Project Document: Project Name
NeuralNexus

Apala Pramanik
Karissa Jelonek
Avhishek Biswas

Contents

Abstract

For this project, we researched several current research questions in machine learning and reviewed the previous literature on those topics before selecting three potential project ideas based on our interests and research goals. Thus, in this paper we propose three potential deep learning problems including Spanish news classification, sentiment analysis, and image captioning. For each project, we summarize the problem, list possible applications of the problem, summarize previous work in that area, and discuss possible resources and models that we could use for the project.

Chapter 1

Milestone 1: Project Ideas

1.1 Introduction

Deep Learning is a subset of machine learning that uses algorithms inspired by the structure and function of the brain, called artificial neural networks. It has revolutionized the field of artificial intelligence and is responsible for many recent advancements, such as breakthroughs in computer vision, natural language processing, and speech recognition. The need for deep learning arises due to the increasing amount of data generated by various sources and the limitations of traditional machine learning algorithms to handle such complex and large datasets. Deep learning algorithms can automatically learn features from the data and make predictions with high accuracy, making it a valuable tool for solving problems in various domains such as finance, healthcare, and transportation.

Moreover, deep learning is capable of handling non-linear and non-structured data, such as images, text, and speech, which makes it an ideal solution for real-world problems where data can be messy and difficult to process.

1.1.1 Apala:

I discovered the project idea from the "Project Idea Sources" page on Canvas under "Miscellaneous Applications" about using CNN for Natural Language Processing, which sparked my interest and motivated me to perform further research on the topic. During my investigation, I came across the term BERT and was fascinated by its capabilities in natural language processing. My interest in learning Spanish led me to search for Spanish datasets for NLP, as I saw the potential for applying my knowledge to real-world problems in a language that I am passionate about. This combination of academic curiosity and personal interest has driven me to delve deeper into the field of NLP and deep learning, and I am eager to explore the potential applications of these technologies.

1.1.2 Karissa:

For this project, I reviewed the recommended project ideas provided and choose to explore natural language processing problems based on my personal interest in languages and linguistics. From the NLP-Progess website (`http://nlpprogress.com`), I choose to explore current applications and research in the field of sentiment analysis. Section 1.3 summarizes my findings on previous approaches to sentiment analysis and lists resources that I found that could be used for this project.

1.1.3 Avhishek:

I have been interested to work on Image Processing but the challenge of not only identifying and captioning of those images felt very challenging and an interesting project that I can work on as a part of this course. So I looked for research and previous works that have been done in this area. From reading [9] [13] I feel inspired to work and produce some equally good results. Image captioning has numerous practical applications, such as assisting visually impaired individuals, improving image search engines, and creating educational materials for students. I reviewed Kaggle and found the COCO dataset as an very powerful dataset with lots of annotated Images. I find a CNN/RNN or a CNN/Transformer architechture can work as a possible solution for this problem.

1.2 Spanish News Classification: [Apala]

Classifying news articles into different categories is an important task in the field of natural language processing. With the increasing amount of news articles being generated daily, it is essential to develop models that can automatically classify news articles into different categories, such as politics, sports, technology, etc. This can help make it easier for users to find the news articles that are most relevant to their interests.

Classifying news articles into different categories is a challenging task due to the large variability in writing styles, language use, and subject matter. Furthermore, many of the available datasets for news classification are in English, making it difficult to develop models that can accurately classify Spanish news articles

In this project, I aim to develop a model that can accurately classify Spanish news articles into different categories. To do this, I will use a pre-trained Spanish BERT model called BETO [2] and fine-tune it on a labeled dataset of Spanish news articles (see: `https://www.kaggle.com/datasets/kevinmorgado/spanish-news-classification`) that was found on Kaggle. BERT (Bidirectional Encoder Representations from Transformers) [3] is a powerful language model that has been pre-trained on a large corpus of text and has shown excellent results on a variety of NLP tasks, including text classification. As mentioned in their documentation(see: `https://github.com/dccuchile/beto`), "BETO

is a BERT model trained on a big Spanish corpus. BETO is of a size similar to a BERT-Base and was trained with the Whole Word Masking technique".By fine-tuning the Spanish BERT model on Spanish news data, I hope to develop a model that can accurately classify Spanish news articles into different categories.

