Crowdsourcing as a Future Collaborative Computing Paradigm

Jie Wu, Chao Song, and Wei Chang

1 Definition and History

Crowdsourcing is a word coming from the combination of “crowd” and “outsourcing,” particularly through the Internet. Crowdsourcing coordinates a group of people called crowds to perform small jobs that solve problems that computer systems or a single user could not easily solve. Presently, crowdsourcing techniques are the key to turning smartphone sensing into a powerful tool and leveraging massive user engagement to gather data and perform required tasks. Advantages of crowdsourcing include, but are not limited to, reduced costs, higher efficiency, more flexibility, higher quality, better scalability, and more diversity. Crowdsourcing applications include virtual labor markets, tournament crowdsourcing, open collaboration, and others, including data donation [118].

According to the literature, several definitions of crowdsourcing have been proposed [7, 12, 14, 57]. Howe in [44] introduces the first definition of crowdsourcing, which represents the activity of an organization in outsourcing its tasks previously performed by employees to an external crowd of persons. Various incentive mechanisms have been proposed in crowdsourcing. Several crowdsourcing studies used Amazon’s Mechanical Turk [4] and its associated incentive mechanisms to

J. Wu (✉)
Department of Computer and Information Sciences, Temple University, Philadelphia, PA, U.S.A.
e-mail: jiewu@temple.edu

C. Song
School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China
e-mail: chaosong@uestc.edu.cn

W. Chang
Department of Computer Science, Saint Joseph’s University, Philadelphia, PA, USA
e-mail: wchang@sju.edu

J. Wu, E. Wang (eds.), *Mobile Crowdsourcing*, Wireless Networks,
https://doi.org/10.1007/978-3-031-32397-3_1

solve assigned tasks [51, 85]. Based on eYeKa, there has been a high adoption of crowdsourcing in business (e.g., 85% of the top global brands).

The remainder of this section reviews the basic components of crowdsourcing and gives an overview of several historical events associated with crowdsourcing. Section 2 provides several motivating examples of crowdsourcing applications, discusses tasks completed by humans via HPU vs. machines via CPU, and introduces the events of humans vs. machines in chess and Go games. Section 3 offers an overview of crowdsourcing workflow and different types of crowdsourcing. Section 4 lists several popular crowdsourcing platforms. Section 5 looks at several crowdsourcing applications. Section 6 discusses challenging issues with a focus on algorithmic and theoretical aspects. Section 7 examines future opportunities associated with crowdsourcing. Section 8 concludes the chapter by presenting our view on the human vs. machine debate through the lens of HPU vs. CPU.

1.1 HPU and CPU

The benefits of crowdsourcing are its inexpensiveness and fast speed. In addition, it supports the notion of the whole (i.e., crowd) being greater than the sum of its parts (i.e., individuals). Crowdsourcing can be viewed as consisting of Human Processing Units (HPUs) (i.e., human brains) that are analogous to CPUs in the traditional computer system [17, 77]. In this case, a small job executed by a worker equates to one instruction to the HPU, and the time of execution equates to the physical time in real life. For certain applications, HPU is more effective than CPU in the following areas: (1) verification and validation, such as image labeling, (2) interpretation and analysis, such as language translation, and (3) surveys, such as social network surveys. The networks formed by the interactions between humans (HPUs) and machines (CPUs) are termed as Human–Machine Networks (HMNs) [91].

HPU as human computation has emerged as a new paradigm of computation [17] relatively recently. In many applications, human computation is a useful complement to computer computation, and it enables tasks, like identifying emotions from human speech, recognizing objects in images, and so on, to become possible or to be done more efficiently. With the help of HPU, many data-intensive applications have been proposed, including (1) crowd-powered databases [31, 68, 80] and useful operators like filtering [79] and max [40], group-by [24], (2) various data processing technologies like image labeling [131], entity resolution [97], and schema matching [134], and (3) combinatorial problems like mining [6] and planning [65].

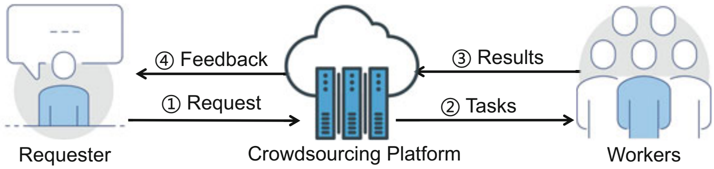

Fig. 1 Components of crowdsourcing

1.2 Basic Components

Crowdsourcing involves three components, which are requester, worker, and platform. Figure 1 shows the basic components of crowdsourcing as follows:

- **Requester**: A requester is a person who submits a request with a task to the platform. In Amazon Mechanical Turk (AMT), a requester can publish a Human Intelligence Task (called HIT) with some requirements, such as the price, the time constraint for solving, the expiration deadline, and the qualification requirement.
- **Worker**: Workers in crowdsourcing are people who will perform HIT issued by the requester. Workers either can choose tasks or are assigned to tasks by the platform. In some cases, the qualifications of workers need to meet the requirement in order to be eligible.
- **Platform**: The platform connects the requester and the workers. The main functions of the platform include assigning tasks to appropriate workers, filtering and merging multiple outputs from the assigned workers, setting up the rewards, and privacy projection for both the requester and the workers.

Note that the last step of the workflow is the final feedback from the platform back to the requester. Feedback can either be raw results or the processed result of raw results from the workers.

1.3 History

Even though crowdsourcing was proposed in 2006, the basic idea of using a crowd to solve a problem was applied in the early eighteenth century, and since then, crowdsourcing appeared in many applications. In 1714, the British government provided £20,000 to reward those who could solve the Longitude Problem [110], which killed thousands of sailors per year [89]. When a ship was damaged, its accurate location coordinates to help seamen are unavailable to be obtained due to the problem. The Longitude Prize was set up at that time for the public to find a method to measure the longitudinal position of a ship. This event is considered to be the first example of crowdsourcing, and John Harrison, an English carpenter, won the award.

The following lists several major events that can be considered crowdsourcing applications [22]. Note that before the invention of the Internet, information gathering in crowdsourcing was done through other means, rather than through the Internet. However, the Internet has lowered barriers to entry for crowdsourcing activities.

- The Oxford English Dictionary was written in 1884 based on a large crowd across the country to catalog English words [66]. Overall, 800 volunteers contributed to creating the original Oxford English Dictionary.
- Japanese car manufacturer Toyota held a logo designing competition held in 1936 [109] and chose the current logo from close to 30,000 candidate logos.
- Wikipedia, started in 2001, is a free content Internet encyclopedia open to the public, and it acquires collective crowd wisdom through crowdsourcing.
- TopCoder, a crowdsourcing software development company, was founded in 2001. It offers a platform for collaboration and competition.
- Threadless.com, started in 2005, has the members creating their own designs. It is considered to be the first crowdsourcing example in the modern era.
- YouTube is an example of crowdsourced entertainment, which was founded in 2005. Similarly, there are many large companies in Fortune 500 relied on crowdsourcing for various business-related tasks [86].
- In 2006, Howe and Robinson were the first to introduce the term “crowdsourcing” in the June issue in Wired.
- Brabham published the first scholarly work using the term crowdsourcing in 2008.

All the above examples show the power of crowdsourcing and the evolution of its applications.

2 Crowdsourcing Events in Recent History

In this section, we introduce four samples of crowdsourcing in recent history as motivation. These failed and successful samples include Help Find Jim Gray, Malaysia Airlines Flight MH 370, DARPA Network Challenges, and Tag Challenges. At the end of this section, we will discuss an important debate on human vs. machine through the lens of HPU vs. CPU.

2.1 Help Find Jim Gray (2007)

Jim Gray, Turing Award laureate, went missing when he sailed outside San Francisco Bay in January 2007. Dr. Gray is known most notably for his contribution to several major database and transaction processing systems. After 4 days of extensive searching, the search effort called off the effort. However, private efforts for searching for Jim Gray were ongoing for a while with the help of a special Amazon link to the search effort. This private effort involves collecting and reviewing individual satellite images to determine if there are any photos that warrant further investigation. However, all these efforts generated no results.

2.2 Malaysia Airlines Flight MH 370 (2014)

Flight MH 370 Malaysia Airlines appeared with a trace during the flight between Manila and Beijing on 8 March 2014. DigitalGlobe, a Colorado-based tech company, immediately posted over 3100 square kilometers of imagery that could be examined by the crowd. Within the first 24 h, thousands of volunteers were viewing 2 million pages of satellite imagery every 10 min. They tagged more than 60,000 objects that have the potential for further examination, and these results were made available to the public to review. However, the crowdsourcing search effect was not fruitful. In fact, the whereabouts of MH 370 remains a mystery.

2.3 DARPA Network Challenges (2009)

In 2009, DARPA, a US military agency, issued an interesting challenging problem for the public to solve. It called on participating groups to find the locations of 10 red balloons scattered in the air around the country (i.e., the USA). A total prize of $40,000 would go to the first participating team to find all balloons. This purpose is to test the effectiveness of social networking (including crowdsourcing) and web-based technologies that can complete a time-critical large-scale task. The MIT team won the competition by finding all 10 balloons in under 9 h with the help of social networking.

The winning team adopted a special incentive mechanism called multi-level marketing to recruit participants, with the prize money ($4000 per balloon) to be distributed in the chain of participants leading to an identified balloon, with $2000 allocated to the person who spots the balloon, $1000 (half of the remaining $2000) to the person who recruited the winner in the chain, and so on. For a short chain, the leftover funds will be donated.

