Université de Lorraine IDMC IECL

UNIVERSITÉ DE LORRAINE

SUPERVISED PROJECT

Text Emotion Analysis

Authors:
Wenjun SUN
Adrien CLAUDEL
Guillaume RICHEZ

Supervisor:
CLAUSEL Marianne
Examiner:
COUCEIRO Miguel

A thesis submitted in fulfillment of the requirements
for the degree of NLP Master of Sciences

in the

Institut des sciences du Digital — Management & Cognition

June 18, 2021

UNIVERSITÉ DE LORRAINE

Abstract

Institut des sciences du Digital — Management & Cognition

NLP Master of Sciences

Text Emotion Analysis

by Wenjun SUN
Adrien CLAUDEL
Guillaume RICHEZ

Text emotion analysis refers to mining and analyzing the opinions and emotions of texts through computer technology. It has become one of the most active research fields in NLP, and has expanded from computer science to management and sociology, such as marketing, communication, health science, and even history. Existing research has produced a large number of techniques that can be used for multiple tasks in sentiment analysis, including supervised and unsupervised methods. The supervised method uses supervised machine learning methods and feature combinations. Unsupervised methods include different methods using emotion dictionaries, grammatical analysis, and syntactic patterns. This paper analyzes the emotion recognition effects of various classifiers based on the ISEAR dataset. At the same time, we also performed a cluster analysis based on the French corpus.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

First, for the ISEAR data set, we conducted a word-embedding experiment and tried to use a variety of methods for sentiment analysis. Besides, our work is part of a project which aims at training a classifier capable of analysing French sentences among a panel of 25 emotions. This article introduces the relevant basic knowledge of sentiment analysis, the data set and word embedding we selected, and introduces in detail the experiments we conducted and the analysis of experimental results including classification task and clustering task.

Chapter 2

Emotion Analysis

Emotions can be expressed through human language expressions and words, etc. And text emotion recognition is to judge the emotional state of experimental objects through text information. Nowadays it is a recent essential research area in Natural Language Processing (NLP) which may reveal some valuable input to a variety of purposes[2]. Emotion recognition has many application scenarios in our lives such as psychological counseling, product evaluation, film reviews, and early warning of social security incidents.

2.1 Emotion Model

Emotion models are the foundation of the recognition task. Models express how to define a kind of emotion. The models assume that emotions exist in various states thus the need to distinguish between the various emotion states[1]. Today, there are two main types of sentiment classification models: categorical model and dimensional model. The first one is made of six fundamental kinds of emotions: anger, disgust, fear, joy, sadness, and surprise. Sometimes researchers also use characteristic emition classes such as confusion and boredom. The second dimensional model denotes affects in a dimensional form. A common set of dimensions link the various emotional states in this model. They are defined in a two(valence and arousal) or three (valence, arousal, and power) dimensional space[7].

These two emotion models are shown below.

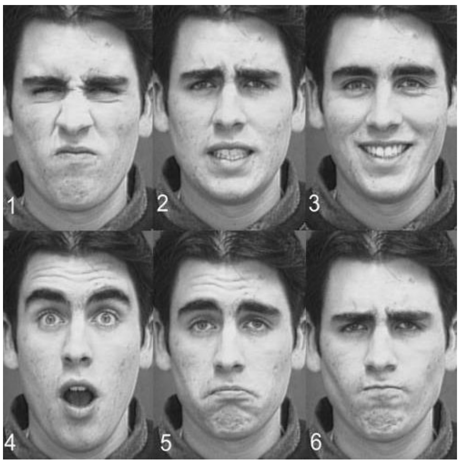

FIGURE 2.1: Categorical Model

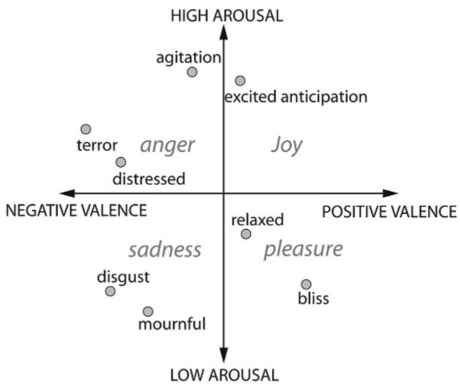

FIGURE 2.2: Dimensional Model

2.2 Analysis Methods

There are two main methods to achieve emotion recognition: Rule-based method and Machine-learning(ML) method. The former one explains grammatical and logical rules to follow in order to detect emotions from documents. Rules for few documents may be easily created; however, with large amounts of documents, complexities may result. ML method applies kinds of classification algorithms to associate text data to different emotion labels. In this paper, we used the ML methods and compared their effect.

Chapter 3

Corpus

3.1 ISEAR Dataset

Intercultural Study on Emotional Antecedents and Reactions (ISEAR) is a text dataset for emotion recognition. It is a worldwide project that started in the 1990s, guided by Klaus R. Scherer and Harald Wallbott, and its data comes from psychologists around the world. All subjects were asked to tell about their experiences about these seven emotions (joy, fear, anger, sadness, disgust, shame, and guilt) and answer a series of questions to explain their reactions at the time. Previous studies using the ISEAR dataset try to find relationships among emotions and different cultures, genders, ages, and religions. But this corpus is well suited to use for emotional text classification purposes[3]. Table 1 shows the sample of joy samples. ISEAR dataset contains 7666 items. The distribution of seven emotions is shown in the below figure. We can see ISEAR is a balanced dataset.

