INMCA - Text Based Affective Analysis

Alessandro Ballerini[1] and Francesco Dettori[2]

[1]alessandro.ballerini1@studenti.unimi.it, Matr: 927412
[2]francesco.dettori@studenti.unimi.it, Matr: 938320

***Abstract*—The goal of this analysis is to study the correlations between gaze features and human emotions, using NLP techniques.**

The theories and studies underlying emotions are different, making it impossible to unify their representation. For this reason we have decided to carry out this analysis using the two main theories of emotions: Paul Ekman's theory on discrete emotions and Russell's theory on core affect.

We decided to structure the analysis following two different approaches. The first approach consists on analyzing the emotional score values (both continuous and discrete) of the individual words to see whether they could provide meaningful information even if taken without context. The second approach consists of analyzing the whole sentences, their emotional values, and their relations with the eye-tracker features.

The sentence-based analysis discovered meaningful relationships between some eye-tracker features and discrete emotions, which were also partially confirmed by the word-based analysis. On the other hand, our models performed poorly when applied to the core affect context, not being able to distinguish between the VAD features.

I. INTRODUCTION

Philosophers and scientists have studied and formulated various theories on emotions, which have begun to take scientific validity with Darwin's studies on the evolutionary approach. From these, two main psychological theories concerning emotions have emerged.

The first theory is based on Paul Ekman's studies[1][2]. He proposes a discrete interpretation of the emotions, and identifies six fundamental emotions: surprise, anger, sadness, disgust, fear, and joy. In particular, he identified some relationships between emotions and eyes movements, which are fundamental in our work.

The second theory is based on Russell's studies [3] on a continuous emotional state: the core affect. The core affect turns out to be a "space" in which the emotional state varies according to whether the values of Valence and Arousal increase and decrease. According to Russel's theory, the single emotions are identified when certain regions of this space are "reached" multiple times are labeled with a specific name.

In this paper we wanted to understand if there are relationships between emotions, both discrete and continuous, and the gaze patterns during the reading.

Eye trackers allows us to record and analyze eyes movements. In particular they can track the quick movements called saccades, and the moments when the eyes stops on the same spot for longer periods, called fixations. These two types of eye actions are the main tools we have to perform our gaze analysis.

II. RELATED WORKS

We used three data sources to perform our analysis: two english lexicons were used to associate words to emotions, and a dataset for the gaze features.

A. NRC Sentiment Lexicons

The NRC lexicon [4] provides different datasets that associate the words of multiple languages to different emotion representations. For our discrete emotion analysis, we choose the "NRC Emotion Intensity Lexicon" dataset for the English language, that associates 6000 English words to a numerical value, between 0 and 1, for eight different discrete emotions. In the lexicon, the terms were extracted from Twitter.

For each of the emotions, the authors created separate lists of terms that satisfied either one of the two properties listed below:

- The word is already known to be associated with the emotion (although the intensity of emotion it conveys is unknown).
- The word tends to occur in tweets that express emotion.

Regarding the continuous emotion analysis, we choose the "NRC VAD Lexicon" that associates 20.000 English words to three continuous variables which are Valence, Arousal, and Dominance that would represent the complete spectrum for continuous emotion representations. The VAD values of each word are numeric and included between 0 and 1. These words are common in English and they include, in particular, terms that denote or connote emotions.

B. ZuCo 2.0

The gaze features for our analysis have been retrieved from the "ZuCo 2.0[5]" dataset, which provides the recordings of the eye features of 18 subjects during the reading task of 739 English sentences. The data used were obtained from a second reading of 390 sentences by the subjects in a task specific paradigm, in which the 18 participants actively search for a semantic relation type in the given sentences. For that reason there are 63 duplicates between the normal reading and the task-specific sentences: they want to provide a set of sentences read twice by all participants with a different task in mind. The gaze data were obtained through eye trackers which recorded the eye movements of each participant during the reading. The raw data obtained was then processed and provided with the dataset we used.

The recorded features provided by the dataset are the following:

- Gaze duration (GD): The sum of all fixations on the current word in the first-pass reading before the eye moves out of the word;
- Total reading time (TRT): The sum of all fixation durations on the current word, including regressions;
- First fixation duration (FFD): The duration of the first fixation on the prevailing word;
- Go-past time (GPT): The sum of all fixations prior to progressing to the right of the current word, including regressions to previous words that originated from the current word;
- Mean pupil size (MPS): The mean size of the pupil during the reading of the current word;
- Number of fixations (nFix): The total number of fixation for the current word;

III. MODELS

A. Multiple Linear Regression

As we said before, with this analysis we want to investigate the correlations between gaze features and emotions. To do this, we decide to use multiple linear regression (MLR).

