Crowdsourcing and Its Applications in Cultural Heritage

Andrew Huynh
University of California, San Diego
Department of Computer Science and Engineering
a5huynh@cs.ucsd.edu

August 20, 2013

Abstract

In the past decade, "human computation" and "crowdsourcing" has emerged as a new field of research leveraging the vast human connectivity offered by the Internet to solve problems that are often too large for individuals or too challenging for current automatic methods. While the physical preservation of cultural heritage has been an on-going process for decades, digital preservation techniques have only recently been applied to the protection, documentation, and interpretation of the world's cultural heritage. Crowdsourcing offers a unique way to involve the beneficiaries of the heritage being preserved in the process of the actual preservation. We present several examples of how crowdsourcing techniques is being utilized to protect, document, and understand cultural heritage. Finally, we discuss the implications of crowdsourcing on the future of cultural heritage preservation.

1 Introduction

The rise of "human computation" [54] and "crowdsourcing" [25] as a field of research emerged from attempts to harness and motivate individuals to finish small tasks by either creating enjoyable experiences [57], appealing to scientific altruism [14], payment [31], or related services in exchange for participation [27].

The term crowdsourcing was first coined in 2006 by Jeff Howe in a Wired magazine article [25], referring to the concept as a method and business model [26] of outsourcing tasks to the crowd. This outsourcing attempts to harness the power of a distributed network of cheap labor; people who are motivated to use their spare time to finish small tasks traditionally performed by an employee. In recent years, the small tasks outsourced have become increasingly complex, relying more and more on human intelligence to discover a solution.

Related to, but not synonymous with crowdsourcing is the concept of human computation. Human computation, a term used as early as in 1838 [44] and more recently and appropriately in Luis Von Ahn's Ph.D. dissertation, referring to the idea of using human cognitive abilities to "solve problems that computers cannot yet solve [54]." These human "computations" can refer to not only to numerical computations but those that require creativity, intuition, pattern recognition, and other forms of human cognitive processing. Thus, while human computation does not necessitate the use of crowdsourcing, it is often paired together in order to accomplish the end goal.

Human computation systems using crowdsourcing have been applied to tackle enormous problems in multiple fields ranging from image annotation [56], galaxy classification [36], music annotation [53], text transcription [57], protein folding [14], and remote damage assessment for disaster response [7]. They have systematically demonstrated that reliable data can be collected in large amounts through incremental contributions from huge numbers of participants.

Growing in parallel with the rise in data collection capabilities [15], digital cultural heritage refers to a variety of activities from the development of digital technologies used in exhibits to the collection [43], preservation [13], and interpretation of cultural heritage objects and locations from around the world. Over the years, human computation and crowdsourcing have been incorporated in more and more cultural heritage projects. One of its greatest successes, the digitization and transcription of ancient texts [42, 13, 61] have led to the preservation and digitization of numerous important historical texts, including the Dead Sea Scrolls [1], ancient Greek scrolls [2], and Civil War Diaries [42]. The very essence of cultural heritage invites public dissemination and accessibility, allowing the masses to explore and connect with the past. Indeed, the locality of cultural heritage and the impact it may have on the world presents unique human computation tasks wherein the beneficiaries of the tasks may often be the participants themselves.

To understand how crowdsourcing and human com-

[1] http://www.google.com/culturalinstitute/about/deadseascroll/
[2] http://ancientlives.org/

putation has become the field it is today and its applications to cultural heritage, we start by examining the different methods of crowdsourcing and how they distribute small tasks that benefit from human involvement (Section 2 & 3). We can then examine how different approaches, ranging from simple clustering techniques to machine learning, can be applied to the crowdsourced data to extract insight (Section 4). Finally, we examine how cultural heritage has benefited from the distributed human involvement of crowdsourcing and human computation (Section 5).

2 Active Crowdsourcing

Traditional, "active", crowdsourcing techniques first described by Howe in 2006 rely on a platform in which certain tasks can be requested for completion by the public [25]. This platform can be represented as a singular website dedicated to a specific task, such as galaxy classification in GalaxyZoo [36], or a larger market place in which many different types of tasks can be requested, such as the human intelligence tasks (HITs) in the Amazon Mechanical Turk [3] marketplace.

At a very fundamental level, these different crowdsourcing platforms must adhere to a set of principles that enable participants to finish a task in a meaningful manner and accomplish the overarching goals of the platform. Specifically, a platform must be able to engage and attract participants, design tasks in such a way that participants can efficiently complete the task, and include constructs to ensure a high-quality of data output.

2.1 Crowd Engagement

Engaging a large pool of labor and presenting compelling motivation to complete tasks is essential to a platform's success. Here we examine three different types of engagement: "Value Incentives", "The Greater Good", and "Games With A Purpose."

