Technology Digest

Bulletin of telecom technology

Issue , January 2018

Crowdsourcing

In this paper

What is crowdsourcing

Crowdsourcing is a specific sourcing model that involves many disparate individuals to perform services, generate ideas or content, and submit their data via Internet, social media or smartphone apps. It was a termed coined in 2006 by *Jeff Howe* and *Mark Robinson* as a portmanteau of crowd and outsourcing to describe businesses using the internet to 'outsource work to the crowd'. The idea of crowdsourcing is to tap into the collective intelligence of the crowd to complete business transactions that organizations would normally either perform themselves or outsource. The advantage of the model is that it lowers the fixed costs such as salaried employees and broadens the talent pool. This also provides the organizations insights into what customers really want.

Data about circumstances affecting the crowd can also be generated by crowd itself that can further be leveraged for the benefit of the crowd by providing information or solutions faster than traditional means. However, it is hard to yield any fruitful information or solution from the crowdsourced data directly. Analysis and processing of crowdsourced data can be helpful to generate useful information from the data which, in turn, may benefit the crowd itself. There are varieties of situations where crowdsourced data can be used to give people a better insight into events that impact them or their communities. For example, Google Map predicts the traffic based on some historical data and the real-time data sent by sensors and the smartphones of the crowd that report how fast cars are moving. TRAI also uses its benefits for analytics through various apps e.g., TRAI MySpeed App, TRAI MyCall App, etc.

Background

Although the term 'crowdsourcing' term was popularized on the internet to describe Internet-based activities, the practice of integrating crowd to find a solution to a problem or to achieve a common goal has a long history. There are events in the history where government organizations and other firms have been trying to channelize human efforts to achieve the certain common goal through monetary or non-monetary incentives. For example, in the year 1800, the publication of first Oxford English Dictionary (OED) was a crowdsourced effort where 800 volunteers cataloged words to create first OED.

Another form of crowdsourcing, historically, is open competitions, where the target is to attract novel solutions to various challenging problems. In 2009, crowdsource term got thrust with the DARPA (Defence Advanced Research Projects Agency) Network Challenge. In that challenge, at a prescribed time, DARPA dropped 10 red weather balloons at public locations across the U.S. The first group to identify all of them received 40K USD in prize money.

The winning team from MIT has used the crowdsourced model to get the locations of the balloons with a proper thought to incentivize both the task completion and recruitment. The team found 10 balloons in less than 9 hours.

While crowdsourcing has a long history, the advancements in ICT sector provides a new opportunity and a number of technologies over the web that can be used effectively and efficiently, for work decomposition and data synthesis to dispatch and collect crowd work. For example, Netflix increased the accuracy of their collaborative filtering algorithm by 10.06%, by running open competition "*Netflix Prize*" to the international community between 2006 & 2009.

How Crowdsourcing works?

Crowdsourcing can be seen as an interactive strategy of outsourcing knowledge generation and problem-solving to external actors through a public or semi-public call for inputs. Such calls typically address creative tasks and topics and are usually realized through a website or platform. Crowdsourcing typically involves three categories of actors: the crowd, organizations that benefit from inputs of the crowd and an intermediate platform which links the crowd and the organization and serves as a Crowdsourcing enabler. The key to why crowdsourcing can deliver results better, faster and cheaper than industrial-era organization forms is that it organizes processes around platforms instead of pipes/chains.

Types of Crowdsourcing systems

1. **Intentional Human Computing (IHC) Systems**

   In IHC systems, participants make explicit contributions (expertise, human intelligence, etc.) to achieve the desired goal. The motivating factors behind this may be intrinsic (entrepreneurial mindset, etc.) or extrinsic (financial rewards, etc.). IHC systems may involve social campaigns, scientific/educational purposes, or commercial purposes.

2. **Unintentional Human Computing (UHC) Systems**

   In UHC systems, participants do not care about the final products that matter to system designers, rather human brainpower is channeled through appropriate platforms to achieve meaningful goals. e.g., Human Interaction Proof System - Animal Species Image Recognition for Restricting Access (ASIRRA)

Other categorizations of crowdsourced systems could be based on the crowd motivations, workforce organization, Decision-making process, etc. Based on crowd motivations, crowdsourced systems can be categorized as paid, altruistic, reputation-based, implicit work, etc.

