

CrowdCafe - Mobile Crowdsourcing Platform

Author 1

University name
Address
ZIP, Country
author@university.edu

Author 2

University name
Address
ZIP, Country
author@university.edu

Author 3

University name
Address
ZIP, Country
author@university.edu

Author 4

University name
Address
ZIP, Country
author@university.edu

Author 5

University name
Address
ZIP, Country
author@university.edu

ABSTRACT

In this paper we present a mobile crowdsourcing platform CrowdCafe, where people can perform microtasks using their smartphones while they ride a bus, travel by train, stand in a queue or wait for an appointment. These microtasks are executed in exchange for rewards provided by local stores, such as coffee, desserts and bus tickets. We present the concept, the implementation and the evaluation by conducting a study with 52 participants, having 1108 tasks completed.

ACM Classification Keywords

H.5.3 Group and Organization Interfaces: Computer-supported cooperative work

General Terms

Human Factors, Design, Performance

Author Keywords

Crowdsourcing, User Interfaces, Motivation

INTRODUCTION

Crowdsourcing is the practice of outsourcing work to an unknown group of people via the Internet, instead of assigning it to internal employees [5]. Crowdsourcing has been so far very successful in performing tasks which are still hard to automate using algorithms, while they can be relatively easily solved by humans, such as image object recognition, annotations, feedback collection, and similar.

Requestors are the people who want to crowdsource their work. They publish tasks on crowdsourcing platforms where requestors meet potential workers - people who solve tasks for monetary reward, curiosity or other motivations. Some examples of crowdsourcing platforms are Amazon Mechanical Turk (MTurk), CrowdCloud, MicroWorkers, Mobileworks, CrowdFlower. In general workers need to perform

tasks from their desktops or laptops, as tasks are designed for non-mobile screens. However, many people, when they have time with their computer, prefer to perform some things related to their job or just to have fun watching something.

People spend everyday some amount of time riding a bus, standing in a line at a grocery store, waiting for a doctor appointment. During this time they can read a pocket book or use their smartphones. Many people end up checking their social network profiles. We present a crowdsourcing platform CrowdCafe, where people perform microtasks, specifically designed for mobile execution, in exchange for non-monetary rewards provided in local stores, such as coffee, desserts and bus tickets.

STATE OF THE ART

Vaish et al. [8] presented an Android application *Twitch*¹, which in order to unlock a phone, asks its owner to answer a simple question, such as: how many people are around, or which activity the owner is doing now. Heimerl et al. [4] presented *Umati* – communitysourcing vending machine, which helps to attract a specific local group of people (e.g. people with deep knowledge in computer science) to perform tasks on the screen of the vending machine in exchange for snacks. Luon et al. [7] proposed a mobile system *Rankr* for crowdsourcing opinions via pair comparison of images and sentences on mobile phones. In [6] Kittur et al. analyzed how different aspects of crowdsourcing could be improved, they also challenge the community to revolutionize the conception of what a crowdsourcing platform is.

With CrowdCafe we want to investigate how during semi-occupied situations (such as riding a bus, traveling by train or waiting in a line) people can perform microtasks using their smartphones, not for the purpose of making income, but to have fun and to benefit out of this time.

CONCEPT

The concept of CrowdCafe affects three aspects of crowdsourcing: *tasks user interfaces*, *tasks classification* and *workers motivation*.

Paste the appropriate copyright statement here. ACM now supports three different copyright statements:

- ACM copyright: ACM holds the copyright on the work. This is the historical approach.
- License: The author(s) retain copyright, but ACM receives an exclusive publication license.
- Open Access: The author(s) wish to pay for the work to be open access. The additional fee must be paid to ACM.

This text field is large enough to hold the appropriate release statement assuming it is single spaced.

¹<http://twitch.stanford.edu/>

Tasks User Interfaces

In order to provide a good user experience of tasks execution on CrowdCafe we want to apply the best user interface (UI) practices from current mobile applications, such as: *feed* to present all the content as a list, without sidebars; *big full width buttons* to make it comfortable to press them with a thumb; *swipe for action* to keep a user interface very clean without buttons, where, depending on a direction and a distance of swiping a UI element, different actions are triggered (was announced with the MailBox mobile application²).

Tasks Classification

As described by [1] on the platforms such as MTurk or CrowdFlower it is hard for workers to select a task to work on, because descriptions are not informative enough and they never know how much time they will spend on execution. On CrowdCafe we decided to split all the tasks by completion time in 3 clear categories:

- “Espresso” - about 10 seconds to be completed, with mostly only clicking and swiping actions required (e.g. to identify the sentiment of tweets or to compare pairs of images);
- “Cappuccino” - about 2 minutes to be completed, with some typing and learning required (e.g. to fill up a short survey, to annotate images);
- “Wine” - more than 5 minutes to be completed, requiring a worker to be in a specific context or a location (e.g. to go to a grocery store and to make a photo of a particular product with its price).

