

# Sustainable Employment in India by Crowdsourcing Enterprise Tasks

Shourya Roy  
Xerox Research Centre India  
Bangalore, India  
Shourya.Roy@xerox.com

Chithralekha Balamurugan  
Xerox Research Centre India  
Bangalore, India  
Chithralekha.Balamurugan@xerox.com

Sujit Gujar  
Xerox Research Centre India  
Bangalore, India  
Sujit.Gujar@xerox.com

## ABSTRACT

This paper explains how crowdsourcing would constitute for a promising and a successful alternate business model in India, especially at the juncture where the advantages of outsourcing is beginning to fade. The success of this alternate business model depends on the solutions for two challenges – by taking work to people to leverage large educated population in India and making enough work available for the workforce to work and earn. Though solution to the first challenge is implicitly available through increasing penetration of Information and Communication Technologies (ICT) in India, the second challenge requires enough tasks to be available by enabling business organizations to adopt crowdsourcing. Since enterprise tasks are not readily *crowdsourcable* owing to security, compliance and contractual reasons, this paper proceeds to describe an end to end system which encompasses technical solutions that could help crowdsourcing business tasks, by tactfully overcoming the existing business constraints. For business tasks, we consider *Insurance Claim Form Digitization* which is one of the most common tasks taken up by outsourcing enterprises.

## Keywords

Crowdsourcing, Data digitization, Employment models, Microtasking, Amazon Mechanical Turk

## 1. INTRODUCTION

Outsourcing and India have been synonymous with each other over the last couple of decades. However, increasing cost of delivery of outsourced services has been eroding the benefit of labor arbitrage based on differential wages. Secondly, increasing number of countries from different parts of the world are offering outsourced services at decreasing cost challenging India's position at the pinnacle. Finally, outsourcing is also contributing to tremendous disparity of employment between bigger cities like Bangalore and other parts of the country. This is because outsourced jobs are mostly technical-oriented addressing the employment requirements of the technically skilled and mostly urban based population. Ironically, about 60% to 70% [of BPO

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.

DEV '13, January 11-12, 2013 Bangalore India

Copyright © 2013 ACM 978-1-4503-1856-3/13/01... \$15.00

recruitment] happens in smaller towns, where the BPO doesn't

have a presence. And of those selected, just 20% to 27% accept; the rest don't move for [a variety of] reasons. Even after moving to the cities, a significant fraction of workers eventually go back to their bases by not being able to adjust with different societal lifestyle. This leads to high degree of attrition and absenteeism in the outsourcing industry.

Crowdsourcing<sup>1</sup> which has been emerging as a paradigm of massive scale distribution of knowledge work can play a big role there. Enterprises and business organizations have started using crowdsourcing for a variety of tasks including ideation, solving complex problems, data tagging etc. We bring in the notion of leveraging crowdsourcing for service delivery in outsourcing enterprises. We consider a typical outsourced task viz. *Insurance Claim Form Digitization* and propose that a large fraction of the same do not need to happen in cities like Bangalore but can be crowdsourced for people to work from anywhere in the country. This will not only generate employment across the country but also will not uproot people which was one of the major contributors of high attrition in this industry. Crowdsourcing has been observed to be significantly more time and cost efficient than traditional models. Hence, thereby leveraging the large population and increasing penetration in online platforms[2] can help India to leap ahead of competition.

Form Digitization tasks involve manually, assisted by software tools such as Optical Character Recognition(OCR), digitizing data from scanned images of insurance forms. This does not require high-skilled labor force and hence, can be considered for crowdsourcing. However, insurance claim forms are not easily *crowdsourcable* because of security, compliance and contractual challenges which arise while moving beyond the physical and digital boundaries of enterprises. While previous work has demonstrated feasibility of digitizing images by microtasking[1], our contribution is an end to end system which describes how enterprise tasks could be taken to the crowd for data entry while still fulfilling business constraints. In the next section we describe how digitization of insurance forms, can be crowdsourced while matching enterprise compliance requirement of time, cost and quality of work execution.