Based on workforce organization, crowdsourced systems can be categorized into a collection based, contest based, and collaborative systems. In collection-based systems, independent work done by individuals is collected by project organizers, e.g., YouTube. In contest-based systems, only limited crowd submissions are to be accepted based on some criteria set by project organizers, e.g., InnoCentive. In a collaborative system, crowd works together to accomplish a certain goal, e.g., Wikipedia.

On the basis of Decision-making process, crowdsourced systems can be categorized as crowd decision based and hierarchical systems. In crowd decision-based systems, crowd voting is done to evaluate the quality of new content. However, in hierarchical systems, hierarchical privileges as associated to the crowd for quality control of the content, Wikipedia, where a hierarchy is used for quality control.

Crowdsourced Data

Crowdsourcing a dataset means building it with others. Everyone contributes their own data points to create the dataset. Due to asynchronous nature of communication over the web, individuals may participate at a time and location that is convenient for them. It is considered to be an effective way to harness crowd capabilities to apply human computation for tasks like sentiment analysis, entity resolution, etc.

There are basically three important problems in crowdsourced data management; 1.Quality control: data collected from the crowd may have noisy or irrelevant data;2. Cost control: data collection from the crowd should have some incentive (monetary or non-monetary) for the crowd; and 3. Latency control: latency control techniques optimize computing time-scales. Therefore, a proper data management framework (data processing and data mining) should be deployed to maximize benefits from crowdsourced data.

Applications of crowdsourcing in TRAI

Crowdsourced data can be used to ascertain the general sentiment of a crowd. Based on this, TRAI took several initiatives in the form of different apps and portals that are helpful to get better insights into the service quality being offered to subscribers in Indian telecom market and formulation of policies, subsequently, for better QoS which is one of the most important objectives of TRAI as per TRAI Act, 1997.

QoS measurements for data should be spatially distributed and collected under comparable conditions as per the Standards of Quality of Service for Mobile Data Services Regulations, 2012, which wasn't always the case. If data of measurement of QoS of wireless data is collected using crowdsource approach with the mobile app installed in user's device then a large number of data points can be collected which may be spatially and temporally distributed. These data points may be post-processed to assess network level performance. However, data points collected using crowdsource approach may be quite different in numbers and the points in time when such measurements were conducted if compared with the approach of measurement on selected points by TSPs as mentioned in the regulation. Appropriate filtering, aggregation, and analysis of data points may be required to provide results in accordance with the requirements.

TRAI decided to crowdsource the test samples, for measurement of mobile data service, to users' mobile phones, for which the TRAI MySpeed App was developed and launched in 2016. TRAI MySpeed App is available on app stores for iOS and Android platforms. It allows users to conduct speed tests to measure the network speed and submit the results to TRAI portal. In addition to download and upload speeds, the app also measures latency and packet loss.

So far, TRAI MySpeed App has been installed by more than 3 Million users and about 400 thousand users are actively using this app to conduct tests and report it to TRAI. In a period of August to October 2017, more than 3 Million test results were available using this TRAI MySpeed App, measuring data throughput and other related parameters from mobile networks of different TSPs.

Recently, TRAI introduced the provision of auto testing in TRAI MySpeed App, with an objective to get more data points which are spatially and temporally distributed. In auto testing, data speed measurements are scheduled and triggered by application logic without user's intervention. In the month of December, on an average, around 4 thousand successful auto tests per day have been conducted by TRAI MySpeed App.

