A Transformers Approach to Create Emotional Profiles From Text

Amirreza Dashti Genave

Abstract

In this project, we explore the application of text mining methodologies to predict emotional profiles from text inputs. Specifically, we utilize a pre-trained BERT model combined with regression techniques to predict the valence, arousal, and dominance dimensions of emotions. We present an overview of our approach, including data preprocessing, model training, and evaluation. Furthermore, we demonstrate how the trained model can be applied to analyze emotional content in text, providing insights into the emotional states conveyed by the input text. Additionally, our transformers-based model can be employed to analyze the emotional profiles of characters in movies or TV series, offering a nuanced understanding of their psychological trajectories and emotional dynamics. This tool has the potential to enhance our comprehension of psychological profiles and emotional journeys solely through text analysis, contributing to deeper insights into character development and narrative themes.

1 Introduction

Text mining, a subfield of natural language processing (NLP), involves extracting useful information and patterns from textual data. One area of interest in text mining is sentiment analysis, which focuses on understanding the emotions expressed in text. Sentiment analysis has gained significant attention in recent years due to its applications in various fields such as marketing, customer feedback analysis, and social media monitoring. [1]

Emotions are complex psychological constructs that encompass various dimensions, including valence (positive/negative), arousal (intensity), and dominance (control). These dimensions are often used to characterize the affective content of text and understand the underlying emotional states of individuals. Valence, often referred to as the positivity or negativity of emotions, represents the degree of pleasantness or unpleasantness associated with an emotional state. Arousal measures the intensity or activation level of emotions, ranging from calmness to excitement or agitation. Dominance reflects the perceived control or power dynamics within emotional experiences, indicating whether the individual feels in control or subordinate in a given situation. [2]

Predicting these emotional dimensions from text can provide valuable insights into the underlying sentiments and affective states conveyed by the text. By analyzing emotional dimensions, researchers and practitioners can better understand user preferences, behavior patterns, and psychological well-being. Moreover, it enables personalized recommendation systems, targeted advertising strategies, and sentiment-aware decision-making processes.

In this report, we present a text mining project focused on predicting emotional profiles from text inputs. Leveraging deep learning models, particularly BERT [3] (Bidirectional Encoder Representations from Transformers [4]), we aim to learn rich representations of textual data that capture the nuanced semantics and emotional nuances present in the text. BERT, a state-of-the-art language representation model, has shown remarkable performance in various NLP tasks, including sentiment analysis, text classification, and question answering.

Our approach involves fine-tuning a pre-trained BERT model on a dataset of annotated emotional expressions. Fine-tuning allows us to adapt the model's parameters to the specific task of predicting emotional dimensions from text. By learning from annotated data, the model can capture the relationships between textual features and emotional labels, enabling it to generalize to unseen text inputs effectively.

Our goal is to develop a model using deep learning techniques, specifically BERT, to accurately predict the emotional profiles of actors/actresses in movies or TV series. By analyzing character emotional dynamics throughout a storyline, the model offers insights into the evolution of fictional personalities. Once trained, it can analyze dialogue transcripts, screenplay texts, or subtitles to infer character emotions at different narrative points, aiding researchers, filmmakers, and industry professionals in understanding character development, plot dynamics, and thematic elements. This approach uncovers patterns in character emotions, enriching the understanding of psychological journeys and enhancing audience engagement and storytelling effectiveness.

2 Research question and methodology

This project aims to utilize transformer models to predict the three parameters of emotions throughout an entire movie or TV series, thereby enabling the exploration of how these emotional parameters evolve over the course of the narrative. By leveraging a transformer-based regressor model, we seek to predict the values corresponding to each norm of Valence, Arousal, and Dominance, which collectively constitute the emotional landscape of the characters. Drawing from the framework of discrete emotion models proposed by Oana Bălan et al. [5], our objective is to translate these predicted values into a human-understandable emotional profile for each character within the movie or TV series. Our methodology involves several key steps: firstly, employing a regressor model atop the transformer architecture to derive numerical predictions for the emotional norms; subsequently, interpreting these predictions within a three-dimensional representation to identify the most probable manifestations of the six basic emotions; and finally, synthesizing the frequent occurrences of predicted emotions to construct comprehensive emotional profiles for the characters, offering insights into their psychological journeys and affective states throughout the narrative. Through this approach, we aim to provide a nuanced understanding of character emotions and their narrative significance within cinematic and storytelling contexts.