---

<sup>1</sup> <http://en.wikipedia.org/wiki/Crowdsourcing>

## 2. TECHNOLOGY DESIGN AND DESCRIPTION

In this section, we describe how we address issues arising from disparity in security, quality, and compliance controls between outsourcing and crowdsourcing settings. A basic crowdsourcing model comprises of the following three steps:

- Create TASKs, where TASK refers to microtasks posted on crowdsourcing platforms
- Distribute TASKs
- Validate and Integrate Results

### Creation of TASKs – Intelligent Microtasking

The following are the steps to be carried out for creation of TASKs which can be collectively called as *Intelligent Microtasking*:

- **Context removal** - the labels (e.g. *patient name, hospitalization date*) associated with the various fields are to be removed in order to partially shield the semantics of the data available in that field – for enforcing a minimal level of confidentiality of data
- **Split form into segments** - This function takes a filled form with the labels removed and splits it into image pieces corresponding to the various fields.
- **Group segments** - images corresponding to multiple related fields (e.g. all *Name* fields) are grouped together to form a composite field to be distributed to the crowd.
- **Label Segments** – Segments are assigned labels based on the level of security required, probable rejection frequency of that segment and the dependency nature of the segment w.r.t other segments. Rejection refers to the scenario where workers have the option of rejecting a form owing to poor quality of scanning of forms, illegible handwriting etc.
- **Create TASKs with easy User Interface** – TASKs are created using simple and minimal on-screen instructions for performing the data entry, providing appropriate UI boxes for data entry, on-Screen Validation, enabling TASK acceptance and rejection of the individual group fields at worker's discretion.

### Distribution of TASKs - Automated Optimized Distribution

The automated distribution function is designed in such a way that it helps for an optimized distribution of form processing task to the crowd leveraging on the following:

- **Type and contents of Fields** – Check box fields (e.g. Gender) and blank text boxes are processed by automatic image processing techniques and not sent to the crowd, thereby reducing cost of crowdsourcing.
- **Controlled redundancy** – While we depend on agreement based on redundant distribution but redundant distribution happens in a staged manner. If we are looking for majority agreement in a redundancy of 5 then only 3 TASKs are posted initially and more are posted only if required.
- **Intersection dependency logic** – This leverages on the interdependency that exists between the TASKs and dispatch them in an order that would help to minimize distribution, if the required values could be obtained / inferred from already fields.

### Validation and Integration

At this phase, the data entry for the TASKs worked by the crowd is completed with a subset of business rules being enforced through on-screen validation. For enforcing the remaining set of business rules, post-processing modules work upon the entered data. For example, an entered name field has to be split to 'first name', 'last name' and inserted in appropriate database fields. In addition to post-processing, automatic validation for certain fields could be executed for ensuring correctness of data. For example, checking if both gender options are ticked, checking if any two fields do not conflict on their values could be done using automatic validation and flagged for manual validation on violation.

## 3. EXPERIMENTAL RESULTS

The feasibility of the use of crowdsourcing for enterprise application is demonstrated by deploying the proposed system over the most popular crowdsourcing platform, Amazon Mechanical Turk (AMT). The digitization of CMS1500 health insurance claim forms is the task chosen. We created around 50 forms by volunteers from our organization. There are three key highlights of the results that was obtained:

1. Average time taken to complete a form is around 10 hours but nearly 90% of tasks get done in half of that time. This is in line with findings from earlier work which showed the rate of completion of work on crowdsourcing platforms is very high in the beginning[3].
2. The cost to process each form was around 15 cents on average which is significantly less than what companies charge today (of the order of a dollar(\$) per form).
3. While we did not have enough data to calculate accuracy numbers of crowd workers but we were impressed with crowd performance on some of the very poorly written data points.

Thus, the technical solution describes how enterprise tasks could be crowdsourced by overcoming the security challenges. The experimental results help to prove that compliance requirements could still be achieved in a crowdsourced task execution model which provides enterprises with encouraging evidences for crowdsourcing their tasks. However, more extensive experiments will be required to validate the claims. Ultimately, this would constitute for a technology enabled sustainable employment solution for India by leveraging on the extensive mobile phone consumerism characteristic of India.

## 1. REFERENCES

- [1] K Chen, A Kannan, Y Yano, JM Hellerstein, TS Parikh, Shreddr: pipelined paper digitization for low-resource organizations. In the proceedings of ACM DEV'12, March 12
- [2] Ipeirotis, P.G. 2010a. Demographics of Mechanical Turk. CeDER Working Papers, CeDER-10-01, <http://hdl.handle.net/2451/29585>
- [3] Your estimated waiting time: Infinite! <http://behind-the-enemy-lines.blogspot.com/2009/02/your-estimated-completion-time-infinite.htm>