Value Incentive

Participants who can be engaged and motivated through *Value Incentives* expect compensation that has a tangible value, which may come in one of two forms:

- **Cold Hard Cash** - Marketplaces such as Amazon Mechanical Turk present a financial reward in the exchange for completion of tasks. This financial reward is set by the task assigner and can vary depending on the difficulty or length of the task.
- **Products and/or services** - The recent introduction of "crowdfunding" [26, 8, 46] platforms such as Kickstarter [4] allow the outsourcing of money contributions instead of tasks in exchange for an end product or service that is created utilizing the monetary contributions. Crowdfunding is less about utilizing human intelligence for computation tasks and more about harnessing the power of a particular community to fund and connect them to services or products.

An impressive example of how motivating a value incentive can be is the 2009 Red Balloon Challenge. Teams competed for a $40,000 prize to find 10 weather balloons deployed at undisclosed locations across the continental United States. The only team to find all balloons accomplished the task in the span of 8 hours with crowdsourcing and a unique monetary incentivisation method. Their hierarchal incentivisation divided the prize money with participants who found the balloons [51], creating a motivation to discover the balloons and share the task with as many participants as possible. This method quickly motivated thousands of participants to search for the balloon and disseminate the task on social media sites with the hopes of claiming a piece of the prize. However, due to the public nature and competitiveness of the task, the team experienced plenty of potentially inaccurate and fraudulent submissions. This negative outcome demonstrates that while value incentives map quickly motivate a large number of participants, it is important to note that the quality of the data received does not increase with the value offered [37] and constructs must be placed to ensure the highest quality of collected data.

The Greater Good

In contrast to participants in *Value Incentive* category, participants who are engaged and motivated to contribute to a platform incentivized by only the idea of pushing forward humanity can be seen as "citizen scientists", serving *The Greater Good*. These citizen scientists often act as collaborators with the task assigners, volunteering their time to map streets around the world [24], classify galaxies [36], translate ancient texts [5], and even assess satellite imagery to aid emergency disaster response [7, 23].

The earliest forms of citizen scientists can be traced back to the early 1900s with the Christmas Bird count run by the National Audubon Society [49], where amateur birdwatchers contributed a major source of data used to determine the status of bird species in North America [11]. More recently, the powerful connectivity of the internet has lead to larger scale citizen science projects contributing useful data to a number of fields. These collaborations between citizens and scien-

[3] http://www.mturk.com/mturk/

[4] http://www.kickstarter.com

[5] http://www.ancientslives.org

tists have shown enormous potential [41] as a enhanced method of collecting and interpreting scientific data.

Games With A Purpose

Games With A Purpose (GWAPs) [57] attempts to engage individuals through personal enjoyment or social reward by creating a game environment in which crowdsourced tasks take place. GWAPs are often related to the process of "gamification", wherein game elements are added to what may be a mundane task, such as image annotation [56], in order to motivate and engage participants in the task.

While there exist GWAPs where participants may contribute solely for their own entertainment value, such as those built around image annotation [56, 58, 60] and collecting "common-sense facts" [59], GWAPs can be and often are coupled with the previously mentioned *Value Incentive* and *The Greater Good* methodology as a means for additional motivation and engagement.

A recent GWAP, Foldit [14], created a crowdsourced effort to produce accurate protein structure models, successfully using this engagement and motivation model to attract players who were not only interested in their potential contributions to a scientific field but the interaction with other competitive players in a game environment as well. In cultural heritage, *Exploration: Mongolia* [6] uses this approach to motivate "armchair archaeologists" to annotate satellite imagery to aid the search the Tomb of Genghis Khan in a game environment.

2.2 Quality Control

A major issue often discussed in crowdsourcing research is the interesting data quality control problems that occur when accepting potentially inaccurate or even fraudulent data from human participants. Verifying the quality of every submitted data point without existing automated methods can be as expensive, if not more, as the process of collecting the initial data [29]. As such, researchers have used a variety of methods to ensure a higher quality of output from participants which we explore below.

Automatic Verification

Some crowdsourcing efforts like Foldit [14] have difficult solutions but involve problems that can be easily verified. Automatic verification can be used to check the quality of solutions as soon as a task is completed, often times awarding the participant who found the solution.

Agreement

Perhaps one of the first methods of quality control, different examples of agreement can be seen in the ESP game [56] and TagATune [34]. Quality control with agreement accepts input if *all* participants agree on the same solution for a task. The number of participants required to finally accept a solution is a minimum of two and can vary according to the task.

Agreement can occur before or after a task, respectively known as input agreement and output agreement. Input agreement, utilized in TagATune, involves giving independent participants inputs that may or may not be the same and asked to describe the inputs to each other to determine the solution to the task. Contrastingly, output agreement, utilized in the ESP game, accepts a solution only after two or more participants agree on the same solution.

Redundancy

Building upon agreement, redundancy as a form of quality control requires multiple participants to complete the same task [48] selecting a final solution through majority voting or averaging aggregate solutions. Efforts such as reCAPTCHA [57] have successfully utilized this method to ensure high quality output, at the cost of additional crowdsourcing work.

Reputation

In reputation systems, a participant's quality of work is determined through previous inputs. The quality of a given past input can be determined by the task assigner or reviewed by other participants.

Reputation can be used to immediately filter out participants who have a high potential of inaccurate input. Amazon Mechanical Turk employs reputation systems to allow task assigners to specify a minimum reputation level for any HITs that are created, thereby ensuring a certain level of quality in the data that is collected.

