Workshop on Linguistic and Neurocognitive Resources (LiNCr2020), pages 50–58
Language Resources and Evaluation Conference (LREC 2020), Marseille, 11–16 May 2020

Towards a Multi-Dataset for Complex Emotions Learning based on Deep Neural Networks

Belainine Billal, Sadat Fatiha, Boukadoum Mounir, Lounis Hakim
Computer Science, UQAM, Quebec, Canada
belainine.billal@courrier.uqam.ca,
{sadat.fatiha,boukadoum.mounir, lounis.hakim}@uqam.ca

Abstract

In sentiment analysis, several researchers have used emoji and hashtags as specific forms of training and supervision. Some emotions, such as fear and disgust, are underrepresented in the text of social media. Others, such as anticipation, are absent. This research paper proposes a new dataset for complex emotion detection using a combination of several existing corpora in order to represent and interpret complex emotions based on the Plutchik's theory. Our experiments and evaluations confirm that using Transfer Learning (TL) with a rich emotional corpus, facilitates the detection of complex emotions in a four-dimensional space. In addition, the incorporation of the rule on the reverse emotions in the model's architecture brings a significant improvement in terms of precision, recall, and F-score.

Keywords: Complex Emotions, Emotional Intelligence, Data Augmentation, Machine Learning, Natural Language Processing

1. Introduction

Several works in natural language processing (NLP) have addressed the recognition of expression of emotions. They can be divided into two approaches. The first one assesses emotions by using quantitative metrics such as the level of intensity or valence, arousal, domination, etc. For example, the emotion carried by a text is measured as very joyful, a little angry, fearful, etc., with the metric value referring to the degree of emotion (Posner et al., 2005). The second approach starts from a dictionary of basic emotions, considered as atomic and irreducible, to build more complex ones. This is the case of the Plutchik model (Plutchik, 1980), which allows to represent a complex emotion as a combination of several basic emotions (De Bonis, 1996).

Regardless of the approach used, a relevant corpus of examples is required for training and/or validation.
Many researchers have considered social media with emoji and hashtags as a source of training data. However, Some emotions, such as fear and disgust, are underrepresented in those media, and others such as anticipation are absent.

This research proposes the following contributions:

1. Construction of a novel annotated dataset for emotion-related work, created by mixing several existing corpora, that addresses the previous limitations. This annotated corpus is then used in a system designed to detect complex emotions based on the Plutchik model.
2. Introduction of a formal method for reading and interpreting complex emotions based on basic emotion vectors. This vector is reduced in a 4-dimensional space.
3. Introduction of a rule for reverse emotions in the model's architecture, stating that an emotion cannot be present at the same time as its opposite.

The structure of the present paper is described as follows: Section 2. introduces the Plutchik model in the context of this study, Section 3. surveys the state of the art on the analysis and detection of emotions, Section 4. describes our approach to the recognition of complex emotions with a deep neural network, Sections 5. and 6. describe the experiments that help evaluate our model and compare its performance to other models, along with an error analysis and a discussion. Finally, Section 7. concludes this work and offers perspectives for future research.

2. Overview of the Plutchik Theory

Plutchik (Plutchik, 2003) proposed a model based on a dictionary of emotions similar to the color dictionary. Indeed, since there are secondary colors derived from primary colors, there would be secondary emotions derived from primary emotions, and each combination of certain primary emotions can generate secondary emotions (Plutchik, 1980).
According to Plutchik (Plutchik, 1980), there are four pairs of opposite emotions: *(Joy, Sadness), (Trust, Disgust), (Fear, Anger), (Surprise, Anticipation)*. The eight dimensions of these fundamental emotions are adjacent and arranged like a cone, with the terms that designate the maximum intensity of each emotion at the top.
In relation to complex emotions that are added to the primary ones, first we can find the emotions that are a result of the combination of two adjacent emotions. These are the primary dyads (Plutchik, 2003). Moreover, there are emotions that are the result of a combination of two adjacent primary emotions, but separated by an emotion. These are the secondary dyads. Finally, the emotions that are the result of a combination of two adjacent primary emotions, but separated by two emotions, these are the tertiary dyads (Plutchik, 1980). Table 1 represents all possible combinations of the primary dyads, the secondary dyads, as well as the tertiary dyads, with the generated emotions according to the Plutchik model.

Primary Dyads	Results	Secondary Dyads	Results	Tertiary Dyads	Results
Joy + Trust	Love	Joy + Fear	Guilt	Surprise + Joy	Delight
Trust + Fear	Submission	Surprise + Trust	Curiosity	Sadness + Trust	Faintness
Surprise + Fear	Alarm	Sadness + Fear	Despair	Disgust + Fear	Shame
Surprise + Sadness	Disappointment	Surprise + Disgust	Horror	Surprise + Anger	Outrage
Sadness + Disgust	Remorse	Sadness + Anger	Envy	Sadness + Anticipation	Pessimism
Disgust + Anger	Contempt	Disgust + Anticipation	Cynicism	Disgust + Joy	Morbidity
Anticipation + Anger	Aggressiveness	Anger + Joy	Pride	Anger + Trust	Domination
Anticipation + Joy	Optimism	Anticipation + Trust	Fatalism	Anticipation + Fear	Anxiety

Table 1: Combinations of Plutchik's emotions (Plutchik, 2003).

3. Related Work

Because of the absence of annotated data, manually or otherwise, many NLP tasks related to sentiment analysis and emotion mining use co-occurring emotional expressions for remote supervision of social media, to allow models to learn directly useful textual representations before modelling these tasks (Mohammad et al., 2013; Nida et al., 2019).

