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

Jefferson Sankara
Senior Data Scientist
jefferson.sankara@gmail

ABSTRACT

I propose the use of Machine Learning 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. Studies have shown that access to information can help people assess the magnitude of the climate change challenge, possible options and those feasible within the relevant context (Nyahunda & Tirivangasi, 2019) I endeavour to demonstrate that if this information was available in a language the locals can understand, it would result in local empowerment and as a result inspire action.

1 INTRODUCTION

I propose the use of Machine Learning(ML) techniques such as Active Learning(AL) and Transfer Learning(TL) with improvements Liu et al. (2018) to translate climate information and adaptation techniques from major Western and Asian languages to thousands of low resource languages in the developing world. Studies have shown that access to information can help people assess the magnitude of the climate change challenge, possible options and those feasible within the relevant context Nyahunda & Tirivangasi (2019). I endeavour to demonstrate that if this information was available in a language the locals can understand, it would result in local empowerment and as a result inspire action. Without an automated approach, building evaluation sets on low-resource languages is both expensive and time-consuming because the pool of professional translators is limited, as there are few fluent bilingual speakers for these languages Guzmán et al. (2019), Ambati (2012). This solution would be very impactful considering that these languages are spoken by a large fraction of the world population Guzmán et al. (2019). Global warming, which occurs when carbon dioxide (CO_2) and other air pollutants and greenhouse gases collect in the atmosphere and absorb sunlight and solar radiation that have bounced off the earth's surface Pappas (2017), is one of the greatest challenges facing humanity today Rolnick et al. (2019). It continues to affect natural habitats and biodiversity with disruption on growing season, phenology, primary production, and species distributions and diversity (Bellard et al., 2012). The proposed ML applications will be pivotal to the communities in Northern Kenya that continue to be affected by Climate change Huho (2015) through prolonged droughts that have threatened their livelihoods and as result caused conflict van Baalen & Mobjörk (2017) Schilling et al. (2012) since these communities scramble for the increasingly less available natural resources like water and pasture for their animals, by availing climate information in a language locally understood and techniques to adapt to the effects of the phenomenon. The effect is well pronounced by the fact that about a third of the Kenyan population depends on pastoralism for food and income security, KNBS (2010).

2 OBJECTIVES

The objectives of the proposal are;

- to demonstrate that communities in Northern Kenya contribute to and are affected by Climate change,
- to examine the gap that exists in climate information and techniques to adopt to the phenomenon due to language barrier and possible impacts to local empowerment
- to show that ML techniques can be used to translate information from English to low resource languages like those used in Northern Kenya
- to explore the impact ML applications to bridge this gap would have on the communities

3 Livestock Contribution towards Climate Change

In as much as pastoralism communities in northern Kenya have undergone evolution they largely continue to keep large numbers of livestock as per Kaye-Zwiebel & King (2014) and Hauck & Rubenstein (2017) that contribute towards global warming (Hauck & Rubenstein, 2017). It is estimated that the livestock sector contributes 14.5% of global greenhouse gas (GHG) emissions, driving further climate change (Rojas-Downing et al., 2017). The risks of this phenomenon are apparent in agriculture, fisheries and many other components that constitute the livelihood of rural populations in developing countries (Adger & Kelly, 1999). According to Hurst et al. (2005), the livestock sector contributes to the livelihoods of one billion of the poorest population in the world and employs close to 1.1 billion people out of which 1 billion people live in the areas like Northern Kenya (Rojas-Downing et al., 2017).

4 Gap in Climate Information due to Language barrier

A study by otieno & Pauke (2010) has shown that climate change terminology is poorly understood and does not have standard translations in Swahili, Luo and other local Kenyan languages. This prevents many people from having a voice on the issue or even fully grasping it. Further studies by otieno & Pauke (2010) have shown that African citizens' response to climate change is hampered by a fundamental shortage of relevant, useful information for African audiences. Additionally, due to low literacy levels as in Duba et al. (2001) and high levels of poverty in these communities, climate information is largely not easily accessible even in its simplest form through TV (Huho, 2015). Most analysis and findings published and accessible through the internet are in English and a few major western and Asian languages according to the Bank (2014) and that makes the information not easily understood by these communities even if there was wide coverage of the internet. Finally, the indigenous knowledge systems that these communities relied upon to understand weather patterns like other parts of Africa are largely unpredictable due to global warming (Nyahunda & Tirivangasi, 2019). There is ample evidence that reveals that Africa and other developing countries face more challenges from climate change because of poor adaptation mechanisms in place (Enete & Achike, 2008); (Jagtap, 2007); (Nwafor, 2007). As a result of all the effects of global warming Nyahunda & Tirivangasi (2019), there is an urgent need to bridge the gap in climate information.

