Proof of Concept

Yi (Michelle) Wang, Echo Liu, Yueqiao Dai, Mianchun Lu,
Yirou Ge, Marshall Zhao, Yuefu Wu, Rachel Li

Introduction

In order to achieve wide-scale implementation of climate targets, the values and priorities of multi-sectoral stakeholders must be aligned and integrated with other development agenda. Although misalignment across scales of policy and between stakeholders are well-known barriers to implementing restoration, fast-paced policy making in multi-stakeholder environments complicates the monitoring and the analysis of governance and policy. As the World Resources Institute (WRI) aims to facilitate sustainable development of environment, our team plans to utilize NLP and machine learning to monitor worldwide restoration initiatives and detect incentives and disincentives across diverse policy documents.

Frameworks

The current work of WRI focuses on converting the policy documents in plain text into datasets with labels indicating the financial incentives and disincentives in each sentence, which helps them gain a deeper understanding about the worldwide land restoration. This workflow includes two key methods: Label Model and Neural Networks. In the Label Model method, the manually labeled gold standard dataset was used to train the Snorkel Label Model, resulting in a trained model and a new dataset with labels generated by Snorkel. Then the trained model was used to predict labels for unlabeled samples to generate a noisy dataset. After that, the workflow enters the Neural Networks method, where the sentences in both datasets are encoded by roBERTa and the gold standard dataset was first split into training and validation data to train a shallow recurrent neural network. Then the noisy dataset was used as training data to train another shallow RNN as well. Therefore, the current workflow eventually generated two trained neural network models and one of them serves as the baseline model.
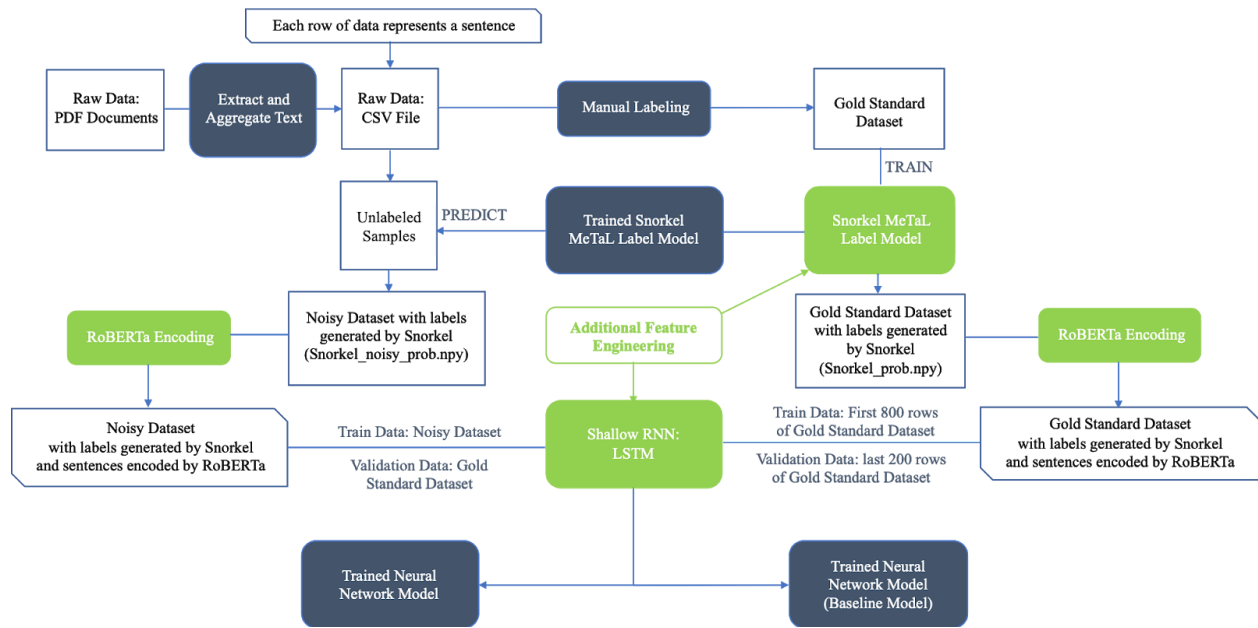
Current Workflow of WRI

In order to optimize the current workflow, our team will be focusing on Label Model optimization and Neural Networks optimization. In the Label Model optimization, our team is planning to compare the performance of Snorkel Model and BabbleLabble. In the Neural Networks optimization, our team is expecting to build a deeper and more complicated neural network and fine tune roBERTa encoding as well. Meanwhile, our team will also utilize additional feature engineering for both models to further improve their performance.

