An Intelligent Mobile Crowdsourcing Information Notification System for Developing Countries

Arun Singh[1], YueXin (Sophia) Li[2], Yu Sun[1], Qingquan Sun[3]

[1]Department of Computer Science
California State Polytechnic University, Pomona
Pomona, CA 91768
{alsingh,yusun}@cpp.edu

[2]Branksome Hall
Toronto, Ontario, M4W 1N4
sli2@branksome.on.ca

[3]School of Computer Science and Engineering
California State University San Bernardino
San Bernardino, CA, 92407
qsun@csusb.edu

Abstract. Crowdsourcing is an important computing technique that taps into the collective intelligence of the public at large to complete business-related tasks and solve many real-time problems. It is changing the way we work, hire, research, make and market. Many developing nations are trying to take advantage of crowdsourcing for information notification to make cost effective system, like real-time transit system, disaster notification system and other services which are available to the masses. However, many of them are still not able to completely benefit from it compared to developed nations. In this paper, we have identified a series of limitations of using crowdsourcing for information gathering and providing real-time notification in developing countries due to their unstable electronic communication infrastructure, their lack of contribution, lack of crowdsource (participating people), less exposure to English language, and unawareness of crowdsourcing. We proposed, and demonstrated, a solution to overcome these limitations by developing a prototype which uses SMS as a reliable method for providing real-time notification and information gathering. Our prototype uses prediction algorithms to fill the gaps in real-time notification. It also uses the prediction of a user's behavior to provide a better reward and motivational platform, as well as good usability.

Keywords: Mobile Crowdsourcing, Machine Learning, Information Prediction, Reliable Communication

1 Introduction

Notification systems have been a crucial part of any system in recent decades. A notification system is a set of protocols and procedures that can involve both human and computer or mobile components. The purpose of these systems is to generate and

send timely messages to a person or group of people. Simple notification systems use a single means of communication, such as an email or text message [1]. More complex information notification systems involve the processing of bulk information and providing some meaningful result from processed information. For example, a transit system gets information from different sources to notify user about the real-time schedules. Data reins as the great equalizer and democratizer in an era of the global economies. Vast amounts of data in developed societies are being handled by the leading tech companies such as Google [8] and Facebook [9]. Crowdsourcing can be seen as a catalyst for global innovation and is something businesses should keep an eye on going forward [2][3]. Figure 1 shows the main participants and abstract dataflow for crowdsourcing. The main participants include: a contributor, which is the crowd, a consumer which could be anyone from a single business to the whole world, an administrator/business that manages all of the activity and an operator who does monitoring and management. One of the great examples is Wikipedia, which has opened a number of possibilities for crowd sourcing learning resources [4].

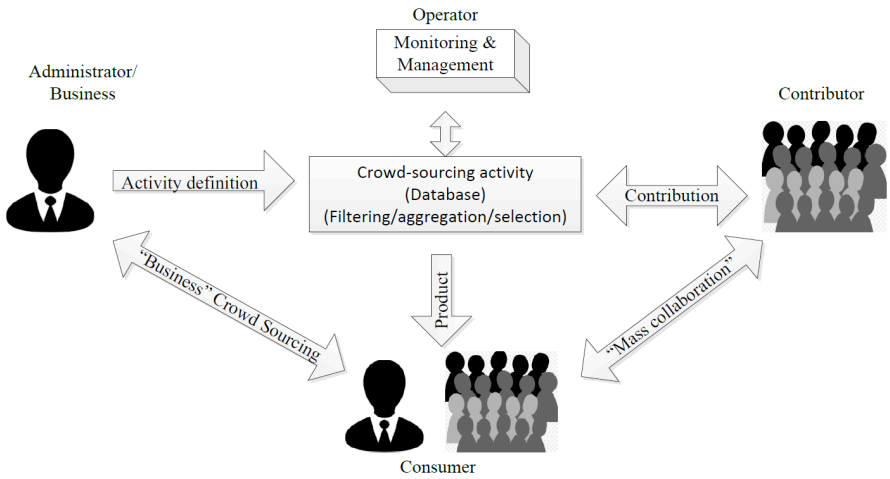

Figure 1. Crowdsourcing main participant and abstract dataflow (adapted from [25])

