# Are Words in Movies Biased? Exploring the Boundary Between Art and Commercial Films with Word2Vec and WEAT

1st MYUNG JUN SUNG
*dept. NLP*
*AIFFEL*
Seoul, Korea
smjhym@gmail.com

***Abstract*—This study investigates the linguistic bias between art and commercial films using Word2Vec models and the Word Embedding Association Test (WEAT). By analyzing a dataset of Korean film synopses, we examine how different word embedding methods (CBOW and Skip-gram) and vocabulary selection techniques (TF-IDF and LDA) reveal biases in word associations across various film genres.**

**The findings suggest that the Skip-gram model, which captures infrequent word relationships, shows a stronger tendency to highlight nuanced differences between art and commercial films compared to CBOW. Among vocabulary generation methods, TF-IDF demonstrates clearer genre-specific word associations than LDA or their combinations. Notably, the Skip-gram and TF-IDF combination reveals pronounced biases, with genres such as drama, performance, and adult films aligning more with art films, while genres like action, comedy, and fantasy show stronger associations with commercial films.**

**These results illustrate how linguistic patterns in film genres can reflect varying cultural and commercial inclinations. The study underscores the importance of methodological choices in identifying and interpreting semantic biases within textual data, offering insights into the interplay between language use and cultural framing in the film industry.**

***Index Terms*—Word2Vec, WEAT, TF-IDF, LDA, Film Genres, Textual Analysis**

## I. Introduction

Language encapsulates cultural and commercial dynamics, especially in creative industries like film. Art films and commercial films are often characterized by distinct linguistic patterns that reflect their thematic and cultural orientations. This study investigates these linguistic distinctions using Word2Vec embeddings and the Word Embedding Association Test (WEAT) to measure semantic biases across various film genres.

Using a dataset of Korean film synopses, we examine how different Word2Vec training methods—Continuous Bag of Words (CBOW) and Skip-gram—combined with vocabulary generation techniques such as TF-IDF and LDA, influence the identification of semantic biases. The analysis highlights how these methodological choices shape the representation of art and commercial films, revealing genre-specific word associations and biases.

By quantifying these linguistic patterns, this research contributes to understanding the interplay between cultural framing and language in the film industry. The findings provide a foundation for exploring how linguistic biases reflect broader societal and industry dynamics.

## II. Related Work

### A. Word Embeddings and WEAT

Word embeddings, such as Word2Vec, have been widely used to capture semantic relationships in textual data. Mikolov et al. introduced Word2Vec with two key architectures: Continuous Bag of Words (CBOW) and Skip-gram, both of which excel in learning contextual relationships between words[1]. WEAT (Word Embedding Association Test), developed by Caliskan et al., extended the application of word embeddings by quantifying biases between target and attribute word sets[2]. While WEAT has been used in various fields, including psychology and social sciences, its application to creative industries, such as film, remains underexplored.

### B. Film Data Analysis

Prior studies on film data have focused on audience preferences, box office success, and narrative structures. For example, Bolelli et al. used text mining to analyze movie reviews for sentiment prediction[3], while Mei et al. explored linguistic patterns in screenplay texts to predict movie success[4]. However, these studies primarily address commercial metrics and lack a focus on semantic biases in linguistic representations of film genres.

[1] Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *arXiv preprint arXiv:1301.3781.*

[2] Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science, 356*(6334), 183-186.

[3] Bolelli, L., Ertekin, Ş., & Giles, C. L. (2019). Text mining for movie review sentiment analysis. *Journal of Information Science, 45*(1), 28-44.

[4] Mei, Q., Ling, X., Wondra, M., Su, H., & Zhai, C. (2007). Topic sentiment mixture: Modeling facets and opinions in weblogs. *Proceedings of the 16th international conference on World Wide Web*, 171-180.

## III. Methods

### A. Data Description

The primary dataset, `synopsis.txt`, contains synopses of films produced between 2001 and August 2019. This dataset, approximately 17MB in size, provides textual descriptions that reflect the thematic and cultural orientations of various films. To analyze linguistic biases, we classified the dataset into two categories: art films and commercial films, and further divided it by film genres. The structure is as follows:

- **Film Classification**:
  - `synopsis_art.txt`: Art films
  - `synopsis_gen.txt`: Commercial films
- **Genre Classification**: Genres include *SF*, *Family*, *Performance*, *Horror*, *Documentary*, *Drama*, *Melodrama*, *Musical*, *Mystery*, *Crime*, *Historical*, *Western*, *Adult*, *Thriller*, *Animation*, *Action*, *Adventure*, *War*, *Comedy*, and *Fantasy*. Each genre is stored in separate files, such as `synopsis_SF.txt` for SF films.

