Using Language Model for Emission Factor Mapping

Stanford CS224N Custom Project

Esteban Barrero-Hernandez, Gerald Kang, Amar Gadkari
Department of Computer Science
Stanford University
ejbh24@stanford.edu, gkang03@stanford.edu, amarga@stanford.edu

This template is built on the NeurIPS 2019 template[1].

1 Key information to include

- (Optional) External collaborators:
- (Custom project only) Mentor: We have no particular mentor, but would really like to be assigned one for our project.
- (Optional) Sharing project:
- Short description: In response to the urgent need for businesses to address climate change by enhancing sustainability and efficiency, our proposal aims to improve carbon emission tracking through a language model. By automating the process of collecting business activity data and applying emission factors, our project seeks to drastically reduce human labor in emission calculations, improve accuracy by identifying correct data mappings, and as a stretch goal, possibly pinpoint emission reduction opportunities.

2 Research paper summary (max 2 pages)

Title	Unsupervised Cross-lingual Representation Learning at Scale
Venue	Association for Computational Linguistics (ACL)
Year	2020
URL	`https://aclanthology.org/2020.acl-main.747/`

Table 1: Sample table for bibliographical information (**?**).

Background. One of the most daunting tasks of a company is reporting their carbon emissions for the year. The sheer amount of data to parse through along with the confusion of whether these details were even important or not leads to inconsistencies and difficulty in addressing current sustainability problems. Scope 3 is a particular area for this issue as many companies successfully report Scopes 1 and 2 (which relate to their direct operations), but fail with 3. Furthermore, Scope 3 accounts for "80% of a company's greenhouse gas impacts" Jain et al. (2023) emphasizing its importance to be managed. The authors of this research article search to make data collection of Scope 3 easier so that companies would be more inclined to report their carbon emissions and sustainability problems can be more directly tackled.
Scope 3's difficulty in ability to be managed stems from the complexity of tracking a company product's carbon footprint. With product descriptions varying across companies, the variability of data reporting discouraged companies from reporting Scope 3 emissions. During this paper, the authors sought to find a large language model that extracted title descriptions of reports that could be used in calculating carbon emissions. They tested different models (roberta, bert, distilroberta) along

[1] `https://www.overleaf.com/latex/templates/neurips-2019/tprktwxmqmgk`

with various vector featurization (TF-IDF, word2vec) to see which setup performed the best on the testing dataset.

Summary of contributions. This paper contributes greatly to the sustainability space. The authors compared regular machine learning algorithms to supervised fine tuning to explore which method performs best in emissions classification based on a product description. More specifically, they researched the zero-shot classification for regular machine learning algorithms and measured its performance. They also studied two different vector featurization methods (TF-IDF and word2vec) and found that word2vec performed better. Furthermore, they tested 3 pretrained LLMs (Bert, Roberta, ClimateBert) and found that Bert performed the best. But overall, no matter how poorly an LLM did in comparison to another, all LLM's performed with greater accuracy than the zero-shot classification. This article contributes a new starting point for carbon emissions researchers, proving that supervised fine-tuning is proving more effective in this space than regular machine learning algorithms and even provides which models/methods seem to prove the most efficient.

Limitations and discussion. This project we will finetune and validate recommendations as suggested by the referenced research paper. The referenced research paper focuses on finetuning and measuring which language model performs better. That is necessary and important work. Goal of our paper is to improve quality of EEIO category mapping for businesses. What we have noticed is just looking at description of business transaction, the EEIO category mapping is low quality. Though the paper focuses on best language model and tuning of models giving the best mapping. The quality can be improved and enhance further by including other parameters associated with that business transaction. e.g. if you include vendor name and vendor category that will help do the EEIO mapping correctly, specially when description is vague. When packaging material description list the item that is being packaged, the LLM will get confused and do the incorrect mapping, that will result in incorrect EEIO categorization and hence incorrect emission calculation. Paper does not focus on quality issue due to this limitation. Given our focus on quality of EEIO mapping, we believe paying attention to other information that is available in business transaction is important along with the description text.