2.4 Tag Challenges (2012)

The Tag Challenge in 2012 extends the idea of the DARPA Network Challenge, but in a much harder way. The objective of each participating team was to be the first to locate and photograph five volunteer “suspects” in five different cities in the world: New York City, Washington DC, Stockholm, London, and Bratislava (capital in Slovakia). Again, social networking and crowdsourcing play an important role; the challenge is how information propagates through social networks and what it takes for a message to spread out widely. The winning team from UCSD, which found 3 out of 5 individuals, included one team member from the winning team of the DARPA Network Challenge.

The winning team adopted a unique incentive mechanism for the award distribution: recruiters of the first 2000 recruits will receive $1 for each recruit. The success of the UCSD team gave some interesting insights into designing incentive mechanisms for crowdsourcing. The question is which factor plays a dominating role: money, personal satisfaction, or justice under a shared belief?

2.5 Kasparov vs. IBM Deep Blue (1997)

If HPU refers to human intelligence and capability, while CPU refers to computer intelligence and capability, which one will dominate in the future? This is a rather philosophical question that goes beyond crowdsourcing through people vs. problem-solving using computers. We focus first on a narrow field of human vs. machine in chess and Go games.

In May 1997, Deep Blue, a computer built by IBM, beat the then world-chess champion Garry Kasparov. If AI could beat the world's sharpest chess mind, it seemed that AI would overtake humans in everything. Still, some people felt that the chess game is still not sophisticated enough and AI may find it hard to beat humans in more complex games, such as the Go game. However, in October 2015, DeepMind technologies' AlphaGo became the first AI program to beat a professional Go player on a full-sized board. In 2017, AlphaGo beat the then No. 1 player in the world. In fact, the self-taught AlphaZero, without using any human knowledge, is currently the world's top player in the Go game.

2.6 A Big Picture: Human vs. Machine

Since the historical event of Kasparov vs. IBM Deep Blue in 1997, people introduced the *freestyle chess game* [50] that allows humans unrestricted use of computers during the games. This freestyle format introduced a powerful idea for the future mix of HPU/CPU and applications. In Freestyle chess tournaments dated 2005, there are four types of players: Grand-master (>2500 points), Machine (Hydra, then best AI chess machine), Grand-master + machine, and Amateur (>1500 points) + machine. The winner came from the last group, where an Amateur teamed up with the machine. How can such a team become the champion? Perhaps the amateur knows how to collaborate with the machine. He knew when and how to explore ideas, i.e., when to take a machine's suggestion and when to ignore it.

Note that a chess or Go game has a clear set of rules for achieving a well-defined goal. In addition, a player has all the information needed to make the right decision toward the ultimate goal. In many crowdsourcing applications discussed in this chapter, such information is vaguely presented or incomplete.

3 Crowdsourcing Overview

We give a quick overview of crowdsourcing basics in terms of workflow in crowdsourcing and three major types of crowdsourcing.

3.1 Workflow of Crowdsourcing

As shown in Fig. 2, the workflow of crowdsourcing can be divided into three steps: *preparation*, *execution*, and *termination* [22].

The preparation includes all preparation before a crowdsourcing task is given to the platform to be solved by a group of workers. First, the requester should design the task adequately and calculate the workforce needed. If a task is too complex, a divide-and-conquer approach can be used to partition the task into several small ones. Each small task should be easy to solve and independent of any other small tasks, which means the execution of a small task should not affect any other tasks. Once the requester creates the task suitably, incentives should also be designed properly so that a sufficient number of workers will sign up for the tasks.

The execution part starts after the preparation step. A requester needs to find the workers for his/her task. The requester recruits the workers through different crowdsourcing platforms. The selection of the workers is essential for the success of crowdsourcing tasks which depend on the quality of the results performed by these workers. Based on the requirements of the task, the requester via the platform will recruit workers with some skill levels and assign the sub-tasks to them.

Termination is the last step of crowdsourcing where various steps are executed to complete the task. After receiving all the results from workers, in addition to

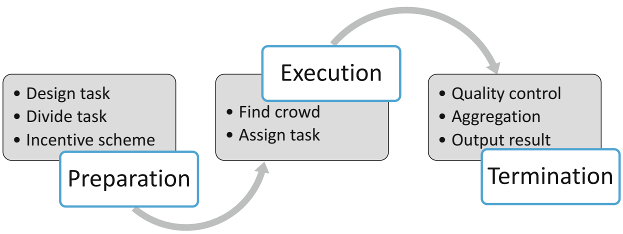

Fig. 2 Workflow of crowdsourcing

passing the raw results directly to the requester, in many cases, the platform refines the raw results by separating the outputs provided by ordinary workers from the outputs provided by experts, with the methods of quality control (such as majority voting). Finally, these outputs are merged and calculated to find the right result for the requester.

3.2 Types of Crowdsourcing

There are three major types of crowdsourcing based on the nature of applications.

- **Virtual labor markets**: this crowdsourcing provides a platform where workers can complete tasks for monetary compensation, e.g., Amazon Mechanical Turks.
- **Tournament crowdsourcing**: this crowdsourcing offers a platform for ideas competitions, where usually only the winner is compensated, e.g., CrowdFlower and TopCoder.
- **Open collaboration**: this crowdsourcing serves as a platform for collaboration to complete an assigned task. Typically, it does not provide monetary compensation; people who are willing to help are recruited through social media, e.g., Wikipedia.

According to [13], there is no consensus on the scope of crowdsourcing. There are nearly forty different interpretations of crowdsourcing. Here, we focus on the above three types of applications throughout this chapter when we cover the history, platforms, and applications of crowdsourcing.

4 Platform

Nowadays, there are many crowdsourcing websites and applications. Essentially, they provide a virtual marketplace where a requester can post a variety of tasks online and workers are able to seek, join, and complete some tasks at their discretion. In general, these websites and applications are called crowdsourcing platforms. This section introduces some important crowdsourcing platforms.

4.1 Amazon Mechanical Turk

Among all the existing general-purpose crowdsourcing platforms, Amazon Mechanical Turk (AMT) is the most famous one. It provides business owners and developers with an on-demand workforce at a small monetary cost. The idea of AMT was first mentioned by Venky Harinarayan in an Amazon patent filed on October 12, 2001

[1, 45]. On November 2, 2005, AMT was launched officially as part of Amazon Web Services (AWS). Since then, it has received worldwide attention, and by the year 2011, workers of AMT are from over 190 countries [56, 99].

AMT creates a flexible and convenient labor force market. According to AMT, each worker can freely select a HIT (i.e., Human Intelligence Task) based on the task's description, keywords, expiration date, the amount of reward, and time allotment. Some tasks even require a qualification test before selecting them. Workers are able to find their submission status and rewards once they complete the assigned workload.

The AMT platform supports various types of tasks. The following list summarizes six commonly used types and corresponding examples:

- *Content Creation*: Composing a description of a local spot
- *Data Collection*: Finding and collecting data that satisfies certain features
- *Labeling and Categorization*: Interpreting content, such as labeling an image
- *Surveys*: Taking certain actions based on the survey
- *Transcription*: Writing word descriptions from a given video
- *Verification*: Checking the authenticity of the information

The AMT platform is well maintained by Amazon and it is updated continuously. The current version allows requesters to build HITs in three different ways [58]: the web-based user interface, command-line tools, and APIs. AMT is rich in templates, and a requester can easily build his task page by using those templates. By the time of writing this chapter, there is a total of 11 project templates available on AMT.

4.2 Crowd4U

Unlike the commercial crowdsourcing platforms, such as AMT, whose internal technique details are hidden from the public, Morishima et al. [71] constructed an open, generic, and non-profit platform called Crowd4U [103]. Crowd4U was first opened to the public in November 2011, and by June 2022, there are over 2 million tasks that have been performed on Crowd4U.

One unique feature of Crowd4U [70] is its compatibility with other platforms: via an HTTP-based API, Crowd4U is able to retrieve data from others. Supporting complex crowdsourcing tasks is another feature of Crowd4U. It is achieved by using CyLog, a language that can implement complex logic and dataflows. Since Crowd4U is all-academic open, researchers can test various incentive mechanisms, task assignments, or recruitment strategies on it. Besides conducting research experiments, Crowd4U can also be used as a general-purpose crowdsourcing tool [123], and several interesting crowdsourcing tasks have been done on it, such as identifying the course of tornadoes. Due to all these special characteristics, Crowd4U gets a lot of attention, especially from academia.

4.3 gMission

gMission [21] is a smartphone-based platform, which focuses on spatial crowdsourcing tasks, e.g., detecting the crowdedness level in a cafeteria. It was designed by a group of researchers from HKUST. In gMission, smartphones are used as sensors, and tasks are bound to specific locations. To accomplish a task, each worker has to physically appear at a specific place so that his smartphone's internal sensors can collect data.

gMission consists of four components: an interface module, a data manager, a function manager, and a quality control module. Each module handles different aspects of the crowdsourcing procedure. The interface module is responsible for the user interface, the data manager stores data and action information, the function manager handles location-sensing tasks' management, task recommendation, and work allocation, and last but not least, the quality control module verifies locations and controls the life cycle of tasks. Unlike the existing commercial web-based crowdsourcing platforms, such as AMT, gMission is one of the first few attempts at smartphone-based mobile crowdsourcing systems. Its design inspired the implementation of other mobile crowdsourcing platforms.

