Mobile Image Search via Local Crowd: a User Study

Yefeng Liu, Todorka Alexandrova, Tatsuo Nakajima
Department of Computer Science
Waseda University, Tokyo, Japan
{yefeng, toty, tatsuo}@dcl.info.waseda.ac.jp

Vili Lehdonvirta
Helsinki Institute for Information
Technology, Aalto, Finland
vili.lehdonvirta@hiit.fi

Abstract

In this paper we present a on-field study for evaluating a crowdsourcing mobile social search application. With the help of the local crowd via social medias, this application assists foreign visitors in Japan by answering their image-based questions at hand in a timely fashion. We ran a controlled field experiment for 6 weeks with 55 participants. We found that the mobile crowdsourcing model demonstrated a reliable performance on response speed and response quantity: half of the requests were answered within 10 minutes, 75% of requests were answered within 30 minutes, and on average every request had 4.2 answers. Especially in the afternoon, evening and night, nearly 88% requests were answered in average approximately 10 minutes, with more than 4 answers per request. In terms of participation motivation, we found the top active crowdworkers were more driven by intrinsic motivations rather than any of the extrinsic incentives (gamification incentives and social incentives) we designed.

1. Introduction

Crowdsourcing [3] is a recent term that describes the act of outsourcing tasks, which are traditionally performed by an employee or contractor, to a large group of the Internet population (the wise crowd) by means of an open call. The tasks are typically ones that humans are good at but machines are not, such as annotating pictures, recognizing images, or ranking search results. Mixing mobile platforms and the crowdsourcing model potentially offers vast resources for computation. For instance, today, in most urban areas it is common that people allot a large amount of time for commuting or waiting for various events. Their time is actually fragmented into numerous small pieces of time and most of them are occupied with meaningless activities like playing with their mobile phones. We believe that, in this case, the crowdsourcing model provides a win-win solution for better use of this time by engaging people through their networked and ubiquitous mobile devices. However, a majority of current human computation and crowdsourcing systems (e.g., Amazon Mechanical Turk[1]) are passive services that are using worker-pull strategy to allocate tasks, and require relatively complex operation to create a new task. Consequently such systems fail to adapt to the mobile context where users require simple input process and rapid response.

We explore crowdsourcing platform designed for mobile users. Our design principle is simple: build the system around existing social media platforms — the popular services that inherently have the culture of sharing and participation, and are already well known and used by a large and growing number of users. In this paper, we practice the platform in the context of one specific application area – mobile image translation/search. This refers to camera phone applications that attempt to solve the problem of translating text written in an unfamiliar script. This kind of a system is particularly useful for travellers and short-term residents in a foreign country. Ordinary digital pocket translators are useless if the user is unable to input the text they see (e.g., Japanese or Chinese text for Western visitors). Image-based mobile translation/search systems typically employ Optical Character Recognition (OCR) algorithms to extract text out from image, and apply Machine Translation (MT) technologies to translate the text. Several systems like this have been proposed during the past decade [2, 5]. However, only a handful of them have gone beyond pilot development. Even the state-of-the-art in this field, such as Word Lens[2] or Google Goggles[3], demonstrate very limited performance in real-world situations (i.e., complex background, dark environment, blurred photos, irregular fonts, size or formats, etc.), especially for non-Latin scripts.

To solve the problem, we present *UbiAsk*, a social media crowdsourcing application built on top of existing social networking infrastructure. UbiAsk provides translation services and situational advice to mobile users in unfamil-

[1] https://www.mturk.com Last checked: February 2011
[2] http://questvisual.com/ Last checked: March 2011
[3] http://www.google.com/mobile/goggles/ Last checked: March 2011

1533-2306/11 $26.00 © 2011 IEEE
DOI 10.1109/RTCSA.2011.10

iar environments. Instead of applying machine algorithms, we draw on the power of ordinary people in the cloud via social networks to solve the difficult computational problems such as image recognition and text translation. Since the workload of each task in the image based mobile translation/search service is lightweight enough to be described as a micro-task, the tasks are perfectly suitable to be distributed to large groups of casual workers.