Why this paper? As outlined above, We decided to work on using language model for emission factor mapping. When started doing research on this topic, we found quite a few individuals and companies are already digging deeper in this area. It is not surprising given effect of climate change on the environment and us as individuals. There are also various regulatory requirements that puts focus on emission calculations for companies. That has created new wave of opportunities for companies. There are individuals, researchers, and companies joining the bandwagon to fill the void. This paper particularly interested us because it compares various language models. We can build on top of it, by selecting the top performing model.

What interested you about the topic? The research paper, reiterated the fact that approach we are taking is a good approach. It also did some heavy lifting for us by comparing heavy models. We are interested in taking one of the two best models roberta-base and bert-base as a starting point.

Wider research context. Right EEIO categorization is very important for the companies. Quality of the reporting is very important from regulatory perspective. Additionally, with focus on environment sustainability, the quality of data is even more important as individuals and organizations focused on environment are paying more and more attention to such data. This paper focuses on improving quality of the emission reporting.

2.1 Note on citations

There are two citation commands:

Citation in parentheses `\citep{}` is for when you cite a paper at the end of a clause, and you could read the text out loud and not read the authors' names. For example "We also run our experiments on a multilingual language model (**?**)."

Citation in text `\citet{}` is for in-text citations, when someone reading the text out loud would have to read the names of the authors for the sentence to make sense. For example, "We also run our experiments on the multilingual model described in **?**."

3 Project description (1-2 pages)

Goal. The primary aim of this project is to assess the efficacy of a language model, such as BERT, in automating carbon emission tracking for businesses, focusing on whether our language model can significantly reduce human labor and enhance the accuracy of emission data mapping compared to current methods. Motivated by the critical need for businesses to enhance their sustainability practices amidst escalating climate change concerns, this project seeks to refine and make the carbon accounting process more efficient, thus facilitating more informed and strategic environmental decisions. A critical element of this project is to showcase how our approach enhances the application of language models in carbon emission tracking, focusing on surpassing current methodologies by significantly improving automation and precision in tracking efforts. Unlike prior applications, our project aims to set a new standard in accuracy and efficiency, demonstrating the substantial advancements our methods bring to this crucial environmental challenge. As a stretch goal, pinpointing emission reduction opportunities underlines the project's ambition to not only track but actively contribute to the mitigation of environmental impact, showcasing the transformative potential of AI in environmental sustainability efforts.

Task. The task our project tackles is to optimize the carbon emission tracking process through the use of a language model, such as BERT, improving upon the current standard methodologies. To illustrate, the input would be raw business activity data, such as employees commuting 500 miles daily using cars, and the corresponding emission factors provided by regulatory bodies like the EPA, which in this case is 0.18 kg CO2e per mile. The input will originally in tabular form, but we will convert it into one string, making this an NLP task. The output of our system would be the precise calculation of total emissions from this activity, which would be 90 kg CO2e. The learning model would automate data collection, apply the correct emission factors, and calculate the total emissions, thereby streamlining the entire process. This automation is expected to not only reduce the workload but also increase the accuracy and reliability of emission reporting and potentially highlight areas for emission reduction. Here is a more in depth 3 step process to calculate emissions:

Step 1: Collect business activity data - every business activity translates into some impact to the environment i.e. carbon emission, hence it is important to track every business activity. Specific example is employees commuting from home to work adds up to 500 miles/day and they use a car for that travel.

Date	**Activity**	**Method**	**Amount**
2024-01-08	Employee Commuting	Car travel	500 miles/day

Input String: "2024-01-08 Employee Commuting Car travel 500 miles/day"

Step 2: Apply emission factors - government agencies like EPA have provided guidelines (emission factors) for converting business activity data into CO2 equivalent. In this step you identify right emission factor for activities identified in 1. Continuing with employee commute example, the emission factor in this case is 0.18 kg CO2e/mile 0.18 (kg CO2e/mile)

Step 3: Emission Calculations - this step involves calculations over data from 1 and 2 (usually multiplication of quantity and emission factor) to compute total emission from a particular business activity. Building on the employee commute example, 90 kg CO2e is emitted due to this activity.

Data. Our project will utilize synthetic datasets generated to simulate business transaction data, which serves as a proxy for the private datasets accessible to one of out tea member's company. The data is expected to mimic the structure and complexity of real-world business activities, encompassing various transaction types that impact carbon emissions.

