**Assigning emotion intensities to tweets using Machine Learning**

CodaLab username: Indranuj Banerjee  
Note: Profile created on [new server](https://codalab.lisn.upsaclay.fr/) as [old server](https://competitions.codalab.org/) did not allow profile creation. WASSA-2017 Shared Task on Emotion Intensity (EmoInt) competition not available on new server.

Abstract:

This report focuses on the objective of assigning intensity scores to emotions expressed in textual data. The study leverages the WASSA dataset, comprising text data extracted from tweets, annotated with corresponding emotions, and associated emotion intensity scores regarded as the gold standard. The report introduces machine learning methodologies for emotion intensity score assignment, encompassing both statistical methods and neural networks.

Introduction:

The proposed approach for assigning scores to labelled text was introduced [1]. The motivation behind this method is underscored by practical needs, such as optimizing a commercial customer satisfaction system. In such a system, addressing requests that convey higher intensities of frustration requires more urgent attention compared to those with milder intensities. In addition to outlining the necessity for this concept, the paper also delineates several challenges associated with its implementation, stating that ---

“A notable obstacle in developing automatic affect intensity systems is the lack of suitable annotated data. Existing affect datasets are predominantly categorical. Annotating instances for degrees of affect is a substantially more difficult undertaking: respondents are presented with greater cognitive load and it is particularly hard to ensure consistency (both across responses by different annotators and within the responses produced by an individual annotator).”

To combat annotation inconsistencies, the researchers applied the Best-Worst Scaling (BWS) method, demonstrating its efficacy in assigning scores to annotations and establishing the gold standard for intensity scores. Leveraging these scores, this report introduces machine learning techniques designed to assign scores to emotions in a similar manner. The analysis delves into key factors influencing the scores and explores methods for representing text in a machine-interpretable manner.

Methodology:

As stated in [1], it is specific keywords within a text, particularly in tweets, that are crucial in determining the expressed emotion. For example, words such as "joyful," "frustrated," and "excited" serve as clear indicators of the underlying emotions. An additional advantage of using tweets is the inclusion of highly expressive hashtags, such as #thrilled or #disappointed, which function as special keywords and can be leveraged to ascertain the intensity of the emotion being expressed.

Therefore, the methodology employed in this task integrates natural language processing along with neural networks, involving the following steps:

1. Extract Hashtags: Extract hashtags from tweets, if available.
2. Text Cleaning: Perform regular text cleaning procedures, including stopword removal, lemmatization, lowercasing, and the elimination of other undesired elements related to tweets such as emojis, user handles, and URLs. The primary goal of this step is to filter out significant keywords from the tweet.
3. Word Embedding: Represent the words from the cleaned texts as numbers.
4. Developing a single model to predict the emotion intensity of all emotions.

Given the pre-annotation of tweets with corresponding emotions, the methodology focusses solely on assigning emotion intensities through the presence of keywords, without considering their spatial or contextual relationships. Consequently, the task is regarded as a regression problem, with words serving as features. In this context, TF-IDF vectorization emerges as a reasonable method for embedding words.

The importance of a word in a tweet is conveyed through TF-IDF vectorization, which considers its frequency across all tweets. This process incorporates an element of the Bag of Words (BoW) methodology as well.

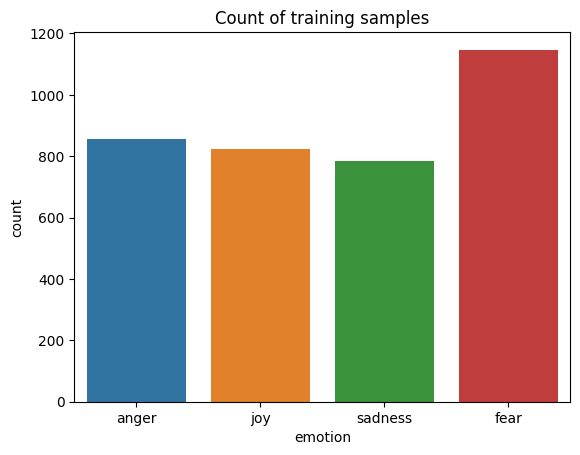
The data:

Fig. 1: Count of training samples per emotion. While the distribution is relatively balanced, this bar chart supports the earlier statements regarding the limited availability of annotated data in this domain.

