Mass Collaborative Knowledge Processing on the Amazon Mechanical Turk

Chaolemen Borjigin[1,2]

[1] Key Laboratory of Data Engineering and Knowledge Engineering (Renmin University of China), MOE, Renmin University of China
Beijing, 100872, P.R. China

[1] Information Resource Management School, Renmin University of China
Beijing, 100872, P.R. China

Abstract

This paper aims to reveal the significant features of Amazon Mechanical Turk (MTurk) from a Mass Collaborative Knowledge Management perspective and to provide some recommendations for the improvement of that platform and the similar practices. This research mainly employs three types of research methodologies: Literature study was conducted to describe the R&D of MTurk and its applications; Case study was conducted for examining its knowledge processing features form a mass collaboration perspective; Software engineering methodology is applied to discuss the underlying principles and application interfaces of that platform. MTurk provided better practices for Mass Collaborative Knowledge Management in terms of harnessing long tail side human intelligence, discomposing knowledge tasks, keeping track of users' working history, providing qualification approaches, building a trusted third-party collaboration platform and introducing a pay-after-go payment approach. However, it has its limitations such as lacking a direct collaboration between the participants on a given task, machine unreadable knowledge presentation and failing to control the quality of knowledge processing. We can remedy those limitations by facilitating collaboration between the long tail-side workers, encouraging workers to submit extra data such as proof data, background information and relevant knowledge, introducing the Semantic Web into it and implementing man machine collaboration.

Keywords: *Mass Collaboration; Mass Collaborative Knowledge Management; Knowledge Engineering; Amazon Mechanical Turk.*

1. Introduction

We are entering into a new era of mass collaboration [1] and there are shifts in knowledge processing practices towards mass collaborative intervention in open knowledge ecosystems. Amazon Mechanical Turk, which is provided by Amazon.com, Inc., is one of the most typical pioneers of such practices. However, the theoretical research in Knowledge Management has not been kept up with these changes in practices. In order to reveal the underlying principles of those emerging practices and to provide theoretical guidelines for the similar practices, the author proposed a novel knowledge management paradigm called 'Mass Collaborative Knowledge Management' [2]. He has pointed out that the players in Mass Collaborative Knowledge Management can be classified into two groups: the professional actors and the professional amateur (Pro-Am) actors. The former mainly consists of internal employees of an organization and locates at the heads of the organizational knowledge chains. They may be professionals or knowledge agents such as software applications and services. The latter is comprised of human or machine located at the long tails, including customers, suppliers, volunteers and amateurs. Hence, a mass collaborative knowledge processing is interactions between those two kinds of participants. A typical mass collaborative knowledge processing includes the following steps (Figure1):

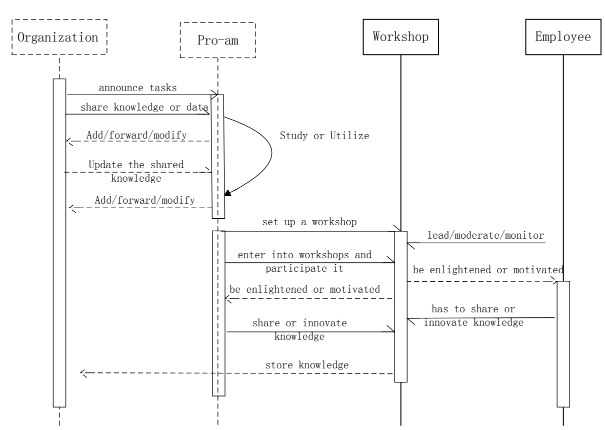

Fig. 1 the main stages of MCKM.

