**Chapter 5:**

**Lexical Study of Emojis**

This chapter aims to evaluate relevance of parameters for inclusion in an emoji lexicon with respect to the most optimal outcomes for later use to detect sarcasm. The most prevalent methodologies in literature are largely unfit for this purpose as sentiment is derived from the corresponding text, however, incongruency in text and emoji sentiment is widely cited as a marker of sarcasm thus there are validity concerns with regards to any works which deploy this strategy.

**Dataset selection**

*Basic Emotional Theory*

By the nature of the problem domain, there is a necessity for human-annotated data, where emojis are presented independent of any text which may influence classifications. The EmoTag1200 dataset contains information regarding volunteers’ association of emojis to basic emotions, based on the Plutchik model. The presentation of emojis without any textual prompts makes this dataset uniquely suitable for this task, and methodologies for its generation are robust; multiple annotators were utilised, and agreement was monitored via Pairwise Pearson correlation and Krippendorff’s α. Steps were taken to mitigate fatigue bias through randomisation and reported results represent averages of the reported values.

*Dimensional Theory*

The dataset used in this section was the Emoji Sentiment Ranking dataset, containing rankings of the 751 most popular emojis annotated by 83 annotators as positive, negative, or neutral. Annotators were presented with the emoji alongside the text, however, were instructed to rank the emoji sentiment only. While bias associated with the textual component of the content cannot be totally ruled out in this case, no datasets presently exist which are created where emojis were annotated independently of text.

**Definition of Sentiment Parameters:**

The following section aims to evaluate optimal approaches to define sentiment of emoji with regards to the two emotional theory models. Successful methodologies should generate parameters can fingerprint sentiment in terms of a wide array of features to facilitate highly accurate machine learning models to be constructed in subsequent work.

*Basic Emotional Theory:*

The EmoTag1200 dataset contains information regarding the affinities of emoji to 8 basic emotions, which enable comparison and acknowledge that emotions are experienced synchronously with varying strengths based on the stimulus in question. Sentiment labels *c*, consist of continuous values representing affinity of the emoji to each emotion within set *e*:

where

Labels are most clearly equated to relative affinities to emotions; however, it is also reasonable to link these values to probabilities that a given emotion is the primary emotion associated with the emoji. Such a definition implies that:

where

The primary basic emotion is , the secondary is and the trend continues with decreasing affinity to the emoji. The former definition aligns more literally with the intention of original authors and provides increasing granularity which may be of value when more nuanced fingerprinting between emoji is necessary. The latter yields reduced granularity however greater accuracy of models generated for prediction of unseen emoji may be possible, and thus the result may be data of overall greater value.

*Dimensional Theory:*

The Emoji Sentiment Ranking dataset contains information regarding the frequency to which emojis are classified as positive, negative, or neutral. These classifications were represented by discrete values which consider logical ordering of categories of equal distance apart to facilitate quantitative analysis:

Using these values, a discrete probability distribution for sentiment label *c*, can be determined where an assumption that the sum of the three respective probability distributions is equal to one is made to normalise data across emojis which appear at varying frequencies:

Where *c* is the sentiment label, and the following abbreviations were used to denote negativity, neutrality, and positivity:

(Negativity)

(Neutrality)

(Positivity)

Due to the low occurrence of annotation (*N* ≥ 5 is true for the lower quartile) presented in the dataset, relative frequency is not an ideal metric for approximation of probabilistic sentiment scores in many cases as will equal zero in events not observed in the data; such situations would result in bias due to the implication that certain events are impossible, whereas a low probability is more likely. For machine learning, contributes to overfitting as data is sparser. Where non-zero probabilities are assigned to unseen events, some information about their potential occurrence is preserved and thus more robust patterns can be captured. Additionally, where *N* is small averages are increasingly subject to skew. Probability distributions were thus determined using a Laplace estimate:

Where *k* is the cardinality of the class, in this case |c| = 3. The assumption of uniform distribution for the estimator is true for the dataset in question. The smoothing model was selected for its ability to address the issue of zero-probabilities and mitigate the impacts of small *N* where relevant. At larger *N*, the output approaches prior to any transformation. Such a feature of the estimator focuses its effects on cases where it is most necessary. Finally, defining as the mean of the probability distributions weighted against their discrete labels, a sentiment score can be determined:

This approach aims to acknowledge the varying perceptions of emoji sentiment, dependent upon an individuals’ personal usage, while retaining ability to evaluate the sentiment with regards to the consensus classification. An overall sentiment score for sentiment using discrete categorisations often employs a majority decision methodology, however in this context an alternative methodology which acknowledges a degree of subjectivity is preferable. The standard error of the mean is one such methodology which enables a more nuanced classification methodology:

The result of such a methodology enables the identification of a most probable classification and inclusion or exclusion of secondary potential classifications at a 95% confidence.

A diagram of emojis

Description automatically generated

*Figure X*Sample of result of emoji sentiment classification using dimensional emotional theory. Coloured bars are proportional to *p+, p0* and *p-*. Markers (black) represent . Grey markers represent the limits of the 95% confidence intervals for the . Note weary face (left) has high confidence that a negative classifier is appropriate, however ghost (right) cannot necessarily be classified as positive with 95% confidence.

**Feature extraction of emoji**

Previous works cite challenges regarding emojis lacking characteristics to enable feature extraction. The following work aims to evaluate several strategies to extract information regarding the sentiment of emoji for the purposes of improving outcomes of sarcasm detection.

*Use of Emotion Vocabulary Embeddings:*

Word embeddings are representations of words as fixed-size vectors in multi-dimensional space, where each dimension captures some information about the word such as semantic and pragmatic relationships. Word2Vec is one such method to generate embeddings which has been observed to be particularly effective as capturing relationships between words (Fontaine, 2013). For example, given the male/female relationship is known:

*King – Man + Woman = Queen*

Relationships as above are effectively demonstrated where the results of the above transformation on the ‘King’ vector is very close to that of the ‘Queen’ vector. Given this observation, it is appropriate to use these vectors for comparative purposes, and by extension feature extraction. Emoji2Vec is an equivalent embedding which may supplements word-based corpus with emoji embeddings (Eisner et al., 2016). Authors cite greatest performance of this embedding with the Google News embedding for Twitter sentiment analysis tasks, thus this is the combination of embeddings which will be used. See appendix X for further discussion on the vectors selected.

