

Quality Control for Real-time Ubiquitous Crowdsourcing

Afra J. Mashhadi

Dept. of Computer Science
University College London
London WC1E 6BT, UK
a.jahanbakhshmashhadi@cs.ucl.ac.uk

Licia Capra

Dept. of Computer Science
University College London
London WC1E 6BT, UK
l.capra@cs.ucl.ac.uk

ABSTRACT

Crowdsourcing has become a successful paradigm in the past decade, as Web 2.0 users have taken a more active role in producing content as well as consuming it. Recently this paradigm has broadened to incorporate *ubiquitous* applications, in which the smart-phone users contribute information about their surrounding, thus providing a collective knowledge about the physical world. However the acceptance and openness of such applications has made it easy to contribute poor quality content. Various solutions have been proposed for the Web-based domain, to assist with monitoring and filtering poor quality content, but these methods fall short when applied to ubiquitous crowdsourcing, where the task of collecting information has to be performed continuously and in real-time, by an always changing crowd. In this paper we discuss the challenges for quality control in ubiquitous crowdsourcing and propose a novel technique that reasons on users mobility patterns and quality of their past contributions to estimate user's credibility.

Author Keywords: Crowdsourcing, Quality Assurance, Participation.

ACM Classification Keywords:
Social and Behavioural Sciences.

General Terms: Human Factors, Verification.

INTRODUCTION

Thanks to the widespread adoption of powerful and networked (i.e., Internet-enabled) handheld devices, consumers of digital content are now taking a more active role in producing content on the go. This trend has allowed a new category of applications to surface, in which data is collected by participants and is collectively used to offer services to citizens [3]. In this paper we focus on a new stream of research known as *ubiquitous crowdsourcing*, in which the contributed information is not limited to passively-generated sensor-readings from the device, but also includes proactively-generated user's

opinions and perspectives, that are processed to offer real-time services to participants.

For example, in an urban city such as London, where there exists a complex transport network with unavoidable disruptions, users engagement in travel updates can be very valuable. By introducing a ubiquitous crowdsourcing application for public transportation, participants could actively contribute *real-time* information related to their journey. These contributions could include, for instance, information about accidents, unplanned road closures, congestions, and other highly dynamic events that affect user's journeys, but that transport authorities do not have the capacity to process as promptly as required. If such information could be gathered in real-time, users could be given useful updates while their journeys execute, so to dynamically and effectively adapt their travel plans [6]. In such scenarios, the real-time information contributed by participants can be invaluable, and thus the more participants are engaged in providing this information, the better such applications can work [4]. However, the very same openness characteristic of such applications can threaten their success and impact the correctness of the results, as they allow *anyone* to contribute information. Indeed, a field trial study of ubiquitous crowdsourcing application has shown that users are concerned with the credibility of data provided by the other participants [21]. Therefore, quality control in crowdsourcing applications is an important issue which cannot be neglected. In the web domain, this challenge has been highlighted and effectively tackled by using various approaches such as aggregation and reporting. However, ubiquitous crowdsourcing exhibits unique properties that lead to different requirements for controlling the quality of contributions. These properties are as follows:

- *Real-time Events*: in ubiquitous crowdsourcing the task of collecting information is often tightly linked to events which are highly dynamic. Furthermore, the collected information needs to be analysed, and the results provided to users, in real-time. For instance, in the above scenario, participants can upload information about the status of the bus journey they are taking, and the collected information has to be processed in real-time to give an estimate of buses *real* arrival time to awaiting users. This requirement of processing contributions in real-time differs from

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

UbiCrowd'11, September 18, 2011, Beijing, China.

Copyright 2011 ACM 978-1-4503-0927-1/11/09...\$10.00.

what is observed in web-based crowdsourcing, where the applications can achieve quality assurance by relying on users (or authorised users) to flag and report poor quality content with some time delay.

- *Dynamic Crowds*: as opposed to web-based crowdsourcing, in ubiquitous crowdsourcing the crowd set (i.e., participants) keeps changing all the time. Let us refer back to the public transport scenario, where the crowd that can contribute travel information is formed by public transport users, undertaking their daily journeys. Such crowd varies throughout the day (e.g., the travellers who can report disruptions on a bus route), and it may not always reach the critical mass required for such applications to function (e.g., night-bus riders may be just a handful). Sparsity of contributions by small crowds is a well-known challenge in web-based crowdsourcing systems too, with severe impact on content quality [4]. However the web-based techniques cannot directly be applied in our domain, because of the the real-time and highly dynamic nature of the applications at hand.

To address the above challenges, we propose a technique which estimates the quality of contributions based on the contributor’s mobility, as well as their trustworthiness score based on their past contributions. We combine these two sets of information to estimate a *credibility weight* for each contributor, allowing us to compute the results based on a weighted average of all the uploaded contributions. We continue this paper with an overview of the current state-of-the-art in quality control within the crowdsourcing paradigm; we then proceed to our novel quality control technique and lay out our evaluation plan.

