Crowd-Sourced Reviews of Aid Delivery

A Distant Dream or a Soon-to-Be Reality?

Challenges and Potential Solutions

Crowd-Sourced Reviews of the Aid Delivery

A Distant Dream or a Soon-to-Be Reality?

December 2017

By

Jeremy M Andrews (jm.andrews@mail.utoronto.ca)
University of Toronto - Munk School of Global Affairs

Andrej Verity (verity@un.org | @andrejverity)
Office for the Coordination of Humanitarian Affairs (OCHA) United Nations

Editor

Alona Volinsky (volinsky@un.org)
Office for the Coordination of Humanitarian Affairs (OCHA) United Nations

Design & Illustrations

Ignacio G. Rebollo (contact@igrebollo.com | gimenez-rebollo@un.org)
Office for the Coordination of Humanitarian Affairs (OCHA) United Nations

made available through support from OCHA

Contents

Crowd-Sourced Reviews of the Aid Delivery

A Distant Dream or a Soon-to-Be Reality?

Background

The concept of gathering feedback directly from affected populations about their experiences with aid agencies has been around for several years. As has the concept of using a technical platform in order to scale this feedback to as many affected as possible.

In 2014, Mark Turner described a concept called AidAdvisor which would be an implementation of a "Trip Advisor"-like rating system within the field of humanitarian aid provision. Such a rating system could be used to help evaluate the effectiveness of aid agencies in meeting the needs of affected populations. About the platform idea, Turner noted that "*ratings are entered on a mobile phone or PC, and aggregated to create an AidAdvisor score. This information is hosted on a simple, informative site, accessible by anyone. The score will guide to help individual and institutional donors in their funding decisions*."

AidAdvisor was envisioned as a tool to encourage aid agencies' accountability to the communities they serve. In mid-2017, the authors decided to revisit this idea with the purpose of examining some of the practical challenges of implementing such a system.

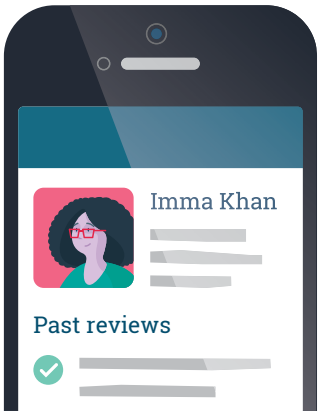
Initial research findings

We first focused on discovering what, if any, similar projects already exist, read academic literature about the challenges associated with rating systems more broadly, and reached out to experts in the field.

Starting with academic literature surrounding websites like TripAdvisor, we found that a considerable amount of research was conducted in the mid-2000s as rating sites grew in popularity. Much of the research was self-funded by the companies involved in using reviews as part of their system. Topics frequently covered were the reliability of reviews, the prevalence of fake reviews, the problem of spam, and fraud.

A common question that arose was whether one can rely on a particular review to be an objective description of the subject of the review. Feedback systems vary in how they handle this question, but generally they use an algorithm to evaluate each review and give it a certain score so that "better" reviews are given more weight than potentially inaccurate or outright fraudulent ones. The precise algorithms are proprietary, but the following are often cited as factors that go into scoring each new review as heuristic properties of the algorithmic function:

1. **The user's history on the site**. Have they been reviewing frequently and for a long time, or is it a new account created just to leave a single review? It is assumed to be more likely that a review coming from someone with more experience will be more fair, balanced and accurate, whereas a one-off review from a new account may be based on an individual complaint that's not representative of the average experience.



2. **How the review compares to others**. There is often assumed to be a bias towards an average, so reviews that look similar to others are given more weight than a review that deviates greatly, since it is generally more important for readers of these sites to know about the average experience than about particular experiences.

3. **User verification**. When possible, the feedback system will use a heuristic function based on various data sources to try and confirm whether or not the reviewer was a genuine customer.

Sites like Expedia and Amazon, which sell directly to customers but also allow outside reviews, give greater weight to reviews from customers that actually purchased products/services from the company they are reviewing. Companies like Google, which collects location data on many of its users, check whether the person leaving a review actually visited the location which s/he is reviewing.

