

txteagle: Mobile Crowdsourcing

Nathan Eagle

MIT Media Laboratory, E-15 383, Cambridge, MA 02139, USA
nathan@mit.edu

and

The Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM 87501, USA
nathan@santafe.edu

Abstract. We present txteagle, a system that enables people to earn small amounts of money by completing simple tasks on their mobile phone for corporations who pay them in either airtime or MPESA (mobile money). The system is currently being launched in Kenya and Rwanda in collaboration with the mobile phone service providers Safaricom and MTN Rwanda. Tasks include translation, transcription, and surveys. User studies in Nairobi involving high school students, taxi drivers, and local security guards have been completed and the service has recently launched in Kenya nationwide.

Keywords: crowdsourcing, mobile phones, appropriate technology, reputation systems.

1 Introduction

The impact of mobile phones throughout the world has been widely documented. These devices now function as fundamental tools instrumental to billions of economic livelihoods. The transformative impact of what is now the fastest technology adoption in human history has had the most dramatic ramifications within some of our most underserved societies. Indeed, while mobile phones were originally designed for Western business executives, the vast majority of mobile phone subscribers today live in the developing world. Africa, with East Africa at the forefront, is currently the fastest growing mobile phone market in the world.

As the price of unlocked GSM phones (complete with sim card and airtime) falls below US\$20 in many of these markets, the mobile phone has become many people's single technology purchase. As such, the expectations for this technology far exceed simply two-way communication. In East Africa, where over 90% of the population live in an area with GSM reception, mobile phones are also expected to serve as flashlights, as music players, and now, even as digital wallets.

The reduction in the price of handsets has allowed a dramatic increase in the number of individuals relying on this technology to do their jobs. Day laborers repairing roads throughout Nairobi no longer need to congregate in the morning in central areas throughout the city waiting for prospective employers to collect them for the day's work. Instead, in many of these regions, daily labor is now organized via text message.

However, with unemployment levels rising to almost fifty percent in countries such as Kenya, there are still hundreds of millions of mobile phone subscribers who are unable to find consistent work. Given high rates of unemployment and marginal income sources, many of the more than 2 billion mobile phone subscribers currently living in the developing world would greatly benefit from even an extra dollar per day. Amazon's Mechanical Turk¹ has successfully introduced "human intelligence tasks" that can be completed by individuals with a personal computer connected to the internet for small amounts of money. We apply the same principle to "txteagle", a mobile phone-based system that untethers these tasks from the PC and offers them to the world's billions of mobile phone users - providing an additional source of supplementary income to rural and low income populations.

2 Related Work

SMS Bloodbank. The idea for txteagle originated in the Kilifi local district hospital on the coast of Kenya, where the author worked for the two years between 2006 and 2007. The village of Kilifi is located on one of Kenya's major thoroughfares between the cities of Mombasa and Malindi – a road that hosts a significant number of serious traffic accidents. The Kilifi local district hospital provides emergency support to the victims of these vehicular accidents, which causes the relatively small hospital's blood supplies to be regularly depleted. Messages periodically circulate around the hospital asking for volunteer donors of a specific blood type, and hospital staff become emergency blood donors far more often than is preferable. This situation is at least partially due the lack of direct communication between the rural hospitals and the officials in Kenya's central blood banks.

Together with University of Nairobi research assistant Eric Magutu, we developed a SMS server application that enabled nurses in the local district hospital to directly provide current blood supply levels directly to officials in the centralized blood bank via text message. The information in the text was incorporated in a visual representation of blood levels at the local hospitals. This would have allowed officials to recognize an upcoming shortage before the rural blood supplies were fully depleted. However, the original SMS Bloodbank system was a total failure. While the officials at the centralized blood repositories could regularly check the website to view the rural blood supply levels, the levels were almost never updated by the nurses in the rural hospitals. While virtually all of the nurses owned a mobile phone, they were very reluctant to use their own airtime to regularly pay for the text message necessary to update the system.

In the summer of 2007, we made a slight modification to the SMS Bloodbank system such that it would automatically transfer a small amount of airtime to each nurse after they sent an appropriately formatted SMS with the day's blood supplies. With the simple additional feature, we changed our failed system into an information gathering system that is now being considered for national deployment. Our experience with this initial automatic compensation system led us to consider other types of work that can be completed on (extremely) low-end phones.

¹ <http://mturk.com>

Mobile Airtime & Money Transfers in East Africa. Originally developed to enable subscribers to transfer small amounts of airtime to others, transfer protocols like Safaricom’s Sambaza² or MTN Rwanda’s me2u³ systems became quickly popular throughout East Africa. It was common practice in many rural areas for mobile phone subscribers to accept an airtime transfer to their phones *in lieu* of a cash payment.

