Article can be accessed online at http://www.publishingindia.com

Crowdsourcing: A Survey of Applications

Jayshri Namdeorao Ganthade*, Sunil R. Gupta**

Abstract

Crowdsourcing, itself a multidisciplinary field, can be well-served by incorporating theories and methods from affective computing. We present a various applications which are based on crowdsourcing. The direction of research on principles and methods can enable to solve a general problem via human computation systems. Crowdsourcing is nothing but an act of outsourcing tasks to a large group of people through an open request via the Internet. It has become popular among social scientists as a source to recruit research participants from the general public for studies. Crowdsourcing is introduced as the new online distributed problem solving model in which networked people collaborate to complete a task and produce the result. However, the idea of crowdsourcing is not new, and can be traced back to Charles Darwin. Darwin was interested in studying the universality of facial expressions in conveying emotions. For this, it required large amount of database and for this he had to consider a global population to get more general conclusions.

This paper provides an introduction to crowdsourcing, guidelines for using crowdsourcing, and its applications in various fields. Finally, this article proposes conclusion which is based upon applications of crowdsourcing.

Keywords: Crowdsourcing, Mechanical Turk, Human Intelligence Tasks, General Computation, Human Program Synthesis, Emergency Management, Volunteered Geographic Information, Crowdsourcing, Web 2.0, Wildfire, Santa Barbara.

1. Introduction

As per the Human Participant Research Guidelines by the University of Waterloo, online labour markets, known as crowdsourcing, are becoming popular mechanisms for recruiting potential research participants. Crowdsourcing is an act of outsourcing tasks to a large group of people (we are using the wisdom of "crowd") through an open request via the Internet in exchange for monetary remuneration (Human Participant Research Guidelines, 2012) or simply you can take the opinion of crowd.

Researchers studying the use of crowdsourcing have found that individuals who sign up to complete tasks tend to use the monetary remuneration they receive as a supplementary source of income. Some websites report the vast majority of individuals state they use the site casually as a way to spend time exchange or share their thoughts.

These guidelines have been created to assist University of Waterloo researchers when planning online studies using a crowdsourcing service. These guidelines are useful for detail information about the most common researchers may face problems such as: contractual obligations required by the crowdsourcing service, recruitment and information-consent, risks, use of deception or partial disclosure, inclusion/exclusion criteria, remuneration, privacy and confidentiality, details concerning how to withdraw from the study, appropriately estimating the time duration for the study, and contacting participants (Human Participant Research Guidelines, 2012; Policy statement of Canadian Institute of Health Research, 2010).

* Lecturer, Amrutvahini College of Engineering, Computer Science and Engineering Department, Sangamner, Ahmednagar, Maharashtra, India. Email: jayshri_ganthade@rediffmail.com

** Assistant Professor, Department of Computer Science and Engineering, PRMIT&R, Badnera, Amravati, Maharashtra, India. Email: srg_99@rediffmail.com

Crowdsourcing

The most common crowdsourcing service used by University of Waterloo researchers is a service based in the USA called Amazon Mechanical Turk. Researchers who are not citizens of the USA are not able to post a study directly to Amazon Mechanical Turk therefore many work through an intermediary service called Crowdflower. These two services are described in detail next (Attari et al., 2010).

Amazon Mechanical Turk

This service started in 2005 mainly as a service to "crowdsource" tasks that require human intelligence to complete (e.g., video or audio transcription). Groups, businesses, or individuals who use the crowdsourcing service Amazon Mechanical Turk are known as "requesters" and they post tasks to be completed as HITs (Human Intelligence Tasks). "Workers" can browse among the posted HITs and complete them for monetary remuneration set by the "requester". "Workers" can be located anywhere in the world, but about 80% of "workers" using Amazon Mechanical Turk reside in the USA or India (Berinsky et al., 2010; Buhrmester et al., 2011; Policy statement of Canadian Institute of Health Research, 2010)

After completing tasks the remuneration provided to the workers can be redeemed by them on Amazon.com via a gift certificate or be transferred to a bank account. Some University of Waterloo researchers have reported that the vast majority of Amazon Mechanical Turk participants complete surveys and tasks conscientiously and completely; typically about 90%. However, researchers should keep in mind that online participants often have shorter periods of time and attention in which to complete a study in comparison to an in-lab participant and can be less tolerant to complete open-ended responses or complicated tasks (Berinsky et al., 2010; Buhrmester et al., 2011).