We will begin by creating a small dataset of approximately 10 rows using ChatGPT, to generate realistic examples of business activity data. This preliminary dataset will be used for initial unit testing to ensure our language model is capable of emission factor mapping for business activities, including travel, energy consumption, and supply chain operations. For our training and development data, we will scale our data generation efforts to produce a larger dataset of around 100 rows. Finally, for our training data, we will scale our data generation even more to produce around 10000 rows. Each data generation phase will be conducted with stringent measures to ensure there is no overlap between datasets, preserving the integrity of our training, development, and testing phases. Again these input rows will be represented as long strings. The timeline for data generation and preprocessing is estimated to finished in less than a day!

Examples of data rows. These will be turned into long strings:

Date	Fuel Source	Unit	Quantity	Emission Factor	CO2 Emissions	Notes
2024-01-05	Natural Gas	m³	2,500	2.2 (kg CO2e/m³	5,500	Furnace operation during peak production
2024-01-12	Gasoline	L	650	2.3 (kg CO2e/L)	1,495	Company vehicle fleet for deliveries
2024-01-19	Propane	kg	1,200	3.0 (kg CO2e/kg	3,600	Backup generator usage during power outage
2024-01-26	Diesel	L	300	2.7 (kg CO2e/L)	810	Forklift and maintenance equipment operation
2024-01-31	Natural Gas	m³	2,200	2.2 (kg CO2e/m³	4,840	Furnace operation during standard production
2024-02-29	Natural Gas	m³	2,700	2.2 (kg CO2e/m³	5,940	Furnace operation during increased production
2024-03-07	Gasoline	L	700	2.3 (kg CO2e/L)	1,610	Company vehicle fleet with new fuel-efficient model added
2024-03-14	Propane	kg	0	3.0 (kg CO2e/kg	0	Backup generator not used this month
2024-03-21	Diesel	L	250	2.7 (kg CO2e/L)	675	Reduced forklift use due to production optimization
2024-03-28	Natural Gas	m³	2,400	2.2 (kg CO2e/m³	5,280	Furnace operation with efficiency improvements implemented
2024-06-20	Gasoline	L	600	2.3 (kg CO2e/L)	1,380	Continued efficient fuel use in company vehicles
2024-06-27	Natural Gas	m³	2,000	1.9 (kg CO2e/m³	3,800	Furnace operated using renewable energy blend (20% biogas)
2024-06-30	Diesel	L	200	2.7 (kg CO2e/L)	540	Further reduction in diesel use due to process automation

Methods. Using the conclusions made from the research article, we will use Bert as our LLM. Initially, we will train Bert on our EPA table where each input would be the tabular data of the product description. Our variant in comparison to the original model is our input data and how it is structured. Furthermore, we are utilizing the best fine-tuned model in our paper and expanding it to account for more features when classifying the product description. The vector embedding we will use will be word2vec as from the research model, it was concluded that this was the best-performing method. Finally, we will generate more data using ChatGPT to use to fine-tune our model further.

Baselines. We decided to have our baseline method to be the ChatGPT responses. More specifically, what an untuned LLM would predict each product description to be and evaluate that accuracy. This will be implemented ourselves by feeding in test product description queries into ChatGPT and observing what summarizing classifications we obtain.

Evaluation. Our numerical evaluation will be what percentage of product description queries the model could correctly classify. The existing score we will compare against will be how ChatGPT performs on the testing dataset. We will feed the testing data into a fresh session of ChatGPT (replicating a fresh pretrained LLM with no current context) and log its performance percentage. This calculated score will be used to compare against our fine-tuned Bert LLM's performance percentage on the same testing dataset. We predict that the fine-tuned LLM will perform better than the ChatGPT classifier as it will have more optimized layers/context to perform a better prediction in comparison to the general, untrained ChatGPT model.

References

Ayush Jain, Manikandan Padmanaban, Jagabondhu Hazra, Jagabondhu Jagabondhu, and Kommy Weldemariam. 2023. Supply chain emission estimation using large language models. In *Fragile Earth: AI for Climate Sustainability - from Wildfire Disaster Management to Public Health and Beyond*, Online. ACM, New York, NY, USA.