To evaluate the UbiAsk system, we conducted a controlled, between-groups field experiment for the duration of six weeks (from late January 2011) with 55 participants. 19 participants were foreign visitors in Japan, the majority from North America and Europe. They served as requesters. 36 participants were Japan-based Japanese/English speakers, who served as local experts. In this evaluation, the main focus was on response speed and quantity. Quality of the answers and how to assure it was mainly left for the next stage of the project, although requesters were asked to evaluate the overall quality of the answers. In terms of user engagement, we also investigated how the crowd-worker participation is affected by different incentive mechanisms.

2 The Crowdsourcing Mobile Social Search Application

Working with potential users we designed UbiAsk, a crowdsourcing platform for helping travelers in Japan by answering their image based questions at hand in a timely fashion. In the UbiAsk system, users can issue requests via several channels that use a common API. Native mobile applications and email are the currently implemented channels. The requested task is pushed to a community of voluntary local experts in the form of an open call via different social media platforms (Twitter, Facebook, etc.) and email. The crowdsourced result data is not only returned to requesters but also visualized on location based social mapping and augmented reality (AR) platforms (Sekai Camera[4] and Ushahidi[5], etc.). This gradually results in an information pool that constitutes a public good.

After several rounds of design iteration including formative evaluation with potential users [4], we implemented a prototype version of UbiAsk. The proxy server was built as a REST style web service with open API to clients. A task assigner was developed as the component with responsibility for assigning task to and receiving answers from local experts. The task assigner also connects to the social media platforms APIs. All back-end programs are implemented in Java. Requester can use an iPhone application or mobile email to make a request. The questions were pushed to local experts via email and Twitter. The local experts can submit their answer by simply replying the email or tweet. The maximum living period of a request was 12 hours and server will reject any later answers. All result data were visualized on an Ushahidi based interactive map.

[4]http://sekaicamera.com/ last checked: February 2011

[5]http://www.ushahidi.com/ last checked: February 2011

3 Expertment

After UbiAsk was implemented, we launched a field user study with 55 end-users (19 requesters and 36 local experts) for six weeks from January 2011. On the requester end, the user study is designed to find out if UbiAsk could provide answers of satisfactory speed and quantity; on the crowd-worker end, the relevance of the activating participation to different motivational mechanisms is examined.

3.1 Participants

The 19 foreign requesters were mainly recruited from a total of thirteen popular international travel forums such as Lonely Planet thorn tree[6], Tripadvisor[7], etc. The majority of the requesters were short-term travelers, the exceptions included Japanese returnees, foreign students, foreign employees, foreign housewife and visiting scholars. All, but one, of them were from Western countries, the only exception was from Southeast Asia. Of the 36 local experts, 17 were recruited from Internet by twitter adverts or posts on local forums. The rest of the local experts were mainly undergraduate students, graduate students, and staffs from Waseda University that were recruited by emails.

3.2 Procedure

For collecting demographic information, all participants completed a pre-test questionnaire before the requesters left their countries for Japan. The iPhone app with the usage instructions and the information of the user study were emailed to the requesters before their departures. When requesters arrived in Japan, they were told to use the service freely.

3.2.1 Overall Performance

In the pre-test questionnaire two important questions were asked: the user's *expected* response time, and the user's *bearable* response time. Based on the answers we can further identify three time periods with regards to user satisfaction: we assume a user will be satisfied if the first answer comes before his/her expected response time; a user will be unsatisfied if the first answer comes after his/her bearable response time; and it is acceptable if the first answer come

[6]http://www.lonelyplanet.com/thorntree/ Last checked: February 2011

[7]http://www.tripadvisor.com/ Last checked: February 2011

between the two time points. The survey result is shown in Figure 1, if the response time is less than 10 minutes, we can satisfy all users. It is beyond our expectations that even if it takes 30 minutes to have the response, only 25% of user will be unsatisfied. But if the waiting time becomes more than one hour, it will be unacceptable for most of the users.