During the period of falling in love, each time that we met and especially when we had not met for a long time.
When I got a letter offering me the Summer job that I had applied for.
On days when I feel close to my partner and other friends.
After my girlfriend had taken her exam we went to her parent's place

TABLE 3.1: ISEAR Joy Samples

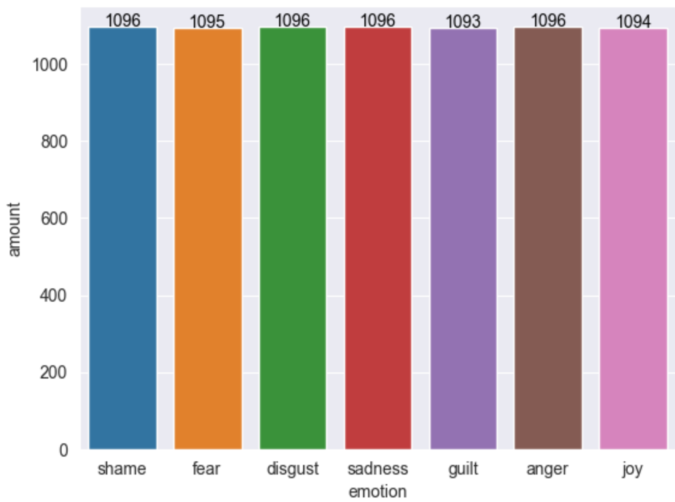

FIGURE 3.1: Amount of Each Emotion

The questions that subjects were asked included:

- Describe a situation or event in which you felt the given emotion
- When did this happen
- How intense was this feeling
- How long did you feel the emotion

We only focus on the emotion and the corresponding description. The distribution of the length of sentences is shown below.

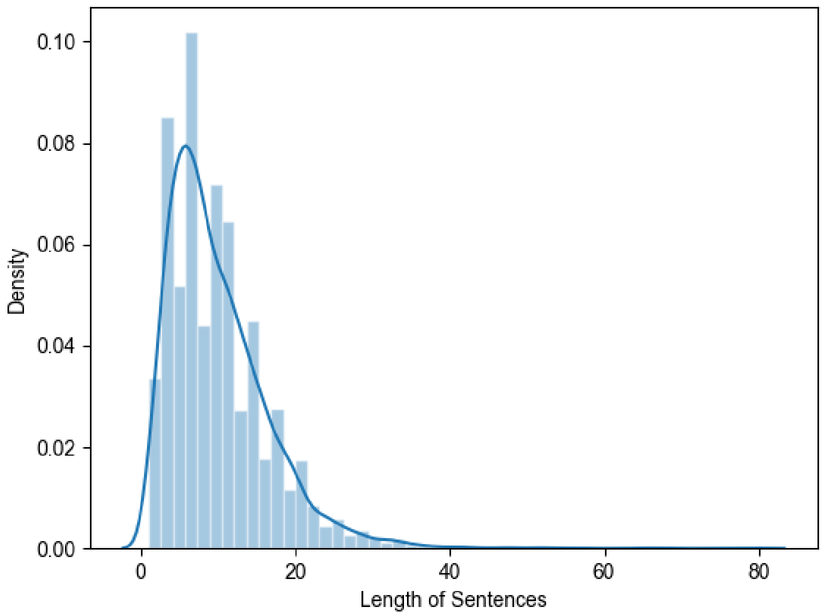

FIGURE 3.2: Distribution of Length

Table 3.2 shows the four most frequent words in each emotion.

emotion	word
joy	happy, friend, good, passed
fear	afraid, night, alone, scared
anger	angry, anger, time, boyfriend
sadness	died, death, failed, sad
disgust	disgusted, people, worn, stink
shame	aloud, ashamed, spell, smear
guilt	guilty, pinch, decline, caretaker

TABLE 3.2: ISEAR Samples

3.2 French Corpus

3.2.1 Sentiment analysis Corpus

The corpus used in the classification task is a transcript of interviews conducted with several students and then separated into separate sentences. Each of these sentences is annotated with one of the following 25 emotions:

- Relaxation
- Disgust
- Empathy
- Pride
- Skepticism
- Reflection
- Admiration
- Concentration
- Surprise
- Guilt
- Embarrassment
- Amusement
- Confusion
- Worry
- Contempt
- Pain
- Joy
- Anger
- Disappointment
- Sympathy
- Sadness
- Annoyance

Initially, the corpus was split into 25 text files (one for each emotion). We retrieved these files in order to store their data in a csv file with two columns, one for the emotions and one for the corresponding sentences, all in the form of character strings. This allowed us to define several characteristics of the corpus. The corpus contains 11560 sentences of which 10813 are unique. The sentence that appears with the highest frequency is "What?" which appears 67 times.

