5th International Conference on Advanced Research Methods and Analytics (CARMA2023)
Universidad de Sevilla, Sevilla, 2023
DOI: http://dx.doi.org/10.4995/CARMA2023.2023.16433

Food insecurity trends in the Famine Early Warning Systems Network

Bia Carneiro[1], Chiara Perfetto[2], Giuliano Resce[2], Giosuè Ruscica[1], Giulia Tucci[1]
[1]Alliance of Bioversity International and CIAT, [2]Department of Economics, University of Molise, Italy.

Abstract

Over last 30 years, periodic country analyses elaborated by FEWS NET (Famine Early Warning Systems Network of the United States Agency for International Development) enabled creation of a unique source of knowledge comprising consistent reporting in over two dozen countries. This paper proposes to systematically assess documentation from historical perspective to provide comprehensive overview of food insecurity in FEWS NET covered countries. We propose an integrated machine learning approach to systematically analyse available documentation and generate knowledge. In particular text mining algorithms have been implemented to analyse reports: automated retrieval of high-quality information from text, by finding patterns and trends through machine learning, statistics and linguistics. This enables analysis of large amounts of unstructured text to derive insights. Results show that there is a wide heterogeneity in what is relevant, and in what reports focus on at the territorial level. Many country-level topics are persistent over time with some interesting exception, as Guatemala, Malawi, Niger, and Somalia with more instability. Overall, the evidence show that advances in machine learning and Big Data research offer great potential for international development agencies to leverage the vast information generated from reports to gain new insights, providing analytics that can improve decision-making.

Keywords: *Food insecurity; Early Warning Systems; Text Mining.*

1. Introduction

The Famine Early Warning Systems Network (FEWS NET) is a leading provider of early warning and analysis on acute food insecurity around the world. Created in 1985 by the United States Agency for International Development (USAID) in response to devastating famines in East and West Africa, FEWS NET provides evidence-based analysis to governments and relief agencies who plan for and respond to humanitarian crises. FEWS NET analyses support resilience and development programming as well. FEWS NET analysts and specialists work with scientists, government ministries, international agencies, and NGOs to track and publicly report on conditions in the world's most food-insecure countries.

The FEWS NET reporting includes:

- Monthly reports and maps detailing current and projected food insecurity
- Alerts on emerging or likely crises
- Special reports on factors that contribute to or mitigate food insecurity, including weather and climate, markets and trade, agricultural production, conflict, livelihoods, nutrition, and humanitarian assistance
- Access to data, learning, and analysis of the underlying dynamics of recurrent and chronic food insecurity and poor nutritional outcomes, to improve early warning and better inform response and program design

Over last 30 years, periodic country analyses elaborated by FEWS NET enabled creation of a unique source of knowledge comprising consistent reporting in over two dozen countries (Figure 1). Longitudinal breadth of knowledge has potential to provide information and insights into long-term trends regarding shocks over time, coping capacity, and livelihoods. However, time and resource constraints and focus on current humanitarian assistance mean such outputs underutilized as sources of evidence.

This paper proposes to systematically assess documentation from historical perspective to provide comprehensive overview of food insecurity in FEWS NET covered countries. The insights can also support current operations by providing another evidence-base from which to carry out analyses.

Extracting knowledge from extensive text-based sources poses several challenges (Garbero *et al.* 2021). Conventional approaches such as manual coding or keyword searches time and resource intensive (Carneiro *et al.* 2022). This is the reason why we propose a text mining approach: automated retrieval of high-quality information from text, by finding patterns and trends through machine learning, statistics and linguistics (Resce, Maynard, 2018). This enables analysis of large amounts of unstructured text to derive insights. Our proposal is an integrated machine learning approach to systematically analyse available documentation and generate knowledge.

