Collecting Expert Opinions in Non-Profit Sectors in Bangladesh for Poverty-Alleviation Plans in South Asia

Dr. Alejandro Sánchez-Becerra, James Park

Emory University

1. Abstract

Social investments can be quite costly when it comes to extreme scales, such as implementing poverty-alleviation plans in multiple densely populated countries. Thus, it is important to obtain accurate data to explain the attainability and plausibility of these plans and justify costs. This large-scale study looks into collecting expert opinions regarding these plans in multiple sectors within certain countries in South Asia (India, Pakistan, and Bangladesh) and finding similarities and differences between organizations that hold these experts. For this research, the focus is primarily on Bangladesh and non-governmental organizations (NGOs). Through pre-processing of existing data and collecting contact information through code and data collection tools, the final product consists of summary statistics regarding the updated dataset, the dataset with information regarding these NGOs and experts, and a summary paper reiterating results and how this process took place.

2. Preprocessing

Given two PDF files that hold Bangladesh NGO information, one local and one international, the first task was to convert this data into files that could be easily accessed through Python and R. We looked into PDF to CSV conversion tools on the internet and were able to obtain our first two datasets. Both the local and international NGO files needed further preprocessing and thus led to additional tasks.

Preprocessing Steps:
1. Convert PDF to CSV
2. Move information to accurate columns
3. Remove empty columns
4. Check for empty rows and repeats
5. Change column names
6. Concatenate rows
7. Combine both datasets

The initial files consisted of 10 columns:

- *Sl. No.* (serial number)
- *Name of NGOs*
- *Address*
- *Reg. No.* (registration number)
- *Reg. Date* (registration date)
- *Renewed On* (NGO renewal date)
- *Valid upto* (NGO expiration date)
- *District*
- *Country*
- *Remarks* (comments)

During the PDF to CSV conversion stage, information leaked into different sections, ultimately pushing values to the right. This led to both files having more columns than needed and often having information in the wrong sections. Thus, we needed to go through the dataset and ensure that accurate data points were in their respective fields. Afterwards, once confirming that the data were in the right columns and that these additional sections were empty, we removed these unnecessary columns. We also removed rows that did not have an NGO name and repeats of different NGOs. Some column names were changed as well for easier access when data processing: *num_var, ngo_name, ngo_address, reg_no, reg_date, renewed_on, and valid_until*. Rows needed to be concatenated for each NGO, since each line in the original PDF file created a separate row in the CSV file. After polishing the two files, both datasets were combined and a new column, *type_ngo*, was added to indicate whether or not an NGO is local or international.

---

3. Results

The initial combined dataset had 2328 entries of local NGOs and 267 entries of international NGOs, leading to a total of 2595 entries. Out of these 2595 values, 969 NGOs were located within Dhaka, the capital of Bangladesh. The district with the second largest count for NGOs was Chittagong, holding 87 entries. This showed us that the majority of registered NGOs were located within Dhaka.

Given these NGOs, we pulled a random sample of 20 organizations and looked for information on these companies online to see how significant this dataset could be.

To log findings, we created two new columns in the random sample file: *info_found* ("Yes" or "No") and *loc_found* (website where contact was found). We searched for NGO information using these steps.

Searching Steps:
1. Search for given NGO name on a search engine, like Google
2. If link is not found, look up NGO name with words "NGO Bangladesh"
3. If link is still not found, log "No" in *info_found* for NGO
4. If link is found, click on available link and search for contact information
5. If no contact information is found, log "No" in *info_found* for NGO
6. If an email is found, log "Yes" in *info_found* and record website link in *loc_found*

For the 20, we were only able to find contact information for 10: a 50% success rate. This led to a revision of the data to see why this might be the case, and we realized that many NGOs in the dataset have already expired (*valid_until* date has passed the current date). Thus, we went back into cleaning the data and removing all NGOs that have expired. This removed 643 entries and brought the combined dataset down to 1480 local NGOs and 205 international NGOs, leading to a total of 1685 entries. For this new dataset, we were able to find information for 17 entries given a random sample of 20 companies: an 85% success rate. When given a random sample of 50 companies for this updated dataset, we were able to find contact information for 34 NGOs: a 68% success rate. These high rates showed that

this new dataset could provide useful information for collecting experts' opinions in Bangladesh. Furthermore, when looking into 41 different sector types, ranging from Animal Husbandry to Science & Technology (these fields come from the official NGO Darpan website, where we downloaded the initial PDF files), the majority of these NGOs were in the fields of Human Rights (12 out of 34) and Rural Development and Poverty Alleviation (11 out of 34). Other NGO fields did not account for a significant portion of the 50. This showed that these NGOs could potentially host experts that can provide significant responses for the proposed plans.

After validation of the data and seeing that it indeed results in a high success rate, we had to look for ways to search for these NGOs and their contact information efficiently; a manual search of 1685 entries would not be ideal for a single individual. Thus, we looked into using Amazon Mechanical Turk, a crowdsourcing website where businesses can hire crowdworkers to perform discrete on-demand tasks that computers are currently unable to do as economically, to complete this task. In order for this operation to begin, we needed to create an interactive page for Amazon Mechanical Turk where crowdworkers could receive information regarding the NGO and how to look for it. Furthermore, we needed this page to be able to obtain information from these crowdworkers and log this data into a separate file. Therefore, we used HTML to create a webpage that would provide the worker with relevant details, such as *ngo_name*, *type_ngo*, and *valid_until*, and can simultaneously take inputs regarding NGO information, if found. To expedite the search process, we added four new columns to the updated dataset: *text_ngo_name, text_ngo_name_words, link_ngo_name, link_ngo_name_words.* The first two columns include pre-made Google search links (the former with the NGO name, the latter with the NGO name and the additional words "NGO Bangladesh"). The last two columns include hyperlinks to these respective fields. Furthermore, we included a checklist of the 41 different sector types from NGO Darpan to gather information on what industry most of these NGOs are a part of. Once uploading the HTML and CSV files and specifying the dataset that needs to be traversed, Amazon Mechanical Turk will pull each NGO information from the CSV file and report it to each crowdworker. Then, Turk will record the information and links the crowdworkers provide it on a separate file. Ultimately, the information that we would be receiving are as follows.

Logging Details:
- *website_type* (is the website local, international, social media, etc.)
- *email_entered* (email of NGO, if found)
- *website_email* (website where email was found, if found)
- *type_contact* (local, international, NA)
- *ngo_field* (what does the NGO specialize in, if found)
- *ngo_field_specific* (in case NGO field is not listed above, log predicted field)

With these details, we can obtain contact information of experts within NGOs in Bangladesh on a wide scale efficiently, supporting the poverty-alleviation plans.