Crowdsourcing Human-Based Computation

Doug Wightman
Human Media Lab
Queen's University
Kingston, ON K7L 3N6, Canada
wightman@cs.queensu.ca

ABSTRACT

Thousands of websites have been created to crowdsource tasks. In this paper, systems that crowdsource human-based computations are organized into four distinct classes using two factors: the users' motivation for completing the task (direct or indirect) and whether task completion is competitive. These classes are described and compared. Considerations and selection criteria for systems designers are presented. This investigation also identified several opportunities for further research. For example, existing systems might benefit from the integration of methods for transforming complex tasks into many simple tasks.

Author Keywords

Crowdsourcing, Human-based computation, Distributed knowledge acquisition

ACM Classification Keywords

I.2.6 [**Learning**]: Knowledge acquisition. H5.3 [**HCI**]: Web-based interaction.

INTRODUCTION

Human-based computation is the technique of outsourcing steps within a computational process to humans [11]. Alex Kosorukoff, who coined the term, designed a genetic algorithm that allows humans to suggest solutions that might improve evolutionary processes [11]. His description of human-based computation includes a division of labor between humans and computers. Labor is divided into two roles: selection and innovation. Selection refers to the selection of tasks and innovation to the performance of a task. A human or a computer can act as a selection agent or an innovation agent. For example, electronic systems that administer GRE tests could be considered selection agents and the human users innovation agents. Human-based computation can also involve multiple humans.

Crowdsourcing is the practice of outsourcing tasks to a group of humans [31]. In some cases, computers may be used to administer crowdsourced tasks, creating human-based computation systems. Many online systems (websites and other online applications) feature this form of human-based computation. In this paper, examples of crowdsourced human-based computation (CHC) are grouped into four distinct classes using two factors: the users' motivation for completing the task (direct or indirect) and whether task completion is competitive.

This paper is a preliminary inquiry into the differences between classes of CHC systems. This inquiry was conducted to support the development of future systems by informing designers' CHC class selection decisions. Image labeling games, news aggregation websites, and Wikipedia [32] are non-competitive CHC examples in which the users' primary motivation for participation is to perform the task itself (direct motivation). reCAPTCHA [21], online surveys, and websites for volunteers can be non-competitive CHC systems in which the users' primary motivation for participation is not the task itself (indirect motivation). Mechanical Turk [17] and InnoCentive [9] are competitive examples with users who are indirectly motivated to participate. Google search ranking and Yahoo! Answers [37] are competitive CHC examples with users who are directly motivated to participate. Each of these examples are described and compared in the following sections. This paper concludes with an analysis of advantages and disadvantages that may be associated with different classes of CHC systems.

NON-COMPETITIVE DIRECT MOTIVATION TASKS

Image Labeling Games

Most of the images that are publicly available on the Internet are not associated with keywords describing their depictions. If image keywords were available, they might be used to improve algorithms for image search and the filtering of inappropriate content. The ESP game [29] is an online game that was designed to label images with keywords. Humans, who play the game for enjoyment, enter the image keywords.

The ESP Game groups players into pairs, shows them both the same image, and awards points when the players type the same word on their keyboards. Every 2.5 minutes, points are awarded and a new image is displayed. The

players are not able to communicate directly with one another, hence the game of the name. The game may indicate that certain words are taboo, meaning that points cannot be earned by entering these words.

Once a certain number of pairs of players have entered the same word for an image, the ESP game will notify future players that encounter the image that this word is not to be used to describe it. This feature forces players to enter different keywords, providing different labels for images. The accuracy of the labels generated by participants playing the ESP game was evaluated. The results indicated that 85% of the keywords associated with images "would be useful in describing" them [29].

The ESP game uses a number of techniques to encourage players to label images accurately. Players are allowed to skip images. This feature is also used to determine when an image is finished being labeled. The rationale is that players will skip an image if all of the keywords are taboo. Players' IP addresses are required to be different. The authors suggest that cheating might be further curtailed by requiring players to use different words to describe each image during a particular session playing the game. This might further reduce opportunities for player collusion.

In their 2004 paper, the creators of the ESP Game estimated that 5,000 individuals playing continuously could label each of the 425 million images indexed by Google in 31 days [29]. The authors note that popular online games websites, such as Yahoo! Games, feature many games that appear to have more than 5,000 individuals playing at the same time. The 31-day estimate is for labeling each of the images with a single keyword. In six months, the authors anticipate that each image could be labeled with six keywords.

Phetch is a game that produces natural language descriptions for images [30]. Natural language descriptions can be more useful than keyword lists, particularly for describing complex images. Natural language descriptions may be used to describe images to individuals who are blind. In each round, one player, called as the Describer, is shown an image. The Describer provides a textual description for the image to help the other players, who are called Seekers, select the image from among many different images using a search engine. Points are awarded to the Seeker who finds the image. The authors estimated that 5,000 individuals playing the game continuously could create captions for all of the images indexed by Google in ten months.