MLR is a statistical technique that uses several explanatory variables to predict the outcome of a response variable. The goal of multiple linear regression (MLR) is to model the linear relationship between the explanatory (independent) variables and response (dependent) variable. To fit our model, we used the ordinary least square (OLS).

The MLR formula is the following:

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_i x_i + \epsilon \tag{1}$$

There are:

- i = number of observations (1,...,n)
- y_i = dependent variable.
- x_i = explanatory variables.
- β_0 = y-intercept (constant-term).
- β_i = slope coefficient for each explanatory variable.
- ϵ = error of the model (residual).

The value on which we focused most is the p-value. It is the probability of obtaining test results at least as extreme as the results observed, under the assumption that the null hypothesis is correct.

We used the P-value to build our MLR model via Stepwise Regression, which is a method of fitting regression models in which the choice of predictive variables is carried out by an automatic procedure.

B. Backward Elimination

The backward elimination is one of the different approaches that can be implemented for the Stepwise Regression.

It is a procedure that starts with all candidate variables, testing the deletion of each variable using a chosen model fit criterion, deleting the variable (if any) whose loss gives the most statistically insignificant deterioration of the model fit, and repeating this process until no further variables can be deleted without a statistically insignificant loss of fit.

This technique consists of 5 steps:

- Select a significant level (SL) of the parameter to stay in the model (we chose a SL = 0,05).
- Fit the full model with all possible predictors.
- Consider the predictors with the highest P-value: if the $P - value > SL$ go to Step 4, otherwise End.
- Remove the predictor.
- Fit model without this variable. Then go back to step 3.
- END: the model is ready.

C. R-Squared

Another important value we get from this model is the R-Squared coefficient. It is a statistical metric that is used to measure how much of the variation in outcome can be explained by the variation in the independent variables. R-squared can only be between 0 and 1, where 0 indicates that the outcome cannot be predicted by any of the independent variables. On the other hand, 1 indicates that the outcome can be predicted without error from the independent variables. Since our model involves multiple independant variables we are goin to use the adjusted R-squared. This is to avoid that the mere fact of involving more variables is an added value of the models we are going to consider, as simple R-squared tends to reward models with more variables.

IV. ANALYSIS

The analysis consists of using a regressor model (OLS) to search for meaningful relationships between the selected eye-tracking features (from the ZuCo dataset) and the lexicon's emotional values (from the NRC lexicons).

The eye-tracking features are used as independent variables, while the different emotional features are involved as the dependent ones. The analysis is repeated both using the discrete emotion score values, and the VAD values, each of them taken singularly (8 discrete emotions and 3 VAD features).

Backward elimination has been used to select the most relevant features and obtain the best model.

A. Data preprocessing

The eye-tracker features have been extracted from the MATLAB files provided by the ZuCo dataset. Numerical values and punctuation marks have been removed from the sentences to reduce the number of strings to process. Missing values entries have been removed. The resulting collection of words have been then exported into CSV files.

B. Word base analysis data handling

For the word-based analysis we compute the mean value of each ZuCo feature for the restricted set of words that belong to the intersection between the two dataset (ZuCo and NRC). Since the same word could compare into multiple sentences, we try to isolate every word from its context by taking the mean values of each feature and search for meaningful relationships with the discrete emotions or VAD features.

	Word	GD	FFD	GTP	nFix	MPS	Score
0	losing	-1.018	-1.169	-0.268	-0.803	0.575	-0.724
1	whip	0.340	1.245	-0.460	-0.714	0.083	-0.153
2	gore	-0.323	-0.001	-0.537	-0.232	0.750	-0.756
3	socialist	0.825	1.427	-0.282	-0.605	0.388	-1.202
4	politics	-0.831	-0.647	-0.438	-0.642	-0.203	-1.279

TABLE I: Snapshot of the resulting dataset obtained after the data preprocessing and data handling, for the word-based analysis. Independent variables: gaze duration (GD), first fixation duration (FFD), go past time (GPT), number of fixations (nFix) and mean pupil size (MPS).mm

C. Sentence base analysis data handling

To obtain a set of feature values associated with each sentence, we sum the values of the same features associated with each word that compose the sentence. In other words, the feature ”A” value for the sentence ”X” will be the summation of the values of the feature ”A” associated with each word that compose the sentence ”X”. The only exception is for the ZuCo feature MPS (mean pupil size), since due to the nature of this specific feature, the summation of the value of each word would not be a meaningful operation to perform.