Expert/Crowdsourced Review

Quality control using expert or crowdsourced review involve additional sets of participants who review and check solutions for accuracy during or after the initial tasks are completed.

An example of crowdsourced review, Soylent [9] employs this method as the *Find-Fix-Verify* pattern, splitting tasks into three stages that filter out inaccurate solutions at each stage through redundancy. Monetary rewards, such as those offered for HITs on Amazon Mechanical Turk, can be withheld until a solution is reviewed and accepted.

Statistical Filtering

If the solution for a task is expected to match some distribution, such as in Quality of Experience (QoE) evaluations [12], researchers can filter out inaccurate data

[6] http://exploration.nationalgeographic.com

points. QoE indicates the degree of a user's subjective satisfaction with a particular piece of media, a task that can be easily crowdsourced to a large number of participants. By using paired comparisons, wherein a participant is asked to compare two stimuli of varying degrees of quality simultaneously, the results of the task can be verified and filtered through simple transitivity.

Annotation Models

Modeling annotation processes [30, 62] using a set of variables provide a rich method of quality control through the representation of different factors that may lead to inaccurate data. These models can group participants of certain skill levels together and allow researchers to filter or combine participants to improve accuracy.

For instance, Welinder et. al [62] look at a set of variables (competency, expertise, and bias) to represent the annotation ability of participants and the difficulty of the task, which in this case was annotating an image. The idea is that competent participants tend to provide accurate annotations while less competent participants provided annotations with inconsistent accuracy. Participants may also have certain "strengths" or levels of expertise that make certain solutions easier for them than others. Finally, the difficulty of certain images may lead to more inaccuracies even by the more competent participants. Through a Bayesian generative probabilistic model of the annotation process, Welinder et. al found that they could infer factors such as the difficulty of the image and participant competency and expertise.

Karger et. al [30] takes this even further and models both tasks and participants to determine an optimal task assignment. Through comparing participants' solutions to the same task, we can infer a participant's reliability and weigh their answers accordingly. This comparison is done by constructing a bipartite graph where each edge corresponds to a task-participant assignment. An iterative algorithm updates the reliability of a participant through each task, ending with a final estimate of the weighted sum of solutions using the reliability weights calculated.

2.3 Discussion

In this section, we examined how active crowdsourcing requires a large distributed network of participants and different models of engagement and motivation to successfully complete tasks. Generally, a combination of different engagement methods are used to attract the largest number of participants. Furthermore, it is important to note that while financial incentives have the ability to motivate large number of participants and are utilized as the business models for platforms such as Amazon Mechanical Turk, it was found that greater financial reward does not necessitate greater motivation and quality from the participants [37].

Once the task is scaled up to enough participants, we arrive at the question of data quality and how to ensure the highest level of output when receiving data from participants with differing levels of experience and knowledge. The quality control methodologies discussed have evolved from simple agreement [34, 56] and majority voting [48, 57] methods which require very little post-processing to the more state-of-the-art methods of annotation models [30, 62] that can represent participants and tasks as a set of variables to optimize. Depending on the tasks, simpler methods may be more attractive than the more involved methods of annotation modeling. While simpler methods of quality control are easier to implement and administrate, many require additional costs [48]. These costs are typically from additional tasks for participants or experts to verify and accept solutions.

3 Passive Crowdsourcing

Crowdsourcing often invokes the image of large groups of participants in front of their computers independently working on a task. Passive crowdsourcing, "work for nothing" [1], relies on being able to gather data or perform computations as a result of the existing behavior of a group of participants with or without their knowledge. Due to its passive, often hidden nature, passive crowdsourcing bypasses many of the problems due to crowd motivation and engagement present in active crowdsourcing techniques, while also raising questions about privacy and the ethics of gathering data without a participant's knowledge.

3.1 Parasitic Computing [6]

First coined by Albert-Laszlo Barabasi in 2001, parasitic computing leverages existing behavior of individuals to solve pieces of a complex computation problem, often without their explicit permission. This has involved harnessing small amounts of a participant's computational power [5, 39, 40] or by embedding the computations into an underlying standard protocol [6, 32].

Initially, parasitic computing utilized underlying standard protocols as a layer in which to do small pieces of computational work. For example, Barabasi [6] utilized the Hyper Text Transmission Protocol (HTTP) to perform calculations to solve a NP-complete satisfiability problem. The calculations were accomplished using a standard characteristic of TCP, a checksum which is checked by a receiving computer and can be formulated such that the receiving computer performs any arithmetic operation. A special message is created to represent a potential solution to an NP-complete problem, such as 3-SAT, and sent to many targets. Since TCP messages with an incorrect checksum are dropped, only those that contained a valid solution will receive a reply.

Kohring [32] further explored this concept, embedding the computations as part of the standard IP communication protocol through the use of carefully constructed ICMP packets. The use of IP and ICMP increases the overall target surface as the targets are no longer restricted to HTTP servers like Barabasi implementation and any Internet capable device is required to implement ICMP. Much like Barabasi, Kohring utilized a checksum in the ICMP message to encode the calculation needed. However, since ICMP does not contain a guarantee that a message will arrive in order or even arrive at all, the problems that can be solved are more restricted. In this case, Kohring simulated a stochastic neural network due to its built-in resilience to noise.