News Aggregation Websites
Slashdot [23] and Digg [7] are two examples of news aggregation websites. These websites maintain lists of user-submitted stories. Stories typically consist of a web link and a short text description. Moderation systems filter the user-submitted stories, removing duplicate entries and ranking them. Slashdot inspired the Digg moderation system. Digg now receives more than forty million unique monthly visits

[7]. The traffic generated from a link that appears on the main page of Digg or Slashdot is often sufficient to overload the web server that is linked.

The Slashdot moderation system consists of users, moderators, and meta-moderation. Users submit stories. Moderators assign descriptors to stories. Descriptors include: normal, offtopic, flamebait, troll, redundant, insightful, interesting, informative, funny, overrated, and underrated. Each descriptor is associated with numeric values. A score for the story is determined by calculating the sum of the scores associated with the assigned descriptors. This same information is also used to generate a characterization of the story, such as "20% insightful, 80% funny". The user who submitted the story has their karma value (a measure of their reputation on the site) updated based upon score for the story. Moderators are assigned semi-randomly. The algorithm selects a subset of the users who view comments associated with a new story the opportunity to act as a moderator for that story. Meta-moderation is a process whereby users are selected to review the correctness of eight to ten moderation decisions. Meta-moderation assigns scores to moderators, affecting their karma values.

Digg allows users to vote for stories. Votes are called diggs. Users can also vote to "bury" a story. Stories with a large number of diggs (typically over one hundred) appear on the main page of the website. Stories with a very large number of diggs (typically over one thousand) appear in a special section of the main page that is updated less frequently. This provides increase exposure for stories with a large number of diggs. uSocial advertises the opportunity to pay a fee to have a story appear on the main page for Digg [24].

Wikipedia
Wikipedia is an online, user-moderated encyclopedia [32]. Any individual with Internet access can act as an editor, modifying encyclopedia articles even if they have not created a user account. The accuracy of the encyclopedia is maintained by a set of policies and guidelines that are enforced by volunteer editors who act as moderators.

Wikipedia's policies and guidelines include written rules for resolving disputes and for working with other editors. Articles are expected to be written from a neutral point of view, contain only verifiable information, and not include original research. Each Wikipedia article has an associated discussion page that acts as a forum for editors to organize and debate revisions.

Editors can place notices on articles to advertise that they do not appear to follow policies and guidelines. Editors can also indicate which articles they are planning to work on next, to avoid overlap. By creating a watchlist, an editor can quickly survey recent changes to articles that are of interest. Editors receive commendation in a number of different forms. Some editors have user pages on Wikipedia to which other editors can post messages or award them virtual

badges. An internal peer-review process is also used to label some exemplary articles as "Featured Articles".

Some editors are provided with access to restricted technical tools. These editors are also called administrators. Administrators are able to reverse edits to articles and remove editors. Administrators are also called upon to mediate disputes between editors. Editors can request a comment from an administrator, or request arbitration, depending upon the severity of the dispute.

Wikipedia has many policies in place to address vandalism. Individuals who have not created an account are not able to edit certain flagged articles, due to high rates of vandalism. IP addresses that have been used by vandals are also regularly banned. Some high profile articles require an editor to have a certain "edit count" (a numerical measure of their editing experience) before they are permitted to make modifications. Edit count refers to the number of revisions that an editor has made that have not been reversed.

A history flow visualization of Wikipedia edits has been developed to help researchers identify editing trends [27]. This visualization has been used to identify four patterns of cooperation and conflict: vandalism and repair, anonymity versus authorship, negotiation, and content stability. Vandalisms are categorized mass deletion, offensive copy, or phony copy. The authors indicate that "there is no clear connection between anonymity and vandalism" [27]. As of May 2003, anonymous editors had conducted approximately 31% of the edits to Wikipedia. The negotiation pattern refers to sequences of conflicting edits in which two or more editors effectively revert each other's revisions. Finally, the authors also note that most edits consist of insertions or deletions, rather than moving text within articles.

Novice Wikipedia editors primarily locate information and fix mistakes [4]. They often begin editing in order to fix a perceived shortcoming in an article. Experienced Wikipedia editors, often referred to as Wikipedians, typically view themselves as members of a community. Experienced editors who are not administrators often still perform administrative roles. For example, answering novice editors' questions at the help and reference sections within Wikipedia.

Analysis
These examples demonstrate that computer systems can be used to coordinate many humans, each performing a small task. Further, the humans who perform these computer-mediated tasks are primarily motivated by the task itself. They are not provided with compensation, beyond acknowledgement of the work they have performed.

The task does not need to be of great importance to a human for it to be performed. It seems likely that the easier the task is to perform, the less important the task can be and still be performed. Novice Wikipedia users are often enticed to edit articles when they find a mistake and then notice the "Edit this page" link. The accessibility of the task is sufficient to engage participation.

The Wikipedia and Slashdot user communities demonstrate that users can moderate themselves. Moderation can require a higher level of engagement than performing a task that is more directly applicable to an individual. Experienced users have a higher level of engagement with these websites. This is a reason to nurture experienced users.