To allow for a better interpretation of the data and the resulting relationships, each feature is standardized (mean = 0, standard deviation = 1).

Notice that the independent variables adopted are a subset of the features made available by the ZuCo dataset. For both word based and sentence based analysis a selection has been made to avoid multicollinearity.

D. Multicollinearity and features selection

Since the goal of our analysis is to search for meaningful relationships between the emotional features and the eye tracker recordings using regression techniques, we need to ensure that the independent variables are not affected by multicollinearity.

Multicollinearity occurs when the independent variables in a regression model are correlated. A high degree of correlation between the independent variables makes it difficult to estimate the relationship between every single independent variable and the dependent one since the independent variables tend to change in unison.

This directly affects the coefficient estimation of our model, and it undermines our attempts to draw meaningful conclusions from data analysis. Multicollinearity also negatively affects the statistical power of our regression model. It makes the interpretation of the p-value potentially meaningless to select the most statistically significant variables.

The causes of multicollinearity could be different: It may be dependent on the model structure and on how it handles the data (structural multicollinearity). If one or more features derived from the combination of other ones, this necessarily introduces a strong dependence between the variables in question. But multicollinearity could also depend on the data itself, by how data are retrieved and what they mean (data multicollinearity).

E. Variance Inflation Factor (VIF)

To verify whether our model's variables are affected by multicollinearity and to which degree, we use the Variance Inflation Factors (VIF).

The variance inflation factor is calculated for each independent variable and identifies relationships between the selected feature and all the others.

VIF starts at 1 and has no upper limit. The higher the value, the higher is the correlation between the selected variable and all the others.

A value of 1 indicates no correlation, while 5 is the conventional threshold to moderate correlation levels, and a value of 10 or superior indicates high correlation. For values of 10 and above p-value and the coefficient estimation became less and less meaningful.

VIF formula:

$$VIF_i = \frac{1}{1 - R_1^2} \tag{2}$$

Where “i” refers to the i-th independent variable and R-squared is the coefficient of determination for such variable.

F. VIF calculation

We performed the VIF calculation for both the word based and sentence based analysis on the ZuCo features (GD, TRT, FFD, GPT, nFix, MPS).

VIF calculation for the word-based analysis:

	feature	VIF
0	GD	3.169
1	TRT	16.660
2	FFD	2.878
3	GPT	1.139
4	nFix	11.729
5	MPS	1.051

TABLE II: VIF values for word-based model features.

The other features VIF indicator is under 5, which means that those features correlation is not significant and won't affect our analysis. The reasons behind the presence of a high correlation between the TRT and the nFix variables is due to the nature of the data and not being imputable to the model. It is easy to infer that a higher number of fixations on a word corresponds to a higher time spent on reading, hence the high correlation between the two features. By discarding one of the two features from the model we should solve the multicollinearity problem. We choose to discard the TRT for two main reasons: it's VIF value is higher than the nFix one, this means that by discarding TRT we would decrease the correlation between all variables more significantly. The second reason is since TRT is much more similar to the other features, they all measure reading times in different ways; hence keeping nFix would add more information to the analysis.

	feature	VIF
0	GD	2.709
1	FFD	1.968
2	GPT	1.124
3	nFix	1.604
4	MPS	1.039

TABLE III: VIF values for word-based model features after the features selection.

The VIF calculation performed after the feature selection shows that the multicollinearity issue has been successfully handled.

VIF calculation sentence based analysis result:

	feature	VIF
0	GD	12.346
1	TRT	92.300
2	FFD	124.645
3	GPT	12.346
4	nFix	80.860
5	MPS	1.158

TABLE IV: VIF values for sentence-based model features.

In this second case, the VIF calculation showed very high correlationì with all the features, except for the mean pupil size one. This could be due to the fact that most of the variables concern the reading time of a word and each of them adds some variations to these measurements. For example, the feature gaze duration measures the duration of the first gaze on the word, while the total reading time measures also the regressions on the word (the sum of all gazes on that word). This means that in all those cases when the reader fixates only once on the word those two features will have the same value. In this case only one variable, beside mean pupil size, can be kept, since the correlation between the remaining variables is far above the threshold of tolerance. In this case we decided to keep TRT since we think that it is one of the most comprehensible and significant features.

	feature	VIF
0	TRT	1.025
1	MPS	1.025

TABLE V: VIF values for sentence-based model features after the features selection.