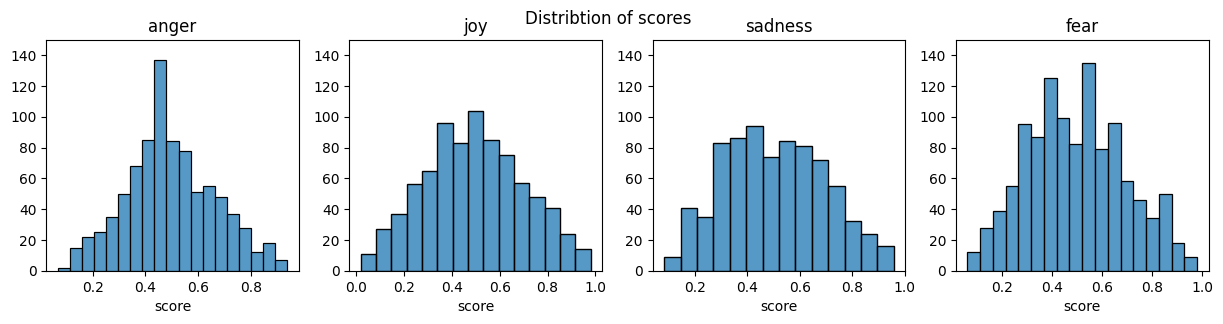
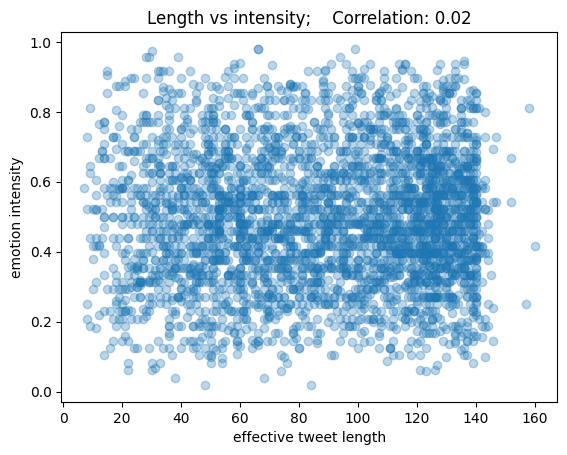


Fig. 2: Distribution of scores per emotion. It is important to clarify that these scores do not represent probabilities of the emotions but rather their intensities. The emotions themselves are considered certain.

Fig. 3: Correlation between the effective length of a tweet (excluding user handles, URLs, etc.) and its associated score. This figure disproves the idea of a positive correlation between these two variables.

Observations:

1. Statistical model – K Nearest Neighbor regressor

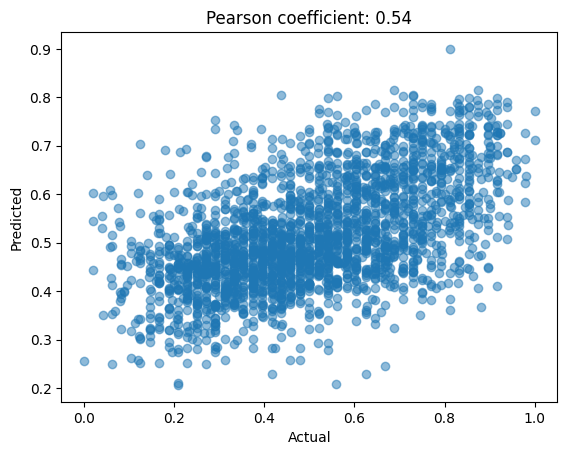
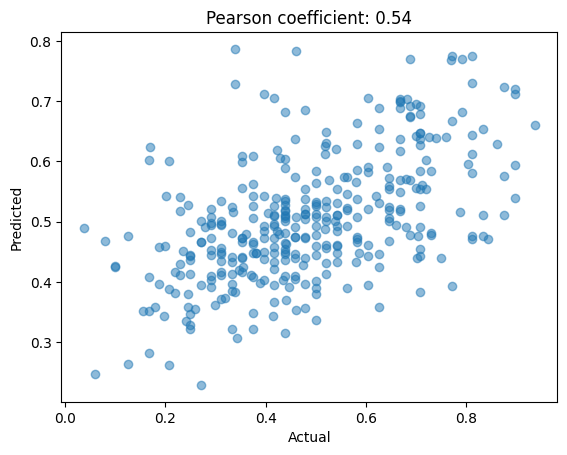