There are a number of successful applications where crowdsourcing is used for information notification; one of popular apps is Waze [5]. Its mapping and traffic information are built from 70,000 volunteer map editors and more than 15 million active users. These users contribute their live driving data by default so others can benefit by seeing how fast they are going. Given the increasing number of people who carry smartphones, it seems likely other services could be built leveraging willing users who contribute a small portion of the data from their travels. That data can be mapped and analyzed for the common good [6]. While the west is doing its best to tackle the immense flow of data, the case is very different in developing countries where entire societies are cut off from this transfer of data. Action needs to be taken for the global disparity related to the data that fuels crowdsourcing.

Open Problems: Numerous technical and non-technical barriers exist in developing countries that limit the effective usage of crowdsourcing platforms.

One such barrier is the lack Internet availability, which is typically needed for mobile crowdsourcing. Public Internet services may be available but that not conducive to real-time information acquisition and dissemination. Other barriers include gaps between real time notifications, understanding of user behavior due to technology limitations, the lack of skill in understand complex English instructions, minimum awareness of crowdsourcing, and motivations to use crowdsourcing platform.

Proposed Solution ➔ An intelligent mobile-based information notification system applicable for developing countries. In order to address the problem, our proposed solutions are using SMS as a reliable method for providing real-time notification and information gathering, using prediction algorithms to fill the gaps in real-time notification, using the prediction of user's behavior to provide better a reward and motivational platform.

The solution presented in this paper focuses on stable communication protocol between mobile devices and server, prediction algorithm to encourage crowd to use crowdsourcing applications in developing countries. The main contribution of this paper is to provide the practical crowdsourcing solution for developing countries with a completely intelligent mobile-based crowdsourcing platform.

The remainder of this paper is organized as follows: in Section 2, we provide a motivating example for this paper and in section 3 we list the challenges that are faced to institute a crowdsourcing application in developing countries. In Section 4, we present our solution to address these challenges. We analyze the related work in Section 5 and conclude this paper as well as provide some insight into future work and scope of crowdsourcing in developing countries in section 6.

2 Motivating Example – Public Bus Notification System

Buses are indispensible part of the public transport system. Bus transport services help in reducing private car usage and fuel consumption. Public transport bus notification services are generally based on predictable operation of transit buses along a route arriving at the specified bus stops according to a published public transport timetable. Many public bus services are run to a precise timetable giving specific departure and arrival times. These are often difficult to maintain in the event of traffic congestion, breakdowns, on/off bus incidents, road blockages or bad weather. Predictable effects such as morning and evening rush hour traffic are often accounted for in timetables using past experience of the effects. This prevents the drafting of a clock face time schedule where bus arrival is predictable at any time throughout the day. Providing precise arrival times of buses will advance user experience and attract more users to use buses. Nowadays, most of the public bus transport agencies provide their bus schedules on web but the bus schedule information are not real-time.

Increasingly, technology is being used to improve the information provided to bus users [10]. Vehicle tracking technologies are used to assist with scheduling and to achieve real time integration with passenger information systems. These information systems display service information at stops, inside buses, and to waiting passengers through personal mobile devices or text messaging, but this type of systems are

costly. There are a few crowdsourcing applications such as Moovit [11] and Tiramisu [12], which collect bus passengers' real-time movements and share them with other users to track arriving bus. However, the existing crowdsourced applications does not work very well in developing countries because of various challenges listed in next section.

3 Challenges

The United Nations is using digital media and mobile phone technology to enable people from across the world to take part in setting the next generation of anti-poverty goals. Crowdsourcing applications are changing the way we work, hire, research, make and market. Many developing nations are trying to take advantage of crowdsourcing for information notification to make cost effective systems like real-time transit system, disaster notification system and other services which are to be available to the masses. Despite the fact that crowdsourcing is a well-known technique for information collection and providing real-time notifications, developing countries are still not able to completely benefit from it compared to developed nations. There are many limitations to using crowdsourcing for information gathering and providing real-time notification in developing countries.