Crowdflower

Crowdflower, formerly known as Dolores Labs, located in San Franciso, California, was founded to create tools to manage internet crowdsourcing. Crowdflower works with "requesters" i.e. the user to take large, heavy projects which require heavy data and break them into small tasks that are then distributed to various on-demand workforces around the world. Out of these, one of the workforces is Amazon Mechanical Turk. If needed, the Crowdflower system will aggregate the results which are collected from various workforces and control for quality of the work for the "requester" (Human Participant Research Guidelines, 2012).

The USA based company, Amazon Mechanical Turk, requires "requesters" or users to provide a USA billing address and a USA credit card, Amazon Payments account or USA bank account to post a HIT.

Therefore, it is not possible for "requesters" (i.e. users/ researchers) from outside of the USA to post a HIT directly on Amazon Mechanical Turk. Therefore, the University of Waterloo researchers including non-USA researchers often work with Crowdflower to coordinate their requests and listing HITs on Amazon Mechanical Turk (Human participant research guidelines, 2012).

Data Collection Systems

Typically University of Waterloo researchers use Crowdflower only as an intermediary to recruit participants through Amazon Mechanical Turk, even though the Crowdflower platform allows researchers to create their own surveys hosted by Crowdflower Researchers tend to create the surveys themselves using Survey Monkey™, Qualtrics™ or a University of Waterloo departmental system. Mainly the crowdflower is used as a third-party platform for accessing the Amazon Mechanical Turk labourmarket. Researchers may use programs where the computer servers are housed in the USA, such as Survey Monkey™ or Qualtrics™, to collect a participant's response to the study tasks or questions (Human Participant Research Guidelines, 2012).

Case Studies

As we know that there are the various applications of crowdsourcing. Some applications related to crowdsourcing are explained here.

Crowdsourcing for Enterprises

In first application we will see how the crowdsourcing is used in enterprises. Now a day, there is no general-purpose

Table 1: Classification of Existing Crowdsourcing Examples by Function and Mode

Crowdsourcing Mode	*Crowdsourcing function*			
	Design and Innovation	*Development and testing*	*Marketing and sales*	*Support*
Competition	Muji Threadless 99De-signs	People Per Hour	Marketocracy	Askville by Amazon
Marketplace	Innocentive I Stock Photo	Top Coder Crowdspirit Mob 4 Hire U Test	Peer To Patent Spot us Predictify	Amazon Mechanical Turk* Get Satisfaction Fixya

crowdsourcing platform, which embeds the task-based services to support crowdsourced activities. We require a good understanding of the right number of emerging crowdsourcing solutions, in order to build such Cloud-enabled service. In addition, to analyze the capabilities of existing crowdsourcing systems, we need a structured approach also.

In this paper, he had the categorised the crowd sourcing plat forms (Maja Vukovi, 2009). The first section presents categorisation of existing crowdsourcing platforms, in order to identify the required capabilities of a general-purpose crowdsourcing service. Heproposes two dimensions: (1) crowdsourced function and (2) crowdsourcing mode.

A subset of existing crowdsourcing solutions according to the proposed categorisation is shown in Table 1. Crowdsourced function represents the part of the product and/or service lifecycle that is being crowdsourced. Crowdsourced function may take one of the following forms: design, development and testing, marketing and sales and support (MajaVukovi, 2009). By crowdsourcing design processes, enterprises benefit from innovation and creativity of crowds. Furthermore, they can employ members of the crowd to evaluate their own design, before proceeding with new product development. One of the examples in this category is Threadless.com, a popular community oriented T-shirt company, where crowds submit their T-shirt designs (Maja Vukovi, 2009). In another example of Japanese design shop, Muji used crowdsourcing to both obtain design ideas and evaluation feedback from over half a million people. 99 designs is a i Stock photo is a digital photo marketplace. Innocentive is a market place for solving business, science, and product development problems (Maja Vukovi, 2009). The next group of this system supports the product development and testing and it also gathers the benefits of scalable workforce, and expertise matching. The u Test.com, mob4 hire. com, Top Coder.com, Crowd Spirit and People Per Hour are examples of this group. Crowdsourcing uses numerous platforms formarketing and sales functions. In marketing and sales functions, employing crowdsourcing for enterprises can benefit from crowd analytics. Amazon Mechanical Turk, FixYa. com, GetSatisfaction.com, and Ask ville are examples of this group. Amazon provides platforms such as crowdsourcing supporting functions. They represent various examples such as information centric systems, a next generation of online help system that have evolved from traditional user groups and online forums, due to this social networking capabilities will be integrated. Amazon Mechanical Turk is a special case because it focuses on providing a marketplace for the micro-tasks, in which content-creation, testing, and micro-development may take place (Maja Vukovi, 2009).