3.1. Previous works on emotion recognition

Some works use binarized emojis as noisy labels (Read, 2005; Nakov et al., 2016; Yang et al., 2016; Nikhil and Srivastava, 2018), but emojis can be ambiguous as they can serve both as comments or to set emotional state of a text. This ambiguity was addresses by Kunneman et al. (2014) with emotional hashtags such as *#nice* and *#lame*. Nevertheless, DeepMoji has succeeded in showing that emoticons can be used to accurately categorize the emotional content of texts in many cases (Felbo et al., 2017). But DeepMoji requires more than one billion pieces of data for training (1 246 million of tweets), and it has two limitations: a) The analyzed text must contain emoticons; b) the emojis do not always reflect the emotional state behind the writing of the text, since they can also be used to complete the writing text.Other works use emotion theories such as Ekman's six basic emotions and Plutchik's eight basic emotions (Mohammad et al., 2013; Suttles and Ide, 2013; Felbo et al., 2017). The categorization is also done manually, and it requires requires an understanding of the emotional content of each expression, which is difficult and time-consuming for sophisticated combinations of emotional content.

The work of Suttles and Ide (2013) uses a binary classifier that indicates the existence of an emotion according to the representation of Plutchik. However, this method suffers from ambiguity when the emotion is presented with its opposite, for example the binary classification in a multi-label context can indicate *joy* and *sadness* at the same time, an impossible representation by Plutchik's theory.

The authors Felbo et al. (2017) used transfer learning (Bengio, 2012), which does not require access to the original dataset, but only to the model of an already trained deep learning classifier. This allowed them to classify *sarcasm* (Gal and Ghahramani, 2016) and the 7 emotions of the PsychExp dataset (Wallbott and Scherer, 1986). Others works using transfer learning (Barbieri et al., 2018; Gee and Wang, 2018; Park et al., 2018) demonstrated a great performance in detecting emojis in shared tasks such as SemEval[1].

The authors Barbieri et al. (2017) studied the relationship between words and emoticons. They also proposed an approach to predict the most likely emoji associated with a tweet. This proposed approach was based on a Bidirectional Long Short-Term Memory (BiLSTM) architecture (*BiLSTM*).

Zhong and Miao (2019) used a model that extends the Recurrent Convolutional Neural Network (RCNN) using finely-tuned external word representations and DeepMoji phrase representations on the emotion detection task in *SemEval-2019*.

Other work (Tang et al., 2014) proposed a method to learn to incorporate specific words in Word Embeddings and showed an improvement in the performance especially when combining other sets of existing features.

In our knowledge, none of the previous works considered the case of texts with conflicting emotions, hence the need for such a model.

3.2. Datasets Overview

In this section, we present the existing emotional English datasets in chronological order.

The dataset ISEAR, published by (Scherer and Wallbott, 1994) uses the responses of people from different cultures to questionnaires in social media. The final dataset contains about 3,000 reports, for 7,665 sentences labeled with unique emotions. The set uses the labels "joy", "fear", "anger", "sadness", "disgust", "shame" and "guilt".

The WordNet-Affect Lexicon (Valitutti, 2004) is a collection of emotion related words (nouns, verbs, adjectives, and adverbs), classified as "Positive", "Negative", "Neutral", or "Ambiguous", and categorized into 28 subcategories ("Joy", "Love", "Fear", etc.).

The dataset Tales, published by Alm et al. (2005; Bostan and Klinger (2018) is based on literature and consists of 15,302 sentences, with its annotators only agreeing on 1,280 sentences. The goal of this resource is to help build emotion classifiers for literature. The annotation scheme includes Ekman's six basic emotions. Labels 'angry' and 'disgust' are merged.

The dataset AffectiveText, published by Strapparava and Mihalcea (2007; Bostan and Klinger (2018), is built from news headlines. The main objective of this resource is the classification of emotions and valence in news headlines using the basic emotions of Ekman, supplemented by enumerate valence between 0 to 100.

The dataset Blogs, published by Aman and Szpakowicz (2007), includes 5,205 sentences. Each instance annotated with one label. The used annotation scheme corresponds to Ekman's six fundamental emotions.

[1] (International Workshop on Semantic Evaluation)

The dataset EmoTxt , published by Ortu et al. (2015), includes 4000 comments posted by software developers. This corpus contains sentences manually labelled with the emotions "Love", "Joy”, "Surprise", "Anger", "Sadness" and "Fear".

The dataset Electoral-Tweets, published by Mohammad and Kiritchenko (2015) for the field of elections, contains more than 100,000 responses to two detailed online questionnaires (questions focused on the emotions, purpose, and style of the electoral tweets). These tweets are annotated via Crowdsourcing and the labels for emotions are nonstandard, examples: polite,impolite . The tweets are annotated with emotional words (Bostan and Klinger, 2018).

The dataset Emotion-Stimulus, published by Ghazi et al. (2015), contains 820 sentences that are annotated with both emotions and their causes, and 1,549 sentences that are uniquely marked with emotions. The annotators used FrameNet (Fillmore et al., 2003) to annotate this dataset using the Ekman’s theory alimented with the Shame label.

The dataset fb-valence-eveal, published by Preoţiuc-Pietro et al. (2016), is a data set of 2,895 Social Media posts rated by two psychologically-trained annotators on two separate ordinal nine-point scales. These scales represent valence (or sentiment) and arousal (or intensity), which defines each post’s position on the circumplex model of affect, a well-established system for describing emotional states.

The dataset Grounded-Emotions, published by Bostan and Klinger (2018), is built on tweets and contains 2,557 instances published by 1,369 users. The labels is "happy" and "sad". The tweets are annotated by the authors.

The dataset TEC, published by Mohammad et al. (2013)(Bostan and Klinger, 2018), includes 21,051 tweets. The main objective of this resource is to use emotion word hashtags as a source of annotation for emotions. The annotation scheme corresponds to Ekman’s basic emotion model. They collected tweets with hashtags corresponding to Ekman’s six basics emotions: anger, disgust, fear, happy, sadness, and surprise.