3.1 The Lack of a Reliable Mobile Internet Infrastructure

Developed countries have decent Internet infrastructure, which can be used in combination with GPS technology to provide real-time notifications that are commonly used by many crowdsourcing applications [13]. For instance, in bus tracking systems, mobile apps can provide riders' locations when they are in transit by using their mobile data and GPS. Crowdsourcing systems can utilize this information to provide bus arrival times. On the other hand, however, developing countries do not always have the reliable Internet infrastructure which restricts the usage of the traditional crowdsourcing solutions.

3.2 The Lack of Contribution Motivation

Users participation drives the success for any online crowdsourcing applications. Participation becomes more difficult in developing countries where people have less knowledge of crowdsourcing. It is a big challenge to understand what motivates people to participate in online crowdsourcing platform. Even though it is easy to introduce a platform but how to attract a large number of crowds especially in this competitive market requires close study. For bus tracking system we can undertake people who use bus tracking systems to find the bus arrival time. However, in order to provide real-time bus arrival information, we need riders inside bus to use app to provide real-time notification and updates.

3.3 The Lack of Active Crowd Source

When working with real-time mobile crowdsourced based applications, it is highly probable to see gaps in real-time notification because of lack of crowd source (participating people). It is a challenge to retain existing users (crowd) if system does not provide real-time information. To give an example in bus tracking system, there could be possible time or day when there are none or only a few people travelling in bus and proving tracking information and system would not sufficient data to provide bus arrival times.

4 Solution

We have implemented smart mobile crowdsourcing platform for developing countries which includes reliable communication infrastructure, reward system for motivating users and intelligent system which can predict information and interested rewards.

4.1 Reliable Communication Infrastructure

We know with Internet we can easily develop any crowdsourcing application because of the ability to communicate information (like device location and small data) from crowd (people) to system and vice-versa. It would take decades to replicate a developed country's equivalent infrastructure in developing countries. Based on our reviews and our experience of using Short Message Service (SMS), we know SMS is still widely used and remains one of the most reliable channels of communication so far [14]. We have designed and implemented a SMS protocol to send data over mobile networks. SMS can work in combination with other methods of communication in areas of developing countries that have fractional infrastructure for Internet. Figure 2 is the high-level flow of a crowd-sourced bus tracking system where users with mobile apps are requesting and providing tracking information.

In our implementation we identify each rider by their phone number used in mobile phone. To get the bus arrival information, registered rider uses our mobile app to select the bus number, mobile app internally checks for internet connectivity and if internet is not available then it forms a SMS which is combination of request type, user location and bus number and send it to server. Server process the request based on request type and send the response back in the form of SMS. Bus arrival response SMS include response type, bus number, bus arrival time, distance of user from nearest bus stop and some advertisement text. When user's mobile phone receives the SMS, mobile application recognizes and reads it then displays message in user interface based on response type. If rider's mobile device has internet connection then mobile application can directly make web API call to communicate with server instead of communicating via SMS.

4.2 Reward System for Contribution Motivation

Users' participation in mobile crowdsourcing platforms is vital as the success of platforms largely depends on the presence of their members. Researchers found that the most frequently mentioned motives of users participating in crowdsourcing are: money, altruism, fun, reputation/attention, and learning. Many scholars of crowdsourcing suggest that there are both intrinsic and extrinsic motivations that cause people to contribute to crowdsourced tasks and that these factors influence different types of contributors [7]. We implemented a reward-based system in the crowdsourcing application to motivate people in developing countries. This reward should be mapped to financial incentives like bonus, coupons, deals with lower cost, free product, free service etc. The reward value should be divided based on type of activity. For example, reward value should be in the descending order of: user referral, users helping to provide notification, users using applications for some need etc.

To make reward system more interesting, we have managed to get commercial companies to participate in offering the rewards. For example, when users request for bus arrival time we would process the request, but along with bus arrival response we would add advertisement data to display on user's screen.

