Emotion Patterns in Reviews for Top-Rated Movies on Letterboxd

Victor Verma

Boston University

Introduction

Letterboxd is an extremely popular social platform which enables users to log, rate, and review movies. We wanted to learn if there were patterns in the emotions that were expressed in reviews for the top-rated movies on the app. To explore this, we collected nearly 70,000 reviews from 65 of the highest rated movies and built a custom feedforward neural network to classify each review as one of the six Ekman emotions: happiness, sadness, anger, fear, disgust, and surprise. Then we looked for hidden patterns.

Movie Review Data

The movie review data was sourced from Letterboxd's Official Top 250 Narrative Feature Films list. The processed dataset contained 65,967 movie reviews, with each receiving an average of 153 likes and 3 comments. Most of the ratings associated with reviews were high, which was expected given that the reviews were taken from the top-rated movies on Letterboxd. This led to the expectation that positive emotions would be expressed most often.

Emotion Data

The initial training data for the emotion classification model was sourced from the Twitter Emotion Corpus (TEC). The processed TEC dataset consisted of 19,333 text-label pairs. In practice, it was found that the small training dataset and the class imbalance increased the variance of the prediction accuracy across emotion classes and resulted in a low macro F1-score. To improve classification performance, supplementary labeled data was found on GitHub. The additional data was merged with the TEC data to create a combined dataset containing 36,464 text-label pairs.

Label	Count	Percentage
Happiness	14,264	39.12%
Sadness	8,918	24.46%
Anger	4,025	11.04%
Fear	4,798	13.16%
Disgust	733	2.01%
Surprise	3,726	10.22%

Table 1: Combined Emotion Class Distribution

Text Embedding Model

The sentence-transformers/all-MiniLM-L6-v2 embedding model from Hugging Face encodes text as a 384-dimensional vector that captures semantic information. It is intended for sentences and short paragraphs, which made it ideal for both the training data and its application in classifying movie reviews.

Neural Model

The neural model was trained on the combined emotion dataset using cross-entropy loss and further optimized by using SGD with momentum. During development, grid searches were conducted over the combinations of training datasets, embedding models, and training parameters. The best model was trained for 100 epochs using a learning rate of 0.01, momentum of 0.9, and batch size of 64.

EEC Model Architecture

The Ekman Emotion Classifier (EEC) was created by combining the embedding and neural models. The input layer takes in a 384-dimensional vector embedding, which is passed through a linear layer with 256 hidden units. The ReLU activation function is applied to the output of the first linear layer, and the result is passed through a second linear layer which maps the 256-dimensional hidden representation to a 6-dimensional output.
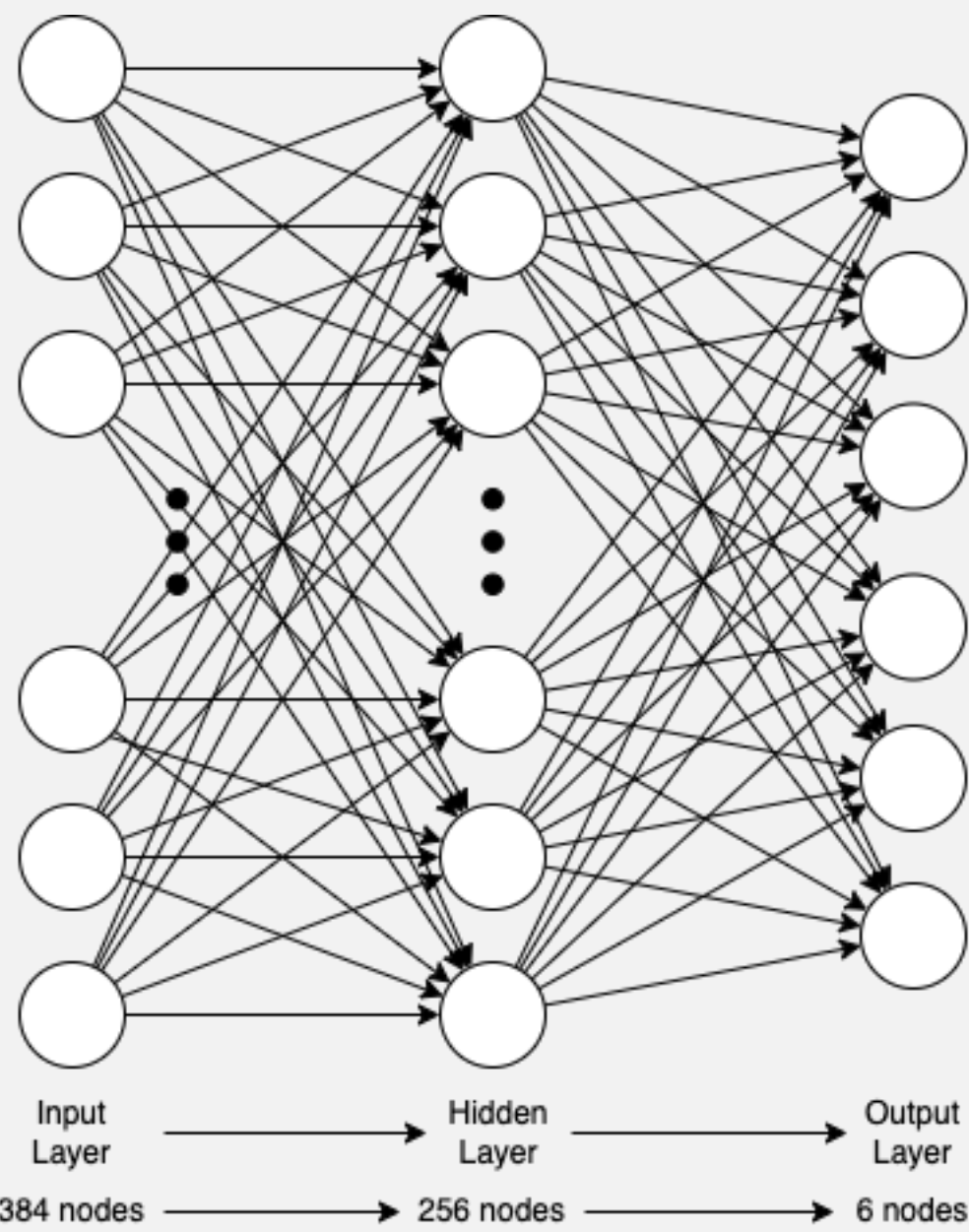