In summary, our solution to the problem of Spanish news classification is to fine-tune a pre-trained Spanish BERT model on a labeled dataset of Spanish news articles. We believe that this approach will lead to a highly accurate model that can effectively classify Spanish news articles into different categories.

1.3 Project Idea 2: [Karissa]

Sentiment analysis is a branch of natural language processing which can be used to determine the sentiment, or opinion, of a text. For example, sentiment analysis could be used to determine if the tone of a text has an overall positive, negative, or neutral tone. This is a classification problem that classifies the input as either positive or negative or determines a rank of how positive or negative the input is. Sentiment analysis is a challenging task because it involves the issues commonly seen in natural language processing due to the ambiguity of natural language. These issues include negation, ambiguous words (words with multiple meanings depending on the context), multi-polarity words (words that can have positive or negative meaning depending on the context), sarcasm, and others [1].

There are several potential applications of sentiment analysis. One such example is using sentiment analysis for the detection of fake online reviews. Online reviews for businesses and products allow consumers to make informed decisions about their purchases, but could be manipulated by malicious fake negative reviews or fake positive reviews that could be purchased by businesses. As fake reviews are more likely to be highly negative or highly positive, determining the tone of the review can help in identifying potentially fraudulent reviews [6]. Another potential application of sentiment analysis is to gauge the public reception of a product or topic. Social media sites like Twitter are often used by individuals to express their opinions. Analyzing tweets that contain a certain word or use a certain tag could be used to determine how individuals feel about those topics.

The input for a sentiment analysis model is some text such as reviews or tweets. There are several datasets available for sentiment analysis including the IMDb dataset (`https://ai.stanford.edu/~amaas/data/sentiment/`) [10] which contains 50,000 movie reviews which are labeled as highly positive or highly negative (25,000 reviews each). There is an additional 50,000 unlabeled reviews available for unsupervised learning. The output of a sentiment analysis model is the predicted polarity of the text. This could be either binary (positive or negative) or multi-class (ranked according to how positive or negative the text is).

Several previous approaches to sentiment analysis have used the BERT model developed by Google AI Language [14, 12]. BERT (Bidirectional Encoder

Representations from Transformers) has a two step framework that pre-trains and fine-tunes the model. In the pre-training step, the model is trained on unlabeled data. In the fine-tuning step, the parameters are fine-tuned using labeled data. A key feature of BERT is that it is bi-directional meaning that the model can consider the text to both the left and right of a word in order to determine its context [4].

Other approaches to sentiment analysis have used a Long Short-Term Memory (LSTM) neural network [5, 8, 11] to classify the dataset. LSTM uses a memory cell to allow it to store long-term dependencies in data which makes it more suitable for language processing tasks than other recurrent neural networks (RNNs) [7].

For this project, I believe BERT would be the best model to use for this task as it is open-source and has performed with high accuracy for other natural language processing task due to its ability to consider the bi-directional context of text which is important for natural language processing due to the ambiguity of languages.

1.4 Project Idea 3: [Avhishek]

Image captioning is a task in computer vision where a model is trained to generate a descriptive sentence for an input image. The goal is to generate a textual description that accurately summarizes the content of an image. Image captioning has the goal to generate a sentence that describes the content of an image. This is a more complex problem than either regression or classification, as it involves both the extraction of features from an image and the generation of a sequence of words that accurately describes the image. My idea for this project was inspired from reading [9] [13]. Image captioning has numerous practical applications, such as assisting visually impaired individuals, improving image search engines, and creating educational materials for students.

Image captioning is different from both regression and classification, as the goal is to generate a sentence that describes the content of an image. This is a more complex problem than either regression or classification, as it involves both the extraction of features from an image and the generation of a sequence of words that accurately describes the image.

I want to use The Microsoft Common Objects in Context (MS COCO) data set(see: `https://cocodataset.org/#download`). Which is a widely used data set for image captioning, as well as for other computer vision tasks such as object detection and semantic segmentation.The COCO dataset contains 330K images, each annotated with 80 object categories and a variety of attributes, such as instance segmentations, keypoints, and captions. The captions in the COCO dataset are provided in natural language, making it a valuable resource for image captioning research.

