



LANGUAGE MODELING AND TEXT CLASSIFICATION

Edson Bladimir Pinto Luque
MINERIA DE DATOS

Ingeniería Estadística e Informática
Universidad Nacional del Altiplano

Mayo de 2023

DEFINITION

Text classification and language modeling are two related but separate NLP (natural language processing) activities.

language modeling

Building a statistical or computational model that can anticipate the next word or phrase. The basic purpose of language modeling is to capture the structure and grammatical relationships in the text. This can be applied to a variety of tasks, including automatic text production, word autocompletion on mobile devices, and machine translation.

Classification text

Its focus is on assigning one or more predefined labels. The goal is to classify the text into specific categories or classes based on its content. Some popular instances of text categorization are email spam detection, social media sentiment analysis, dividing news into distinct groups, and document subject detection.

LIST OF TECHNIQUES

Modeling Languages

- N-gram models.
- Statistical-Based Language Models
- Neural Language Models
- Transformer-Based Language Models
- Neural Language Models
- Transformer-Based Language Models

LIST OF TECHNIQUES

Text Classification

- Word Vector Representations.
- Conventional Classification Models
- Neural Language Models
- Neural Networks for Text Classification
- Transfer of Learning

Code and Results

GoogleColab https://colab.research.google.com/drive/1EZgR-mjh_67enWq8S4m3jINNp6lzs8Pl?usp=sharing

```
-----  
Texto: Socially distanced and back for another supermarket trip. One day for one f  
Etiqueta real: 2  
Etiqueta predicha: Neutral  
-----  
Texto: Gold Prices Set for Record Highs Says Most Bearish Forecaster as Covid-19 S  
Etiqueta real: 2  
Etiqueta predicha: Neutral  
-----  
Texto: Heading to the grocery store... I should probably shower and put deodorant  
  
#Covid_19 | #SeattleTogether  
Etiqueta real: 2  
Etiqueta predicha: Neutral  
-----  
Texto: I think @ReginaKing knew about #TheRona before we all did. Care to explain  
Etiqueta real: 0  
Etiqueta predicha: Positive  
-----  
Texto: Better to have and not need than need it and not have it. Essential jobs re  
Etiqueta real: 3  
Etiqueta predicha: Positive  
-----  
Texto: @TheBlinkingOwl joined the growing list of distilleries producing and/or do  
#COVID19  
Etiqueta real: 3  
Etiqueta predicha: Extremely Positive  
-----
```

Figura: Result for predictions

TOPICSTRIKER: A TOPIC KERNELS-POWERED APPROACH FOR TEXT CLASSIFICATION

The article describes two well-known techniques one of machine learning and natural language processing.

The goal is to provide better text representations, including text categorization. String kernels are mathematical operations on text strings that allow constructing feature matrices for classification

Theme modeling is used to discover latent semantic patterns in vast collections of texts.

The method combines the two methods to maximize their benefits and produce superior text renderings.

CAN AI LANGUAGE MODELS REPLACE HUMAN PARTICIPANTS?

The article explores the possibility of using language models as participants in psychological research, to capture human-like moral judgments.

The advantages and limitations of language models in relation to human participants are discussed, and areas in which they can complement research are highlighted.

It is concluded that an understanding of a human is still needed and that language models cannot fully replace human participants in psychological science.