RELATED WORK

Web-based Crowdsourcing

Web-based crowdsourcing systems have gained popularity in recent years; in this domain, content quality is assured by means of: manual edition, aggregation and user’s reputation. In [9], the authors propose a decision matrix for selecting one of the quality assurance techniques based on tasks complexity. In practice (i.e., Slashdot, Reddit, and Digg) content quality is often achieved by relying on users and their social networks to report and filter inappropriate content from the vast volume of online stories. Other systems such as Wikipedia and IMDB restrict the quality control to only a pre-defined set of editors (i.e., *experts*), who can delete and correct poor quality articles (e.g., spam, misinformation, abusive language). In these cases, the editors are often chosen based on their profile and historical data on their previous contributions, such as the number of edited articles.

In addition to examples of centralised web-based crowdsourcing applications, semi-distributed crowdsourcing systems have also been extensively studied by the research community. An example of this research stream

is Voluntary Geographical Information (VGI), in which the data is contributed by participants from the physical world and used to maintain and enhance the overall body of environmental knowledge (e.g., OpenStreetMaps and WikiMapia). In [14], the authors investigate the quality of voluntarily tagged Points Of Interests (POIs), and propose an algorithm which aggregates contributed data to retain the POIs that are only consistently repeated. In [5], Flanagan et al. discuss the issue of credibility of VGI by arguing that credibility is a measure of trustworthiness more than of expertise (i.e., data accuracy). That is, credibility is less about data accuracy and more about which information, or perspective, people believe in. In [1], a need for a trust model that takes into account subjectivity of geographic information and user’s perspective has been discussed.

Participatory Sensing

A rich body of participatory sensing systems has been proposed in the literature, examples of which include: noise level monitoring [11, 15] and traffic monitoring [13, 6]. A common problem faced by most of these applications is known as *data pollution*. That is, malicious users can compromise the quality of the results by uploading forged data or interfering with the data collection process (e.g., creating deliberate noise to impact the readings by the noise sensing application).

In [17], a novel architecture for participatory sensing, which enables data consumers to assign trust scores to the data they access, is proposed. The authors investigate the challenge of verifying data integrity in terms of verifying participant’s context based on authenticating their location. In [19], the problem of fabricating the recorded data by malicious users is investigated. The authors assume the existence of a threat model in which malicious users can corrupt the sensors data after the reading. In order to overcome these types of threats, they propose a solution based on a trusted platform module (TPM), which confirms the integrity of sensing devices. Similarly, in [7] a TPM based model is proposed which enables the service providers to trust the content generated on mobile devices. In [10], the co-located infrastructure is used for issuing a timestamped location certificate to the participant’s device, which can later be used as the proof of the task’s location.

Common to all the above approaches is their focus on *data integrity*; that is, they verify and confirm that the contributed data is indeed from the participant device and was collected at the claimed location. However, unlike participatory sensing, where the data comes from automatic readings from devices, in ubiquitous crowdsourcing contributed data is more subjective and includes users’ opinions. Therefore, there is a need for assessing the *quality* (i.e., correctness) of contributions in addition to data integrity.

The only other work that tackles a similar problem to ours is [8]. In [8] Huang et al. consider scenar-

ios in which some of the participants may deliberately affect the result of the readings from the sensor, and thus introduce poor quality information. They propose a reputation-based trust model which assigns a trust value to all participants and assesses the overall results of readings by taking into account participants credibility. The trust model is then evaluated using a participatory sensing application designed for assessing the noise level in office spaces, in which malicious users produce fake data by positioning their devices in various ways that can hinder the readings. Their result shows a clear improvement over the previous non-trust based approach, where quality was assessed using aggregation.

We propose to extend this reputation-based line of work by considering scenarios in which users move and produce data within urban spaces such as the described crowdsourcing public transport application. Recently, various works have started studying the correlation between users *mobility* and their role as contributors in crowdsourcing applications. In particular, research has shown that users are more likely to perform a crowdsourcing task when close to home or a familiar place [2]. Similarly, [16] has shown that crowdsourcing participants selected based on their mobility pattern can offer much more valuable contributions. In this work, we apply the same reasoning and propose a quality control model for real-time ubiquitous crowdsourced information, based on participants mobility pattern as well as their historical reputation score.

