Final Project Submission
Association of Power and Agency with Character Relationships in Film

Submitted by: *Ishaan Saxena, Nikita Rajaneesh*

Abstract

Gender bias is prevalent in our society and has been for a long time. Although the male-female dynamic and traditional roles in society may slowly be shifting, the effects of the historical bias linger in modern gender roles. In order to study such biases, we utilize the connotation frames of power and agency to present an analysis of the relationships between pairs of male and female characters in film as an extension to the work done by Sap et al. [9]. We combine the descriptors obtained by the Relationship Modeling Network proposed by Iyyer et al. [5] with information about characters' power and agency. Using this, we determine the kinds of relationships (as characterized by the descriptors) where characters are portrayed with varying levels of power and agency. This reveals a bias in relationships with such descriptors. We found that there were some relationship descriptors for which the difference in the male and female power/agency is significantly higher (indicating females have more power/agency) or significantly lower than what would be a usual deviation.

1 Introduction

How a character is written plays an important part in how they are perceived by an audience. Even the same kinds of actions may be projected differently by altering the use of language describing those actions. For instance, an actor who 'demands' something will be seen as one with more power than one who 'requests' something [9]. Such a difference in the framing of actions can influence the impressions the audience has, as well as the assumptions they make, about the actor. This has a significant influence on how people form stereotypes on things such as gender norms [2]. Furthermore, how a character is written itself may be influenced by stereotypes already pervasive in the society. We wish to use such a notion of framing of actions to better understand the presence of gender bias in film.

A more traditional framework to study gender bias in film might exploit prescriptive measures such as the Bechdel test [1] or character tropes [6]. Instead, Sap et al. [9] present a more in-depth, quantitative analysis. They introduce connotation frames of power and agency as an extension of the connotation frame lexicon introduced by Rashkin et al. [8]. Using these connotation frames, they reveal that there exist subtle, but prevalent gender bias in how characters are portrayed in modern films. Specifically, they show that female characters tend to have less agency than male characters.

1.1 Novel Aspects

We utilize these connotation frames of power and agency to present an analysis of the relationships between pairs of male and female characters in film as an extension to the work done by Sap et al. [9]. We combine the descriptors obtained by the Relationship Modeling Network proposed by Iyyer et al. [5] with information about characters' power and agency. Then we determine the kinds of relationships (as characterized by the descriptors) that are portrayed by characters with varying levels of power and agency. By studying such character dyads, we identify the presence of gender bias in the portrayal of characters in specific kinds of relationships in modern film.

2 Problem Definition

We use data from the Cornell Movie Dialog Corpus [3] and scripts from IMSDb[1]. We use GloVe [7] trained on the common crawl to obtain embeddings for words.

2.1 Preliminaries

Let S be the set of all characters. We obtain a set of relationships $R_{mf} \subset S \times S$ such that for all such pairs, the first character is female, the second is male, and the two have a significant level of interaction with each other. This allows us to focus on relationships between male-female character dyads.

We obtain a set $D = \{\mathbf{d_1}, \ldots, \mathbf{d_l}\}$ of l distinct relationship descriptors which are used to characterize the relationship between a pair of characters. For any relationship $r \in R_{mf}$ between a pair of characters, we may obtain the probabilities $\Pr(\mathbf{d_i} \mid r)$ for each descriptor such that $\sum_i \Pr(\mathbf{d_i} \mid r) = 1$. We say that a descriptor is the most likely descriptor for a relationship r iff

$$\mathbf{d_j} = \max_i \Pr(\mathbf{d_i} \mid r) \tag{1}$$

Note that each of these descriptors is a d-dimensional vector learned in the same embedding space as words used. As such, we can infer the kind of relationship by looking at the words which are cosine similar to each descriptor [5].

Finally, we compute maps $m_a : S \to [-1, 1]$ and $m_p : S \to [-1, 1]$ to the character's agency and power over the theme as defined by Sap et al. [9].

2.2 Task

Given a pair of characters $(C_f, C_m) \in R_{mf}$, and probabilities associated with each descriptor for their relationship $\mathbf{p} = (\Pr(d_i \mid C_f, C_M))_{i \in [l]}$, we wish to see if there is any association between the characters' relationship and their agency and power values $m_a(C_f), m_a(C_m)$ and $m_p(C_f), m_p(C_m)$.

Furthermore, we wish to find relationship descriptors $\mathbf{d_j}$ such that power and agency values reveal gender bias in relationships where it is the most likely descriptor.

