Process write-up: Smallholder text mine

Introduction

The web scraping and text mining completed for this project were intended to discover the extent to which large agricultural projects financed by major international finance institutions (IFIs) have focused on smallholder farmers in recent years, as well as compare the intensity of this focus between institutions over time.

Process

We looked at projects within the agricultural sector from 2015-2021 from the World Bank, Inter-American Development Bank (IDB), International Fund for Agricultural Development (IFAD), Asian Development Bank (ASDB), and African Development Bank (ADB). For each institution, the loan agreements for relevant projects funded by the institution itself (and not outside organizations) above USD $5 million were downloaded from the institution website. A number of IFAD documents from 2015-2016 which were unavailable online were accessed through the institution itself. The specific navigational processes for the websites of each institution are detailed here.

A metadata file was created as an Excel spreadsheet to attach organization name, country, region, year, and financing amount to each project document. A folder of the collected PDF documents and metadata was put through the Distant Reader[1], a software created by Eric Lease Morgan of the Navari Family Center for Digital Scholarship[2] at the Hesburgh Libraries, which takes an input of text and outputs structured data describing it, including plain text files for each document.

Extraneous text such as annexes, tables of contents, and lists of abbreviations or contact information was eliminated from each text file before they were put through the Distant Reader again.

In order to analyze the text for mentions of "smallholder farmer" specifically, we created a list of terms that captured this idea in various ways. This list included the terms small farm, small holder, smallholder, small scale, smaller scale, and small producer. The software also catches plurals and altered forms of words such as "farmer" from "farm."

In conjunction with the second round of Distant Reader outputs, our programs[3] searched for these queries within the text and outputted them as frequencies relative to the amount of words in a given corpus. Using the metadata, we separated the full corpus of text into corpora by year, institution, and region as well as sub corpora defined as Institution by Year (ie World

[1] https://distantreader.org/
[2] https://cds.library.nd.edu/
[3] https://github.com/ericleasemorgan/ifad-toolbox

Bank-2015, IDB-2015, etc). In this way, the output was divided into the corpora and sub corpora to compare mentions of smallholders from different years and/or institutions.

Outputs

The numbers visible in the corpora graphs (Figures 1-3) represent the amount this collection of terms appears relative to the size of each corpus. For example, the output for IFAD in Figure 1 shows how frequently the set of terms was written relative to all the words written by IFAD. Figure 2 repeats this by year, and Figure 3, by region. The original net values of target words divided by total words are multiplied by 10,000 to get the relative weights seen in Figures 1-3.

The sub corpora of institutions by year breaks down the full value of an institution shown in Figure 1 into its mentions by year from 2015 to 2021. That is, the sum of all years for each institution in Figure 4 equals the relative weight of that institution as shown in Figure 1. This shows how mention of smallholders fluctuates *within an institution* over the time period studied. It is important to mention that the IDB data ends after 2019, as all IDB projects in agriculture after this year were too small to meet the USD $5M minimum. Similarly, though all of IFAD's projects are agriculturally focused and many presumably meet the USD $5M requirement, the project documents for IFAD past 2018 are not yet publicly available.

We also used the financing amount to weigh the number of smallholder mentions by the amount of money spent on the projects. In Figure 1, we see that IFAD mentions smallholders more than twice as much as the World Bank. However, it is clear in Figure X, that the World Bank invests far more money into their projects in total, meaning that each mention of smallholders in their project documents carries more weight in monetary value. In other words, it may be more valuable for smallholders to be mentioned in World Bank projects than in IFAD projects because of the amount of financing attached to them. Nevertheless, the primary aim of this study was to compare the relative *focus* on smallholders rather than the net amount of *money* directed at smallholders, and the proportion of project money spent on smallholders may not be exactly equal to the proportion of mentions in project documents. Still, it is important to note the discrepancies in funding between institutions as part of understanding the results and their implications.

Conclusions

As previously mentioned, this text mine sought to understand the amount of focus that different international finance institutions place on investing in smallholder farmers. It found that IFAD is significantly more invested in smallholders than other IFIs, with almost twice the number of text mentions than the World Bank at the second highest. Additionally, it found that mentions of smallholders have been steadily declining across the board from 2015 to 2020, though shot up in 2021, presumably in response to the COVID-19 crisis. Throughout all years, smallholders received greater focus in Africa relative to Latin America and the Caribbean, Asia, and the Middle East. Along with these results, the scale of financing from

different IFIs, namely, the larger scale of the World Bank's financing relative to the other four institutions, should be noted. Nevertheless, IFAD seems to be most committed to investing in smallholder farmers as a general rule.