Cosine similarity is a measure of similarity between two vectors, defined as follows:

Where:

And:

Which in the context of natural language processing provides an interpretable metric appropriate for the use-case it measures directional similarity, without consideration to magnitude and is robust to high dimensionality. As word frequency influences vector magnitude, this strategy eliminates related noise while conserving semantic and pragmatic relationships between words.

A diagram of a straight line

Description automatically generated

*Figure X* Cosine Similarity.

*Word Vectors of Basic Emotion Words*

This strategy aims to compare emojis () to each basic emotion () and defines emotion scores based on the cosine similarity between respective vectors ():

This approach eliminates noise related to distinctions for related vocabulary however this may be too strict where the goal is correlation to human-annotated data given the varied nuance that individuals assign to emotions. To address the strict parameters bounds of this strategy, several alternatives are proposed which employ less stringent boundaries on vocabulary for comparison.

*Binary Word-Association Lexicon*

This method draws upon the EmoLex lexicon which assigns binary labels to convey association to basic emotions for each word. Vocabulary tagged for association with a basic emotion (), association to the basic emotion is ranked using cosine similarity and words (, where top- words) with greatest association are determined. Finally, the emotion score is determined as the average of resulting cosine similarities:

The method addressed the nuance related to individuals’ understanding of basic emotions, however words not contained within the EmoLex vocabulary are not considered, potentially disregarding relevant vocabulary. Additionally, this model filters for inclusion based upon cosine similarity to the emoji which fails to account for the degree of association between the basic emotion and the words in the subset. This likely results in negative skew in distribution of results for words uncorrelated to a given emotion.

*Word-Emotion Intensity Lexicon*

This method proposes an alternative approach which addresses the limitations associated with the use of a binary word lexicon, through the utilisation of the NRC-EIL lexicon which tags vocabulary with an intensity score for affinity to a given basic emotion. This model operates on a similar principle, however, selects the top-k words based upon intensity score, increasing the relevance of the considered vocabulary.

*Table X* Pearson Correlation scores for all considered prediction methods. Bolded Scores represent the greatest correlation (positive or inverse) observed for the emotion in the respective column, excluding human agreement scores. Average values are reported with respect to absolute correlation.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Source | Variant | Anger | Anticip | Disgust | Fear | Joy | Sadness | Surprise | Trust | Average |
|  | ***Basic Emotion Word Similarities*** | | | | | | | | | |
| Word2Vec |  | **0.36** | -0.01 | **0.30** | 0.10 | **0.44** | **0.36** | **0.09** | **0.15** | 0.23 |
|  | ***Emotion Lexicon Corpus Word Similarities*** | | | | | | | | | |
| EmoLex | k=5 | 0.35 | -0.16 | 0.25 | **0.39** | 0.36 | 0.30 | -0.04 | 0.14 | 0.25 |
| EmoLex | k=10 | 0.31 | -0.13 | 0.24 | 0.37 | 0.39 | 0.29 | **-0.09** | **0.15** | 0.25 |
| EmoLex | k=50 | 0.26 | **-0.18** | 0.28 | 0.34 | 0.40 | 0.29 | -0.02 | **0.15** | 0.24 |
| EmoLex | k=100 | 0.26 | -0.17 | **0.30** | 0.30 | 0.41 | 0.31 | 0.01 | 0.14 | 0.24 |
| NRC-EIL | k=10 | 0.26 | -0.16 | 0.13 | 0.20 | 0.30 | 0.32 | 0.05 | -0.01 | 0.18 |
| NRC-EIL | k=100 | 0.19 | -0.12 | 0.24 | 0.24 | 0.33 | 0.33 | -0.01 | 0.01 | 0.18 |
| NRC-EIL | k=300 | 0.18 | -0.13 | 0.23 | 0.25 | 0.31 | 0.33 | 0.02 | -0.02 | 0.18 |
|  | ***Human Annotation*** | | | | | | | | | |
|  |  | 0.67 | 0.74 | 0.65 | 0.59 | 0.78 | 0.69 | 0.64 | 0.72 | 0.69 |

Table X compares the mean human-annotated emotion ratings against predicted scores for the proposed methods. Direct comparison to basic emotion word vectors yielded the overall highest performance. Notably the basic emotion was the most significant factor in performance overall, with some emotions achieving some level of correlation and others displaying zero, or weak negative correlation to their predicted values. The impact of context window size was varied across all subsets tested.

These results may be explainable by the nature of the relationships described by the vectors used. For example, *happiness* and *sadness* have a cosine similarity score of 0.42 using Google News Word2Vec model. This is capturing an intuitive relationship between the pair; however, within the context of this methodology, the implication that happiness is a core component of *sadness* is invalid.

*Sentiment-Aware Vector Space Modification*

This section aims to modify Googles News Word2Vec Embeddings to emphasise characteristics relevant to sentiment analysis while mitigating influence of contradictory semantic relationships. Two proposed strategies which leverage word-pair polarities or correlations between basic emotions to modify vectors were evaluated. The strategies implement a stochastic gradient descent model and Adam optimiser with two possible loss functions:

Where:

The goal of the above function is to maximize the similarity score of synonymous pairs of words ( where ), while minimizing the score for antonymous pairs () by considering their respective word vectors (). Following tuning, obtaining a min value ≈-7500 was achieved which approaches a theoretical minimum ≈-8000. Given the loss function which cannot account for degree of similarity, but rather optimises for polarity, several functions alternative loss values were considered.