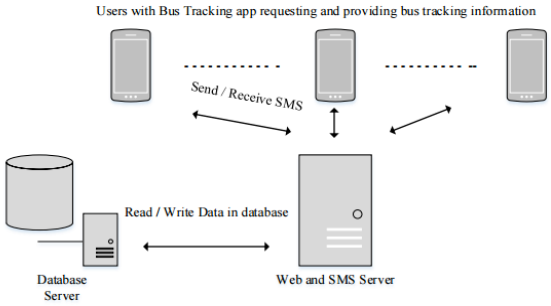

Figure 2. Bus Tracking System in Developing Countries based on SMS Protocol

4.3 Prediction of the Information and Interested Rewards

The lack of crowdsource or participating people is one of the biggest factors for failure of a crowdsourcing application that needs real-time information from participants. We cannot completely rely on the crowd to provide real-time information in to minimize any response gaps. For example, if a user is requesting a bus arrival time and the system does not have any real-time update, then the system would not be able to provide a response. This limitation would discourage the crowd from using the application. This would again decrease real-time notifications. Machine learning algorithms work best to predict information based on historical data. Here we implemented information prediction based on the historical data using Naïve Bayes [26] algorithm. For an example user requesting Bus No "A1" arrival time, if system does not find real-time information of Bus No "A1" then machine learning algorithm would predict the Bus No "A1" arrival based on historical Bus No

"A1" historical arrival data. It is really important to pick the correct machine learning algorithm that fits this particular need. The accuracy of algorithm should be known before it is put in live application. Figure 3 shows the basic workflow of the request and response to understand the bus tracking example. In the figure we have request and response on right side which is sent to the tracking request web API which has logic inside to handle the response.

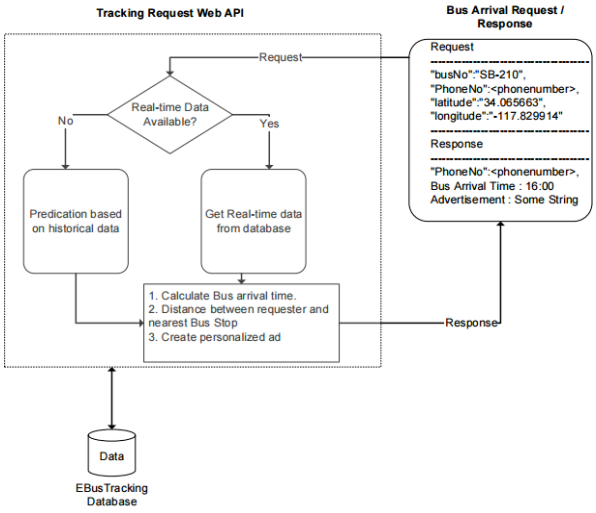

Figure 3. Bus Tracking Request Work Flow

The system also predict user's interested reward category for the reward based system. For example, in the bus tracking application, we can categorize riders based on their ride actions such as bus stops, riding days, riding times, information from user profiles captured during registration. If rider is requesting bus arrival times for a bus stop which is near a college, and request is within college working hours, and the rider's age is 18, then it is highly possible the rider is student. Figure 4 shows the machine learning basic workflow to predict the user's category for bus tracking system.

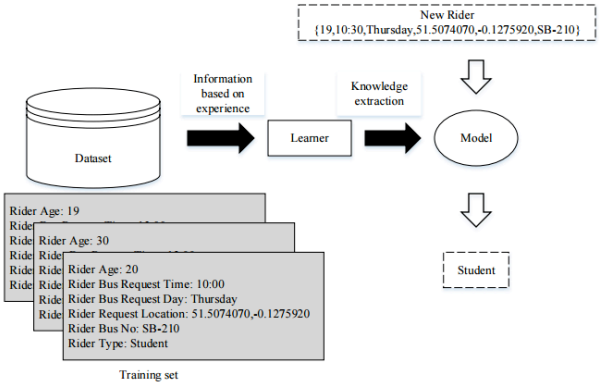

Figure 4. Predict user's category for interested reward

To find the optimal training approach for the user's category prediction, a group of sample data containing 220 instances is processed using Weka. In total, 35 different classifiers are tested under the same qualification of 10-fold cross-validation, and the percentage of correctly classified instances is recorded as the accuracy of the classifier. The results are summarized in Figure 5.