3 Technical Approach

For any relationship, a span, analogous to a scene, is the set of words (unigram features) used by the two characters in a certain time-frame. For any relationship $(c_1, c_2) \in R_{mf}$, we define $S_{c_1,c_2} = \{s_t\}$ as the set of spans where characters c_1 and c_2 are present. We obtained the set of relationship descriptors using the Relationship Modeling Network (RMN) proposed by Iyyer et al. [5]. The RMN consists of an input layer, a hidden layer which uses ReLU activation, and an output layer which uses the softmax function. For each span, the input features are $v_{s_t} = \frac{1}{|s_t|} \sum_{w \in s_t} v_w$ where v_w is the embedding for the word w, and the character and movie embeddings v_{c_1}, v_{c_2}, v_b.

Through a variant of dictionary learning the model learns a descriptor matrix R of size $l \times d$. l is the number of descriptors/topics for the set of spans S_{c_1,c_2}. d is the length of a relationship descriptor. Given the input features, the objective of the RMN is to reconstruct S_{c_1,c_2} using a linear

[1] Available at http://imsdb.com/

combination of relationship descriptors: rows of R. d_t is the d-dimensional vector that represents the relationship state at span s_t. The model's loss function ensures that $r_t = R^T \cdot d_t$ is similar to v_{s_t}.

For each character, we obtain values of power and agency using the methods proposed by Sap et al. [9]. The authors show that verbs used by characters can be used to quantify their power and agency. We extract the verbs used by each character via dependency parsing and co-reference resolution using SpaCy [4]. Then we use the annotations provided by Sap et al [9] to assign a value between $[-1, 1]$ for each character's power and agency. The annotations provide a list of verbs which have positive or negative power and agency. For each character we get the aggregate power and agency by counting the number of "positive" and "negative" verbs[2] used by them. We normalize this value with the total number of verbs used by that character.

4 Evaluation

4.1 Rationale

We wanted to understand whether power and agency have any association with the relationship descriptors. To do so, we conducted statistical analyses to check for any correlation between the difference in power and agency of a male and female and the type of their relationship. We also wanted evaluate whether power and agency have any predictive power: Can a given set of power and agency for a pair of characters predict the type of relationship?

4.2 Predictive Power of Agency and Power

To explore the predictive power of power and agency we modified one of the inputs to Mohit Iyyer's neural network: the character embedding. We added power and agency as additional features to the character embedding to see if the relationship trajectories would be more accurate.

Due to the nature of the loss function used in the neural network, proposed by Mohit Iyyer, adding power and agency to the character embedding hardly made any difference to the relationship trajectories. The loss function, see below, is a hinge loss that minimizes the inner product between r_t and the average of the negative samples v_{s_n} and maximizes the inner product between r_t and v_{s_t}.

$$
\begin{aligned}
J(\theta) &= \sum_{t=0}^{|S_{c_1,c_2}|} \sum_{n \in N} \max(0, 1 - r_t v_{s_t} + r_t v_{s_n}) + \lambda X(\theta) \\
X(\theta) &= ||RR^T - I||
\end{aligned}
$$

The negative samples are randomly sampled spans from the dataset. The term $X(\theta)$ ensures that the relationship descriptors are not similar. By maximizing the the inner product between r_t and v_{s_t} the model makes r_t similar to v_{s_t}. Hence, for each span the relationship trajectories represents span t. So, the loss function mainly focuses on the span embedding. Hence, by adding power and agency to the character embedding we aren't achieving a smaller loss and more accurate relationship trajectories. The following table consists of stats for 5 runs:

[2]More specifically, these are verbs which amplify or diminish the agency and power over theme for characters. Positive and negative is an over-simplification, but is done for the sake of clarity

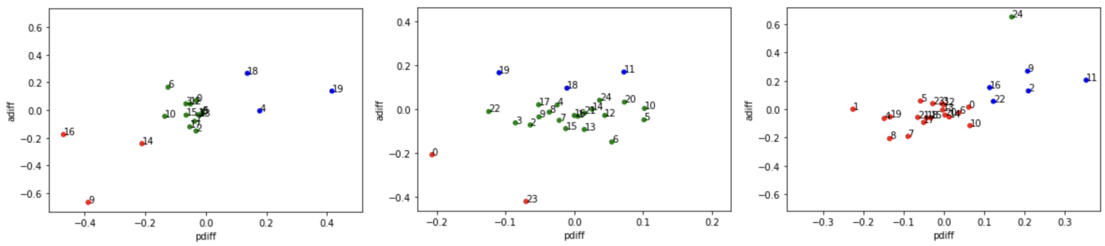

Figure 1: Descriptors with their average power and agency differential values. The plot on the left is 20 descriptors obtained with a word dropout rate of 0.70, the center is 25 descriptors obtained with a word dropout rate of 0.67, the right is 25 descriptors obtained with a word dropout rate of 0.75.