The alternative loss function which directly considers Pearsons’ correlation () between basic emotions for relevant datasets:

The goal of this strategy is to optimise vector placement based upon Pearsons Correlation for all permutations () of Plutchiks basic emotions () using the human annotated data contained in the EmoTag1200 and NRL-EIL datasets respectively. Optimisations were weighted with respect to the correlation values to avoid shortcomings highlighted for the alternative strategy and thus loss functions could be minimised rather than considering varying degrees of loss. A depth discussion of the vector optimisation models is described in section X.

*Evaluation parameters*

The quality of outcomes was quantified by means of considering the Pearson’s correlation between the approximated values for each basic emotion, using methods outlined in section X against the human-annotated values. Pearson’s correlation measures the strength of association between variables:

Where:

As both total similarity and opposites provide valuable information for the purposes of a regression task, the best outcome is defined as the maximum absolute value for correlation, rather than the greatest correlation. While ideally an approximation method would yield a linear relationship to the human annotated data which can be captured by Pearson’s correlation, it is possible that more complex relationships, which are useful for regression modelling but not evident using this metric may arise. In this case, suitable results could be obtained used this metric only however it may be of value in expansions upon this work to consider alternative relationships between approximations and human-annotations to improve outcomes.

*Results*

Tables X-X summarise the results of all vector space modification tests using previous estimators and drawing correlation to human-annotations for each emoji in the EmoTag1200 dataset. In the case of every basic emotion except for *anger*, modified vectors displayed improved suitability for the purpose of sentiment analysis using the Plutchik model. Results indicate the Emoji2Vec vectors were not limited by contradictory semantic relationships to the same extent as the word vectors, evident by the improved outcomes using modified word vectors alongside the original Emoji2Vec vectors. Given the disproportionate representation of emoji relative to words as a manner to convey sentiment rather than information, this outcome is reasonable. The quality of outcomes across the emotions seems to be related to the strength of ties between the basic emotion and sentiment polarity. For example, *joy* and *sadness* are intuitively linked to *positive* and *negative* sentiments respectively and yielded reasonably favourable results in the analysis. Conversely *surprise* is more ambiguous to classify based on polarity and yielded lower correlation. Such an observation highlights the limitations of the proposed objective functions; they rely on the polarity of the emotions and cannot necessarily acknowledge the core feeling that they represent. While this is a relevant component of the emotions, additional emphasis on the pragmatic component may improve outcomes in future improvements upon the proposed models. Improved correlation was achieved to human-annotated values in the EmoTag1200 dataset which will improve outcomes for the regression model discussed in section X. Additionally and importantly, this result leverages the emoji vector only to obtain results, thus enabling the extrapolation of data to all emojis in the Emoji2Vec vocabulary. If additional parameters were considered, improved outcomes were possible however this would reduce the quantity of emojis that could be considered for the regression task to those which the additional parameters were available.