These categorizations provide the foundation for generating target and attribute word sets necessary for the WEAT analysis.

### B. Word2Vec Training

To represent the semantic relationships between words in the synopses, Word2Vec models were trained using two architectures: Continuous Bag of Words (CBOW) and Skip-gram.

- **CBOW:** Predicts the target word based on its surrounding context. This approach captures broader contextual meanings and works effectively with frequent words[5].
- **Skip-gram:** Predicts surrounding words given a target word, making it more sensitive to infrequent words. This characteristic allows Skip-gram to capture nuanced relationships, which is critical for distinguishing subtle biases in the data[6].

Both models were trained on the full synopsis dataset using a window size of 5 and a vector dimension of 300.

### C. Target and Attribute Word Set Generation

The target and attribute word sets were generated based on the film classifications and genres, leveraging two different methods: TF-IDF and LDA.

- **TF-IDF (Term Frequency-Inverse Document Frequency):** TF-IDF was applied to identify words that are most representative of each category (art films, commercial films, and genres). Words with high TF-IDF scores were selected as they reflect the uniqueness of a category[7].
- **LDA (Latent Dirichlet Allocation):** LDA was used to extract topic-specific words for each genre and classification. For each topic, the top words with the highest probability scores were included in the attribute and target word sets[8].

This dual approach ensured that both genre-specific and general thematic patterns were captured.

### D. WEAT Score Calculation

The Word Embedding Association Test (WEAT) quantifies the association between target word sets ($T_1, T_2$) and attribute word sets ($A_1, A_2$). The WEAT score is calculated as follows:

1) **Association of a word with an attribute set:**

$$s(w, A_1, A_2) = \text{mean}_{a \in A_1} \cos(w, a) - \text{mean}_{a \in A_2} \cos(w, a)$$

where $\cos(w, a)$ represents the cosine similarity between word vectors.

2) **Differential association of target sets:**

$$s(T_1, T_2, A_1, A_2) = \sum_{t \in T_1} s(t, A_1, A_2) - \sum_{t \in T_2} s(t, A_1, A_2)$$

3) **Effect size ($d$):**

$$d = \frac{\text{mean}_{t \in T_1} s(t, A_1, A_2) - \text{mean}_{t \in T_2} s(t, A_1, A_2)}{\text{std\_dev}_{t \in T_1 \cup T_2} s(t, A_1, A_2)}$$

The WEAT score and effect size quantify the strength and significance of the association between target and attribute word sets[9].

## IV. Results

In this study, we analyzed the semantic biases between art and commercial films using WEAT scores derived from Word2Vec models. The results across eight model configurations highlight notable distinctions in how different training methods and vocabulary generation techniques influence the measurement of biases. Tables and figures are provided to illustrate these findings in detail.

### A. Word2Vec Model Comparison

*a) CBOW:* CBOW-based configurations displayed broader contextual patterns, capturing general relationships between genres rather than distinct biases. For example, in the CBOW + TF-TF configuration, traditional and narrative-heavy genres such as *historical drama* and *drama* were consistently associated with art films (e.g., *historical drama* vs. *animation*, score = 1.044). Conversely, highly commercial genres like *animation* were aligned with commercial films.

[5] Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *arXiv preprint arXiv:1301.3781*.

[6] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed Representations of Words and Phrases and Their Compositionality. *Advances in Neural Information Processing Systems (NeurIPS)*.

[7] Ramos, J. (2003). Using TF-IDF to Determine Word Relevance in Document Queries. *Proceedings of the First International Conference on Machine Learning*.

[8] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research, 3*, 993–1022.

[9] Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science, 356*(6334), 183-186.

*b) Skip-gram:* Skip-gram-based configurations revealed sharper distinctions and stronger biases. In Skip-gram + LDA-LDA, for instance, *performance* was strongly associated with art films (*performance* vs. *comedy*, score = 1.365), while *family* was closely linked to commercial films (*family* vs. *western*, score = -1.359). This demonstrates Skip-gram's ability to capture subtle, infrequent relationships.