(1) The organization announces knowledge processing tasks to the Pro-Ams who are located at the long tails of its knowledge chains.
(2) The organization shares its some internal knowledge or data, which may be used in dealing with the tasks. Then, the Pro-Ams can study, utilize, comment, forward or modify them. Pro-Ams will be encouraged or facilitated by the employees or professionals who are located at the short head of the knowledge chains and the interaction logs will be created and stored into the organizational knowledge base for the purpose of accumulating more knowledge.
(3) The organization updates the shared knowledge or data without delay when they are modified or remarked. If there are no changes in the content, the organization will also have to update the metadata, such as the clicked rate, the remarked rank and the forwarded frequency. Mass Collaborative Knowledge Management regards metadata as an important component of organizational knowledge base and Semantic Web technologies are supposed to formalize the meanings of it[2].
(4) The organization encourages the Pro-Ams to attend workshops and to participate in its mass collaborative knowledge interventions. The same Pro-Am is allowed to enter into more than two workshops at the same time. Each workshop elects its chair to lead the discussions in it. The designing, running and adjourning of a workshop may be completed by Pro-Ams and this enables mass collaborative knowledge processing to be self-organized.
(5) The Pro-Ams are enlightened or motivated by other participants, including the professionals and other Pro-Ams, and encouraged to share their own knowledge or data. This knowledge or data will be automatically stored in the organizational knowledge base via the Semantic Web technologies.
(6) The employees who are located at the head of organizational knowledge chains can take part in a mass collaborative knowledge processing, and have the responsibility to lead, moderate or facilitate it as well.
(7) The Pro-Ams gets some extra knowledge or data from mass collaborative knowledge intervention. Further, they will become alternative human resources for the organization when they have accumulated as much knowledge as its internal staffs. This will force the internal staffs to share their knowledge and to participate in knowledge intervention positively. Finally, there will be significant changes in the organizational Knowledge Management and the cost and risk will decrease because it will be able to find or foster a candidate professional or employee at the long tails of its knowledge chains. As a result, the motivations of employees to take part in organizational knowledge interventions will become diverse. Organizational regular job appraisals, substitution threat from the long tails and personal interests push the employees to play an active role in organizational knowledge interventions. This is one of the most essential steps towards building organizational knowledge ecosystem.
(8)A new knowledge ecosystem will be established in the organization, and the cost and risk of its knowledge management will be decreased remarkably. Naturally, the bottlenecks of current knowledge management, for instance, challenges to convert personnel knowledge into that of the organization will be addressed.
The rest of this paper is organized as follows. Section 2 describes Amazon Mechanical Turk (MTurk) and its knowledge processing features. Section 3 further discusses the main stages and the underlying technologies of mass collaborative knowledge processing on it. Then, section 4 introduces typical applications of the mass collaborative knowledge processing on MTurk. Section 5 further discussed the strengths and weaknesses of the platform in line with MCKM paradigm. Finally, Section 6 provides some recommendations to remedy the weaknesses of the platform from a mass collaboration perspective.

2. MTurk and its knowledge processing features

MTurk is a knowledge processing platform developed by Amazon.com, Inc in 2006. However, its name stems from a chess-playing "robot" called Mechanical Turk that was designed by Wolfgang von Kempelen, a Hungarian inventor in the late 18th century [3]. The "robot" was toured Europe as an automatic playing machine, but was revealed later that it was controlled by a chess master hidden in a special compartment. Amazon named its new product after Mechanical Turk in that both of them comply with the same design principle: human-machine cooperative knowledge processing.
Figure2 shows the main page of MTurk[4].It provides a collaborative working platform between workers (also known as Turkers) and requesters of the same knowledge processing task. Contrary to traditional knowledge processing platforms, the tasks on MTurk are carried out by humans instead of machines. The requester breaks his or her knowledge processing tasks into HITs (Human Intelligence Tasks), which are more specific and easier to be completed, and published them via the platform. And then, workers pick up some of the HITs according to their own willingness and MTurk automatically creates an assignment for the purpose of keeping trace of the accepted HITs. After the workers have been conducted the tasks described in the assignment, they will upload their results

to the platform, and it will further reorganize the submitted results into a bigger result. As the requester accepts the result in an assignment, the worker who submitted it will get some money in return. In addition, the platform also provides APIs and Web Services so that developers can build their own application on it.

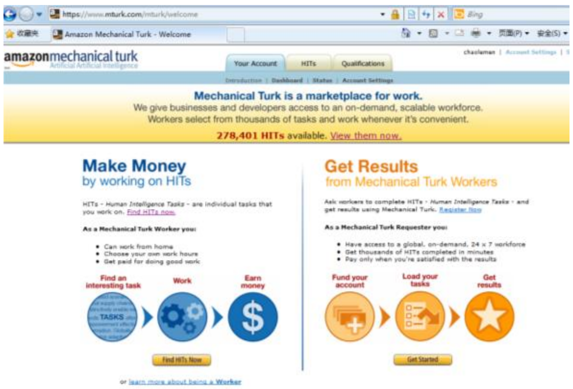

Fig. 2 The Amazon Mechanical Turk.

