
MACHINE LEARNING APPLICATIONS THAT CAN HELP PASTORAL COMMUNITIES IN NORTHERN KENYA AND ELSEWHERE ADAPT TO CLIMATE CHANGE

By Jefferson Sankara
Senior Data Scientist and NLP Researcher

Introduction

The paper is a proposal;

1. to use of Machine Learning(ML) techniques such as Active Learning(AL) and Transfer Learning(TL) to translate climate information and adaptation techniques from major Western and Asian languages to thousands of low resource languages in the developing world.

OBJECTIVES

The objectives of the proposal are;

1. to demonstrate that pastoralist communities in Northern Kenya and other parts of the world contribute to and are affected by Climate change,
2. to examine the gap that exists in climate information and adaptation techniques due to language barrier
3. to show that ML techniques can be used to translate information from major western and Asian Languages to low resource languages like Turkana and others that are used in Northern Kenya
4. to explore the impact Machine Learning applications to bridge this gap would have on these communities

LIVESTOCK CONTRIBUTION TOWARDS CLIMATE CHANGE

1. Even though pastoralist communities in northern Kenya have undergone evolution in terms of their economic activities for instance by embracing agriculture, they still keep large numbers of livestock.
2. These animals contribute towards global warming by releasing greenhouse gases like methane.
3. It is estimated that the livestock sector contributes 14.5% of global greenhouse gas (GHG) emissions that contribute to global warming
4. Climate change has affected agricultural sectors like farming, fisheries and other components that constitute the livelihood of rural populations in developing countries like Kenya.

GAP IN CLIMATE INFORMATION DUE TO LANGUAGE BARRIER

- Studies conducted in Kenya have revealed that climate change terminology and concepts are poorly understood and do not have standard translations in Swahili, Luo and other local Kenyan languages.
- This as a result creates a knowledge gap that prevents many people from understanding the phenomenon and having a voice on the issue.
- Their response to climate change or its adaptation techniques are therefore hampered
- Additionally, low literacy levels and high levels of poverty in these communities make it impossible to access climate information that might easily be accessible even in its simplest form through Television.
- The language barrier would make it hard for most analysis and findings published and accessible through the internet in major western and Asian languages are to be easily understood by these communities even if they had access to the internet.
- Since the effects of global warming are real and evident in their daily lives, there is an urgent need to bridge the gap in climate information.

APPLICATIONS OF ML TO LOW RESOURCE LANGUAGES ACTIVE LEARNING (AL)

1. Low resource languages lack elaborate monolingual or parallel corpus that can be used to build NLP applications like Machine Translations models
2. Generation of a corpus is an intensive and expensive human effort.
3. AL is a technique of selecting the most informative examples from unlabeled data in order to reduce human effort as an oracle seeks to annotate the data.
4. AL is relevant for translating these African Languages particularly because of the dearth of corpus.
5. Improvements on AL have led to;
 - a. Possibility of getting the highest translation quality of the retrained model.
 - b. Introduction of weighting schemes to allow for the sorting of sentences based on the frequency of unseen n-grams.
 - c. The ability to select smaller training corpora that proved the need for much less training data with competitive performance

APPLICATIONS OF ML TO LOW RESOURCE LANGUAGES

TRANSFER LEARNING (TL)

1. TL is the technique where a model is first trained on a high-resource language pair, then the child model's parameter values are copied from the parent's and are fine-tuned on its low-resource data. E.g English-based model applied to Turkana.
2. There have been improvements on TL where the idea was to share the parent and child's source vocabularies, so that when source word embeddings are transferred, a word that appears in both vocabularies keeps its embedding.
3. There is evidence that a combination of TL with NMT witnessed improved performance on a low-resource language pair by exploiting its lexical similarity with another related low-resource language.
4. There was consistent improvement in two Turkic languages.
5. This would be a technique that could be adopted to Translate closely related languages of Turkana and Njemps in Northern Kenya.

POTENTIAL IMPACT OF TRANSLATING KNOWLEDGE INTO LOCAL LANGUAGES

1. Studies have shown that
 - Access to climate information in a language communities can understand would help them assess the magnitude of the climate change challenge
 - As a result of the above, they would embrace adaptation techniques and be inspired to act and empower them to plant food types that can survive in prolonged dry conditions.
2. This would lead to food security in pastoralist communities like those in Northern Kenya.
3. It is estimated that approximately 95% of the household income would emanate from livestock keeping and agricultural activities.
4. These adaptation techniques would help them through extended dry seasons and during heavy rain too by learning
 - a. to prepare enough food and preserve it
 - b. Plan accordingly in order to deal with floods by storing water and putting measures in place to prevent deaths and loss of crop yield

POTENTIAL IMPACT CONTD.

4. Adaptation initiatives that build on local knowledge most of the time communicated by locals in their local languages and integrate scientific findings have a higher chance of leading to sustained and effective change.
5. Availability of this information in local languages would further make easier the collaboration between scientists, policy makers and other experts in an attempt to understand and appreciate the challenges locals are facing.
6. As a result, they would be able to refine their approaches and the solutions they would be working on.
7. Ability to automate translation of information to local languages and vice versa would also empower people who work with these vulnerable groups, such as extension officers, empowerment programmes, local and national governments, and Nongovernmental Organisations(NGO).

CONCLUSION

1. These communities;
 - a. Contribute towards climate change
 - b. Are affected by climate change
 - c. are vulnerable to the effects of climate change due to their direct dependency on the natural resources being affected.
2. There is a knowledge gap that needs to be filled urgently due to language barrier
3. Invitation to build models leveraging on existing ML techniques to help with the translation