Other noteworthy findings in the research are:

4. **Typical reviewers**. People are more likely to use a feedback system if they have complaints about the product or service rather than if they have praise for it, so ratings tend to present a more negative view of the product than what the average customer could expect to experience. As a result, some systems now use reward systems for encouraging users to leave reviews regardless of experience. Examples include FourSquare and Google which use various status levels to attribute public recognition based on how frequently a user leaves reviews. Many sites also use algorithms to account for this problem, including Google, Amazon and Expedia.

5. **Platform reliability**. One also has to consider the nature of the feedback system to assess the reliability of reviews. Websites that do not sell directly to consumers and rather have an advertising-based revenue model such as TripAdvisor and Google, are generally considered more reliable than websites which have commission-based revenue models, such as Amazon and Expedia. These in turn are considered more reliable than individual companies that sell their own services directly, such as hotel chains like Marriott and IHG.

6. **Fraud**. There is a positive correlation between the level of competitiveness in a given market and the prevalence of fraudulent reviews. It has been hypothesized that business-owners threatened by direct competition are more inclined to leave fake negative reviews of their competitors, as well as fake positive reviews of themselves. Yelp and Expedia, for example, employ algorithms to try to filter out reviews that appear to be coming from businesses rather than real individuals.

Potential challenges

Much of the research done previously relates to reviews of a commercial nature, such as travel services and consumer goods. Our next step was to identify additional considerations that may arise for an aid agency feedback system versus commercial systems. Below, we review these additional considerations, each of which could be the subject of further research. These challenges should not be considered as entirely comprehensive--rather, it highlights some of the largest concerns that are specific to aid agency feedback as opposed to commercial feedback.

TECHNICAL CONSIDERATIONS

One of the greatest challenges to the success of such a project would be access to adequate technology. While great advances have been made over the past years, there still exists a great disparity between technology in the developed and developing worlds. For example, people in the developing world more frequently rely on cellphones for their primary access to the outside world, and may not have desktop/laptop computers. Furthermore, data connections are less common and are slower.

This has important implications for how a feedback system would have to be designed. Any website would need to be optimized for low-bandwidth connections. The feedback system would also ideally accept SMS-based reviews.

One of the most recent developments with regards to messaging services has been the proliferation of new cloud-based instant messaging (IM) services such as WhatsApp and Facebook Messenger. These are low-bandwidth services where messages are stored and transmitted through the cloud, as opposed to traditional IM services which worked on a peer-to-peer basis. These services can be used over a WiFi internet connection, and therefore don't require a cellphone contract as SMS does. While SMS would still be a crucial platform to support, any solution would need to support some cloud IM platforms and provide reviewers with the choice of how to submit a review.

There would also be significant numbers of people who simply do not have any access to the Internet or cellphones. To serve them, it would be important to provide an option for collecting feedback at the point of delivery. One possible solution is mentioned in the section on "HappyOrNot", below.

PRIVACY AND REPRESENTATION

There are also social implications related to privacy and representation that would have to be considered in implementing a feedback system. Any system which handles user-created content, inherently raises **privacy concerns** for the reviewers, and potentially even for members of the aid agencies.

One expert we spoke with voiced concerns related to the handling of information collected through feedback mechanisms. They were concerned that when the United Nations (UN) gathers data about people in areas experiencing ongoing political instability, that data may end up falling into the hands of groups which are hostile to UN-affiliated agencies. Such an information leak could jeopardize the safety of individuals providing feedback as well as UN-affiliated agency staff. This type of scenario could result from inadequate information storage, inappropriate inter-agency sharing, or insecure information transmission methods, among other reasons. While the authors found no concrete evidence that such incidents have occurred in the past as a result of feedback from the affected populations, the concern is theoretically a valid one and provides a compelling illustration in support of the importance of privacy. In light of such concerns, any system would have to take the privacy of its users very seriously.