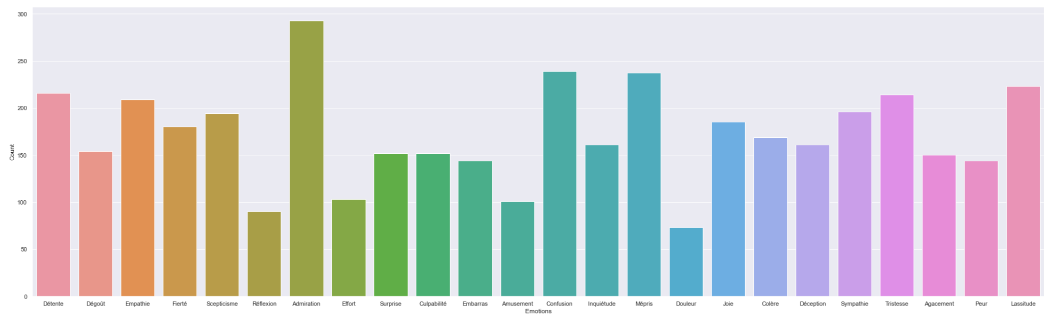

FIGURE 3.3: Distribution of emotions on gobal corpus

3.2.2 Corpus for the test: frenchtweets

French Twitter Sentiment Analysis is a corpus that we used to compare different embeddings. *frenchtweets* consists of 1,526,724 values of which 1,471,857 are unique, the phrase that appears with the highest frequency is: "I thank you" which appears 785 times.

The corpus used for this clustering consists of tweets and is generally used for sentiment analysis. As the final project has the same objective, this seemed to us to be a coherent choice. Initially, a positive or negative label was attached to each tweet of this corpus in the form of a binary number: 0 for tweets with a negative connotation and 1 for tweets with a positive connotation.*frenchtweets* is very balanced since

49.46% of the tweets have a negative connotation and 50.34% a positive connotation.

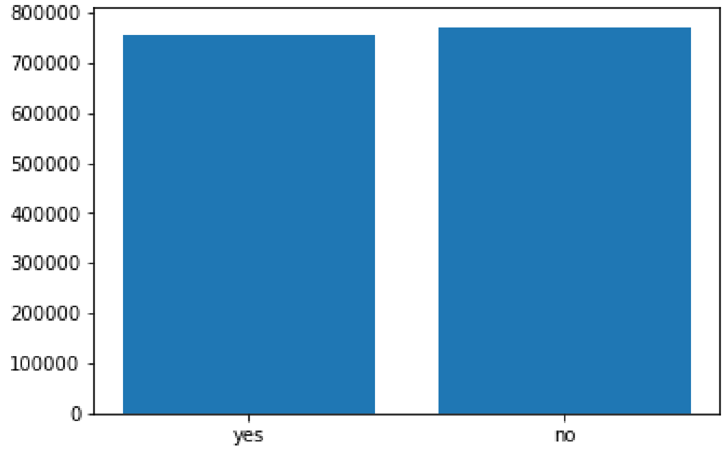

FIGURE 3.4: Distribution of emotions on French tweet corpus

However, this corpus has a major drawback in that the tweets in it are not written by native French speakers, they are French translations of English tweets.

Chapter 4

Word Emdedding Methods

4.1 Method used on ISEAR

4.1.1 Bert

Bidirectional Encoder Representations from Transformers(BERT) is a data pre-processing algorithm developed by google in 2018 to improve the understanding of user queries on their search engines. This model processes words according to all the other words in the sentence. This method significantly improves the results of the current models for data mining tasks. The essence of BERT is to learn a good feature representation for words by running a self-supervised learning method on the basis of a large amount of corpus. The so-called self-supervised learning refers to supervised learning that runs on data that is not manually labeled. In future specific NLP tasks, we can directly use the feature representation of BERT as the word embedding feature of the task. Therefore, BERT provides a model for migration learning of other tasks, which can be used as a feature extractor after being fine-tuned or fixed according to the task. BERT's network architecture uses the multi-layer Transformer structure proposed in "Attention is all you need"[9]. Its biggest feature is that it discards the traditional RNN and CNN, and converts the distance between two words at any position into 1 through the Attention mechanism, Which effectively solves the thorny long-term dependence problem in NLP.

BERT's model architecture is a multi-layer bidirectional Transformer en-coder based on the original implementation described in Vaswani et al.(2017)[4].

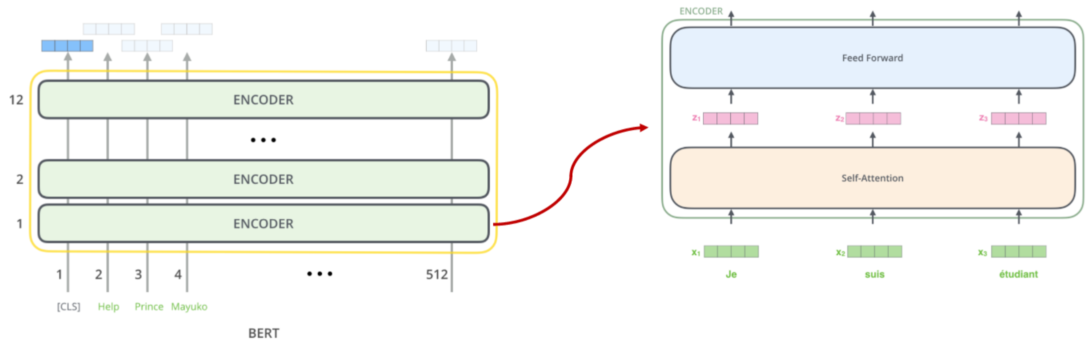

FIGURE 4.1: Bert Model Structure

There are two steps in the Bert framework: pretraining and fine-tuning. Unlabeled data was used on the Bert model to achieve pretraining tasks. Then for fine-tuning, reachers firstly initialize the model using parameters from pretraining step, finally modified these parameters with labeled data which is from specific task.