Rewards and performance tracking may be effective methods to foster a sense of identity in a community of users. By helping users identify with the community, they may be encouraged to continue to participate, increasing the number of experienced users.

It can also be important to design systems to filter out intentionally inaccurate user submitted information. The ESP game verifies the accuracy of keywords by comparing the keywords created by different pairs of users who viewed the same image. Digg compares stories to other stories that have previously been submitted. If the match is exact, the story is not accepted. If there is a partial match, the submitted is prompted to review similar stories and verify that there is a difference. The submitter's privileges may be restricted if the story is later flagged as a duplicate by other users. The Wikipedia system's moderation tools allow both users' accounts and IP addresses to have restricted access privileges.

The benefit to the user can be different than the benefit to the owner of the system. The ESP game is an example of such a system. This approach requires designers to address potentially divergent interests between the users and the system owner. It is also expected that this approach would be more difficult to scale to more complex tasks. More complex tasks might be more difficult to translate into fun games. However, designers may also find that there is a happy medium for many tasks, in which users may be willing to perform useful tasks that they also find moderately enjoyable to complete.

Designers who are evaluating whether to create a non-competitive direct motivation task might want to consider:

- The difficulty of the task

- The accessibility of the task for the humans who might be willing to complete it

- Methods to filter inaccurate user-submitted information

- Rewards that might be offered to encourage user-moderation

NON-COMPETITIVE INDIRECT MOTIVATION TASKS

reCAPTCHA
A CAPTCHA is a Completely Automated Public Turing test to tell Computers and Humans Apart [13]. CAPTCHAs

are used to verify that a user is human. Google, Yahoo, and many other providers of free email accounts require individuals to complete a CAPTCHA as part of their sign up processes. This step provides some confidence that a human, rather than a machine, is completing the sign up process. It also makes it difficult to create a large number of email addresses at once, which increases the cost of sending spam email messages. Most CAPTCHAs are images of words or phrases that have been distorted so that computer vision algorithms will be unlikely to be able to correctly identify the text.

reCAPTCHA is a web service that harnesses the human mental effort that is required to decipher a CAPTCHA [21]. Each reCAPTCHA image consists of two images that have been distorted so that they are difficult to read. One of the images is a word that is known to the system. The other image contains a word that is not known to the system. The word that is known to the system acts as the control. If a user does not enter this word correctly, they are considered to have failed the CAPTCHA. The word that is not known to the system is a word that is being transcribed. The reCAPTCHA system compares the text that different users entered. If a number of users have entered the same text for an unknown image, this text is considered to be the transcription for that image. In an evaluation, reCAPTCHA was found to be more than 99% accurate. A standard Optical Character Recognition algorithm was found to be 83.5% accurate on the same data set.

The images are presented in random order. This prevents users from being able to determine which word is the control. Users are also able to indicate that they would like to try a different set of images. This allows users to opt-out, rather than entering arbitrary text if they are unable to identify a word. It also provided an indicator that an image may be unreadable. These features may reduce inaccurate transcriptions.

If six individuals opt-out on an image, it is considered unreadable, and removed from the set of words to be transcribed. In an evaluation, only 4.06% of the images required six or more users to enter the same text for the transcription to be accurate. A post-processing step corrects text for predictable humans errors. These errors include transposing letters, incorrect capitalization, failing to enter a space between the two words, extra letters accidentally appended to words, and Non-English characters.

reCAPTCHA is used on over 40,000 websites and has transcribed over 440 million words [21]. CAPTCHAs are necessary on many different websites. By creating reCAPTCHA as a web service that can be embedded on any website, the designers have managed to harness to the mental effort of a large number of individuals. These individuals are willing to perform this task for the indirect benefit of verifying that they are human.

Online Surveys with Participation Incentives

Online surveys are a standard method to gain insights from prospective customers and other target audiences. Many online surveys provide individuals with participation incentives. These surveys are examples of non-competitive indirect motivation tasks. Search engine listings, banner advertisements, and newsgroups are common methods to attract survey participants.

One study found that search engine listings were significantly more successful than banner ads and newsgroups at attracting individuals to complete a survey for a chance to win a mobile phone [20]. The authors found that their newsgroup postings were sometimes considered unwelcome even if they were not off-topic. Less intrusive approaches that are only likely to be noticed by interested parties are recommended. Of course, these approaches may also amplify effects from non-random sampling.

Advertisements that mentioned the incentive were three times more likely to attract a participant. However, even when the incentive was mentioned, the highest response rate was less than one in five thousand. This study also found that females were significantly more likely to respond than males.

Another study, which tracked participants' email addresses, found that 4% of the responses were included the duplicate email addresses [5]. Researchers have also found evidence that individuals completing mailed print questionnaires may be more motivated but will not necessarily do a more thorough job 19.