Workers Motivation

We do not position CrowdCafe as a source for primary or secondary income. We consider CrowdCafe as a way to convert time, which will be otherwise wasted, to enjoyable rewards. According to Dan Ariely [2] the smaller reward now is more desirable than the bigger one later. So we want to minimize the time frame between a worker starting task execution and a worker enjoying a reward, by providing micro rewards from local stores, such as coffee or dessert in a university bar. This can help workers to start working on tasks and in 15 minutes to feel an outcome of their work by drinking a coffee.

IMPLEMENTATION

We have implemented the CrowdCafe as a website³. It has two main components: i) “Kitchen”, where requestors design and publish tasks from a desktop; ii) “Cafe”, where workers perform tasks using their mobile phones and get reward coupons. The CrowdCafe code is open-sourced and is available on GitHub⁴, where more details about the implementation can be found.

Requestors Interface

In “Kitchen” requestors create tasks (Figure 1), defining: 1) title, 2) instructions, 3) category, 4) preselection logic, 5) user interface, 6) input dataset and 7) some settings for the quality

²<http://www.mailboxapp.com/>

³<http://crowdcafe.io/>

⁴<https://github.com/CrowdCafe/crowdcafe>

control. The first three parts are trivial and we focus on the other four.

The screenshot shows a web form for creating tasks. It is divided into seven numbered sections: 1. Title (with a sub-note 'Be specific. 1 - 5 words only.' and a text input containing 'Sentiment Analysis'), 2. Instructions (with a sub-note 'Clear instructions. Not more than 140 characters.' and a text input), 3. Category (a dropdown menu), 4. Preselection (a dropdown menu with 'worked on a task' selected and a text input with '#20 Instagram #Trento'), 5. User Interface (a text input with a URL and a sub-note 'URL leading to raw HTML file, check existing templates on GitHub.'), 6. Data (with radio buttons for 'No data (survey)', 'Dataset .CSV', and 'Social Network', and a 'dataset template' label), and 7. Settings (with input fields for 'Data items/instance' (5), 'Min answers/data item' (1), and 'Min confidence, %' (50)). At the bottom of section 7 are two buttons: 'Ask us to design a task for you (free!)' and 'Save task'.

Figure 1. Requestors user interface

Preselection

Requestors can define to which workers the task will be visible by adding a set of restrictions, such as “worked” or “did not work” on particular tasks. This simple preselection logic is powerful enough to route surveys to particular workers, to create tasks which are only visible to workers who performed some skill test tasks.

User Interface

We decided to leave a lot of freedom for requestors, so they can design the UI of their tasks using HTML and CSS. In order to not design from scratch and accumulate the best practices such UI templates are stored in the public GIT repository. To apply a particular template, requestors need to insert its URL, which refers to a raw HTML file.

Input Dataset

There are three options for input data: 1) no input data (a survey task), 2) data uploaded from a .csv file, 3) data from a social network feed (e.g. Twitter, Instagram) on a particular topic, defined by a hash-tag (such as #helsinki). Requestors also define how many data items (rows in case of .csv file, or tweets in case of Twitter) a worker should process in one time.

Quality Control

In order to have some redundant data for calculating agreement level, requestors define a required amount of judgments per each data item. Later we plan to let requestors to define minimum and maximum amount of judgments and required agreement level, so the platform will accept extra judgments if necessary.

Workers Interface

The mobile website for tasks execution is very light. It has four main sections: 1) home page and side menu (Figure 2a), 2) tasks listing page (Figure 2b), 3) task execution page (Figure 2c), 4) reward page (Figure 2d).

