Emotion and Affect Representation in Sentence Embeddings

Author:
Luis Alberto
BARRADAS CHACÓN
278183

Supervisors:
Prof. Dr. Christian WARTENA
MSc. Rafael DRUMOND

18th May 2020

Thesis submited for
MASTER OF SCIENCE IN DATA ANALYTICS

WIRTSCHAFTSINFORMATIK UND MASCHINELLES LERNEN
STIFTUNG UNIVERSITÄT HILDESHEIM
UNIVERSITATSPLÄTZ 1, 31141 HILDESHEIM

Statement as to the sole authorship of the thesis:

Emotion and Affect Representation in Sentence Embeddings.
I hereby certify that the master's thesis named above was solely written by me and that no assistance was used other than that cited. The passages in this thesis that were taken verbatim or with the same sense as that of other works have been identified in each individual case by the citation of the source or the origin, including the secondary sources used. This also applies for drawings, sketches, illustration as well as internet sources and other collections of electronic texts or data, etc. The submitted thesis has not been previously used for the fulfilment of a degree requirements and has not been published in English or any other language. I am aware of the fact that false declarations will be treated as fraud.

18th May 2020, Hildesheim

Abstract

Word embeddings exist to abstract features of text into a numeric space. In this project we explore the representation of different emotions and affects from established labeled datasets in common word embeddings.

Acknowledgements

Acknowledgements Here

Contents

List of Figures

List of Tables

Listings

Chapter 1

Introduction

1.1 Emotions and Affect

Within the context of this project it is important to distinguish between emotion and affect. Affect, in the context of this project will be treated as a term to associate predisposition towards stimuli. Thus, affect is in a sense, a general term that can be even used to describe animal, and other non-human entities. Emotions, on the other hand, are treated in this project as a state inherent to humans. This state is multidimensional, and every dimension, or emotion, can either be present in a certain amount, or not be present at all. Emotions present an affect value, but not necessarily otherwise.

Models of Emotions

When trying to detect emotion, it is relevant to know what emotions to look for. This is called a model of emotions, and is still a very discussed subject is psychology. Although there are several models of emotions the most persistent are Ekman's model, Plutchikc's model. The main assumption this work does follows the theory of constructed emotions, which recognizes affect as a physiological response to positive or negative stimuli, but emotions as a cognitive form of context-giving.

Why is it important to study emotion? Emotions are considered a human state that influences behaviour and decision making. Many times, when expressing thoughts in a written or spoken form, one or several emotions are present. Detecting these emotions is an important task for human interaction. Automatic emotion detection on text is thus a machine learning task required for comprehensive human-computer interaction.

Eckman's model of Emotions

What are emotions? What is affect? Why study emotions? The role of emotions in communication Measurements of emotions:

- Facial Expressions
- Biosignals
- Language or self report

Constructed theory of emotions Emotions in Language Semantic Fields Emotion Lexicon Emotion Networks Learning word semantics from context

1.2 Emotion and Machine Learning

Language in Machine Learning NLP workflow? Language representation Tokenization The problem of large dictionaries Dimensionality reduction/ autoencoders Deep learning for language: automatic feature extraction Machine Learning meaning from context Word Embeddings Transformers

1.3 Problem setting

1.4 Project Description

1.5 Objective

Objetivo general: Análizar la representación de emociones en modelos del lenguaje en machine learning. Esto de manera objetiva y medible, y subjectiva, pero disponible al público para fomentar discución sobre la representación de emociones en modelos de machine learning.

Objectivos secundarios Plantear una metodología para análisar un dataset de emociones, basado en modelos de lenguaje previamente entrenados.

Cómo consecuencia del objetivo anterior, a marco teoríco y práctico se puede crear para el análisis de eficiencia de modelos previamente entrenados en datasets de clasificación de textos cortos, acompañados de una sola etiqueta. Ya que el resultado no está vinculado a la semántica de la etiqueta: (emociones). Pero la representatividad adecuada de las etiquetas en el espacio abstracto propuesto por el modelo, resulta en la más fácil clasificación de las observaciones y sus respectivas etiquetas.

1.6 Justification

Word Embeddings Numerically representing words is commonly known as word embedding. This allows for Machine Learning (ML) models to easily manipulate text data that would otherwise be an arbitrary encoding of text. With the advances in machine learning automatic word embedding became a possible solution to avoid crowdsourcing. Several machine learning approaches try to automatically learn the best numeric representation for characters, words, or sentences given the context of a dataset. Recently there have been many efforts from research institutions to generalize these embeddings though the use of powerful models, and bigger datasets. Such is the case of BERT, a model created by Google with massive datasets.

Affect learning and its relevance The relevance of affect in text has increased since the popularization of text-based social networks, like twitter. There, individuals and organizations openly express their opinions. This creates an environment where implicit feedback about entities is present. An easy way to abstract popular opinion about a named entity is learning the affect expressed in text, such as a tweet. Affect can be a multidimensional phenomenon, but the most important dimension of it is valence: whether a text expresses positive or negative emotion.

Figure 1.1: Example of an image

In Figure 1.6 you can see an example of an image.