Crowdsourcing General Computation

In this application, Haoqi Zhang et al. (2011), present a direction of research on principles and methods that can enable for general problem solving via human computation systems. They discuss the interplay between algorithmic paradigms and human abilities, and demonstrate through examples how members of a crowd can play diverse roles in an organised problem-solving process, How they are serving 'data oracles' at the end points of computation, as modules for decomposing problems, also controlling the algorithmic progression and performing human program synthesis.

Taking the perspective that coordinating a crowd can be considered as a thought of programming and organisation design, they were investigating the interplay between algorithmic paradigms and human abilities and interests for the purposes of crowdsourcing, via human computation general computation enables the solution of new classes of tasks.

The existing crowdsourcing systems are built, supporting a set of specific micro-tasks in a distinct domain and a

specific part of the product lifecycle. To address the identified gaps in architectural support for building crowdsourcing service they had begun on building a platform for crowdsourcing e.g., demonstrating how the Cloud infrastructure can be used as hosting that can be scalable and application development environment for dynamic, task-based crowd teaming (MajaVukovi, 2009).

They identify three interrelated subareas of study, each focused on a particular way that the crowd can be harnessed in more general problem solving:

By applying known computational procedures with opportunities and challenges to human problem solvers, recognising the potential and need for designing algorithms whose subroutines are tailored to the crowd's abilities and interests (Maja Vukovi, 2009).

Studies of the ability of crowd to guide the control flow of an algorithm, taking advantage of people's ability to make decisions on whether to solve the problem now or further decompose the problem, in search process when to backtrack or to continue, and what should be the subgoals to communicate in organising the efforts of other contributors. Investigation using a crowd as a general problem solver, where given a problem statement as input or a goal, by defining the problem-solving process and executing the plan crowd synthesizes a solution to the problem (Haoqi Zhang et al., 2011).

Haoqi Zhang et al. (2011) had presented an ongoing research and a larger research agenda moving forward for harnessing human computation in general problem solving. In this application, Haoqi Zhang et al.(2011) had introduced three subareas of study that focus on particular ways that a crowd can be harnessed for general problem solving.

Crowdsourcing Predictors of Behavioural Outcomes

The aim of this research was to test an alternative perspective to modeling in which the wisdom of crowds is utilized to both propose potentially predictive variables to study by asking questions, and use the responses of those questions, in order to develop a predictive model.

According to Bongard et al.(2012), crowdsourcing is Human Intelligence Task. It is nothing but the wisdom of crowd. The characterizing data (Sorokin and Forsyth, 2008), transcribing spoken language (Marge et al., 2010), or creating data visualisations (Kong et al., 2010) are the examples of "Human Intelligence Task". It is very difficult to complete tasks to accomplish with computers alone, so by involving large groups of humans in many locations it is possible to make it easier and it would be prohibitively expensive to accomplish the same task through traditional expert-driven processes (Kittur et al., 2008).

They had used crowdsourcing for two applications such as electric energy consumption and body mass index. Because of policy efforts to increase energy efficiency, many are working to provide consumers with better information about their energy consumption. Research on consumer, interpretation of energy efficiency indicates that electricity customers often misjudge the relative importance of various activities related to energy consumption and devices which can be used to reduce energy consumption (Attari et al., 2010).

Crowdsourcing Geographic Information for Disaster Response: A Research Frontier

This application focuses on a specific and rapidly evolving area of geospatial data and tools, which is a subset of social networking and web content which is user-generated that has been termed volunteered geographic information VGI (Goodchild & Glennon, 2010) and that is the focus of an emerging body of research. The events in Santa Barbara which are a series of experience of recent wildfire are used to examine the key issues associated with volunteered geographic information and its potential role in disaster management.

Geographic information is information linking to a location on or near the Earth's surface and a time. Because of many of these components must be measured, and since the potential amount of such information is infinite, it is unavoidable that all geographic information be subject to unpredictability.

For anyone encountering VGI (Goodchild & Glennon, 2010), quality is the first topic that suggests itself for the first time.

No one can expect the results of VGI creation and publication to be accurate, if the providers of VGI are not experts, and if they operate under no institutional or legal frameworks. There are several grounds above which

one can believe the quality of VGI can approach and even exceed that of authoritative sources. First, in some circumstances, there is evidence that the crowdsourcing mechanism works.