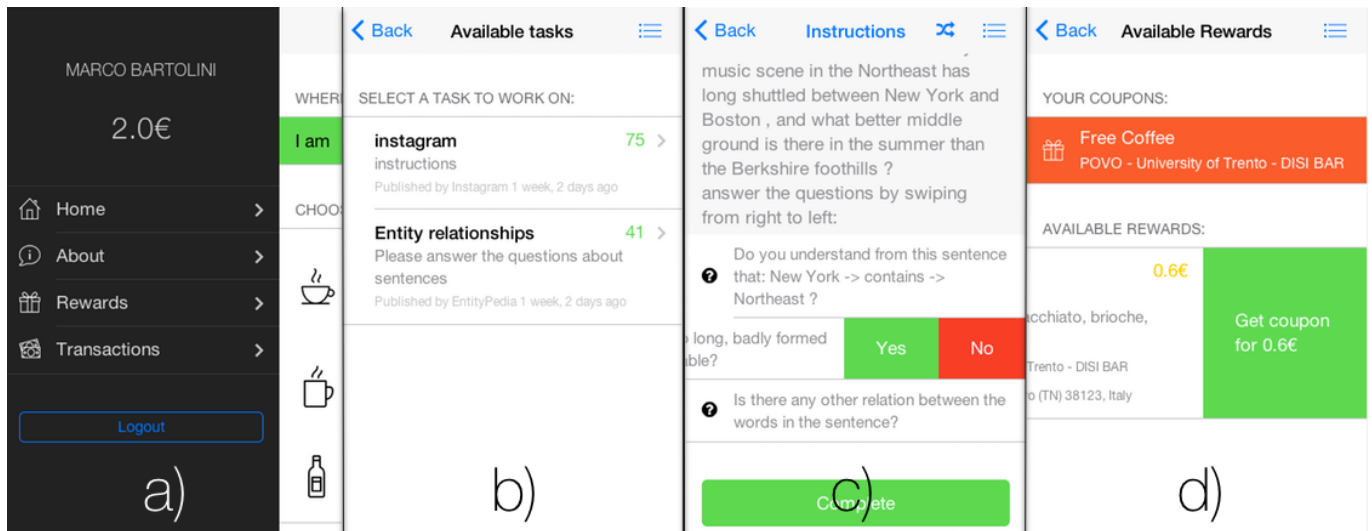


Figure 2. Workers User Interface. a) Task categories and side menu, b) Available tasks, c) Example of a task, d) Reward page

Home page

On the home page workers see the list of task categories. In the right part of the top bar on every page there is a button which opens the side menu. There is a context select box where we ask workers to define where they currently are (e.g. in a bus, having lunch). On the side menu workers see their name, the amount of money they have earned and four links: 1) “Home”, 2) “About”, where the service is described in details, 3) “Rewards”, 4) “Transactions” - where the history of all earnings (related to tasks execution) and spendings (related to purchasing coupons) is presented. This side menu also has a logout button.

Tasks listing page

On the tasks listing page workers see a list of tasks with its description and amount of instances available.

Task execution page

When workers start executing a task, they first see a pop up window with instructions. Workers are expected to read them and after they can start to perform a task. Workers should fill up all the necessary fields otherwise they can not submit the task. After they submit one task, they are redirected to another instance of this task. If there are no any instances available workers are redirected back to the tasks listing page.

Rewards page

On this page a worker can see a list of available rewards with its price and address where they can get them.

EVALUATION

In order to evaluate the concept of CrowdCafe along with our implementation we decided to post two tasks and one follow up survey:

- “Instagram #Trento” – in this task workers were expected to look on images potentially about Trento and provide two actions: 1) add three relevant tags to each image, 2) specify whether this image really represents Trento as a city.

- “Sentence Analysis” – in this task workers were expected to read a short text and to answer two questions: 1) is a given relationship between two nouns correct from the content of the text? (yes, no, i don’t know) 2) does the text consist of only one clear sentence? (yes, no)
- Survey – in this task we wanted to collect the workers feedback about tasks and the CrowdCafe platform in general.

For the first task we uploaded a dataset of 1000 sentences, splitting them in 334 tasks of 3 sentences, asking for at least 3 judgments for each sentence. For the second task we took a feed of 231 images from Instagram with a hashtag #trento, splitting them in 77 tasks of 3 images, asking for at least 3 judgments. Both tasks were qualified as “Espresso” tasks with a reward of 0.03 euro. The final survey was qualified as “Cappuccino” task with a reward of 0.33 euro.

We prepaid 84 coffees (0.60 euro each) at the bar on our faculty and left there a list of 84 unique codes. When workers earn enough money, they can purchase a coupon which they exchange for a coffee or a dessert in the bar.

In order to approach the first users we sent email invitations to people from our research group (30 people) and distributed 20 printed posters around our faculty building. It helped to get 80 sign ups on the platform in 1 day.

Results

We collected all the judgments for all tasks in two days from 52 workers. Two workers were identified which used the vulnerability in the code and submitted extra 400 equal judgments. These judgments were removed from the analysis. Some people did not specify a place where they performed tasks, so about a half (46.9%) of all responses did not have associated place. Out of those, which had: 56.40% were performed on a workplace, 14.10% – outside, 13.13% – in a bus, 11.83% – at home, 4.38% – in a train, 0.16% – walking.

Task 1 - “Instagram #Trento”

The average execution time for this task was 107.31 seconds (317 task responses, median 87, standard deviation 88.03 sec). We received 791 judgments for 231 image. There were 737 (93.17%) images according to instructions and included three or more tags, while 54 included only 1 or 2 tags.

