Project: Emotion detection in movie scripts

Course: Information retrieval and knowledge extraction

Name: Xaver Brückner

MAT: 989819

Xaver.brueckner@studenti.unimi.it

November 2024

Abstract

This project explores sentiment analysis in movie scripts using Valence-Arousal-Dominance (VAD) modeling. Using machine learning techniques and pre-trained GloVe embeddings, we estimate emotional characteristics of movie dialogues and investigate their relationships with movie genres and main characters.

Introduction

Sentiment analysis have become essential tools for understanding human emotions through textual data. This project focused on developing a machine learning model to predict three key emotional dimensions: valence, arousal, and dominance. These dimensions are derived from the circumplex model of affect, which conceptualizes emotions on continuous scales rather than discrete categories.

The model utilized GloVe embeddings, LSTM layers, and multi-task learning to address these regression tasks simultaneously. The focus was to optimize model performance while maintaining generalizability and avoiding overfitting. This report details the approach taken, the design decisions, and the evaluation of the model.

The primary objectives of this study are:

1. Modeling Dialogue Sentiments: Predict VAD scores using LSTM models trained on movie dialogue
2. Genre Analysis: Correlate VAD scores with movie genres to discern emotional patterns
3. Character Analysis: Determine the main character's emotional spectrum and its variations across time

Data and Methodology

Data Sources

The project integrates two datasets:

1. EmoBank Dataset: Provides text with annotated VAD scores, used for training.

id	split	V	A	D	text	word_counts
10CYL068_1036_1079	train	3.00	3.00	3.20	remember said last letter	9
10CYL068_1079_1110	test	2.80	3.10	2.80	wasnt working	5
10CYL068_1127_1130	train	3.00	3.00	3.00		1
10CYL068_1137_1188	train	3.44	3.00	3.22	goodwill helps people get public assistance	8
10CYL068_1189_1328	train	3.55	3.27	3.46	sherry learned future works class could rise m...	23

2. Cornell Movie Dialogues Dataset: Contains dialogues from movies, enriched with metadata on genres and characters.

	LineID	Character	Movie	Name	text
0	1045	0	0	BIANCA	They do not!
1	1044	2	0	CAMERON	They do to!
2	985	0	0	BIANCA	I hope so.
3	984	2	0	CAMERON	She okay?
4	925	0	0	BIANCA	Let's go.

	Movie	Name	Year	Rating	Number	Genre
0	0	10 things i hate about you	1999	6.9	62847	['comedy', 'romance']
1	1	1492: conquest of paradise	1992	6.2	10421	['adventure', 'biography', 'drama', 'history']
2	2	15 minutes	2001	6.1	25854	['action', 'crime', 'drama', 'thriller']
3	3	2001: a space odyssey	1968	8.4	163227	['adventure', 'mystery', 'sci-fi']
4	4	48 hrs.	1982	6.9	22289	['action', 'comedy', 'crime', 'drama', 'thril...

Preprocessing

Key preprocessing steps included:

- Text Normalization: Conversion to lowercase, punctuation removal, and stop-word filtering
- Tokenization: Conversion of text into sequences using TensorFlow’s tokenizer
- Scaling: Min-max normalization of numerical features between 0 and 1

Embedding and Model Training

Pre-trained GloVe embeddings (100-dimensional) were used for initializing word representations. A bidirectional LSTM model with three output layers was trained for Valence, Arousal, and Dominance predictions. The Huber loss function ensured robust learning with minimal sensitivity to outliers.

Performance

The model’s performance was evaluated using Mean Absolute Error (MAE) for each of the Valence, Arousal, and Dominance dimensions. The results indicate that the model achieved an MAE of 0.0634 for Valence, 0.0850 for Arousal, and 0.0977 for Dominance. These results show that the model performed best in predicting Valence, likely due to its more stable patterns in the dataset, while Arousal and Dominance, being more dynamic, presented slightly higher error rates. The metrics demonstrate the model's effectiveness in capturing the emotional dimensions of movie dialogues.