4.4 UpWork

UpWork [115] (formerly called Elance-oDesk) is another well-known crowdsourcing platform. There are a few unique features that distinguish it from AMT. First, UpWork was designed to support job hunting and remote collaboration. A requester can interview, hire, and work with freelancers. Second, unlike AMT, where a task typically consists of a collection of questions, each task in UpWork is indecomposable. Third, UpWork uses a unique payment strategy by which the platform gets commission from the freelancers: in UpWork, a worker is allowed to set a price and submit proposals for jobs [75]. The platform connects the worker with employers. Once the task is done, the platform cuts a 5% to 20% service fee, depending on the total amount the worker billed the employer. Due to all those features, UpWork has become a very popular online freelancing platform. By 2017, UpWork had 14 million users from 180 countries [116].

4.5 CrowdFlower

In 2007, Lukas Biewald and Chris Van Pelt together founded CrowdFlower [105], an aggregator crowdsourcing platform that aims at involving human beings in machine learning. Unlike general-purpose crowdsourcing systems, CrowdFlower is mainly used by students and data scientists. Typical types of CrowdFlower tasks are data collection, transcription, sentiment analysis, labeling, etc. CrowdFlower asks the requester to appraise a task. Like UpWork, when the task is done, CrowdFlower takes approximately 20% of the payment as a commission [76]. What makes Crowdflower unique is its recruitment strategy: Crowdflower is partnered with multiple other platforms, and a Crowdflower task can be delegated to the workers on those platforms [25]. However, perhaps because of the delegation and commissions for the related platforms, some workers are discontented with their amount of payment [104].

5 Sample Applications

The crowdsourcing platform organizes the crowd workers with assigned tasks and provides services to different applications. In this section, we introduce the typical applications on image processing, commonsense knowledge, smart city, and other science projects, with a focus on software engineering and coding. Finally, we discuss several projects that use a mix of HPU and CPU, i.e., both crowd workers and traditional computer-based solutions.

5.1 Image Processing

For some applications, humans are better at labeling than machines [22]. This subsection discusses three applications of image labeling and classification.

The work in [130] focuses on real-time applications for searching for target images on smartphones. A requester sends a target photograph with buildings and some candidate photos to the platform and queries which candidate photos contain the same building in the target photograph. The workers receive the task with the target and candidate photographs and answer "yes" or "no" to indicate whether a candidate photograph has the same building as the target photograph.

For the challenge of classifying galaxies, astronomers hope to classify millions of galaxies (based on color and shape) in pictures taken by the Hubble Space Telescope [138]. While computers are good at recognizing colors in images, they find difficulty in recognizing shapes. However, recognizing shapes is very simple for humans. Imagine a case where scientists could use crowdsourcing in an entertaining game format and get help from ordinary people to classify these galaxies. This project (called Galaxy Zoo) was the starting point for the Zooniverse [119], a website that connects scientists with research collaborators and has since expanded to many other science projects [137].

Tohme [113] provides a customized interface to find the curb ramps from Google Street View scenes, by the technologies of computer vision (CV) and machine learning. Tohme dynamically schedules the two workflows of human (worker) labeling and computer vision with human (worker) validation according to the prediction of performance. The performance of Tohme has been verified with 1086 Google Street View scenes in four North American cities and 403 recruited workers.

5.2 Commonsense Knowledge

In many applications, it is hard for a machine to accomplish a given task, while it is common sense knowledge for human beings. Therefore, it can easily be solved through crowdsourcing [22].

Gwap [95] is a website with games for users to play, and the users collaboratively do the tasks provided in the games. For example, two users as the players are presented with a photograph in the game, and each user inputs a list of tags for that photograph. Thus, the system will feed back the points when one tag of the two users is matched. Therefore, the system can help to label the photograph with tags by the users as workers in crowdsourcing.

The technology of CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) displays the out-of-order sequence of letters for the users to transcribe, in order to distinguish whether the user is human or machine [94]. Moreover, by the method of CAPTCHA, the transcribing of old books can be performed by users around the world. To improve the performance of the CAPTCHA system, Google created and developed a new system reCAPTCHA in 2007. This system utilizes artificial intelligence and helps protect the website from malicious attacks by robots [108].

5.3 Smart City

A smart city [120] utilizes different types of information and communication technologies to increase operational efficiency through data collection and information sharing. The information gained from the smart city could help to manage resources and services efficiently to improve the quality of both government services and citizen welfare. Here, we discuss three particular areas that are related to crowdsourcing: traffic monitoring, trajectories of mobile users, and monitoring of air pollution.

Waze [117] is a crowdsourcing application that provides services for transportation. In Waze, the drivers share the events (such as accidents or traffic jams) where they meet along their trajectories. In 2013, Digital China [107] also proposed an "Integrated Citizen Service Platform" for the users to publish the events they encounter in the city. In 2021, Didi Chuxing [106] in China proposed a new application called "Long-Distance Eyes," which also allowed drivers to share traffic events with photographs or short videos to improve the travel plans of users and migrate traffic congestion.

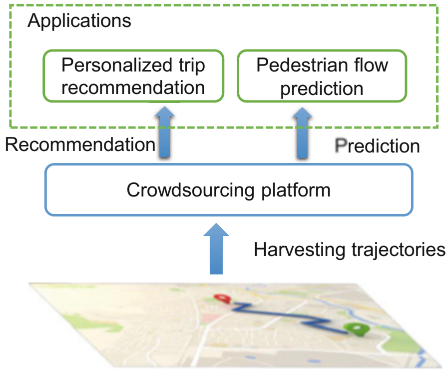

Fig. 3 Crowdsourcing platform harvests the mobile trajectories

The crowdsourcing platform harvests the trajectories of mobile users and provides services to the applications of individual users or business institutions, as shown in Fig. 3. In order to improve user experience in the personalized trip recommendation, the work in [39] fully takes advantage of the Foursquare dataset, a crowdsourced large-scale check-in LBSNs dataset, and discovers attractive routes to improve personalized trip recommendation. In detail, apart from POIs, the routes' connected POIs also attract visitors. These routes have many crowd flows and have high popularity. This kind of route is termed an Attractive Route and brings extra experience to users. Pedestrian flow prediction is used to help the operators make decisions. Moreover, the operators held events (such as sales promotions) to attract nearby crowds. This kind of event is termed a business event. This work in [38] investigates the impact of business events on pedestrian flows from the crowdsourcing trajectories and proposes a model for pedestrian flow prediction.

For air pollution problems (for example, PM2.5), many crowdsourcing platforms are proposed to monitor air conditions. Third-Eye [61] is a *crowdsensing* platform to monitor fine-grained PM2.5, which is developed by BUPT and Microsoft Research Asia. The platform utilizes the photographs taken by the users' smartphones and identifies PM2.5 levels of air conditioning by deep learning algorithms. The system helps the government to collect information on air conditioning and adopt related strategies for protection.

5.4 *Other Science Projects*

Crowdsourcing finds many applications in science [118] in the areas of astronomy, energy system research, genealogy research, ornithology, and seismology. Here, we discuss its applications in software engineering and coding in the field of computer science.

The crowdsourcing platform of software engineering recruits global software engineers, to execute various kinds of software engineering tasks, including requirements development, detailed design, programming, and testing. Many platforms implement such crowdsourced software engineering, such as TopCoder, uTest, and TestFlight [54, 67].

TopCoder [114] is a famous crowdsourced software engineering company, which has a process model called TopCoder Competition Methodology [67]. The platform decomposes complex software development into multiple sub-tasks. With a waterfall model, the development is divided into phases. After the online competitions with crowd developers, the qualified winning solutions of each development phase are accepted by the platform.

5.5 Mixed HPU and CPU Applications

This subsection discusses several applications that make use of a mixed HPU and CPU approach, i.e., crowdsourcing that draws resources from both humans and machines.

One of the early adoptions of both HPU and CPU is collaborative crowdsourcing language transition over the web. In general, the requirement of professional translators or the lack of bilingual speakers makes translation difficult and costly. In [5], a three-step cost-effective approach is used to minimize the cost of hiring bilingual speakers: (1) context-sensitive lexical translation by CPU only for initial word-by-word translation, (2) assistive translation by bilingual HPU who know both original and target language for accurate sentence-level transactions, and (3) target synthesis by monolingual HPU for final polishing in the target language.

A mix of HPU and CPU has found its use in databases. Databases give incorrect answers when there are missing data or when an understanding of the data semantics is required. In this case, humans can provide inputs for the missing information. In [30], an extended database called CrowdDB is provided that combines both CPU (for normal database functions) and HPU (for abnormal cases). CrowdDB uses SQL and can be operated on two platforms: AMT and custom-made mobile phones.

Nowadays, in medical imaging research, AI is the most discussed topic, both in diagnostic and in therapeutic contexts [81, 90]. The mix of HPU and CPU can start with CPU-based diagnoses for initial screening and elimination, followed by HPU-based decisions by medical doctors. AI can also help in precision medicine and risk assessment. However, fully machine-based approaches still have a long way to go for two reasons: (1) unlike a quantitative task, the knowledge of decision-making with medical imaging requires long-term experience and medical training. (2) There are legal implications when fully machine-based diagnostic and therapeutic treatments are adopted for medical imaging.