We used Fleiss Kappa [3] to assess the reliability of the provided tags. We could not use the tags directly to estimate the Kappa values due to the open vocabulary of the tags provided by workers. Therefore, 3 experts categorized all the tags into 10 predefined clusters and voting system has been used to select the ground truth cluster for each tag. Finally, cluster names were replaced with the real tags and Kappa values were calculated. The overall Fleiss Kappa value is 0.4416 with 0.0154 error that is an indication of a moderate agreement. The 95% confidence interval of Fleiss Kappa is [43.4, 44.9]. Also, the *p-value* is less than 0.0001 which shows that the observed agreement is statistically significant.

Task 2 - "Sentence Analysis"

The average execution time for this task was 62.85 seconds (1006 task responses, median 16 sec, standard deviation 276.76 sec). For each sentence we received from 3 to 5 judgments. These judgments have very low agreement level (we did not find any agreement in the majority of sentences).

Survey task

We sent email invitations to 18 people who provided at least one judgment to both "Instagram #Trento" and "Sentence Analysis" tasks. There are 15 people completed this survey. Out of these people 66% responded that 0.03 euro is enough reward for such tasks, 80% responded that they will use the platform in future, the average interest on scale from -3 (very negative) to +3 (very positive) in "Instagram #Trento" is 0.93, in "Sentence Analysis" is -0.60. The average overall satisfaction about CrowdCafe platform is 1.93 on the same scale.

Discussion

When we designed our two tasks we did not expect that it would take people so much time (107.31 and 62.85 seconds) to complete them. It showed that we classified this tasks not correct. Still we had workers, providing many judgments and in the survey the majority of workers responded, that 0.03 euro is enough reward for completing such tasks (which is only about 2 euro per hour). Several people mentioned that they performed tasks thinking about coffee and not money.

The overall quality of tags provided in "Instagram #Trento" task is high. All the tags were relevant. Even when workers did not follow the instructions and provided less than 3 tags or tags in other language than English (54 judgments), tags were still relevant. We found very low agreement between workers stating that an image characterizes Trento. There are two possible reasons: 1) instructions were not clear enough and some people marked only images which have some famous Trento building on it, while others marked all images which they believed were made in Trento, 2) some users pointed an issue that this button did not work well in the native Android browser.

In the "Sentence Analysis" task there is very low agreement between workers, because many workers did not understand the instructions clearly and some workers did not pay enough attention and simply provided random results. This is also the reason of the big standard deviation and big difference between mean and median execution time for this task. In addition the survey results show that the interest in this task was much lower than in the "Instagram #Trento" task.

CONCLUSION

In this paper we have described the concept and evaluated the implementation of the mobile crowdsourcing platform CrowdCafe, where people can and are willing to perform microtasks during short spans of free time in exchange for tangible rewards such as coffee. We showed that for well-designed tasks with clear instructions, even without any specific control (e.g. gold data, skill tests), workers provide results of a very high quality (93.17%). In tasks with ambiguous instructions the quality of results is poor.

In future we plan to investigate: i) the variety of tasks people can perform on their smartphones with better or the same quality as on regular "desktop" crowdsourcing platforms, ii) the workers productivity with different user interface approaches in tasks design (e.g. radio buttons, swiping to action, set of buttons), iii) how different motivation strategies (e.g. cash, coffee, donation to charity, no reward) affect the workers productivity.

REFERENCES

1. Anonymous, Anonymous, Anonymous, Anonymous, and Anonymous. Toward effective tasks navigation in crowdsourcing. AVI '14, ACM (New York, NY, USA, 2014), 1843–1852.
2. Ariely, D., and Wertenbroch, K. Procrastination, deadlines, and performance: Self-control by precommitment. *Psychological Science* 13, 3 (2002), 219–224.
3. Fleiss, J. L. Measuring nominal scale agreement among many raters. *Psychological bulletin* 76, 5 (1971), 378.
4. Heimerl, K., Gawalt, B., Chen, K., Parikh, T., and Hartmann, B. Communitysourcing: Engaging local crowds to perform expert work via physical kiosks. CHI '12, ACM (New York, NY, USA, 2012), 1539–1548.
5. Howe, J. The rise of crowdsourcing. *Wired* 14, 14 (October 2006), 1–7.
6. Kittur, A., Nickerson, J. V., Bernstein, M., Gerber, E., Shaw, A., Zimmerman, J., Lease, M., and Horton, J. The future of crowd work. CSCW '13, ACM (New York, NY, USA, 2013), 1301–1318.
7. Luon, Y., Aperjis, C., and Huberman, B. Rankr: A mobile system for crowdsourcing opinions. vol. 95. Springer Berlin Heidelberg, 2012, 20–31.
8. Vaish, R., Wyngarden, K., Chen, J., Cheung, B., and Bernstein, M. Twitch crowdsourcing: Crowd contributions in short bursts of time.