators.

Protocols	Performance indicators
Offline: use questionnaires to measure worker-centric indicators *Online:* observe workers during task completion and measure performance	Worker-Centric: *Worker Expected Throughput, Worker Satisfaction* Requester-Centric: *Worker Retention, Task Throughput, Task Payment, Task Quality*

Table 2: Evaluation protocols and performance indicators in crowdsourcing

Longer worker retention does not necessarily imply higher outcome quality. The same could be said about all other indicators (task throughput, payment, workers' psychological state). For collaborative tasks, while it has been shown that team size affects outcome quality in the case of collaborative editing, there is no study on how different group interaction models affect quality. In practice, quality is evaluated in one of two ways: against a known ground-truth as in Information Retrieval, or using crowdsourcing. Note that a golden standard is not always available or possible. For example, in the case of text creation tasks, there does not exist a ground-truth and one has to resort to text evaluation criteria such as word error rate, clarity and completeness [38]. A more general approach is to crowdsource quality evaluation by asking another set of workers to evaluate potential ground truth answers. In the case of text creation for example, the accuracy of a text translation and the quality of the output text could be evaluated using traditional independent and comparative approaches.

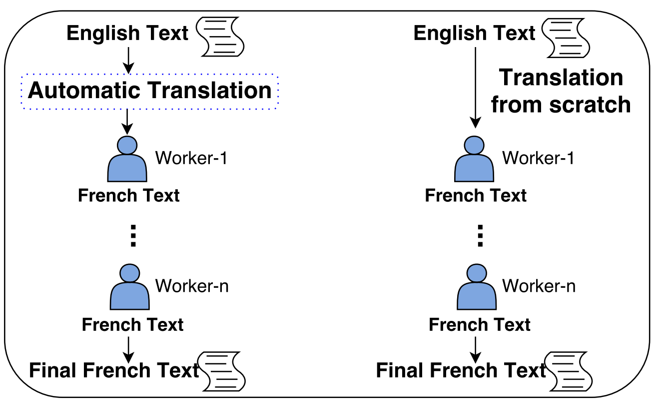


Figure 4: SEQ-IND-HYB & SEQ-IND-CRO

Much work is still needed in evaluating outcome quality for collaborative tasks. In Section 3.2, we discussed recent work that leverages group size and affinity between group members in the assignment of collaborative tasks to workers under quality and budget constraints [33]. In a recent piece of work [31], we also examined the usefulness of different skill aggregation functions in practice and validated them for a variety of tasks. A promising direction is the study of how different group dynamics and team interaction models [14] affect worker performance and outcome quality in crowdsourcing.

A deployment strategy defines the choices made to deploy a task. In worker-centric crowdsourcing, the question of how to organize the workforce becomes essential - thus deployment should become a center stage activity. A requester wishing to deploy a task makes a choice of how to combine algorithms and humans (crowd-only, or CRO vs hybrid, or HYB), a choice between sequential and simultaneous work structures (SEQ vs SIM), and a choice between an independent and a collaborative workforce organization (IND vs COL). In recent work [18, 30], we characterized different deployment strategies for text creation tasks such as translation and summarization. Figures 4 and 5 show some deployment options for English-to-French translation tasks. Our experiments measured outcome quality for different strategies and resulted in a set of guidelines to deploy

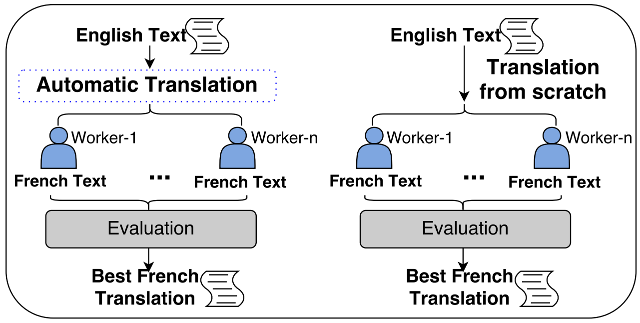


Figure 5: SIM-IND-HYB & SIM-IND-CRO

text translation and text summarization. For example, SEQ-IND-HYB (Figure 4), produces the best quality translation for long texts because workers prefer to start with an automatically translated text and are effective at improving each others' contributions in a sequential manner. The study of deployment strategies for other kinds of tasks and their impact on human factors, remains an open question.

Last but not least, *experimental repeatability* is a major concern in crowdsourcing and to the best of our knowledge, it has not received much attention. Repeatability in crowdsourcing is complex given the volatility of the crowd. Different times of day attract workers with different socio-demographics and in different time zones. Different platforms will also attract different workers. Different compensation schemes result in different behaviors [8]. Given this diversity, a pressing question is to define statistical significance in crowdsourcing. Some recent work [20,21] studied how to evaluate workers' quality with confidence interval - but using these proposed methods to obtain statistically significant output from the crowd remains an open question.