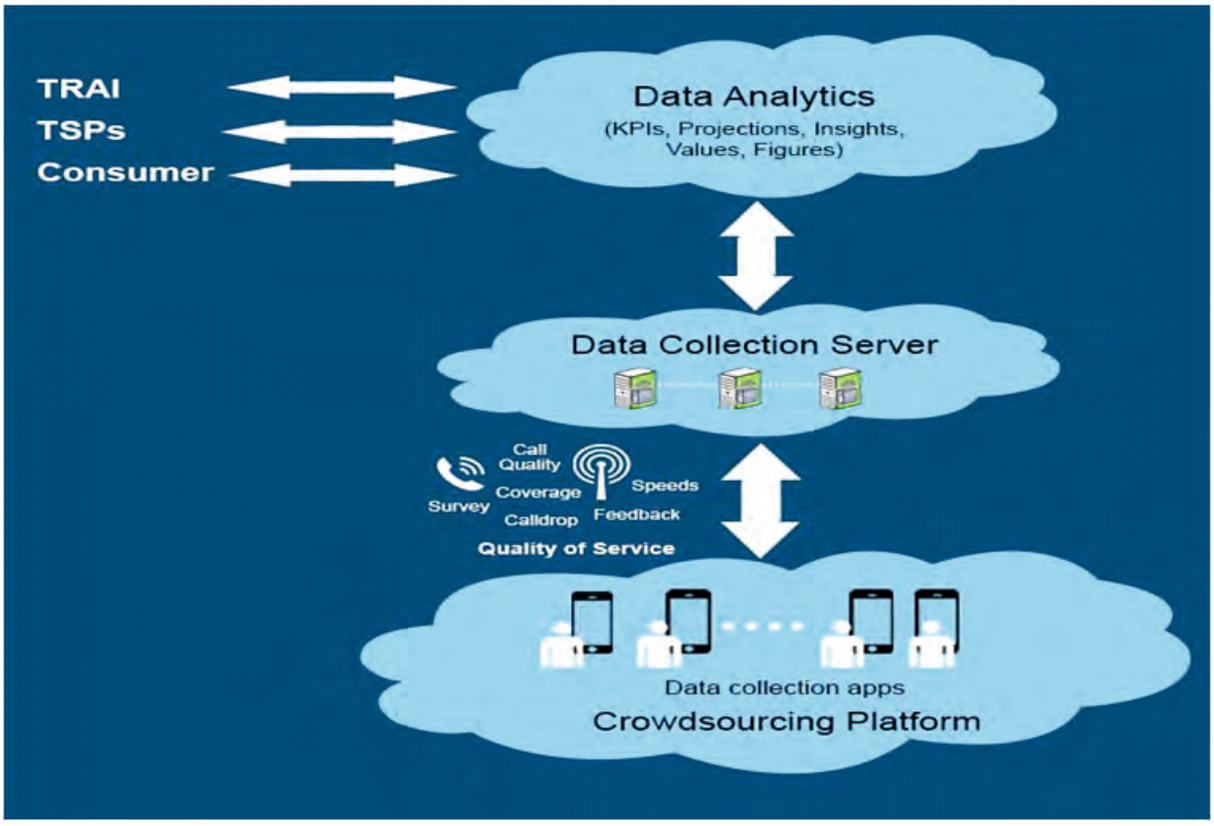

Figure 1: High-level architecture of TRAI's crowdsource platforms

As depicted in Figure 1, TRAI manages the crowdsourcing platforms i.e. apps and portals, which are available for end users. Mobile apps installed on user's device are used to collect data points to a data collection server, which are far more than what may be possible through field measurements. The data collected on the server are then retrieved, processed, and analyzed and the aggregated results, related to various QoS parameters, are then shown the analytics portal of TRAI. The aggregated results of the data collected through crowdsourcing may further be used by TRAI for the formulation of policies related to QoS of telecom in India. The results shown on the portal may be used by consumers to make informed choices while selecting their service providers. On the other hand, TSPs may also use these results while provisioning their services, more effectively and efficiently.

On similar grounds, TRAI released another mobile app – TRAI MyCall – with an objective to provide a platform to all telecom subscribers in India to express their opinion on a centralized platform through a feedback rating process about the quality of calls on their mobile phone. In case of poor quality, users may also give their feedback in terms of Poor Voice quality or Dropped Call. The feedback submitted by users are then analyzed and aggregated and presented on TRAI's analytics portal. The information available on the portal may be helpful for TSPs to get an idea about the general perception of consumers about voice calls offered by them.

So far, TRAI MyCall App has been installed by more than 463 thousand users and about 22 thousand users are actively using this app to express their opinion on the quality of calls. From the launch of the app until the end of Dec 2017, more than 2 million feedbacks have been received.

To take the advantage of crowdsourcing TRAI developed DND 2.0 as an open source initiative where different developers on Cloud contributed to the development of the app. The App is used for consumer complaint management related to Unsolicited Commercial Communication (UCC) to the service providers.