Followed by equation 1.1

$$
\begin{aligned}
A = \{&'a' : ['d','e','f','g'],\\
&'b' : ['d','e','f','g'],\\
&'c' : ['d','e','f','g'],\\
&'d' : ['h'],\\
&'e' : ['h'],\\
&'f' : ['h'],\\
&'g' : ['h'],\\
&'h' : []\}
\end{aligned}
\tag{1.1}
$$

This is now a URL:
http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm

1.7 A Section

A subsection

A sub-subsection

Listing 1.1: A codeblock

```
for i in aList:
  print(i)
l = [i for i in aList if i % 2 == 1]
```

As we can see in the Listing 1.1 it can be referenced.

1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20

Table 1.1: An Example of a table

As we can see in the Table 1.1 it can be referenced

Chapter 2

Related Work

2.1 Lexicons

Mohamad and Turney created an Emotion Lexicon through crowdsourcing [Mohammad and Turne
In this way an emotional word embedding was created by subjectively asking participants whether or not a word was related to a specific emotion.

2.2 Automatic Approaches

Vo and Zhang created an automatic approach to learning sentiment lexicons for short texts through the use of emojis [Vo and Zhang, 2016]. This method uses the intrinsic usage of emojis to express positive or negative valence in a sentence, and exploded this to expand that valence to words used in the same context.

Maas et al.created a method to learn word vectors for sentiment analysis [Maas et al., 2011].

By applying machine learned automatic embeddings, the creation of word embeddings based only on text data was open as a possibility. This is also a method that later became the popular Word2Vec method [Mikolov et al., 2013].

A refining of word embeddings has been suggested by Yu et al.by means of a clustering algorithm on the vector space [Yu et al., 2017].

Rothe et al.suggested an orthogonal transformation to word embeddings used on SemEval2015 which yielded ultradense word embeddings for affect [Rothe et al., 2016].

A further exploration of transformations of a word vector space was done by Hollis et al.by means of component analysis, thus creating models of semantics from text. These were applied to affect[Hollis and Westbury, 2016].

These studies have mostly been done with affect: positive and negative valences, but have mostly ignored other emotional dimensions.

2.3 Word Embeddings

2.4 Language Models

For word embeddings, different pre-trained models will be used:

- Fasttext
- Word2Vec
- GloVe
- BERT

A word embedding can also be trained from data. This is considered an optional model to analyze.

Selected Language Models

FastText

Word2Vec

GloVe

BERT

2.5 Analysis Algorithms

The following techniques have been used:

- TSNE
- PCA

2.6 Datasets

There is the possibility of using the datasets from the unified dataset of emotion in text. (An Analysis of Annotated Corpora forEmotion Classification in Text). This includes:

- AffectiveText
- Blogs
- CrowdFlower
- DailyDialogs
- Electoral-Tweets
- EmoBank
- EmoInt
- Emotion-Stimulus
- fb-valence-arousal
- Grounded-Emotions
- ISEAR
- Tales
- SSEC
- TEC

The link to these datasets can be found under the github repository for the unified emotion datasets. https://github.com/sarnthil/unify-emotion-datasets/tree/master/datasets

During the exploratory phase of this project, a subset of these datasets will be selected, giving priority to those with a more standardized, cleaner, or easier to access data. Models

2.7 Research Question

As a general research question we propose to answer the following: When using pre-trained models for word and sentence embedding, **is the information about the emotional and affective content or context of the word or sentence represented in the vector space?**

This question can be approached in three different ways:

- Is there a direct correlation between any of the dimensions of the vector space and human-labeled emotions and affect?

- Is there a linear transformation that will yield a direct correlation to the same human-labeled emotions?
- Is there a hierarchical structure that accurately represents the embedding of said labels?

Chapter 3

Methodology

To analyze the representation of emotions in different word embeddings, a high emphasis on dimensionality-reducing visualizations was done.

The steps to do so are the following:

1. First, the selected datasets must be embedded in their vector space. This will allow for a faster processing of the data, but requires making the technical decision of how the word vectors will be turned into sentence vectors to share the label in the case of datasets that contain a label for every sentence.

2. A supervised clustering must yield accuracy similar to that of a classification task. This means testing the accuracy of a classifier on the emotion recognition task. Even though this step could be seen as optional, failing to reproduce the baseline accuracy scores for this simple task might mean that the embeddings are not even capturing the basic information about the task.

3. A correlational analysis will be done between every dimension of the vector space and the emotions present in the dataset. This will tell us about any linear representation of emotions in the vector space.

4. A second study must show if a linear transformation of the vector space dimensions will yield a better correlation with the emotions of the datasets. This means either LDA, or PCA.

5. A hierarchical clustering will be used to analyze any possible structure in the embedded dataset in relation to emotions.

6. A final approach to this problem can be done through the study of oppositeness.

The mentioned methodology is based on answering the research question with progressive approximations. It is highly unlikely that a simple embedding model represents emotions in a single dimension in a linear manner, but it is increasingly more likely that some correlation is found with a linear transformation of the aforementioned. In case these two approaches present no information about emotions, a hierarchical clustering can extract the intrinsic information of affect in emotions. Since previous works have already shown that affect can be represented in vector spaces, created with a linear transformation of word embeddings[**?**], it would be contradictory to not find a hierarchical structure of emotions in this last step.