G. *Word-based analysis: Discrete emotions*

After fitting an OLS multiple regression model for each one of the discrete emotions (selecting the most relevant eye-tracker features by backward elimination), only the fear model showed interesting signs of relationship with two of the independent variables: number of fixations and mean pupil size.

	coeff	std err	P-value
const	-0.497	0.183	0.009
nFix	-0.598	0.252	0.021
MPS	0.317	0.154	0.044

Adj. R^2
0.091

TABLE VI: Word-based analysis - Fear model summary

The low p-value indicates that the fear emotion seems to be strongly related to the Number of Fixations and Mean Pupil Size features. By interpreting the coefficients, we may assume that fear is associated with fewer fixations and greater dilation of the pupil. The low R-square value indicates that our model is not able to explain the variability in the data, so it is not suitable to perform predictions, but may still be useful for analytical intents.

H. *Word-based analysis: VAD*

The model showed hardly any linear relationship between the emotions scores and the eye-tracker features. The data distribution is very similar for the three VAD features. This means that probably either our methods of data handling or the eye-tracking features themself are badly suited to highlight the VAD features differences. Only the arousal model showed the presence of relationship signs with the gaze duration feature. Beside, as in the previous case, the low R-squared value denotes the low predictive capability of the model.

Model	Features	coeff	std err	P-value	Adj. R^2
Arousal	GD	0.092	0.040	0.022	0.005
Dominance	GD	0.151	0.045	0.001	0.013
	MPS	-0.0785	0.035	0.025	

TABLE VII: Word-based analysis - Arousal and Dominance models summary

The analysis showed the presence of relationship signs with the gaze duration feature. Nevertheless, as the R-squared suggests, hardly any variation in the data can be precisely explained by our model. A similar situation resulted from the dominance model. In this case, the model found some significant relationship signs with both the GD and the MPS features. As in the previous case, however, the low R-squared value suggests the lack of predictive power of our model.

I. Sentence-based analysis: Discrete emotions

As we can notice by the plots, a considerable number of entries have zero as a dependent variable. This phenomenon occurs since, in a single sentence, rarely more than three or four emotions are present. For this reason, for each discrete emotion analysis, we are going to take into consideration only the sentences where that emotion is present.

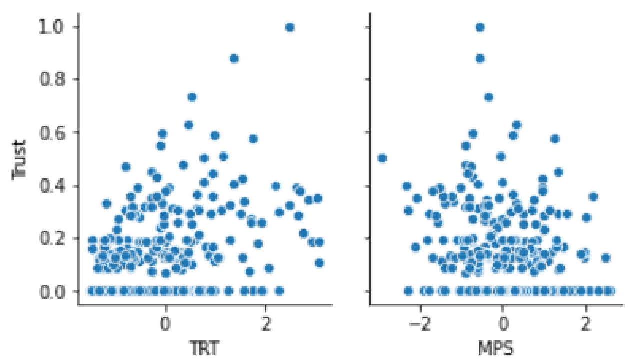

Fig. 1: Sentence-based analysis scatter plot for trust. We can see how a considerable amount of sentences does not contain the trust emotion.

This analysis, in particular, led to interesting results. Meaningful relationships have been shown by the model for the emotion Anticipation, Trust, Fear, and Sadness.

Emotion	variables	coef	std err	P-value	Adj. R^2
Anticipation	TRT	0.042	0.015	0.005	0.067
Trust	TRT	0.0606	0.030	0.048	0.121
Fear	MPS	0.0509	0.025	0.043	0.060
Sadness	MPS	0.0606	0.030	0.048	0.057

TABLE VIII: Sentence-based - discrete emotions analysis results summary

Is interesting to notice how “negative” emotions as fear and sadness seem to be linked to the mean pupil size feature, while neutral emotions such as trust and anticipation seem to have a connection with the total reading time feature.

J. Sentence-based analysis: VAD

All three analyses (Fig. 2) show strong linear relationship signs, in particular with the TRT feature. The R-square value also is around 50% for each model, this is an improvement from the previous analysis since the model seems to have a much greater predictive strength.

However the scatter of the data points is very similar for the three VAD features. Since data has been handled differently compared to the word-based/VAD analysis, this could lead to the conclusion that VAD features are poorly distinguished by the eye-tracker values.