Volunteers

Many websites facilitate the exchange of information between users. Some websites facilitate the organization of volunteers. One example is a website that was created to support the completion of an image search task for a missing aviator named Steve Fossett [33]. A second example is a website that was created by the The Guardian to crowdsource the identification of government expense documents that might be of interest to the general public [25].

More than 50,000 volunteers helped search for Steve Fossett by viewing and flagging satellite images of the 17,000 square-mile area in which his plane is believed to have crashed [34]. The website which hosted this distributed search task was built using Amazon's Mechanical Turk web service [17]. Each of the more than 300,000 images was viewed over ten times. The satellite images on the website were updated as new images became available. This online search was ultimately unsuccessful. Afterwards, one volunteer said "It was so exciting and new when we started it and it seemed like it could really help them, but eventually it was disheartening, and I realized I had no idea what I was actually looking for". A Major in the Civil Air Patrol said that the online search "added a level of complexity that we didn't need, because 99.9999

percent of the people who were doing it didn't have the faintest idea what they're looking for" [33].

Following media coverage of a major expense account scandal, The Guardian, a national British newspaper, downloaded over two million printed documents and receipts that the government made available online. The Guardian paid one software developer for one week to develop an online system that would allow Internet users to volunteer their time identifying receipts that might be of interest to the general public [25].

The website allows users to search for documents by postcode, or Member of Parliament or constituency. Users click one of a set of buttons after viewing an image of a document. The buttons allow the user to indicate the document type (claim, proof, blank, or other) and whether it may be interesting (not interesting, interesting, interesting but known, investigate this!). The main page features a progress bar and statistics about the number of documents that have been reviewed.

More than 20,000 volunteers have reviewed the documents using this system [25]. In the first 80 hours, over 170,000 documents were reviewed.

Analysis

reCAPTCHA demonstrates that tasks which humans perform can be modified to provide additional human computation data. The Internet allows for easy integration of web services. There may be many other opportunities to harness existing human computations. For example, Internet users who use social networking sites spend many hours browsing photos of their friends. This browsing data could be used to rank interest in the images.

It is likely that only a small percentage of advertising impressions will be effective. Providing an incentive to prospective online survey participants may be an effective method to attract more people. For these reasons, when possible, it may be more effective to use other approaches to entice participation. For example, building the task into an existing task, encouraging free word-of-mouth advertising by creating a community around performance of the task, or integrating the task an as optional activity on web sites that may attract people who are interested in performing similar tasks.

The volunteer examples illustrate that humans are willing to work together to perform computer-mediated tasks that may help them achieve a goal. Some people seem to have significant trust in the capacity of CHC systems to solve problems. This trust may not yet always be warranted, but compelling applications exist. The relatively low costs required to create a CHC system, along with the strong interest from the general public to participate in solving problems, may enable many more applications in the near future.

Designers who employ indirect motivation may approach web site owners or provide incentives to entice users to complete tasks. A web site owner may have sufficient leverage to convince a large number of individuals to complete the task. It is also possible to convince a large number of users to perform a task by providing an incentive. Incentives can be monetary or physical prizes, or more altruistic outcomes that may appeal to potential volunteers. Incentives can be an effective method to engage a large number of participants in a short period of time.

Designers who are evaluating whether to create a non-competitive indirect motivation task might want to consider:

- Existing tasks that might be modified to also achieve a desired CHC goal

- Providing an incentive

- Tasks that can be associated with major media events may be more likely to attract large audiences

- Response rate may improve if less intrusive advertising approaches are used

- Monitoring the results: people may participate even if their actions are not effective

COMPETITIVE INDIRECT MOTIVATION TASKS

Mechanical Turk

Mechanical Turk is a web service that provides a marketplace for Human Intelligent Tests (HITs) [17]. A Human Intelligence Test is a task that humans can perform more cost-efficiently than computers. HITs include image and audio processing and subjective rating and ranking tasks. Mechanical Turk is advertised as being well suited to handle photo and video processing, data verification, information gathering, and data processing tasks.

Several companies outsource aspects of this business to Mechanical Turk. CastingWords [5], an audio transcription company that has been employed by the Wall Street Journal, hires and manages freelance transcribers and editors using Mechanical Turk. Most HITs pay users less than 5 cents USD. Amazon, the company that owns and operates Mechanical Turk, is paid a 10% commission on each HIT.

Amazon provides APIs for companies to integrate their online systems into Mechanical Turk. This allows the process of creating and approving HITs to be automated. Once a user has accepted a HIT, they are typically given a short period of time (less than an hour) to complete it. Creators can specify qualifications that a user must have before they can accept a particular HIT. Qualifications are earned by completing online tests called qualifiers. Users can complete qualifiers on the Mechanical Turk website. These tests allow users to demonstrate their capacity to perform specific tasks. For example, there are qualifiers that

test users' abilities to transcribe audio, to label images with keywords, and to write Java source code.