6 Algorithmically and Theoretically Challenging Issues

This section discusses challenging issues, with a focus on algorithmic and theoretical aspects that are related to computer science. The coverage is by no means complete. Some aspects are not included, such as privacy in crowdsourcing [27, 129].

6.1 Paradigm

The tasks to be accomplished can sometimes be complex and difficult to solve, and the requester takes a different approach to solving such the tasks, such as breaking the task down to reduce complexity [22]. There are three typical methods for task decomposition: (1) by sequential implementation, each task is partitioned into several small sub-tasks which are executed sequentially, and the output of a sub-task is set as the input of the next sub-task. (2) By parallel implementation, each task is also partitioned into several small sub-tasks which are executed in parallel, and the outputs of these sub-tasks will be aggregated to the final output for the task. (3) By the method of divide and conquer, the original problem of the task is divided into several smaller problems recursively which can be solved easily. According to the results of these solved smaller problems, the original problem can be solved and the final result is generated. Moreover, we will discuss multi-armed bandit methods and incentive mechanisms in crowdsourcing.

6.1.1 Sequential Implementation

In the method of sequential implementation, the original task is partitioned into several small sub-tasks, which are executed sequentially. The platform will select and execute these sub-tasks one by one until the last one is finished and generates the final result. The output of a sub-task is set as the input of the next one, as shown in Fig. 4.

For sequential execution of sub-tasks, [10] adopts identify, filter, and extract stages. Hirth et al. [42] first divides the original task into many sub-tasks by a method of the control group. Then, the platform assigns the sub-tasks to the workers

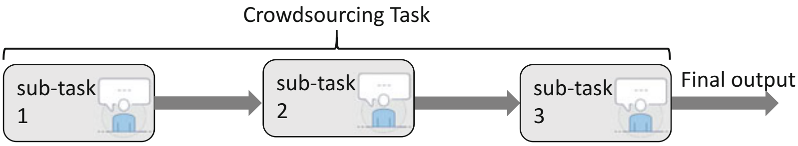

Fig. 4 Sequential implementation

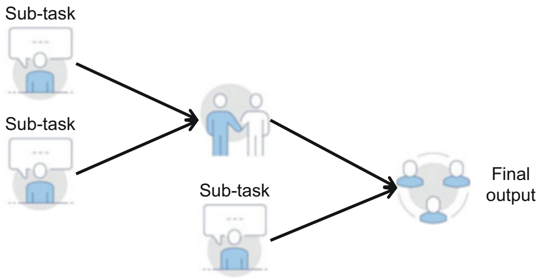

Fig. 5 Parallel implementation

for evaluation. At last, the platform obtains the outputs from the workers for the final output. The model in [135] has the sequential stages of training, refinement, and evaluation of the results. The platform in [74] first tags all the photographs of food. Then, each photograph is identified. At last, the qualities of the identified foods are evaluated.

6.1.2 Parallel Implementation

The method of parallel implementation divides a task into several small sub-tasks which are executed in parallel, and the outputs of these sub-tasks will be aggregated to the final output of the task, as shown in Fig. 5.

SCRIBE [55] is a system to provide instantaneous captioning for deaf people. The system recruits the captionists as workers from AMT [4] or quickTurk [11]. Each recording is divided into several parts for capturing by different workers. Then, the results are aggregated and sent back to the user by the server. The system in [92] recruits the users to take different photographs at different locations and then aggregates these photographs into a three-dimensional model. Moreover, the systems in [42, 74] also divide the original task into multiple sub-tasks, and the sub-tasks are executed in parallel.

6.1.3 Divide and Conquer

The method of divide and conquer recursively decomposes the problem of the original task into smaller problems until the problem cannot be divided anymore. Each problem can be solved easily, and the outputs of the problems will be aggregated for the final output, as shown in Fig. 6.

The method of divide and conquer in Turkomatic [3, 53, 72] has been used in many applications, such as labeling. First, the requester submits an original task to the platform. Due to the complexity of such a task, the platform recursively decomposes it into several smaller tasks which are much easier to be solved than the original one. Then, the recruited workers execute these sub-tasks and feed the outputs back to the platform. At last, the platform aggregates these outputs for the final result of the original task.

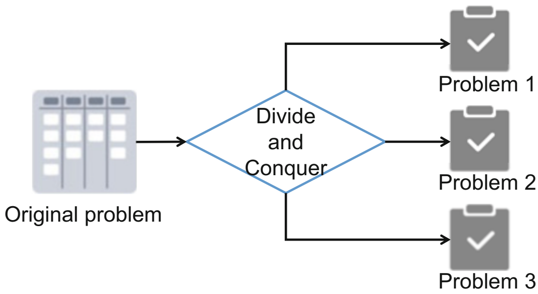

Fig. 6 Divide-and-Conquer implementation

6.2 *Multi-Armed Bandit (MAB)*

A dynamic procurement using multi-armed bandit (MAB) defines the following problem: when a gambler in a casino faces a row of slot machines, he/she needs to decide which one to play, by how many times, and in which order when multiple machines are selected. Suppose a machine has a random reward with a fixed distribution, the objective is to maximize the sum of rewards earned through a series of sequential lever pulls. The total number of pulls is restricted to a given number.

Worker recruitment in crowdsourcing resembles the above MAB process with uncertainty regarding the reward from each worker (e.g., the probability of completion or the quality of a worker). Badanidiyuru et al. [8, 9] proposed a worker selection algorithm to maximize the total rewards through a sequence of worker selections, similar to MAB. In some cases, tasks may not even be successfully completed. For example, the task assigned to a worker can be completed with probability. Hassan et al. [41] used the MAB process for task assignment among all the workers to maximize the success rate under a given budget.

In another extension, workers are associated with a unique completion cost, but unknown completion quality. The objective of crowdsourcing is to maximize the total completion quality under the given budget constraint [33, 35]. More specifically, the platform would divide the task allocation process into multiple rounds. In each round, the platform will select several workers (i.e., each worker is assigned one task [34]). Here, the platform has to face the dilemma between exploitation and exploration. The term "exploitation" means that the platform would select the users who had good performance in the past, while the term "exploration" indicates that the platform would try some users with not-so-good performance in the past to discover the potentially optimal users who will generate high completion quality in the future. For example, the traffic monitoring application can leverage various network users' terminals to collect real-time traffic data. However, since the external environment is changing, the data quality for different users is usually unknown in priority.

6.3 Incentive Mechanisms

Incentive mechanisms are the main motivation to do a task in crowdsourcing. There are two kinds of incentive mechanisms, which are *monetary* and *non-monetary* [22]. In the monetary incentive mechanism, a worker selects a task sent from the requester from the platform, and the worker will get a payment when the requester accepts the result of the work. It is usually assumed that workers are rational and will go after tasks that offer more monetary reward [22, 28, 43, 69, 128]. Not all workers are looking for monetary compensation.

Among non-monetary incentives, the motive may come from personal satisfaction, such as free contributions to knowledge-sharing sites like Wikipedia [122] and wikiHow [121] and question-and-answer sites like Quora [111], and Stack Overflow [112]. Some incentives are more natural, such as fun and entertainment. Making tasks more entertaining will help to increase user participation and ease the recruitment effort. Like Gwap [95], the platform provides games to attract users for solving various complex tasks. Foldit is a special type of multi-player game used to predict the structure of proteins [23]. Other non-monetary incentives come from social recognition, self-esteem, or honor, such as a $1 award for the first 2000 participants in the winning team of Tag Challenges. Still, other incentives come from actions that benefit a cause that is considered to be just, based on some common briefs or religions, such as in the case of searching for debris of Malaysia Airport MH 370.

7 Opportunities and Future Directions

This section discusses some opportunities and future directions of crowdsourcing. It focuses on the traditional model that selects workers for tasks issued by the requester. Issues related to task partition and uncertainty associated with workers in terms of their qualities and honesty will be covered.

7.1 Beyond Simple Workflows

The workflow structure controls how a task is completed by a crowd of workers. It influences not only the completion time of a task but also the quality of the results.

One of the possible research opportunities in crowdsourcing is to include more complex workflow structures. Most of the existing crowdsourcing systems adopt sequential, iterative, or parallel workflow structures. By blending bottom-up and open processes with top-down organization goals, many difficult tasks can be accomplished by crowdsourcing. For instance, instead of presenting a fixed set of task questions to crowdsourcing workers, HumanGS [78] selects optimal sets of questions to minimize the overall cost. It uses a graph search approach to optimally allocate crowdsourcing questions to different workers. Jigsaw Percolation [16] designs a new model to let workers collaboratively solve a puzzle on social networks: each worker is assigned a piece of the puzzle, and acquaintances' puzzle pieces can be merged if they are compatible. If all pieces eventually merge together, then the puzzle is solved. This model essentially represents the natural dynamic of how compatible ideas and innovations are merged in the real world, and the underlying workflow structure forms a graph.

We strongly believe that the research on workflow design will be a fruitful research direction. The readers should consider not only the workflow for task allocations but also the workflow for collecting results. This problem becomes more important especially since many distributed crowdsourcing applications have emerged. In addition, in Crowdsourcing 2.0, human intelligence has been integrated with other advanced technologies, such as artificial intelligence, big data, cybersecurity, and the Internet of Things. The workflow design for those integrated new crowdsourcing applications may also need to be addressed.