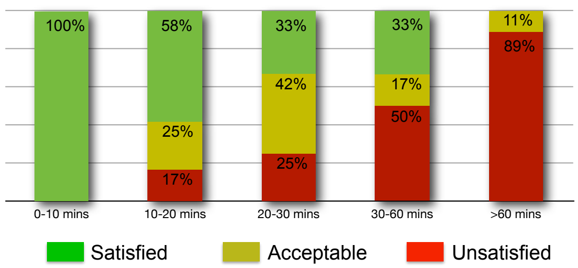

Figure 1. Requesters' Overall Requirements on Response Time Results

3.2.2 Incentive Mechanisms Comparison

In terms of participation motivation mechanism design, local experts were randomly placed in three incentive groups, among which there were two experimental groups and one control group.

A – derived from the GWAP concept, a location based mobile social game was designed and implemented as participation incentive. The main interface of the game was a Google map based real world map, which was divided into non-overlapping hexagons. The goal of the game is to conquer territories. Every hexagon has one owner, who is the player with the highest number of "task-done" in the area. In other words, crowdworkers need to compete with each other to get the "Local Expert" title of an actual location on the map.

B – a simple feedback mechanism - social psychological incentive that was known to be effective [1] - was applied: when a local expert provided an answer, the system will rapidly reply him the number of existing answers of that question and previous answers' content.

C – control group, no additional motivational method was applied.

For performance measures, we instrumented our prototype to log the timestamp and the question/replies ratio. To measure user satisfaction, ease of use, and overall experiences, we administer a post-test questionnaire with Likert scales. The questions covered typical usage patterns, content's quality, features preferences, likes, dislikes, and suggestions.

4 Study Results

4.1 Performance Results

In all, the system recorded 180 interactions, covering 33 questions and 147 answers. We expected to see a much bigger number of requests, however eventually there were only 11 (58%) requesters that submitted their questions to the system. On the other end, 23 (64%) local experts answered at least once question. Of the 147 answers, the local experts that recruited from Internet provided 93 of them. The seven most active local experts accounted for nearly 70% of the answers. The requests are relatively equally distributed across the day expect early in the morning and mid-night.

Figure 2 shows the overview of the response time (the first answer). Approximately 50% of the questions were answered within 10 minutes, and 27% of them were responded within just 5 minutes. Three-quarters of the first answers arrived within half an hour. Only 6% of the answers were arrived after 60 minutes, and most of them were asked during midnight or in the morning. The 9% of the requests were never being answered, the main reasons were bad timing (e.g., local experts were busy) and boring question's content (e.g., translation or explanation of a long text).

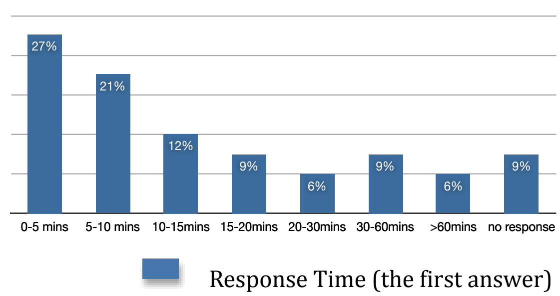

Figure 2. The Performance Results: Response TIme

According to the pre-test questionnaire's result, we set a strict rule for the response time: if the first answer of a question arrives after 60 minutes, we consider this question has no answers. Based on this rule, we show a time-of-the-day breakdown of the average response speed, average response rate, and answers per request in Figure 3. Generally speaking, the overall performance is reliable: except midnight and early morning, nearly 88% of the requests could be answered in approximately 10 minutes, with more the 4 answers per request.