The dataset DailyDialogs, published by Li et al. (2017), is based on conversations and includes 13,118 sentences. The annotation used is from Ekman, with a label of "no emotion". A single label by utterance via an expert annotation. This dataset contains annotations about the user’s intent and the topic of the dialog.

The dataset EmoBank, published by Buechel and Hahn (2017), is based on several genres and domains. It consists of 10,548 sentences, each one annotated manually according to the emotion expressed by the author and the readers.

The dataset EmoInt, published by Mohammad and Bravo-Marquez (2017) (Bostan and Klinger, 2018), consists of 7,097 tweets. It associates each text with different intensities of emotion. The tweets are annotated via crowdsourcing with intensities of anger, joy, sadness, and fear.

As the previous list shows, there exist many emotional data set to work with. However, they all have the following limitations with regards to Plutchik’s theory : 1) The labels are not based on the fundamental emotions of Plutchik’s theory; 2) the size of the data may be too small to train an efficient emotion detection model.

Plutchik’s theory offers many advantages for the detection of complex emotions. 24 complex emotions can be modeled with just 8 basic emotions; while, the model proposed by Ekman offers 16 complex emotions, and needs a larger data set for its implementation(Ekman, 2004). Our motivation in the present research is to use the Plutchik’s theory for the detection of basic and complex emotions. For this purpose, a new dataset was constructed and annotated with the complex emotions.

4. The Proposed Approach

Our ultimate goal is to create an emotion classifier that is capable of detecting complex emotions based on the Plutchik model, and that introduces an implicite rule to handle conflicting emotion representations. This rule forces the classifier to detect either an emotion like joy or sadness but not both at the same time.

The overall process is summarized in Figure 1 and consists in three phases :

(1) collection of annotated data to construct a training annotated corpus from several types of corpora in order to cover the eight basic labels of the Plutchik theory;

(2) detection of basic emotions and representation with a four-dimensional emotion vector. The proposed strategy for the emotion detection relies on multi-label classification using transfer learning (Felbo et al., 2017);

(3) learning and interpretation of complex emotions using multi-label classification.

4.1. Corpus Construction

Our training corpus combines several English data sets from different sources. Table 2 represents the details of the source and types of labels considered. As the table shows, all eight basic emotions according to Plutchik’s theory are considered, plus three complex emotions that are generally associated with our model. For instance, we break down complex emotions *Love* into the basic emotions *Trust* and *Joy* in the whole corpus. By repeated the operation for all complex emotions, the initial corpus becomes a multi-label one. Table 2 shows the different corpora, as components of the data set used in this research to detect the basic and complex emotions. These corpora are described in the following paragraphs.

Dataset	Labels
EmoTxt [1]	joy, anger, sadness, love, surprise, fear
PsychExp [2]	joy, fear, anger, sadness, disgust, shame, guilt
DailyDialog [3]	no emotion, anger, disgust, fear, happiness, sadness, surprise
NRC_Emotion_Lexicon_ v0.92 [4] emotion_proposition_store [5]	joy, fear, disgust, anger, sadness, surprise, trust and anticipation
WordNet-Affect (Valitutti, 2004)	joy, fear, disgust, anger, sadness, surprise, trust and anticipation

Table 2: Sources of each component of the data set

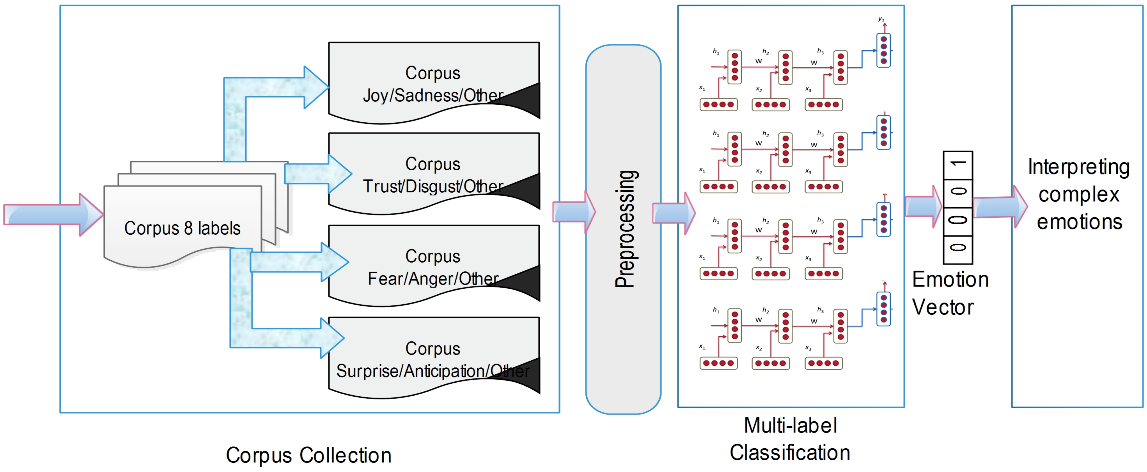

Corpus 8 labels
Corpus Joy/Sadness/Other
Corpus Trust/Disgust/Other
Corpus Fear/Anger/Other
Corpus Surprise/Anticipation/Other
Corpus Collection
Preprocessing
Multi-label Classification
0 0 0 1
Emotion Vector
Interpreting complex emotions

Figure 1: General presentation of the proposed method

Dailydialog (Li et al., 2017) is annotated with the Big Six emotions of Ekman, and it is a multi-turn corpus built for human dialogue. We extracted sentences containing between 5 and 12 words, and deleted the sentences that do not contain emotions in the big Six of Ekman, since they can have emotions that can be represented by the Plutchik model but are absent in Ekman model.