7.2 Beyond Simple Worker Selection

In crowdsourcing, the quality of the workers affects the results' accuracy and the tasks' completion time and costs. Usually, a requester has no choice but to hire more workers and extend the tasks' expiration time if the existing ones are doing poorly. Therefore, the selection of workers is a crucial step in crowdsourcing.

The initial version of crowdsourcing simply assigns tasks to any worker who is willing to participate. However, considering the fact that workers may have distinct accuracy and spend different amounts of time working on the tasks, this simple worker selection mechanism usually cannot optimize the system's overall rewards (e.g., the completion time of the entire project, costs of hiring workers, or the accuracy of the results.)

To design a better worker selection strategy, several attempts have been made, including the MAB-based approaches discussed earlier. However, in the real world, a crowdsourcing platform does not know workers' reliability a priori. Also, workers may be dishonest and report a higher cost in order to receive a better payment. Gao et al. [36, 126] explored this situation and designed an auction-based combinatorial multi-armed bandit mechanism. This approach can help a crowdsourcing platform select workers and optimize the overall quality achieved under a limited budget.

Challenges still remain in crowdsourcing when multiple workers are recruited and their areas of coverage may overlap. Some form of extended MAB beyond superarm [20] may be needed. There are still open issues in MAB-based approaches, including the way to introduce fairness in addition to merit. We believe that more theoretical works about the recruitment strategy of workers are needed in the field of crowdsourcing.

7.3 Beyond Independent Workers

The majority of the existing crowdsourcing platforms hire independent workers, which results in a common problem that they can only process simple and independent tasks [52]. To resolve the limitation of independent workers, new crowdsourcing systems have been designed.

For instance, Social Crowdsourcing (SC) [19] explores the social networks of individual workers, and it outsources tasks to a crowd of socially related workers, instead of a single independent one. In SC, a task is accomplished by repeatedly recruiting new workers through their social connections. In other words, a task's owner simply outsources the workloads to a few workers and leaves them to further propagate the sub-tasks to their friends, friends of friends, and so on.

However, since SC is a fully distributed system, there is no central control. How to estimate the amount of workload that propagated from a worker to his friends is a crucial problem, which significantly influences the completion time of the entire project. In addition, the distributed workload allocation problem is not trivial since there are common friends of some workers. Chang et al. [18] found that the workload allocation strategy should be changed at different stages of the workload propagation, and they proposed an adaptive approach to assign workload to SC workers. Xiao et al. designed an optimal offline method and a greedy online scheme to solve the problem [125].

When shifting from independent workers to teams, many new problems emerge, such as workload allocation, incentive strategy, and task results aggregation from workers. Recently, several novel mechanisms have been designed for social networks-based crowdsourcing. For example, Jiang et al. [46] study how to measure the overall capacity of a group of workers on social networks and propose a new algorithm for team-based task allocation. Wang et al. [98] consider the impacts of selfish workers and design new incentives to promote workers to behave honestly.

In short, the research about social networks-based crowdsourcing is a new trend. Since more complicated worker recruitment schemes have been developed, such as hiring a crowd of friends, the workload assignment, result collection, and incentive mechanism should be re-designed completely.

7.4 Beyond Simple Training

Since the tasks of crowdsourcing are completed by independent workers, their responses contain large variances in terms of quality [48]. In the real world, some workers are even malicious and may aim to corrupt a crowdsourcing structure [64]. To resolve this issue, many papers have proposed new crowdsourcing models or schemes, such as hiring redundant workers, involving experts [132], designing an assessment for workers' quality [22], or constructing statistical models to reduce the influence of the variance in responses [37]. Besides those traditional approaches, some platforms [32, 60, 96] even provide initial training or tests for the non-expert workers [26, 32].

However, there are two open questions about the training in crowdsourcing. The first question is about the construction of proper training data for a variety of users. The majority of the existing systems only provide a single set of data for all workers. However, people with different backgrounds may need diverse training before participating in some crowdsourcing tasks. In other words, the state-of-the-art crowdsourcing systems focus only on the tasks themselves but ignore the opportunity to analyze workers, learn their learning needs, and adaptively create different training examples. There are some existing works [29, 93] on classifying workers and aggregating responses or assigning new tasks based on the community feature of workers. However, as far as we are aware, the research on adaptively generating differential tutoring is limited. Instead of simply discarding the responses from workers with lower accuracy, one may explore them to generate more efficient training data for future workers with similar backgrounds. We believe this could be a promising research direction.

Secondly, in the field of crowdsourcing, the impacts of subjectivity are underestimated widely. Not all variance in workers' responses is related to the error. In fact, many of the crowdsourcing questions are subjective to a certain degree, but most people simply regard an uncommon response as mislabeling. The lack of the study of subjectivity further brings in three sub-questions: (1) how to quantitatively measure the subjectivity in crowdsourcing questions and responses, (2) how to design training data to minimize the influence of subjectivity, and (3) how to aggregate subjective responses and produce a more comprehensive result. To date, only a few papers [47, 48, 73] mathematically model the subjectivity in crowdsourcing. It is apparent that a theoretical study of subjectivity is necessary for the field.

To conclude, as more and more crowdsourcing platforms provide training and qualification tests to workers, the creation of subjectivity-sensitive and differential training data for various groups of workers is an open question.

7.5 Beyond Simple Interactive Mode

Traditional crowdsourcing platforms are web page-based, where individual workers proactively find a task from a web page, accept it, and complete a series of subtasks alone. This kind of interaction makes workers bored and tired soon [82]. In addition, their accuracy rate may drop due to boredom or tiredness. In recent years, the increasing popularity of conversational agents has enabled a more interactive communication mode between workers and crowdsourcing platforms. The Conversational Crowdsourcing platforms [15, 49, 88] adopt some conversational agents to assist the workers. Based on their preferences, many conversational styles are available, such as "High-Involvement," "High-Considerateness," casual chatting, or formal description styles [83]. Some agents can even invite workers to take a break after a certain number of tasks. Based on the studies [83, 84], Conversational Crowdsourcing platforms not only enhance the engagement of workers but also improve their retention rate.

The appearance of conversational agents in crowdsourcing brings many interesting research problems, ranging from the design of conversational agents to the auto-recognition of workers' preferred interactive modes. For instance, can we include some social features when designing the agents? Can we let crowdsourcing workers be the agents or machines only? How can initial training exercises be added to the interactions between workers and agents? Is there some way to automatically find out the preferred conversational styles of a new worker based on his/her social profile? In short, conversational crowdsourcing is a newly emerged topic, and there are plenty of open questions.

7.6 AI Applications

In recent years, the blooming of AI significantly affects the field of crowdsourcing. Many novel AI-based crowdsourcing applications have emerged. Those applications range from the environment and traffic monitoring to the preservation of data privacy.

Crowdsourcing has wide applications in Federated Learning (FL) [59], which is a machine learning (ML) technique that trains an ML algorithm across multiple decentralized devices, i.e., IoT devices of edges in a crowdsourcing application, holding local data samples, without exchanging them directly. One important aspect of FL is the ever increasing demand for privacy-preserving AI [2] and truthfulness for each participant in collective contribution.

In the application of sparse crowdsensing, the spatiotemporal matrix of a sensing task has the following spatiotemporal correlations: the adjacent rows (or columns) are linearly dependent. The matrix with spatiotemporal correlations is called a low-rank matrix. Therefore, several data inference methods combining data-driven methods, and unique attributes of a low-rank matrix have been proposed for urban traffic flow [87, 136], air quality [62], humidity and temperature, and road speed [100–102, 127] to solve the problems of sparse data in large-scale datasets. One future direction is to introduce active learning for the process of job selection and worker matching. The objective of sparse crowdsensing is then to dynamically recruit workers to actively collect an optimal subset of more valuable data to speed up the sampling process.

7.7 Crowdsourcing 2.0

The next generation of crowdsourcing seamlessly combines multiple advantageous technologies together and provides more powerful functionalities to solve many complex problems in the real world.

The emergence of AI, IoT, Blockchain, and advanced sensing and edge computing techniques bring new opportunities to crowdsourcing. Many cross-platform novel applications have been developed. Third-Eye [61] monitors PM2.5 pollution by analyzing the outdoor images taken by citizens' phones [133] This system integrates AI, Crowdsourcing, and Image processing together to protect the environment. LiFS [124] uses radio signals, sensors, and crowdsourcing to build an indoor floor plan and future provide localization service.

As new technologies have been integrated into crowdsourcing, there are more issues to be tackled. Unlike traditional crowdsourcing, the next generation employs both human intelligence and machine computing capability. The workers consist of not only human beings but also a variety of types of sensors. How to efficiently manage those heterogeneous workers, how to allocate and schedule tasks, and how to coordinate and orchestrate them in an appropriate manner become open questions. In addition, the existing crowdsourcing applications usually address specific problems independently, and therefore compatibility inevitably becomes a crucial issue. CrowdOS [63] is an initial attempt at it, but the problem of compatibility is far from sufficiently explored, especially since other technologies are integrating into crowdsourcing.

