Crowdee: Mobile Crowdsourcing Micro-task Platform for Celebrating the Diversity of Languages

Babak Naderi, Tim Polzehl, André Beyer, Tibor Pilz, Sebastian Möller

Quality and Usability Lab, Telekom Innovation Laboratories, TU Berlin, Germany

{babak.naderi, tim.polzehl, andre-beyer, tibor.pilz, sebastian.moeller}@telekom.de

Abstract

This paper introduces a novel crowdsourcing platform provided to the community. The platform operates on mobile devices and makes data generation and labeling scenarios available for many related research tracks potentially covering also small and underrepresented languages. Besides the versatile ways for commencing studies using the platform, also active research on crowdsourcing itself becomes feasible. With special focus on speech- and video recordings, the mobility and scalability of the platform is expected to stimulate and foster data-driven studies and insights throughout the community.

Index Terms: mobile phone app, crowdsourcing, field tests, scalable studies, tools, labeling, recording

1. Introduction

Crowdsourcing micro-task platforms open up unique opportunities to generate large and specific data for researchers and developers in many ways. They use human intellectual power to solve problems that are still extremely complex for computer algorithms, e.g. image tagging or transcribing. The main advantages of using crowdsourcing systems are the high speed of data collection, the flexibility of design, and low costs.

In this paper we introduce *"Crowdee"* - a micro-task crowdsourcing platform which provides a native mobile application as front end for crowd workers. As a result, it provides a unique opportunity for communities to collect realistic data such as impressions, opinions, audio-visual recordings and ample sensor data in any location or environment. Thus, studies on Crowdee can be designed to cover many of the diverse conditions required for advanced data generation from studies in the field.

To give some examples, crowdsourcing micro-task platforms facilitate data labeling and data generation for researchers. Affective labeling, semantic extraction and transcription of text, speech and video data can be provided by the scalable and diverse human workforce. In addition, Crowdee offers audio and visual data recording by its crowd members. It brings not only fast, diverse and inexpensive data generation but also gives the opportunity to capture field quality dataof realistic conditions including, e.g. background noise, illumination, and all kinds of interference happening in the field.

2. Functionalities

In the following some general functionalities, as well as functionalities of Crowdee that are expected to be beneficial specifically to the speech community are described.

In general, *jobs* are created by *job providers* and are performed by *workers*. Each job can be allocated a number of repetitions. During job creation, *tasks* are generated as instances of a job. These tasks can include and iterate on *dynamic contents* as well, e.g. iterating through a number of different database entries. In addition, the job provider can specify a maximum number of tasks that can be carried out by one individual worker.

Principally, a job is formally similar to an HTML form. Each job contains one or more *questions* that can correspond to items of questionnaires. Each question consists of two parts: a) an information part that can contain text, images, audio or video content, which can be provided as static or dynamic contents; and b) an answer part including selected answers types like free text input, multiple choice, radio-buttons or sliders. Question organization includes different methods for ordering and randomizing items and answer options.

2.1. API and Basic Job Parameters

Using the comprehensive RESTful API provided by Crowdee it is possible to create, update, extent, and stop any job at any time. Introducing sequentially nested job-chains, it is possible to develop systems that push the outcome of a task (e.g. voice recorded by a worker) as an input of a new job (e.g. transcription job). Jobs can be given a start and end date. This functionality can be used to automatically enable work at any time or schedule repetitive jobs. In addition, each job length is defined by an amount of minutes the workers can actively work on the task.

2.2. Question and Answer Specifications

Four types of answers are currently supported in this system i.e. text input, selection, file generation, and external activities. Workers can type a text freely or select one or more items for answering selection questions. Additional parameters determine the minimum and maximum number of multiple choices. Items can be text, images, videos, or audios. In file generation questions, workers take photos or record audio or video content. In Figure 1, a question with voice record answer type has been illustrated. External activities are a type of answers which refers the user to a task out of the Crowdee platform like filling a survey on external sites or conducting a phone call to external mailboxes.

Question types and answer types can be determined by respective options. The order of appearance of items of any list inside a job can be randomized for each task. Same applies to all answer options. In addition, the order of items and answer options can be reversed. This option can be set to alternate from one task to the other. Moreover, it is possible to pin an item to a specific position in the item list by setting a fixed position.

2.3. Data Logging

Comprehensive log data is provided from the Android app by integrating a data logging system developed for performing

dynamic multivariate GUI testing. As a result, app session information containing the screen flow and the time spent in each screen is logged. The web analytics platform Piwik [1] is modified for visualizing collected log data.

In addition, the time spent by a worker on each question is reported in the log data.