BERT's pretraining includes two tasks:Masked LM and Next Sentence Prediction. It proposes a new pre-training goal: masked language model (MLM) to overcome the unidirectional limitation mentioned above. MLM was inspired by the Cloze mission. MLM randomly masks some tokens in the input of the model. The goal is to predict the original vocabulary id based only on the context of the masked word. Different from left-to-right language model pre-training, the MLM target allows characterization to fuse the context of the left and right sides, thereby pre-training a deep two-way transformer. In addition to masking the language model, the author of this article also introduces a "next sentence prediction" task, which can pre-train the representation of text pairs with MLM. And the second pretraining task aims to let the model understand the relationship between sentences. In order to achieve this task, researchers chose two sentences A and B from corpus, 50% of the time B is the next sentence of A and 50% not.

Because the Transformers' self-attention, the fine-tuning process is straightforward.

4.1.2 DistilBert

As Transfer Learning from large-scale pre-trained models becomes more prevalentin Natural Language Processing (NLP), operating these large models in on-the-edge and/or under constrained computational training or inference budgets remainschallenging[8]. Because of the limitation of our PC, we selected DistilBert to get the embedding result. The authors of this paper took advantage of knowledge distillation in the pretraining step to reduce size of Bert by 40% but remaining 97% capabilities of language understanding and 60% faster.

4.2 Method used on French Corpus

4.2.1 Pre-trained spacy templates

fr_core_news_sm
fr_core_news_sm is a polyvalent model that performs many modifications on text such as tagging, parsing, lemmatization and named entity recognition. This model has been trained on a news corpus.

fr_dep_news_trf
The model *fr_dep_news_trf* performs the same operations, except the named entity recognition. It is trained by the Camembert[6] method which is an adaptation of the BERT algorithm (proposed by Google Inc) but based on the French language.

4.2.2 Dimension reduction

Dimension reduction with the PCA method consists of taking a vector with a very large number of dimensions (in our case a sentence vectorization), to look at the variation of the data contained in these vectors in order to merge the most similar dimensions until reaching the number of dimensions requested. This allows us to select the dimensions that contain the most information at the expense of those whose data has a more negligible variance. The *PCA* method proposed by the *Scikit-learn* library allows the implementation of this algorithm.

Chapter 5

Experiment

5.1 Classification Task

In this section, we use traditional statistical-based classification methods and neural network methods to classify different sentences into corresponding emotions. For statistical methods, we applied SVM, RandomForest and Logistic Regression. And for neural network, there are original network, CNN and RNN. Our goal is to compare the performance of different classification methods on the ISEAR dataset and find ways to improve performance.

First, we preprocess the ISEAR corpus, and then obtain the word vector and sentence vector through the DistilBert model, and finally we use different classification methods to complete the task of emotion recognition.

The sentence-vectors and word-vectors are both 768-dimension.

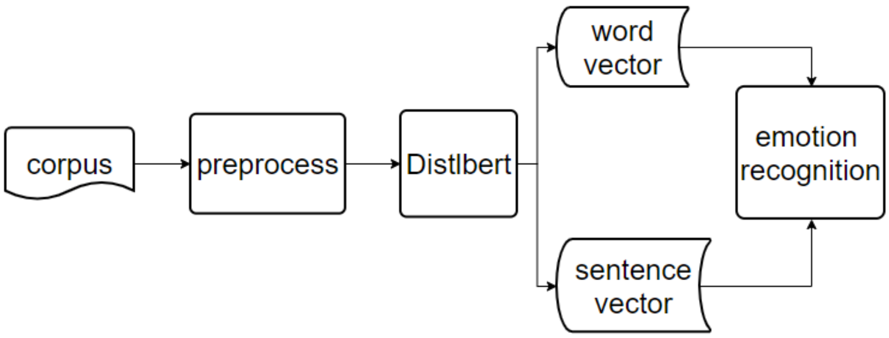

FIGURE 5.1: Experiment Process

5.1.1 Preprocess

For preprocessing, we first converted all letters to lowercase, and then remove all the stop words which are from NLTK and finally remove all the all non-Latin letters.