I am interested in both a CNN/RNN and a CNN/Transformer based architechture taking inspiration from [13]. The goal is to train a model end-to-end on the MS COCO dataset, starting from randomly initialized weights and op-

timizing the parameters of both the CNN and the RNN/Transformer together. As the model can learn to extract features that are specifically relevant to the task of image captioning.

In summary, the image captioning project involves building a deep learning model that can take an image as input and generate a natural language description of the image, using a combination of CNNs and RNNs or a related architecture, and training the model on the MS COCO dataset or a similar dataset.

1.5 Conclusions

In conclusion, in this paper we propose three possible project ideas – Spanish news classification, sentiment analysis for online reviews, and image captioning with CNN-Transformers. The contributions of each team member for this milestone are listed in Table 1.1. We believe each of these project will allow us to explore beyond the scope of what is covered in this class, but can be completed within the required time frame. We are hoping to get instructor feedback on whether these problems are within the intended scope of this project.

Table 1.1: Contributions by team member for Milestone 1.

Team Member	**Contribution**
Apala	Spanish News Classification
Karissa	Sentiment Analysis for Online Reviews
Avhishek	Image Captioning with CNN-Transformers

Bibliography

[1] Ahmed H Aliwy, Ayad R Abbas, and Mustafa J Hadi. Key challenges and proposed solutions to design sentiment analysis system.

[2] José Cañete, Gabriel Chaperon, Rodrigo Fuentes, Jou-Hui Ho, Hojin Kang, and Jorge Pérez. Spanish pre-trained bert model and evaluation data. In *PML4DC at ICLR 2020*, 2020.

[3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2018. URL: `https://arxiv.org/abs/1810.04805`, https://doi.org/10.48550/ARXIV.1810.04805 `doi:10.48550/ARXIV.1810.04805`.

[5] Scott Gray, Alec Radford, and Diederik P Kingma. Gpu kernels for block-sparse weights. *arXiv preprint arXiv:1711.09224*, 3:2, 2017.

[6] Rakibul Hassan and Md. Rabiul Islam. Impact of sentiment analysis in fake online review detection. In *2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD)*, pages 21–24, 2021. https://doi.org/10.1109/ICICT4SD50815.2021.9396899 `doi:10.1109/ICICT4SD50815.2021.9396899`.

[7] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997. https://doi.org/10.1162/neco.1997.9.8.1735 `doi:10.1162/neco.1997.9.8.1735`.

[8] Rie Johnson and Tong Zhang. Supervised and semi-supervised text categorization using lstm for region embeddings, 2016. URL: `https://arxiv.org/abs/1602.02373`, https://doi.org/10.48550/ARXIV.1602.02373 `doi:10.48550/ARXIV.1602.02373`.

[9] Wei Liu, Sihan Chen, Longteng Guo, Xinxin Zhu, and Jing Liu. Cptr: Full transformer network for image captioning. *arXiv preprint arXiv:2101.10804*, 2021.

[10] Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. Learning word vectors for sentiment analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 142–150, Portland, Oregon, USA, June 2011. Association for Computational Linguistics. URL: `http://www.aclweb.org/anthology/P11-1015`.

[11] Bryan McCann, James Bradbury, Caiming Xiong, and Richard Socher. Learned in translation: Contextualized word vectors, 2017. URL: `https://arxiv.org/abs/1708.00107`, https://doi.org/10.48550/ARXIV.1708.00107 `doi:10.48550/ARXIV.1708.00107`.

[12] Chi Sun, Xipeng Qiu, Yige Xu, and Xuanjing Huang. How to fine-tune bert for text classification? In Maosong Sun, Xuanjing Huang, Heng Ji, Zhiyuan Liu, and Yang Liu, editors, *Chinese Computational Linguistics*, pages 194–206, Cham, 2019. Springer International Publishing.

[13] Yiyu Wang, Jungang Xu, and Yingfei Sun. End-to-end transformer based model for image captioning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 2585–2594, 2022.

[14] Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Ruslan Salakhutdinov, and Quoc V. Le. Xlnet: Generalized autoregressive pretraining for language understanding, 2019. URL: `https://arxiv.org/abs/1906.08237`, https://doi.org/10.48550/ARXIV.1906.08237 `doi:10.48550/ARXIV.1906.08237`.