Second, geographic information is unusually rich in context. Information about a location x can always be compared to other information that is likely to be available from authoritative sources, and about the same location from and to information about the surroundings of x.

Third, discussions of geographic information quality (Mason & Suri, 2011) give the importance of completeness as a dimension of quality, the degree up to which the data are true and at the time of use, report all existing features (Goodchild & Glennon, 2010).

Unfortunately traditional methods of map-making by government agencies, require expert teams to travel to every part of the area and were constantly subject to funding constraints, it will take lengthy delays in the updating of maps, and as a result the average map may have been years or even decades out of date by the time it was used.

From July 2007 to May 2009 there was a series of four large and damaging fires occurred in rapid succession, even though as we know that the wildfire has always been a part of life in Southern California, but due to strong wind that developed from the north, the fire threatened the city of Santa Barbara. In the end no inhabited structures were destroyed and though the costs of fighting the fire ran into the tens of millions, the fire never threatened the city (Good child & Glennon, 2010).

Crowdsourcing in Emotion Studies across Time and Culture

According to Mahmoud et al. (2012), to study the recognition of emotions, on a set of photographs of facial expressions; a very modern approach at that time, Darwin used a judgment study. He showed some Duchenne's photographs (Duchenne & Cuthbertson, 1990) of facial expressions of emotions to people to find out if they agreed about the emotion shown in each expression. Although Darwin's experiments lacked some of the accuracy or precision which was expected of psychology experiments today, his studies were cutting edge at the time. He used external observers and 'realistic' stimuli such as photographs. The main concept of Darwin's experiment is very similar to many contemporary studies on expressions and emotions (Mahmoud et al, 2012).

Conclusion

The existing crowdsourcing systems are built, supporting a set of specific, micro-tasks in a particular domain and a distinct part of the product lifecycle. To address the identified gaps in architectural support for building crowdsourcing service the researcher had begun on building a platform for crowdsourcing, demonstrating how Cloud infrastructure can be used as a scalable hosting and application development environment for dynamic, task-based crowd teaming (Maja Vukovi, 2009).

According to Haoqi Zhang et al. (2011), as they move from parallelisation based on simple partitioning and distribution to more sophisticated problem-solving procedures, they found that it is useful to view the coordination of a crowd's problem-solving as programming, and also, as organisation design. Little et al. (2009) developed Turkit, a toolkit that enables requesters to write programs executed by human workers on Mechanical Turk, showing how basic programming constructs can be applied to human computation.

Haoqi Zhang et al. (2011) had seen great opportunities in the design of human computation systems, for leveraging algorithmic paradigms and much research ahead on principles and methods. These methods and principles can enable new classes of application.

According to Bongard et al. (2012), this paper introduced a new approach to social science modeling in which the participants themselves are motivated to find out the correlates of some human behaviour outcome, such as home owner electricity consumption or body mass index. In both cases participants successfully uncovered at least one significant predictor of the outcome variable. For the body mass index outcome, the participants successfully formulated many of the correlated factors which are useful to predict BMI, and should provide honest values for those correlated factors sufficiently to become predictive during the experiment.

Predictors of a behavioural outcome may be stimulated under several conditions. Such as, if subjects are incurring a health or financial cost for electricity consumption as a result of the outcome under study, crowd may be

motivated to contribute. But whatever correlates the crowd should be meaningful and relevant, so that the next peer of questions should be found from the outcomes.

During an emergency (Goodchild & Glennon, 2010), especially when one that threatens a large community with loss of life and property, agencies are automatically stretched thin. As agencies have limited staff, they have limited ability to acquire and synthesize the geographic information that is vital to effective response. On the other hand, the common citizen is equipped with powers of observation, and is now empowered with the ability to georegister those observations, to transmit them through the Internet, and to produce electronically them into readily understood maps and status reports. Thus the fundamental question raised in this paper is: Howcan society employ the eyes and ears of the general public (crowd), their willingness and eagerness to help, and their recent digital empowerment (i.e. Internet), to provide effective assistance to responders (may be crowd) and emergency managers (again may be crowd)? (Goodchild & Glennon, 2010).

From the recent experience of the Santa Barbara fires, it was suggested that a community can indeed contribute effectively. Of course, there are risks, and more research is urgently needed to understand and minimize them. The critical importance of this, is the role of the citizen is important in those parts of the world that lie beyond the 'digital divide' (i.e. Internet), where the Internet and its services are mostly unavailable (Goodchild & Glennon, 2010).