5 Applications of ML to Low resource Languages

Low resource languages lack elaborate monolingual or parallel corpus that can be used to build NLP applications(Jiatao Gu, 2018). Generation of a corpus for low resource languages is an intensive human effort that would require the availability of fluent bilinguals or expert translators (Ambati, 2012), (Settles & Craven, 2008). ML techniques like AL with improvements such as Neural Machine Translation(NMT) and TL have been proven as applications to Machine Translations for Low resource scenarios. Ambati (2012), Liu et al. (2018), Nguyen & Chiang (2017) and (Ambati, 2012) defined AL as 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. AL is extremely relevant for translating African Languages particularly because of the dearth of labelled data and resources on the said languages (Garrette et al., 2013), (Fang et al., 2017). Latest improvements on AL Liu et al. (2018) have demonstrated possibility of getting the highest improvements in the translation quality of the retrained model. Research by Zoph et al. (2016) has demonstrated that Neural Machine Translation (NMT) can be used for low-resource languages as well, by introducing more local dependencies and using word alignments to learn sentence reordering during translation. Examples of

NMT as per Zoph et al. (2016) have demonstrated innovative ways such as training a model using a high-resource language pair, then using it to initialize a child model which is further trained on a low-resource language pair. Thompson et al. (1999) demonstrated that its applications to Natural Language Processing tasks has minimized expenses for annotating data. Further advancements of AL like in Hildebrand et al. (2005) introduced weighting schemes to allow for the sorting of sentences based on the frequency of unseen n-grams. The output of this was ability select smaller training corpora that proved the need for much less training data with competitive performance compared to baseline systems using all available training data. In Gangadharaiah et al. (2009) a pool-based strategy was used to maximize a measure of expected future improvement, to sample instances from large parallel corpora.

As per Zoph et al. (2016), 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. There have been improvements on TL that Zoph et al. (2016) worked on by Nguyen & Chiang (2017) where the idea was to share the parent and child's source vocabularies, so that when source word embedding are transferred, a word that appears in both vocabularies keeps its embedding. Nguyen & Chiang (2017) demonstrated that by combining TL with BPE, NMT witnessed improved performance on a low-resource language pair by exploiting its lexical similarity with another related low-resource language. there was consistent improvement in two Turkic languages. This would be a technique that could be adopted to Translated closely related languages of Turkana and Njemps in Northern Kenya. TL has also exhibited successful results in other areas like Automated Speech recognition (Julius Kunze, 2017).

6 Potential Impact of Translating Knowledge into local languages

Studies have shown that access to information can help people assess the magnitude of the climate change challenge, possible options and those feasible within the relevant context (Nyahunda & Tirivangasi, 2019). Access to adaptation techniques and climate information in local languages would inspire action within these communities and empower them to plant plants and food types that can survive in prolonged dry conditions. This would lead to food security in Northern Kenya where approximately 95% of the household income is from agricultural activities (Huho, 2015). Further findings Heath (2019) have demonstrated that communities adapt and use techniques to help them through extended dry seasons and during heavy rain too when they have knowledge and awareness of climate change. 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 adaptation (Gina Ziervogel & Scodanibbio, 2016). This would additionally make it easy for collaboration between scientists, policy makers and other experts to understand and appreciate the challenges locals are facing and as a result refine their approaches and the solutions they would be working on. 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 like this UNFCCC (2020), local and national governments, and Nongovernmental Organisations(NGO) practitioners.

7 Conclusion

Climate change is already wreaking havoc in arid and semi arid areas of Northern Kenya. These communities are vulnerable due to their direct dependency on the natural resources being affected. I hope that our proposal has demonstrated a gap that exist in climate change adaptation information for Pastoral communities such as those in Northern Kenya and other parts of Africa that can be addressed through the applications of ML. As AL, TL and NMT continue to improve as techniques for Machine Translation, there is urgent need to apply them to translate key climate change adaptation information from Western and major Asian languages to local Northern Kenya languages and equip these communities with requisite knowledge to adapt as their normal order of natural environments continue to be disrupted by global warming.