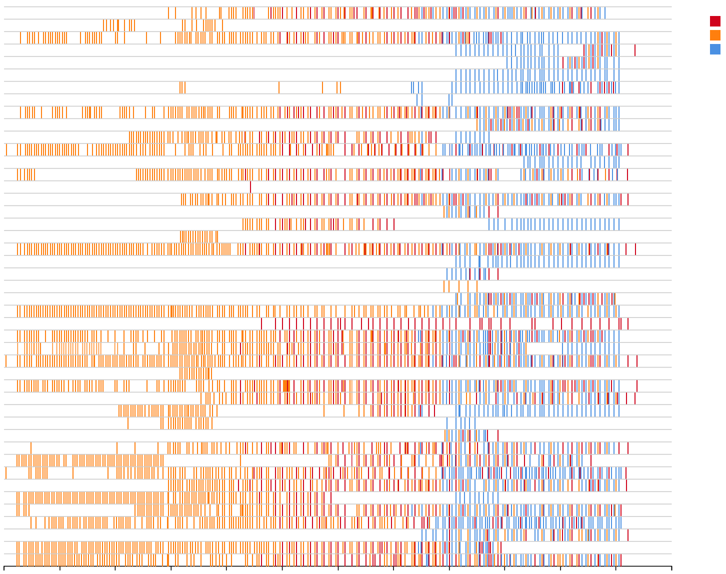

Figure 1 FEWS NET Reports Timeline

The research questions of the present analysis are the following:

- What historical trends (or anomalies) can be identified regarding the prevalence of various dimensions from the FEWS NET framework for acute food insecurity (i.e. reported hazards/shocks and outcome/impacts, particularly focused on indicators related to infectious diseases, agroclimatology, markets, and conflict)?
- How do trends compare by region and type of hazard/shock?
- How are the dynamics, such as the interactions within and between the different dimensions, represented? How do these representations transform in time and space?

2. Material and methods

We start systematizing the 5217 total Documents available in FEWS NET (1017 Food Security Outlook; 2923 Food Security Outlook Update; and 1276). For each report in machine readable PDF format, relevant sections were extracted into spreadsheet with section headers and column title. We only focus on the section National overview, which contains the description of the present situation of the country the report is referring to.

The corpus of analysis was prepared using functions from the R package "tm" (Feinerer, Hornik, 2018; Feinerer *et al.* 2008): punctuation, stop words (i.e. in English, words like "the", "is", "of", etc), and numbers were removed from the corpus. The words were then converted to lowercase and stemmed. The most common formats for representing a corpus of texts (i.e., a collection of texts) in a bag-of-words format is Term Document Matrix (TDM). A TDM is a matrix in which rows are terms, columns are document/report, and cells indicate how often each term has occurred in each document/report. The advantage of this representation is that it allows you to analyse data with vector and matrix algebra, effectively moving from text to numbers. In the TDM the frequently occurring terms are assigned a higher score than the rarely occurring terms. Starting from TDF with the number of occurrences it is possible to quantify what a document is about. The TF-IDF score is another useful metric used to populate the TDM. One measure of a word's importance is its term frequency (TF), which counts a word's occurrence in a document. Another approach is to look at a term's inverse document frequency (IDF), which decreases the weight of commonly used words and increases the weight of words that do not appear frequently in a collection of documents. The two can be combined to calculate a term's TF-IDF (the two quantities multiplied together), which measures the frequency of a term adjusted for how rarely it is used (Silge, Robinson, 2017). Formally:

$$(1) \qquad IDF(term) = ln\left(\frac{n_{documents}}{n_{documents\ containing\ term}}\right)$$

In our case, the TF-IDF combines frequency, i.e., how many times a word is associated to a document, and the inverse of ubiquity, i.e., how exclusive the association is between a word and a document (Hidalgo and Hausmann, 2008; 2009). To this regard, it is worth stressing that more ubiquitous words are more likely to have less informative power than exclusive words.

From the TDM perform correlation analysis to understand to what extent does the usage of a topic change (increase or decrease) in relation to the usage of the same topic in a previous text. A correlation analysis aims to determine the extent to which there is a relationship, or linear dependence, between two sets of points (Jockers, 2014). The correlation is a measure of the strength of the linear dependence between the word frequency in one report and the word star frequency in another report. This result, called the Pearson moment-product absorbing coefficient, is expressed as a number between -1 and +1. A negative coefficient one (-1) negative represents a perfect negative; if the correlation between the reports is -1, then we would know that the higher frequency of a word in one report corresponds proportionally to the lower frequency of the same word in the other report. This means that something changes in the report A positive one (+1) represents a perfect positive correlation (when one variable goes up and down, the other variable does so ideally).