VAD Prediction Pipeline

Using the trained model, VAD scores were predicted for each line of dialogue. Averages were computed at multiple levels:

1. Movie-Level: Aggregate VAD scores across all dialogues in a movie
2. Genre-Level: Aggregate VAD scores for movies within the same genre
3. Character-Level: Focused analysis on the most frequently occurring character in each movie i.e. the main character

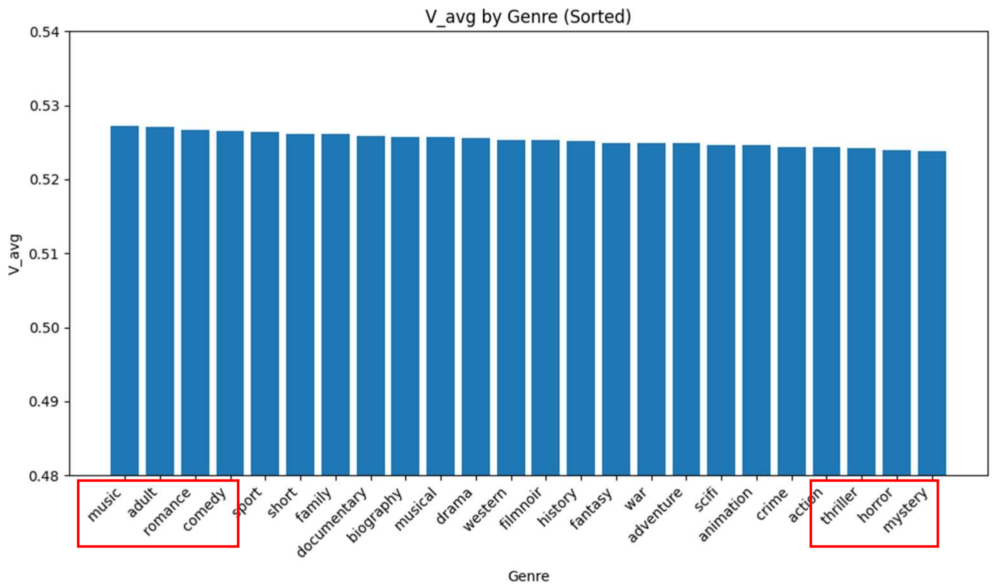
Results and Discussion

Genre-Level Insights

The analysis revealed small but consistent and humanly interpretable differences in the average VAD scores across genres, at least in the extremes. The inbetween zones are harder to interpret. For example:

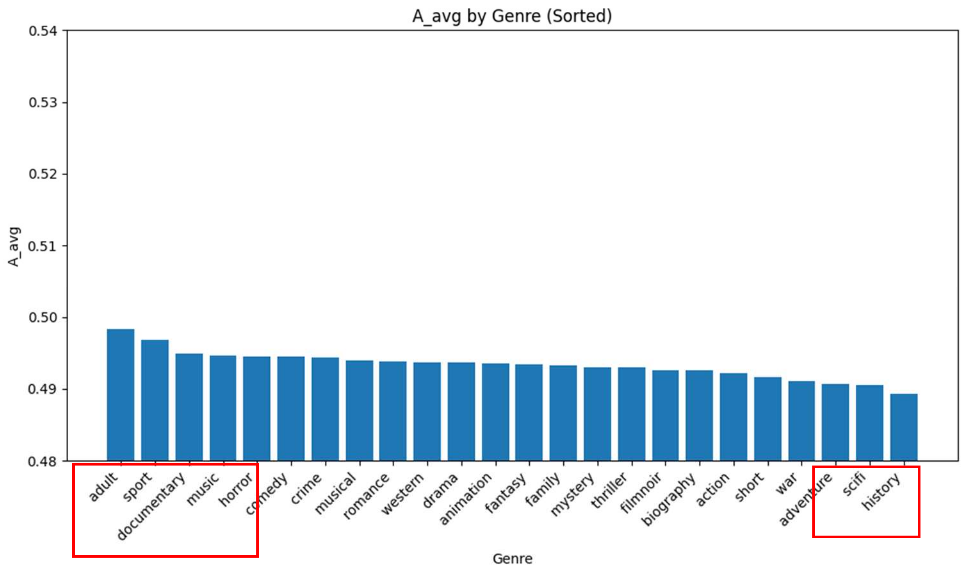
Valence

Music, Adult, Comedy and Romance had higher valence scores, indicating pleasant and uplifting and joyful dialogues, whilst Thriller, Horror and Mystery had the lowest average scores signaling negative emotions like sadness.