2.4. Adaptability to Other Languages

All platform components were carefully designed to be highly dynamic and flexible in terms of languages we support. In principle, Crowdee can be switched to any other language including Persian or Mandarin languages. The current version supports German and English.
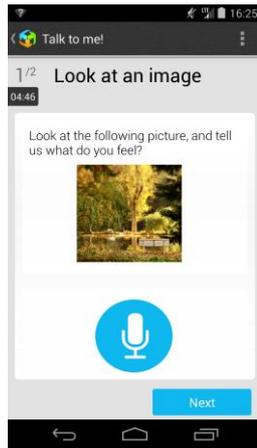

Figure 1: *A voice recording question in the Crowdee app (Android version)*

3. Prior Work and Current Research

Being a development of the Quality and Usability Lab of Technische Universität Berlin, the Crowdee platform also serves as an active research place. Much of the prior work of the Crowdee development team has focused on automatic quality and usability estimation using online or remote mobile testing, cf. [2, 5]. [6] has focused on affect interference and mobile questionnaire evaluation. [7] analyzed personality from test participants and developed an automated system for estimating personality as well as its online assessment. [3] evaluated scale dependencies in speech assessment. [4] evaluated tactile feedback from mobile devices.

Current research focusses on the automated characterization from speech and text data, motivation of crowdsourcing workers, and automated quality control in crowdsourcing platforms. With respect to the latter, manual verification of crowdsourced data is oftentimes cumbersome. Still, results from anonymously paid micro tasks may be corrupted by different kinds of noise, e.g. demotivated workers, wrong job-design, adverse job delivery, or noisy transmission.

In comparison to repetition-based approaches on quality control for item-based questions, which require more tests and participants, several techniques were proposed in the literatures that aim to identify unqualified responses. Frequently, any kind of gold-standards has been used. Further, [8] showed a reduction of invalid responses from 48.6% to 2.5% after inserting verifiable answers into the tests. Also automatic procedures using machine learning techniques were

developed. [9] showed that unreliable assessors exhibit unusual patterns, e.g. with regard to periodic or abnormal timings, or ratings.

The authors' research focusses on the development of an automatic quality prediction model for crowdsourcing data collections on basis of manifold observations, e.g. user characteristics, task requirements, log-data patterns, motivational match between tasks and workers, as well as executing environment.

4. Conclusions & Outlook

This paper introduced the novel mobile micro-task crowdsourcing platform Crowdee, provided to the community by the Quality and Usability Lab of TU Berlin. General functionalities as well as specific opportunities beneficial to the speech community are given in more detail. As a core-set of functionalities, Crowdee allows for very realistic audio and visual data recording in the field, as well as for fast labeling including various scales. Its application flexibility makes it very beneficial especially to small language communities. Simultaneously, the system also provides unique opportunities for research on crowdsourcing itself, such as automatic quality prediction or motivation steering. Both topics are active research fields in the authors' lab.

5. Acknowledgements

6. References

[1] Piwik. (2014, April 11). Piwik [online]. Available: http://www.piwik.org

[2] R. Schleicher, T. Westermann, and R. Reichmuth, "Mobile Human–Computer Interaction," Quality of Experience. Springer International Publishing, 339-349, 2014.

[3] J.-N. Antons and S. Möller, "Effects of Rating Scale Order on Subjective Quality Ratings of the Speech Signal Alone, the Background Noise Alone, and Overall Quality," International Telecommunication Union, Geneva, ITU-T Contribution COM 12 – C 113 – E, 2013.

[4] J. Seebode, R. Schleicher, I. Wechsung, and S. Möller, "Investigating the Affective Impression of Tactile Feedback on Mobile Devices," in Proc. *Proc. 27th Int. British Computer Soc. Human Computer Interaction Conference*, 1-10, 2013.

[5] S. Möller, Quality Engineering: Qualität kommunikation-stechnischer Systeme. Heidelberg, Germany: Springer, 2010.

[6] R. Schleicher, T. Westermann, B. Weiss, I. Wechsung, and S. Möller, "Research On Mobile HCI –Taken Out Of Context?," Research and Design Innovations for Mobile User Experience. IGI Global K. Rizvanoğlu, G. Cetin, 76–93, 2014.

[7] T. Polzehl, K. Schoenenberg, S. Möller, F. Metze, G. Mohammadi, and A. Vinciarelli, "On Speaker-Independent Personality Perception and Prediction from Speech," in Proc. *Proc. 13th Ann. Conf. of the Int. Speech Communication Assoc*. (Interspeech 2012), 2012.

[8] A. Kittur, E. H. Chi, and B. Suh, "Crowdsourcing user studies with Mechanical Turk", in *CHI '08*, Florence, Italy, 2008.

[9] D. Zhu, and B. Carterette, "An analysis of assessor behavior in crowdsourced preference judgments". in *SIGIR 2010 workshop on crowdsourcing for search evaluation* , 17-20, 2010