Mahmoudet al. (2012) recreated Darwin's original experiment using a crowdsourcing approach. The goal of this was to study if the labels given to the photographs now that would correspond to labels given by Darwin's subjects. Mahmoud et al. (2012) displayed the photographs that were used by Darwin in his original experiments. Besides interest in Darwin's work, the subjects for this experiment were anonymous and had no material incentive for task completion. The task was too short and took around two minutes on average.This experiment was publicized on the BBC news website, and acquired 202, 303 labels in total, or per photograph it was 18, 391.

Crowdsourcing can be considered an indispensable research tool, in the fields of psychology and social sciences. Ideas that shared properties of crowdsourcing existed since 1866, even though the term is relatively new, in studying emotions, realizing the importance of collecting data from large and diverse population in order to reach reliable conclusions that generalize across different cultures. Their recreation experiment reinforces the belief that the ability to perceive emotions remains stable through time and cultures.

From all the above examples, we can conclude that whatever task is done by the experts or computers, may the outcome result will be correct, but sometimes they have the limitations. When we are considering the survey applications for various fields, then crowdsourcing can be the alternative for getting more accurate results.

References

1. Attari, S. Z., DeKay, M. L., Davidson, C. I. & De-Bruinc, W. B. (2010). Public *Perceptions of Energy Consumption and Savings*. Proceedings of the National Academy of Sciences, August, (pp. 16-20).
2. Berinsky, A. J., Huber, G. A. & Lenz, G. S. (2010). Using Mechanical Turk as a Subject Recruitment Tool for Experimental Research.Retrieved from http://huber.research.yale.edu/materials/26_paper.pdf.
3. Bongard, J. C., Paul, D., Hines, H., Conger, D., Hurd, P. & Lu, Z. (2012). *Crowd-sourcing Predictors of Behavioral Outcomes*. IEEE Transactions on Systems, Man, and Cybernetics.
4. Buhrmester, M., Kwang, T. & Gosling, S. D. (2011). Amazon's mechanical Turk: A new source of in-expensive, yet high-quality, data? *Perspectives on Psychological Science*, 6, 3-5.
5. Duchenne, G. & Cuthbertson, R. (1990). The Mechanism of Human Facial Expression. Cambridge University Press.
6. Goodchild, M. F. & Glennon, J. A. (2010). Crowdsourcing geographic information for disaster response: A research Frontier. *International Journal of Digital Earth, September*, 3(3), 231-241.
7. Human Participant Research Guidelines.Revised September 2012, November 2012, July 2013.
8. Kittur, A., Chi, E. & Suh, B. (2008). *Crowd-sourcing User Studies with Mechanical Turk*. In Proceedings Twenty-sixth Annual SIGCHI Conference on Human Factors in Computing Systems.
9. Kong, N., Heer, J. & Agrawal, M. (2010). Perceptual *Guidelines for Creating Rectangular Tree-maps*. IEEE Transactions on Visualization and Computer

Graphics, 16(6).

10. Little, G., Chilton, L. B., Goldman, M. & Miller, R. C. (2009). Turkit: Tools for Iterative Tasks on Mechanical Turk. In KDD-HCOMP '09.
11. Mahmoud, M., Baltrušaitis, T. & Robinson, P. (2012). *Crowd-souring in Emotion Studies across Time and Culture*. Crowd MM' 12, October 29, Nara, Japan.
12. Maja, V. (2009). Crowd-sourcing for Enterprises. IBM T. J. Watson Research.
13. Marge, M., Banerjee, S. & Rudnicky, A. (2010). *Using the Amazon Mechanical Turk for Transcription of Spoken Language*. In Proceedings of IEEE International Conference on Acoustics Speech and Signal Processing.
14. Mason, W. & Suri, S. (2011). Conducting behavioral research on Amazon's Mechanical Turk. Behavioral Researchers.
15. Sorokin, A. & Forsyth, D. (2008). Utility Data Annotation with Amazon Mechanical Turk. In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops.
16. Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans. (2010). Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, and Social Sciences and Humanities Research Council of Canada. Ottawa, Ontario.
17. Zhang, H., Horvitzy, E., Millerz, R. C. & Parkes, D. C. (2011). *Crowd-sourcing General Computation*. May, pp. 7-12, Vancouver, BC, Canada.