In comparison with knowledge processing on traditional platforms, that on MTurk has the following features:

2.1 Long tail-side workforce

An organization can lengthen its knowledge chains by announcing its knowledge processing tasks through MTurk. The Professional –Amateurs (Pro-ams) who come from all over the world can be able to find those published tasks and be encouraged to take participate in it. Therefore, most workers on MTurk are always located at the long tails of organizational knowledge processing chains instead of its small heads that mainly consist of its internal employees or professionals.

2.2 High scalable human resource market

Traditionally, human source management always need a fixed routine for getting a new worker for knowledge processing in an organization, including steps such as publishing recruitment announcement, selecting candidates, conducting an interview, providing pre-job training, and assigning tasks for the worker. As a result, the labor market in traditional knowledge processing paradigms lacks scalability. MTurk, however, breaks through these conventional beliefs and follows an alternative pattern that coordinates one-time, task- oriented and loosely coupled workers. That is to say, the requesters can find on-demand and scalable workforces via MTurk.

2.3 Task breakdown structure

In order to make it easier for the workers to complete knowledge processing tasks published via MTurk, the requesters are supposed to decompose their bigger tasks into small tasks called as HITs (Human Intelligence Tasks). Although we can also find some other kinds of crowd sourcing like Wikipedia, the tasks published on them are too big to be completed or to be monitored. HITs are appropriate for long tail- side workforce in that those are convenient to be completed with a little effort.

2.4 Pay after delivery pattern

MTurk abides by pay after delivery pattern, and the requesters are allowed to pay only for approved work. This pattern not only avoids the workers submitting low quality results but also eliminates the requester's worries about the high cost of knowledge processing. MTurk further makes it possible for the requesters to define qualifications for candidate workers, and only the desired workers are allowed to participate in the knowledge processing tasks. The qualification can be defined by the candidate workers' geo-location, their work history, and predefined qualification tests as well.

2.5 Cost effective practices

MTurk provides its users with a low-cost knowledge processing market in that it can cut the costs for recruiting and training rookies. The requester can directly find a competent worker via the platform. Further, it can also save operating costs since the collaborative processing isn't completed in physical environments. In addition, the average wage that the requester pays for workers who have submitted a satisfactory result is lower than that for a similar practice in the current labor market. Therefore, where the number of HITs corresponding to a knowledge processing task is not too large, we will be able to complete it via MTurk with less cost than traditional labor markets.

3. Mass collaborative knowledge processing on MTurk

3.1Main processes

A typical workflow for knowledge processing on MTurk is in line with Mass Collaborative Knowledge Processing discussed in the Section 1 and includes the following stages (Figure 3):

(1) The requester initiates a new knowledge processing intervention by publishing their knowledge processing tasks. The requester is supposed to identify the main purpose of his or her knowledge processing task as a starting point and then set the classifications of it. This not only helps the requester to choose an appropriate user interface for the task, but also makes it easier for workers to retrieve or understand it. Where the task is structured in a manner that allows workers to complete it efficiently, they are able to submit its results at higher velocities and higher quality [6]. MTurk provides the requesters with some basic task types, such as categorization, research, writing, survey, tagging, moderation, transcription, translation. The requesters are able to not only use one of these pre-defined types to publish their tasks for the purpose of automatically displaying them, but also to specify their own task types to have more control over their user interfaces.

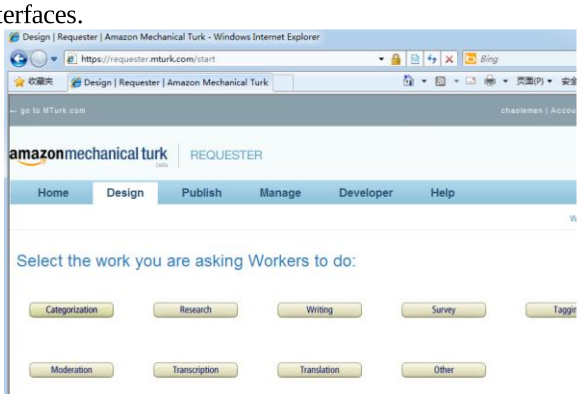

Fig. 3 the common task types provided by MTurk.