3.1 Preliminaries

This research was managed as both, a research project, and a software development project. As a setup for this project, two steps were taken. Setting up the physical and virtual environments for the experiments, and converging the different datasets and model embeddings into a single data source.

With scientific rigor, order, and reproducibility in mind, a git repository has been setup, where not only the working environment is provided, but also the history of the project development.

Environment Setup

Organizational

The project planning was layed out throught three months: March, April and May of 2020. A total of twelve weeks were divided into four equal sprints, where the four main tasks in the project were equaly separated in time: Exploration and Preparation, Programming, Experiments, and Writing. The four sprints were described by tasks, further divided by sub-tasks. These were kept in track and followed me, and both Supervisors through the Asana application[**?**].

The repository is accessible through github: `TODO`

Hardware

OS: Manjaro Linux x86˙64 Kernel: 5.6.11-1-MANJARO Uptime: 1 day, 21 hours, 56 mins Packages: 2095 (pacman) Shell: zsh 5.8 Resolution: 1920x1080 DE: Plasma WM: KWin Theme: Breath [Plasma], Breath-Dark [GTK2/3] Icons: Materia-Manjaro-Dark-2 [Plasma], Materia-Manjaro-Dark-2 [GTK2/3]

Terminal: yakuake CPU: Intel i7-8700K (12) @ 5.000GHz GPU: NVIDIA GeForce GTX 1080 Ti Memory: 11200MiB / 15937MiB

Software

Development

ML Environment

Emdedding the Datasets

The datasets were downloaded and stored under the project folder "data". Since every dataset is provided in different format and under different folder structures, every dataset is simply stored inside a folder with it's name. Under the datasets folder, every selected dataset is accompanied by folders with the embedding model used to embed the dataset. Thus every dataset folder has several subfolders. On these subfolders, a python script called "embed.py". This script varies for every model and dataset. In general terms, it extracts the text and label from the dataset, embeds the text into the desired model, and stores it in a "csv" file under the same folder. The "csv" file is stored under the name "embedded.py", except for the FastText model. In this case, there are two embedding approaches, one supervised and one unsupervised. Thus the names of the FastText embedding files are "embeddings˙supervised.csv", and "embeddings˙unsupervised.csv". Every other script creates a single "csv" file called "embeddings.csv".

Since every model embedds text in different

Thie file structure of the embedded files allows for exploration and experimental scripts to access the embedded data of different datasets, by building a single string with the dataset and model selected. This string must be prepended by the "./data/" folder name, and appended with the "embeddings.csv" string to generate a path that creates accesibility to the different datasets via a python coma-sepparated-value library, such as the built in `csv`, or Pandas [?] and it's `read_csv` function. This effectively create a data source to be used in a data pipeline. This approach was selected due to it's simplicity.

Embedding with FastText

Due to the two methods for the usage of the FastText python library: supervised and unsupervised, the process of embedding a dataset with it requires two extra text files one with a sentence per line, and a second one, which includes the label as the last word of every line, prepended by two underscores

(__).

In both ways of training, the language model is being trained specifically for the

Embedding with Word2Vec

Embedding from W2V means that the vocabulary is dependant on the training set,

Embedding with GloVe

Embedding with BERT

3.2 Analysis

Correlational Analysis

Linear Dimentionality Reduction

Non-Linear Dimentionality Reduction

Clustering Analysis

3.3 Experiments

Emolex

Selecting Emotions

Class Inbalance

Valence

Chapter 4

Experiments

4.1 Performance

4.2 Results

Chapter 5

Conclusion

Bibliography

[Hollis and Westbury, 2016] Hollis, G. and Westbury, C. (2016). The principals of meaning: Extracting semantic dimensions from co-occurrence models of semantics. *Psychonomic Bulletin & Review*, 23(6):1744–1756.

[Maas et al., 2011] Maas, A. L., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., and Potts, C. (2011). Learning Word Vectors for Sentiment Analysis. page 9.

[Mikolov et al., 2013] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. page 9.

[Mohammad and Turney, 2013] Mohammad, S. M. and Turney, P. D. (2013). Crowdsourcing a Word-Emotion Association Lexicon. *arXiv:1308.6297 [cs]*. arXiv: 1308.6297.

[Rothe et al., 2016] Rothe, S., Ebert, S., and Schütze, H. (2016). Ultradense Word Embeddings by Orthogonal Transformation. *arXiv:1602.07572 [cs]*. arXiv: 1602.07572.

[Vo and Zhang, 2016] Vo, D. T. and Zhang, Y. (2016). Don't Count, Predict! An Automatic Approach to Learning Sentiment Lexicons for Short Text. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 219–224, Berlin, Germany. Association for Computational Linguistics.

[Yu et al., 2017] Yu, L.-C., Wang, J., Lai, K. R., and Zhang, X. (2017). Refining Word Embeddings for Sentiment Analysis. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 534–539, Copenhagen, Denmark. Association for Computational Linguistics.