As a concluding remark, we envision that for the long term sustainability of crowdsourcing. In particular, we believe that deployment strategies should be made a primary focus of interest, followed by other worker-centric directions of research, such as collaboration, adaptability, and repeatability.

References

[1] Yael Amsterdamer and Tova Milo. Foundations of crowd data sourcing. In *SIGMOD Record*, 43(4):5–14, 2014.

[2] Sihem Amer-Yahia and Senjuti Basu Roy. Human Factors in Crowdsourcing. In *PVLDB*, Tutorial, 9(13):1615–1618, 2016.

[3] Anand Inasu Chittilappilly and Lei Chen and Sihem Amer-Yahia. A Survey of General-Purpose Crowdsourcing Techniques. In *IEEE Trans. Knowl. Data Eng.*, 28(9), pages 2246–2266, 2016.

[4] B. B. Bederson and A. J. Quinn. Web workers unite! addressing challenges of online laborers. In *CHI*, pages 97–106, 2011.

[5] L. Chen, D. Lee, and T. Milo. Data-driven crowdsourcing: Management, mining, and applications. In *31st IEEE International Conference on Data Engineering, ICDE 2015, Seoul, South Korea, April 13-17, 2015*, pages 1527–1529, 2015.

[6] Peng Dai and Mausam and Daniel S. Weld. Artificial Intelligence for Artificial Artificial Intelligence. In *AAAI*, 2011.

[7] P. Dai, J. M. Rzeszotarski, P. Paritosh, and E. H. Chi. And now for something completely different: Improving crowdsourcing workflows with micro-diversions. In *ACM CSCW*, pages 628–638, 2015.

[8] Y. Gao and A. G. Parameswaran. Finish them!: Pricing algorithms for human computation. *PVLDB*, 7(14):1965–1976, 2014.

[9] D. Geiger. Personalized task recommendation in crowdsourcing-Current state of the art. In *Decision Support Systems* 2014.

[10] B. Groz, T. Milo, and S. Roy. On the complexity of evaluating order queries with the crowd. *IEEE Data Eng. Bull.*, 38(3):44–58, 2015.

[11] S. Guo, A. G. Parameswaran, and H. Garcia-Molina. So who won?: dynamic max discovery with the crowd. In *SIGMOD Conference*, pages 385–396, 2012.

[12] Daniel Haas and Jiannan Wang and Eugene Wu and Michael Franklin. CLAMShell: Speeding up Crowds for Low-latency Data Labeling In *PVLDB* 2016.

[13] J. Hackman and G. R. Oldham. Motivation through the design of work: Test of a theory. *Organizational Behavior and Human Performance*, 16(22):250–279, 1976.

[14] G. Hertel and G. Hertel. Synergetic effects in working teams. *Journal of Managerial Psychology*, 2011.

[15] C. Ho and J. W. Vaughan. Online task assignment in crowdsourcing markets. In *AAAI*, 2012.

[16] C. Ho, S. Jabbari, and J. W. Vaughan. Adaptive task assignment for crowdsourced classification. In *ICML*, pages 534–542, 2013.

[17] Ho, Chien-Ju and Slivkins, Aleksandrs and Suri, Siddharth and Vaughan, Jennifer Wortman. Incentivizing High Quality Crowdwork. In *WWW*, 2015.

[18] K. Ikeda, A. Morishima, H. Rahman, S. B. Roy, S. Thirumuruganathan, S. Amer-Yahia, and G. Das. Collaborative crowdsourcing with crowd4u. *Proceedings of the VLDB Endowment*, 9(13), 2016.

[19] J. J. Horton and L. B. Chilton. The labor economics of paid crowdsourcing. In *ACM EC*, pages 209–218, 2010.

[20] M. Joglekar, H. Garcia-Molina, and A. Parameswaran. Evaluating the crowd with confidence. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '13, pages 686–694, New York, NY, USA, 2013. ACM.

[21] M. Joglekar, H. Garcia-Molina, and A. Parameswaran. Comprehensive and reliable crowd assessment algorithms. In *2015 IEEE 31st International Conference on Data Engineering*, pages 195–206. IEEE, 2015.

[22] R. Kenna and B. Berche. Managing research quality: critical mass and optimal academic research group size. In *IMA Journal of Management Mathematics*.

[23] D. Karger et. al. Budget-optimal task allocation for reliable crowdsourcing systems. In *Operations Research* 2014.