### *B. Vocabulary Generation Technique Comparison*

*a) TF-IDF:* TF-IDF highlighted clear genre-specific biases, providing the most distinct and interpretable results. In CBOW + TF-TF, *historical drama* and *animation* exhibited the highest contrast (score = 1.044), reinforcing the value of this approach for pinpointing specific associations.

*b) LDA:* LDA-based techniques tended to capture broader, more generalizable patterns, which occasionally diluted genre-specific biases. However, certain configurations, such as Skip-gram + LDA-TF, still produced significant results (e.g., *performance* vs. *comedy*, score = 1.305).

### *C. Notable Genre Associations*

- **Art Films**: Repeatedly linked with *historical drama*, *performance*, and *science fiction (SF)*. These genres emphasize artistic depth, cultural narratives, and creative storytelling.
- **Commercial Films**: Dominated by *animation*, *documentary*, and *comedy*. These genres prioritize accessibility, entertainment, and mass appeal.

### *D. Figures*

- **Figure 1**: Heatmap for CBOW + TF-TF
- **Figure 2**: Heatmap for Skip-gram + TF-TF
- **Figure 3**: Heatmap for CBOW + TF-LDA
- **Figure 4**: Heatmap for Skip-gram + TF-LDA
- **Figure 5**: Heatmap for CBOW + LDA-TF
- **Figure 6**: Heatmap for Skip-gram + LDA-TF
- **Figure 7**: Heatmap for CBOW + LDA-LDA
- **Figure 8**: Heatmap for Skip-gram + LDA-LDA