The creator of a HIT has the right to decide whether or not to pay the user who accepts the HIT, regardless of the user's performance completing the task. A HIT that has been indicated by its creator to be successfully performed is called an approved HIT. Online message boards have been created to provide experienced Mechanical Turk users, often called turkers, with venues for rating companies that post HITs [27]. Turkers also use message boards to warn each other about scams.

Turkers warn each other about HITs that require personal information to be entered. HITs requesting users to click on the banner advertisements displayed on particular web pages have also been created. Spammers have created HITs that require users to sign up for email accounts and then share their passwords. Turkers have also indicated that HITs have been created for the specific purpose of completing CAPTCHAs.

Over 100,000 users from over 100 countries have performed HITs. There is data indicating that there are more female users than male users, and most users are under the age of 40 18. Form posts and articles written about turkers indicate that most users earn significantly less than minimum wage by performing HITs. Despite the relatively low pay, most Mechanical Turk users are located in the United States [18].

Reasons for participation vary. Turkers who have been interviewed have cited many different reasons for participating. These reasons include: the ease of performing the tasks while in distracting environments, the ability to earn some money in one's spare time at home, and preference to work rather than watch television in their free time [22].

Little et al. evaluated models for HITs that are iterative steps in the refinement of a solution [14]. Tasks included improving the descriptions of images, improving a letter to (better) convey an outline, deciphering handwriting, and sorting images. Instructions requested users to complete small improvements towards goals. Goals included "make it succinct", "use the present tense", and several others [14]. The paper discusses the possibility of automating the selection of different goals based upon other HITs that moderate the progress made between the iterative steps.

The authors were initially concerned that turkers might try to game the system. Their first evaluations include review HITs that required a majority of the users to agree that tasks had been completed properly. This review process did not prove to be necessary. Subsequent evaluations instead only required users to have a 90% approval rating on previously performed HITs.

Once submitted to Mechanical Turk, review HITs were generally completed in 7-15 minutes and improvement HITs in 15-30 minutes. Many of the results from the evaluations were included in the paper, indexed by iteration number. Although the quality of the results was not formally evaluated, the iterative improvement method appeared to be effective.

InnoCentive

InnoCentive is a web service that provides a marketplace for companies to outsource research and development work to individuals [9]. Similar to Mechanical Turk, companies, called seekers, post problems and individuals, called solvers, post solutions to the problems. Innocentive collects a fixed fee once a solution is selected. Proctor & Gamble and Eli Lilly are companies that have posted multiple problems on InnoCentive. Awards for solving problems are typically between $10,000 to $100,000 USD. InnoCentive has paid out over $3.5 million in awards to over 300 solvers.

Once solvers have submitted solutions to a problem, the seeker evaluates the solutions and selects the one that best meets the stated criteria. Seekers are typically given 2-6 months to evaluate solutions. InnoCentive reserves the right to "audit the decision making process of the Seeker on any specific awards where there might be an issue or question around payment" in order to ensure a fair outcome is reached [9].

InnoCentive supports four different types of problems: ideation, theoretical, reduction to practice (RTP), and electronic request for proposal (eRFP). Ideation problems are brainstorming problems to come up with new market opportunities, such as a product or service. The seeker received a non-exclusive perpetual license to use all of the submitted solutions. These solutions are typically two pages in length.

Solutions to theoretical problems are typically awarded larger payments than ideation solutions. In most cases, the terms of the agreement require the solver to transfer or license the intellectual property rights. Solvers are provided with technical evaluations of their submissions regardless of their selection by the seeker.

RTP problems require a precise description and evidence of solution optimality. These solutions typically require the most time to prepare and have the largest cash awards. eRFP problems do not have cash awards. Terms are directly negotiated between seekers and solvers.

Analysis

Mechanical Turk features a very different form of competition than InnoCentive. A Mechanical Turk users' performance is often measured by the number of HITs that he or she has had approved and rejected. Many HITs require users to have completed a certain number of HITs successfully and to have a minimum ratio of approved to rejected HITs. Users scan the Mechanical Turk website, racing each other to be the first to accept easy or profitable HITs. Competition is primarily on time to find and time to

complete HITs. Most users appear to be motivated by the opportunity to perform a useful task in their spare time, rather than, for example, watching television, and to earn a relatively small amount of money.

Process of producing and accepting an InnoCentive solution is complex. Solutions can require technical evaluation by experts and months to prepare. Competition is not on the basis of time to find or complete, but instead on the quality of the product. Most users are motivated by the opportunity to earn a considerable amount of money and to receive credit for having performed a task of significant value to a large corporation.

Mechanical Turk demonstrates that a large number of individuals are willing to perform simple tasks for less than minimum wage. Many Mechanical Turk tasks are created by dividing larger tasks into smaller tasks, moderating fulfillment of the smaller tasks, and then combining them to create a solution to the larger task. It is possible that this approach might also yield useful results for problems that are submitted to InnoCentive. On Mechanical Turk, multiple users could be working on different aspects of the problem at the same time. For some problems, this approach might also significantly reduce the time required to receive an acceptable solution.