5.1.2 Statistical Methods

The method based on statistics is to use a given limited number of sample sets, under the conditions of the known statistical model of the research object or the known discriminant function class, according to certain criteria, the d-dimensional feature space is divided into c regions, each region corresponds to each category. When the pattern recognition system is working, as long as it judges which area the recognized object falls into, it can determine the category it belongs to. A classification system based on statistics should include preprocessing, feature extraction, and classifiers. The performence of different classifiers is shown below.

method	accuracy
SVM	0.53
Random Forest	0.45
Logistic Regression	0.52

TABLE 5.1: Accuracy of each Classifier

SVM

Support vector machine (SVM) is a classic classifier. It is a class classifier. The formal definition is a hyperplane that can separate samples of different classes in the sample space. In other words, given some labeled training samples (supervised learning), the SVM algorithm outputs an optimal separation hyperplane. The essence of the SVM algorithm is to find a hyperplane that can maximize a certain value. This value is the minimum distance between the hyperplane and all training samples. This minimum distance is called margin in SVM terminology. To summarize, the SVM classifier is the optimal segmentation hyperplane to maximize the interval of training data.

emotion	precision	recall	f1-score
anger	0.39	0.39	0.39
disgust	0.53	0.52	0.53
guilt	0.34	0.39	0.36
shame	0.43	0.38	0.40
joy	0.64	0.74	0.69
sadness	0.66	0.60	0.63
fear	0.69	0.67	0.68

TABLE 5.2: Performance of SVM

FIGURE 5.2: Confusion Matrix of SVM

According the experimental results, SVM has a better classification effect for joy, sadness and fear while for the other four emotions, its performence is poor especially anger. SVM confuses at anger and guilt.

Random Forest

Random forest is an ensemble algorithm based on decision tree classifiers. By combining multiple independent decision trees, the final prediction result is obtained according to voting or averaging. It is often higher than a single tree. Accuracy and greater stability. Compared with decision trees, the superior performance of random forest mainly depends on randomly sampled samples and features and integrated algorithms. The former makes it more stable against overfitting, and the latter makes it more accurate. We set 200 trees in the random forest.

emotion	precision	recall	f1-score
anger	0.33	0.28	0.30
disgust	0.44	0.48	0.46
guilt	0.34	0.37	0.35
shame	0.34	0.24	0.28
joy	0.55	0.69	0.61
sadness	0.59	0.55	0.57
fear	0.53	0.57	0.55

TABLE 5.3: Performance of Random Forest

We can see that Random Forest classifier performs bad except joy and sadness emotion. It can not distinguish correctly (anger, disgust) and hardly identify shame emotion.

FIGURE 5.3: Confusion Matrix of Random Forest

Logistic Regression

Logistic regression is a widely used classification machine learning algorithm that fits data to a logit function (or called logistic function), so as to predict the probability of an event.

emotion	precision	recall	f1-score
anger	0.40	0.37	0.39
disgust	0.50	0.49	0.49
guilt	0.38	0.41	0.39
shame	0.41	0.41	0.41
joy	0.67	0.73	0.70
sadness	0.63	0.62	0.62
fear	0.66	0.63	0.64

TABLE 5.4: Performance of Logistic Regression

The effect of Logistic Regression classifier is roughly the same as SVM, but has less power in distinguishing joy and sadness.

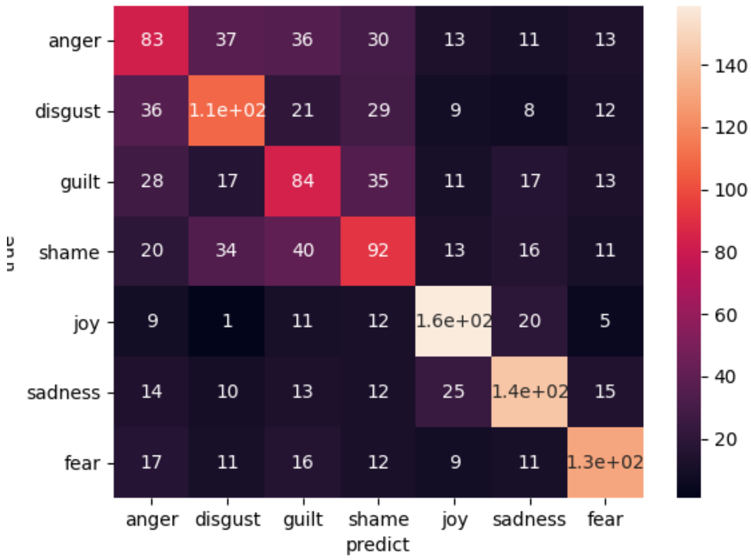

FIGURE 5.4: Confusion Matrix of Logistic Regression

5.1.3 Neural Networks

Neural network is a wide range of flexible nonlinear regression and discriminant models, data reduction models and nonlinear dynamic systems. They are composed of usually a large number of "neurons", that is, simple linear or non-linear computing elements, usually interconnected in complex ways, and usually organize large content. We set epoch equals 2 for every network.

method	accuracy
Linear network	0.3783
TextCNN(word-vector)	0.6216
TextCNN(sentence-vector)	0.6164
LSTM	0.6320

TABLE 5.5: Accuracy of each Classifier

Linear Network

We only use two linear layers and selected sentence-vector as the input of the model. The first layer is to convert the sentence vector to the same dimension, and the second is to map the sentence to corresponding emotion. The performance of Linear Network is very poor, it can hardly be used to do the task.