3 Experimental results

3.1 Dataset

To train our model, we employed the Emobank dataset [6, 7], a comprehensive text corpus annotated with emotions based on the psychological Valence-Arousal-Dominance scheme. Developed by the JULIE Lab at Jena University, this dataset encompasses various columns including id, split, V, A, D, and text. With a collection of 10,000 sentences spanning diverse genres, each sentence is annotated by both writers and readers. However, for our purposes, we focused on a file that presents a weighted average of annotations from both parties, ensuring a balanced perspective for our model training.

3.2 Data Splitting

The dataset used in this project already contains a predefined column named `split`, which specifies whether a given data point belongs to the training set or the test set. The values in this column are either `train` or `test`. We utilized this column directly to separate the dataset into training and test sets, ensuring consistency with the original dataset's design.

3.3 Creating the `EmoBankDataset` with PyTorch

The `PyTorch` library was employed to create a custom dataset class, `EmoBankDataset`. This class inherits from `torch.utils.data.Dataset` and was designed to handle the EmoBank data, which contains annotations for valence, arousal, and dominance. The `__getitem__` method was overridden to return tokenized text data along with their corresponding emotional labels. The `transformers` library was used for tokenization, specifically utilizing BERT's tokenizer to convert text into tokens suitable for model input.

3.4 Backbone Model

The backbone of our model is a pre-trained BERT model, specifically `bert-base-uncased` from HuggingFace's Transformers library. BERT was chosen for its proven effectiveness in various NLP tasks, providing a robust foundation for emotion prediction.

3.5 Creating and Integrating `EmoRegressor`

We designed a custom regressor, `EmoRegressor`, to predict the emotional dimensions from the output of the BERT model. The regressor consists of a fully connected layer with 512 nodes, followed by a `ReLU` activation function, and finally, an output layer producing three values corresponding to valence, arousal, and dominance.

Choice of Activation Function and Node Count: The `ReLU` activation function was selected for its simplicity and effectiveness in handling non-linearity. We chose 512 nodes as a balanced choice to ensure sufficient capacity for the model to learn complex relationships without overfitting.

Attention Mask: The `attention_mask` parameter in the BERT tokenizer was utilized to indicate which tokens should be attended to by the model. This mask ensures that padding tokens are not considered in the self-attention mechanism, thereby preventing them from influencing the model's predictions.

3.6 Training Procedure

The model was trained using the Mean Squared Error (MSE) loss function, `nn.MSELoss()`. MSE was chosen as the criterion because it directly measures the difference between predicted and true values, making it suitable for continuous-valued predictions like valence, arousal, and dominance.

Optimizer: The AdamW optimizer was used for training, with weight decay applied to prevent overfitting. AdamW was chosen for its efficiency and effectiveness in handling large-scale models like BERT.

Training Details: The model was trained for 10 epochs with a learning rate of 5×10^{-5}. We used a batch size of 16, balancing memory constraints and the need for efficient training.

3.7 Evaluation Metrics

To evaluate the performance of the model, we employed several metrics:

- **Mean Squared Error (MSE):**

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

  MSE measures the average squared difference between the predicted and actual values. A lower MSE indicates better performance. For our model, the MSE was 0.0471.

- **Root Mean Squared Error (RMSE):**

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

  RMSE provides a measure of error in the same units as the original values. Our model achieved an RMSE of 0.2170.

- **Mean Absolute Error (MAE):**

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

  MAE calculates the average magnitude of errors in predictions, offering a straightforward interpretation. The MAE for our model was 0.1652.

- **R-squared ($\mathbf{R^2}$):**

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

  R^2 represents the proportion of variance in the dependent variable that is predictable from the independent variables. Our model had an R^2 of 0.3021, indicating a moderate fit.

Interpretation: The results indicate that while the model shows a reasonable fit, there is still room for improvement in capturing the variance in emotional dimensions.

3.8 Application to Real-world Data: Joker (2019) Analysis

To further evaluate our model, we extracted the dialogues of the main character from the movie *Joker* (2019) [8]. We developed a function named `extract_dialogues` to isolate and prepare the dialogue text for analysis.

Prediction and Scaling: The model predicted valence, arousal, and dominance for each dialogue line. The predicted values were then scaled from -1 to 1, following the standard practice in emotion analysis.