Benefits

In order to attain worldwide execution of land restoration, the priorities and responsibilities of international environmental agencies must be coordinated and aligned. Unfortunately, there are some remaining huge gaps between the policies and each country's audience, and the rapid legislative development worldwide further complicates the situation. Thus, how to popularize the policy content to the audience, and support the implementation of the policy becomes one of the

biggest business problems that the restoration project currently face. WRI seeks to utilize technical tools including NLP and machine learning to resolve the challenge and how to build a model framework for accurately analyzing the policy documents is one of the most essential problems for the organization right now. Therefore, the two optimization tasks that our team is focusing on will bring benefits to WRI in the following three ways.

First, by optimizing the key methods in the current workflow, our team is expecting to generate a more accurate model which can classify the policy documents from different countries based on the financial incentives and disincentives it provides for land restoration. This improved model will help WRI understand whether each policy document supports land restoration in a more efficient way, leading to more effective and positive impacts on the farmers community.

Second, the classification model can be applied in a more global scale in the future as it can be used to analyze the policy documents of different countries. Furthermore, by modifying the current language modeling and encoding method, WRI can apply this model to classify documents in other languages as well.

Third, more aspects of each financing mechanism can also be identified utilizing the classification model (neural networks) through transfer learning. For example, the model can be adjusted to identify additional characteristics of the financing policy including type, audience, financier, stipulations, and amount. In this way, the organization will be able to understand and extract the information of the financial mechanisms in policy documents automatically, which can not only help popularize the policy content to the general audience more efficiently but also enhance the implementation of each policy as the audience gains a better and deeper understanding of it.

Methodologies Illustration

Labeling Model Optimization

In the current Snorkel model that WRI trained, they listed some important words that express different emotions, and considered them as well as the relative position of these words as main features to write the functions and train the model. To enhance the performance of the Snorkel, our team thinks that more features, including N-grams, Part-of-Speech tags generated by Hidden Markov Model, LIWC Features, Named Entity Features, Word Frequency and similarity coefficient, should also be added as features to the Snorkel model.

In order to prove that adding more features can enhance the performance of Snorkel model, our team plans to select some samples with different labels from our training data firstly and do a quick calculation. If there are significant differences among all the three categories in terms of the features that our team selects, it can be concluded that the features have high possibility to affect the labeling results, and should be added to the model.

For example, our team selected three "positive" sentences, three "negative" sentences, and three "neutral" sentences from the training data, and calculated the similarity matrix. The hypothesis our team set was that the similarity coefficient in the same class should be larger than the similarity coefficient between different classes. Our team applied TFIDF (term frequency–inverse document frequency) with cosine similarity, as well as Google Word2Vec model with cosine similarity to do the calculation respectively, and their results are shown as follows. Based on these results, it can be concluded that the average similarity coefficient between two classes is smaller than the average similarity coefficient in each one class, which means that the similarity coefficient differs in different classes and it can be used as a feature to

help with the labeling model. Then, some "positive" sentences are selected as well as some "negative" sentences as baselines, and calculate their similarity coefficients with other sentences that we want to label.

TFIDF	Positive 0	Positive 1	Positive 2	Negative 0	Negative 1	Negative 2
Positive 0		0.1487	0.1165	0.0329	0.1227	0.0127
Positive 1			0.1634	0.0598	0.0382	0.0690
Positive 2				0.0562	0.1027	0.1575
Negative 0					0.1945	0.0183
Negative 1						0.0390
Negative 2						

average similarity coefficient
positive class   0.1429
negative class   0.0839
between two classes   0.0724

word2vec	Positive 0	Positive 1	Positive 2	Negative 0	Negative 1	Negative 2
Positive 0		0.7614	0.7190	0.7448	0.7206	0.7467
Positive 1			0.8236	0.7213	0.7464	0.7032
Positive 2				0.7910	0.8155	0.7449
Negative 0					0.8788	0.8313
Negative 1						0.7940
Negative 2						

average similarity coefficient
positive class   0.7680
negative class   0.8347
between two classes   0.7483

For the other features that were mentioned before, our team should also test them in a similar way, and use all the features that has been proved to be useful to train our final Snorkel model. After that, our team will compare the new results with initial results that WRI provided to see if our feature engineering successfully help with the Labeling Model Optimization.