Whenever possible, feedback should be collected anonymously in the first place, so that the collectors and the feedback system are not burdened with the responsibility of data protection. Anonymous data collection may have the disadvantage of being less accountable and lacking the possibility of follow-ups, but it has the advantage of not creating safety risks, and of being easily shareable among the various concerned parties without requiring any special data handling or encryption. Certain general metadata, such as demographics, could be collected while still maintaining anonymity.

In instances where identifying information is required as part of the feedback process, data would have to be handled confidentially, would need to be stored in an encrypted format, would have to be transmitted using encrypted communications methods, and may have to be redacted if shared with external parties. The UN Secretariat provides guidelines related to handling sensitive

information in bulletin ST/SGB/2007/6[1] and an "Information Sensitivity Toolkit" from the UN's Archives and Records Management Section[2]. These procedures would be important to safeguarding individuals' privacy, but could create a significant administrative burden, which lends further weight to prioritizing anonymous data collection.

Another social challenge to consider is the question of **representation**: does a reviewer accurately represent the affected population? Is the sample size large enough? Will the feedback represent both genders in a balanced way? How will children be included? What about minorities or remote/hard-to-reach groups? Will literacy rate impose a challenge?

In regards to sample size, one expert we spoke to discussed leveraging diaspora and countless social media channels to spread news of their available support and ask for information from the local population. By all intents, the processed worked well but did come in the wake of a major international disaster with lots of press coverage.

Once contact is made with an affected person, the question becomes of whom is s/he representative. Frequently a single cellphone will be shared by an entire household, so feedback may be more representative of the householder than of other members of the household. An open question stands as to whether the experiences of other members of the household, such as women, children and the elderly, might be excluded.

Another barrier related to representation is literacy. A text-based feedback mechanism would clearly require literacy, meaning that written feedback could introduce a bias where ratings are only reflective of the literate population.

REACTION FROM AID AGENCIES AND DONORS

Aid agencies often face a competitive atmosphere. Multiple agencies apply for the same funding, which, in turn, requires increasing their publicity and lobbying of donors. A feedback system, such as AidAdvisor, would likely influence this competition among aid agencies in potentially surprising ways.

[1] *https://archives.un.org/sites/archives.un.org/files/ST_SGB_2007_6_eng.pdf*

[2] *https://archives.un.org/sites/archives.un.org/files/files/Information_sensitivity/Information%20Sensitivity%20Toolkit_2010.pdf*

As noted above, in an attempt to gain publicity, agencies handling private information may end up compromising it. Another factor to consider is that if aid agencies were to follow a similar pattern to corporations, we could expect that some would try to falsify reviews in order to improve the image of their own agency. Finally, if aid agencies perceive a feedback system as an obstacle to getting funding, they would likely resist its implementation.

On the positive side, a feedback system could be a powerful tool for donors. When an individual or an organization plans to grant funding to an aid agency, they want to have confidence that their funds will be used responsibly and with the greatest possible impact on affected populations. Giving donors feedback about the various agencies could influence where funds are directed, and could even increase the total donations, given donors' increased confidence in the system as a whole.

Having this feedback would certainly be a disruptive change and would undoubtedly face resistance, but making aid agencies more accountable and placing more weight on the experiences of aid recipients rather than on the ability of an agency to market itself, would ultimately be a net positive outcome.

Potential solutions

Throughout the research, two feasible options to resolve many of the challenges mentioned above came to the surface - artificial intelligence and customer reaction buttons.

ARTIFICIAL INTELLIGENCE

One important technology to leverage will be artificial intelligence (AI), a.k.a. machine learning; or more specifically, natural language processing (NLP).

One of the most expensive aspects of a feedback system would be processing the reviews once they come in. Traditionally they would be processed by a team of people who read every single submission and summarize the content. However, this is not feasible at scale. NLP offers a cheaper and faster solution by enabling computers to read the feedback and automatically detect the general attitude of each review, recurring themes and several other aspects. While the upfront cost of developing the necessary applications could be high, one can safely assume that over time it

would be much, much cheaper than hiring people to do this work manually.