Wordnet samples will help us generate the missing labels in other corpora such as Surprise and Anticipation. In addition, we enriched our corpus with the sources of WordNet and WordNet-Affect (Valitutti, 2004) as follows:

First, we have extracted all the effective examples of WordNet that have a word-annotated relationship in WordNet-Affect. Second, we manually annotated the examples using Crowdsourcing, three users chose emotions that correspond to the 8 basic plutchik emotions.

Then,we choose only the examples to the three evaluators agree on the same emotions.

Word	Examples	label WordNet-Affect
love	She loves her boss and works hard for him	Joy + Trust
love	he has a very complicated love life	Joy + Trust
sad	feeling sad because his dog had died	Sadness
surprise	The news really surprised me	Surprise

Table 3: Examples of annotated WordNet data where the three annotators agreed

Table 3 presents some examples where all of the annotators agreed on the same label.

Table 4 presents the complex emotions that exist in our corpus and that we replaced by the corresponding basic emotions in the Plutchik model. *Love* represents the Primary Dyads, *Guilt* represents Secondary Dyads, and *Shame* represents Tertiary Dyads. Moreover, we augmented our corpus with words associated with the emotions extracted from Wordnet-Affect. Thus, all examples associated with these words have the same affect. Hence, the emotions associated with these words reflect the emotions already present in the examples used in Wordnet.

Complex emotion	Basic emotion	Composition type
Love	Joy + Trust	Primary Dyads
Guilt	Fear + Joy	Secondary Dyads
Shame	Fear + Disgust	Tertiary Dyads

Table 4: Decomposition of complex emotions

The corpus is divided into four sub corpora, each one making use of three labels for emotion representation, its opposite and the absence of the two (e.g., Joy/Sadness/No, Anticipation/Surprise/No, Disgust/Trust/No, Anger/Fear/No). then, The instances of each sub corpus are mixed randomly. We divided each sub-corpus into three parts related to: (1) training 70%, (2) development 15% and (3) testing 15%. We used the same validation process as the one used for DeepMoji (Felbo et al., 2017), using the provided code [2]. The DeepMoji model uses an embedding layer of 256 dimensions to represent each word in a vector space model. A hyperbolic tangent activation function is used to enforce a constraint of each embedding dimension being within [-1, 1]. To capture the context of each word, DeepMoji uses two bidirectional LSTM layers with 1024 hidden units in each (512 in each direction). Finally, the attention mechanism lets the model decide the importance of each word for the prediction task by the projection on 64 outputs of emojis. Our model uses the same architecture with changing the output layer to 3 outputs.

The test phase is done after the generation of the final model. Table 5 represents the statistics by the average number of words per sentence for each label that exist in the corpus. Figure 2 illustrates the distribution of the eight emotions in our corpus by percent.

[1] https://github.com/collab-uniba/EMTk

[2] https://github.com/bfelbo/DeepMoji/tree/master/data

[3] http://yanran.li/dailydialog

[4] Lexicon of the NRC Word Emotion Association.

[5] https://github.com/sebastianruder/emotion_proposition_store

[2] https://github.com/huggingface/torchMoji

Emotion sub data	Train	Eval	Test	Total	Average Sentence Length
Anticipation	1572	336	336	2246	6.1
Joy	5640	1208	1208	8058	6.7
Trust	1653	354	354	1653	7.1
Fear	5859	1255	1255	8370	7.3
Surprise	1317	282	282	1881	6.2
Sadness	3357	719	719	4795	7.4
Disgust	4067	871	871	5810	7.2
Anger	2231	478	478	3188	6.9

Table 5: Statistics by number of labels in the corpus

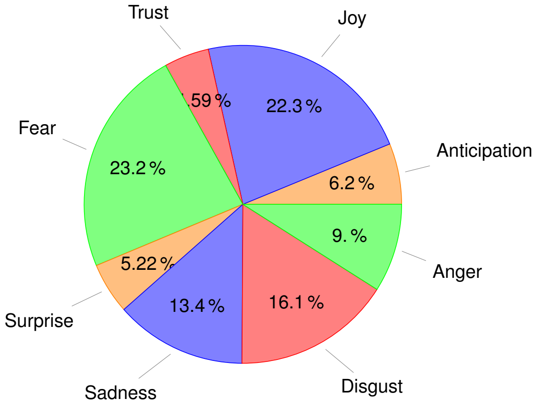


Figure 2: Distribution of emotions in the corpus

4.2. Using the corpus for emotions detection

In the case of the presence of the emotion, we mark 1 and in the case of the presence of the opposite emotion, we mark -1. If the emotion with its inverse are absent we mark 0. Our main objective is to avoid having an emotion with its opposite at the same time, either 1 or -1. In addition, if the model detects 0, then we have no emotion.

With the proposed corpus, the emotion recognition problem can be seen as a problem of learning multi-labels where each of the four dimensions is represented by a label with three values (1,0,-1), each label detected by a Sequence to Vector model (Seq2vec). The seq2vec model used is DeepMoji, as shown in figure 3.

To detect each label, we use transfer learning of a DeepMoji model shown in the figure 3.

DeepMoji model is learnt on 50,000 words of inputs and 65 outputs that correspond to emojis. The model contains two BiLSTM layers that can learn the sequential structure of the sentence. These two layers were kept during the transfer learning. On the other hand, the layers of the attention and the output are replaced by a layer of three outputs.

Our modelling of emotional states is based on representing of emotional states in the form of vectors. For each emotional state, there is a vector in a 4-dimensional space, each dimension representing a pair of contradictory basic emotions (eg. Joy and Sadness and No).