(2) The requester breaks his or her knowledge processing task into HITs. After identifying the type of the knowledge processing task, the requester is supposed to decompose it into smaller sub tasks so that the workers complete them more easily and quickly. Breaking the task into HITs also makes it possible to complete it in parallel. A HIT template commonly consists of title, description, keywords, lifetime, detailed instructions, assignment duration, assignment number, answers format, worker's qualifications, rewards for each satisfactory assignments.

(3) The requester publishes his or her HITs. The requester can upload his or her HITs into MTurk templates and make it available for the workers to retrieve them. The published HITs may be assigned to any workers or some restricted users. It is notable that the requester can use CSV (Comma Separated Values) format to define the key and value of HITs. In addition, MTurk also provides a preview function for the requesters, and they can justify the HITs prior to publishing it.

(4) Workers find the published HITs and accept some or all of them. MTurk provides the workers with a search system and recommended lists so that they can find desired HITs easily. Workers can access more details of a HIT and make a decision about whether or not to accept it, as they find a candidate HIT. Where requesters set specific qualifications for their HITs, the system can automatically filter out unexpected workers. Geo-location, work history and qualification test have been most commonly used for qualifications on MTurk.

(5) MTurk creates an assignment for the accepted HIT and traces it until the worker submits his or her result or the result is automatically submitted. After accepting a HIT, the worker starts to accomplish the task within the assignment. Accepted workers are supposed to complete the assignment before the deadline. If the assignment duration expired, the assignments would be automatically submitted by the system.

(6) The requester checks the results submitted by the workers and makes a decision to approve or reject them. If the result is approved by the requester, the reward would be transferred from the requester's Amazon.com account to that of the workers automatically. The requesters are allowed to refuse to pay for the rejected work. In addition, the requesters are also encouraged to give their excellent workers some extra bonus for their outstanding contributions.

(7) The requester terminates the task and further utilizes the accepted results. If the requester believes that a satisfactory answer has been found, he or she can end the intervention and prevent any other workers from accepting it. Otherwise, the requester can put back the HITs and republish them to other workers. Finally, the requester may also export these result sets and further conduct deep insights into them by utilizing third-party applications such as Microsoft Excel or SPSS.

3.2 Underlying technologies

Amazon Mechanical Turk (MTurk) has implemented some of the key technologies for mass collaborative knowledge processing. These new technologies involve management of HITs, assignments, qualifications, results, workers and requesters. Programming interfaces are also indispensable for the advanced users to develop their own knowledge processing application on it.

(1)HIT management. MTurk allows the requesters to create, track, abandon, end and review their HITs. The requesters can specify the title, description, keyword, reword, life span, the number of workers, and qualification requirements, as they create a HIT. Once the HIT accepted by a worker, the platform will create an assignment to track the process of being completed. When all of the assignments of a HIT are submitted or their life time is expired, the system will forward the results to the requester who submitted the HIT.

(2)Assignment management. Once a HIT accepted by a worker, an assignment is automatically created by MTurk in order to track the completion of it. A HIT may have multiple assignments allowing for collaborative working. By default, each HIT has ten assignments. The platform provides the requester with an interface to specify the maximum number of assignments corresponding to his or her HITs. Where the requesters call the web services provided by the platform, for instance, they can change the max number by specifying the *MaxAssignements* prosperity. If the worker uploads his or her results, rejects the assignment or fails to complete it before its deadline, then MTurk will terminate the assignment. If an assignment is rejected by the workers, the platform will once again make it available for other workers to accept it.

(3)Qualification management. Qualification Type (QT) is one of the most commonly used technologies for quality assurance of mass collaborative knowledge processing on MTurk. The platform provides some default QT templates in order to help the requesters to define their specific qualification requirements. The platform also allows its requesters to define their own QT template. The requesters can easily set a QT for their HITs via those pre-defined templates. QT commonly comprises the properties such as ID or name, subject, duration, grade and status of a qualification. The requester is supposed to allocate QTs for their HITs in advance of publishing their knowledge processing tasks and candidate workers who fails to meet the qualification will be rejected automatically. Three of the most commonly used QTs are geo-location, work history of the candidates and qualifications tests.