Visualizing the Evolution: The evolution of each VAD parameter throughout the movie is illustrated in the plots below. These plots reveal the emotional trajectory of the character as the narrative progresses.

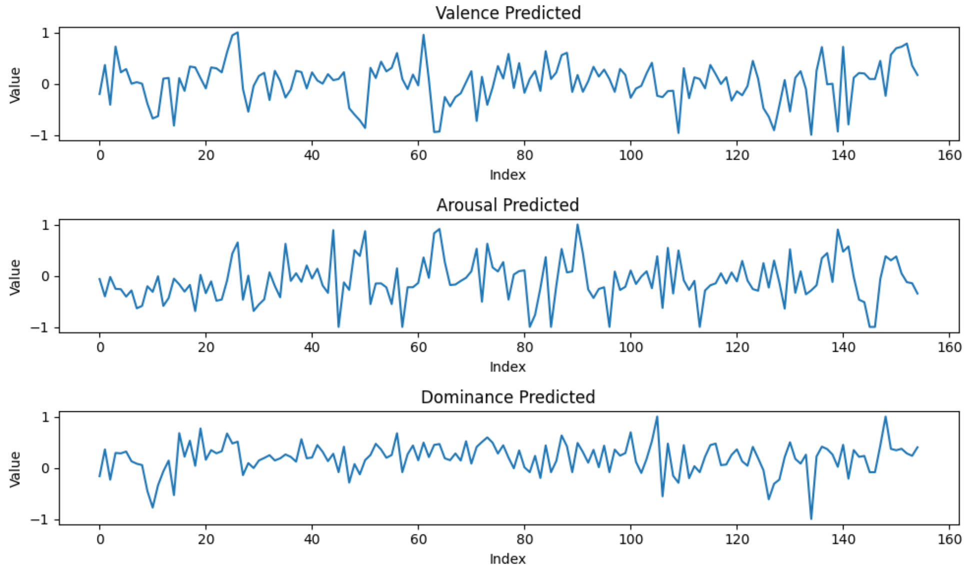

Figure 1: Evolution of Valence, Arousal, and Dominance over the course of the movie *Joker* (2019). These plots highlight the emotional fluctuations experienced by the main character throughout the film.

3.9 Mapping to Categorical Emotions

We referenced the work of Russell and Mehrabian [2], which provided a mapping between the categorical representation of emotions and the VAD dimensions. These six emotions were further elaborated in a more recent study by Bălan et al. [5], who utilized this mapping to classify emotions using biophysical signals and machine learning techniques. This dual reference provides a robust basis for interpreting the predicted VAD values in terms of categorical emotions.

3D Plot of Predicted Emotions: We created a 3D plot marking the positions of the reference emotions according to Bălan et al.'s work. Our predicted VAD points were then plotted in the same space, allowing us to visualize the character's emotional profile throughout the movie.

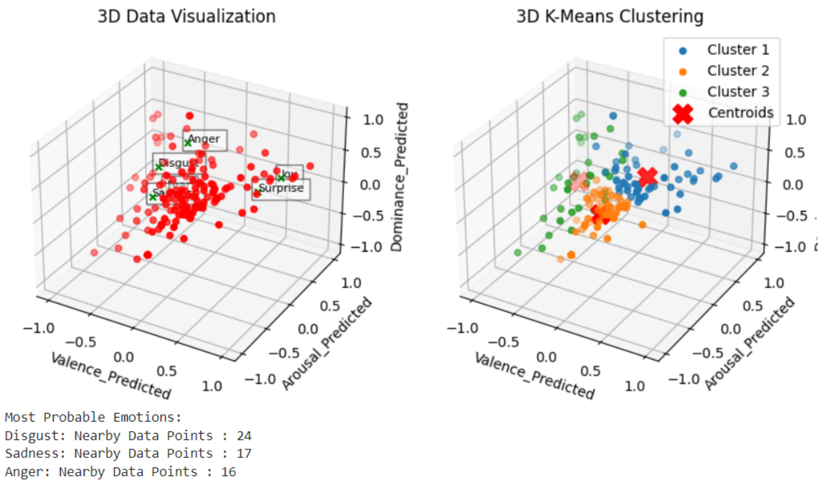

Figure 2: 3D Plot of Predicted Valence, Arousal, and Dominance values alongside categorical emotions derived from the reference work. This visualization helps in interpreting the emotional profile of the character.