More recent approaches harness idle or underutilized CPUs to do pieces of computational work. Acting as a screensaver for idle computers SETI@home [5] has been able to harness millions of computers around the world to analyze radio signals from space with the explicit permission of participants. Merelo et al. [39, 40] has explored this concept as a web application by running genetic algorithms in web browsers using a small percentage of computation resources as the result of web page being loaded.

3.2 Crowdsensing

In contrast to parasitic computing, crowdsensing [22] involves the collection of data from participants through existing behavior as a method of providing a solution to those same participants. This often involves, but is not limited to, embedded sensors on consumer-centric mobile devices or other computing devices with wireless capabilities that produce data as a participant interacts with the device or world through their existing behavior.

A remarkable example of crowdsensing without the use of mobile sensors is Google Flu Trends (GFT) [7]. GFT uses search query data for influenza related search terms provided by millions of Google users to estimate influenza activity. These results have been shown as an accurate predictor of influenza trends [18] tracking influenza even before major national disease centers.

Building upon the existing need and activity of checking traffic, Google Traffic [28] and VTrack [52] crowd sense real-time traffic information through the act of using their map or traffic checking application.

A non-intuitive, yet powerful example of crowdsensing relates to 3D reconstruction from photographs [2, 20], which requires large amounts of imagery, has benefitted from crowdsensing through the use of publicly available photographs on websites such as Flickr [8]. This example is especially useful in digital preservation of cultural heritage, where publicly available photos, such as those taken by tourists, of important cultural heritage sites could be used to construct models.

3.3 Discussion

In this section, we explored the concept of passive crowdsourcing which relies on the ability to gather data and/or perform computations as a result of existing participant behavior with or without their knowledge. The rise of embedded sensors in everyday computing devices such as mobile devices and the ever expanding network of Internet enable devices have created a largely untapped potential source of data and computational power. Parasitic computing taps into existing Internet infrastructure to perform computations, such as using HTTP or ICMP checksums to harness the computation power of other computers. Crowdsensing looks to using embedded sensors available in everyday computing devices to capture useful information present in the existing behavior of participants, such as traffic or flu information.

However due to the opportunistic nature of passive crowdsourcing, especially in crowdsensing, sensitive data pertaining to individuals must be kept private and secure while also ensuring passive crowdsourcing applications still provide some level of utility. Some simple procedures, such as running analytics on the aggregate [22] of all data and only keeping data that is anonymous and unattached to any single participant goes a long way to protecting the identity of the participants on the platform. Alternatively, differential privacy [19] methods could be used to encrypt sensitive data while still provided statistically accurate analysis.

4 Learning from the Crowd

Crowdsourcing has the potential to collect an enormous amount of information, all of which can be harnessed to partially or fully automate the tasks from which the data originated through the use of machine learning. Machine learning approaches excel at recognizing consistent patterns among labeled training examples, such as those provided by crowdsourcing participants, which can then be used to label new, unlabeled data.

Pattern recognition techniques rely on the extraction of features that describe the properties of the concept being recognized. For example, feature extraction on images can refer to edges, corners, shapes, or even color distribution. Machine learning algorithms can then identify statistical consistencies between sets of features

[7] http://www.google.org/flutrends

[8] http://www.flickr.com

and can apply this knowledge to future sets of new, unlabeled data.

While traditional machine learning approaches focus on a single, fixed training set to learn static models, crowdsourced data presents a source of temporally dynamic data which can be used to continually create new, improved models. This methodology, known as active learning [47, 17], is able to query some outside source for additional information that can be used to improve a model. Generally, the additional information comes in the form of additional labels from participants that can be applied to instances with no clear classification. Active learning has been shown to produce reliable, accurate models [7, 10].

4.1 Active Learning

Active learning techniques can be categorized into two overarching methodologies based on how additional information is acquired to improve the underlying model.

Stream-based [21] or *online* sampling approaches require a decision as each unlabeled instance is queried. The key assumption in online active learning is that obtaining an unlabeled instance is free or inexpensive so it can be sampled from a source one at a time. Due to the sequential querying nature of stream-based approaches, applications of stream-based approaches must decide of whether or not to continue querying for instances. This decision can rely on some "informative measure" or "query strategy" such that more informative instances are more likely to be queried [47].

In contrast, *pool-based* [38] sampling approaches allow an algorithm to choose from a typically large set of unlabeled data based on some quantitative measure and a small set of labeled data. For many real-world learning problems, a large set of unlabeled data can be gathered at once, for instance using crowdsourcing. Pool-based methods are typically studied in more detail, used in approaches stemming from text classification [35, 63], natural language processing [50], and more.

Furthermore, we can categorized active learning algorithms based on the types of training data available [7]. *Discriminative* methods harness both positive and negative training data to create a model. In contrast, *generative* methods require only positive training data while negative training data does not increase the accuracy of the model.