*Table X* Pearson Correlation scores for all considered prediction methods following vector-space modification- method 1.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Source | Variant | Anger | Anticip | Disgust | Fear | Joy | Sadness | Surprise | Trust | Average |
|  | ***Basic Emotion Word Similarities (Loss ≈-7500)*** | | | | | | | | | |
| Word2Vec |  | -0.01 | 0.10 | 0.04 | 0.13 | 0.21 | 0.10 | -0.02 | 0.05 | 0.08 |
|  | ***Emotion Lexicon Corpus Word Similarities (Loss ≈-7500)*** | | | | | | | | | |
| EmoLex | k=5 | -0.10 | 0.10 | 0.00 | 0.05 | 0.16 | 0.06 | -0.01 | -0.01 | 0.06 |
| EmoLex | k=10 | 0.00 | 0.04 | -0.04 | 0.05 | 0.17 | 0.07 | -0.09 | 0.00 | 0.06 |
| EmoLex | k=50 | 0.01 | 0.16 | -0.02 | 0.09 | 0.19 | 0.00 | **-0.10** | 0.04 | 0.08 |
| EmoLex | k=100 | 0.01 | 0.16 | -0.03 | 0.12 | 0.15 | 0.02 | -0.09 | 0.02 | 0.08 |
| NRC-EIL | k=10 | 0.04 | 0.08 | 0.12 | 0.00 | 0.20 | -0.01 | 0.03 | 0.02 | 0.06 |
| NRC-EIL | k=100 | -0.01 | 0.14 | 0.00 | 0.00 | 0.17 | 0.08 | -0.03 | 0.00 | 0.05 |
| NRC-EIL | k=300 | 0.00 | 0.19 | -0.04 | 0.04 | 0.17 | 0.05 | -0.05 | 0.01 | 0.07 |
|  | ***Basic Emotion Word Similarities (Loss ≈-5000)*** | | | | | | | | | |
| Word2Vec |  | 0.01 | 0.07 | -0.06 | -0.04 | 0.04 | 0.10 | 0.00 | -0.06 | 0.05 |
|  | ***Emotion Lexicon Corpus Word Similarities (Loss ≈-5000)*** | | | | | | | | | |
| EmoLex | k=5 | -0.04 | 0.17 | 0.03 | -0.03 | 0.00 | 0.03 | 0.04 | -0.04 | 0.05 |
| EmoLex | k=10 | -0.03 | 0.15 | 0.03 | -0.02 | 0.05 | 0.02 | 0.03 | -0.02 | 0.04 |
| EmoLex | k=50 | -0.05 | 0.15 | 0.00 | -0.03 | 0.02 | 0.01 | 0.02 | 0.03 | 0.04 |
| EmoLex | k=100 | -0.05 | 0.14 | -0.01 | -0.03 | -0.01 | 0.01 | 0.02 | 0.03 | 0.04 |
| NRC-EIL | k=10 | -0.08 | 0.17 | 0.00 | -0.04 | 0.05 | -0.04 | -0.02 | 0.05 | 0.06 |
| NRC-EIL | k=100 | -0.06 | 0.15 | -0.01 | -0.03 | 0.00 | 0.02 | 0.02 | 0.03 | 0.04 |
| NRC-EIL | k=300 | -0.06 | 0.14 | -0.01 | -0.04 | 0.01 | 0.02 | 0.03 | 0.05 | 0.05 |
|  | ***Basic Emotion Word Similarities (Loss ≈-3000)*** | | | | | | | | | |
| Word2Vec |  | 0.04 | -0.23 | 0.11 | 0.06 | **-0.25** | 0.12 | 0.04 | -0.13 | 0.12 |
|  | ***Emotion Lexicon Corpus Word Similarities (Loss ≈-3000)*** | | | | | | | | | |
| EmoLex | k=5 | 0.07 | -0.24 | 0.05 | 0.06 | -0.17 | 0.12 | 0.01 | -0.14 | 0.11 |
| EmoLex | k=10 | 0.06 | -0.23 | 0.04 | 0.08 | -0.04 | 0.12 | 0.00 | -0.15 | 0.09 |
| EmoLex | k=50 | 0.06 | -0.23 | 0.09 | 0.11 | -0.12 | 0.14 | 0.03 | -0.16 | 0.12 |
| EmoLex | k=100 | 0.08 | -0.22 | 0.12 | 0.14 | -0.11 | 0.13 | 0.03 | -0.17 | 0.13 |
| NRC-EIL | k=10 | 0.06 | **-0.25** | 0.16 | 0.16 | -0.10 | 0.07 | -0.01 | -0.15 | 0.12 |
| NRC-EIL | k=100 | 0.10 | -0.21 | 0.16 | 0.15 | -0.16 | 0.10 | 0.03 | -0.17 | 0.14 |
| NRC-EIL | k=300 | 0.11 | -0.21 | 0.17 | 0.16 | -0.15 | 0.09 | 0.03 | **-0.18** | 0.14 |
|  | ***Basic Emotion Word Similarities (Loss ≈-1000)*** | | | | | | | | | |
| Word2Vec |  | **0.16** | 0.08 | 0.10 | 0.19 | 0.15 | 0.02 | 0.01 | 0.16 | 0.11 |
|  | ***Emotion Lexicon Corpus Word Similarities (Loss ≈-1000)*** | | | | | | | | | |
| EmoLex | k=5 | 0.03 | -0.08 | 0.12 | 0.08 | 0.07 | 0.06 | 0.05 | -0.08 | 0.07 |
| EmoLex | k=10 | 0.03 | -0.08 | 0.11 | 0.10 | 0.06 | 0.07 | 0.05 | -0.10 | 0.08 |
| EmoLex | k=50 | 0.01 | -0.09 | 0.09 | 0.10 | 0.06 | 0.07 | 0.06 | -0.14 | 0.08 |
| EmoLex | k=100 | 0.01 | -0.08 | 0.09 | 0.10 | 0.05 | 0.08 | 0.06 | -0.15 | 0.08 |
| NRC-EIL | k=10 | 0.03 | -0.09 | 0.04 | 0.10 | 0.06 | 0.08 | 0.06 | -0.16 | 0.08 |
| NRC-EIL | k=100 | 0.00 | -0.08 | 0.07 | 0.10 | 0.07 | 0.09 | 0.05 | -0.15 | 0.08 |
| NRC-EIL | k=300 | 0.00 | -0.08 | 0.08 | 0.09 | 0.06 | 0.08 | 0.05 | -0.16 | 0.08 |
|  | ***Basic Emotion Word Similarities (Loss ≈-500)*** | | | | | | | | | |
| Word2Vec |  | **0.16** | 0.08 | 0.10 | **0.19** | 0.15 | 0.02 | 0.01 | 0.16 | 0.11 |
|  | ***Emotion Lexicon Corpus Word Similarities (Loss ≈-500)*** | | | | | | | | | |
| EmoLex | k=5 | 0.06 | -0.13 | 0.11 | 0.16 | 0.06 | **0.22** | -0.08 | -0.08 | 0.11 |
| EmoLex | k=10 | 0.06 | -0.13 | 0.14 | 0.14 | 0.07 | 0.21 | -0.06 | -0.11 | 0.12 |
| EmoLex | k=50 | 0.04 | -0.14 | **0.18** | 0.13 | 0.09 | 0.20 | -0.02 | -0.15 | 0.12 |
| EmoLex | k=100 | 0.05 | -0.15 | 0.17 | 0.12 | 0.08 | 0.19 | -0.02 | -0.16 | 0.12 |
| NRC-EIL | k=10 | 0.03 | -0.14 | 0.09 | 0.15 | 0.13 | 0.20 | -0.01 | -0.11 | 0.11 |
| NRC-EIL | k=100 | 0.03 | -0.15 | 0.14 | 0.12 | 0.09 | 0.21 | -0.05 | -0.16 | 0.12 |
| NRC-EIL | k=300 | 0.04 | -0.15 | 0.14 | 0.12 | 0.08 | 0.21 | -0.04 | -0.17 | 0.12 |
|  | ***Human Annotation*** | | | | | | | | | |
|  |  | 0.67 | 0.74 | 0.65 | 0.59 | 0.78 | 0.69 | 0.64 | 0.72 | 0.69 |