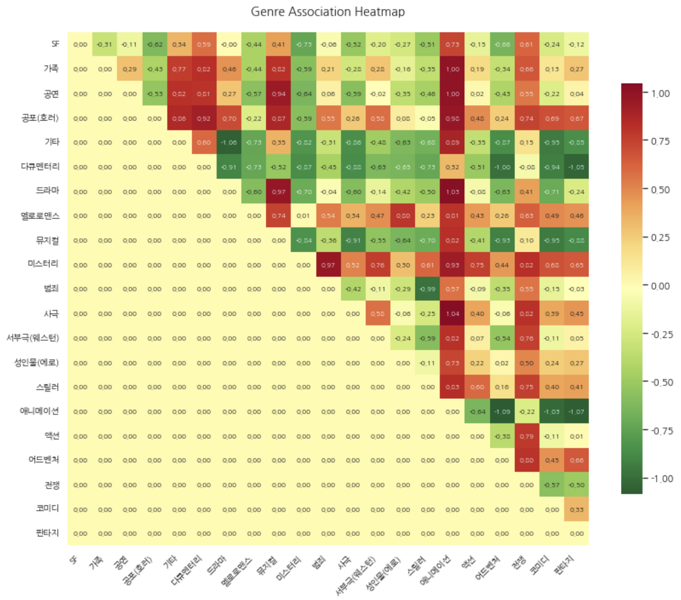

Fig. 1. Heatmap for CBOW + TF-TF

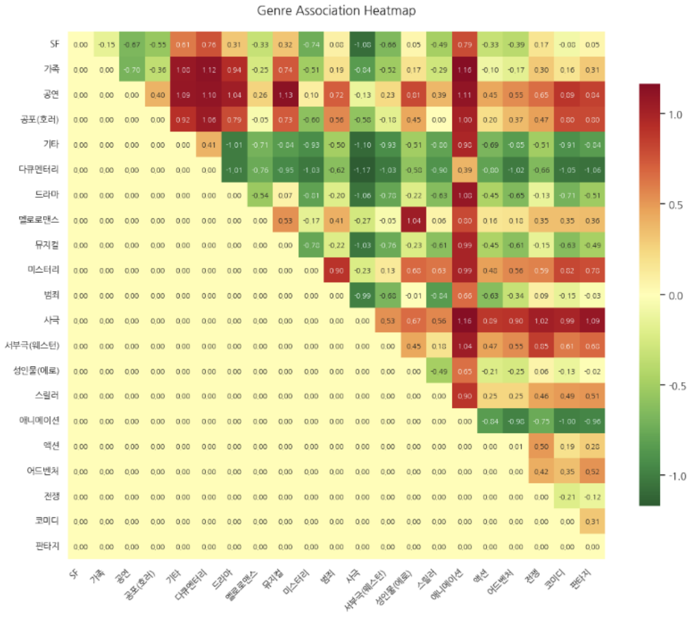

Fig. 2. Heatmap for Skip-gram + TF-TF

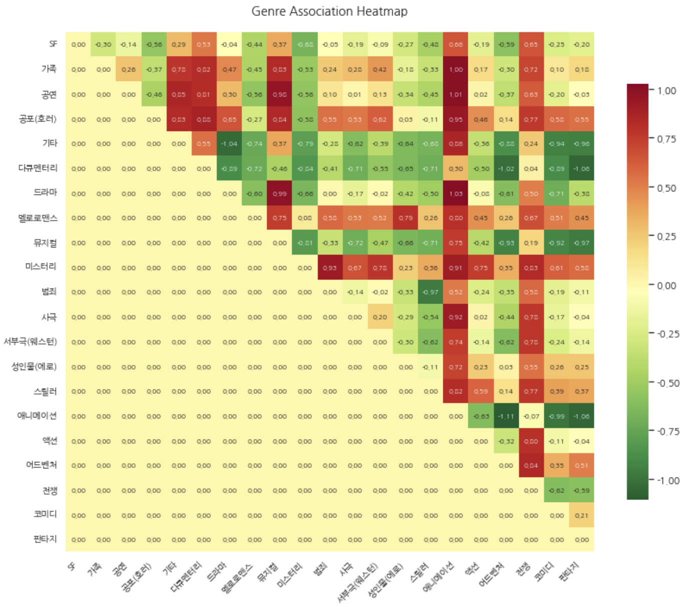

Fig. 3. Heatmap for CBOW + TF-LDA

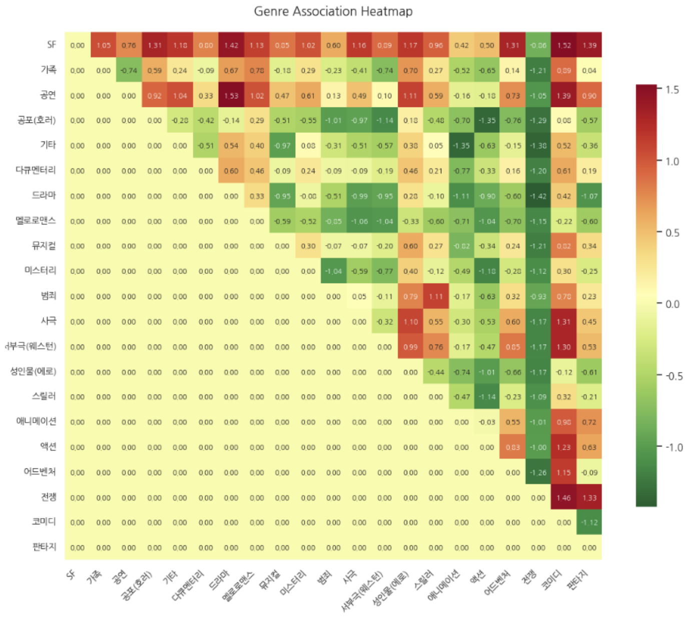

Fig. 5. Heatmap for CBOW + LDA-TF

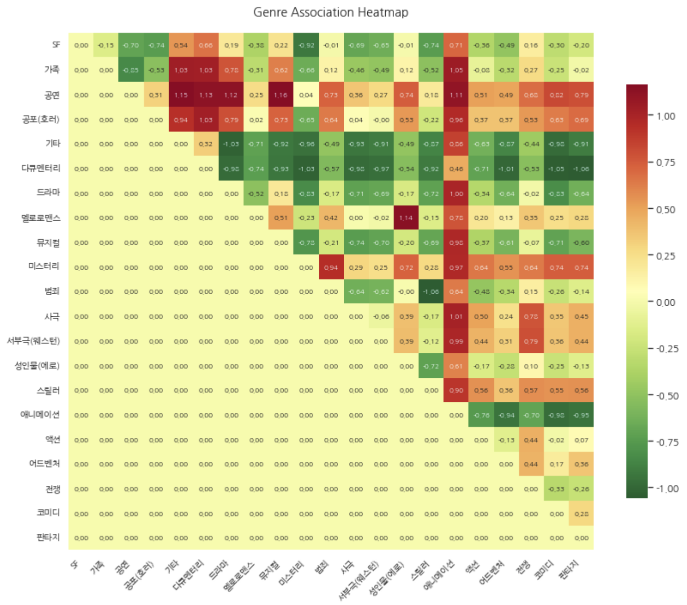

Fig. 4. Heatmap for Skip-gram + TF-LDA

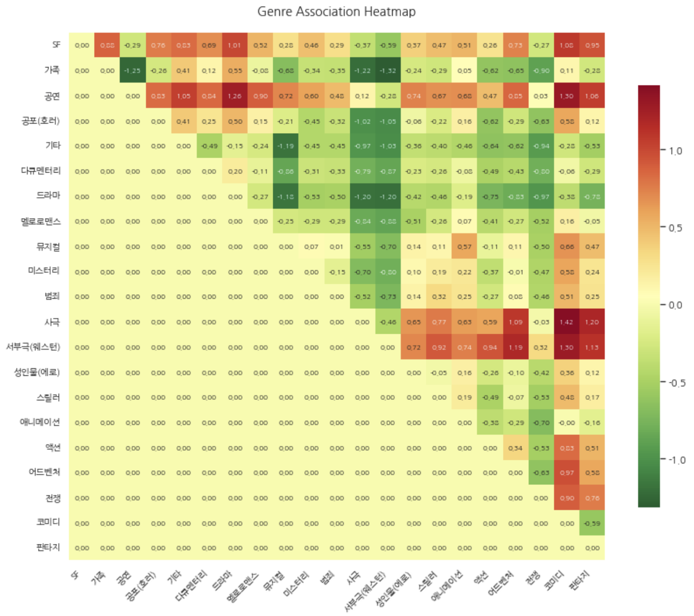

Fig. 6. Heatmap for Skip-gram + LDA-TF

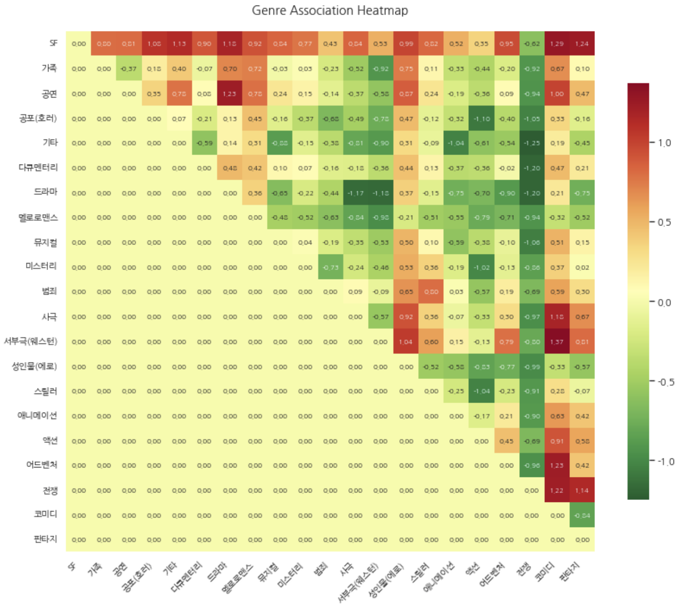