Challenges in Crowdsourcing

Due to inter-disciplinary characteristics of the crowdsourcing model, there are certain challenges in the implementation of crowdsourced systems – System Design, Human Data Analysis, and Human Computation Theory.

- **System Design**: While designing any crowdsourced system, the system designers need to carefully think along three dimensions – Accuracy, Cost, and Latency. To improve the quality of human submissions, participants may need to be compensated more generously or wait for a longer time. On the other hand, to shorten the human processing latency (the time span between task release and completion), either incentive be increased to motivate crowd participants or confidence on the accuracy of the submissions collected be lowered.
- **Human Data Analysis**: Crowdsourcing model should also consider the algorithm and procedure to be followed for proper cleaning of human inputs with varied quality. There are also concerns related to the generation of trustworthy results from the inputs received.
- **Human Computation Theory**: In a proper crowdsourced model, human computation theory should also be employed to train machines using human intelligence to achieve artificial intelligence that can further be used to attain a higher level of computation.

Next Generation Crowdsourcing

Advancements in techniques leveraging IT-mediated crowds such as Crowdsensing, Situated Crowdsourcing, Spatial Crowdsourcing, and Wearable Crowdsourcing have now materially emerged. These techniques represent modified configurations of hardware, software, and people as IT-mediated crowds. These crowdsourcing techniques are different from typical desktop computing paradigm and are termed as next generation crowdsourcing.

- **Crowdsensing**

  In Crowdsensing techniques, smartphone sensor hardware passively collects data and autonomously supplies the data content through device networks to the system controller. For example, GPS and accelerometer data can be used to locate potholes in cities,

- **Situated crowdsourcing**

  In Situated Crowdsourcing technique display hardware (e.g., tablets and other touch screens) fixed to a geographic location are employed, to generate inputs from human beings engaged at the geo-fenced terminal. It focuses on employing human skills at fixed IT installations, therein, engaging human problem solving and idea generation skills.

- **Spatial crowdsourcing**

  In Spatial Crowdsourcing multiple forms of IT (e.g., Mobile apps & Web platforms, Smartphone devices, and Device networks) are used to perform specific actions in the physical environment. An example of spatial crowdsourcing is Waze.com, which is a free GPS navigation app on iPhone/Android with spatial-

crowdsourced features. The users (drivers) can report traffic jam, accident or police so that the other users who are driving ahead this road be aware of those incidents in advance.

- **Wearable crowdsourcing**

  In wearable crowdsourcing, wearables collect and transmit data about the specific wearer of the device and supply such individually-focused data to Crowdsourcing efforts incorporating them. For example, translation of bio-signals into meaningful gauges can inform on how to operate at peak efficiency, recommending for example when to slow down, speed up, or take a break. In other words, bio-signal-driven wearables provide you with a personal dashboard to help steer your life.

Conclusion

Crowdsourcing is a useful approach because it can be utilized by virtually any group or organization for a variety of causes. Crowdsourced data has emerged as a valuable resource to retrieve useful information that can be leveraged for the benefit of the crowd in different situations, e.g., crisis events, day-to-day activities, etc. However, it may be more helpful to generate better insight after application of data processing techniques to clean-up and analyze the data for end-users because crowdsourced data is often noisy and may have redundant or irrelevant data. There may also be a need for validation of framework that is adapted for data processing. Data-mining techniques like trust assessment, feature selection, vectorizing, etc., can also be applied to crowdsourced data.

Sources:

[1] ZhiZhai, Technology-facilitated Crowdsourcing Systems, Notre Dame, Indiana, August 2013

[2] Daren C. Brabham, Crowdsourcing, The MIT Press Essential Knowledge Series, 2013

[3] Geoffrey Barbier, Reza Zafarani, Huji Gao, Gabriel Fung, and Huan Liu, Maximizing Benefits from Crowdsurced Data, Comput Math Organ Theory (2012) 18:257–279, June 2012.

[4] Guoliang Li, Jiannan Wang, Yudian Zheng, and Michael Franklin, Crowdsourced Data Management: A Survey, IEEE 33rd International Conference on Data Engineering, 2017

[5] John Prpic, Next Generation Crowdsourcing for Collective Intelligence, Lulea University of Technology, 2016

Published by: Telecom Regulatory Authority of India
Editorial responsibility: TD Division, TRAI
Contributions, comments and suggestion: sroit@trai.gov.in