8 Conclusion

In this chapter, we gave a systematic review of crowdsourcing, its definition, history, applications, key challenges, and future research directions. We took a view of human capability and knowledge in crowdsourcing as the Human Processing Unit (HPU) that complements the widespread applications that resort mainly to computers, i.e., the Central Processing Unit (CPU). Our future seems to be moving in the direction of a combination of both HPU and CPU. For example, the current state of the art in medical imaging is a kind of CPU-assisted HPU application, where the CPU plays an important role, but the final decision is still made by the doctors (special HPUs). Through studying the history of chess and Go games between humans and machines, it seems that the future challenges lie in finding methods to combine human intelligence (HPU) and machine intelligence (CPU) to reach new heights. Perhaps, we can take the view of humans racing with machines side by side, rather than humans racing against machines.

Acknowledgments The work of Jie Wu was supported in part by NSF grants CPS 2128378, CNS 2107014, CNS 2150152, CNS 1824440, CNS 1828363, and CNS 1757533.

The work of Chao Song was supported in part by the National Key R&D Program under grants No. 2021YFB3101302 and No. 2021YFB3101303, the NSFC under grant 62020106013, the S & T Achievements Transformation Demonstration Project of Sichuan No. 2018CC0094, and the Fundamental Research Funds for the Central Universities No. ZYGX2019J075.

This chapter is expanded based on Jie Wu's keynotes at IEEE IPCCC 2014, IEEE APDCM 2017, and EAI CollaborateCom 2021.

References

1. https://en.wikipedia.org/wiki/Amazon_Mechanical_
2. M. Al-Rubaie, J.M. Chang, Privacy-preserving machine learning: threats and solutions. IEEE Secur. Priv. **17**(2), 49–58 (2019)
3. A. Amato, A.D. Sappa, A. Fornés, F. Lumbreras, J. Lladós, Divide and conquer: atomizing and parallelizing a task in a mobile crowdsourcing platform, in *Proceedings of the 2nd ACM International Workshop on Crowdsourcing for Multimedia (CrowdMM '13)* (2013)
4. Amazon Mechanical Turk. http://www.mturk.com/ [Online]
5. V. Ambati, S. Vogel, J.G. Carbonell, Collaborative workflow for crowdsourcing translation, in *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (2012)
6. Y. Amsterdamer, Y. Grossman, T. Milo, P. Senellart, Crowd mining, in *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data (SIGMOD '13)* (2013)
7. E.E. Arolas, F.G.L. de Guevara, Towards an integrated crowdsourcing definition. J. Inf. Sci. **38**, 189–200 (2012)
8. A. Badanidiyuru, R. Kleinberg, A. Slivkins, Bandits with knapsacks: dynamic procurement for crowdsourcing, in *The 3rd Workshop on Social Computing and User Generated Content, co-located with ACM EC*, vol. 9 (Citeseer, Princeton, 2013), p. 2
9. A. Badanidiyuru, R. Kleinberg, A. Slivkins, Bandits with knapsacks. J. ACM (JACM) **65**(3), 1–55 (2018)
10. M.S. Bernstein, J. Teevan, S.T. Dumais, D.J. Liebling, E. Horvitz, Direct answers for search queries in the long tail, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2012)
11. J.P. Bigham, C. Jayant, H. Ji, G. Little, A. Miller, R. Miller, R. Miller, A. Tatarowicz, B.A. White, S. White, T. Yeh, Vizwiz: nearly real-time answers to visual questions, in *Proceedings of the 23rd annual ACM Symposium on User Interface Software and Technology* (2010)
12. D.C. Brabham, Crowdsourcing as a model for problem solving. Convergence **14**, 75–90 (2008)
13. D.C. Brabham, *Corwdsourcing (the MIT Press Essential Knowledge Series)* (The MIT Press, New York, 2013)
14. D.C. Brabham, *Crowdsourcing* (MIT Press, New York, 2013)
15. L. Bradeško, M. Witbrock, J. Starc, Z. Herga, M. Grobelnik, D. Mladenić, Curious cat–mobile, context-aware conversational crowdsourcing knowledge acquisition. ACM Trans. Inf. Syst. **35**(4), 1–46 (2017)
16. C.D. Brummitt, S. Chatterjee, P.S. Dey, D. Sivakoff, Jigsaw percolation: what social networks can collaboratively solve a puzzle?, in *The Annals of Applied Probability* (2015), pp. 2013–2038
17. C. Cao, J. Tu, Z. Liu, L. Chen, H.V. Jagadish, Tuning crowdsourced human computation, in *2017 IEEE 33rd International Conference on Data Engineering (ICDE)* (2017), pp. 1021–1032
18. W. Chang, J. Wu, Progressive or conservative: Rationally allocate cooperative work in mobile social networks. IEEE Trans. Parallel Distrib. Syst. **26**(7), 2020–2035 (2014)
19. W. Chang, J. Wu, Reliability enhanced social crowdsourcing, in *2015 IEEE Global Communications Conference (GLOBECOM)* (IEEE, New York, 2015), pp. 1–6
20. W. Chen, Y. Wang, Y. Yuan, Combinatorial multi-armed bandit: general framework and applications, in *International Conference on Machine Learning* (PMLR, Singapore, 2013), pp. 151–159
21. Z. Chen, R. Fu, Z. Zhao, Z. Liu, L. Xia, L. Chen, P. Cheng, C.C. Cao, Y. Tong, C.J. Zhang, gMission: a general spatial crowdsourcing platform. Proceedings of the VLDB Endowment **7**(13), 1629–1632 (2014)
22. A.I. Chittilappilly, L. Chen, S. Amer-Yahia, A survey of general-purpose crowdsourcing techniques. IEEE Trans. Knowl. Data Eng. **28**, 2246–2266 (2016)

23. S. Cooper, F. Khatib, A. Treuille, J. Barbero, J. Lee, M. Beenen, A. Leaver-Fay, D. Baker, Z. Popovic, F. players, Predicting protein structures with a multiplayer online game. Nature **466**, 756–760 (2010)
24. S.B. Davidson, S. Khanna, T. Milo, S. Roy, Using the crowd for top-k and group-by queries, in *Proceedings of the 16th International Conference on Database Theory (ICDT '13)* (2013)
25. J.C. de Winter, M. Kyriakidis, D. Dodou, R. Happee, Using crowdflower to study the relationship between self-reported violations and traffic accidents. Procedia Manufacturing **3**, 2518–2525 (2015)
26. S. Doroudi, E. Kamar, E. Brunskill, E. Horvitz, Toward a learning science for complex crowdsourcing tasks, in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (2016), pp. 2623–2634
27. Y. Duan, G. Gao, M. Xiao, J. Wu, A privacy-preserving order dispatch scheme for ride-hailing services, in *2019 IEEE 16th International Conference on Mobile Ad Hoc and Sensor Systems (MASS)* (IEEE, New York, 2019), pp. 118–126
28. C. Eickhoff, A. de Vries, How crowdsourcable is your task, in *Proceedings of the Workshop on Crowdsourcing for Search and Data Mining (CSDM) at the Fourth ACM International Conference on Web Search and Data Mining (WSDM)* (2011), pp. 11–14
29. J. Fan, G. Li, B.C. Ooi, K.l. Tan, J. Feng, iCrowd: An adaptive crowdsourcing framework, in *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data* (2015), pp. 1015–1030
30. A. Feng, M.J. Franklin, D. Kossmann, T. Kraska, S. Madden, S. Ramesh, A. Wang, R. Xin, CrowdDB: Query processing with the VLDB crowd. Proc. VLDB Endow. **4**, 1387–1390 (2011)
31. M.J. Franklin, D. Kossmann, T. Kraska, S. Ramesh, R. Xin, CrowdDB: answering queries with crowdsourcing, in *Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data (SIGMOD '11)* (2011)
32. U. Gadiraju, B. Fetahu, R. Kawase, Training workers for improving performance in crowdsourcing microtasks, in *European Conference on Technology Enhanced Learning* (Springer, Berlin, 2015), pp. 100–114
33. G. Gao, H. Huang, M. Xiao, J. Wu, Y.E. Sun, Y. Du, Budgeted unknown worker recruitment for heterogeneous crowdsensing using CMAB. IEEE Trans. Mob. Comput. **21**(11), 3895–3911 (2021)
34. G. Gao, J. Wu, Z. Yan, M. Xiao, G. Chen, Unknown worker recruitment with budget and covering constraints for mobile crowdsensing, in *2019 IEEE 25th International Conference on Parallel and Distributed Systems (ICPADS)* (IEEE, New York, 2019), pp. 539–547
35. G. Gao, J. Wu, M. Xiao, G. Chen, Combinatorial multi-armed bandit based unknown worker recruitment in heterogeneous crowdsensing, in *IEEE INFOCOM 2020-IEEE Conference on Computer Communications* (IEEE, New York, 2020), pp. 179–188
36. G. Gao, H. Huang, M. Xiao, J. Wu, Y.E. Sun, S. Zhang, Auction-based combinatorial multi-armed bandit mechanisms with strategic arms, in *IEEE INFOCOM 2021-IEEE Conference on Computer Communications* (IEEE, New York, 2021), pp. 1–10
37. P. Germain, A. Lacasse, F. Laviolette, M. Marchand, J.F. Roy, Risk bounds for the majority vote: From a PAC-Bayesian analysis to a learning algorithm. arXiv preprint arXiv:1503.08329 (2015)
38. J. Gu, C. Song, W. Jiang, L. Shi, H. Gong, M. Liu, Pedestrian flow prediction with business events, in *2019 15th International Conference on Mobile Ad-Hoc and Sensor Networks (MSN)* (IEEE, New York, 2019), pp. 43–48
39. J. Gu, C. Song, W. Jiang, X. Wang, M. Liu, Enhancing personalized trip recommendation with attractive routes, in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34 (2020), pp. 662–669
40. S. Guo, A.G. Parameswaran, H. Garcia-Molina, So who won?: dynamic max discovery with the crowd, in *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data* (2012)