Active learning presents a simple, powerful approach to harness crowdsourced human annotators in collaboration with machine learning techniques to continually formulate new, improved models as more and more data is collected.

4.2 Closing the loop

Querying new training instances through uncertainty sampling [35] involves the selection of instances that have low classification or prediction confidence. Using this selection criteria, we can have new human annotators label instances where we have low confidence, retrain our model, and repeat the process thus completing the loop between machine learning and human annotators.

While uncertainty sampling provides us with a useful criterion for selecting unlabeled instances, it does not provide a method to select a annotator. Annotation modeling methods discussed in Section 2.2 can potentially be used to pair new unlabeled instance with an optimal annotator. Karger et. al [30] used bipartite graphs to model tasks and annotators. Yan et. al [63] presents an optimization problem to select the best possible annotator based on the current iteration of data and model.

Due to the fact that human annotations are inherently noisy, an open research question [47] is the decision of whether to label a new unlabeled instance or repeat the labeling of an existing instance to potentially improve the model. Research [48] using heuristics representing uncertainty in both the human annotators and the active learning model have shown that repeated labeling improves the overall data collected but assumes that all human annotators represent the same level of expertise and that there exists some true label for any instance. Addressing the first assumption, Donmez et. al [16] allowed human annotators to have different "noise" levels showing that both true labels and the expert level of individual human annotators can be estimated, thus allowing them to query only the more expert human annotators in later active learning iterations.

4.3 Discussion

By itself, crowdsourcing and human computation tasks are powerful but difficult to scale without constantly active engagement and motivation. This section explored the utility of using crowdsourcing to gather a small, but still substantial set of annotated data to be used for training machine learning algorithms. Additionally, if we come across data that can not be classified by our model we can further employ the concept of active learning and uncertainty sampling to close the loop where data goes through humans ⇒ active learning ⇒ humans. Active learning coupled with inexpensive crowdsourced data has been shown to be a cost effective method that is as accurate as more expensive expert-based machine learning [7, 63], requiring only a small number of non- expert participants to match the performance of an expert participant [50].

5 Applications in Cultural Heritage

While the majority of crowdsourcing techniques have thus far focused on fields outside of cultural heritage, the massive data collection capabilities of the crowd have started to find ways to improve the interpretation, digital preservation, and exploration/discovery of cultural heritage objects and sites of interest.

More importantly, the engagement of the public with the interpretation, preservation, and discovery of cultural heritage offers an opportunity for a participant to do something more than simply consume information, "engag[ing] them in the fundamental reason that these digital collections exist in the first place [42]."

5.1 Interpretation

In museums, collections of cultural heritage objects are traditionally organized into galleries based around a common theme by professional curators. This organization helps museum visitors interpret the collections [3] For digital collections, there may not be an clear organization or theme making large collections of digital collections overwhelming. Crowdsourcing different interpretations of a set of cultural heritage creates a compelling human computation task to creatively link together related pieces of human history.

BBC *WW2 People's War* [9] was a platform where the public was allowed to submit "memories" of World War 2 collecting over 47,000 stories and 14,000 images of how an entire generation interpreted the war and preserving it for future generations. This platform provided a unique opportunity for a generation of people to provide their own interpretations and get in touch with others we shared the same experiences.

The datasets that make up large digital collections make curation by the crowd a viable solution albeit one that has not fully been explored for the cultural heritage domain.

5.2 Digital Preservation

Perhaps the earliest example of using crowdsourcing for cultural heritage was the use of CAPTCHAs [55], an automated test to tell humans and computers apart, to transcribe and digitize old texts. CAPTCHAs typically consist of an image containing random distorted characters, used as a challenge response during Web form registrations or submissions to prove that the submitter is human. The distorted characters present a significantly difficult problem for computers or "bots" that attempt to automatically submit Web forms, while only a minor inconvenience for humans.

Foundations and projects such as the Google Books Project have been scanned in an attempt to preserve human knowledge and make it more accessible. While optical character recognition (OCR) technology can transcribe a large majority of the imagery scanned, there exists a significant chunk (20% [61] of all scanned text) that require human transcription, who tend to be expensive. Through the combination of CAPTCHAs and the scanned imagery from text digitization projects, reCAPTCHAs [61] replaces the random distorted characters used in CAPTCHAs with words that OCR programs can not recognize. reCAPTHCAs have enabled historical data that have no digital equivalent to now be fully preserved and accessible for millions of humans across the world.

Perhaps a more meaningful example that demonstrates the impact of crowdsourcing cultural heritage preservation, the University of Iowa libraries crowdsourced the transcription of a set of Civil War diaries [42]. As a surprising byproduct of the transcription, the project dramatically increased the number of visitors to the library's website and was shown to be a novel meaningful engagement between the human annotators and this piece of history. These connections between the participants and the content allow them to engage with the collections and with the past, the "apex of user experience for cultural heritage collections" [42], while offering the useful byproduct of transcription.