We propose to use the same basic emotions of the Plutchik model to define the dimensions of our base. Therefore, the number of dimensions of our basic emotion is four pairs of emotions and is formally defined by the base B = ((Joy, Sadness), (Trust, Disgust), (Fear, Anger), (Surprise, Anticipation)). Thus, any emotion can be realized using a combination of the other fundamental emotions that define our base B. Our model represents the following axes, as defined in Table 6:

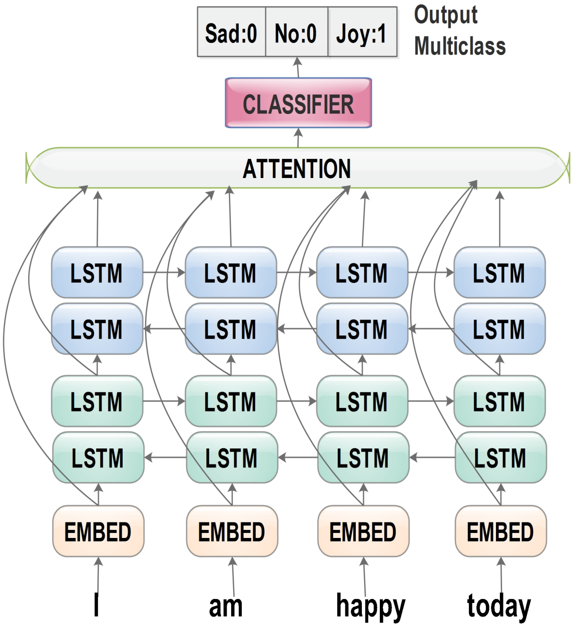


Figure 3: The architecture used to transfer learning based on the DeepMoji model for each classifier multiclass.

Positive axis(+)	Negative axis(-)
Joy	Sadness
Trust	Disgust
Fear	Anger
Surprise	Anticipation

Table 6: Combinations of two by two conflicting emotions in 4 dimensions

Each basic positive emotion is in the interval [0,1] and every basic negative emotion is in the interval [-1,0].

This allows on the one hand to represent an infinite number of complex emotions, because our model is a continuous one, and on the other hand, to offer high-performance mathematical tools for the analysis and processing of these emotions.

4.3. Learning complex emotions

Table 8 shows a representation of primary complex emotions using the Plutchik model, with the combinations of 2 adjacent emotions separated by no emotion constituting the primary dyads.

Table 7 shows a representation in 8 dimensions equivalent to Table 8. The latter represents the emotion in 4 dimensions and prevents the representation of the emotion with his inverse that will serve as a transition matrix W to detect the main complex emotions.

The numerical contents of Table 8 are used as a transition matrix W to detect complex emotions for primary dyads. To this end, we converted each type of dyad in table 1 into a W transition matrix.

Complex emotions Primary Dyad	Anticipation	Joy	Trust	Fear	Surprise	Sadness	Disgust	Anger
Optimism	1	1	0	0	0	0	0	0
Love	0	1	1	0	0	0	0	0
Submission	0	0	1	1	0	0	0	0
Apprehension	0	0	0	1	1	0	0	0
Disappointment	0	0	0	0	1	1	0	0
Remorse	0	0	0	0	0	1	1	0
Contempt	0	0	0	0	0	0	1	1
Aggressiveness	1	0	0	0	0	0	0	1

Table 7: Combinations of 2 adjacent emotions that make the primary dyads in 8 dimensions.

Complex emotions Primary Dyad	Anticipation-Surprise	Joy-Sadness	Trust-Disgust	Fear-Anger
Optimism	1	1	0	0
Love	0	1	1	0
Submission	0	0	1	1
Apprehension	-1	0	0	1
Disappointment	-1	-1	0	0
Remord	0	-1	-1	0
Contempt	0	0	-1	-1
Aggressiveness	1	0	0	-1

Table 8: Combinations of 2 adjacent emotions that make the primary dyads in 4 dimensions.

Equations 1 and 2 show how one can detect the presence of a complex emotion by multiplying matrix W by the vector V that represents the emotion coordinates in our vector space (equation 1).

$$\underset{Primary\ Dyad}{S} = \underset{Primary\ Dyad}{W} V = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ -1 & 0 & 0 & 1 \\ -1 & -1 & 0 & 0 \\ 0 & -1 & -1 & 0 \\ 0 & 0 & -1 & -1 \\ 1 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} \quad (1)$$

The result for the complex emotion obtained should be the result that maximizes a component of the vectors. A problem that can be faced is that the components can exceed the value 1. To fix this problem, we propose to seek the value greater than 1. Does it mean to convey that no complex emotion is detected when $S_i < 1$.

Equation 2 presents our objective function for reading the complex emotion. The complex emotions generated by the index i correspond to the emotions in the transition matrix W given in table 8.

$$\begin{cases} Emotion\hat{\ }complex & = \underset{i}{argmax}(S_i) \\ \quad and & \\ \quad S_i \geq 1 & \end{cases} \quad (2)$$

i∈ (*Optimism* =0, *Love* =1, *Submission* =2, *Alarm*=3, *Disappointment*=4,*Contemptment*=5, *Remord*=6, *Aggressiveness*=7)

5. Experiments and Results

We conducted two sets of experiments. The first experiments considered the emotion space in four dimensions, each one having three labels that reflect the presence of an emotion and its inverse, or the absence of both. As a result, the classifiers consider the vector of four labels: *Joy/Sadness/No* , *Trust/Disgust/No*, *Anticipation/Surprise/No*, *Anger/Fear/No*.

The second experiments turn the problem into binary classification, we modeled as the baseline approach. This method, called the binary relevance method, models the emotion space in 8 dimensions, each one having two classes that reflect the presence of emotion and its absence. Thus, The classifiers consider the vector of 8 labels: *Joy/No* ,*Sadness/No*, *Trust/No*, *Disgust/No*, *Anticipation/No*, *Surprise/No*, *Anger/No*, *Fear/No*.

Both sets of experiments are based on transfer learning and can be represented by table 9.