References

W Neil Adger and P Mick Kelly. Social vulnerability to climate change and the architecture of entitlements. *Mitigation and adaptation strategies for global change*, 4(3-4):253–266, 1999.

Vamshi Ambati. *Active Learning and Crowdsourcing for Machine Translation in Low Resource Scenarios*. PhD thesis, School of Computer Science Carnegie Mellon University, 2012.

World Bank. Internet access, yes, but in my mother language! 07 2014. URL `https://www.worldbank.org/en/news/feature/2014/07/03/internet-access-yes-but-in-my-mother-language`.

Céline Bellard, Cleo Bertelsmeier, Paul Leadley, Wilfried Thuiller, and Franck Courchamp. Impacts of climate change on the future of biodiversity. *Ecology Letters*, 15(4):365–377, 2012. doi: 10.1111/j.1461-0248.2011.01736.x. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1461-0248.2011.01736.x`.

Huka Duba, Ingrid Mur-Veeman, and Arno Raak. Pastoralist health care in kenya. *International journal of integrated care*, 1:e13, 02 2001. doi: 10.5334/ijic.21.

Anselm A Enete and Anthonia I Achike. Urban agriculture and urban food insecurity/poverty in nigeria: The case of ohafia, south-east nigeria. *Outlook on agriculture*, 37(2):131–134, 2008.

Meng Fang, Yuan Li, and Trevor Cohn. Learning how to active learn: A deep reinforcement learning approach. *CoRR*, abs/1708.02383, 2017. URL `http://arxiv.org/abs/1708.02383`.

Rashmi Gangadharaiah, Ralf D. Brown, and Jaime G. Carbonell. Active Learning in Example-Based Machine Translation. *NEALT Proceedings Series*, 4, 1 2009. doi: 10.1184/R1/6620912.v1. URL `https://kilthub.cmu.edu/articles/Active_Learning_in_Example-Based_Machine_Translation/6620912`.

Dan Garrette, Jason Mielens, and Jason Baldridge. Real-world semi-supervised learning of POS-taggers for low-resource languages. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 583–592, Sofia, Bulgaria, August 2013. Association for Computational Linguistics. URL `https://www.aclweb.org/anthology/P13-1057`.

Kate Kloppers Gina Ziervogel and Lucia Scodanibbio. Lessons from semi-arid regions on how to adapt to climate change. *Academic rigor, journalistic flair*, 2016.

Francisco Guzmán, Peng-Jen Chen, Myle Ott, Juan Pino, Guillaume Lample, Philipp Koehn, Vishrav Chaudhary, and Marc'Aurelio Ranzato. Two new evaluation datasets for low-resource machine translation: Nepali-english and sinhala-english. *CoRR*, abs/1902.01382, 2019. URL `http://arxiv.org/abs/1902.01382`.

Stephanie Hauck and Daniel I. Rubenstein. Pastoralist societies in flux: A conceptual framework analysis of herding and land use among the mukugodo maasai of kenya. *Pastoralism*, 7:1–30, 2017.

Hilary Heath. Empowered communities adapt to climate change. *Climate Justice Resilience Fund*, 2019.

Almut Hildebrand, Matthias Eck, Stephan Vogel, and Alex Waibel. Adaptation of the translation model for statistical machine translation based on information retrieval. *Proceedings of EAMT*, 01 2005.

Julius Huho. Climate change knowledge gap in education system in kenya. *International Journal of Innovation and Research in Education Sciences*, 2:2349–5219, 06 2015.

P Hurst, P Termine, and M Karl. Agricultural workers and their contribution to sustainable agriculture and rural development. fao, rome, 2005.

S Jagtap. Managing vulnerability to extreme weather and climate events: Implications for agriculture and food security in africa. In *Proceedings of the international conference on climate change and economic sustainability held at Nnamdi Azikiwe University, Enugu, Nigeria*, pp. 12–14, 2007.

Jacob Devlin Victor O.K. Li Jiatao Gu, Hany Hassan. Universal neural machine translation for extremely low resource languages. *Association for Computational Linguistics*, pp. s 344–354, 2018.

Ilia Kurenkov Andreas Krug Jens Johannsmeier Sebastian Stober Julius Kunze, Louis Kirsch. Transfer learning for speech recognition on a budget. *Computer Science, Mathematics*, 1, 2017.

