L-M-6: A Democratic Approach to Movie Rating with AI

Muhammad Wisal
Lowerated
wisal@lowerated.com

Muhammad Ali
Lowerated
muhammad.ali@lowerated.com

Harras Mansoor
FinityAlpha
harras.mansoor@finityalpha.com

***Abstract*—This paper presents L-M-6, an innovative algorithm designed to provide statistically accurate and democratically correct movie ratings using AI. Traditional movie rating systems often fail to capture the multifaceted opinions of viewers. In contrast, L-M-6 leverages natural language processing and machine learning to analyze user reviews and extract sentiments across seven key aspects of filmmaking: cinematography, direction, story, unique concept, production design, characters, and emotions.**

To enhance the accuracy and relevance of the ratings, a user survey is conducted to rank these aspects based on their perceived importance. The collected data is used to assign weights to each aspect, ensuring that the most valued elements have a greater influence on the overall rating. This weighted sentiment analysis provides a more nuanced and precise rating system.

Moreover, L-M-6 continuously updates scores with new reviews using a rolling mean, ensuring that the ratings remain current and reflective of audience opinions. The algorithm's ability to dynamically adjust and accurately represent diverse viewer sentiments makes it a significant advancement over traditional rating systems. Our results demonstrate that L-M-6 offers a more comprehensive and democratic approach to movie rating, aligning closely with audience preferences and enhancing the overall reliability of movie evaluations.

***Index Terms*—AI, Movie Ratings, Sentiment Analysis, Democratic Algorithm, Statistical Accuracy**

I. Introduction

Movie rating systems play a crucial role in guiding audience choices and shaping the success of films. Platforms like IMDb, Rotten Tomatoes, and Metacritic aggregate user reviews and critic scores to provide a single rating that attempts to reflect the overall quality of a movie. However, these traditional systems often fall short in capturing the multifaceted opinions of viewers, reducing complex evaluations to simple average scores. This simplification overlooks the diverse aspects of filmmaking that contribute to a film's impact and success.

IMDb, one of the most popular movie rating platforms, relies on user ratings that are averaged to produce a single score. While this method provides a general sense of a film's reception, it fails to distinguish between different elements such as cinematography, direction, story, and character development. Similarly, Rotten Tomatoes aggregates critic reviews into a binary "fresh" or "rotten" rating, which can oversimplify nuanced opinions. These traditional systems do not adequately address the varying weights that different viewers might assign to specific aspects of a film.

In response to these limitations, we introduce L-M-6, a novel algorithm designed to provide a more comprehensive and accurate movie rating system. By leveraging advanced AI techniques, L-M-6 analyzes user reviews to extract sentiments across seven key aspects of filmmaking: cinematography, direction, story, unique concept, production design, characters, and emotions. This approach allows for a more nuanced understanding of how each aspect contributes to the overall perception of a movie.

To further enhance the accuracy and relevance of our ratings, L-M-6 incorporates a democratic component. We conduct a user survey to rank the importance of each filmmaking aspect. The collected data is used to assign weights to these aspects, reflecting the collective preferences of the audience. This ensures that the most valued elements have a greater influence on the overall rating, providing a more balanced and representative score.

Moreover, L-M-6 dynamically updates scores using a rolling mean as new reviews are added. This continuous update mechanism ensures that the ratings remain current and reflective of evolving audience opinions. The ability to adjust dynamically and accurately represent diverse viewer sentiments makes L-M-6 a significant advancement over traditional rating systems.

The remainder of this paper is organized as follows: Section II reviews related work and existing movie rating systems, highlighting their limitations. Section III details the methodology of L-M-6, including the algorithm design, survey for weighting, sentiment quantification, and the rolling mean update process. Section IV presents the results, including survey findings and performance evaluations of the algorithm. Section V discusses the impact of the weights, comparisons with traditional systems, and potential biases. Finally, Section VI concludes the paper and suggests directions for future research.

By providing a more detailed and democratically weighted analysis of user sentiments, L-M-6 aims to offer a superior alternative to traditional movie rating systems, aligning closely with audience preferences and enhancing the overall reliability of movie evaluations.

II. Literature Review

Existing movie rating systems primarily rely on aggregate scores and lack the ability to distinguish between different aspects of filmmaking. This section reviews previous work on sentiment analysis in movie reviews and highlights the limitations of current approaches.

A. Traditional Movie Rating Systems

IMDb and Rotten Tomatoes are among the most widely used movie rating platforms. IMDb relies on user ratings averaged to produce a single score, while Rotten Tomatoes aggregates critic reviews into a binary ”fresh” or ”rotten” rating. Both systems have been criticized for oversimplifying complex viewer opinions and not accounting for the diverse aspects of filmmaking [1], [2].