The popularity of airtime transfer systems in East Africa and the use of airtime as a surrogate for currency in rural regions led local mobile phone service providers to allow subscribers to transfer currency as well as airtime. MPESA⁴, the first of such a system, was originally developed by Safaricom in Kenya to enable any mobile phone subscriber to send or receive Kenyan Shillings from another mobile phone subscriber. A recipient can withdraw money at any Kenyan post office or one of the thousand MPESA agents throughout the country. Virtually overnight Safaricom became the largest bank in East Africa, and MPESA is rapidly scaling to neighboring countries.

Crowdsourcing on the Web. Enabling the public to earn small amounts of money as compensation for completing simple tasks is not a novel idea. Crowdsourcing, a term coined in 2006 [8], has taken off in a wide range of application domains (see Table 1). Amazon’s Mechanical Turk (AMT) is a web service that enables anyone to post ‘human intelligence tasks’ (HITs) to the AMT’s users (‘turkers’), as well as be paid for completed tasks. Typical HITs include tasks that are easy for humans but typically extremely difficult for computers such as image tagging, natural language processing, and even survey responses. Most tasks require little time to complete and payments are generally on the order of cents.

Table 1. Types of crowdsourcing applications

Application	Example
Image Tagging	Google, Flickr, ESP Game
Natural Language Processing	OpenMind.org
Photo Sales	iStockPhoto
Coding	Rent-a-Coder, Innocentive

Inferring ‘Accuracy’ from Noisy User Responses. Inferring the correct answer from the responses of multiple error-prone respondents has been a problem addressed in detail throughout a variety of academic literature. Dawid and Skeene approached the problem in 1979 when attempting to infer a patient’s history based on potentially biased reports from different clinicians [2]. They introduce an expectation-maximization (EM) model that simultaneously estimates the bias of these different clinicians as well as the underlying latent variable, in this case the patient’s medical record. Variants of this approach have been used for a variety of other applications including linguistic annotations [7], image categorization [5], and biostatistics [1]. While these methods generally assume that all respondents complete all of the available tasks, it is fairly trivial to adjust these models to a crowdsourcing scenario. Snow et al. employ a similar

² <http://www.safaricom.co.ke/index.php?id=244>

³ <http://www.mtn.co.rw/me2u.htm>

⁴ <http://www.safaricom.co.ke/index.php?id=745>

EM model to infer respondent bias in categorical data [6], while Sheng et al. discuss the problem of response uncertainty and methods to estimate the number of samples required to achieve a given confidence of a correct response [3].

3 Tasks

Humans can outperform computers in many types of simple tasks. Below are some examples of potential tasks using only text and voice communication channels that can be completed on every one of the 4 billion GSM phones in the world today [9].

Transcription. The global transcription market is expected to grow to 18 billion dollars by 2010. With currently over \$12B in annual revenue, standard rates in the medical transcription industry are \$100 for every 1,000 lines of text, or approximately 10 cents per 65 character line, with a 98% average accuracy rate [10]. Using SMS concatenation, we have shown that 5 lines of audio text can be written down by hand and then copied into an SMS in less than 2 minutes. Paying proficient users \$3/hour to do this work on their mobile phone drops the cost to 2 cents per line. Even collecting 100% overhead to cover payments to the operator and our other partners, this still results in a price reduction of over 60% from today's transcription rates, corresponding to an annual savings within the medical transcription industry of over 7 billion dollars.

Software Localization. There are over 60 distinct languages in Kenya alone. It is impossible for software companies to incorporate these languages into their interfaces because no translation service exists currently. We are generating a 'phrase book' of relevant words in every Kenyan language and to date, txteagle users in Kenya have translated these words into more than 15 local languages.

Citizen Journalism. While the idea of citizen journalism is not new, it has yet to have significant penetration within the developing world due to low connectivity and literacy rates. Furthermore, the need for compensation is critical to aspiring journalists in the developing world. By leveraging our existing mobile payment system we enable anyone with a phone to submit a story for free and have compensation directly transmitted back to the handset.

Search Relevancy. The major players in the multi-billion dollar internet search industry all have similar approaches to generating search results and as a result, all perform fairly similarly. By augmenting the machine learning search algorithms with human input, it may become possible to make dramatic improvements in relevancy for specific queries.