PROPOSED APPROACH

We propose to reason upon user’s mobility patterns in order to estimate the quality of their contributed information. Indeed, research has shown that a great degree of information can be learned by monitoring human mobility [18]. For instance, studies demonstrate that urban travellers exhibit a high level of regularity in their daily journey patterns and this property has been extensively exploited in mobile content dissemination networks [12]. We propose to record a user’s mobility pattern in terms of the locations (POIs) he travels through (e.g., the bus stops the user covers). In addition, for each recorded location, we log the *logical* time segment of the recording (we are not interested in precise timestamps, but rather application-meaningful times, such as, for the transport domain, *early morning*, *morning*, *afternoon*, *evening* and *night*). We then define, for each user, a set of spatio-temporal tuples, $T(loc_{POI_x}, t_i)$, where loc_{POI_x} corresponds to a specific POI that the contribution was recorded from at logical time t_i . For instance, in the case of crowdsourcing travel information, the application can allow users to “check in” (à la FourSquare) to public services and locations, and the tuples can be: T_1 (“BUS No.1-Stop B”, Monday Early Morning), T_2 (“BUS No.1-Stop D”, Monday Evening), T_3 (“Oxford Street”, Tuesday Afternoon).

As this mobility data is recorded every day, a *regularity* value, based on the frequency of the repeated locations for each logical time, can be calculated for the dura-

tion of the observation period. For the transport domain, data is separately aggregated for weekdays and weekends, as users tend to travel differently Monday-to-Friday, and Saturday/Sunday. Figure 1 exemplifies the frequency with which each tuple is generated during a given observation period, with the x axis representing the different time segments of a day, and the y axis corresponding to the frequency with which the associated tuple was observed. In this example, we can see that the user travels from bus stop B on every single morning of the observation period (high regularity), while T_j (“BUS No.1 – Stop D”, *Afternoon*) is produced less regularly. This pattern could reflect, for instance, user’s adherence to a strict arrival time at work every morning, and a less strict leave time from the office.

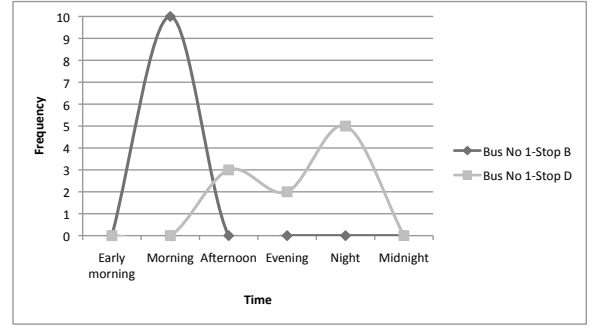


Figure 1. Frequency distributions of POI for a given user

Based on this information, we define a regularity function $Reg(T_j)$ whose value is determined based on *implicit* (location) readings from the user’s device. Although at bootstrapping time there is no available record over which to compute regularity patterns, after just a few days of usage of the application the logged records are sufficient to compute initial regularities. It is worth noting that, as illustrated in the above example, the regularity value is per user per tuple, that is, the familiarity of a user with a specific location at a specific logical time. Based on the regularity function, we can then define user’s expertise for any tuple as *local*, *familiar*, or *stranger*, and use application-dependent thresholding to delimit each category.

In addition to the regularity function, for each user we estimate and maintain a reputation score which corresponds to their trustworthiness based on their past interactions with the ubiquitous crowdsourcing application. In so doing, whenever participants upload a piece of information, the *usefulness* of their contribution is ranked and the score is fed into a reputation-based trust function, $Trust(u_i)$. In order to estimate such usefulness, the web-based systems have relied on *explicit* rating by users. However, in ubiquitous crowdsourcing, where the reputation score needs to be calculated in real-time, a more dynamic way of estimating content’s usefulness is needed. For instance, we could compare the uploaded contribution with those provided by local experts (i.e., the highly regular users at that

POI). Furthermore, in situations where there does not exist a benchmark (i.e., due to sparsity of contributions and lack of contributions by local experts), we can still assess the usefulness of a contribution by proactively asking colocated users to explicitly rate the content in *real-time* through game strategies [20]. Based on these two sets of information, we can now compute a *credibility weight* for user u_i as follows:

$$\text{credibility weight}(T_j) = \alpha \cdot \text{Reg}(T_j) + (1 - \alpha) \cdot \text{Trust}(u_i),$$

where α can be dynamically adjusted to give precedence to either user's regularity or trustworthiness, and is to be set by the application. For example, when contributions are sparse, the public transport crowdsourcing application might give a higher weight to the regularity function than to the user's trustworthiness. Once computed, the *credibility weight* is then used to assess the quality of contributions, and the result of the crowdsourcing task can be provided to the application in the form of weighted average on all uploaded contributions.

EVALUATION PLAN

To evaluate the proposed approach, we are building an Android-based public transport crowdsourcing application that enables users to check in to their current journey, rate it, as well as provide comments about it (e.g., status of their bus journey, dynamically occurring events such as traffic jams and accidents). The application also includes a game component, in which users are challenged to compete against each other to become local "experts" of their bus routes and for a given time segment (à la FourSqaure, but with the time component added in). We plan to deploy this application in London, UK, starting in autumn 2011, with a pool of 100 users. To avoid unmanageable data sparsity, users will be selected so that their mobility mainly covers a restricted area in London (e.g., university staff and students). The quality of their contributions, as dynamically computed by our proposed approach, will then be qualitatively assessed.