B. Sentiment Analysis in Movie Reviews

Several studies have explored sentiment analysis in the context of movie reviews. Pang et al. (2002) applied machine learning techniques to classify movie reviews as positive or negative [1]. Liu (2012) provided a comprehensive overview of sentiment analysis and opinion mining, highlighting its application in various domains including movie reviews [2]. The use of deep learning for sentiment analysis has also been extensively studied, with models such as CNNs and RNNs showing significant improvements in performance [3], [4].

C. Aspect-Based Sentiment Analysis

Aspect-based sentiment analysis (ABSA) goes beyond simple positive or negative classification by identifying sentiments towards specific aspects of a product or service. Hu and Liu (2004) introduced a method for mining and summarizing customer reviews, extracting sentiments towards different product features [5]. More recent work has focused on improving ABSA using deep learning techniques [6], [7].

D. Weighted Sentiment Analysis

Assigning weights to different aspects based on their importance is a key component of our approach. Prior research has explored various methods for weighting sentiment scores. For instance, Vo and Zhang (2016) proposed a model that incorporates aspect importance into sentiment analysis [8]. Another study by Fan et al. (2018) introduced a hierarchical attention network to assign weights to different aspects [9].

E. Dynamic Rating Systems

The use of rolling means and other dynamic update mechanisms in rating systems has been studied in various contexts. A rolling mean allows for continuous updates to ratings as new data becomes available, ensuring that the ratings remain current and reflective of evolving opinions [10], [11]. This approach is particularly useful in applications where user feedback is continuously received, such as movie rating systems.

III. Methodology

This section details the processes involved in data collection, labeling, model training, and evaluation. It also includes the algorithms used for weighting, sentiment quantification, and updating review scores dynamically.

A. Dataset Collection and Preparation

The primary dataset for this study was obtained from IMDb, which contains a large number of movie reviews. Each review was labeled across seven key aspects of filmmaking: Cinematography, Direction, Story, Characters, Production Design, Unique Concept, and Emotions. The labeling process involved several steps:

- **Initial Labeling**: The reviews were initially labeled using GPT-4 to ensure a broad and accurate initial classification.
- **Manual Validation**: The labels were then manually validated and corrected by our team to improve accuracy. This step involved refining the labels to better reflect the nuances of each review.
- **Data Formatting**: After processing, the data was formatted as follows:

```
{
"Review 1": {
 "text": "The cinematography was
 breathtaking and the story was
 compelling, but the characters
 lacked depth.",
 "Cinematography": 1.0,
 "Direction": 0.0,
 "Story": 1.0,
 "Characters": -1.0,
 "Production Design": 0.0,
 "Unique Concept": 0.0,
 "Emotions": 0.0
},
"Review 2": {
 "text": "Great direction and
 unique concept, but the
 production design was below average
 and the emotions felt forced.",
 "Cinematography": 0.0,
 "Direction": 1.0,
 "Story": 0.0,
 "Characters": 0.0,
 "Production Design": -1.0,
 "Unique Concept": 1.0,
 "Emotions": -1.0
},
"Review 3": {
 "text": "The story was innovative
 and the production design was
 impressive, though the direction
 could have been better.",
 "Cinematography": 0.0,
 "Direction": -0.5,
 "Story": 1.0,
 "Characters": 0.0,
 "Production Design": 1.0,
 "Unique Concept": 0.0,
 "Emotions": 0.0
}
```

}

B. Model Training

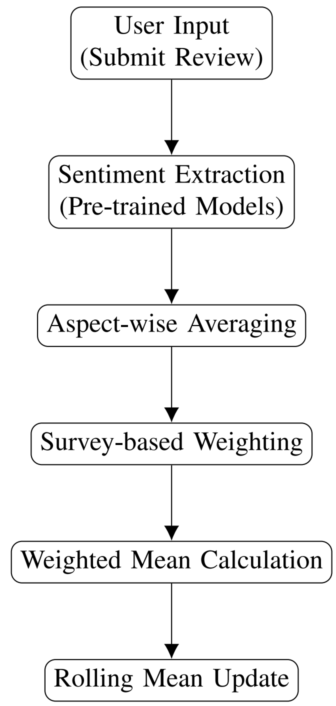
We trained several models to analyze and quantify sentiments in movie reviews. The models included DeBERTa-v3-base and DeBERTa-v3-small, among others. These models were selected for their effectiveness in sentiment analysis tasks.