3.10 Ranking and Analysis

To further understand the emotional profile, we ranked the top three most predicted points using a specific threshold. The table below lists the top emotions along with the number of nearby data points for each emotion, which helps to interpret the dominant emotional states experienced by the character.

Emotion	Nearby Data Points
Disgust	24
Sadness	17
Anger	16

Table 1: Ranking of Emotions Based on Nearby Data Points. The table shows the number of data points that are closest to the emotional states of Disgust, Sadness, and Anger, indicating the character's most dominant emotions.

3.11 Evolution of Character's Emotions

Finally, to capture the evolution of the character's emotional state throughout the movie, we created a diagram that labels each dialogue with its corresponding emotion. This diagram offers a clear depiction of how the character's emotions shift as the story unfolds.

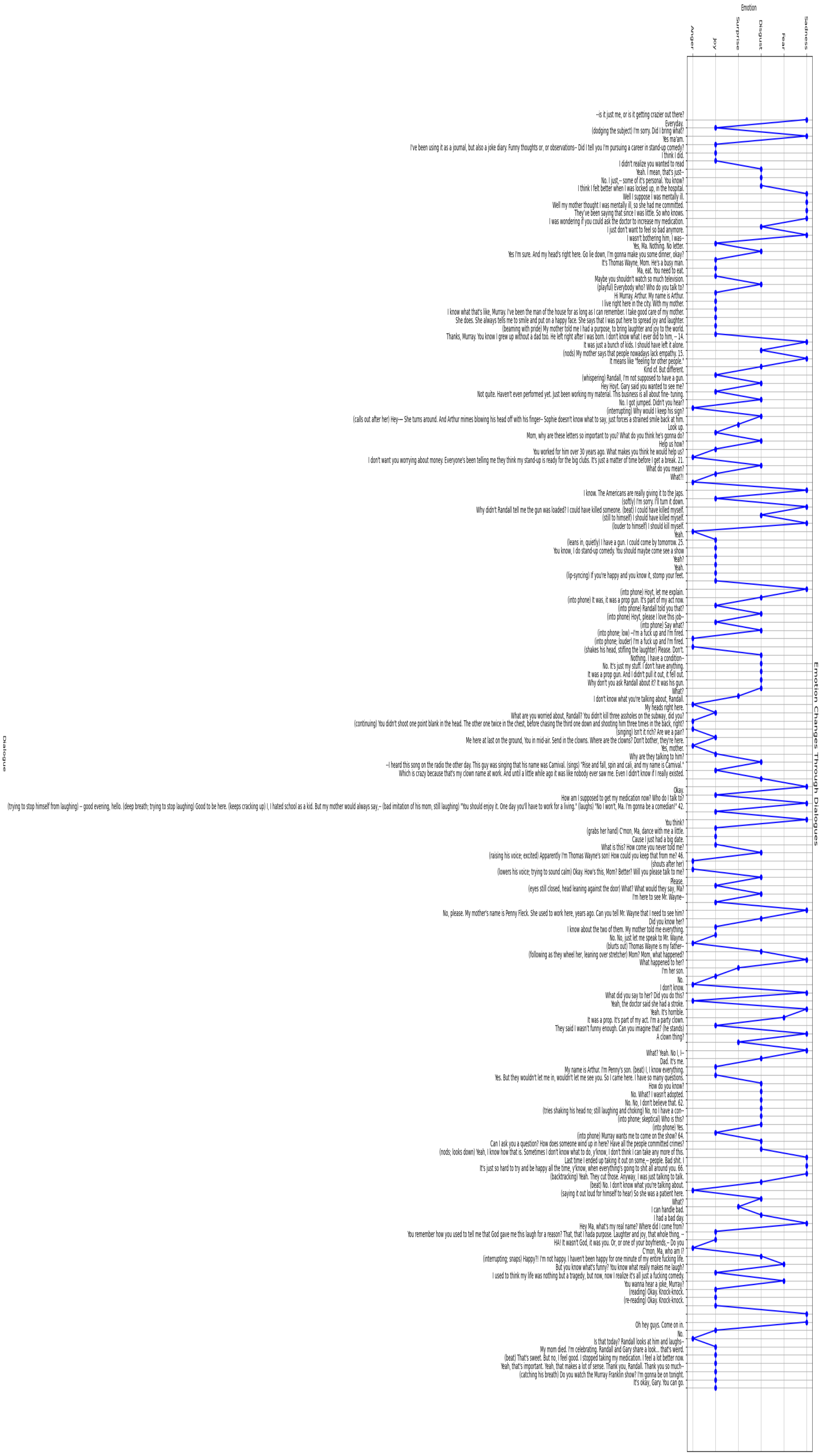

Figure 3: Diagram showing the evolution of labeled emotions for each dialogue throughout *Joker* (2019). This visualization provides a detailed look at the character's emotional journey.