Designers who are evaluating whether to create a competitive indirect motivation task might want to consider:

- High paying or low paying tasks

- Opportunities to reduce large tasks to sequences of smaller tasks

- Costs of paying users

- Opportunities to get results without paying users

COMPETITIVE DIRECT MOTIVATION TASKS

Google Search Ranking

Google search results appear in a specific order, determined by a web page ranking algorithm. The ranking of a web page is partially dependent on the number of other web pages that link to it, and the rankings of the web pages that link to it [3]. HTML links from web pages that are highly ranked are more influential in determining the ranking of the web pages to which they link.

The ranking of a web page in Google's search results is often important to the web site's owners. Web sites that sell products or feature advertising, among others, have a strong incentive to attract as many visitors as possible. Companies create, or remove, HTML links to improve search result rankings [35]. For these reasons, Google search ranking is a competitive, direct motivation task.

Many techniques have been developed to improve search result rankings. 302 page hijacking and link farms are two examples of search engine index spamming. 302 page

hijacking can trick web crawlers into believing a particular web address displays the same content as another web address. Since Google removes pages that contain duplicate content from its results, 302 page hijacking can cause a legitimate web address to be replaced with an illegitimate address. A link farm is set of websites that include web links to each other. This technique can inflate the rankings of each of the pages. Some embodiments of this technique disguise the exchange of links by having only some of the set of websites link to any specific other website. Automated programs have been developed to create and maintain link farms [35].

Yahoo! Answers

Yahoo! Answers is a web site that allows users to ask and answer each other's questions [37]. A user can only ask a few questions before he or she must answer some questions in order to be allowed to ask more questions. The specific number of questions that must be answered is dependent the quality of their answers, as points are awarded based upon other users' grading of the answer. In the default cases, 2 points are earned for answering a question and 5 points are taken away for asking a question.

When a question is created, it is categorized using a fixed, three-level hierarchy. The categories cover a broad range, including makeup and mathematics. Users who ask or answer questions in specific categories will be prompted to answer questions in those categories. The creator of a question selects one of the answers as the best answer. That answer is awarded additional points. The creator can also increase the points that are awarded to the chosen answer by offering some of his or her points. This may increase the number and quality of answers. All users who have at least 250 points can vote either for or against any other users' answers. These votes, and other factors, such as how often the user visits the site, determine the number of points they are given [36].

Each user is assigned a level, depending upon his or her number of points. A user's privileges increase as their level number increases. For example, a user with 5,000 or more points is not subject to any daily limits on the number of questions asked or answered.

Adamic et al. found that Yahoo! Answers users who focused their answers in fewer categories tended to have answers selected as the best answer more often [1]. Categories favoring factual answers were found to be more likely to have fewer and shorter answers. Users participating in these categories were also found to be unlikely to both ask and answer questions in the same category.

Liu and Agitchein found that as Yahoo! Answers has grown the complexity of the questions which are asked has increased [15]. They also found that users are becoming more likely to be passive, voting on each other's answers rather than answering questions. Further investigation

would be required to determine the overall effect on the quality of the answers. Bouguessa et al. have taken a step in this direction by creating an algorithm to identify authoritative users [2]. These authoritative users were demonstrated to "contribute significantly to the generation of high-quality content".

Analysis

When users are directly motivated to be competitive, it may be especially important that there is a robust heuristic for gauging the quality or accuracy of task performance. Indirectly motivated users of competitive systems are not primarily motivated to compete. Users of competitive direct motivation systems may be more likely to consider competition to be their task.

Google search results are subject to carefully researched and organized collusion among webmasters. Web links that are created for the purpose of manipulating search result rankings, rather than directing web site visitors to related content, can be considered noise on the signal that is interpreted by PageRank. The success of Google suggests that the search algorithm is sufficiently robust to filter out most of the noise in this signal.

Yahoo! Answers allows the creator of a question to indicate which of the answers is best. If the purpose of the system is considered to be answering each user's questions to their satisfaction, the selection of a best answer (or decision not to select) may be a highly robust heuristic. Of course, it is possible that the user has unknowingly selected an inaccurate or less accurate answer. One advantage of the voting system is to allow other users to help the question creator select from among the answers. Users might collude to vote up an inaccurate answer, however, the selection decision is still entirely within the control of the question creator.

In the case of Google search, the robust heuristic is an algorithm. The algorithm appears to be effective because it models the primary reason that web links appear on most web pages. Most web links are created to provide web site visitors with access to related content. In the case of Yahoo! Answers the heuristic is a moderation system. This heuristic appears to be effective because it is relatively easy for a question creator to judge content quality.

Designers who are evaluating whether to create a competitive direct motivation task might want to consider:

- How collusion may be controlled

- If there is a robust heuristic for measuring quality/accuracy

    o If the heuristic is an algorithm, the accuracy of the use case model

    o If the heuristic is user moderation, how difficult it may be to judge quality

CLASSIFICATION ANALYSIS

Direct and Indirect Motivation

Designers who employ indirect motivation may approach webmasters or users with incentives to increase user participation. CHC direct motivation examples can also use these methods, but it may be difficult to formulate direct motivation tasks such that indirectly motivated users will be effective participants. For example, the quality of Wikipedia articles would likely differ if users were paid to contribute. A considerably more complex moderation system might be required to prevent collusion.