(4)Result management. Result management on MTurk involves handling the results submitted by workers, such as checking the status of results, filtering the unexpected ones, and evaluating them. Evaluation strategies are crucial to result management for mass collaboration. Chen J. J., Menezes N. J., and Bradley A. D. (2011) proposed that the requesters could evaluate the quality of results through strategies such as single assignment, forced agreement, plurality, expert review and known answer question. These strategies can be used independently or in combination.

(5) Worker management. Worker management on MTurk involves the analysis of demographics, motivation and credibility of the users. Ipeirotis P. G.[11] found that the majority of the users are from the United States and India, which comprise 46.8% and 34.0% of the total users respectively. The motivations of them, however, are different: most of the Indian users choose it as their main source of income, but the American users participate in it for diverse purposes such as killing time, or supplementary income. Downs J. S. found that professionals, students, and non-workers seem to be the most likely to take the task seriously[7]. He also argued that man over 30 and woman at any age are much more likely to be qualified for these activities.

(6)User Interfaces. MTurk provides three kinds of user interfaces for mass collaborative knowledge processing on it: webpage, Application Programming Interface (API), and Command Line Tool (CLT). Web pages are human oriented and facilitate the human users to use it, but its usability is commonly limited. API provides more versatile interfaces than web pages in that the users can develop their own mass collaborative knowledge processing applications on it. CLT gives the user with more parameters than webpage, but its user experiences are not as good as webpage for ordinary users. Besides, the platform also provides Amazon Web Services (AWS) for its users and the programmers can interact with the AWS through SOAP protocol. When the consumer sends a request message to AWS server in order to call the expected services on it, the AWS server accepts and processes it and returns an XML data that contains the results. Services provided by AWS server are described in WDSL and supports two kinds of programming pattern: SOAP and REST.

4. Applications of mass collaborative knowledge management on MTurk

Amazon Mechanical Turk (MTurk) provides two kinds of access methods for its users: graphic user interface and programming interface. The former usually serves for the workers in that their application requirement is simple. A user-friendly interface is an indispensable tool for them to participate in the mass collaborative knowledge processing. User graphic interface is enough for most workers, but the requesters sometimes needs more versatile services. Hence, the requester tends to use programming interface in preference of graphic user interface. Programming interface not only enables the users to design their own graphic interface, but also makes it possible for requesters to develop their own applications.

4.1 Graphic user interface

Graphic User Interface (GUI) of the MTurk is widely adopted by the users for the purpose of collecting, cleaning, transcribing, validating, categorizing, or annotating data. Goodman, J. K., Cryder, C. and Cheema, A. [10] discussed the strengths and weakness of data collection via MTurk.

Researchers[8] proposed how to use the platform to transcribe or annotate spoken language. Winter M. and Siddharth S.[14], Paolacci, G., Chandler, J. and Ipeirotis, Panagiotis G.[15], Rand, D. G.[16] described how Amazon's Mechanical Turk can help researchers conduct behavioral experiments. Ipeirotis P. G.[12] collected a data set of MTurk marketplace from January 2009 to April 2010 and proposed its main features and directions for future research. Rand, D. G. [17] discussed how MTurk is revolutionizing innovation and discovery in the social sciences. Besides, AOL, CastingWords and SnapMyLife use it to categorize content, to transcribe audio files and to annotate photos, respectively [5].

4.2 Programming interfaces

MTurk provides two types of programming interfaces: API (Application Programming Interfaces) and WS (Web Services). Software engineers can develop their own application by calling those API or WS. These programming interfaces provide the following functions:
(1) HITs management such as *CreateHIT( ),ChangeHITTypeOfHIT( ), DisableHIT( ), DisposeHIT( ), ExtendHIT( ), ForceExpireHIT( ), GetHIT( ), GetReviewableHITs( ), GetReviewResultsForHIT( ), RegisterHITType( ), and SearchHITs( ).*
(2) Assignment management including *ApproveAssignment( ), ApproveRejectedAssignment( ), GetAssignment( ), GetAssignmentsForHIT ( ), and RejectAssignment().*
(3) Qualification management such as *AssignQualification ( ), CreateQualificationType( ), Dispose QualificationType( ), GetHITsForQualificationType( ), GetQualificationsForQualificationType ( ), GetQualificationRequests( ), GetQualificationScore( ), GetQualificationType( ), Reject Qualification Request ( ), RevokeQualification( ), SearchQualificationTypes( ), UpdateQualificationScore ( ), and UpdateQualificationType().*
(4) Results management including *GetRequesterStatistic( ), GetRequesterWorkerStatistic( ), GetBonus Payments( ), GrantBonus( ), GrantQualification().*
(5) Worker management such as *BlockWorker( ), GetAccountBalance( ), GetBlockedWorkers ( ), NotifyWorkers( ), and UnblockWorker().*
Programmers can access these operations via SOAP and REST and develop their own mass collaborative knowledge processing applications. For instance, Feedback Army developed an application for reviewing the usability of websites based upon those functions provided by MTurk [9].