Eva Kaye-Zwiebel and Elizabeth King. Kenyan pastoralist societies in transition: Varying perceptions of the value of ecosystem services. *Ecology and Society*, 19, 09 2014. doi: 10.5751/es-06753-190317.

KNBS. *The 2009 Kenya population and housing census*, volume 1. Kenya National Bureau of Statistics, 2010.

Ming Liu, Wray Buntine, and Gholamreza Haffari. Learning to actively learn neural machine translation. In *Proceedings of the 22nd Conference on Computational Natural Language Learning*, pp. 334–344, Brussels, Belgium, October 2018. Association for Computational Linguistics. doi: 10.18653/v1/K18-1033. URL https://www.aclweb.org/anthology/K18-1033.

Toan Q. Nguyen and David Chiang. Transfer learning across low-resource, related languages for neural machine translation. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pp. 296–301, Taipei, Taiwan, November 2017. Asian Federation of Natural Language Processing. URL https://www.aclweb.org/anthology/I17-2050.

JC Nwafor. Global climate change: The driver of multiple causes of flood intensity in sub-saharan africa. In *International Conference on Climate Change and Economic Sustainability held at Nnamdi Azikiwe University, Enugu, Nigeria*, pp. 12–14, 2007.

Louis Nyahunda and Happy M. Tirivangasi. Challenges faced by rural people in mitigating the effects of climate change in the mazungunye communal lands, zimbabwe. *Journals Of Disaster Risk Studies*, 11, 2019.

Sam otieno and Ed Pauke. The public understanding of climate change. *Kenya Talks Climate*, 2010.

Stephanie Pappas. What is global warming? August 2017. URL https://www.livescience.com/37003-global-warming.html.

M. Melissa Rojas-Downing, A. Pouyan Nejadhashemi, Timothy Harrigan, and Sean A. Woznicki. Climate change and livestock: Impacts, adaptation, and mitigation. *Climate Risk Management*, 16:145 – 163, 2017. ISSN 2212-0963. doi: https://doi.org/10.1016/j.crm.2017.02.001. URL http://www.sciencedirect.com/science/article/pii/S221209631730027X.

David Rolnick, Priya L. Donti, Lynn H. Kaack, Kelly Kochanski, Alexandre Lacoste, Kris Sankaran, Andrew Slavin Ross, Nikola Milojevic-Dupont, Natasha Jaques, Anna Waldman-Brown, Alexandra Luccioni, Tegan Maharaj, Evan D. Sherwin, S. Karthik Mukkavilli, Konrad P. Körding, Carla Gomes, Andrew Y. Ng, Demis Hassabis, John C. Platt, Felix Creutzig, Jennifer Chayes, and Yoshua Bengio. Tackling climate change with machine learning. *CoRR*, abs/1906.05433, 2019. URL http://arxiv.org/abs/1906.05433.

Janpeter Schilling, Francis Opiyo, and Jürgen Scheffran. Raiding pastoral livelihoods: Motives and effects of violent conflict in north-western kenya. *Pastoralism*, 2, 01 2012. doi: 10.1186/2041-7136-2-25.

Burr Settles and Mark Craven. An analysis of active learning strategies for sequence labeling tasks. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pp. 1070–1079, Honolulu, Hawaii, October 2008. Association for Computational Linguistics. URL https://www.aclweb.org/anthology/D08-1112.

Cynthia A. Thompson, Mary Elaine Califf, and Raymond J. Mooney. Active learning for natural language parsing and information extraction. In *Proceedings of the Sixteenth International Conference on Machine Learning*, ICML '99, pp. 406–414, San Francisco, CA, USA, 1999. Morgan Kaufmann Publishers Inc. ISBN 1558606122.

UNFCCC. Empowering climate resilient women through community based adaptation – the adaptation learning programme for africa (alp). *Climate Change*, 18, 2020.

Sebastian van Baalen and Malin Mobjörk. Climate Change and Violent Conflict in East Africa: Integrating Qualitative and Quantitative Research to Probe the Mechanisms. *International Studies Review*, 20(4):547–575, 11 2017. ISSN 1521-9488. doi: 10.1093/isr/vix043. URL `https://doi.org/10.1093/isr/vix043`.

Barret Zoph, Deniz Yuret, Jonathan May, and Kevin Knight. Transfer learning for low-resource neural machine translation. *CoRR*, abs/1604.02201, 2016. URL `http://arxiv.org/abs/1604.02201`.

style=alphabetic