*Table X* Pearson Correlation scores for all considered prediction methods following vector-space modification- method 1. Original emoji vectors and transformed word vectors used.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Source | Variant | Anger | Anticip | Disgust | Fear | Joy | Sadness | Surprise | Trust | Average |
|  | ***Basic Emotion Word Similarities (Loss ≈-7500)*** | | | | | | | | | |
| Word2Vec |  | 0.13 | 0.00 | 0.31 | 0.10 | 0.41 | 0.22 | 0.11 | 0.01 | 0.16 |
|  | ***Emotion Lexicon Corpus Word Similarities (Loss ≈-7500)*** | | | | | | | | | |
| EmoLex | k=5 | 0.29 | -0.03 | 0.23 | 0.35 | 0.41 | 0.20 | -0.15 | -0.14 | 0.23 |
| EmoLex | k=10 | 0.28 | -0.03 | **0.34** | 0.37 | **0.50** | 0.27 | -0.11 | 0.04 | 0.24 |
| EmoLex | k=50 | 0.31 | -0.15 | 0.32 | 0.32 | **0.50** | 0.33 | -0.07 | 0.20 | 0.28 |
| EmoLex | k=100 | 0.32 | -0.17 | **0.34** | 0.36 | 0.44 | 0.36 | 0.09 | 0.11 | 0.27 |
| NRC-EIL | k=10 | 0.28 | -0.18 | 0.32 | 0.29 | 0.35 | 0.29 | **0.20** | -0.10 | 0.25 |
| NRC-EIL | k=100 | **0.35** | -0.04 | 0.28 | 0.39 | 0.43 | 0.40 | -0.07 | 0.17 | 0.27 |
| NRC-EIL | k=300 | 0.33 | -0.03 | 0.30 | **0.42** | 0.48 | **0.44** | 0.03 | 0.20 | 0.28 |
|  | ***Basic Emotion Word Similarities (Loss ≈-5000)*** | | | | | | | | | |
| Word2Vec |  | 0.30 | 0.07 | 0.02 | 0.36 | 0.16 | 0.03 | -0.08 | -0.01 | 0.13 |
|  | ***Emotion Lexicon Corpus Word Similarities (Loss ≈-5000)*** | | | | | | | | | |
| EmoLex | k=5 | 0.12 | -0.18 | 0.19 | 0.28 | 0.30 | 0.24 | -0.11 | -0.09 | 0.19 |
| EmoLex | k=10 | 0.17 | -0.17 | 0.23 | 0.26 | 0.27 | 0.27 | -0.07 | -0.07 | 0.19 |
| EmoLex | k=50 | 0.11 | -0.17 | 0.21 | 0.17 | 0.20 | 0.24 | 0.00 | -0.13 | 0.15 |
| EmoLex | k=100 | 0.10 | -0.18 | 0.21 | 0.15 | 0.20 | 0.21 | 0.04 | -0.12 | 0.15 |
| NRC-EIL | k=10 | 0.09 | -0.12 | 0.15 | 0.11 | 0.30 | 0.16 | 0.03 | -0.06 | 0.13 |
| NRC-EIL | k=100 | 0.09 | -0.14 | 0.18 | 0.16 | 0.24 | 0.20 | 0.02 | -0.09 | 0.14 |
| NRC-EIL | k=300 | 0.07 | -0.14 | 0.18 | 0.15 | 0.21 | 0.21 | 0.03 | -0.10 | 0.14 |
|  | ***Basic Emotion Word Similarities (Loss ≈-3000)*** | | | | | | | | | |
| Word2Vec |  | -0.17 | -0.10 | -0.02 | -0.11 | 0.04 | -0.09 | -0.16 | -0.17 | 0.11 |
|  | ***Emotion Lexicon Corpus Word Similarities (Loss ≈-3000)*** | | | | | | | | | |
| EmoLex | k=5 | -0.12 | -0.08 | 0.09 | -0.03 | 0.11 | -0.09 | 0.06 | -0.21 | 0.10 |
| EmoLex | k=10 | -0.09 | -0.08 | 0.09 | -0.03 | 0.09 | -0.06 | 0.04 | -0.19 | 0.08 |
| EmoLex | k=50 | -0.10 | -0.12 | 0.06 | 0.00 | 0.08 | -0.04 | 0.03 | -0.19 | 0.08 |
| EmoLex | k=100 | -0.11 | -0.12 | 0.04 | 0.00 | 0.09 | -0.03 | 0.04 | -0.16 | 0.07 |
| NRC-EIL | k=10 | -0.18 | -0.15 | 0.01 | -0.02 | 0.19 | -0.07 | 0.05 | -0.13 | 0.10 |
| NRC-EIL | k=100 | -0.13 | -0.13 | 0.05 | -0.03 | 0.09 | -0.04 | 0.06 | -0.17 | 0.09 |
| NRC-EIL | k=300 | -0.11 | -0.12 | 0.06 | -0.03 | 0.09 | -0.04 | 0.06 | -0.17 | 0.09 |
|  | ***Basic Emotion Word Similarities (Loss ≈-1000)*** | | | | | | | | | |
| Word2Vec |  | 0.31 | -0.04 | 0.21 | 0.29 | 0.19 | 0.09 | -0.03 | 0.11 | 0.16 |
|  | ***Emotion Lexicon Corpus Word Similarities (Loss ≈-1000)*** | | | | | | | | | |
| EmoLex | k=5 | 0.09 | -0.15 | 0.17 | 0.20 | 0.13 | 0.14 | -0.03 | -0.20 | 0.14 |
| EmoLex | k=10 | 0.09 | -0.15 | 0.17 | 0.18 | 0.14 | 0.15 | -0.02 | -0.17 | 0.13 |
| EmoLex | k=50 | 0.06 | -0.17 | 0.17 | 0.12 | 0.14 | 0.15 | 0.01 | -0.16 | 0.12 |
| EmoLex | k=100 | 0.06 | -0.16 | 0.18 | 0.11 | 0.14 | 0.15 | 0.03 | -0.17 | 0.13 |
| NRC-EIL | k=10 | 0.10 | -0.16 | 0.07 | 0.08 | 0.17 | 0.16 | 0.04 | -0.13 | 0.11 |
| NRC-EIL | k=100 | 0.04 | -0.15 | 0.15 | 0.12 | 0.16 | 0.16 | 0.02 | -0.15 | 0.12 |
| NRC-EIL | k=300 | 0.04 | -0.15 | 0.15 | 0.11 | 0.15 | 0.15 | 0.02 | -0.16 | 0.12 |
|  | ***Basic Emotion Word Similarities (Loss ≈-500)*** | | | | | | | | | |
| Word2Vec |  | 0.30 | 0.03 | 0.28 | 0.28 | 0.39 | 0.11 | 0.05 | **0.22** | 0.21 |
|  | ***Emotion Lexicon Corpus Word Similarities (Loss ≈-500)*** | | | | | | | | | |
| EmoLex | k=5 | 0.09 | **-0.20** | 0.18 | 0.24 | 0.19 | 0.19 | -0.08 | -0.16 | 0.17 |
| EmoLex | k=10 | 0.11 | -0.19 | 0.25 | 0.21 | 0.18 | 0.22 | -0.06 | -0.14 | 0.17 |
| EmoLex | k=50 | 0.08 | -0.18 | 0.22 | 0.16 | 0.17 | 0.19 | 0.00 | -0.13 | 0.14 |
| EmoLex | k=100 | 0.08 | -0.17 | 0.23 | 0.14 | 0.17 | 0.19 | 0.02 | -0.14 | 0.14 |
| NRC-EIL | k=10 | 0.06 | -0.16 | 0.10 | 0.18 | 0.25 | 0.21 | 0.04 | -0.08 | 0.14 |
| NRC-EIL | k=100 | 0.07 | -0.15 | 0.17 | 0.14 | 0.21 | 0.21 | 0.00 | -0.11 | 0.13 |
| NRC-EIL | k=300 | 0.07 | -0.15 | 0.16 | 0.14 | 0.18 | 0.21 | 0.02 | -0.13 | 0.13 |
|  | ***Human Annotation*** | | | | | | | | | |
|  |  | 0.67 | 0.74 | 0.65 | 0.59 | 0.78 | 0.69 | 0.64 | 0.72 | 0.69 |