3. Results and Conclusions

Figures 2, 3, 4, and 5 show the TF-IDF word-cloud by countries, the shape of the clouds reflects the map of the country. The figures are organised by Regions: Figure 2 is East Africa, Figure 3 is Southern Africa, Figure 4 is West Africa, and Figure 5 is MENA & LAM.

Figure 2 TF-IDF by country (East Africa)

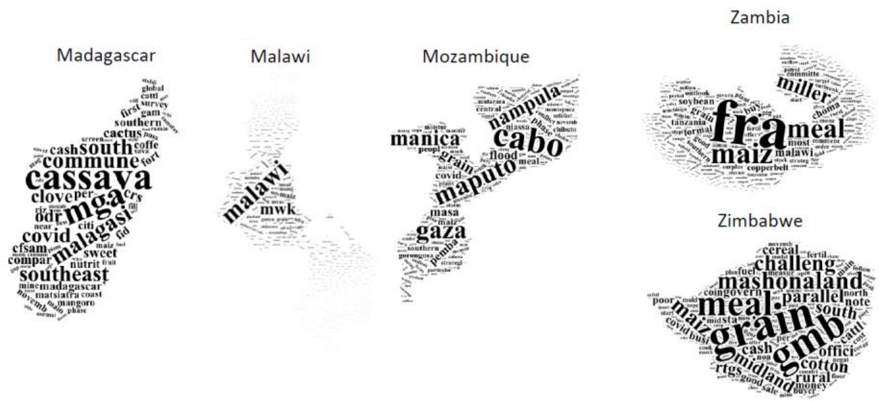

Figure 3 TF-IDF by country (Southern Africa)

Figure 3 TF-IDF by country (West Africa)

Figure 4 TF-IDF by country (MENA & LAM)

Overall, Figures 2-5 show that there are specific terms associated to each country, meaning that there is a wide heterogeneity in what is relevant, and in what reports focus on at the territorial level. Figure 6 shows the average year correlation for a subset of countries having reports in almost all years. The correlation is estimated between a report and the previous report for the same country. We report year average to clean up the analysis for the seasonality that is intrinsic in many events. Figure 6 shows that the correlation is quite high in almost every country, meaning that country-level topics are persistent over time, but there are some interesting exception, such as Guatemala, which shows a flat low correlation. Other countries show a non linear trend, suggesting an internal instability, they are: Malawi, Niger, and Somalia.

Overall, the preliminary evidence show that advances in machine learning and Big Data research offer great potential for international development agencies to leverage the vast information generated from reports to gain new insights, providing analytics that can improve decision-making.