41. U.U. Hassan, E. Curry, A multi-armed bandit approach to online spatial task assignment, in *2014 IEEE 11th International Conference on Ubiquitous Intelligence and Computing and 2014 IEEE 11th International Conference on Autonomic and Trusted Computing and 2014 IEEE 14th International Conference on Scalable Computing and Communications and Its Associated Workshops* (IEEE, New York, 2014), pp. 212–219
42. M. Hirth, T. Hoßfeld, P. Tran-Gia, Analyzing costs and accuracy of validation mechanisms for crowdsourcing platforms. Math. Comput. Model. **57**(11–12), 2918–2932 (2013)
43. J.J. Horton, L.B. Chilton, The labor economics of paid crowdsourcing, in *Proceedings of the 11th ACM Conference on Electronic Commerce* (2010), pp. 209–218
44. J. Howe, The rise of crowdsourcing. https://www.wired.com/2006/06/crowds/ (2006)
45. Hybrid machine/human computing arrangement. U.S. Patent (2001)
46. J. Jiang, B. An, Y. Jiang, C. Zhang, Z. Bu, J. Cao, Group-oriented task allocation for crowdsourcing in social networks. IEEE Trans. Syst. Man Cybern. Syst. **51**(7), 4417–4432 (2019)
47. Y. Jin, M. Carman, Y. Zhu, W. Buntine, Distinguishing question subjectivity from difficulty for improved crowdsourcing, in *Asian Conference on Machine Learning* (PMLR, Singapore, 2018), pp. 192–207
48. Y. Jin, M. Carman, Y. Zhu, Y. Xiang, A technical survey on statistical modelling and design methods for crowdsourcing quality control. Artif. Intell. **287**, 103351 (2020)
49. J.Y. Jung, S. Qiu, A. Bozzon, U. Gadiraju, Great chain of agents: The role of metaphorical representation of agents in conversational crowdsourcing, in *CHI Conference on Human Factors in Computing Systems* (2022), pp. 1–22
50. G. Kasparov, The chess master and the computer. The New York Review of Books **57**(2), 16–19 (2010)
51. A. Kittur, E.H. Chi, B. Suh, Crowdsourcing user studies with Mechanical Turk, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)* (2008)
52. A. Kittur, B. Smus, S. Khamkar, R.E. Kraut, CrowdForge: Crowdsourcing complex work, in *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology* (2011), pp. 43–52
53. A. Kulkarni, M. Can, B. Hartmann, Collaboratively crowdsourcing workflows with turkomatic, in *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (2012), pp. 1003–1012
54. K.R. Lakhani, D.A. Garvin, E. Lonstein, Topcoder (a): Developing software through crowdsourcing, in *Harvard Business School General Management Unit Case (610–032)* (2010)
55. W. Lasecki, C. Miller, A. Sadilek, A. Abumoussa, D. Borrello, R. Kushalnagar, J. Bigham, Real-time captioning by groups of non-experts, in *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology* (2012), pp. 23–34
56. M. Lease, J. Hullman, J. Bigham, M. Bernstein, J. Kim, W. Lasecki, S. Bakhshi, T. Mitra, R. Miller, Mechanical Turk is not anonymous. Available at SSRN 2228728 (2013)
57. J.M. Leimeister, M.J. Huber, U. Bretschneider, H. Krcmar, Leveraging crowdsourcing: Activation-supporting components for it-based ideas competition. J. Manag. Inf. Syst. **26**, 197–224 (2009)
58. G. Li, J. Wang, Y. Zheng, M.J. Franklin, Crowdsourced data management: A survey. IEEE Trans. Knowl. Data Eng. **28**(9), 2296–2319 (2016)
59. W.Y.B. Lim, N.C. Luong, D.T. Hoang, Y. Jiao, Y.C. Liang, Q. Yang, D. Niyato, C. Miao, Federated learning in mobile edge networks: a comprehensive survey. IEEE Commun. Surv. Tutorials **22**(3), 2031–2063 (2020)
60. C.J. Lintott, K. Schawinski, A. Slosar, K. Land, S. Bamford, D. Thomas, M.J. Raddick, R.C. Nichol, A. Szalay, D. Andreescu, et al. Galaxy zoo: morphologies derived from visual inspection of galaxies from the Sloan digital sky survey. Mon. Not. R. Astron. Soc. **389**(3), 1179–1189 (2008)
61. L. Liu, W. Liu, Y. Zheng, H. Ma, C. Zhang, Third-eye: a mobilephone-enabled crowdsensing system for air quality monitoring. Proc. ACM Interact. Mobile Wearable Ubiquitous Technol. **2**(1), 1–26 (2018)

62. T. Liu, Y. Zhu, Y. Yang, F. Ye, alc^2: when active learning meets compressive crowdsensing for urban air pollution monitoring. IEEE Internet Things J. **6**(6), 9427–9438 (2019)
63. Y. Liu, Z. Yu, B. Guo, Q. Han, J. Su, J. Liao, CrowdOS: a ubiquitous operating system for crowdsourcing and mobile crowd sensing. IEEE Transactions on Mobile Computing **21**(3), 878–894 (2020)
64. J. Liu, M. Li, W.C. Tang, S. Islam, A cyber physical system crowdsourcing inference method based on tempering: an advancement in artificial intelligence algorithms. Wirel. Commun. Mob. Comput. **2021**, 1–11 (2021)
65. I. Lotosh, T. Milo, S. Novgorodov, CrowdPlanr: planning made easy with crowd, in *Proceedings of the 2013 IEEE 29th International Conference on Data Engineering (ICDE)* (2013), pp. 1344–1347
66. A. Lynch, Crowdsourcing is not new—the history of crowd-sourcing (1714 to 2010) (2015). http://blog.designcrowd.com/article/202/crowdsourcing-is-not-new--the-history-of-crowdsourcing-1714-to-2010
67. K. Mao, L. Capra, M. Harman, Y. Jia, A survey of the use of crowdsourcing in software engineering. J. Syst. Softw. **126**, 57–84 (2017)
68. A. Marcus, E. Wu, D.R. Karger, S. Madden, R. Miller, Human-powered sorts and joins. Proc. VLDB Endow. **5**, 13–24 (2011)
69. A. Moreno, J.L. de la Rosa, B.K. Szymanski, J.M. Barcenas, Reward system for completing FAQs, in *CCIA* (2009), pp. 361–370
70. A. Morishima, N. Shinagawa, T. Mitsuishi, H. Aoki, S. Fukusumi, CyLog/Crowd4U: a declarative platform for complex data-centric crowdsourcing. Proc. VLDB Endow. **5**(12), 1918–1921 (2012)
71. A. Morishima, S. Amer-Yahia, S.B. Roy, Crowd4u: An initiative for constructing an open academic crowdsourcing network, in Second AAAI Conference on Human Computation and Crowdsourcing (2014)
72. M. Negri, L. Bentivogli, Y. Mehdad, D. Giampiccolo, A. Marchetti, Divide and conquer: crowdsourcing the creation of cross-lingual textual entailment corpora, in *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing* (2011), pp. 670–679
73. A. Nguyen, M. Halpern, B. Wallace, M. Lease, Probabilistic modeling for crowdsourcing partially-subjective ratings, in *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, vol. 4 (2016), pp. 149–158
74. J. Noronha, E. Hysen, H. Zhang, K.Z. Gajos, Platemate: crowdsourcing nutritional analysis from food photographs, in *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology* (2011), pp. 1–12
75. https://www.nerdwallet.com/article/finance/how-to-make-money-on-upwork [Online]
76. https://digital.hbs.edu/platform-digit/submission/crowdflower-powering-the-human-side-of-artificial-intelligence/ [Online]
77. M. Pantic, A.S. Pentland, A. Nijholt, T.S. Huang, Human computing and machine understanding of human behavior: a survey, in *Proceedings of the 8th International Conference on Multimodal Interfaces (ICMI'06)* (2006)
78. A. Parameswaran, A.D. Sarma, H. Garcia-Molina, N. Polyzotis, J. Widom, Human-assisted graph search: it's okay to ask questions. arXiv preprint arXiv:1103.3102 (2011)
79. A.G. Parameswaran, H. Garcia-Molina, H. Park, N. Polyzotis, A. Ramesh, J. Widom, Crowdscreen: algorithms for filtering data with humans. *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data* (2012)
80. A.G. Parameswaran, N. Polyzotis, Answering queries using humans, algorithms and databases, in *CIDR* (2011)
81. F. Pesapane, M. Codari, F. Sardanelli, Artificial intelligence in medical imaging: threat or opportunity? radiologists again at the forefront of innovation in medicine. European Radiology Experimental **2**, 1–10 (2018)
82. S. Qiu, U. Gadiraju, A. Bozzon, Estimating conversational styles in conversational microtask crowdsourcing. Proc. ACM. Hum. Comput. Interact. **4**(CSCW1), 1–23 (2020)