4 Conclusion

In this study, we successfully developed a machine learning model using a transformer-based approach to create an emotional profile from text. Our work focused on predicting the Valence, Arousal, and Dominance (VAD) dimensions for each dialogue spoken by the main character of the movie *Joker* (2019). By leveraging the EmoBank dataset and using the BERT model as our backbone, we fine-tuned the model to generate accurate predictions for the VAD parameters.

The experimental results demonstrated the effectiveness of our approach, as indicated by the performance metrics. The model achieved a Mean Squared Error (MSE) of 0.0471, Root Mean Squared Error (RMSE) of 0.2170, Mean Absolute Error (MAE) of 0.1652, and an R-squared (R^2) value of 0.3021. These metrics suggest that while the model shows promise, there is room for improvement, particularly in enhancing the model's ability to explain the variance in the data (R^2).

Using the trained model, we extracted and analyzed the emotional profile of the main character throughout the movie. This analysis allowed us to observe the character's emotional evolution, providing valuable insights into their psychological state. By mapping the predicted VAD values to categorical emotions, based on the work of Russell and Mehrabian, we identified that emotions such as Disgust, Sadness, and Anger were predominant in the character's profile, with nearby data points of 24, 17, and 16 respectively.

The ranking and analysis of these emotions further enriched our understanding of the character's emotional journey. The creation of 3D plots and time series charts provided a visual representation of how these emotions fluctuated throughout the film, revealing key moments of emotional intensity.

Moving forward, there are several avenues for future work within the field of text mining. First, improving the model's performance by experimenting with different text-based architectures or incorporating additional textual data could lead to more accurate predictions. Second, applying this approach to a broader range of characters and movies could validate the model's generalizability and provide comparative emotional profiles across different genres or storytelling styles. Lastly, refining the model by exploring alternative NLP techniques or leveraging more advanced transformer models could further enhance its ability to accurately predict emotional trajectories from textual inputs.

In conclusion, this project not only demonstrated the feasibility of creating an emotional profile from textual data but also opened up new possibilities for exploring and understanding the emotional dynamics of characters in narrative forms.

References

[1] Charu C. Aggarwal and ChengXiang Zhai (Eds.). *Mining Text Data*. Springer Science & Business Media, 2012. DOI: 10.1007/978-1-4614-3223-4.

[2] J.A. Russell, A. Mehrabian, *Evidence for a three-factor theory of emotions. J. Res. Personal.*, 1977, 11, 273–294. DOI: 10.1016/0092-6566(77)90037-X.

[3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 4171–4186. Association for Computational Linguistics, Minneapolis, Minnesota. DOI: 10.18653/v1/N19-1423.

[4] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Scao, Sylvain Gugger, Alexander Rush. *Transformers: State-of-the-Art Natural Language Processing*. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, pages 38–45. DOI: 10.18653/v1/2020.emnlp-demos.6.

[5] Oana Bălan, Gabriela Moise, Livia Petrescu, Alin Moldoveanu, Marius Leordeanu, and Florica Moldoveanu. *Emotion Classification Based on Biophysical Signals and Machine Learning Techniques*. Symmetry 2020, 12(1), 21. DOI: 10.3390/sym12010021.

[6] Sven Buechel and Udo Hahn. *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 (EACL 2017), Volume 2, Short Papers, pages 578–585, Valencia, Spain, April 3-7, 2017. Available: http://aclweb.org/anthology/E17-2092.

[7] Sven Buechel and Udo Hahn. *Readers vs. Writers vs. Texts: Coping with Different Perspectives of Text Understanding in Emotion Annotation*. In Proceedings of the 11th Linguistic Annotation Workshop @ EACL 2017, pages 1–12, Valencia, Spain, April 3, 2017. Available: https://sigann.github.io/LAW-XI-2017/papers/LAW01.pdf.

[8] Sripaad Srinivasan. (2020). *The Joker Script*. Retrieved from Kaggle: `https://www.kaggle.com/datasets/sripaadsrinivasan/the-joker-script/data`.