Figure 1: EEC Model Architecture

Model Performance

The EEC model achieved a macro F1-score of 0.57 on the validation dataset. It was able to adequately identify happiness, sadness, and fear, but struggled with the other emotions. Disgust was regularly misclassified, and there remained confusion between surprise and happiness. This variance was expected due to the imbalanced training dataset.

Emotion	Precision	Recall	F1-Score
Happiness	0.75	0.76	0.76
Sadness	0.58	0.77	0.66
Anger	0.61	0.46	0.53
Fear	0.66	0.62	0.64
Disgust	0.49	0.32	0.39
Surprise	0.57	0.36	0.44

Table 2: EEC Model Performance Statistics

Classifying Reviews

The EEC model was used to classify the emotions expressed in the movie review dataset, a collection of 65,967 reviews gathered from top-rated movies on Letterboxd. As expected, happiness was the most common emotion, expressed in nearly 37% of all reviews. The least frequent emotions were disgust and anger, combining for less than 8%.

Emotion	Count	Percentage
Happiness	24,404	36.99%
Sadness	17,924	27.17%
Anger	4,261	6.46%
Fear	12,754	19.33%
Disgust	633	0.96%
Surprise	5,991	9.08%

Table 3: Distribution of Emotions in Movie Reviews

Emotion Patterns

Beyond the raw counts, the emotions expressed in reviews were stratified across different categories to identify less obvious patterns. First, the distribution of emotions expressed in reviews was found to be approximately the same regardless of the decade in which the movie was released.
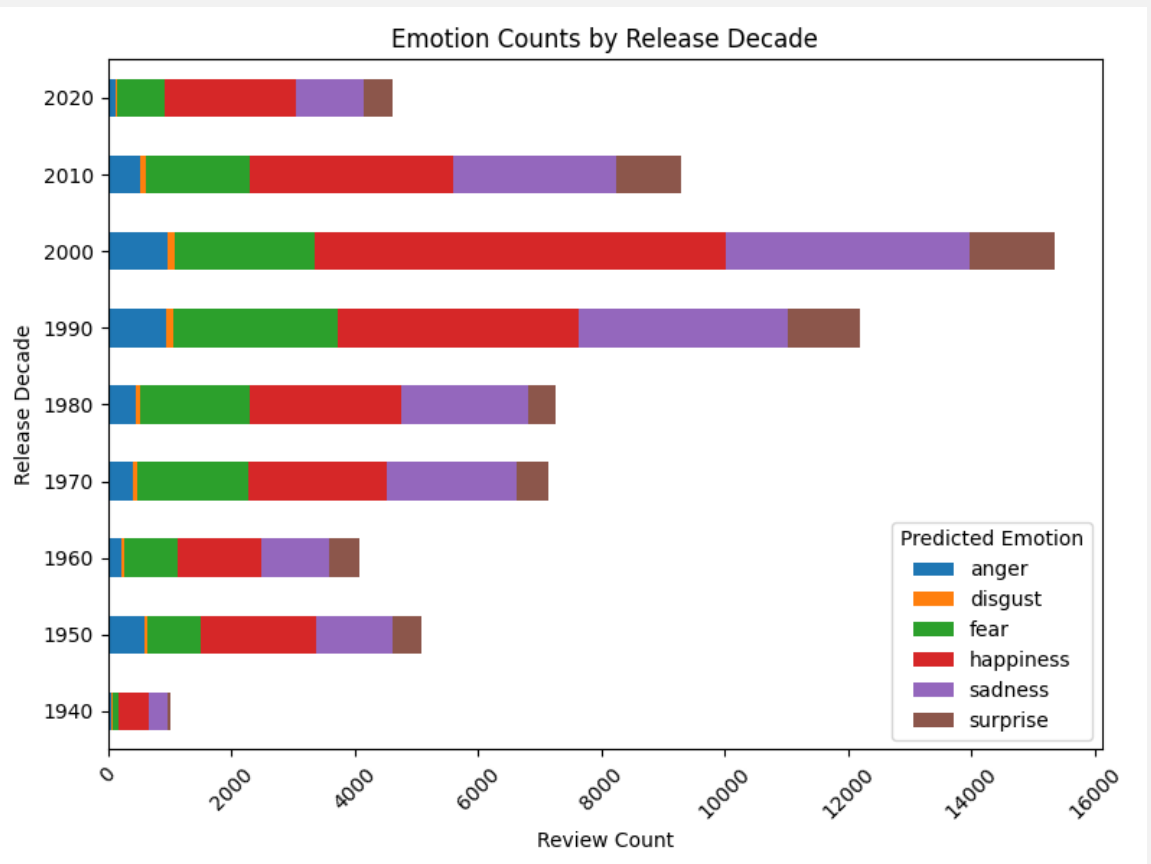

Figure 2: Emotion Counts by Release Decade

Next, it was observed that the distribution of emotions expressed in reviews remained nearly identical to the overall distribution regardless of the movie's country of origin.
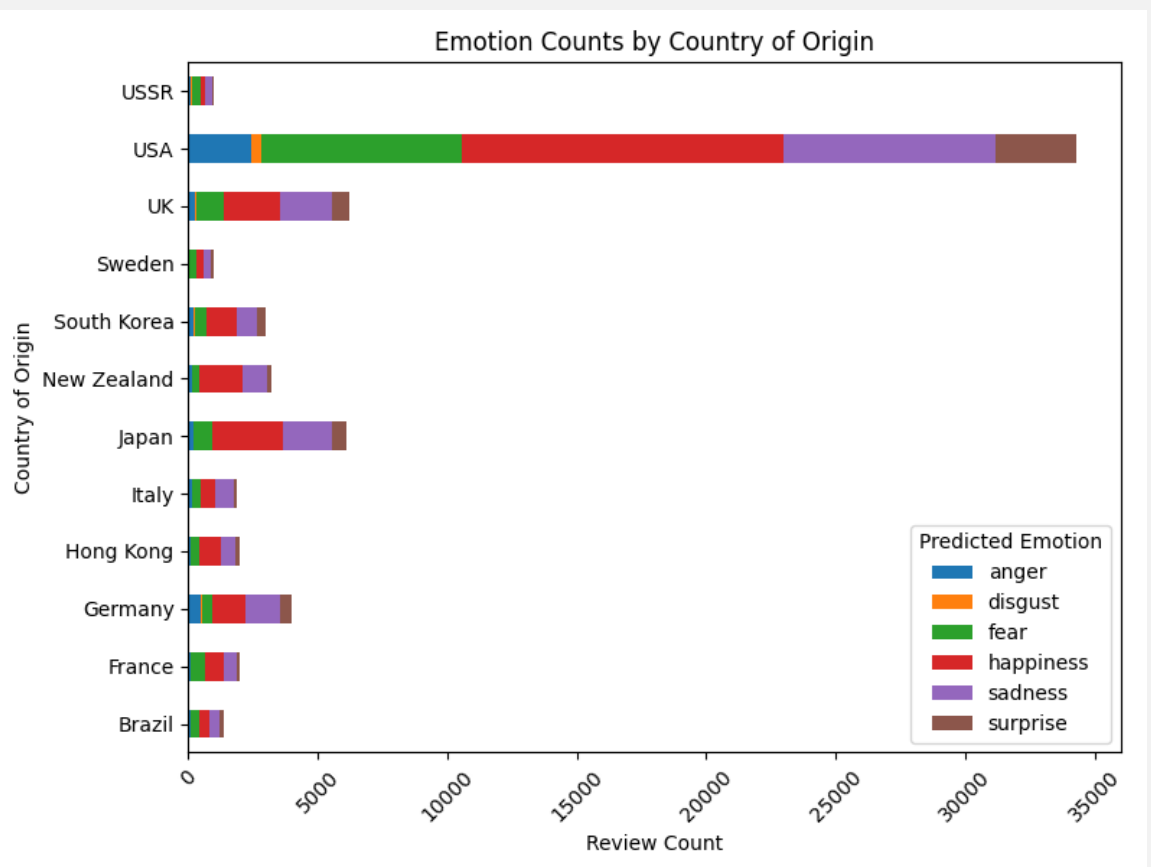