[24] N. Kaufmann, T. Schulze, and D. Veit. More than fun and money. worker motivation in crowdsourcing - A study on mechanical turk. In *AMCIS*, 2011.

[25] A. Marcus. et. al Human-powered sorts and joins. In *VLDB* 2011.

[26] Winter A. Mason and Duncan J. Watts. Financial incentives and the "performance of crowds". In *SIGKDD Exploration*, 11(2), pages 100–108, 2009.

[27] I. B. Myers and M. H. McCaulley. *Myers-Briggs Type Indicator: MBTI.* Consulting Psychologists Press, 1988.

[28] A. Parameshwaran et. Al CrowdScreen: Algorithms for filtering data with humans In *VLDB* 2010.

[29] Julien Pilourdault, Sihem Amer-Yahia, Dongwon Lee and Senjuti Basu Roy. Motivation-Aware Task Assignment in Crowdsourcing. In *Technical Report*, University of Grenoble-Alps, 2016.

[30] Ria Mae Borromeo, Maha Alsayasneh, Sihem Amer-Yahia and Vincent Leroy. Hybrid Deployment Strategies for Crowdsourced Text Creation Tasks. In *Technical Report*, University of Grenoble-Alps, 2016.

[31] Dalila Koulougli, Julien Pilourdault, Sihem Amer-Yahia, Allel Hadjali. Effect of Human Factors and Work Structure on Aggregating Workers Contributions in Crowdsourcing. In *Technical Report*, University of Grenoble-Alps, 2016.

[32] H. Rahman, S. Thirumuruganathan, S. B. Roy, S. Amer-Yahia, and G. Das. Worker skill estimation in team-based tasks. *Proceedings of the VLDB Endowment*, 8(11):1142–1153, 2015.

[33] H. Rahman, S. B. Roy, S. Thirumuruganathan, S. Amer-Yahia, and G. Das. Task assignment optimization in collaborative crowdsourcing. In *IEEE ICDM*, pages 949–954, 2015.

[34] V. Raykar, C. Vikas C. and Y. Shipeng. Eliminating Spammers and Ranking Annotators for Crowdsourced Labeling Tasks. In *J. Mach. Learn. Res.*, pages 491–518, 13(1), 2012.

[35] V. Raykar, S. Yu, L. Zhao, A. Jerebko, C. Florin, G. Valadez, L. Bogoni, and L. Moy. Supervised Learning from Multiple Experts: Whom to trust when everyone lies a bit. In *ICML*, pages 889–896. ACM, 2009.

[36] R. Rebeiro et al. Crowdsourcing subjective image quality evaluation *Int Conference on Image Processing 2011*.

[37] J. Rogstadius, V. Kostakos, A. Kittur, B. Smus, J. Laredo, and M. Vukovic. An assessment of intrinsic and extrinsic motivation on task performance in crowdsourcing markets. In *Proceedings of the Fifth International Conference on Weblogs and Social Media, Barcelona, Catalonia, Spain, July 17-21, 2011*, 2011.

[38] Senjuti Basu Roy and Ioanna Lykourentzou and Saravanan Thirumuruganathan and Sihem Amer-Yahia and Gautam Das. Task assignment optimization in knowledge-intensive crowdsourcing. In *VLDB J.*, 24(4): 467–491, 2015.

[39] J. M. Rzeszotarski, E. H. Chi, P. Paritosh, P. Dai. Inserting Micro-Breaks into Crowdsourcing Workflows. In *AAAI* 2013.

[40] Shaw, Aaron D. and Horton, John J. and Chen, Daniel L. Designing Incentives for Inexpert Human Raters. In *CSCW* 2012.

[41] Soberón, Guillermo and Aroyo, Lora and Welty, Chris and Inel, Oana and Lin, Hui and Overmeen, Manfred. Measuring Crowd Truth: Disagreement Metrics Combined with Worker Behavior Filters. In *CrowdSem Workshop* 2013.

[42] Beatrice Valeri and Shady Elbassuoni and Sihem Amer-Yahia. Crowdsourcing Reliable Ratings for Underexposed Items. In *WEBIST* 2016.

[43] J. Wang et al. Crowder: Crowdsourcing entity resolution In *VLDB* 2012.

[44] Whitehill, J., Ruvolo, P., Wu, T., Bergsma, J., and Movellan, J. R. Whose vote should count more: Optimal integration of labels from labelers of unknown expertise. In *NIPS*, pages 2035–2043, 2009.

[45] R. Yan, M. Gao, E. Pavlick, and C. Callison-Burch. Are two heads better than one? crowdsourced translation via a two-step collaboration of non-professional translators and editors.