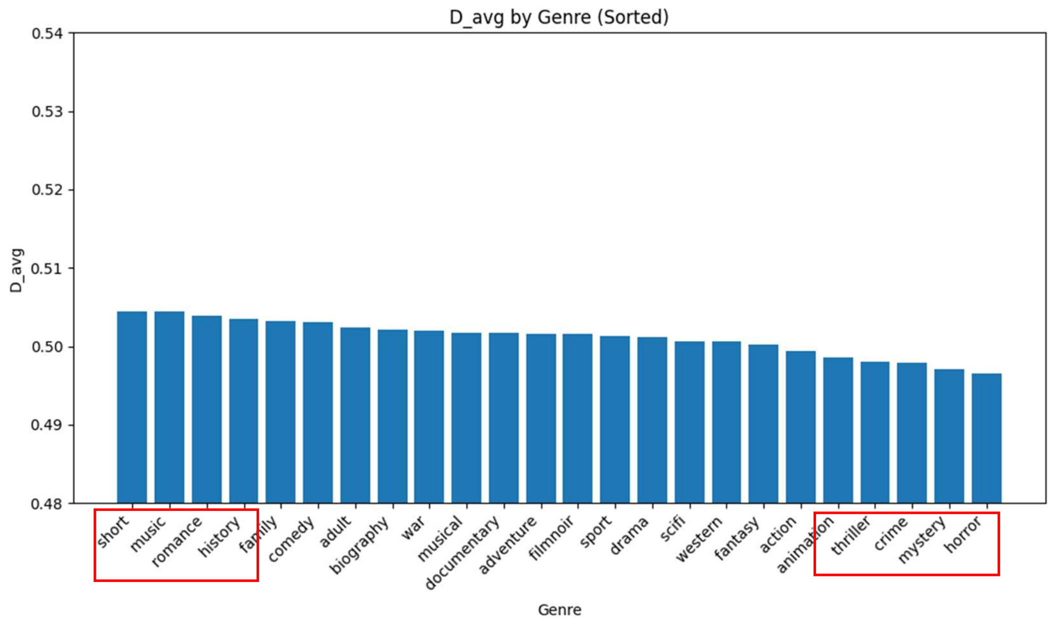
Arousal

Adult, Sport, Documentary and Music were associated with high arousal, aligning with their dynamic narratives. Also History exhibiting the lowest average valance score makes sense showing it's usually calm narrative. The Sci-fi and Adventure genre were the second and thrid lowest ones respectively, which was a bit surprising.



Dominance

Thriller, Crime, Mystery and Horror also exhibited lower average dominance scores, reflecting characters' feelings of vulnerability, fear and lack of control. For the genres Short, Music, Romance and History the opposite was the case reflecting confidence, pride and control.



Character-Level Dynamics

Analyzing main characters' average VAD scores and the standard deviations of those was used to provide insights into their emotional arcs and to choose a movie to do more intricate analsis with. The main character level averages mostly followed what could be seen in the genre level analysis already.

The movie Final Destination 2 has been chosen based on the low average values of Valence and Dominance and also high standard deviation of those two for he main character Kimberly. This was done to achieve a volatile and interpretable time series of her emotions along the movie.