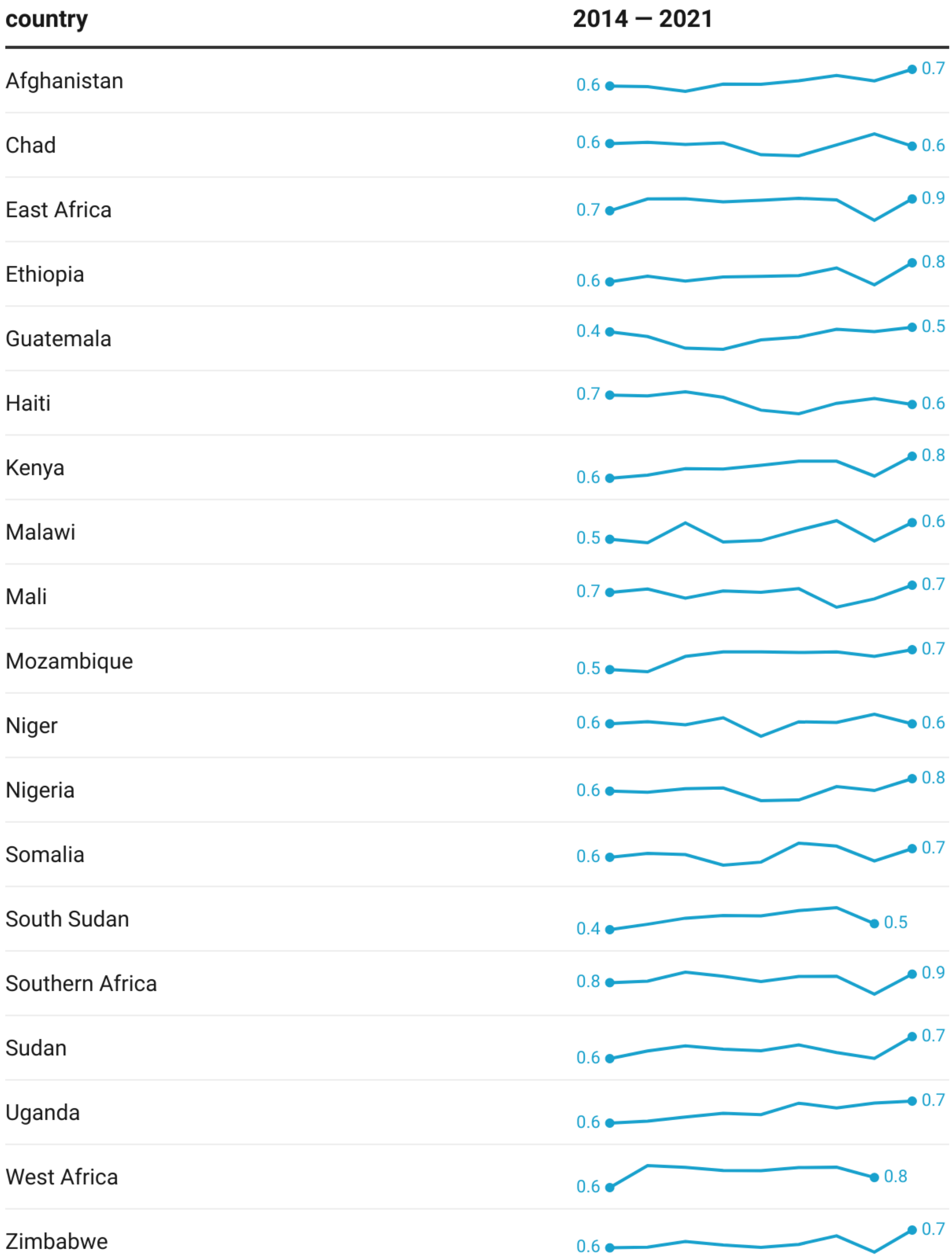

Figure 5 Annual trend of correlation with previous report for a subset of representative Country

References

Carneiro, B., Resce, G., & Sapkota, T. B. (2022). Digital artifacts reveal development and diffusion of climate research. *Scientific Reports*, 12(1), 14146.

Feinerer I, and Hornik K (2018). tm: Text Mining Package. R package version 0.7-5.

Feinerer I, Hornik K, and Meyer D (2008). Text Mining Infrastructure in R. Journal of Statistical Software 25(5): 1-54. URL: http://www.jstatsoft.org/v25/i05/

Garbero, A., Carneiro, B., & Resce, G. (2021). Harnessing the power of machine learning analytics to understand food systems dynamics across development projects. *Technological Forecasting and Social Change*, 172, 121012.

Hidalgo, CA, and Hausmann, R (2008). A network view of economic development. *Developing alternatives*, 12(1), 5-10.

Hidalgo, CA, and Hausmann R. (2009). The building blocks of economic complexity. *Proceedings of the national academy of sciences*, 106(26), 10570-10575.

Jockers, M. L. (2014). *Text Analysis with R for Students of Literature*. Cham: Springer International Publishing.

Resce, G., & Maynard, D. (2018). What matters most to people around the world? Retrieving Better Life Index priorities on Twitter. *Technological Forecasting and Social Change*, 137, 61-75.

Silge, J., & Robinson, D. (2017). Text mining with R: A tidy approach. O'Reilly Media, Inc.