Indirect motivation tasks may require different moderation. The Wikipedia, Digg, Slashdot, and Yahoo! Answers moderation systems are reliant upon experienced users. Users who participate in indirect motivation tasks may be less likely to be concerned with the community of users, as their primary reason for participation is an incentive. The moderation systems for most of the direct motivation tasks are optional. Users are not forced to moderate if they are not concerned about the community or the quality of its output. When moderation is required, the quality or the accuracy of the moderation may also differ between indirect and directly motivated participants. Tasks that use Mechanical Turk, an indirect motivation example, often feature multiple levels of moderation. Further, HIT creators have the right to reject users' submissions without explanation.

The success, and relatively inexpensive costs, of operating direct motivation tasks are a compelling argument for their use. Building the task into an existing task, encouraging free word-of-mouth advertising by creating a community around performance of the task, or integrating the task an as optional activity on web sites, may attract people who are willing to perform the task.

Indirect motivations, including incentives, may be an effective alternative when a larger number of participants than would otherwise be likely to perform the task are required within a particular period of time. Depending upon the task, indirect motivation may also require less effort to implement, as the user experience may not need to be enjoyable for participation to occur.

Non-Competition and Competition

Competition can be a useful task feature. Competition on Mechanical Turk decreases the time before HITs are completed. Most of InnoCentive's tasks are inherently competitive. The quality and number of answers to questions on Yahoo! Answers is at least partially dependent on the competitive nature of the task.

Systems that feature competition between users require robust heuristics for measuring the quality or accuracy of the users' contributions. The heuristic may include an algorithm or a moderation system. One approach to the design of heuristics is to create models of the system use cases. Some users may attempt to exploit the heuristic. The

heuristic must be able to extract the signal that it is intended to interpret from the noise generated by exploitation attempts.

Tasks that are not inherently competitive or that are reliant on experienced users are particularly vulnerable to be negatively affected by competition. Introducing competition to a non-competitive task can reduce the sense of community between users. Wikipedia and Digg are examples with non-competitive tasks and moderation systems that rely on experienced users. Moderation systems that are reliant on experienced users will also be more prone to manipulation by collusion if the task is competitive.

Motivation Interaction with Competition

Competitive indirect motivation tasks may be improved by conversion into competitive direct motivation tasks. These tasks can be significantly less expensive to operate, as no incentive may be required to encourage participation. The quality of users' contributions may also increase, as they will be more likely to be concerned with system performance.

In some circumstances, competitive direct motivation tasks may be improved by conversion to indirect motivation tasks. Designers who do not have a robust heuristic for measuring the quality or accuracy of user contributions, or webmasters who find that the heuristic is ineffective after the task has been created, might transform their tasks to be indirect motivation tasks. A new moderation system to control the distribution of the incentive can be introduced. If the indirect motivation incentive is sufficiently compelling, collusion may be reduced.

It can be difficult to directly motivate users to perform complex tasks. InnoCentive, which features highly complex tasks, provides incentives that are commensurate with the difficulty of the tasks. By dividing complex tasks into a large number of easier tasks, it may be possible to encourage communities of users who are highly motivated by a task to complete it using a non-competitive direct motivation system.

CONCLUSION

Four different classes of CHC tasks, with varying motivation (direct or indirect) and competition (competitive or non-competitive), have been described and compared. Considerations for designers and opportunities for future work have been identified.

In particular, methods to improve task performance by transforming complex tasks into many simple tasks should be investigated. Methods to encourage and support CHC contributions from users who are highly motivated to participate may also provide substantial improvements.

REFERENCES

1. Adamic, L. A., Zhang, J., Bakshy, E., and Ackerman, M. S. 2008. Knowledge sharing and yahoo answers: everyone knows something. In *Proc. WWW 2008*. ACM, New York, NY, 665-674.

2. Bouguessa, M., Dumoulin, B., and Wang, S. 2008. Identifying authoritative actors in question-answering forums: the case of Yahoo! Answers. In *Proc. 14th ACM SIGKDD international Conference on Knowledge Discovery and Data Mining* (Las Vegas, Nevada, USA, August 24 - 27, 2008). KDD '08. ACM, New York, NY, 866-874

3. Brin, S. and Page, L. 1998. The anatomy of a large-scale hypertextual Web search engine. *Comput. Netw. ISDN Syst.* 30, 1-7 (Apr. 1998), 107-117.

4. Bryant, S. L., Forte, A., and Bruckman, A. 2005. Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia. *In Proc. 2005 international ACM SIGGROUP Conference on Supporting Group Work* (Sanibel Island, Florida, USA, November 06 - 09, 2005). GROUP '05. ACM, New York, NY, 1-10.