Model	Axis Emotions	Recall	Precision	F1	Macro F1	Exact Match
Our Model in 4 dimensions space	joy/sadness/No	0.56	0.44	**0.49**	**0.54**	**0.43**
	anger/fear/No	0.61	0.56	**0.58**		
	surprise/anticip/No	0.55	0.51	**0.52**		
	trust/disgust/No	0.63	0.59	**0.57**		
Our Model in 8 dimensions space	joy/No	0.48	0.41	0.44	0.46	0.23
	sadness/No	0.46	0.39	0.42		
	anger/No	0.51	0.47	0.48		
	fear/No	0.52	0.44	0.47		
	surprise/No	0.46	0.42	0.43		
	anticipation/No	0.45	0.39	0.41		
	trust/No	0.54	0.49	0.51		
	disgust/No	0.57	0.48	0.52		

Table 9: Results based on precision, recall, F-score for different classifications after using transfer learning.

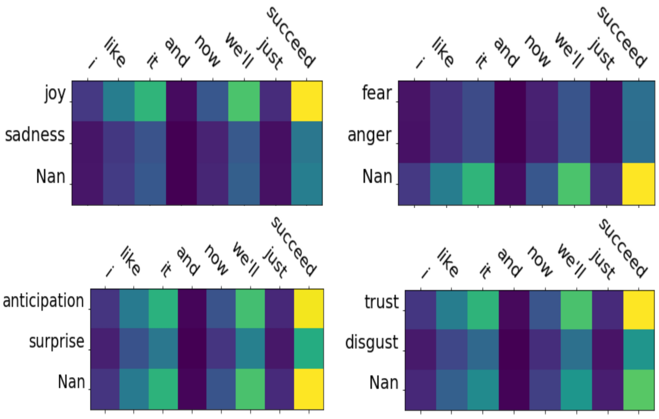


Figure 4: Visualization of the attention for each Multi Class classifier with example '*I like it and now we'll just succeed*'.

Table 9 provides the obtained Precision, Recall, F1 and Macro-F1 values of the model trained with transfer learning, comparing the use four-dimensional space and eight-dimensional space representations.

Figure 4 illustrates four attentions, each one detected by one classifier. They represent the score of participation of each word in the example above, with the model detecting the associated class. The yellow color represents a high probability of contribution, whereas the blue color represents a low probability of contribution. The classifiers *Joy/Sadness* and *anticipation/surprise* identified the labels *Joy*, *Trust* and *Anticipation*. The classifiers represent the absence of other emotions with the label *Nan*. The complex emotion detected in this example is *Optimism*, *Fatalism*, and *Love*, as *Joy+Anticipation=Optimism*,*Trust+Anticipation=Fatalism* and Joy+Trust=Love.

5.1. Comparison with other models

As our model appears to be the first to apply the Plutchik model to a text with conflicting emotions, a precise comparison with other works is not possible. However, as there exit methods that attempt to detect complex emotions by direct means directly such as PsychExp and EmoTxt, a qualitative comparison may give insight into the strengths and weaknesses of the different methods.

Model	Complex Emotion	Average-F1	Exact Match
Our Model	Love (Joy + Trust)	**0.58**	0.52
	Shame (Fear + Disgust)	**0.54**	0.53
	Guilt (Fear + Joy)	**0.54**	0.51
Model	Complex Emotion	F1	Accuracy
DeepMoji(**PsychExp**)	Shame	0.56	**0.59**
	Guilt	**0.54**	**0.60**
DeepMoji (**PsychExp + EmoTxt**)	Love	0.57	**0.63**
	Shame	0.53	**0.58**
	Guilt	0.51	**0.58**

Table 10: Results based on Exact Match, F-score for Love, Guilt and, Shame classification after using Transfer Learning.

Table 10 presents a comparison with the state of the art, which uses public data sets that contain some complex emotions. The *EmoTxt* dataset contains a test with 200 instances of the labels '*Love*' and the PsychExp dataset contains a test with 264 and 427 instances of labels *Guilt* and *Shame*, respectively.

For the first model, we used the DeepMoji model (Felbo et al., 2017) with the *PsychExp* data set, and for the second, we added the *Love* label to the model after training it with *PsychExp* and *EmoTxt* dataset. The *Love* label represents the *Joy* + *Trust* detection found in the Primary Dyads. The *Shame* label represents the '*Fear* + *Disgust*' detection found in the Tertiary Dyads. The *Guilt* label represents the '*Fear* + *Joy*' detection that is in the Secondary Dyads.

6. Discussion

The analysis of our experiments, we notice a correlation between the different loss estimates illustrated in figure 5. An inverse relationship can be detected between the loss and the results shown in table 9: the more we reduce the loss the more we increase the F1 score. In addition, we can notice that the duration of learning depends on the size of the data. Moreover, the convergence towards the local minima collapses quickly, because the DeepMoji parameters used are using Transfer Learning.

The obtained results also reveal a slight difference between the different experiments in table 10. Indeed, the average F1 score of our model for label *Guilt (Fear + Joy)* is greater then the F1 score of the experiment done by the DeepMoji model (PsychExp), but the DeepMoji model accuracy exceeds the Exact match (subset accuracy) of our model by 0.07, because the exact match means that both labels detect it at the same time.

Our model has a better performance in terms of average F1 score for the label *Love (Trust + Joy)* when compared to the DeepMoji model (PsychExp + EmoTxt) which contains the *love* label. However, the accuracy of DeepMoji (PsychExp + EmoTxt) is better than the Exact match of our model.

Table 9 also reveals obvious difference between models. The F1-score in the experiment *Joy/Sadness* improves to 5% (from 0.44 to 0.49) due to the incorporation of the reverse emotions rule, which imposes that the presence of an emotion excludes the existence of its inverse.

Figures 6b and 6a are attention heat maps for two sentences. The first one is an affirmative sentence, '*I am happy*', and it is classified by the label Joy; the second one, ' textit I am not happy', is its negative sentence and is classified by the label Sadness.