*Table X* Pearson Correlation scores for all considered prediction methods following vector-space modification- method 2.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Source | Variant | Anger | Anticip | Disgust | Fear | Joy | Sadness | Surprise | Trust | Average |
|  | ***Basic Emotion Word Similarities (Both Modified Vectors)*** | | | | | | | | | |
| Word2Vec |  | 0.19 | -0.13 | 0.21 | 0.16 | 0.20 | 0.11 | -0.08 | -0.12 | 0.15 |
|  | ***Emotion Lexicon Corpus Word Similarities (Both Modified Vectors)*** | | | | | | | | | |
| EmoLex | k=5 | 0.10 | -0.13 | 0.08 | 0.19 | 0.16 | 0.14 | -0.07 | -0.07 | 0.12 |
| EmoLex | k=10 | 0.09 | -0.11 | 0.10 | 0.16 | 0.16 | 0.16 | 0.02 | -0.04 | 0.11 |
| EmoLex | k=50 | 0.06 | -0.12 | 0.11 | 0.10 | 0.17 | 0.13 | 0.02 | -0.03 | 0.09 |
| EmoLex | k=100 | 0.06 | 0.11 | 0.10 | 0.09 | 0.16 | 0.13 | 0.03 | -0.04 | 0.09 |
| NRC-EIL | k=10 | 0.01 | -0.09 | 0.07 | 0.05 | 0.21 | 0.09 | -0.02 | 0.00 | 0.07 |
| NRC-EIL | k=100 | 0.02 | -0.07 | 0.11 | 0.08 | 0.16 | 0.11 | 0.01 | 0.02 | 0.07 |
| NRC-EIL | k=300 | 0.01 | -0.08 | 0.09 | 0.08 | 0.15 | 0.12 | 0.03 | 0.00 | 0.07 |
|  | ***Basic Emotion Word Similarities (Modified Word Vectors Only)*** | | | | | | | | | |
| Word2Vec |  | **0.36** | -0.11 | **0.38** | 0.41 | **0.44** | 0.18 | -0.11 | **-0.24** | 0.28 |
|  | ***Emotion Lexicon Corpus Word Similarities (Modified Word Vectors Only)*** | | | | | | | | | |
| EmoLex | k=5 | 0.24 | -0.17 | 0.18 | **0.42** | 0.34 | 0.26 | **-0.16** | -0.11 | 0.24 |
| EmoLex | k=10 | 0.22 | -0.16 | 0.25 | 0.33 | 0.30 | 0.28 | -0.06 | -0.05 | 0.21 |
| EmoLex | k=50 | 0.19 | **-0.20** | 0.27 | 0.25 | 0.31 | 0.27 | -0.02 | -0.09 | 0.20 |
| EmoLex | k=100 | 0.20 | -0.19 | 0.29 | 0.24 | 0.30 | 0.28 | 0.00 | -0.12 | 0.20 |
| NRC-EIL | k=10 | 0.12 | -0.14 | 0.14 | 0.23 | 0.40 | 0.16 | -0.04 | -0.04 | 0.16 |
| NRC-EIL | k=100 | 0.14 | -0.11 | 0.22 | 0.22 | 0.31 | 0.27 | -0.01 | -0.06 | 0.17 |
| NRC-EIL | k=300 | 0.15 | -0.13 | 0.21 | 0.22 | 0.27 | **0.29** | 0.03 | -0.08 | 0.17 |
|  | ***Human Annotation*** | | | | | | | | | |
|  |  | 0.67 | 0.74 | 0.65 | 0.59 | 0.78 | 0.69 | 0.64 | 0.72 | 0.69 |

**Emoji Sentiment Prediction- Basic Theory**

*Data Preparation*

*Feature scaling*

The available data for the task falls into several classes: basic emotions (8 features), dimensional theory polar emotions (2 features), degree of neutrality and sentiment score. This presents a challenge with regards to feature scaling. It is unlikely that a simplistic strategy of scaling evenly across all features will appropriately represent the patterns present thus an alternative strategy was implemented. The aforementioned four parameter classes were scaled relative to each other using a simplistic min-max scaler. A secondary scaling was applied within the two classes containing multiple features to scale them proportionate to their class. This approach considers that not every emoji evokes equal magnitude of emotion and prevents the disproportionate consideration of one emotional model based on the quantity of features it contains.