Blog / News Translation & Sentiment. Major products get mentioned online thousands of times every day in many different languages. Gauging the sentiment of what is being written about a particular product is critical information for branding and marketing departments, yet extremely difficult to quantify in an automated way using traditional natural language processing techniques. We will ask users if a section of web-scraped text expresses a positive, negative, or neutral sentiment about a particular product or brand.



Fig. 1. There are particular words or phrases that are necessary to adequately localize software, such as the interface of a mobile phone. The photos above shows a phone receiving a translation task targeted at users who speak Giriama, the local language of coastal Kenya and the user being compensated after the successful completion of a set of translation tasks.

Surveys & Market Research. Conducting large demographic surveys in remote areas of the developing world is a challenging and costly endeavor. However in countries such as Kenya, where even the most isolated areas generally have GSM reception, it becomes possible to remotely administer surveys using text messaging. Phone-based surveys are an efficient and economical method of collecting a wide variety of information including market research, census, health, activity, and commerce data. While there is no ‘correct’ answer when assessing personal opinions, we are also asking users to estimate what the opinion is of the average person in her area. With enough respondents in the area, we will learn if these responses converge.

Beyond txt. We are also developing voice tasks for use during off-peak hours when network usage is significantly below capacity. Our first voice tasks will be rating radio commercials; these ratings will be incorporated into our collaborative filtering algorithms to provide subscribers with targeted advertisements. Additionally, we are working with Nokia to develop tasks for use in training speech recognition engine. Lastly, we are hoping txteagle tasks will soon be able to be completed online, taking advantage of the dramatic growth of the mobile web in many parts of the developing world. Whether it is auditory information for local language transcription, image recognition tasks for video security camera monitoring, or simply more extensive text analysis, we are currently working with operators to offer bandwidth rich services during times when their network bandwidth is underutilized.

4 System Design

txteagle is an active service in contrast to most crowdsourcing services, such as Mechanical Turk, services that are passive. We are building our system to actively select the most appropriate tasks for a txteagle user. By customizing the task difficulty for a given user, we have created a more efficient system to generate high-confidence task responses in the shortest amount of time. The more a user responds to tasks, the more the system learns about the user’s areas of expertise and improves the task assignments.

Improving Accuracy through Task Repetition

To ensure tasks have been completed correctly, we send out the same task repeatedly and verify that we are getting the same response across multiple, independent users. Formally, we note that each response y_{ij} comes from an error-prone user j that has an accuracy p_j drawn from a known distribution of accuracies shown in Figure 2a. We initially assume that p_j is independent of a particular task x_i and true answer y_{Ti} , such that $Pr(y_{ij} = y_{Ti} | x_i) = Pr(y_{ij} = y_{Ti}) = p_j$.

A simple way to assess the validity of a response is through majority voting. Following a notation similar to Sheng et. al [3], we infer a single integrated response, \hat{y} that is simply the most popular response from a set of users. The probability this response is correct is noted as $q_i = Pr(\hat{y}_i = y_{Ti})$ and is referred to as the integrated quality of the most popular response. Assuming there are an odd number of users $(2N + 1)$ to avoid the potential of a tied vote, integrated quality q is defined as a sum of probabilities,

$$q = Pr(\hat{y} = y_T) = \sum_{w=0}^N \binom{2N+1}{w} \cdot p^{2N+1-w} \cdot (1-p)^w \tag{1}$$

where w is the number of potential incorrect labels. Figure 2b represents how integrated quality increases as a function of the number of independent respondents who have an accuracy shown in Figure 2a.

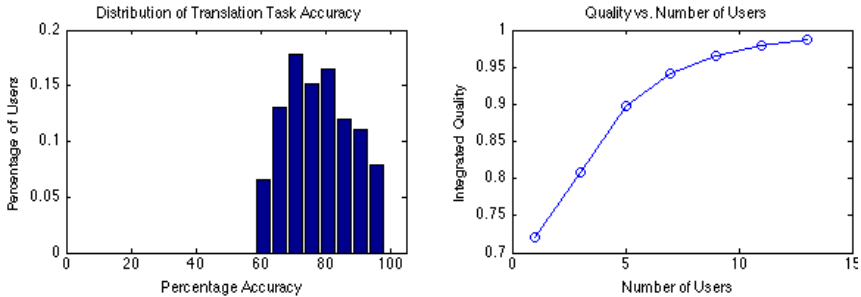


Fig. 2a/b. Figure 2a represents the distribution of initial accuracies of 20 txteagle users for a given set of translation tasks. Figure 2b. depicts how integrated quality q increases as a function of the number of independent users responding to the same task.