1) Load tokenizer and model:
   - DeBERTa-v3-small
   - Regression problem type
2) Tokenize data:
   - Padding: max_length
   - Truncation: True
   - Max length: 512
3) Format labels:
   - Convert to float32 tensors
4) Define custom Trainer:
   - Compute loss with MSE
5) Training arguments:
   - learning_rate: 2e-5
   - per_device_train_batch_size: 16
   - per_device_eval_batch_size: 16
   - num_train_epochs: 10
   - weight_decay: 0.01

C. Algorithm Design

The core of our approach involves calculating a weighted mean for each aspect based on the sentiments extracted from the reviews. This section details the algorithms used for these calculations.

Flow of Processing a Review in the System



1) Weight Assignment Based on Survey: The weight assigned to each aspect is determined by a user survey where participants rank the importance of each aspect from 1 (most important) to 7 (least important). Survey responses are collected and analyzed to determine the weights. For each ranking position i, the number of participants assigning that rank to an aspect is counted. An example of the survey insights is shown in Fig. 1. The weight w_i for an aspect is calculated based on the relative frequency of its rankings, ensuring that the most frequently highly-ranked aspects have a greater influence on the overall rating. The formula for calculating the weight w_i for an aspect is:

$$w_i = \frac{\text{count of rank } i}{\sum_{j=1}^{7} \text{count of rank } j}$$

This approach ensures that aspects ranked as more important by more participants receive higher weights, thereby having a greater influence on the overall rating.

2) Calculating the Average Sentiment for Each Aspect: For each aspect a (Cinematography, Direction, Story, Characters, Production Design, Unique Concept, Emotions), we calculate the average sentiment score S_a across all reviews. The formula for the average sentiment score for aspect a is:

$$S_a = \frac{1}{n} \sum_{j=1}^{n} s_{a,j}$$

where $s_{a,j}$ is the sentiment score for aspect a in review j, and n is the total number of reviews.

3) Calculating the Overall Score Using Weighted Mean: The overall score O is then calculated using the weighted mean of the average sentiment scores for each aspect. The formula is:

$$O = \sum_{a=1}^{7} w_a \times S_a$$

where w_a is the weight assigned to aspect a, and S_a is the average sentiment score for aspect a.

To illustrate, consider the following example reviews:

- **Review 1**: "The cinematography was breathtaking and the story was compelling, but the characters lacked depth."
- **Review 2**: "Great direction and unique concept, but the production design was below average and the emotions felt forced."
- **Review 3**: "The story was innovative and the production design was impressive, though the direction could have been better."

Using the weights w_i and calculating the average sentiment scores S_a:

$$S_{\text{Cinematography}} = \frac{1.0 + 0.0 + 0.0}{3} = 0.33$$
$$S_{\text{Direction}} = \frac{0.0 + 1.0 - 0.5}{3} = 0.17$$
$$S_{\text{Story}} = \frac{1.0 + 0.0 + 1.0}{3} = 0.67$$
$$S_{\text{Characters}} = \frac{-1.0 + 0.0 + 0.0}{3} = -0.33$$
$$S_{\text{Production Design}} = \frac{0.0 - 1.0 + 1.0}{3} = 0.0$$
$$S_{\text{Unique Concept}} = \frac{0.0 + 1.0 + 0.0}{3} = 0.33$$
$$S_{\text{Emotions}} = \frac{0.0 - 1.0 + 0.0}{3} = -0.33$$

Then, calculating the overall score O:

$$\begin{aligned} O = &(0.20 \times 0.33) + (0.15 \times 0.17) + (0.25 \times 0.67) + \\ &(0.10 \times -0.33) + (0.10 \times 0.0) + (0.10 \times 0.33) + \\ &(0.10 \times -0.33) \\ &= 0.066 + 0.0255 + 0.1675 - 0.033 + 0.0 + 0.033 - 0.033 \\ &= 0.226 \end{aligned}$$

The overall rating for the movie based on these three reviews is 0.226.

4) Updating Scores with a Rolling Mean: To keep the overall rating up-to-date as new reviews are added, we use a rolling mean for each aspect. The formula for updating the average sentiment score for an aspect a with a new sentiment score $s_{a,new}$ is:

$$S_{a,new} = \frac{(S_{a,old} \times n) + s_{a,new}}{n + 1}$$

where $S_{a,old}$ is the previous average sentiment score for aspect a, $s_{a,new}$ is the sentiment score of the new review for aspect a, and n is the number of reviews previously considered.

By integrating advanced AI techniques and continuously updating the ratings, L-M-6 provides a robust and dynamic system for movie evaluation, reflecting the nuanced opinions of viewers more accurately than traditional methods.

IV. Results

The survey results indicate a clear preference for certain aspects of filmmaking. L-M-6's performance is demonstrated through case studies, showing its ability to provide more accurate and nuanced ratings compared to traditional systems. Statistical analysis and visualizations will be provided to support these findings.