5.3 Exploration & Discovery

New cultural heritage sites are still being unearthed and analyzed [33, 45, 4], increasingly through more modern methods and often utilizing large amounts of high resolution satellite and/or aerial imagery. As the resolution of these images increase with improved technology, the amount of data that must be sifted through to discover new sites increases significantly. Interpreting ambiguous features in the satellite imagery also present additional difficulties, all of which can be crowdsourcing as human computation tasks.

The *Valley of the Khans* project opens this exploration and discovery process to the public through Exploration: Mongolia [10] an an online platform motivating human annotators to survey high-resolution satellite imagery in the hopes of discovering the lost tomb of Genghis Khan. Annotators are asked to categorize roads, rivers, modern structures, and ancient structures in the satellite imagery to assist field efforts in navigating the large study area. During three field surveys, over 100 sites of interest were visited, resulting in the discovery of fifty-five archaeological sites, ranging a human existence of over 3000 years.

5.4 Discussion

As explored in this section, while crowdsourcing is often though of as an instrument for data collection, it has shown the potential to be so much more. Cultural heritage exists to connect people today with the knowledge

[9] http://www.bbc.co.uk/history/ww2peopleswar

[10] http://exploration.nationalgeographic.com

and experiences of the past and crowdsourcing gives researchers and curators of cultural heritage collections the ability to reach out to new participants while also as a byproduct enriching their own datasets.

Furthermore, as more and more collections turn to digital techniques for preservation and analysis, there will an increasingly amount of data and tasks that can benefit from human computation and crowdsourcing. While still being explored, the crowdsensing techniques described to create 3D models from photographs [2] could someday be used to accurately preserve and analyze historic structures and objects. Crowdsourced methods of exploration and discovery of new sites of interest with high-resolution satellite imagery are also being researched.

6 Towards the Future

Crowdsourcing has shown increasing potential as a platform to perform human computation tasks and its many successes across a number of different fields have shown that serious work can be done using inexpensive, distributed participants. We found (Section 2) that crowd engagement was a large part of the successes in different crowdsourcing platforms, and depending on the type of task, there are multiple methods that can be used to motivate participants to finish tasks.

Furthermore, the use of passive crowdsourcing (Section 3) has side-stepped the issues of motivation and crowd engagement that may plague active crowdsourcing platforms. Passive crowdsourcing is increasingly being used in novel ways that can measure and collect data from existing participant behavior. While limited in the number of potential applications, passive crowdsourcing provides a powerful alternative to traditional crowdsourcing platforms as a means of data collection and distributed computation.

A potential solution and progression in crowdsourced data analytics is the use of machine learning algorithms (Section 4) to search for patterns existing within the collected data. This reduces the number of potential participants and tasks, reducing the financial burden while also providing a way to scale past the limits of crowdsourcing. As a way to continually increase the accuracy of machine learning models, we can use the concept of active learning which allows us to iteratively improve a model with additional information.

While crowdsourcing is often thought of as an instrument for data collection, it has shown the potential to be so much more [42]. As discussed in Section 5, as digitization methodology improves over the years, there will an ever increasing demand for more and more human computation to analyze, interpret, and discovery digital cultural heritage. Cultural heritage has the unique and powerful feature of involving the beneficiaries of the heritage being preserved or analyzed in the process of the actual preservation and analysis. A "crowdsourcing" platform for cultural heritage does not only have to serve as a platform for distributing tasks and collecting solutions but can also serve as a platform for connecting the participants to the history involved.

7 Acknowledgements

Thanks to Professor Yoav Freund for chairing my research exam committee and for Serge Belongie and Bill Griswold for participating on the committee. Thanks for Albert Lin for proofreading and giving feedback on drafts.

References

[1] E. Adar. Why I hate Mechanical Turk research (and workshops). In *Proc. CHI Workshop on Crowdsourcing and Human Computation*, 2011.

[2] Sameer Agarwal, Noah Snavely, Ian Simon, Steven M Seitz, and Richard Szeliski. Building Rome in a day. *2009 IEEE 12th International Conference on Computer Vision*, (Iccv):72–79, September 2009.

[3] Nikolaos Aletras, Mark Stevenson, and Paul Clough. Computing similarity between items in a digital library of cultural heritage. *Journal on Computing and Cultural Heritage*, 5(4):1–19, October 2012.

[4] Dimitrios Alexakis, Apostolos Sarris, Theodoros Astaras, and Konstantinos Albanakis. Detection of neolithic settlements in thessaly (Greece) through multispectral and hyperspectral satellite imagery. *Sensors (Basel, Switzerland)*, 9(2):1167–87, January 2009.

[5] DP Anderson, Jeff Cobb, and Eric Korpela. SETI@ home: an experiment in public-resource computing. *Communications of the ACM*, 45(11), 2002.

[6] a L Barabási, V W Freeh, H Jeong, and J B Brockman. Parasitic computing. *Nature*, 412(6850):894–7, August 2001.

[7] Luke Barrington, Douglas Turnbull, and Gert Lanckriet. Game-powered machine learning. *Proceedings of the National Academy of Sciences of the United States of America*, 109(17):6411–6, April 2012.

[8] Paul Belleflamme. Crowdfunding: Tapping the right crowd. In *International Conference of the French Finance Association (AFFI)*, 2011.