5. Discussions

As one of the pioneers of Mass Collaborative Knowledge Management, MTurk has dealt with some acute challenges in these emerging practices. However, there are also some weaknesses in implementing mass collaboration on it.

5.1 Strengths of mass collaborative knowledge processing on MTurk

According to Mass Collaborative Knowledge Management, Amazon Mechanical Turk (MTurk) has the following advantages over traditional knowledge processing platforms:
(1) Harnessing the intelligence of the humans, especially who located at the long tail of organizational knowledge chains, to complete knowledge processing tasks. How to take advantages of human intelligence is one of the main concerns of Mass Collaborative Knowledge Management. Traditional knowledge management theories, including First Generation Knowledge Management (FGKM) and Second Generation Knowledge Management (SGKM), fail to exploit long-tail side human intelligence. FGKM tends to overestimate the intelligence of machines, which heavily depends on Artificial Intelligence. However, there are many bottlenecks to do breakthroughs in the research of its theoretical foundations such as Psychology, Physiology and Brain Science. As a result, some researchers proposed a new paradigm for knowledge management that is called as Second Generation Knowledge Management (SGKM), and the main concerns shift from machines to humans. One of the main limitations is SGKM also neglects the contributions from the long tails. Mass Collaborative Knowledge Management identifies these limitations and embraces the long tail side workers into organizational knowledge management systems. MTurk abides by this new paradigm and provides its user with a scalable human workforce.
(2) Discomposing a knowledge processing task into many of Human Intelligence Tasks and allowing the same HIT to be completed by different workers. MTurk breaks a knowledge processing task into single, self-contained, and easier to be completed small tasks, which are called as Human Intelligence Tasks (HITs). The requester can predefine the required qualifications of the candidate worker who may be allowed to accept them. Once a qualified worker accepts a HIT, MTurk creates an assignment to track his or her work to its completion. The same HIT may correspond to more than one assignment so that different workers can complete the same HIT. Once the platform gets results, which are submitted by different participants but correspond to the same HIT, it will make a

comparison between them and choose some or all of them as the recommended result(s).

(3) Keeping track of each worker's working history and providing qualification approaches to facilitate the requesters to find expected workers. To qualify candidate workers and to filter out the unexpected one is two of the main challenges which are faced by Mass Collaborative Knowledge Management. The requesters of MTurk can choose one or both of the following methods to exclude unexpected workers from accepting their HITs. One is directly qualifying them by the statistics of their work history, and rejecting the ones who have bad behavioral records or poor performances. The other is choosing them by qualification tests, which are specially designed for a specific knowledge processing task and assigned a threshold score. Each of candidate workers should answer the qualification tests prior to accepting the task successfully, and only the candidates who pass the qualification tests are allowed to continue to complete the task.

(4) Offering a trusted third-party collaboration platform and introducing a pay-after-go payment approach. User's account management and payment mechanism are essential for Mass Collaborative Knowledge Management since the participants need a trusted collaborating platform. The account information about each user on MTurk is managed by Amazon.com, and users' accounts hold their money. Besides, the platform encourages its users by introducing a pay-after-go payment approach so that the requesters are allowed to pay only for the satisfying results. This payment approach also enhances the quality of results submitted by workers.

(5) Providing multiple user interfaces and facilitating the access of different applications. MTurk offers three types of user interfaces: Web Page, Application Programming Interface (API) and Command Line Tool (CLT). Each of them is suitable for their unique application scenarios. Web Page is user friendly so that it can be used by workers to find desired HITs and submit their results. API is programmable, and the requesters can choose it to develop their own applications on it. CLT is powerful in that it allows the users to deal with a small number of HITs easily.