Various companies are currently offering NLP services, including Amazon, Google and Microsoft. These companies also provide tools for the development of AI-based bots that can be integrated into apps such as Facebook Messenger to communicate with people conversationally and gather feedback. Microsoft, for example, has lately been demonstrating how their Azure services can be used to create a medical bot which gives tentative diagnoses by chatting with a patient.

AI tools have been improving dramatically over the last several years. At Amazon, the growth has been exponential. For instance, In 2010, when NLP was used to process SMS messages from Haitians, regional language differences posed a major challenge. Fast forward seven year and we find that Amazon's NLP services can now function regardless of the input language. This is a major leap in just a few years. Based on our conversations with experts, the understanding is that the AI field is expected to see another significant leap in the next 12 months and therefore any desire to commence technical work should be tempered for about a year.

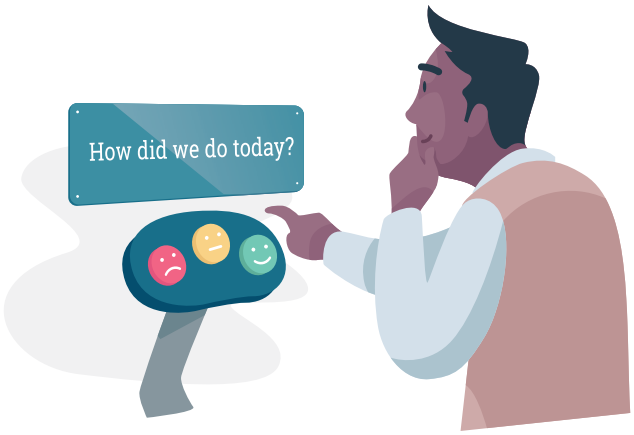
CUSTOMER REACTION BUTTONS

Since the machine learning/NLP solution proposed above might require significant investment (development, support, security, etc), we looked at the other end of the spectrum for a less-involved solution which may be easier and cheaper to implement. HappyOrNot[3] is a Finnish company which popularized the usage of so-called "Smiley Terminals" which offer a quick way of obtaining basic reactions from customers. These have been implemented in a wide range of places ranging from general businesses, to airports, to museums to educational institutions. The buttons are meant to be placed where a customer will see them immediately after having had a specific interaction with the business/organization. S/he can press one of four "smileys" corresponding to their level of satisfaction. The heart of HappyOrNot's product is their data-processing service. They use information from the Smiley Terminals to dynamically generate reports in live-time about

[3] *https://www.happy-or-not.com/*

satisfaction levels. This information can then be used by an organization to take note immediately of where a problem may be arising.

In the case of a feedback system, such an approach could be introduced alongside ongoing delivery activities and other feedback gathering mechanisms. For example, Smiley Terminals (or a mobile app-based feedback tool) could be used to gather basic satisfaction feedback immediately after a person has an interaction with an aid agency. Live-time reports could help the organization, or even remote support, to identify if feedback is skewing more negative than usual and immediately implement a course of action.



Maintaining the collection device (terminal or mobile device) at the front lines of a response, where power, dust or even heat can be problemsome, would be one of the main challenges to address with this solution.

Conclusions

Creating a feedback system would encourage greater accountability by aid agencies for the people they are assisting. It would reduce problems related to substandard interactions, abusive relationships, mismanagement, faulty supply chains, and other issues that may stand in the way of affected populations receiving aid. It would also provide donors with an additional window into how the affected population is being serviced by organizations the donors fund.

At the same time, such a system would raise new challenges. The problem of filtering reviews for accuracy might be largely solved with upcoming technological developments in artificial intelligence. Privacy and the data integrity of confidential information is a major consideration, but could be addressed through adequate encryption or by ensuring that data is anonymized before being transmitted or stored. Building a community-wide solution and handling all the data and support may be well beyond one agency's capacity and therefore dictate the use of a centralized customer-reaction-buttons service to start.

If a feedback system was to be designed and implemented, it is very clear that there would be several benefits, but also potential challenges. Overcoming these challenges would be key, but the authors believe that they can be with political will and emerging technologies.

Overall, a feedback system concept still appears to be a formidable and innovative solution that could be leveraged for positive change.