Valence avg - ascending

Valence	Arousal	Dominance	Character	Name	Genre
0,52785	0,51855	0,47603	KIMBERLY	final destination 2	['horror', 'thriller']
0,52844	0,50204	0,49157	MULDER	the x files	['crime', 'horror', 'mystery', 'sci-fi', 'thriller']
0,52874	0,50203	0,47772	MORGAN	what women want	['comedy', 'fantasy', 'romance']
0,52877	0,50581	0,48353	ALEX	final destination	['horror', 'thriller']
0,52905	0,52317	0,49982	NIKOLAS	vampyr	['fantasy', 'horror']

Dominance avg - ascending

Valence	Arousal	Dominance	Character	Name	Genre
0,52785	0,51855	0,47603	KIMBERLY	final destination 2	['horror', 'thriller']
0,52874	0,50203	0,47772	MORGAN	what women want	['comedy', 'fantasy', 'romance']
0,52912	0,51781	0,47871	MR. WHITE	reservoir dogs	['crime', 'mystery', 'thriller']
0,52877	0,50581	0,48353	ALEX	final destination	['horror', 'thriller']
0,53063	0,51191	0,48558	STEVEN	jason goes to hell: the final friday	['horror', 'thriller']

Valence std - descending

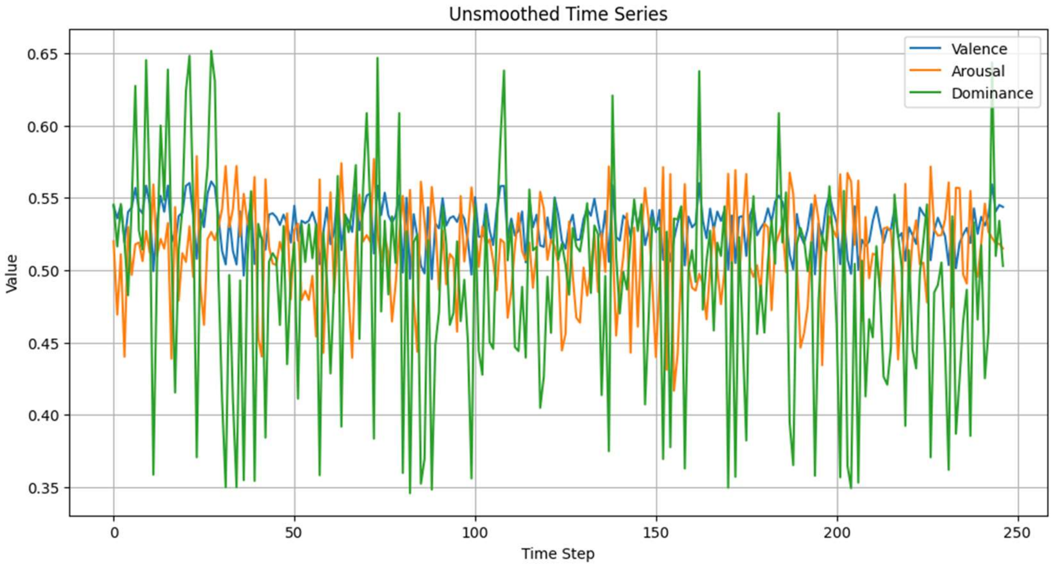
V_std	A_std	D_std	Character	Name	Genre
0,02239	0,02772	0,10242	NIKOLAS	vampyr	['fantasy', 'horror']
0,01918	0,03088	0,09222	LT	bad lieutenant	['crime', 'drama']
0,01771	0,03200	0,08068	KIMBERLY	final destination 2	['horror', 'thriller']
0,01724	0,03272	0,07885	DADE	hackers	['action', 'crime', 'drama', 'thriller']
0,01664	0,03563	0,07664	ROMEO	romeo and juliet	['drama', 'romance']