Figure 3: Emotion Counts by Country of Origin

More Emotion Patterns

The most prevalent emotion in horror movie reviews was fear, and the most common emotion in history movie reviews was sadness. Meanwhile, a relatively high percentage of reviews for both thriller and mystery movies expressed surprise.
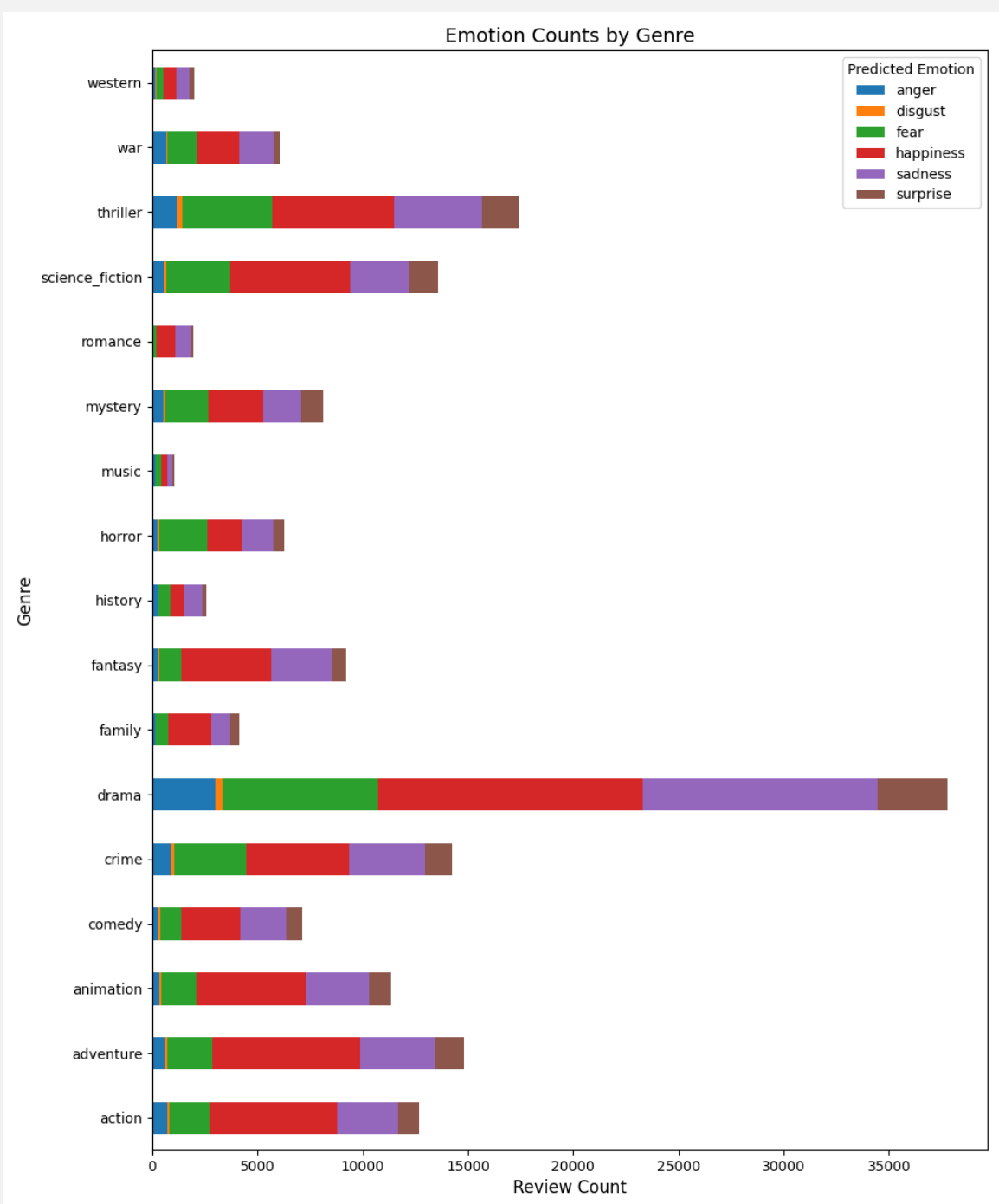