The classifiers detect the labels (Sadness, Joy) through the yellow boxes. The words that caused the detection of sadness are '*not happy*' and the word that caused the detection of joy is '*happy*'. However, the word '*I*' participates less in the generation of emotion, this can be explained by the fact that the words '*happy*' and '*not happy*' are subjective words. The word '*am*' has a weak intensity represented by the blue box. It does not contribute to the generation of the emotion, because it is objective and can be replaced by another entity without affecting the subjectivity of the sentence.

The comparison between the attentions of the figures 6b and 6a illustrates the independence of the emotion from the vocabulary. The replacement of the sentence '*I am happy*' by '*I am not happy*' shows that the system learnt an interesting rule as follows: the reversal of sentences by negation involves the reversal of the *Joy* emotion by the *Sadness* emotion.

The labels *Love*, *Guilt*, and *Shame* in Table 10 represent the detection of Primary Dyads, Secondary Dyads, and Tertiary Dyads, which confirms that our hypothesis worked well on these three test labels with a Macro-F1 exceeding 50%.

7. Conclusions and perspectives

This paper presents a novel approach for the detection of complex emotions, according to the Plutchik model and using multi-label classifiers. These classifiers are divided into 4 multiclass classifiers. Our main contributions are listed as follows:

(1) A new corpus labeled by the 8 basic emotions of Plutchik.

(2) Representation of complex emotions according to the Plutchik theory, in a vector space with four axes.

(3) Learning new rules that the detected emotions do not show up using their inverse emotions in the same axis.

To our knowledge, there exist no previous efforts to automatically detect and recognize complex emotions which was introduced by Plutchik's theory, in a textual data using four dimensions and deep neural networks.

Our proposed research is a crucial step towards building a conversational agent endowed with emotional intelligence. We are also looking forward to transferring the idea of complex emotions to task-oriented dialogs and multi-turn dialog generation problems.

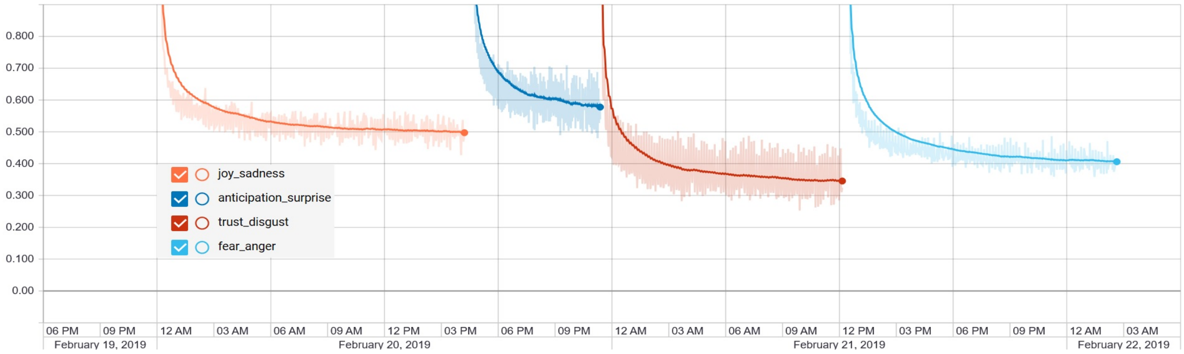
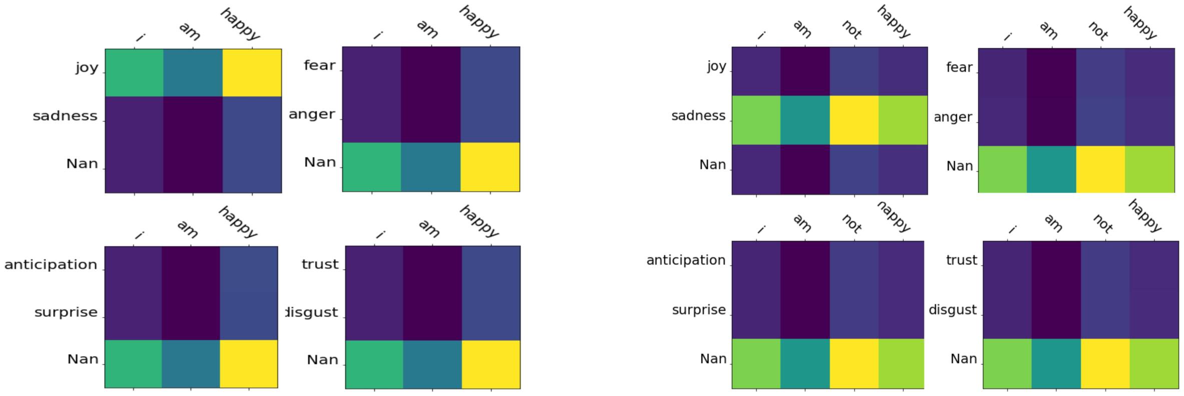


Figure 5: Visualization of loss reduction for each classifier Multi Class in evaluation process.



(a) Visualization of the attention for example '*I am happy*' (b) Visualization of the attention for example '*I am not happy*'.

Figure 6: Visualization of the attention mechanism.

8. Bibliography

Alm, C. O., Roth, D., and Sproat, R. (2005). Emotions from text: Machine learning for text-based emotion prediction. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, HLT '05, pages 579–586, Stroudsburg, PA, USA. Association for Computational Linguistics.

Aman, S. and Szpakowicz, S. (2007). Identifying expressions of emotion in text. In *International Conference on Text, Speech and Dialogue*, pages 196–205. Springer.

Barbieri, F., Ballesteros, M., and Saggion, H. (2017). Are emojis predictable? *CoRR*, abs/1702.07285.

Barbieri, F., Camacho-Collados, J., Ronzano, F., Anke, L. E., Ballesteros, M., Basile, V., Patti, V., and Saggion, H. (2018). Semeval 2018 task 2: Multilingual emoji prediction. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 24–33.