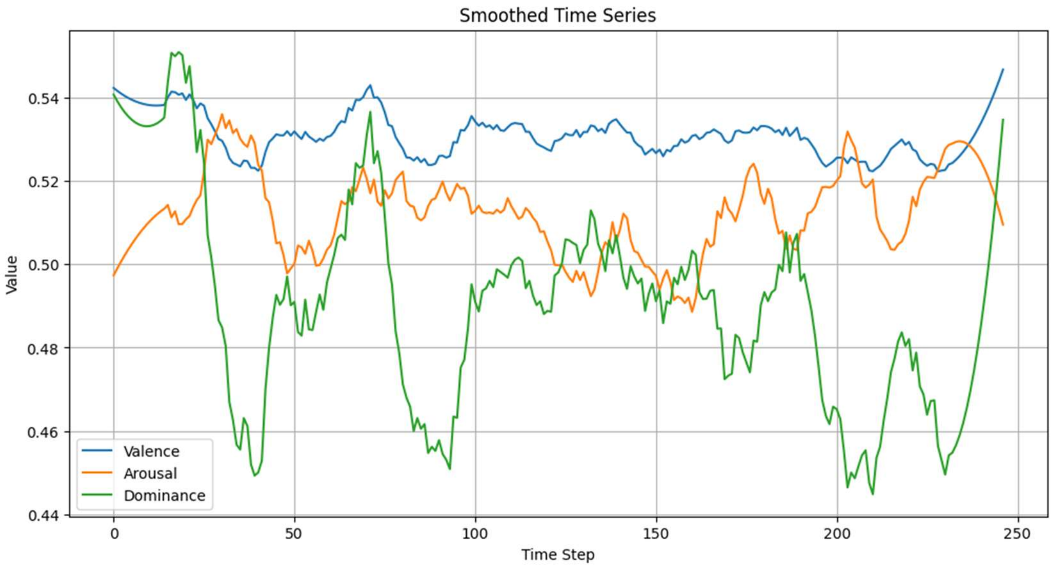
Dominance std – descending

V_std	A_std	D_std	Character	Name	Genre
0,02239	0,02772	0,10242	NIKOLAS	vampyr	['fantasy', 'horror']
0,01918	0,03088	0,09222	LT	bad lieutenant	['crime', 'drama']
0,01771	0,03200	0,08068	KIMBERLY	final destination 2	['horror', 'thriller']
0,01637	0,03181	0,07904	MOM	serial mom	['comedy', 'thriller']
0,01724	0,03272	0,07885	DADE	hackers	['action', 'crime', 'drama', 'thriller']

Final destination 2

The first analysis was on a movie level, predicting the V, A, D values for each individual line in the movie including all characters. To make the noisy looking output more interpretable and let patterns emerge, smoothing with a 30 step window length and a polyorder of 2 was used in the next step.





Movie Level - Analysis

Valence

The relatively consistently declining and slightly fluctuating pattern indicates that the movie maintains a moderate to low and slowly decreasing level of positive emotional tone throughout. The generally low variance suggests that while there might be moments of heightened pleasantness, they are minimal in a film focused on suspense and dread.

Arousal

The orange line shows significant peaks and bottoms, reflecting moments of high tension and calm throughout the film. Peaks in arousal likely correspond to high-stakes action scenes, jump scares, or emotionally intense sequences (e.g., death scenes or moments of realization about the "design" of death which is the key feature in the movie).

Dominance

The green line shows significant fluctuations, with extended periods of low dominance. This pattern aligns with the movie's theme of characters feeling helpless against the inevitability of death. The rare spikes in dominance might represent moments where the characters believe they have gained some control over their fate, although these are fleeting.