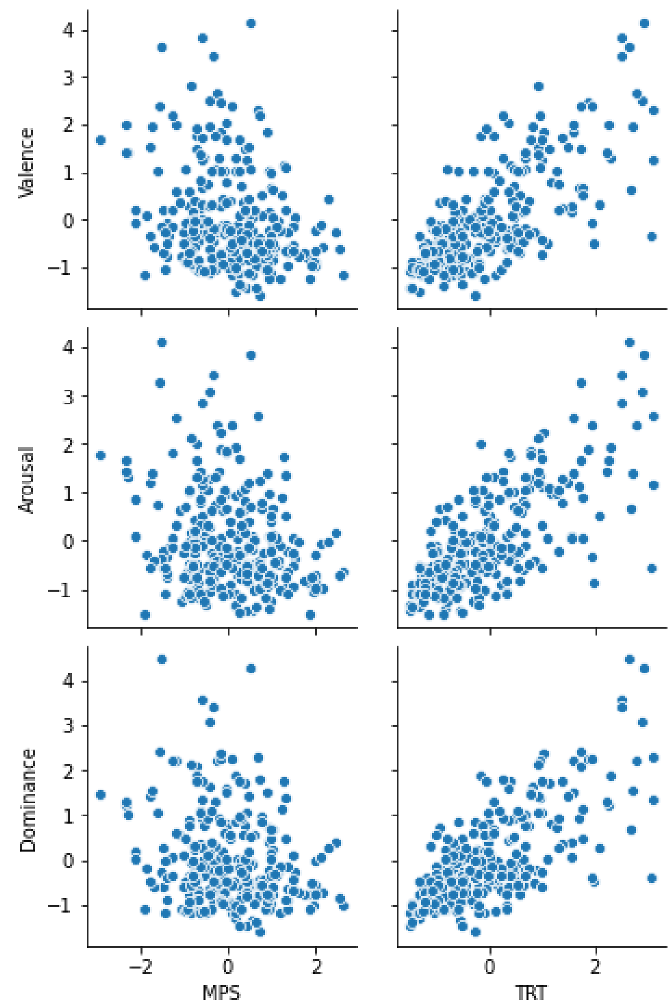

Fig. 2: Sentence-based data points scatter plot over MPS and TRT features. All three VAD features show a very similar data distribution over each feature.

VAD model	variables	coef	std err	P-value	Adj. R^2
Valence	TRT	0.7039	0.043	0.000	0.530
	MPS	-0.1562	0.045	0.001	
Arousal	TRT	0.7078	0.044	0.000	0.525
	MPS	-0.1224	0.045	0.007	
Dominance	TRT	0.7144	0.043	0.000	0.526
	MPS	-0.1006	0.045	0.026	

TABLE IX: Sentence-based - VAD analysis results summary

V. Conclusions

The goal of this analysis was to search for relationships between gaze features and human emotions through written text. Emotional values were retrieved by the use of lexicons, datasets that associate words to emotional scores. Since two are the main approaches to emotions, the discrete and the core affect interpretation, we adopted two lexicons to obtain the emotion score values, one that followed the discrete emotion approach (NRC Emotion Intensity Lexicon) and one that adheres to the core affect one (NRC VAD Lexicon).

We then performed two different analyses based on different data handling. In the word-based approach, we wanted to isolate the words from their context while in the sentence-based approach we were interested in the relationships between eye-tracking features and emotions considering the whole phrase.

The most interesting results were provided by the sentence-based/discrete emotion analysis. The analysis showed significant relationship signs between two ”negative” emotions, sadness, and fear, with the MPS (mean pupil size) feature. The relationship suggests that those emotions are related to a slight increase in pupil size. Those results were also partially confirmed by the word-based analysis, where the relationship between fear and MPS was also highlighted. Another interesting result was the relationship between two much more ”neutral” emotions like anticipation and trust and the TRT (total reading time) feature. This could be interpreted by associating the ”expectation about something” with a slower pace during the readings. The VAD/sentence-based analysis results showed that the way we handled the data was not well suited to distinguish between the VAD features. The three models (one for each one of the VAD features) showed almost identical results, failing to highlight any substantial difference between them.

References

[1] P. Ekman and D. Keltner, “Universal facial expressions of emotion,” *Segerstrale U, P. Molnar P, eds. Nonverbal communication: Where nature meets culture*, pp. 27–46, 1997.

[2] P. Ekman, “Basic emotions,” *Handbook of cognition and emotion*, vol. 98, no. 45-60, p. 16, 1999.

[3] J. A. Russell, “Core affect and the psychological construction of emotion.,” *Psychological review*, vol. 110, no. 1, p. 145, 2003.

[4] S. M. Mohammad and P. D. Turney, “Nrc emotion lexicon,” *National Research Council, Canada*, vol. 2, 2013.

[5] N. Hollenstein, M. Troendle, C. Zhang, and N. Langer, “Zuco 2.0: A dataset of physiological recordings during natural reading and annotation,” *arXiv preprint arXiv:1912.00903*, 2019.