Fig. 4: Output of KNN model on development dataset (left) and test dataset (right)

1. Neural network model – Multi layer perceptron

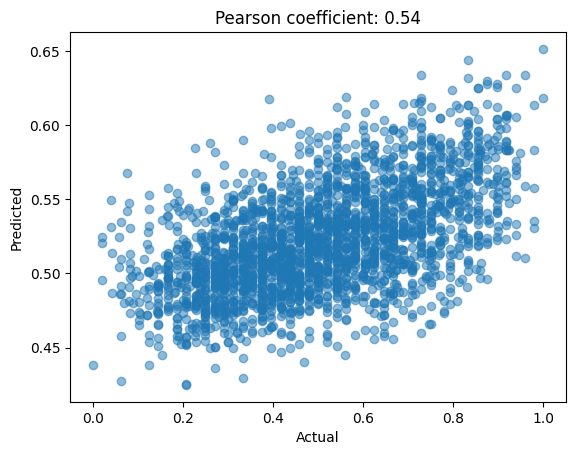
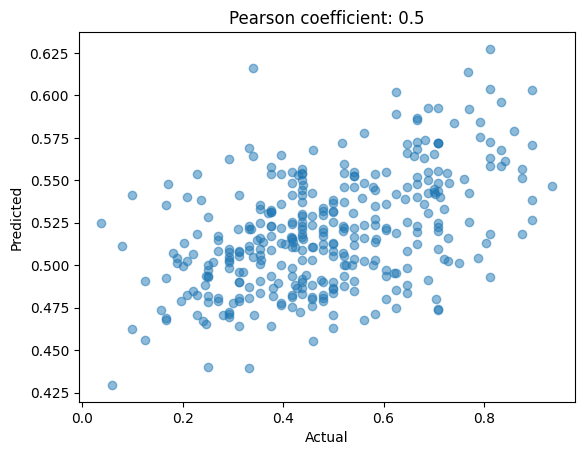


Fig. 5: Output of MLP model on development dataset (left) and test dataset (right)

Summary:

* K nearest neighbors outperforms linear regression or random forest.
* Including hashtags in the tweets is beneficial for obtaining results.
* As the evaluation script provided [here](https://github.com/felipebravom/EmoInt) is not reproducible, model tuning has been performed with the “pearsonr” function from scipy. Additional metrics such as Spearman rank coefficient have not been calculated.
* Limited data is a major challenge for coming up with a generalized model that utilizes dynamic data processing which is robust enough for all 4 emotions.
* A wide range in tweet lengths is a challenge as well, as stopwords comprise a big chunk of most tweets, their removal creates some ambiguity in the remaining data. Additionally, hashtags are sometimes written as words clubbed together (e.g., #JoyfulWeekend) which are difficult to split and therefore, their effective contribution is lost.
* The above methodology is promising and with more annotated data, it will be possible to come up with a model that reliably assigns intensity scores to emotions. As of now, the above models have generated scores with a Pearson correlation coefficient of 0.54 on average, which denotes a moderate positive correlation which means the models are not able to assign very high or very low intensity scores.

# References

|  |  |
| --- | --- |
| [1] | &. F. B.-M. Saif M. Mohammad, *Emotion Intensities in Tweets.,* 2017. |