*Mitigation of skew in target variables*

Four approaches were evaluated for each basic emotion to address the skew in the data:

*Table X* Strategies evaluated to mitigate skew for each basic emotion.

|  |  |
| --- | --- |
| **Transformation** | **Formula** |
| Log transform |  |
| Square root transformation |  |
| Cube root transformation |  |
| Reciprocal transformation |  |

And the degree of improvement was defined by the skewness metric:

Where E is the expectation operator and X is a random variable and the best identified outcome is that which is closest to zero.

*Table X* Results of skewness transformation.

|  |  |  |
| --- | --- | --- |
| **Basic Emotion** | **Best transformer** | **Best skew value obtained** |
| Anger | Reciprocal transformation | -1.176 |
| Anticipation | Cube root transformation | 0.001 |
| Disgust | Reciprocal transformation | -0.706 |
| Fear | Reciprocal transformation | -0.095 |
| Joy | No transformation | 0.040 |
| Sadness | Reciprocal transformation | -1.171 |
| Surprise | Reciprocal transformation | -0.031 |
| Trust | Log transformation | 0.109 |

Each optimal transformation was performed on the respective data and stored in separate columns. This action was performed as the performance of certain regression models is adversely affected by skew in data. Where these parameters are used to yield outcomes for regression tasks a reverse operation can be performed on predictions to revert the data to its original form.

*Assessment of components*

*Data volume:* The dataset contains 150 rows for use in the regression task, a notably small volume of data which will likely limit the model selection to more simplistic options, which perform better on smaller volumes of data.

*Parameters for training:* A significant volume of approximations have been generated prior to and following vector adjustments. Given parameters which display greater linear correlation to the target variable will be of greater value to the model, three subsets of appropriations with the greatest similarity to the target variable will be selected in each case to aid in model training. Based on the same logic, dimensionality reduction during optimisation will also be performed with consideration to similarity.

*Target variables:* Both target variables with skew mitigated and without transformation will be evaluated and the option which yields the best results will be used. Previous data evaluation which focused on correlation data highlighted that basic emotions sometimes display correlation or negative correlation to the others which indicates that a multivariate regression approach may be viable. Such an approach would account for both the relationship between the target and input parameters in addition to the relationships between each target variable. Traditional univariate regression may provide a more accurate result due to the small data volume limiting the bandwidth to increase complexity thus both options will be considered.

*Model selection*

Model selection was carried out using the following approach:

A diagram of a company

Description automatically generated

*Figure X* Process to select model for basic theory regression task.

The goal of the approach was to evaluate each considered model in its most robust, optimised configuration to ensure the most accurate model could be identified. Each step in this process is discussed in section X.

*Model Evaluation*

Table X reports the results of model evaluation for each basic emotion. The goal was to identify the model which yielded an outcome which minimised mean absolute error and mean squared error, while maximising R2 score. Where the optimal metrics for a basic emotion were distributed across multiple models the determination of the optimal model considered relative trade-offs between the metrics in each case. For example, where a model had a slightly lower mean absolute error than its alternative, but a much higher R2 score, a cost-benefit analysis would favour this model. Metrics for training, validation and test were all considered in the context of one another, where close alignment was viewed with preference, as this signals good generalisation.

As the models have varying means for their respective values it was important to consider the mean when assessing model performance in each case. Table X uses the mean values to provide an indication of the accuracy of the models. While this cannot consider the distribution of error across the set, the low mean squared errors achieved in each optimal model indicates that error is broadly distributed in proportion to the value of the label. Most models had moderate abilities to explain the underlying patterns in the training data, evidence by their respective R2 scores. However, this was not the case for *anticipation* and *surprise*, which had poor performance. This consolidates previous hypothesis relating to the limitations of the proposed objective functions to modify the word vectors as these emotions are distinct from the others in that they can apply in both positive and negative contexts and are distinguished by additional features.

Accuracy scores averaging 90% were achieved for the basic emotion regression (range 86-92%). A likely limitation of the performance was the limited data for training. An expansion of the annotated data may improve the outcomes of these models or enable the use of models capable of capturing more complex patterns in future works. Supplementary features gleaned from other sources may have improved outcomes for this task, however the use of word vectors only to for feature extraction makes expansion upon this dataset more accessible as the emoji corpus expands, which is a primary strength of this approach.

*Table X* Results of model development process for basic theory regression displaying the best outcomes for each model, correct to three significant figures. The selected optimal model is in bold for each basic emotion. Note that reported metrics are with reference to validation performance.