[9] MS Bernstein, Greg Little, and RC Miller. Soylent: a word processor with a crowd inside. *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, pages 313–322, 2010.

[10] A Brew, Derek Greene, and P Cunningham. Using Crowdsourcing and Active Learning to Track Sentiment in Online Media. *ECAI*, 2010.

[11] GS Butcher, MR Fuller, LS McAllister, and PH Geissler. An evaluation of the Christmas Bird Count for monitoring population trends of selected species. *Wildlife Society Bulletin*, 18(2):129–134, 1990.

[12] Kuan-Ta Chen, Chen-Chi Wu, Yu-Chun Chang, and Chin-Laung Lei. A crowdsourceable QoE evaluation framework for multimedia content. *Proceedings of the seventeen ACM international conference on Multimedia - MM '09*, page 491, 2009.

[13] Otto Chrons and Sami Sundell. Digitalkoot : Making Old Archives Accessible Using Crowdsourcing. *Proceedings of the ACM SIGKDD Workshop on Human Computation*, pages 20–25, 2011.

[14] Seth Cooper, Firas Khatib, Adrien Treuille, Janos Barbero, Jeehyung Lee, Michael Beenen, Andrew Leaver-Fay, David Baker, Zoran Popović, and Foldit Players. Predicting protein structures with a multiplayer online game. *Nature*, 466(7307):756–760, 2010.

[15] Gregory Crane. Cultural heritage digital libraries: Needs and components. In *Research and Advanced Technology for Digital Libraries*, pages 626–637, 2002.

[16] Pinar Donmez, Jaime G. Carbonell, and Jeff Schneider. Efficiently learning the accuracy of labeling sources for selective sampling. *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '09*, page 259, 2009.

[17] Gregory Druck, B. Settles, and A. McCallum. Active learning by labeling features. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*, pages 81–90. Association for Computational Linguistics, 2009.

[18] Andrea Freyer Dugas, Yu-Hsiang Hsieh, Scott R Levin, Jesse M Pines, Darren P Mareiniss, Amir Mohareb, Charlotte a Gaydos, Trish M Perl, and Richard E Rothman. Google Flu Trends: correlation with emergency department influenza rates and crowding metrics. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 54(4):463–9, February 2012.

[19] Cynthia Dwork, F McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. *Theory of Cryptography*, pages 1–20, 2006.

[20] JM Frahm, P Fite-Georgel, and D Gallup. Building Rome on a cloudless day. *Computer VisionECCV 2010*, 49(3):352, June 2010.

[21] Y Freund, HS Seung, E Shamir, and N Tishby. Selective Sampling Using the Query by Committee Algorithm. *Machine learning*, 168:133–168, 1997.

[22] RK Ganti, F Ye, and H Lei. Mobile crowdsensing: Current state and future challenges. *Communications Magazine, IEEE*, (November):32–39, 2011.

[23] Michael F. Goodchild and J. Alan Glennon. Crowdsourcing geographic information for disaster response: a research frontier. *International Journal of Digital Earth*, 3(3):231–241, September 2010.

[24] M Haklay and Patrick Weber. Openstreetmap: User-generated street maps. *Pervasive Computing, IEEE*, 7(4):12–18, 2008.

[25] J. Howe. The rise of crowdsourcing. *Wired magazine*, 14(14):1–5, 2006.

[26] Jeff Howe. *Crowdsourcing: How the power of the crowd is driving the future of business*. 2008.

[27] Bret Hull, Vladimir Bychkovsky, and Yang Zhang. CarTel: a distributed mobile sensor computing system. *Proceedings of the 4th international conference on Embedded networked sensor systems*, pages 125–138, 2006.

[28] Google Inc. The bright side of sitting in traffic: Crowdsourcing road congestion data, 2009.

[29] Panagiotis G. Ipeirotis, Foster Provost, and Jing Wang. Quality management on Amazon Mechanical Turk. *Proceedings of the ACM SIGKDD Workshop on Human Computation*, page 64, 2010.

[30] DR Karger, S Oh, and D Shah. Iterative learning for reliable crowdsourcing systems. *Advances in neural information processing systems*, pages 1–9, 2011.

[31] Aniket Kittur, E.H. Chi, and Bongwon Suh. Crowdsourcing user studies with Mechanical Turk. In *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pages 453–456. ACM, 2008.

[32] GA Kohring. Implicit simulations using messaging protocols. *International Journal of Modern Physics C*, 14(2):203–213, 2003.

[33] Rosa Lasaponara and Nicola Masini. Detection of archaeological crop marks by using satellite QuickBird multispectral imagery. *Journal of Archaeological Science*, 34(2):214–221, February 2007.

[34] Edith Law and L Von Ahn. Input-agreement: a new mechanism for collecting data using human computation games. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1197–1206, 2009.

[35] DD Lewis and WA Gale. A sequential algorithm for training text classifiers. In *Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 3–12, 1994.