In addition to Snorkel, BabbleLabble is another model that our team can use for labeling. The working function of BabbleLabble approach is to use semantic parser to convert natural language explanations into labeling functions that are "accurate enough" for using with data programming. In other words, BabbleLabble converts natural language (babble) into labels. In order to use BabbleLabble approach effectively, our team needs to first set up explanations. For example, we would write an explanation like "any policy that contain support is an incentive policy" and put this explanation into BabbleLabble. Then BabbleLabble would convert this explanation into codes, which could be further applied in machine learning procedures.

Currently, our team has successfully ran BabbleLabble following the tutorial on GitHub, so the next step is to try it to generate noisy label. Our team is at the stage of adding explanations to BabbleLabble, where meaningful and unbiased explanations need to be selected in order to generate a better and more accurate labeling function. Then, our team would like to compare the overall performance of BabbleLabble and Snorkel, select the one with higher accuracy and then further improve its performance for an even better result.

Neural Networks Optimization

Building a neural network model is the key approach for our team to predict and classify the policy documents based on their financial incentives and disincentives regarding land restoration.

- Develop Deeper and More Complicated Neural Networks

  The current neural network that WRI uses is a shallow RNN which includes only a LSTM layer and a fully-connected layer. In order to improve the performance, our team have tried adding an embedding layer and change the LSTM layer into a plain RNN layer or a GRU laye and the loss function results on the validation dataset showed a small improvement. Although we will need to validate this improvement in prediction accuracy by introducing a test dataset and calculating the F1 score, the current results show that it is feasible to improve the model performance by adding more layers with more kinds of functions. The final neural network that our team is designing may include additional embedding layer, dense layers, plain RNN layers, LSTM layers and GRU layers.

- Fine tune RoBERTa Encoding Method

  Our team has run the roBERTa encoding method locally on some policy samples, which generated a three-dimensional matrix which can be used as one of the features in the

RNN model. The current pretrained roBERTa model that WRI is using is roberta.large and our team is planning to try other alternatives like roberta.large.mnli which is the roBERTa large model fine-tuned on the Multi-Genre NLI Corpus, roberta.large.wsc which is fine-tuned on Winograd Schema Challenge data, and other roBERTa models our team fine-tuned on different corpus or datasets.

- Additional Feature Engineering

  The current shallow RNN takes the roBERTa encoding of sentences as the input and outputs the predicted probabilities of three labels: positive, negative, and neutral. In order to optimize the performance of the RNN, adding additional features is an available resource which include N-grams, Part-of-Speech tags generated by Hidden Markov Model, LIWC Features, Named Entity Features, Word Frequency and similarity coefficient.

This neural networks method will generate two models including one baseline model trained by the gold standard dataset and another model trained by the noisy dataset. Whether the F1 score of the second model is higher than that of the baseline model with a statistically significant difference will indicate whether our trained model is accurate enough to perform the classification. Therefore, the difference between the two models' F1 scores should be regarded as one of the criteria for success in this project.

Actionability

Time Frame

Our team has drafted a time schedule for our analysis as follows, and have strictly followed our schedule. In week 5 and 6, our team spent much time on downloading the packages that will be

used in the following steps and setting the environment in our computers, which means our team can save much time in week 7-11. Besides, our team will RNN on Google Cloud Platform, which enables us to do the calculation with higher speed than running locally. Thus, our plan is actionable under this time schedule.

- Week 1-3: Identify WRI's business problem and challenge and define objectives and scope for the project
- Week 4: Clean and explore the data and submit the first research proposal
- Week 5-6: Select features and models; prepare for the first client-facing team presentation
- Week 7-11: Compare the results of different labelling model, and fine-tuning one of them by adding more strategies after feature engineering. Fine-tune RNN with RoBERTa embedding, conclude the outputs and results, and put forward suggestions and solutions at the same time.
- Week 12-13: Prepare for the final team presentation

Budget

This project uses data programming to algorithmically label training data based on a small, hand-made gold standard. The current pipeline is from noisy labeling, RoBERTa encoding, to LSTM and the main techniques used are Snorkel and RoBERTa. our team plans to further explore the data provided by WRI by focusing on Label Model Optimization (BabbleLabble vs. Snorkel) and Neural Networks Optimization (RNN and RoBERTa). Therefore, no budget is needed for our project to carry out the analysis on data provided by WRI. The goal of this project is to develop a global policy analysis framework for restoration proposals using policies from India, Brazil, and Mexico as references. While these countries were selected based on existing

available WRI data and resources, we believe that the policy analysis framework developed is versatile and therefore can be modified to scrutinize worldwide restoration initiatives despite limited resources and financial budget.