Maximum Likelihood Estimates for Task and User Evaluation. While the majority voting technique described above is adequate for simple task completion, we would like to be able to assess the accuracy of an individual user while simultaneously inferring the correct answer to a given task. Dawid and Skene [2] have demonstrated that maximum likelihood estimates of user error rates can be calculated with expectation-maximization (EM) with noisy responses from a set of error-prone individuals (in their case, clinicians collecting a patient record). This technique was similarly used for

image classification [5] and Mechanical Turk data [6]. We have developed a similar model infer correct answers, estimate txteagle users' accuracy levels, and infer their expertise (individual accuracy conditioned on task type) based on their response history.

Following a similar notation as above, let T be the total number of tasks and n_{yi} be the number of times x_i is labeled y_i . We define C_{yi} as a binary variable with a corresponding value of 1 if y_i is the true label (out of a total of L possible labels) of x_i and 0 otherwise. If we initially assume that we know the true label y_i , then

$$p(y_{i1}, \dots, y_{iW} | y_{Ti}, x_i) \propto \prod_{y=1}^L p(y | y_{Ti})^{n_{iy}} \tag{2}$$

and assuming the user responses are independent, the joint can be described as

$$p(y_{i1}, \dots, y_{iW}, y_{Ti}) \propto \prod_{i=1}^T \prod_{y=1}^L \left(p(y) \prod_{y=1}^L p(y | y_{Ti})^{n_{iy}} \right)^{C_{yi}} \tag{3}$$

Following from Equation 3, the maximum likelihood estimators of $p(y|y_T)$ and $p(y_T)$ can be simply written as

$$\hat{p}(y_i | y_{Ti}) = \frac{\sum_i C_{yi} n_{yi}}{\sum_1^L \sum_i C_{yi} n_{yi}} \tag{4}$$

$$\hat{p}(y_{Ti}) = \frac{1}{T} \sum_i C_{yi} \tag{5}$$

Equation 4 enables us to get an estimate of each users' accuracy and with the application of Bayes' rule we can find the posterior

$$p(C_{yi} = 1 | y_{i1}, \dots, y_{iW}) = \frac{1}{K} \prod_{y=1}^L p(y | y_{Ti})^{n_{yi}} p(y_{Ti}) \tag{6}$$

where K is a normalization constant

However, for the development above, we need to know C_{yi} to be able to infer the probability of the prior and the probability of the label conditioned on the prior. If we do not know the true answers, we can use expectation-maximization (EM) to find a local maximum of the likelihood function, treating C_{yi} as a hidden variable, obtaining an initial estimate of the its expected value

$$E[C_{yi}] = \frac{n_{yi}}{\sum_y n_{yi}} \tag{7}$$

and subsequently selecting values of $p(y|y_T)$ and $p(y_T)$ to maximize the Equation 7. Given these new estimates, we recalculate the conditional expectation of C_{yi} and repeat until the estimates converge.

5 Interaction Scenarios

The following example user scenarios are intended to illustrate how txteagle facilitates financial earnings in rural settings.

Ruth & Betty, Home-Maker / Village Phone Operator, Butare, Rwanda. Ruth is the mother of four and while she reads and writes English fluently, she hasn't been able to find much work in her local village. She'd like to own a phone, but hasn't been able to save up the money. Betty operates a village phone in Ruth's village. By 'renting' the phone to Ruth for 50 cents/hour during off-peak times when Betty has no other customers, Ruth is able to complete 3 hours of transcription tasks – accumulating \$7.50 into her savings account and \$1.50 into Betty's account. A couple of more sessions like that and Ruth will be able to afford her own phone!

David, Maasai Herdsman, Kisumu, Kenya. While David had been unable to complete formalized education, he, along with many of his Maasi peers, does own a mobile phone. David completes voice-tasks, helping Nokia train a speech recognition engine on his native Maasai dialect. When David wishes to complete a task, he 'flashes' the txteagle Asterisk box that calls him back, asking him to repeat specific key words and phrases. After 30 minutes of work, David has earned enough airtime to last him a week (assuming he doesn't trade it again for another necklace to give to his new wife).

Sophie, Unemployed School Teacher, Kilifi, Kenya. Sophie is an active member of her local woman's microfinance institute and understands the importance of savings. However, she rarely is able to make ends meet for her growing family and has no money to open a traditional savings account. Occasionally she is able to borrow her husband's phone and uses it to begin earning small amounts of money by completing translation tasks for Nokia; texting back the correct translation to words such as "Address Book" in her mother-tongue, Giriama, she soon is on the road to accumulating her own savings.