5. Casting Words. http://castingwords.com/.

6. Cheyne, T. L. and Ritter, F. E. 2001. Targeting audiences on the Internet. *Commun. ACM* 44, 4 (Apr. 2001), 94-98.

7. Digg Site Ranking. http://www.alexa.com/siteinfo/digg.com.

8. Digg. http://digg.com.

9. InnoCentive. http://innocentive.com.

10. Kittur, A., Chi, E. H., and Suh, B. 2008. Crowdsourcing user studies with Mechanical Turk. *In Proc. CHI 2008*. ACM, New York, NY, 453-456.

11. Kosorukoff, A. (2001) Human based genetic algorithm. IEEE Transactions on Systems, Man, and Cyber-netics, SMC-2001, 3464{3459

12. Kuznetsov, S. 2006. Motivations of contributors to Wikipedia. *SIGCAS Comput. Soc.* 36, 2 (Jun. 2006), 1.

13. L. von Ahn, M. Blum, J. Langford, Commun. ACM 47, 56 (2004).

14. Little, G., Chilton, L. B., Goldman, M., and Miller, R. C. 2009. TurKit: tools for iterative tasks on mechanical Turk. *In Proc. ACM SIGKDD Workshop on Human Computation* (Paris, France, June 28 - 28, 2009). P. Bennett, R. Chandrasekar, M. Chickering, P. Ipeirotis, E. Law, A. Mityagin, F. Provost, and L. von Ahn, Eds. HCOMP '09. ACM, New York, NY, 29-30.

15. Liu, Y. and Agichtein, E. 2008. On the evolution of the Yahoo! Answers QA community. *In Proc. 31st Annual international ACM SIGIR Conference on Research and Development in information Retrieval* (Singapore, Singapore, July 20 - 24, 2008). SIGIR '08. ACM, New York, NY, 737-738.

16. McDaniel, S.W. and Rao, C.P. An investigation on respondent anonymity's effect on mailed questionnaire response rate and quality. *Journal of the Market Research Society 23*, 3 (1981), 150–160.

17. Mechanical Turk. http://mturk.com.

18. Panos Ipeirotis, *Mechanical Turk: The Demographics.* http://behind-the-enemy-lines.blogspot.com/2008/03/mechanical-turk-demographics.html.

19. Pitkow, J. E. and Kehoe, C. M. 1996. Emerging trends in the WWW user population. *Comm. ACM* 39, 6 (Jun. 1996), 106-108.

20. Pitkow, J. Recker, M. Using the Web as a survey tool: Results from the Second WWW User Survey 1995; http://www.igd.fhg.de/www/www95/papers/79/survey/survey_2_paper.html.

21. reCAPTCHA: Human-Based Character Recognition via Web Security Measures Luis von Ahn, Benjamin Maurer, Colin McMillen, David Abraham, and Manuel Blum (12 September 2008) *Science* 321 (5895), 1465.

22. Salon, "I make $1.45 a week and I love it". http://www.salon.com/tech/feature/2006/07/24/turks/.

23. Slashdot. http://slashdot.com.

24. Smart Company, *Facebook sends uSocial cease and desist warning.* http://www.smartcompany.com.au/web-20/20091123-facebook-sends-usocial-cease-and-desist-warning.html.

25. The Guardian, *Investigate your MP's expenses.* http://mps-expenses.guardian.co.uk/.

26. The Guardian, *MPs' expenses – what you've found so far.* http://www.guardian.co.uk/politics/blog/2009/jun/19/mps-expenses-what-you-ve-found.

27. Turker Nation Message Board System. http://turkers.proboards.com/.

28. Viégas, F. B., Wattenberg, M., and Dave, K. 2004. Studying cooperation and conflict between authors with *history flow* visualizations. *In Proc. CHI 2004.* CHI '04. ACM, New York, NY, 575-582

29. von Ahn, L. and Dabbish, L. 2004. Labeling images with a computer game. *In Proc. CHI 2004*. ACM, New York, NY, 319-326.

30. von Ahn, L., Ginosar, S., Kedia, M., Liu, R., and Blum, M. 2006. Improving accessibility of the web with a computer game. *In Proc. CHI 2006*. ACM, New York, NY, 79-82.

31. Wired, *The Rise of Crowdsourcing*. http://www.wired.com/wired/archive/14.06/crowds.html.

32. Wikipedia. http://wikipedia.org

33. Wired, *50,000 Volunteers Join Distributed Search for Steve Fossett.* http://www.wired.com/software/webservices/news/2007/09/distributed_search.

34. Wired, *Online Fossett Searchers Ask, Was It Worth It?* http://www.wired.com/techbiz/it/news/2007/11/fossett_search.

35. Wu, B. and Davison, B. D. 2005. Identifying link farm spam pages. *In WWW '05*. ACM, New York, NY, 820-829.

36. Yahoo! Answers Scoring System. http://answers.yahoo.com/info/scoring_system.

37. Yahoo! Answers. http://answers.yahoo.com/.