Ph.D. Dissertation

S.S.J University, Almora

Hindi Tweets Sentiment Analysis Using Transfer Learning

Department Of Computer Science

By ANUBHAV MEHRA

Contents

Introduction

Abstract

Sentiment analysis is a **natural language processing** technique to find if the sentiment of the text is positive, neutral or negative. Traditionally, to train a model for sentiment analysis require very dense neural networks to train on very huge datasets. But, here we have used a technique called **Transfer Learning** that stores a model which has learned some knowledge, that we can leverage in solving some other tasks based on the knowledge of the previous model. Here we are using a language model called **BERT(Bidirectional Encoder Representations from Transformers)**. BERT is a pertained model which learns using the learning techniques developed by Google. The BERT multilingual base model that we are using is pertained on the top 104 languages including Hindi. We then leverage the power of this model for the Sentiment analysis of the Hindi texts dataset that we've got. This allows us to achieve moderately high accuracy scores using a comparatively small dataset.

Transfer Learning is a Machine Learning method where a model that is trained for a certain task is utilized as the starting point for solving some other task i.e., to train a second model from the knowledge learned from the first model as well as the dataset. It is a very popular approach in natural language processing domain to solve problems such as getting the context of the text.

The inspiration for transfer learning comes from us - humans, ourselves - where in, we have an inherent ability to not learn everything from scratch. We transfer and leverage our knowledge from what we have learn in the past for tackling a wide variety of tasks.[1]

Deciding when to use Transfer Learning

As with any technique at our disposal we must have a very clear reason on using transfer learning. First and foremost, we must realize that we can't use transfer learning in every situation. This is because for transfer learning to work we must have a model trained on a similar task. As with the problem concerned in our dissertation here, we have a language model available which is trained in recognizing linguistic structure of various languages. Now, we can leverage the power of this model to train a new model for our context analysis problem. If this is the case, you will realize that transfer learning is a very decent technique to use for acquiring better results and that too in very short amount of time than traditional methods. Now, with this out of the way you may want to use transfer learning if you have following conditions:

- You have a smaller dataset

- You are constrained on time

Applications of Transfer Learning

There are various applications of Transfer Learning in variety of domains. In **Image Recognition**, transfer learning can be used to perform various imaging tasks. For e.g., a model trained to identify face can be used for facial recognition. Transfer Learning can also be used in **Speech Recognition** tasks. For e.g., a model trained in a particular language for speech recognition can be used to train a model for speech authentication. But our key domain of interest where we are most interested in using this technique is **Natural Language Processing**. Transfer learning can be used to solve various tasks in NLP. For e.g., a model trained in recognizing the linguistic structure of a language can be used to train a model to predict next word based on the series of words it gets in the previous sentences.

Sentiment Analysis

One such task in NLP is **Sentiment Analysis**. Sentiment Analysis or emotion AI, is the process in natural language processing of subjective emotional analysis of the text. Primarily sentiment analysis finds if the emotional tone of a piece of writing is *positive, neutral or negative*. **Table 1** lists some examples of what a sentiment analysis categorization may look like.

Table 1: Sentiment Analysis Categorization.

Text	Category
That restaurant has a great food	Positive
He is my brother's colleague	Neutral
Bollywood movies are not entertaining	Negative

As we can see it is easily understood by a human brain what sentiments these pieces of writing represent. But, for a computer this is a very challenging problem. It is a challenging problem because the way humans communicate using natural languages is extremely varied depending upon *subjectivity*, *use of irony*, and the *context* of the phrase. Not all verbs, nouns and pronouns can be treated equally when analyzing the emotional tone of the text. The tone of our phrases depends on how we are using them, when we are using them and for what we are using them. For e.g., the phrase *absolutely everything*, can be an answer to a question like *What are you so happy about?* or can equally be an answer to a completely opposite question *What are you so sad about?*. Thus, the emotional tone of

the phrase can have completely separate meaning depending upon the various parameters. In the example above, the context of the phrase can alter the emotional meaning of the phrase completely.

References

[1] D. D. Sarkar. (Dec. 13, 2018). "Deep transfer learning for natural language processing — text classification with universal...," Medium, [Online]. Available: `https : / / towardsdatascience . com / deep - transfer - learning - for - natural-language-processing-text-classification-with-universal-1a2c69e5baa9` (visited on 07/25/2021).