5.2 Weaknesses of mass collaborative knowledge processing on MTurk

Although Amazon Mechanical Turk (MTurk) has five advantages discussed above, it has some disadvantages to date:

(1)There is no direct collaboration between long tails while they carry out the same knowledge processing tasks. The workers are isolated from each other by MTurk in that it doesn't provide interacting services for them. MTurk provides a forum that the workers (or requesters) can discuss some general topics such as general skills for answering (or publishing) questions. However, the forum is not appropriate for interacting on a given HIT. Therefore, it is difficult for the participants, including workers and requesters, to collaborative working on a given HIT. When a worker accepts a HIT, the platform creates an assignment that belongs exclusively to the worker and rejects other workers to participate in it. This rejects collaboration between different workers on the same assignment.

(2) There are insufficient interactions between the head and the long tails. MTurk can forward the message from the requester (the head) to the E-mail address of the workers who are located at the long tails. However, this simple interaction seldom occurs, and its contents are always outside the scope of a given HIT. In Mass Collaborative Knowledge Management, the requesters are supposed to provide more details about the HITs such as context information or domain knowledge. The provision of this extra information can not only enlighten more workers to participate in the knowledge processing tasks, but also enhance the quality of the knowledge processing.

(3) Knowledge representation is just human-oriented, and machines are unable to understand the semantic of knowledge on MTurk. To date, the intelligence of MTurk comes from human intelligence, and its knowledge representation is only human-oriented. For instance, the requester should design the layout and format of their HITs prior to publishing them in order to make them user-friendlily. However, the platform fails to value the machine readability of knowledge representation. This leads to the difficulty of processing knowledge automatically by the platform.

(4) It is difficult for MTurk to control the quality of knowledge processing. The quality of knowledge processing on MTurk heavily depends upon the consciousness of its participants. Because of lacking evaluation mechanisms, context information, and proof data, the quality of knowledge processing on it is almost out of control.

6. Conclusions

Mass Collaborative Knowledge Management (MCKM) is a new direction in knowledge management studies, and it provides a novel theoretical foundation for building an organizational open ecosystem. As one of the pioneers, MTurk provided a better practice to aid in implementing MCKM in terms of harnessing long tail side human intelligence, discomposing knowledge tasks, keeping track of users' working history, providing qualification approaches, building a trusted third-party collaboration platform and introducing a pay-after-go payment approach. Further, according to Mass Collaborative Knowledge Management paradigm discussed in Section 1, we can improve MTurk by:

(1) Facilitating interactions between long tails. MTurk should introduce a new mechanism to facilitate the interactions between different workers while they complete the same HIT. This interaction mechanism should be able to calculate the contribution of each participant according to his or her work history and the requester's feedback as well.

(2) Asking the requesters to provide context information and basic knowledge of a HIT as they publish it. In order to reach a shared understanding between works and requesters on the same HIT, requesters should provide more details about related information and basic knowledge that may be helpful for the workers to understand and complete it correctly. Further, MTurk should encourage workers to add new context information or domain knowledge and enable them to evaluate such extra information provided by others.

(3) Encouraging workers to submit the proof data when submitting the result of a knowledge task. The workers should provide the proof data that are adopted for calculating the results so that the requesters can assess the reliability of results. At the same time, the platform should also allow other workers to utilize, modify or assess the proof data and provide a mechanism to encourage workers to participate in providing, amendment and assessment of the proof data.

(4) Introducing the Semantic Web to formalize its knowledge representation. MTurk should introduce the Semantic Web technologies to improve the machine readability of knowledge representation in order to deal with the difficulties of processing knowledge by machines and make full use of the current AI technologies. Searching HITs, for instance, is conducted by matching keyword literally and can be enhanced by introducing the Semantic Web technologies to match the semantics of keywords and the worker's needs.

(5) Enabling workers to collaborate on the same HIT. If knowledge processing tasks on MTurk are too complex to break into smaller HITs, the platform should allow different workers to collaborate on it. The platform can keep bigger HITs by inviting more workers to complete them collaboratively.

(6) Implementing man-machine collaboration to take advantage of their complementary advantages. Since man and machine is complementary in knowledge processing [18], the platform should make full use of those complementary advantages. MTurk can introduce the Semantic Web to the platform in order to implement collaborations between men and machines. Man-machine collaboration can remedy the weakness in processing knowledge by one of human or machine.

Acknowledgements

This work was funded by National Natural Science Foundation of China (No. 71103020) , Scientific Research Fund of Renmin University of China (No:14XNQ022) , and National Natural Science Foundation of China (Key Program No. 71133006).