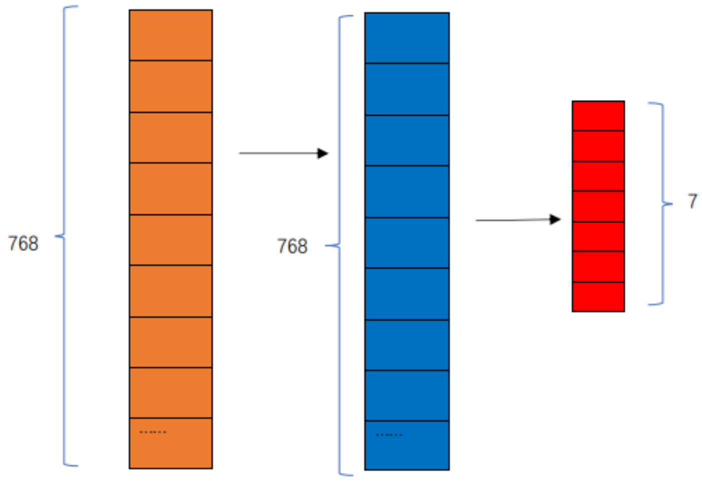

FIGURE 5.5: Linear Network Progress

CNN Network

Convolutional Neural Networks (CNN) was proposed by Yann Lecun of New York University in 1998. Its essence is a multilayer perceptron, which is a type of feedforward neural network that contains convolutional calculations and has a deep structure. . Convolutional neural network is a special multi-layer neural network. Like other neural networks, convolutional neural network also uses a back propagation algorithm for training.

We first tried the TextCNN method[5], created three different convolution kernels, the sizes of kernels are (n*768), n= 2,4,8. That is to say, the kernel scans n words each time. Then did the max pooling to make sure the dimensions of results from these three kernels are same. Finally, we spliced the three max pooled result into one vector and treated this vector as the input of linear layer to achieve classification task. The work progress of TextCNN is as shown.

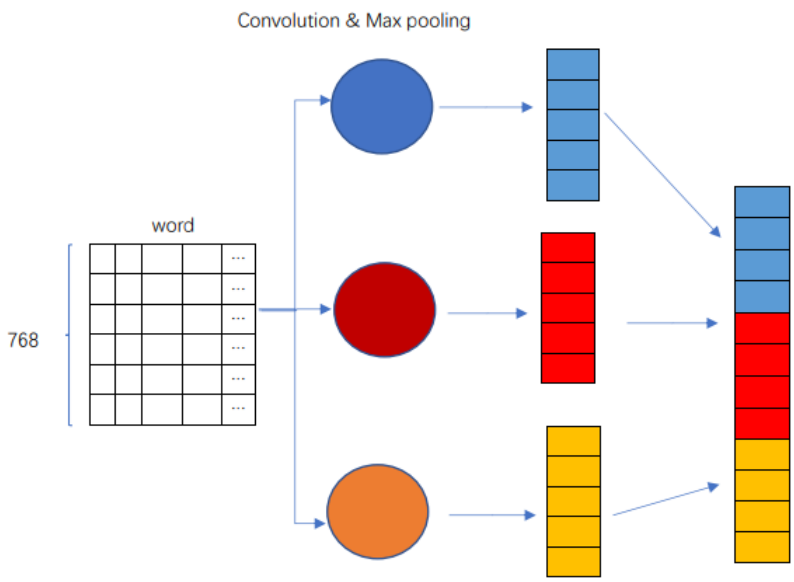

FIGURE 5.6: TextCNN Progress

emotion	precision	recall	f1-score
anger	0.46	0.65	0.54
disgust	0.65	0.65	0.65
guilt	0.53	0.46	0.49
shame	0.51	0.46	0.48
joy	0.75	0.79	0.77
sadness	0.74	0.62	0.68
fear	0.76	0.71	0.74

TABLE 5.6: Performance of TextCNN

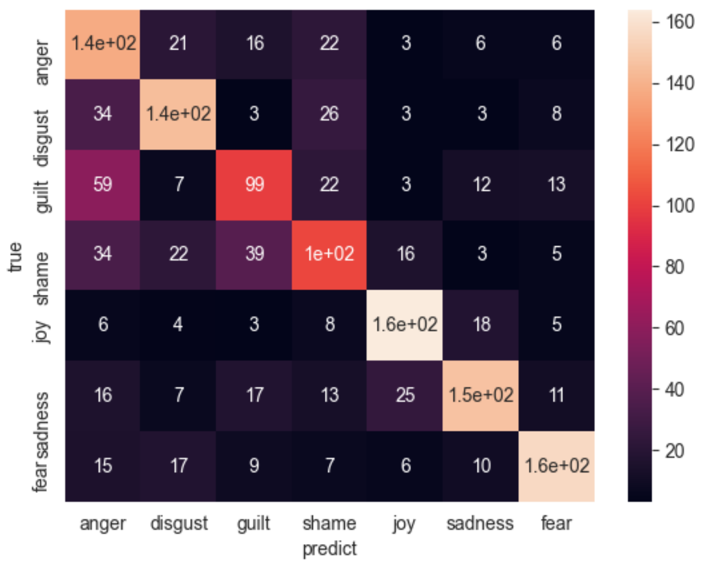

FIGURE 5.7: Confusion Matrix of TextCNN

The performance of TextCNN is much better than the previous statistics-based methods but confuses guilt and anger emotions.

We also tried to use the same idea for convolution on sentence vector, set up 3 one-dimensional convolution kernels, then performed convolution, max pooling and stitching to get final vector as the feature of classification task.