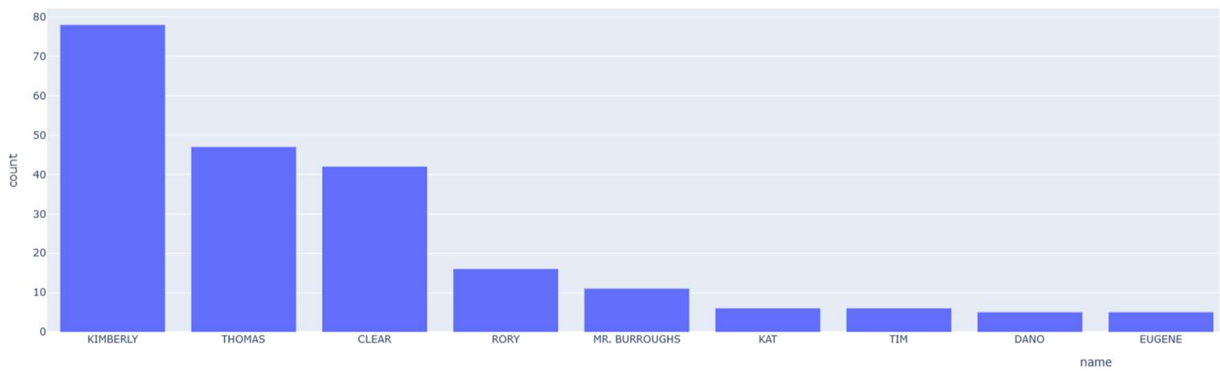
Key Observations

High Arousal, Low Dominance: Frequent moments where arousal is high, but dominance is low, align with scenes of panic, danger, or helplessness (e.g., during the elaborate death sequences).

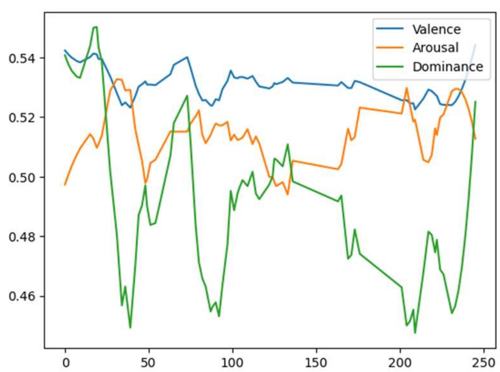
Tension Build-Up: The transitions between arousal and dominance indicate a buildup of tension and moments of perceived control or relief, which are characteristic for movies in the final destination series.

Character Level – Analysis

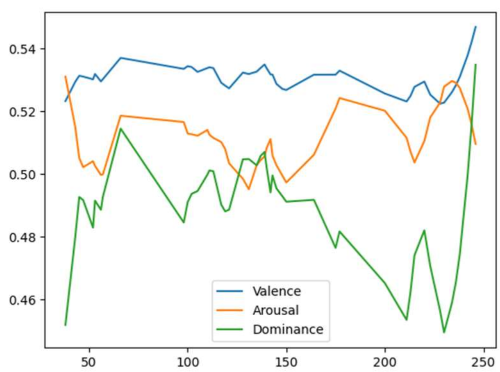
Intuitively the main characters i.e. with the largest amount of lines inhibited the same general pattern as the whole movie, enen though they seem to be more volatile, therefore the analysis and interpretation for the whole movie holds true fort hem as well.



Kimberly



Thomas



Limitations

Challenges

1. Genre Overlap: Many movies belong to multiple genres, complicating precise attributions of VAD patterns
2. Dialogue Context: VAD modeling in this analysis does not account for scene-level context, which could influence the emotional tone
3. The LSTM model semt to plateau very quickly, but giving reasonable results. Other model architectures and hyperparameter tuning could lead to even better ones.

Conclusion

This study highlights the effectiveness of using Valence-Arousal-Dominance (VAD) modeling to analyze emotions in movie scripts. By combining GloVe embeddings, LSTM models, and multi-task learning, we successfully predicted emotional dimensions in dialogues and explored their relationships with genres and characters.

The genre-level analysis revealed clear emotional patterns, such as higher valence in genres like Comedy and Romance, and lower dominance in Horror and Thriller, aligning with their narratives. The character-level analysis showed similar trends, with main characters often reflecting the overall emotional tone of their films. The case study of Final Destination 2 emphasized the movie's relatively low and declining valence and dominance, alongside high arousal, which matched its suspenseful and fear-driven storyline.

While the results were insightful, challenges like overlapping genres and the lack of scene-level context limited the precision of the analysis. The model's performance was reasonable but could benefit from further experimentation with alternative architectures and hyperparameter tuning.

In summary, this project demonstrates the value of VAD modeling for sentiment analysis in movie scripts, offering a nuanced approach to understanding emotional narratives.