Figure 4: Emotion Counts by Genre

On Letterboxd, users are also able to like reviews. Surprisingly, the reviews expressing the least frequent emotions received the most likes on average, and vice versa.
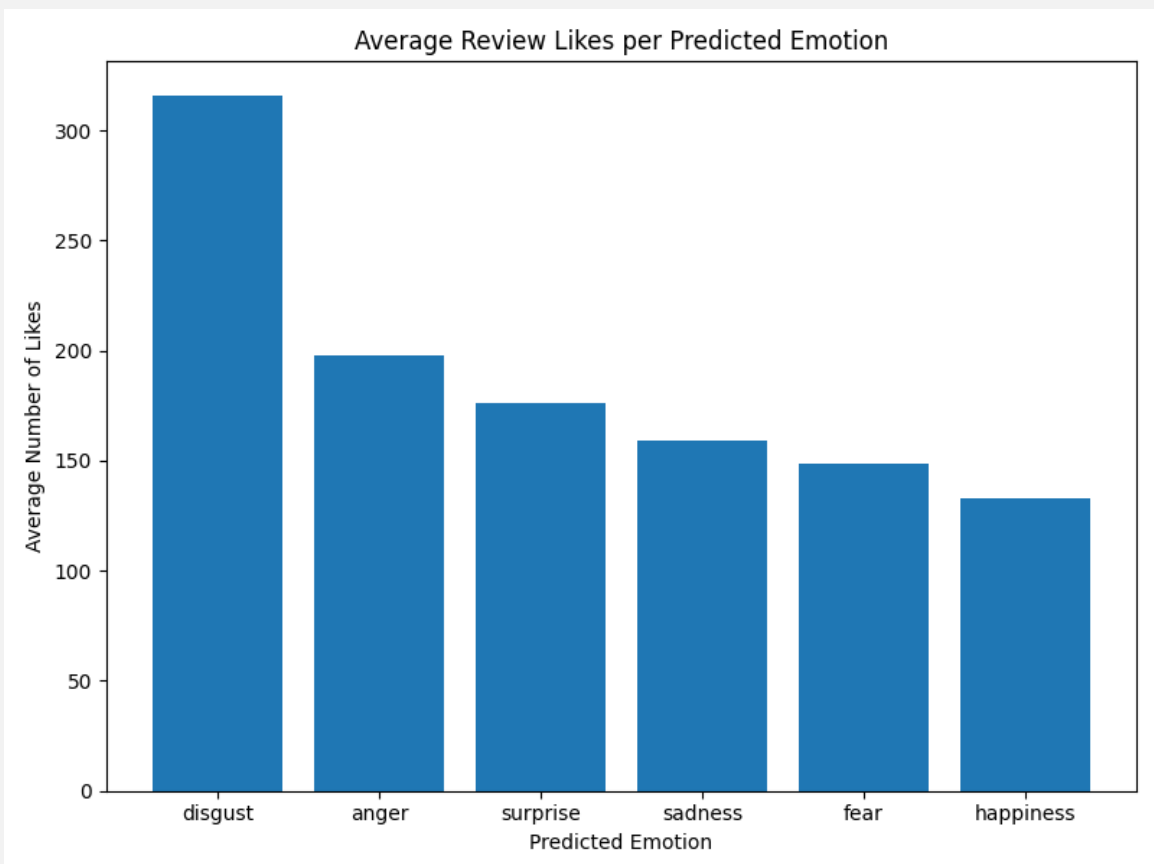

Figure 5.5: Average Review Likes per Emotion

Conclusion

Our analysis revealed that happiness and sadness were the most frequently expressed emotions in reviews, but rarer emotions like disgust and anger tended to receive more likes on average. Surprisingly, the distribution of emotions expressed in reviews remained similar for movies released in different decades or originating from different countries. As expected, that the distribution of emotions varied across genres. Future work could verify these results using a better emotion classification model and further generalize the patterns by analyzing a larger dataset of movie reviews.