No Village Left Behind: A Moroccan Data-driven Platform for Effective Aid Coordination

Abdelaziz Bounhar
Télécom Paris

Adnan Anouzla
ENSMR

Ahmed Lekssays
Qatar Computing Research Institute

Asma Zizaan
UM6P

Basma Chourane
Seven-app

Fatima Zahra Qachfar
University of Houston

Hafsaa Ouifak
UM6P

Imane Momayiz
Equancy

Loubna Ben Allal
Hugging Face

Maroua Razzouqi
FST Mohammedia

Meryem Jabrane

Mustapha Ajeghrir
Heflow.ai

Nissrine Hatibi
ENSIAS Rabat

Nouamane Tazi
Hugging Face

Soundouss Messoudi
Heudiasyc

Yassir Bendou
IMT Atlantique

Youssef Adarrab
EIGSI Casablanca

Abstract

Following the catastrophic earthquake that hit Morocco in September 2023, our platform emerged to optimize relief coordination, efficiently orchestrating resources to aid those in need. This paper presents the various techniques used to collect and process requests and interventions into a clean and actionable dataset, enabling authorities and fellow NGOs to efficiently extend aid to the affected areas.

1 Introduction

On September 8, 2023, a devastating magnitude 6.8 earthquake hit Morocco's High Atlas Mountains, uniting Morocco's government, Non-Governmental Organizations (NGOs), and citizens in an inspiring display of solidarity. Recognizing the need for improved relief coordination, our data-driven platform was created as a centralized hub to consolidate vital earthquake data and relief efforts.

Yet, consolidating diverse, heterogeneous data sources is complex, particularly in low-resource languages like the Moroccan dialect. In this work, we employed Natural Language Processing (NLP) techniques to convert collected data into a refined and usable dataset. Key challenges included authentic data collection in crises, accurately identifying similar-named rural villages ('douars'), and obtaining precise geolocation despite inaccuracies in mapping APIs (e.g. Google Maps and OpenStreetMap). Overcoming these obstacles was essential for effective humanitarian support.

2 Methodology

Figure 1 shows the overall pipeline of our data-driven platform which will be explained further below.

Curating reference douars data : We used two complementary datasets DS1 [1] and DS2 [2] that both contain locations with their geo-ontology categories names ('region', 'province', 'cercle', 'commune' and 'douar'). Our objective was to align douar lists between DS1 and DS2 using specific
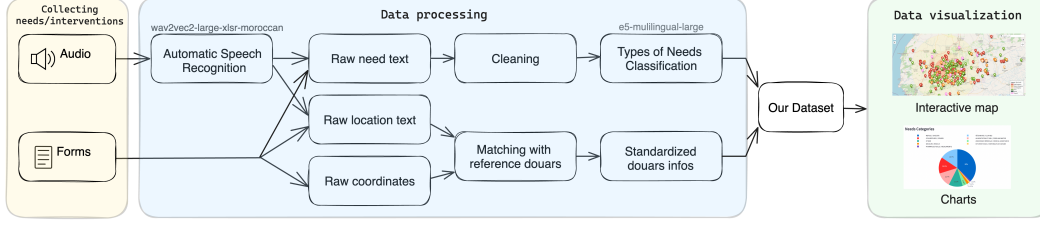
Figure 1: Our pipeline cleans and classifies incoming requests and interventions then affects them to the corresponding reference douars. The resulting dataset is visualized on our online platform.

criteria: i) phonetic distance (using Refined Soundex and Metaphone algorithms) between 'commune' names, ii) same phonetic distance check for douar names, and iii) an edit distance check (Levenstein distance) on douar names. However, these checks, while strict, yielded over 60 douars with more than one match. The final dataset comprises 5197 douars written in both French and Arabic, with their unique geo-ontology categories. This dataset serves as our *reference douars data*.

Collecting needs/interventions: NGOs provided structured information via a form specifying NGO name, intervention dates, supplies provided, etc. Individual requests came through a chatbot, with users sharing their needs in Moroccan Dialect through audio messages or forms specifying their requirements and location. Verification through phone calls ensured accuracy and completeness, addressing missing information and duplicates. In one week, over 2870 requests and 250 interventions were collected.

Data processing: Audios in Moroccan Dialect were transcribed to text using the model `wav2vec2-large-xlsr-moroccan` [3], which is fine-tuned on a Moroccan Dialect dataset. Then, we cleaned data by removing duplicate and inappropriate requests, resulting in a clean textual input.

Following this, we parsed text locations using regex and mapped them to the reference douars' data. To map the parsed input, we generated ngrams and for each ngram: we checked for strict, then fuzzy, then phonetical matching (as explained above). Finally, we selected the matching values with minimum distance per geo-ontology category. This results in entries with standardized locations.

Besides audio transcription and location matching, we used `multilingual-e5-large` model [4] to categorize requests needs and interventions supplies into classes: *shelter, food/water, pharmaceuticals, clothes, covers, rescue, medical assistance* or *other*

3 Evaluation and results

Automatic Speech Recognition: `wav2vec2-large-xlsr-moroccan` achieved an average WER (word error rate) of 53.54% on manually labeled samples from our data containing 306 words.

Matching textual locations with reference douars: We evaluated our algorithm on 66 manually labeled samples from interventions and requests data. Percentage of correct matches over each geo-ontology category: 'province': 22%, 'cercle': 25% 'commune': 52%, 'douar (fr)': 60%, 'douar (ar)':70%

Needs and supplies classification: On 40 manually labeled requests/interventions, we achieved the following accuracies: clothes: 87.5%, shelter: 76.9%, food/water: 76.9%, pharmaceuticals: 70%, covers: 50%, rescue: 50%, medical assistance: 50%, other: 53.8%.

Data analysis and visualization: We performed statistical analysis on the refined dataset. First, we consolidated requests and interventions at both the douar and regional levels, enabling a nuanced understanding of region-specific needs. We also tracked how these statistics evolve over time and evaluate satisfaction levels regarding these requirements in individual douars. This analysis provides valuable insights into the effectiveness of aid efforts. (e.g Globally, in a range of one week the major supply provided to douars evolved from *food/water* to *shelter*). Finally, we showcased these findings using charts and an interactive map on our online platform.

4   Conclusion and future work

This paper presents our efforts for a Moroccan data-driven platform, which is designed to facilitate the coordination of aid for areas in need following the earthquake by collecting data, preprocessing it to get a clean dataset, and visualizing results and statistics on our online platform. Future work aims to involve social media analytics to gather additional updates on the situation of the affected areas and to develop methods for assessing damages through satellite imagery data.

References

[1] Listes localités gps. [Online]. Available: https://files.lebrief.ma/uploads/2023/09/11170540/listes-localites.xlsx

[2] Osint seisme. [Online]. Available: https://www.google.com/maps/d/u/0/viewer?mid=1kZrQiHyjpQDz9Pwhn5oxueTpQXYNlB8&femb=1&ll=0%2C0&z=9

[3] A. Baevski, Y. Zhou, A. Mohamed, and M. Auli, "wav2vec 2.0: A framework for self-supervised learning of speech representations," *Neural Information Processing Systems*, vol. 33, pp. 12 449–12 460, 2020.

[4] L. Wang, N. Yang, X. Huang, B. Jiao, L. Yang, D. Jiang, R. Majumder, and F. Wei, "Text embeddings by weakly-supervised contrastive pre-training," *arXiv:2212.03533*, 2022.