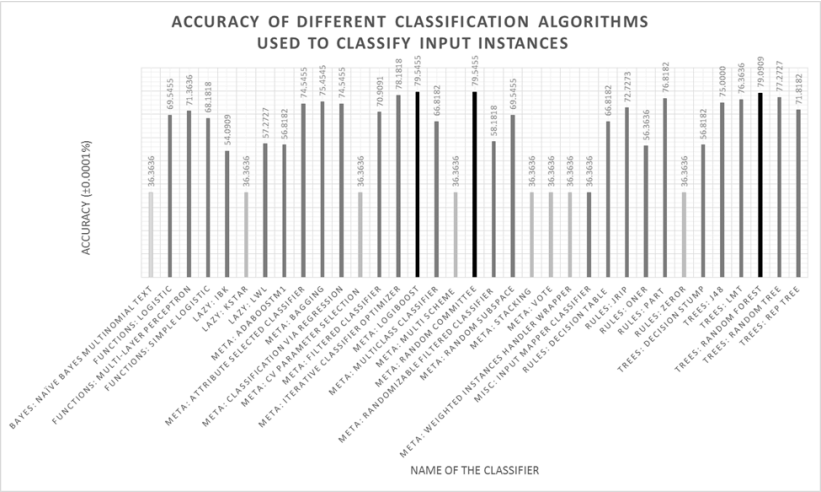

Figure 5. The performance of different machine learning algorithms

The black bars represent the classifiers that yield relatively high accuracies, while the light gray bars represent the classifiers that yield relatively low accuracies. From the diagram, LogiBoost, Random Committee and Random Forest correctly classify the largest amount of instances with an accuracy of 80%, while nine other classifiers, such as Naïve Bayes, KStar, Multi Scheme and several Meta Classifiers, have the lowest accuracy of 36.3636%. Noticeably, the average accuracies of the Functions-based classifiers and the Trees-based classifiers are relatively higher than other types. The main reason is that Random Forest is a highly suitable and accurate algorithm to predict the user's profile as it has an internal system to get an unbiased estimate of test error and a useful tool called proximities [19]. In random forest, as many classification trees as needed based on the number of data are grown which each gives a classification. After each tree "votes" for a class, the forest chooses the classification having the most votes. During this process, out-of-bag error for each point is recorded and averaged over the forest. Meanwhile, proximities are computed for each pair of cases and normalized by dividing by the number of trees. Due to these features, random forest is able to compute prototypes of high accuracy, 79.0909% in this case, to model the relation between the variables and the classification efficiently.

On the other hand, Naïve Bayes Multinomial Text is not the classifier that best fits this set of data due to the severe assumptions it makes that would adversely affect the quality of its results. As designed, it can model the distribution of words in a document as a multinomial with high efficiency [20]. However, its inherent systemic errors lead to the low accuracy of the classification model it gives. One systemic error is the production of biased decision boundary weights due to skewed data. As there are more training examples for the class of student than other classes, the data is

skewed to one direction and therefore the accuracy of Naïve Bayes decreases. Moreover, the production of different magnitude classification weights based on the independence assumption is another problem that impedes the accurate modeling.

However, there are classification models simply inapplicable to this particular training set such as SMO, Additive Regression, and Simple Linear Regression. The primary underlying reason is the mismatch between how the classifier works and the nature of the training data.

5 Related Work

There are quite a few initiatives by organizations and researchers to enable crowdsourcing as a platform to solve developing countries' needs.

NextDrop [15], a social enterprise which is streamlining urban water collection in India. The enterprise collects and shares water delivery information with city residents and water utilities. In this way, efficiency and transparency are improved upon. NextDrop started as a pilot in Hubli-Dharwad. These twin cities in the Indian state of Karnataka, which have seen rapid urban growth in recent years. There have been decent growth like shopping malls and American fast food chains, but access to water is still a challenge. It is therefore unsurprising that NextDrop's Smart Water Supply Message Service has thousands of household subscribers paying a monthly nominal fee to receive advance water alerts. NextDrop's messaging system uses SMS to notify subscribers about when they will get water, when there is an interruption, when pipe damage is likely to affect them, and when someone in the community has water updates to share. However, they do not use machine learning to predict the information.