Acknowledgment This research has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under the Grant Agreement n. 234239.

REFERENCES

- ADLER, B. T., AND DE ALFARO, L. A content-driven reputation system for the Wikipedia. In *Proc. of WWW* (2007), pp. 261–270.
- ALT, F., SHIRAZI, A. S., SCHMIDT, A., KRAMER, U., AND NAWAZ, Z. Location-based crowdsourcing: extending crowdsourcing to the real world. In *Proc. of NordiCHI* (2010), pp. 13–22.
- BURKE, J., ESTRIN, D., HANSEN, M., PARKER, A., RAMANATHAN, N., REDDY, S., AND SRIVASTAVA, M. Participatory sensing. In *Proc. of WSW* (2006), pp. 1–5.
- DOAN, A., RAMAKRISHNAN, R., AND HALEVY, A. Y. Crowdsourcing systems on the world-wide web. *Communications of the ACM* 54, 4 (2011), 86–96.
- FLANAGIN, A., AND METZGER, M. The credibility of volunteered geographic information. *GeoJournal* 72, 3 (2008), 137–148.
- GANTI, R., PHAM, N., AHMADI, H., NANGIA, S., AND ABDELZAHER, T. Greengps: A participatory sensing fuel-efficient maps application. In *Proc. of MobiSys* (2010), pp. 151–164.
- GILBERT, P., COX, L., JUNG, J., AND WETHERALL, D. Toward trustworthy mobile sensing. In *Proc. of HotMobile* (2010), pp. 31–36.
- HUANG, K., KANHERE, S., AND HU, W. Are you contributing trustworthy data?: the case for a reputation system in participatory sensing. In *Proc. of MSWIM* (2010), pp. 14–22.
- KERN, R., THIES, H., BAUER, C., AND SATZGER, G. Quality Assurance for Human-Based Electronic Services: A Decision Matrix for Choosing the Right Approach. In *Proc. of ICWE* (2010), pp. 421–424.
- LENDERS, V., KOUKOUMIDIS, E., ZHANG, P., AND MARTONOSI, M. Location-based trust for mobile user-generated content: applications, challenges and implementations. In *Proc. of HotMobile* (2008), pp. 60–64.
- MAISONNEUVE, N., STEVENS, M., NIESSEN, M., AND STEELS, L. Noisetube: Measuring and mapping noise pollution with mobile phones. *Information Technologies in Environmental Engineering* (2009), 215–228.
- MASHHADI, A. J., MOKHTAR, S. B., AND CAPRA, L. Habit: Leveraging Human Mobility and Social Network for Efficient Content Dissemination in DTNs. In *Proc. of WOWMOM* (2009), pp. 1–6.
- MOHAN, P., PADMANABHAN, V., AND RAMJEE, R. Nericell: rich monitoring of road and traffic conditions using mobile smartphones. In *Proc. of SenSys* (2008), pp. 323–336.
- MUMMIDI, L., AND KRUMM, J. Discovering points of interest from users map annotations. *GeoJournal* 72, 3 (2008), 215–227.
- RANA, R., CHOU, C., KANHERE, S., BULUSU, N., AND HU, W. Ear-phone: an end-to-end participatory urban noise mapping system. In *Proc. of IPSN* (2010), pp. 105–116.
- REDDY, S., ESTRIN, D., AND SRIVASTAVA, M. Recruitment framework for participatory sensing data collections. *Pervasive Computing* (2010), 138–155.
- REDDY, S., SAMANTA, V., BURKE, J., ESTRIN, D., HANSEN, M., AND SRIVASTAVA, M. Mobisensemobile network services for coordinated participatory sensing. In *Proc. of ISADS* (2009), pp. 1–6.
- RHEE, I., SHIN, M., HONG, S., LEE, K., AND CHONG, S. On the Levy-Walk Nature of Human Mobility. In *Proc. of the ICCCN* (2008), pp. 924–932.
- SAROIU, S., AND WOLMAN, A. "i am a sensor, and i approve this message". In *Proc. of HotMobile* (2010), pp. 37–42.
- VON AHN, L., AND DABBISH, L. Designing games with a purpose. *Communication of ACM* 51, 8 (2008), 5867.
- ZIMMERMAN, J., TOMASIC, A., GARROD, C., YOO, D., HIRUNCHAROENVATE, C., AZIZ, R., THIRUVENGADAM, N. R., HUANG, Y., AND STEINFELD, A. Field trial of Tiramisu: crowd-sourcing bus arrival times to spur co-design. In *Proc. of CHI* (2011), pp. 77–86.