Bengio, Y. (2012). Deep learning of representations for unsupervised and transfer learning. In *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, pages 17–36.

Bostan, L.-A.-M. and Klinger, R. (2018). An analysis of annotated corpora for emotion classification in text. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 2104–2119, Santa Fe, New Mexico, USA, August. Association for Computational Linguistics.

Buechel, S. and Hahn, U. (2017). EmoBank: Studying the impact of annotation perspective and representation format on dimensional emotion analysis. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 578–585, Valencia, Spain, April. Association for Computational Linguistics.

De Bonis, M. (1996). *Connaître les émotions humaines*, volume 212. Editions Mardaga.

Ekman, P. (2004). What we become emotional about. In *Feelings and emotions. The Amsterdam symposium*, pages 119–135.

Felbo, B., Mislove, A., Søgaard, A., Rahwan, I., and Lehmann, S. (2017). Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. pages 1615–1625, September.

Fillmore, C. J., Johnson, C. R., and Petruck, M. R. (2003). Background to framenet. *International journal of lexicography*, 16(3):235–250.

Gal, Y. and Ghahramani, Z. (2016). A theoretically grounded application of dropout in recurrent neural networks. In *Advances in neural information processing systems*, pages 1019–1027.

Gee, G. and Wang, E. (2018). psyml at semeval-2018 task 1: Transfer learning for sentiment and emotion analysis. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 369–376.

Ghazi, D., Inkpen, D., and Szpakowicz, S. (2015). Detecting emotion stimuli in emotion-bearing sentences. In Alexander Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing*, pages 152–165, Cham. Springer International Publishing.

Kunneman, F., Liebrecht, C., and van den Bosch, A. (2014). The (un) predictability of emotional hashtags in twitter. In *Proceedings of the 5th Workshop on Language Analysis for Social Media (LASM)*, pages 26–34.

Li, Y., Su, H., Shen, X., Li, W., Cao, Z., and Niu, S. (2017). Dailydialog: A manually labelled multi-turn dialogue dataset. In *Proceedings of The 8th International Joint Conference on Natural Language Processing (IJCNLP 2017)*.

Mohammad, S. M. and Bravo-Marquez, F. (2017). Emotion intensities in tweets. *CoRR*, abs/1708.03696.

Mohammad, S. M. and Kiritchenko, S. (2015). Using hashtags to capture fine emotion categories from tweets. *Computational Intelligence*, 31(2):301–326.

Mohammad, S. M., Kiritchenko, S., and Zhu, X. (2013). Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets. *arXiv preprint arXiv:1308.6242*.

Nakov, P., Ritter, A., Rosenthal, S., Sebastiani, F., and Stoyanov, V. (2016). Semeval-2016 task 4: Sentiment analysis in twitter. In *Proceedings of the 10th international workshop on semantic evaluation (semeval-2016)*, pages 1–18.

Nida, H., Mahira, K., Mudasir, M., Mudasir Ahmed, M., and Mohsin, M. (2019). Automatic emotion classifier. In Bibudhendu Pati, et al., editors, *Progress in Advanced Computing and Intelligent Engineering*, pages 565–572, Singapore. Springer Singapore.

Nikhil, N. and Srivastava, M. M. (2018). Binarizer at semeval-2018 task 3: Parsing dependency and deep learning for irony detection. *CoRR*, abs/1805.01112.

Ortu, M., Adams, B., Destefanis, G., Tourani, P., Marchesi, M., and Tonelli, R. (2015). Are bullies more productive? empirical study of affectiveness vs. issue fixing time. In *2015 IEEE/ACM 12th Working Conference on Mining Software Repositories*, pages 303–313. IEEE.

Park, J. H., Xu, P., and Fung, P. (2018). Plusemo2vec at semeval-2018 task 1: Exploiting emotion knowledge from emoji and# hashtags. *arXiv preprint arXiv:1804.08280*.

Plutchik, R. (1980). Emotion: A psychoevolutionary analysis. *Nueva York: Harper and Row*.

Plutchik, R. (2003). *Emotions and life: Perspectives from psychology, biology, and evolution.* American Psychological Association.

Posner, J., Russell, J. A., and Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, 17(3):715–734.

Preoţiuc-Pietro, D., Schwartz, H. A., Park, G., Eichstaedt, J., Kern, M., Ungar, L., and Shulman, E. (2016). Modelling valence and arousal in facebook posts. In *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 9–15.

Read, J. (2005). Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In *Proceedings of the ACL student research workshop*, pages 43–48.

Scherer, K. R. and Wallbott, H. G. (1994). Evidence for universality and cultural variation of differential emotion response patterning. *Journal of personality and social psychology*, 66(2):310.

Strapparava, C. and Mihalcea, R. (2007). Semeval-2007 task 14: Affective text. In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, pages 70–74.

Suttles, J. and Ide, N. (2013). Distant supervision for emotion classification with discrete binary values. In Alexander Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing*, pages 121–136, Berlin, Heidelberg. Springer Berlin Heidelberg.

Tang, D., Wei, F., Yang, N., Zhou, M., Liu, T., and Qin, B. (2014). Learning sentiment-specific word embedding for twitter sentiment classification. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 1555–1565.

Valitutti, R. (2004). Wordnet-affect: an affective extension of wordnet. In *In Proceedings of the 4th International Conference on Language Resources and Evaluation*, pages 1083–1086.

Wallbott, H. G. and Scherer, K. R. (1986). How universal and specific is emotional experience? evidence from 27 countries on five continents. *Information (International Social Science Council)*, 25(4):763–795.

Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., and Hovy, E. (2016). Hierarchical attention networks for document classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1480–1489.

Zhong, P. and Miao, C. (2019). ntuer at semeval-2019 task 3: Emotion classification with word and sentence representations in rcnn. *arXiv preprint arXiv:1902.07867*.