83. S. Qiu, U. Gadiraju, A. Bozzon, Improving worker engagement through conversational microtask crowdsourcing, in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (2020), pp. 1–12
84. S. Qiu, A. Bozzon, M.V. Birk, U. Gadiraju, Using worker avatars to improve microtask crowdsourcing. Proc. ACM. Hum. Comput. Interact. **5**(CSCW2), 1–28 (2021)
85. J. Ross, L.C. Irani, M.S. Silberman, A. Zaldivar, B. Tomlinson, Who are the crowdworkers?: shifting demographics in Mechanical Turk, in *CHI '10 Extended Abstracts on Human Factors in Computing Systems* (2010)
86. Y. Roth, 11 of the 12 best global brands use creative crowdsourcing [online]. http://yannigroth.com/2012/03/23/xx-of-the-100-best-global-brands-use-creative-crowdsourcing/ (2012)
87. M. Roughan, Y. Zhang, W. Willinger, L. Qiu, Spatio-temporal compressive sensing and internet traffic matrices (extended version). IEEE/ACM Trans. Networking **20**(3), 662–676 (2011)
88. P. Shah, D. Hakkani-Tur, B. Liu, G. Tür, Bootstrapping a neural conversational agent with dialogue self-play, crowdsourcing and on-line reinforcement learning, in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, vol. 3 (Industry Papers) (2018), pp. 41–51
89. G. Smith, H. Rudge-Pickard, Longitude problem. http://crazysquirrel.com/computing/software/vrml/essay/longitude_problem.jspx (2015)
90. X. Tang, The role of artificial intelligence in medical imaging research. BJR Open **2**(1), 20190031 (2020)
91. M. Tsvetkova, T. Yasseri, E.T. Meyer, J.B. Pickering, V. Engen, P.W. Walland, M. Lüders, A. Følstad, G.N. Bravos, Understanding human-machine networks. ACM Comput. Surv. (CSUR) **50**, 1–35 (2017)
92. K. Tuite, N. Snavely, D.y. Hsiao, N. Tabing, Z. Popovic, Photocity: training experts at large-scale image acquisition through a competitive game, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2011), pp. 1383–1392
93. M. Venanzi, J. Guiver, G. Kazai, P. Kohli, M. Shokouhi, Community-based Bayesian aggregation models for crowdsourcing, in *Proceedings of the 23rd International Conference on World Wide Web* (2014), pp. 155–164
94. L. Von Ahn, B. Maurer, C. McMillen, D. Abraham, M. Blum, recaptcha: Human-based character recognition via web security measures. Science **321**(5895), 1465–1468 (2008)
95. L. von Ahn, Games with a purpose (GWAP). http://www.gwap.com/ (2008)
96. J. Wang, P.G. Ipeirotis, F. Provost, Managing crowdsourcing workers, in *The 2011 Winter Conference on Business Intelligence* Citeseer, Princeton (2011), pp. 10–12
97. J. Wang, G. Li, T. Kraska, M.J. Franklin, J. Feng, Leveraging transitive relations for crowdsourced joins. ArXiv abs/1408.6916 (2013)
98. W. Wang, Z. He, P. Shi, W. Wu, Y. Jiang, B. An, Z. Hao, B. Chen, Strategic social team crowdsourcing: forming a team of truthful workers for crowdsourcing in social networks. IEEE Trans. Mob. Comput. **18**(6), 1419–1432 (2018)
99. X. Wang, X. Qin, M.B. Hosseini, R. Slavin, T.D. Breaux, J. Niu, Guileak: tracing privacy policy claims on user input data for android applications, in *Proceedings of the 40th International Conference on Software Engineering* (2018), pp. 37–47
100. E. Wang, M. Zhang, X. Cheng, Y. Yang, W. Liu, H. Yu, L. Wang, J. Zhang, Deep learning-enabled sparse industrial crowdsensing and prediction. IEEE Trans. Industr. Inform. **17**(9), 6170–6181 (2020)
101. E. Wang, M. Zhang, Y. Yang, Y. Xu, J. Wu, Exploiting outlier value effects in sparse urban crowdsensing, in *Proceedings of the 2021 IEEE/ACM 29th International Symposium on Quality of Service (IWQOS)* (IEEE, New York, 2021), pp. 1–10
102. E. Wang, M. Zhang, Y. Xu, H. Xiong, Y. Yang, Spatiotemporal fracture data inference in sparse urban crowdsensing, in *IEEE INFOCOM 2022-IEEE Conference on Computer Communications* (IEEE, New York, 2022), pp. 1499–1508
103. Crowd4u projects. http://crowd4u.org/en/projects [Online]

104. Crowdflower lawsuit could change crowd labor industry forever. https://tinywork.wordpress.com/2013/07/29/crowdflower/ [Online]
105. Crowdflowers. http://www.crowdflower.com/ [Online]
106. Didi chuxing. https://www.didiglobal.com/ [Online]
107. Digital China. http://www.digitalchina.com/en/ [Online]
108. Do you know what the difference between captcha and recaptcha is. https://datadome.co/bot-management-protection/captcha-vs-recaptcha-whats-the-difference/ [Online]
109. From "Toyoda" to "Toyota". http://www.toyota-global.com/showroom/emblem/history/ [Online]
110. The long history of crowdsourcing—and why you're just now hearing about it. http://www.crowd-source.com/blog/2013/08/the-long-history-of-crowdsourcing-and-why-youre-just-now-hearing-about-it/ [Online]
111. Quora. http://www.quora.com/ [Online]
112. Stack overflow. http://stackoverflow.com [Online]
113. Tohme. https://makeabilitylab.cs.washington.edu/project/tohme/ [Online]
114. Topcoder. https://www.topcoder.com [Online]
115. Upwork. https://www.upwork.com [Online]
116. Upwork—Wikipedia. https://en.wikipedia.org/wiki/Upwork [Online]
117. Waze. https://www.waze.com/ [Online]
118. Wiki: Crowdsourcing. https://en.wikipedia.org/wiki/Crowdsourcing [Online]
119. Wiki galaxy zoo. https://en.wikipedia.org/wiki/Galaxy_Zoo [Online]
120. Wiki: Smart city. https://en.wikipedia.org/wiki/Smart_city [Online]
121. wikihow. https://www.wikihow.com/ [Online]
122. Wikipedia. https://www.wikipedia.org/ [Online]
123. Fusioncomp. Available: http://www.kc.tsukuba.ac.jp/~fusioncomp/index.html (2009) [Online]
124. C. Wu, Z. Yang, Y. Liu, Smartphones based crowdsourcing for indoor localization. IEEE Trans. Mob. Comput. **14**(2), 444–457 (2014)
125. M. Xiao, J. Wu, L. Huang, Y. Wang, C. Liu, Multi-task assignment for crowdsensing in mobile social networks, in *Proceedings of the 2015 IEEE Conference on Computer Communications (INFOCOM)* (IEEE, New York, 2015), pp. 2227–2235
126. M. Xiao, B. An, J. Wang, G. Gao, S. Zhang, J. Wu, CMAB-based reverse auction for unknown worker recruitment in mobile crowdsensing. IEEE Trans. Mob. Comput. **21**(10), 3502–3518 (2021)
127. K. Xie, X. Li, X. Wang, G. Xie, J. Wen, D. Zhang, Active sparse mobile crowd sensing based on matrix completion, in *Proceedings of the 2019 International Conference on Management of Data* (2019), pp. 195–210
128. G. Xintong, W. Hongzhi, Y. Song, G. Hong, Brief survey of crowdsourcing for data mining. Expert Syst. Appl. **41**(17), 7987–7994 (2014)
129. X. Xu, Q. Liu, X. Zhang, J. Zhang, L. Qi, W. Dou, A blockchain-powered crowdsourcing method with privacy preservation in mobile environment. IEEE Trans. Comput. Social Syst. **6**(6), 1407–1419 (2019)
130. T. Yan, V. Kumar, D. Ganesan, Crowdsearch: exploiting crowds for accurate real-time image search on mobile phones, in *Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services* (2010), pp. 77–90
131. X.S. Yang, R. Cheng, L. Mo, B. Kao, D.W.L. Cheung, On incentive-based tagging, in *Proceedings of the 2013 IEEE 29th International Conference on Data Engineering (ICDE)* (2013), pp. 685–696
132. Y. Yang, Q. Bai, Q. Liu, Modeling random guessing and task difficulty for truth inference in crowdsourcing, in *Proceedings of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, vol. 19 (2019), pp. 2288–2290
133. Z. Yu, H. Ma, B. Guo, Z. Yang, Crowdsensing 2.0. Commun. ACM **64**, 76–80 (2021)
134. C. Zhang, L. Chen, H.V. Jagadish, C.C. Cao, Reducing uncertainty of schema matching via crowdsourcing. Proc. VLDB Endow. **6**, 757–768 (2013)

135. S. Zhu, S. Kane, J. Feng, A. Sears, A crowdsourcing quality control model for tasks distributed in parallel, in *CHI'12 Extended Abstracts on Human Factors in Computing Systems* (2012), pp. 2501–2506
136. Y. Zhu, Z. Li, H. Zhu, M. Li, Q. Zhang, A compressive sensing approach to urban traffic estimation with probe vehicles. IEEE Trans. Mob. Comput. **12**(11), 2289–2302 (2012)
137. Zooniverse. https://www.zooniverse.org/projects [Online]
138. Zooniverse: scientific research through crowdsourcing. https://digital.hbs.edu/platform-digit/submission/zooniverse-scientific-research-through-crowdsourcing/ [Online]