[36] Chris J. Lintott, Kevin Schawinski, Anže Slosar, Kate Land, Steven Bamford, Daniel Thomas, M. Jordan Raddick, Robert C. Nichol, Alex Szalay, Dan Andreescu, Phil Murray, and Jan Vandenberg. Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey . *Monthly Notices of the Royal Astronomical Society*, 389(3):1179–1189, September 2008.

[37] Winter Mason and DJ Watts. Financial incentives and the performance of crowds. *ACM SigKDD Explorations Newsletter*, 11(2):100–108, 2010.

[38] A McCallum and K Nigam. Employing EM and Pool-Based Active Learning for Text Classification. *ICML*, 98:350–358, 1998.

[39] JJ Merelo, AM García, and JLJ Laredo. Browser-based distributed evolutionary computation: performance and scaling behavior. *Proceedings of the 2007 GECCO conference companion on Genetic and evolutionary computation*, pages 2851–2858, 2007.

[40] JJ Merelo-Guervós. Asynchronous distributed genetic algorithms with Javascript and JSON. *Evolutionary Computation*, pages 1372–1379, 2008.

[41] Greg Newman, Andrea Wiggins, Alycia Crall, Eric Graham, Sarah Newman, and Kevin Crowston. The future of citizen science: emerging technologies and shifting paradigms. *Frontiers in Ecology and the Environment*, 10(6):298–304, August 2012.

[42] Trevor Owens. Digital Cultural Heritage and the Crowd. *Curator: The Museum Journal*, 56(1):121–130, January 2013.

[43] Nancy Proctor. Digital: Museum as Platform, Curator as Champion, in the Age of Social Media. *Curator: The Museum Journal*, 53(1):35–43, 2010.

[44] AJ Quinn and BB Bederson. Human computation: a survey and taxonomy of a growing field. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1403–1412, 2011.

[45] Aled Rowlands and Apostolos Sarris. Detection of exposed and subsurface archaeological remains using multi-sensor remote sensing. *Journal of Archaeological Science*, 34(5):795–803, May 2007.

[46] A Schwienbacher and B Larralde. Crowdfunding of small entrepreneurial ventures. In *HANDBOOK OF ENTREPRENEURIAL FINANCE*. 2010.

[47] Burr Settles. Active learning literature survey. Technical report, 2010.

[48] VS Sheng, Foster Provost, and PG Ipeirotis. Get another label? improving data quality and data mining using multiple, noisy labelers. *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 614–622, 2008.

[49] Jonathan Silvertown. A new dawn for citizen science. *Trends in ecology & evolution*, 24(9):467–71, October 2009.

[50] Rion Snow, B O'Connor, Daniel Jurafsky, and AY Ng. Cheap and fast—but is it good?: evaluating non-expert annotations for natural language tasks. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, number October, pages 254–263, 2008.

[51] John C. Tang, Manuel Cebrian, Nicklaus a. Giacobe, Hyun-Woo Kim, Taemie Kim, and Douglas "Beaker" Wickert. Reflecting on the DARPA Red Balloon Challenge. *Communications of the ACM*, 54(4):78, April 2011.

[52] Arvind Thiagarajan and Lenin Ravindranath. VTrack: accurate, energy-aware road traffic delay estimation using mobile phones. *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems*, pages 85–98, 2009.

[53] Douglas Turnbull, Ruoran Liu, and Luke Barrington. A game-based approach for collecting semantic annotations of music. *Conference on Music*, pages 2–5, 2007.

[54] Luis von Ahn. Human computation. In *Proceedings of the 46th Annual Design Automation Conference*, page 418, New York, New York, USA, 2009. ACM Press.

[55] Luis von Ahn, Manuel Blum, N Hopper, and John Langford. CAPTCHA: Using hard AI problems for security. *ADVANCES IN CRYPTOLOGY EUROCRYPT 2003*, pages 294–311, 2003.

[56] Luis von Ahn and Laura Dabbish. Labeling images with a computer game. *Proceedings of the 2004 conference on Human factors in computing systems - CHI '04*, 6(1):319–326, 2004.

[57] Luis von Ahn and Laura Dabbish. Designing games with a purpose. *Communications of the ACM*, 51(8):57, August 2008.

[58] Luis von Ahn and S Ginosar. Improving Image Search with PHETCH. In *Acoustics, Speech and . . .* , pages 1209–1212, 2007.

[59] Luis von Ahn, Mihir Kedia, and Manuel Blum. Verbosity: a game for collecting common-sense facts. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 75–78. ACM, 2006.

[60] Luis von Ahn, Ruoran Liu, and Manuel Blum. Peekaboom : A Game for Locating Objects in Images. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 55–64, 2006.

[61] Luis von Ahn, Benjamin Maurer, Colin McMillen, David Abraham, and Manuel Blum. reCAPTCHA: human-based character recognition via Web security measures. *Science (New York, N.Y.)*, 321(5895):1465–8, September 2008.

[62] Peter Welinder, Steve Branson, Serge Belongie, and Pietro Perona. The Multidimensional Wisdom of Crowds. In *Advances in Neural Information Processing Systems*, 2010.

[63] Y Yan and GM Fung. Active learning from crowds. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pages 1161–2268, 2011.