Computers in Human Behavior 49 (2015) 213–219

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Research Report

Quanty: An online game for eliciting the wisdom of the crowd

Wahida Chowdhury [a,*], Christopher Burt [b], Ahmad Akkaoui [b], Jim Davies [a]

[a] Institute of Cognitive Science, Carleton University, 1125 Colonel By Drive, Ottawa, ON K1S 5B6, Canada
[b] School of Information Technology, Carleton University, 1125 Colonel By Drive, Ottawa, ON K1S 5B6, Canada

ARTICLE INFO

Article history:
Available online 17 March 2015

Keywords:
Human computation
Wisdom of the crowd
Games
Quantitative magnitudes

ABSTRACT

Quanty is an online game that anonymously pairs players to estimate distances, weights, sizes, frequencies and such from photographs. The degree to which players agree determines the number of points they receive. We hypothesized that this game would generate more accurate aggregated estimates than would singular estimates by exploiting the wisdom of the crowd. Ninety-six participants (50 in group 1 using the metric system, and 46 in group 2 using the non-metric system) estimated height, weight, and distance of various objects; aggregated estimates of each group were more likely to approach accurate answers than were individual estimates, especially when the aggregates were calculated using medians and median absolute deviations. Also, the majority of participants thought that the game was as fun as the popular game Tetris. The results suggest that Quanty can be used to improve the judgment accuracy of professionals.

1. Why we need quantitative data about everyday things

Professionals increasingly rely on computer software to assist them in their practices. For example, doctors often rely on computers to make a diagnosis (see Gilbert & Lemke, 2014), and real estate agents might use computers to estimate the cost of properties (see Von Zur Gathen & Gerhard, 2013). Much of the software they employ also requires input of information. For example, a doctor might be asked to input a patient's weight, or a real estate agent might be asked to input the size of a house. Although professionals can often retrieve the requested information in existing databases, sometime they must resort to estimates based on their own mental models.

Suppose an engineer wants to investigate whether a land is good for installing wireless antennas; she is required to estimate the size of potholes from the photographs of the land. A particular study might be available by a company or a scientist that measured the approximate number of potholes in a given city, but such data are scattered across computers. Also many measures, such as the size of one specific pothole, seem too trivial to collect. The engineer in question might end up guessing, but how can she improve the accuracy of guessed estimates?

One possibility is to utilize *the wisdom of the crowd*, that is, to collect and aggregate estimates of several people. In fact, the Internet is a common platform that utilizes the wisdom of the crowd to gather estimates of things, such as oil, gas, or stock prices in the next month, the likelihood of a terrorist attack, the width of a brain tumor, or who would win the next presidential election. The first purpose of the present study was to design an online game that could be motivating and fun for people to provide their best estimates of various objects shown in photographs, and the second purpose was to analyze how the estimates could be aggregated to elicit the wisdom of the crowd and provide the most accurate answers.

1.1. How wise is the wisdom of the crowd?

The very idea about using the wisdom of the crowd came when a renowned British elitist, Francis Galton, discovered in 1906 that the median guess (1207 lb) of a randomly selected 800 uneducated commoners, about the weight of an about-to-be slaughtered ox at a village fair, was within 1% of the ox's true weight, 1198 lb (Surowiecki, 2005). Since then, many researchers attempted to utilize the wisdom of the crowd to solve complex problems. For example, Nickerson et al. (2009) investigated whether or not crowdsourcing, that is gathering people's independent judgments, could effectively match solutions to problems. The researchers asked students and Internet users to match several problem situations (for example, job search) with likely solutions (for example, social networks), and found that the aggregated wisdom of their crowd was as good as experts. Similarly, Steyvers, Lee, Miller, and Hemmer (2009) elicited the wisdom of the crowd by asking a group of people to order problems, such as listing chronologically the US Presidents, or ranking cities according to their populations.

* Corresponding author. Tel.: +1 613 883 6248.
E-mail address: wahida_chowdhury@carleton.ca (W. Chowdhury).

http://dx.doi.org/10.1016/j.chb.2015.03.004

Surowiecki (2005) outlines four characteristics of the crowd that generates close to accurate wisdom:

1. *Diversity:* The crowd is formed of diverse people with their own knowledge or bias.
2. *Independence:* People in the crowd provide their judgments independently of others.
3. *Decentralization:* People in the crowd are allowed to draw on their own knowledge.
4. *Aggregation:* A mechanism exists to aggregate individual judgments into a collective decision.

If these four characteristics of the crowd are met, people's estimates can be modeled as a probability distribution with a central tendency near the true value of the quantity to be estimated. Large numbers of people are often not needed to improve estimates. For example, in the popular game show *Who Wants To be a Millionaire*, studio audiences are often enough to derive more accurate answers than the answer of an assumed expert. Also, one study showed that the average value of independent guesses of as few as three people produced estimates reliably closer to reality than did just one (Lee & Shi, 2010).