Fig. 7. Heatmap for CBOW + LDA-LDA

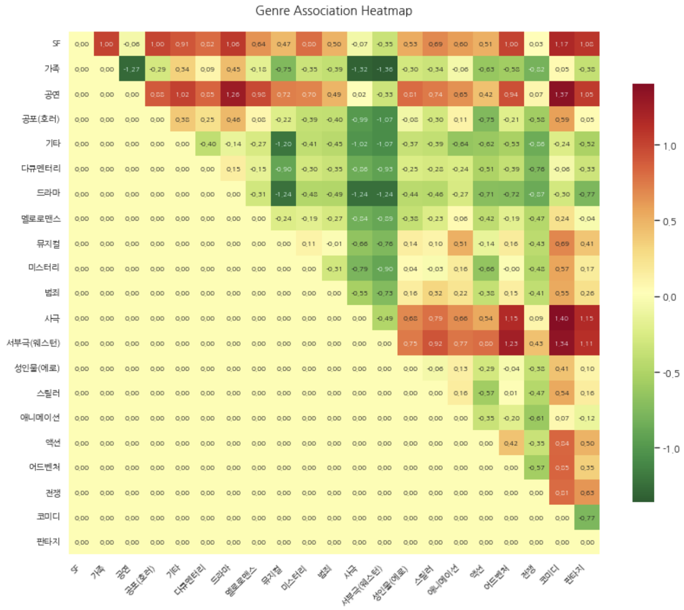

Fig. 8. Heatmap for Skip-gram + LDA-LDA

## V. Discussion

### A. Implications of Model Choice

The comparison between CBOW and Skip-gram highlights the trade-offs in capturing semantic biases:

- **CBOW**: Excels at general contextual relationships but lacks specificity for nuanced biases.
- **Skip-gram**: Better suited for identifying subtle and infrequent associations, making it more effective for analyzing genre-specific semantic biases.