Additionally, our team has tried running the codes locally, which resulted in both huge computation costs and time costs. Therefore, our team plans to continue our work on training the model on Google Cloud Platform, which can reduce the time costs and lead to better model performance without increasing our budget.

Human Capital

Our team is comprised of highly qualified and experienced experts that are enthusiastic and committed to our client. All of our colleagues are top technical and business talents from Columbia University Applied Analytics program with different focuses including Machine Learning, Natural Language Processing, and Data Processing, etc.

Our team has exceptional technical proficiency and are determined to put forward the most optimal data-driven outcome. Leveraging the data provided by WRI, our team will utilize basic and advanced machine learning techniques and ultimately provide analysis and strategy tailored to our client's needs.

Our team believes that the professional elites on the team will contribute greatly to WRI's vision to help local and national governments implement restoration projects utilizing comprehensive statistical analysis, and ultimately make the planet a better and sustainable living space for human beings. Organizational Culture

WRI values and promotes diversity and equity in its organizational culture and world-wide workforce. The company believes that each individual within the organization has the

responsibility to create a welcoming, innovative, and collaborative environment for people of diverse cultures and backgrounds to work together. According to WRI's vision statement on diversity, equity and inclusion, the company is committed to:

- Advancing gender and social equity for human well-being in our organizational and programmatic practices;

- Recruiting and hiring talent from various backgrounds and identities;

- Promoting internal talent and expanding the diversity of leadership within our staff, senior management, board and advisory bodies;

- Creating training opportunities and events for staff to engage in open dialogue about topics aimed at shedding bias and increasing tolerance and respect for others;

- Managing and monitoring compensation increases to ensure pay equity and fairness across the organization;

- Rejecting and eradicating any form of harassment, bullying, discrimination, retaliation or oppression;

- Continually refining and improving our actions to address diversity, equity and inclusion across our global network, as part of a long-term change process.

Competitive Dynamics

WRI doesn't have many direct competitors but there are a few environmental research institutions like WRI but with different focuses:

- The Brookings Institution is a U.S. research group founded in 1916 that conducts research and education in the social sciences, primarily in the global economy and economic development.

- The Urban Institute is a think tank based in Washington D.C. that carries out social and economic policy research.

- The Pembina Institute is a Canadian non-profit think tank founded in 1985 which focuses on energy policy.

- The Center for Strategic and International Studies is a think tank founded in 1962 that conducts policy studies and strategic analysis of political, economic and security issues worldwide.

- RAND Corporation is an American nonprofit global policy think tank created in 1948 to offer research and analysis to the United States Armed Forces on defense and nondefense issues.

WRI could mitigate these competitions by including our team in its global workforce. Our team consists of knowledgeable professionals with diverse cultures and background which serves as our competitive advantage. Our team has a reasonable division of work and each colleague is fully aware of his/her role on the team. For technical issues of the project, the data scientists on the team are experienced experts in selecting features and building models. For planning and logistical issues, the business specialist on the team will conduct efficient strategic planning in order to achieve WRI's objectives.

Conclusion & Critical Analysis

In this project, our goal is to gain a deeper understanding about the worldwide land restoration and detect incentives and disincentives among policy documents using Natural Language Processing (NLP) and machine learning. Our team plans to focus on two optimization tasks including Label Model Optimization and Neural Networks Optimization, which will help

improve the accuracy and efficiency of WRI's current workflow. As a result, we believe that these processes and improvements can lead to a more accurate classification model which can not only be used for classifying policy documents based on financial incentives but also be applied in a more global scale and to identify more aspects of each financing mechanism through transfer learning. The outcome of the models will fill in as a valuable tool for our clients to improve productivity, increase efficiency, and draw conclusions when conducting related studies, which will ultimately align worldwide restoration agenda and achieve climate goals.

However, there are still several unresolved problems in our project that we need to face.

On the one hand, as our team was experimenting on the BabbleLabble labeling method, we have found that composing the explanations used to label the dataset in an accurate and explicit way is still difficult. The reason for this difficulty lies in the fact that writing explanations for BabbleLabble requires the writer/programmer to have a thorough and deep understanding of the policy documents that we are processing and the mechanisms behind it. Therefore, our team will continue researching and studying the policy documents and its unique formation to identify a better way to write explanations and improve the performance of BabbleLabble.

On the other hand, since the entities mentioned in the policy documents can be some local organizations in these three countries, the pre-trained NLP model, such as Spacy, may not include the records of these organizations and cannot directly help us to extract these entities. One of the possible ways to deal with this problem is looking for an open source library or dataset which contains the information about these entities. Therefore, our team will keep doing paper research to find potential practical methods.