References

[1]. Tapscott D, Williams A. D.Wikinomics: how mass collaboration changes everything ,Penguin Group (USA) Inc, New York, NY , 2006.

[2]. Chao L. "Man-machine Mass Collaborative Knowledge Management Model". Journal of Library Science in China,Vol.37, No.5, 2011,pp.101-114.

[3]. Amazon.com, Inc. "Amazon Mechanical Turk FAQs", available at: http://aws.amazon.com /mturk/faqs/(accessed 15 May 2013).

[4]. Amazon.com, Inc. (2013b), "Amazon Mechanical Turk", available at: http: // aws.amazon. com/ mturk/(accessed 15 May 2013).

[5]. Amazon.com, Inc. " Case Studies ", available at: https://requester.mturk.com/ case_studies(accessed 15 May 2013).

[6]. Chen J. J., Menezes N. J., and Bradley A. D."Opportunities for Crowdsourcing Research on Amazon Mechanical Turk", available at: http://www. crowdresearch.org/chi2011-workshop/ papers/ chen-jenny.pdf(accessed 15 May 2013)..

[7]. Downs J. S., Holbrook M. B., Sheng S., Cranor L. F. "Are Your Participants Gaming the System? Screening Mechanical Turk Workers", in CHI 2010:Proceedings of the 28th international conference on Human factors in computing systems, Atlanta, Georgia, USA.2010, pp.2399-2402.

[8]. Evanini K., Higgins D., Zechner K. "Using Amazon Mechanical Turk for transcription of non-native speech", in CSLDAMT '10: Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk.Stroudsburg, PA, USA ,2010,pp. 53-56.

[9]. FeedbackArmy.com. "Using Feedback Army", available at: http://www.feedbackarmy. com/questions.slp (accessed 12 November 2013).

[10]. Goodman, J. K., Cryder, C. and C., A. "Data Collection in a Flat World: Strengths and Weaknesses of Mechanical Turk Samples", available at: http://ssrn.com/abstract=2016308 (accessed 15 May 2013).

[11]. Ipeirotis, Panagiotis G. "Demographics of Mechanical Turk. ", available at:http://ssrn. com/abstract=1585030(accessed 15 May 2013).
[12]. Ipeirotis, Panagiotis G. "Analyzing the Amazon Mechanical Turk Marketplace .XRDS: Crossroads", available at: http://ssrn.com/abstract=1688194(accessed 15 May 2013).
[13]. Marge M., Banerjee S., Rudnicky A. I. "Using the Amazon Mechanical Turk to transcribe and annotate meeting speech for extractive summarization" in CSLDAMT '10: Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk ,Stroudsburg, PA, USA, 2010,pp.99-107.
[14]. Winter M. and Siddharth S. "Conducting Behavioral Research on Amazon's Mechanical Turk ", available at: http://ssrn.com/abstract=1691163(accessed 15 May 2013).
[15]. Paolacci, G., Chandler, J. and Ipeirotis, Panagiotis G. "Running Experiments on Amazon Mechanical Turk". Judgment and Decision Making, Vol. 5, No. 5, 2010,pp.411-419.
[16]. Rand. D. G. " The promise of Mechanical Turk: How online labor markets can help theorists run behavioral experiments ", Journal of Theoretical Biology, Vol. 299,No. 21,2012,pp. 172-179.
[17]. Rand, D. G. "How Online Labor Markets are Revolutionizing Innovation and Discovery in the Social Sciences (May 25, 2011). Gruter Institute Squaw Valley Conference – Innovation and Economic Growth, 2011 ", available at: http://ssrn.com/abstract=1851668.
[18]. Chao L. "Human-Machine Cooperation Knowledge Processing on Semantic Web". Library and Information Service.Vol.53, No.24, 2009, pp.115-119.

Chaolemen Borjigin Chaolemen Borjigen received his Master and PhD degrees from Peking University and Renmin University of China in 2006 and 2010 respectively. From 2010 to 2012, he did his Postdoc at Tsinghua Universtiy, Beijing, China. Currently, he is an assistant Professor at Key Laboratory of Data Engineering and Knowledge Engineering, Renmin University of China, Beijing, China. His main research interests include Semantic Web, knowledge engineering and information studies.