### B. Importance of Vocabulary Selection

The choice of vocabulary generation technique significantly influenced the results:

- **TF-IDF**: Produced clearer and more interpretable biases due to its focus on genre-specific keywords.
- **LDA**: While effective in identifying thematic patterns, it occasionally diluted strong associations due to its broader scope.

### C. Genre-Specific Biases

Our analysis confirms that certain genres inherently align with artistic or commercial paradigms. For instance:

- *Historical drama* and *performance* were consistently linked to art films, reflecting their narrative depth and artistic focus.
- *Animation* and *comedy* were tied to commercial films, underscoring their appeal to mainstream audiences.

### D. Limitations and Future Research

This study focused solely on Korean film synopses, which may limit generalizability to other cultural contexts. Future research could explore:

- Applying similar methodologies to different datasets, such as global film databases.
- Extending the analysis to include temporal trends, examining how genre biases evolve over time.

## VI. Conclusion

This study demonstrated the effectiveness of Word2Vec models and WEAT in uncovering semantic biases between art and commercial films. The findings highlight the impact of model configurations and vocabulary selection on bias measurement, providing valuable insights into the cultural and commercial dynamics of the film industry.

## References

[1] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," *arXiv preprint arXiv:1301.3781*, 2013. [Online]. Available: https://arxiv.org/abs/1301.3781

[2] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed Representations of Words and Phrases and their Compositionality," *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 3111–3119, 2013.

[3] A. Caliskan, J. J. Bryson, and A. Narayanan, "Semantics Derived Automatically from Language Corpora Contain Human-like Biases," *Science*, vol. 356, no. 6334, pp. 183–186, 2017. [Online]. Available: https://doi.org/10.1126/science.aal4230

[4] J. Ramos, "Using TF-IDF to Determine Word Relevance in Document Queries," in *Proceedings of the First International Conference on Machine Learning*, 2003.

[5] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.

[6] L. Bolelli, Ş. Ertekin, and C. L. Giles, "Text Mining for Movie Review Sentiment Analysis," *Journal of Information Science*, vol. 45, no. 1, pp. 28–44, 2019. [Online]. Available: https://doi.org/10.1177/0165551518768797

[7] Q. Mei, X. Ling, M. Wondra, H. Su, and C. Zhai, "Topic Sentiment Mixture: Modeling Facets and Opinions in Weblogs," in *Proceedings of the 16th International Conference on World Wide Web*, 2007, pp. 171–180. [Online]. Available: https://doi.org/10.1145/1242572.1242595

[8] A. Al-Doulat and M. Ramadan, "A Survey on Word Embedding Techniques for Natural Language Processing," *Procedia Computer Science*, vol. 170, pp. 417–426, 2020. [Online]. Available: https://doi.org/10.1016/j.procs.2020.03.115

[9] N. Garg, L. Schiebinger, D. Jurafsky, and J. Zou, "Measuring Social Biases in Social Media Texts using WEAT," in *Proceedings of the 2018 World Wide Web Conference (WWW '18)*, pp. 901–910, 2018. [Online]. Available: https://doi.org/10.1145/3178876.3186130

[10] M. Sap, G. Park, J. Eichstaedt, M. Kern, D. Stillwell, M. Kosinski, L. Ungar, and H. A. Schwartz, "Analyzing Gender Representations in Film through Data Mining," *PLOS ONE*, vol. 9, no. 6, p. e104244, 2014. [Online]. Available: https://doi.org/10.1371/journal.pone.0104244

[11] L. Manovich, "Cultural Analytics: A New Approach to Understanding Films through Text," in *Procedia Computer Science*, vol. 51, pp. 1180–1189, 2016. [Online]. Available: https://doi.org/10.1016/j.procs.2016.06.111