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **Performance Metric** | | |
| **Mean Absolute Error** | **Mean Squared Error** | **R2 Score** |
|  | **Anger** | | |
| Linear Regression | 0.0112 | 0.000230 | 0.148 |
| Random Forest | 0.0136 | 0.000290 | -1.41 |
| XGBoost Regression | 0.0127 | 0.000283 | -1.35 |
| Support Vector Regression | 0.0318 | 0.00112 | -2.63 |
| **Gaussian Process Regression** | 0.00915 | 0.000171 | 0.365 |
| KNN Regression | 0.0110 | 0.000292 | -0.0858 |
|  | **Anticipation** | | |
| Linear Regression | 0.0118 | 0.000217 | -0.238 |
| Random Forest | 0.0122 | 0.000201 | -0.143 |
| XGBoost Regression | 0.0142 | 0.000251 | -0.432 |
| Support Vector Regression | 0.0136 | 0.000241 | -0.373 |
| Gaussian Process Regression | 0.0117 | 0.000185 | -0.0557 |
| **KNN Regression** | 0.0114 | 0.000189 | -0.0735 |
|  | **Disgust** | | |
| **Linear Regression** | 0.0121 | 0.000270 | 0.310 |
| Random Forest | 0.0145 | 0.000419 | -0.0697 |
| XGBoost Regression | 0.0147 | 0.000456 | -0.164 |
| Support Vector Regression | 0.0206 | 0.000543 | -0.385 |
| Gaussian Process Regression | 0.0126 | 0.000273 | 0.303 |
| KNN Regression | 0.0159 | 0.000419 | -0.0691 |
|  | **Fear** | | |
| Linear Regression | 0.0119 | 0.000319 | 0.0607 |
| Random Forest | 0.0162 | 0.000351 | 0.0578 |
| XGBoost Regression | 0.0183 | 0.000477 | -0.282 |
| Support Vector Regression | 0.0263 | 0.000840 | -1.25 |
| **Gaussian Process Regression** | 0.0113 | 0.000170 | 0.543 |
| KNN Regression | 0.0158 | 0.000397 | -0.0670 |
|  | **Joy** | | |
| Linear Regression | 0.0236 | 0.00114 | -0.0148 |
| Random Forest | 0.0313 | 0.00147 | -0.312 |
| XGBoost Regression | 0.0350 | 0.00186 | -0.663 |
| Support Vector Regression | 0.0295 | 0.00115 | -0.0254 |
| **Gaussian Process Regression** | 0.0199 | 0.000614 | 0.451 |
| KNN Regression | 0.0287 | 0.00113 | -0.00802 |
|  | **Sadness** | | |
| **Linear Regression** | 0.0152 | 0.000381 | 0.417 |
| Random Forest | 0.0166 | 0.000669 | -0.0241 |
| XGBoost Regression | 0.0159 | 0.000462 | 0.292 |
| Support Vector Regression | 0.0303 | 0.00106 | -0.619 |
| Gaussian Process Regression | 0.0144 | 0.000408 | 0.375 |
| KNN Regression | 0.0189 | 0.000721 | -0.104 |
|  | **Surprise** | | |
| **Linear Regression** | 0.0103 | 0.000156 | -0.0299 |
| Random Forest | 0.0106 | 0.000178 | -0.0484 |
| XGBoost Regression | 0.0131 | 0.000246 | -1.05 |
| Support Vector Regression | 0.0141 | 0.000270 | -1.25 |
| Gaussian Process Regression | 0.00973 | 0.000141 | -0.173 |
| KNN Regression | 0.0111 | 0.000199 | -0.659 |
|  | **Trust** | | |
| **Linear Regression** | 0.0128 | 0.000303 | 0.202 |
| Random Forest | 0.0169 | 0.000495 | -0.300 |
| XGBoost Regression | 0.0173 | 0.000530 | -0.393 |
| Support Vector Regression | 0.0176 | 0.000410 | -0.0776 |
| Gaussian Process Regression | 0.153 | 0.000519 | -0.411 |
| KNN Regression | 0.0171 | 0.000484 | -0.272 |

*Table X* Assessment of model accuracy considering the mean values of each basic emotion.

|  |  |  |  |
| --- | --- | --- | --- |
| **Basic Emotion** | **Mean Value for Affinity to Emoji** | **Mean Absolute Error** | **% Accuracy Approximation** |
| Anger | 0.122 | 0.00915 | 92.50 |
| Anticipation | 0.137 | 0.0114 | 91.68 |
| Disgust | 0.102 | 0.0121 | 88.14 |
| Fear | 0.120 | 0.0113 | 90.58 |
| Joy | 0.138 | 0.0199 | 85.58 |
| Sadness | 0.118 | 0.0152 | 87.12 |
| Surprise | 0.122 | 0.0103 | 92.20 |
| Trust | 0.131 | 0.0128 | 90.23 |

**Emoji Sentiment Prediction- Dimensional Theory**

The nature of dimensional theory-based sentiment metrics does not facilitate solutions similar to that which was implemented for basic theory parameters as word vectors associated with *positive* and *negative* vocabulary would necessitate the consideration of context and topic to a much greater extent making this a more complex and nuanced problem set. No human-annotated datasets were identified which provide equivalent data to the EmoLex of NRL-EIC lexicons thus the proposed adaptation of the previous methodology (a) compared emoji vectors directly to the vectors of the respective labels, and (b) compared label vectors to the vocabulary of the iSarcasm dataset. Such vocabulary was selected as it represents the vernacular of the population which may use sarcasm in online content thus the words highly associated with the emojis based on their vectors were likely to be relevant choices.

Results obtained using both options performed poorly with respect to generating predictions which correlated to the human-annotated data, and consequently was not useful in predicting information for emojis unseen in the dataset. This approach was unable to obtain accuracies >50% and thus alternative strategies were explored.

Given the limitations in data quantities available to train vectors and the subsequent regression models, pre-trained models were postulated to generate improved outcomes as they benefit from increased volume of training data and computational power available during their training process. Such factors yield a model capable of capturing more complex relationships thus results which mirror human annotations more closely. Several pre-trained models were assessed with regards to their abilities to predict dimensional theory data which correlated to human annotations assessed via Pearson’s correlation. Models selected for assessment based on suitability of architecture and use in similar problem sets within literature (see appendix X).

*Evaluation of results and methodology*

The optimal model identified was the latest Twitter RoBERTa Base Sentiment which achieved a Pearson’s correlation of 0.83 to the human annotated sentiment scores. Fine-tuning the model using the human-annotated data did not result in improvement to this score, likely due to the small quantity of data available for the purpose therefore the model was deployed for the regression task without optimisation.

While the results obtained yielded good correlation to the human annotated data, the input of a single emoji in each case to obtain the results does not necessarily align with the intended use for the model; the attention mechanisms cannot provide contextual information from surrounding information and thus outputs may be limited to some degree by ambiguity. Alternative methodologies would likely involve the input of emoji-containing strings of text, however in the context of the problem set this also has limitations; the sentiment may be skewed by the content of the text. This concept also applies to some extent even where the emoji is inputted in isolation due to the training data. While the approach has limitations, the solution would involve the obtaining of a larger training dataset with annotations. Such a task is outside the scope of this work however may be an avenue for future consideration.