Surowiecki proposes several kinds of problems that could be solved by such a crowd, but for our purposes we were interested in how best the crowd could solve cognitive problems, where each person provides independent estimates of things. But how can we gather diverse people at one place and collect independent and decentralized estimates from them? How can we motivate each person in the crowd to provide their best estimates, and at the same time, make sure that the experience is fun so that they spend time in providing estimates in the first place? One possibility is to construct and deploy an online, serious game.

1.2. Can serious games elicit the wisdom of the crowd?

Serious games refer to those games that are fun, and at the same time, "educational, engaging, impactful, meaningful, and purposeful" (Ritterfeld, Cody, & Vorderer, 2009, p. 3). Ritterfeld et al. reported that the five factors of games that are consistently found to influence the experience of fun are "overall game design, visual representation, audio representation, complexity, and diversity, and control" (p. 36). Most serious games however involve educational or skill training; the serious games that involve coordination or cooperation of players to solve a social problem is less prominent (Ritterfeld et al., 2009).

Von Ahn and Dabbish (2008) attempted social, serious games simply by asking regular Internet users to play an online game that had no external rewards, but were presumably intrinsically entertaining. The authors were able to gather huge amounts of data (labels for images) from diverse Internet users. Others picked up this idea and attempted to elicit the wisdom of the crowd simply by asking amateur Internet users, for example, to map the world (www.openstreetmap.de), or to create geospatial data by playing location-based games (Matyas, Kiefer, Schlieder, & Kleyer, 2011). Instead of explicitly performing quality checks, these games seemed to motivate people to play, by taking into account others decisions to solve a common problem. So can we deploy an online game to generate independent estimates of different objects?

Inspired by Von Ahn and Dabbish's (2008) Games With a Purpose, especially his ESP and Peekaboom games (see also Von Ahn, Liu, & Blum, 2006), we designed a serious, social game called *Quanty* as an enjoyable and competitive way to collect quantitative estimates of physical properties of objects in photographs. Previous studies show that Online Games are a fast way to elicit user's preferences while making it hard for the users to cheat (Hacker & Von Ahn, 2009). Quanty is deployed online, and to further ensure that users do not cheat, Quanty randomly pairs players to estimate quantities such as height and weight; If only one player is available in the game website, Quanty waits for a second player without starting the game. When two players are successfully paired, Quanty starts the game and instructs players that the closer their estimates are to the guesses of their partners, the higher the scores they will receive. Thus the players are assumed to be motivated to produce an estimate in a competitive situation. After players complete the game, estimates are statistically aggregated.

1.3. How to aggregate individual answers

The next question we considered is how to aggregate estimates of a diverse crowd so that the aggregated estimates could be close to accuracy. Most previous studies calculated simple mean or median of estimates (for example, how many jelly beans in a jar); Yi, Steyvers, Lee, and Dry (2012) developed aggregation methods that either combined individuals' judgments into a grand judgment, or identified judgments that is most similar to other individual judgments. Leys, Ley, Klein, Bernard, and Licata (2013) showed when aggregating judgments from a diverse group, it is best to calculate median absolute deviations of estimates (averaging after removing the estimates two standard deviations away from the median). Which method of aggregating is better?

We investigated which of four methods of statistical aggregation might produce the most accurate estimate: (1) simple averaging of all estimated values; (2) averaging estimated values after removing outliers (estimates two standard deviations away from the mean); (3) calculating the simple median estimate; and (4) Calculating Median Absolute Deviation or MAD.

2. Quanty: game design

Quanty is a web-based game, playable at the following URL: http://www.quantygame.com/. See Appendix A for the starting page with the 'how to play' instructions shown to a player. When clicked to start playing, players are randomly paired into teams of two. Players are matched into pairs randomly for the sake of anonymity. A given player does not know who his or her teammate is, nor can teammates communicate, so that they cannot cheat by agreeing on which numbers to enter. For example, if two players knew they would be partners, then they could agree to put the same number in for every question, giving them high scores and worse, bad data for professionals.

Both players are presented with a photograph. One or more objects in the photograph are outlined, each with a different color. The players are prompted with a question regarding some quantitative magnitude associated with the shown object(s), such as "how much does the radiator weigh?" or "what is the distance between the sidewalk to the building?" See Appendix B for an example question with the screenshot. Each player inputs his or her answer. The closer the two answers are, the more points both players get. This mechanism not only gives the game a way to generate a score, but also encourages the players to be as accurate as possible to score higher. This completes a single round. Then a new photograph is shown, and another round begins. Players continue with as many rounds as they can get through before the timer runs out after 3 min.

2.1. Photographs

Quanty in its current version uses a subset of the photographs and data freely available on the website LabelMe (Russell, Torralba, Murphy, & Freeman, 2005), (http://labelme.csail.mit.edu/) which offers an activity where users view photographs, click to trace the outline of an object in the photographs, and label that