Applying Natural Language Processing Methods Towards Wildlife Conservation Efforts

Julia Guo
Columbia University
jzg2110@columbia.edu

May 10, 2022

1 Abstract

From 1800 to the modern day, the global human population has exponentially increased from 1 billion in 1800 to 7.9 billion [9]. As more human beings inhabit the Earth, civilizational needs for land, food, water, and other resources increase as well. Consequently, forests are cut down and water reservoirs are depleted, causing great harm to animal populations and biodiversity in the wild [5]. Biodiversity is important to humankind for many reasons, including food, resources for pharmaceuticals [11], cultural significance, etc. As such, wildlife conservation is an important issue in the modern age. Scientists have contributed many studies in an ongoing effort to identify tangible actions that help and hurt wildlife. However, there are many such studies and actions, and a shortage of human labor to disperse this knowledge to the public. As such, Natural Language Processing (NLP) methods that automatically derive understanding from text, such as summarization or classification, can be helpful to aid in this effort. In this paper, we describe two ways NLP can help with wildlife conservation efforts, and train a large language model to classify actions as helpful, hurtful, or neutral in wildlife conservation.

2 Introduction

Wildlife conservation is an important issue for humankind for many reasons, especially as the preservation of both flora and fauna directly sustains fundamental needs of human civilization. In regards to conservation of plant life, deforestation is associated with nutrient depletion in soil as well as erosion, which can lead to poorer crop yield, landslides, and flooding [12]. In regards to conservation of animal life, ecosystems are built up of complex food networks, where animal species each possess important and unique roles in the food chain. Decreases in the population of one keystone species can generate a domino effect of population collapse across the ecosystem [6], and flourishing of another keystone species may generate abundance across the ecosystem as well [4]. Furthermore, plants and animals involved in wildlife ecosystems satisfy critical needs in human agriculture and medicine, and provide other benefits such as possessing aesthetic beauty that inspires the arts, or preserving traditional cultures [7].

Natural Language Processing (NLP) is the subfield of artificial intelligence concerned with generating and understanding human language. Example applications of NLP include document summarization, essay generation, and social media post classification. In the current paradigm, machine learning models are trained on large textual corpora and then tested on unseen data to complete these textual understanding and generation tasks. Regarding specific model architectures, large language models built on attention-based Transformer models are popular these days, such as BERT [3] or GPT-3 [2].

3 Background

Conservation Evidence is a resource provided by scientists at the University of Cambridge [1]. They collect studies related to wildlife conservation, summarize these studies, and extract actions performed in the studies as well as the consequences of those actions. They have summarized 8161 studies, and collected 3510 actions related to wildlife conservation, making it a rich data resource for NLP model training.

BERT is a transformer-based [10] machine learning model developed by research scientists at Google in 2018. It is considered among the state of the art for encoder models (classification as opposed to generation) in NLP currently.

4 Methods

4.1 Data Collection

We scraped the 3510 actions and effectiveness ratings from the Conservation Evidence website using the Beautiful Soup Python package [8]. The Conservation Evidence team rated action effectiveness using multiple textual categories, which we bucketed into negative, neutral, and positive in the following way:

Action Category Mapping	
Likely to be ineffective or harmful	Negative
Unlikely to be beneficial	Negative
Unknown effectiveness (limited evidence)	Neutral
No evidence found (no assessment)	Neutral
Evidence not assessed	Neutral
Awaiting assessment	Neutral
Trade-off between benefit and harms	Neutral
Likely to be beneficial	Positive
Beneficial	Positive

The data is notably very imbalanced, having 2825 neutral, 596 positive, and 89 negative actions. As such, we divide this data using a 90-10 train-test split so that the model has more training data does not simply classify everything as neutral.

4.2 Model Training

We train BERT on the data (see src/ folder) for 20 epochs, with a batch size of 32, learning rate of 2e-5, Adam epsilon of 1e-8, weight decay of 0.0, and max gradient norm of 1.0.

5 Results

After 20 epochs, the results are as follows:

Results	
Test loss	1.4388
Test accuracy	**0.7236**
Negative Precision	0.2857
Negative Recall	0.2000
Negative F1	**0.2353**
Neutral Precision	0.8264
Neutral Recall	0.8264
Neutral F1	**0.8264**
Positive Precision	0.4177
Positive Recall	0.4342
Positive F1	**0.4258**

Given the small amount of data, the F1 scores for negative and positive categories are considerably high, which is encouraging for future work. The overall accuracy and neutral F1 scores are also very good.

6 Conclusions and Future Work

Through training the BERT classification model on the Conservation Evidence data, we obtain proof of evidence that by looking at actions that are known to be helpful or hurtful to the environment, a model can learn to generalize to new actions. In the future, it would be beneficial to also apply the latest summarization models to wildlife conservation studies in order to create even more conservation action data. As a whole, this initial study provides promising groundwork in a ML $\times$ Climate application area where little to no prior work has been conducted.

References

[1] *About Conservation Evidence*. URL: https://www.conservationevidence.com/content/page/24.

[2] Tom Brown et al. "Language Models are Few-Shot Learners". In: *Advances in Neural Information Processing Systems*. Ed. by H. Larochelle et al. Vol. 33. Curran Associates, Inc., 2020, pp. 1877–1901. URL: https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf.

[3] Jacob Devlin et al. "Bert: Pre-training of deep bidirectional transformers for language understanding". In: *arXiv preprint arXiv:1810.04805* (2018).

[4] Julie A Jedlicka, Russell Greenberg, and Deborah K Letourneau. "Avian conservation practices strengthen ecosystem services in California vineyards". In: *PloS one* 6.11 (2011), e27347.

[5] William F. Laurance, Heraldo L. Vasconcelos, and Thomas E. Lovejoy. "Forest loss and fragmentation in the Amazon: implications for wildlife conservation". In: *Oryx* 34.1 (2000), pp. 39–45. DOI: 10.1046/j.1365-3008.2000.00094.x.

[6] L Scott Mills, Michael E Soulé, and Daniel F Doak. "The keystone-species concept in ecology and conservation". In: *BioScience* 43.4 (1993), pp. 219–224.

[7] ShengJi Pei et al. "Ethnic cultural diversity and nature conservation in Yunnan." In: *Acta Botanica Yunnanica* SUPPL. 15 (2004), pp. 1–11.

[8] Leonard Richardson. "Beautiful soup documentation". In: *Dosegljivo: https://www. crummy.com/software/BeautifulSoup/bs4/doc/.[Dostopano: 7. 7. 2018]* (2007).

[9] Max Roser, Hannah Ritchie, and Esteban Ortiz-Ospina. *World population growth*. May 2013. URL: https://ourworldindata.org/world-population-growth.

[10] Ashish Vaswani et al. "Attention is all you need". In: *Advances in neural information processing systems* 30 (2017).

[11] R. N. Young. "Importance of biodiversity to the modern pharmaceutical industry". In: *Pure and Applied Chemistry* 71.9 (1999), pp. 1655–1661. DOI: doi:10.1351/pac199971091655. URL: https://doi.org/10.1351/pac199971091655.

[12] Fenli Zheng et al. "Effects of erosion patterns on nutrient loss following deforestation on the Loess Plateau of China". In: *Agriculture, Ecosystems Environment* 108.1 (2005), pp. 85–97. ISSN: 0167-8809. DOI: https://doi.org/10.1016/j.agee.2004.12.009. URL: https://www.sciencedirect.com/science/article/pii/S0167880905000095.