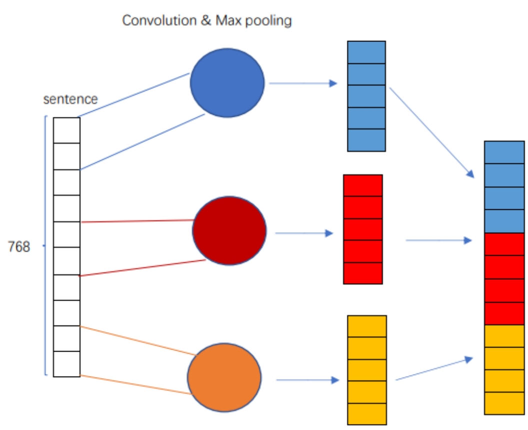

FIGURE 5.8: SentenceCNN Progress

The below table shows the performance of the TextCNN on sentence level.

emotion	precision	recall	f1-score
anger	0.49	0.53	0.51
disgust	0.49	0.53	0.51
guilt	0.49	0.53	0.51
shame	0.49	0.53	0.51
joy	0.74	0.79	0.77
sadness	0.74	0.79	0.77
fear	0.74	0.79	0.77

TABLE 5.7: Performance of TextCNN

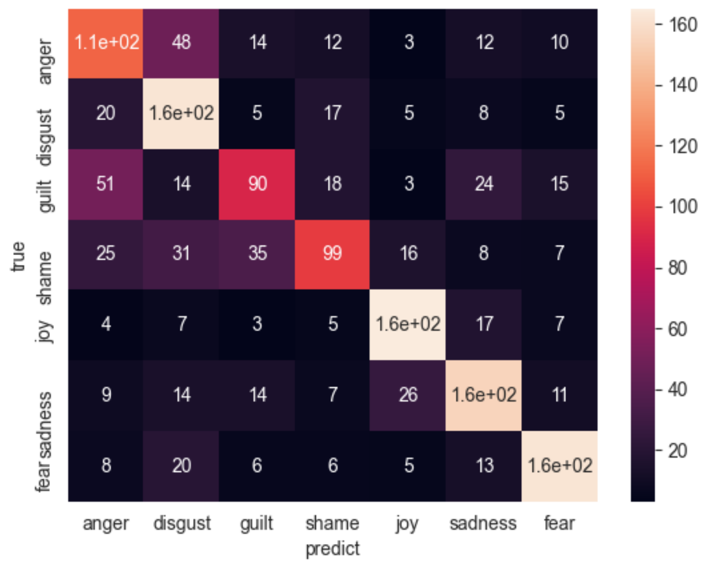

FIGURE 5.9: Confusion Matrix of SentenceCNN

We can see that the TextCNN model on the word level and sentence level obtained almost the same effect. Besides, sentence level TextCNN performanced better in distinguishing sadness emotion.

RNN Network

Recurrent Neural Network (RNN) is a neural network used to process sequence data. Compared with the general neural network, it can process the data of the sequence change. For example, the meaning of a word will have different meanings because of the different content mentioned above, and RNN can solve this kind of problem well.

Long short-term memory (Long short-term memory, LSTM) is a special kind of RNN, mainly to solve the problem of gradient disappearance and gradient explosion during long sequence training. Simply put, compared to ordinary RNNs, LSTM can perform better in longer sequences.

It is can be seen from the data that LSTM has the better performance in distinguishing disgust and guilt emotions.

emotion	precision	recall	f1-score
anger	0.56	0.52	0.54
disgust	0.57	0.70	0.63
guilt	0.53	0.58	0.55
shame	0.61	0.41	0.49
joy	0.69	0.86	0.76
sadness	0.71	0.64	0.67
fear	0.77	0.72	0.74

TABLE 5.8: Performance of LSTM

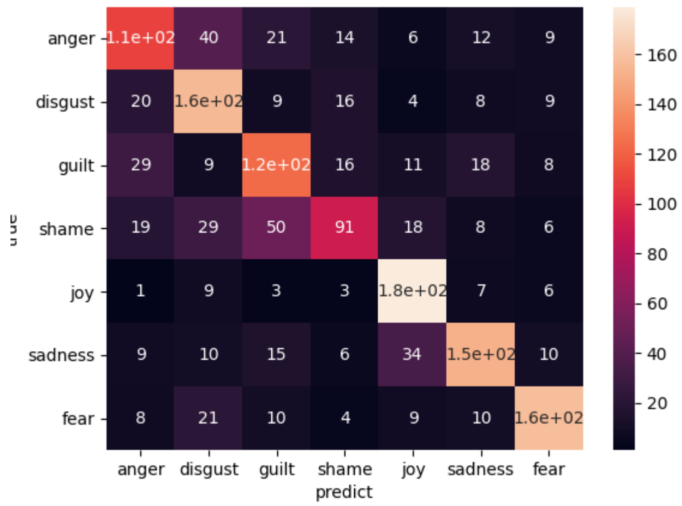

FIGURE 5.10: Confusion Matrix of LSTM

5.2 Clustering Task

We performed a vectorization of the data with the *Spacy* method, using two models *fr_core_news_sm* and *fr_dep_news_trf*. On the suggestions of our supervisor, we implemented a vector dimension reduction with the PCA method of *Scikitlearn* by setting the number of dimensions to 50. This was to avoid the scourge of dimensions[1]which would have considerably increased the computational time required to train the sentiment analysis model. The original vectors were kept in order to compare the representativeness of the data of each.