Medic Mobile [16] uses mobile technology to create connected, coordinated health systems. Medic Mobile is specifically designed for health worker and medical system to offer a free, scalable software toolkit that combines messaging, data collection, and analytics. It is intended for hard-to-reach areas and supports may languages. It works with or without internet connectivity, locally or in the cloud. The tools run on basic phones, smartphones, tablets, and computers. In six months, their pilot program in Malawi had saved hospital staff 1,200 hours of follow-up time and over US$3,000 in motorbike fuel. More than 100 patients started tuberculosis treatment after their symptoms were noticed by community health workers and reported by text message. Medic Mobile uses crowd-sources translation, categorization and geo-tagging which helps to create quick reports for first responders in emergency situations [17]. Compared with our solution, Medic Mobile did not have intelligent reward based system and real-time information prediction.

6 Conclusion and Future Work

In this paper, we have designed and developed an intelligent method to implement crowdsourcing applications in developing countries. A series of solutions have been proposed to address the challenges we identified. Firstly, we chose the use of SMS as a reliable communication channel that is not dependent on the Internet for mobile

crowdsourcing communication. Secondly, we proposed a motivating reward system which predicts the rewards that interest our crowd. Thirdly, we performed predications based on historical data by using relevant machine learning algorithms to fill the gaps in real-time information. The goal of this study helps researchers, nonprofit organizations and commercial companies to plan better crowdsourcing applications in developing countries.

The research presented in this thesis seems to have raised more questions that it has answered. There are several lines of research arising from this work which should be pursued. Firstly, we know the SMS is a reliable communication channel but there is need for dynamic thresholding to minimize the SMS traffic when it is used for crowdsourcing. The second line of research is designing one crowdsourcing application to solve multiple needs in developing countries. This would provide us more information for better information prediction (e.g. multiple user actions would provide better prediction to classify users).

7 References

1. What is a notification system, http://www.wisegeek.com /what-is-a-notification-system.htm
2. Brabham, D. (2008). Crowdsourcing as a model for problem solving: An introduction and cases. Convergence, 14(1), 75-90.
3. Why Crowdsourcing is the Next Cloud Computing, http://blog.innocentive.com/2013/10/14/why-crowdsourcing-is-the-next-cloud-computing.
4. Wikipedia, https://en.wikipedia.org
5. Waze, https://www.waze.com
6. After Waze, What Else Can Mobile Crowdsourcing Do - Liz Gannes, http://allthingsd.com/20130719/after-waze-what-else-can-mobile-crowdsourcing-do
7. M. Hossain, "Users' motivation to participate in online crowdsourcing platforms", Innovation Management and Technology Research (ICIMTR), 2012 International Conference on, Malacca, 2012, pp. 310-315.
8. Google, https://www.google.com
9. Facebook, https://www.facebook.com
10. Camacho, T. , Foth, M. , & Rakotonirainy, A. (2013). Pervasive technology and public transport: Opportunities beyond telematics. IEEE Pervasive Computing, 12(1), 18-25
11. moovit, http://moovitapp.com
12. Tiramisu, http://www.tiramisutransit.com
13. J. Howell and J. Frolik, "An Internet-based, inverse-GPS system for monitoring and tracking mobile aquatic sensors," Sensors, 2002. Proceedings of IEEE, 2002, pp. 1734-1739 vol.2.doi: 10.1109/ICSENS.2002.1037386
14. Why are people still using SMS in 2015, http://thenextweb.com/future-of-communications/2015/02/16/people-still-using-sms-2015/#gref?
15. NextDrop, https://nextdrop.co
16. Medic Mobile, http://medicmobile.org
17. Medic Mobile, http://skoll.org/organization/medic-mobile
18. Main Participants and abstract data flow, https://www.horizon.ac.uk/wp-content/uploads/2015/01/Crowd-sourcing-images-700x539.jpg
19. Additive Model, November 2, 2009. doi:10.1007/0-387-33960-4_5.
20. Breiman, Leo, and Adele Cutler. "Random Forests Leo Breiman and Adele Cutler." http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm#workings.