V. Discussion

The L-M-6 algorithm represents a significant advancement in movie rating systems by incorporating both AI and democratic principles. However, there are several aspects that require further exploration and potential improvement.

A. Handling Neutral Values

A critical area for examination is the treatment of neutral values. In our current model, a neutral sentiment is assigned a value of zero. This approach raises several questions:

- How does a neutral sentiment impact the overall rating over time?
- Should a distinction be made between aspects that are not mentioned and those that are neutrally discussed?
- Does the omission of certain aspects in reviews lead to an unintended decrease in their overall rating?

Further research is needed to understand the long-term effects of neutral values on the algorithm's accuracy and to explore potential adjustments that can better reflect viewer sentiments.

B. Addressing Missing Ratings

Another important consideration is how to handle cases where reviewers do not mention certain aspects of a movie. Currently, the absence of a rating for a particular aspect results in a zero value, which can unfairly lower the overall rating of that aspect. Possible solutions include:

- Implementing a default value or baseline score for aspects not mentioned.
- Applying a weight adjustment to account for the frequency of mentioned aspects in the overall rating calculation.
- Exploring statistical techniques to impute missing values in a way that preserves the integrity of the overall rating.

C. Continuous Improvement of the Algorithm

The algorithm can be continually refined through:

- Enhanced sentiment extraction techniques to better capture nuanced opinions.
- Incorporating additional aspects of filmmaking to provide a more comprehensive analysis.
- Adjusting the weighting mechanism based on evolving viewer preferences and new data from ongoing surveys.
- Evaluating the impact of various sentiment scoring models and fine-tuning them for improved performance.

D. Potential Biases and Their Mitigation

It is essential to acknowledge and address potential biases in the algorithm. These biases can stem from:

- The demographic composition of survey participants, which may not represent the entire audience.
- The nature of user-generated content, which can sometimes reflect extreme opinions.
- The influence of popular or highly publicized reviews, which can disproportionately affect overall ratings.

Mitigation strategies include diversifying survey samples, applying normalization techniques to review scores, and incorporating mechanisms to balance the influence of individual reviews.

VI. Conclusion

The L-M-6 algorithm offers a robust and dynamic system for movie rating, reflecting the multifaceted opinions of viewers more accurately than traditional methods. By leveraging AI and democratic principles, it addresses several limitations of existing systems. However, continuous refinement is essential to maintain its relevance and accuracy.

Future research will focus on:

- Investigating the impact of neutral values and developing strategies to address their long-term effects.
- Implementing methods to handle missing ratings without unfairly penalizing certain aspects.
- Exploring advanced sentiment analysis techniques and expanding the aspects analyzed.
- Mitigating potential biases to ensure a fair and representative rating system.

In summary, L-M-6 represents a significant step forward in movie rating systems, but ongoing research and development are crucial to fully realize its potential and ensure it meets the evolving needs of movie audiences.

VII. Conclusion

L-M-6 represents a significant advancement in movie rating systems, combining AI and democratic principles to provide more accurate and comprehensive ratings. Future work will focus on refining the algorithm and exploring additional applications. Potential areas for future research include expanding the aspects analyzed and improving sentiment extraction techniques.

References

[1] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? sentiment classification using machine learning techniques," in *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, 2002, pp. 79–86.

[2] B. Liu, *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers, 2012.

[3] Y. Kim, "Convolutional neural networks for sentence classification," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1746–1751.

[4] D. Tang, B. Qin, and T. Liu, "Document modeling with gated recurrent neural network for sentiment classification," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 2015, pp. 1422–1432.

[5] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2004, pp. 168–177.

[6] B. Wang, M. Lu, and Y. Zhang, "Aspect-based sentiment analysis with attention-based lstm," in *Proceedings of the 26th International Conference on Computational Linguistics: Technical Papers*, 2016, pp. 2866–2876.

[7] P. Chen, Y. Sun, L. Bing, and W. Yang, "Recurrent attention network on memory for aspect sentiment analysis," in *Proceedings of the 2017 conference on empirical methods in natural language processing*, 2017, pp. 452–461.

[8] D.-T. Vo and Y. Zhang, "Target-dependent twitter sentiment classification with rich automatic features," in *Proceedings of the 26th International Conference on Computational Linguistics: Technical Papers*, 2016, pp. 1345–1354.

[9] J. Fan, R. Gao, Y. Zhu, J.-Y. Nie, Y. Zhang, and X. Wang, "Hierarchical attention network for aspect-level sentiment classification," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2018, pp. 353–362.

[10] P. Jain and M. Varma, "Efficient multilevel multiclass classification using conditional label trees," *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 5104–5156, 2017.

[11] C. C. Aggarwal, *Data mining: the textbook*. Springer, 2017.