Emmanuel, Recent high-school graduate (age 14), Mtwapa, Kenya. Emmanuel has just finished secondary school, but is unsure what to do next. He has a lot of time on his hands but he could really use money of his own. Emmanuel begins experimenting with the txteagle transcription tasks and finds that he gets fairly high accuracy scores. Now instead of spending his time loitering around the village with his peers, he's riding a new bike he purchased with his savings.

6 Deployment

As of February 2009, the txteagle service is currently deployed in Kenya on the Mobile Planet shortcode 3007 and will be soon launching in Rwanda. We have studied usage of the service at local high schools, taxi drivers, and with a group of security guards (askaris) in Nairobi. While the groups successfully completed the translation tasks with similar accuracy (~75%), we found the high-school students were able to complete almost twice as many tasks as the drivers and security guards, presumably due to their proficiency with text entry on numeric keypads.



Fig. 3. Two txteagle users completing tasks in Nairobi. During this particular session, they both completed 20 translation tasks in their native languages (Kikuyu and Luo, respectively) and earned 100 Ksh (US\$1.25) of airtime.

Porting to USSD. Unstructured Supplementary Service Data (USSD) is a common protocol common to all GSM phones that enables services such as checking an airtime balance (dialing *144#) or topping up with an airtime scratch card (*141#12345678912#). Because over 95% the Kenyan and Rwandan mobile phone markets are prepaid, subscribers are extremely familiar with USSD applications and, unlike SMS, there is no fee associated with its usage. (In Kenya, this has meant that balance enquiry events occur more often than actual phone calls!) USSD is also sessions based, as opposed to SMS which is store-and-forward, which means there is very little latency with interactive USSD-based services, making it an ideal platform for menu-driven applications such as surveys. In collaboration with MTN Rwanda and Safaricom, we are developing a USSD txteagle service that is scheduled to launch in both countries during the summer of 2009.

7 Conclusions

Mobile phones are not only empowering most humans on the planet to connect with each other in real-time communication, but soon phones will enable anyone to conduct real-time, peer-to-peer financial transactions. Although the ability for a Western company to remotely ‘hire’ a local Kenyan for the completion of simple tasks may be troubling for some [11], it is the author’s view that providing additional sources of income to remote and impoverished areas of the world has the potential to have a profound and transformative positive impact on the lives of billions. While this project has demonstrated that it is possible to do work and earn money through a phone, we envision txteagle being the first of many such services that help transform billions of mobile phone subscribers into a powerful knowledge workforce.

Acknowledgements. The author would like to acknowledge the following organizations instrumental in txteagle’s initial development and deployment: Mobile Planet (Kenya), Safaricom (Kenya), MTN Rwanda (Rwanda), and Nokia (Finland).

References

1. Albert, P.S., Dodd, L.E.: A Cautionary Note on the Robustness of Latent Class Models for Estimating Diagnostic Error without a Gold Standard. *Biometrics* 60, 427–435 (2004)
2. Dawid, A.P., Skene, A.M.: Maximum Likelihood Estimation of Observer Error-Rates Using the EM Algorithm. *Applied Statistics* 28(1), 20–28 (1979)
3. Sheng, V., Provost, F., Ipeirotis, P.G.: Get Another Label? Improving Data Quality and Data Mining Using Multiple, Noisy Labelers. In: *Proc KDD* (2008)
4. Singh, P.: The public acquisition of commonsense knowledge. In: *Proc. Of AAAI Spring Symposium on Acquiring (and Using) Linguistic (and Wold) Knowledge for Information Access* (2002)
5. Smyth, P., Burl, M.C., Fayyad, U.M., Perona, P.: Knowledge Inferring ground truth from subjective labeling of Venus images. In: *NIPS*, pp. 1085–1092 (1994)
6. Snow, R., O'Connor, B., Jurafsky, D., Ng, A.: Cheap and Fast – But is it Good? Evaluation Non-Expert Annotations for Natural Language Tasks. In: *Proc. of the Empirical Methods in Natural Language Processing, Honolulu*, pp. 254–263 (2008)
7. Wiebe, J.M., Bruce, R.F., O'Hara, T.P.: Development and use of a gold-standard data set for subjectivity classifications. In: *Proc. of ACL* (1999)
8. <http://www.wired.com/wired/archive/14.06/crowds.html>
9. <http://www.gsmworld.com/newsroom/press-releases/2009/2521.htm>
10. Official Newsletter of the Medical Transcription Industry Association of the Philippines (January 2008)
11. Exploitation or globalization,
<http://news.bbc.co.uk/2/hi/technology/7881931.stm>