For this we had to limit ourselves to a randomly selected subset of 10,000 items from this corpus due to the limited computational capabilities of our machines.

To perform this test, we use the k-means[2]clustering method via the *scikit-learn* library. A clustering algorithm groups the data by trying to separate the samples into N groups of equal variance, minimizing intra-group differences. Each group is described by an average of these samples commonly called the "centroids", it does not correspond to a particular sample of the group but has a value numerically close to that of the group it defines. The K-means algorithm aims to choose centroids that minimize the Euclidean distances between each sample and its centroid.

We set the number of clusters to 25, the same number of emotions as the global classifier of the project.

We then performed a test using the *ScikitLearn* silhouette score to measure the consistency of the clusters, these results were used to make a decision on which vectorization method to propose for the sentiment analysis project. This *silhouette_score* method produces a score ranging from -1 to 1.

The results I obtained are not very good overall, but we can clearly see that the method based on Camenbert[6] with PCA (50 components) is the best, so it is the one we adapted for the classifier provided to us by our supervisor.

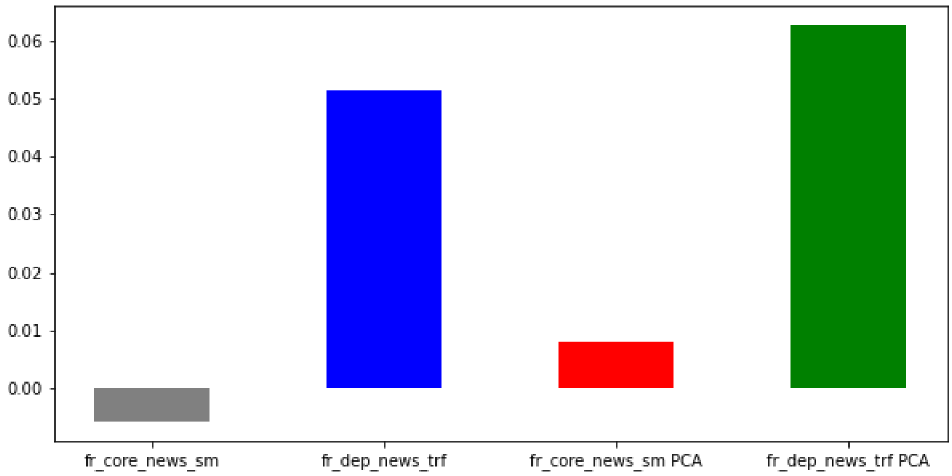

FIGURE 5.11: Clustering result

Chapter 6

Conclusion

Based on the ISEAR data set, we compared the performance of statistical classification methods and neural networks in emotion recognition. We used the DistilBert method to obtain word vectors and sentence vectors. Obviously, the recognition effect of neural network is better. For statistical classification methods, (anger, disgust) and (shame, guilt) are confused, it is difficult to distinguish. For neural networks, they are (guilt, anger) and (shame, guilt). For the French corpus, we achieved clustering task and the effect needs to continue to improve.

Bibliography

[1] Francisca Adoma Acheampong, Chen Wenyu, and Henry Nunoo-Mensah. "Text-based emotion detection: Advances, challenges, and opportunities". In: *Engineering Reports* 2.7 (2020), e12189.

[2] Erdenebileg Batbaatar, Meijing Li, and Keun Ho Ryu. "Semantic-emotion neural network for emotion recognition from text". In: *IEEE Access* 7 (2019), pp. 111866–111878.

[3] Taner Danisman and Adil Alpkocak. "Feeler: Emotion classification of text using vector space model". In: *AISB 2008 Convention Communication, Interaction and Social Intelligence*. Vol. 1. 2008, p. 53.

[4] Jacob Devlin et al. "Bert: Pre-training of deep bidirectional transformers for language understanding". In: *arXiv preprint arXiv:1810.04805* (2018).

[5] Yoon Kim. *Convolutional Neural Networks for Sentence Classification*. 2014. arXiv: 1408.5882 [cs.CL].

[6] Louis Martin et al. "CamemBERT: a Tasty French Language Model". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Online: Association for Computational Linguistics, July 2020, pp. 7203–7219. DOI: 10.18653/v1/2020.acl-main.645. URL: https://www.aclweb.org/anthology/2020.acl-main.645.

[7] Sreeja PS and G Mahalakshmi. "Emotion models: a review". In: *International Journal of Control Theory and Applications* 10 (2017), pp. 651–657.

[8] Victor Sanh et al. "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter". In: *arXiv preprint arXiv:1910.01108* (2019).

[9] Ashish Vaswani et al. "Attention is all you need". In: *arXiv preprint arXiv:1706.03762* (2017).

[10] Vel. *LaTeX Templates » Masters/Doctoral Thesis*. In collab. with Steve R. Gunn and Sunil Patel. Version 2.5. Aug. 27, 2017. URL: https://www.latextemplates.com/template/masters-doctoral-thesis.