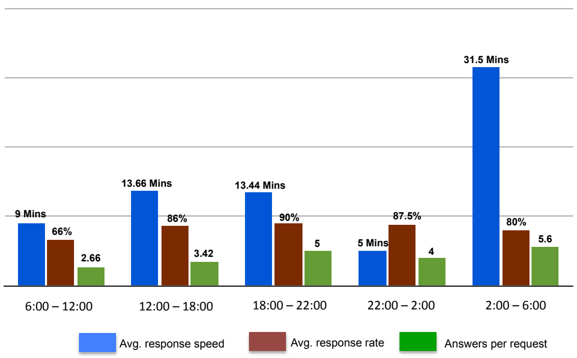

Figure 3. The Time-of-the-Day Breakdown of the Average Response Speed, Response rate, and Answers per Request

4.2 Incentives Comparison Results

The incentives comparison results were more complicated. Figure 4-I and Figure 4-II shows the response time and number of answers by groups. Local experts from Group A demonstrated a much more active engagement of the system, 67% of the first answers were provided by them. In terms of quantity, more than half of the total number of answers was likewise produced by people from group A. In the meanwhile, the control group produced a better outcome than group B. However, the web access logs did not support these findings. The access history reveals that the page of the location-based ranking only drew participants' attention in the beginning stage of the study. The access number shows an obvious decrease over time, and eventually dropped to zero per day in the last phase of the experiment. Moreover, in the post-test questionnaire there was no participant that claimed that he/she answered those questions because of the location-based game.

To better interpret this phenomenon, we further analyzed the top active users. We found four of the top five most frequent local experts were from group A, and the exception is from group C. Based on their replies in the questionnaire, all of them have demonstrated strong intrinsic motivations, i.e., *"I want to help the people in trouble"* or *"I want to introduce Japanese cultural to foreigners"*, which may have a stronger impact than the extrinsic motivations we provided.

On the hand, however, if we measure the user's participation as the number of users in the system (i.e., users who answered at least one question), and give equal weight to every user no matter the amount of the answers they provided (i.e., minority highly active user's data will have less power of influence to the final comparison result), the result will be more explainable. In this case (see Figure 4-III), we observed the same number (38%) of active users were

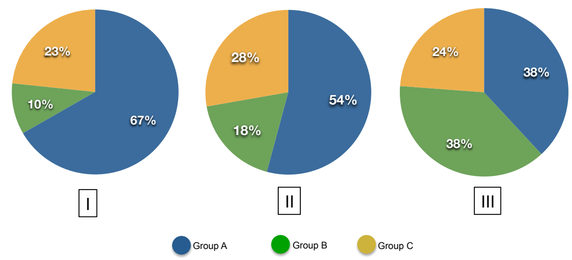

Figure 4. I: Response Time by Groups; II: Number of Answers by Groups; III: Number of Active Users by Groups

from group A and B, and fewer users (24%) were from the control group.

Overall, we argue that although the most frequent participants might be more motived by intrinsic incentives, the effectiveness of the designed extrinsic incentives to the rest of less self-motived users was still verified. Nevertheless, base on the results, we could hardly come to the conclusion that the proposed game-based incentive has a greater impact than the social psychological incentives. We believe it is not only because the game may not very interesting for the local experts, but also because the expected fierce competition did not happen due to insufficient requests' number.

References

[1] CHESHIRE, C., AND ANTIN, J. The social psychological effects of feedback on the production of internet information pools. *Computer Mediated Communication*, 13 (2008), 705–727.

[2] HARITAOGLU, I. Infoscope: Link from real world to digital information space. In *Proceedings of Ubicomp 2001: International Conference on Ubiquitous Computing* (2001).

[3] HOWE, J. The rise of crowdsourcing. *Wired Magazine*, 14.06 (2006).

[4] LIU, Y., LEHDONVIRTA, V., KLEPPE, M., ALEXANDROVA, T., KIMURA, H., AND NAKAJIMA, T. A crowdsourcing based mobile image translation and knowledge sharing service. In *Proceedings of the 9th International Conference on Mobile and Ubiquitous Multimedia* (2010), MUM '10.

[5] NAKAJIMA, H., MATSUO, Y., NAGATA, M., AND SAITO, K. Portable translator capable of recognizing characters on signboard and menu captured by built-in camera. In *Proceedings of the ACL 2005 on Interactive poster and demonstration sessions* (2005).