Model	Avg Loss	Std Dev
RMN	507.32	7.769
RMN with P/A	504.88	6.412

4.3 Power and Agency Differential Analysis

In order to find relationship descriptors $\{\mathbf{d_i}\}$ such that power and agency values reveal the presence of gender bias in character portrayals, we aggregate the power and agency differences for all relationships grouped by their most likely descriptors. That is, for each relationship $(C_f, C_m) \in R_{mf}$, we obtain the most like descriptor $\mathbf{d_j} = \max_i p_i(r)$, and then group the relationships by the most likely descriptor. For each of these groups, we evaluate the mean difference between the female and male power and agency:

$$R_j := \{r \in R_{mf} : \mathbf{d_j} = \max_i p_i(r)\} \tag{2}$$

$$\Delta p_j = \frac{1}{|R_j|} \sum_{(C_f, C_m) \in R_j} m_p(C_f) - m_p(C_m) \tag{3}$$

$$\Delta a_j = \frac{1}{|R_j|} \sum_{(C_f, C_m) \in R_j} m_a(C_f) - m_a(C_m) \tag{4}$$

If for some descriptor $\mathbf{d_j}$, the values of Δp_j or Δa_j deviate significantly from expectation, it indicates that for such relationships the female or male had lopsided power or agency values.

Figure 1 shows the relationship descriptors and their $\Delta a_j, \Delta p_j$ values for three separate sets of descriptors obtained from independent executions of the RMN. While most of the relationship descriptors are seen to have reasonable $\Delta a_j, \Delta p_j$ values, some relationship descriptors deviate significantly. In the trial represented in Figure 1 on the left, for instance, Descriptors 14 represents work based relationships, and Descriptor 9 represents girlish relationships. In each, we can see clearly that the female character has lesser power and agency than the male characters. There is a similar observation in the second trial (central plot in Figure 1) with Descriptor 23 which represents relationships in context of military and other organizations. On the flip side, descriptors with higher female agency/power depict food-based relationships, sexual relationships, moody relationships, etc. These observations show the presence of gender bias in the portrayal of such relationships which reflects existing biases in the real word. The following table contains some examples of these descriptors from the first trial (left plot in Figure 1):

Descriptor	Cosine Similar Words	Δp_j	Δa_j
4	cake, vegetables, noodles	0.177988	-0.005806
9	perky, woah, girlish	-0.388	-0.667
14	database, invoice, photocopy	-0.211	-0.242
16	generations, lived, countryside	-0.470588	-0.176471
18	newborn, remarried, widower	0.137105	0.265771

However, not all descriptors are easy to infer from. Some descriptors, though reasonable, do not reflect similar biases as real life, or if they do, it is not entirely clear why. Other descriptors may often be entirely nonsensical. The nonsensical descriptors mostly result in biased values of $\Delta a_j, \Delta p_j$. Improving the quality of descriptors may be done by increasing the amount of data, having richer features, changing the word dropout probabilities, or through other methods. This is an interesting line of research for future work.

5 Summary

From the power and agency in combination with relationship descriptors obtained by the RMN, we show that there exist certain kinds of gender biases in film in the portrayal of characters with specific kinds of relationships. However, for most relationships, the bias measure is well within expected deviation. Furthermore, a few descriptors may sometimes be nonsensical. Obtaining better descriptors is an interesting topic for future work. We weren't able to show that power and agency have any predictive power.

An interesting extension to our analysis would be to implement a model that takes a set of power and agency values for a pair of characters and predicts whether a relationship is positive or negative. An important part of this approach is to come up with a reasonable method to label a relationship as positive or negative.

5.1 Difference from proposal

In the project proposal, we said that we would also look at how a power and agency dynamic effects relationships differently over genres and time. Unfortunately, we did not have time to do this. This could definitely be an interesting extension to our analysis.

6 Team member contribution

Nikita Rajaneesh extracted verbs for each character via dependancy parsing and co-reference resolution using SpaCy [4]. She also modified the relationship modelling network proposed by Iyyer et al [5] to include power and agency as additional features in the character embedding. Additionally, she attempted another approach for understanding the predictive power of power and agency: a logistic regression model that takes as input power and agency of a pair of characters and predicts whether the given relationship is positive or negative.

Ishaan Saxena, utilized the set of verbs used by characters to learn the power and agency maps using the method proposed by Sap et al [9]. He obtained relevant data and features for the RMN model, and performed hyperparametertuning of the RMN model. He conducted an analysis on the association between power/agency differentials and relationship descriptors.

References

[1] A. Agarwal, J. Zheng, S. Kamath, S. Balasubramanian, and S. A. Dey. Key female characters in film have more to talk about besides men: Automating the bechdel test. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 830–840, 2015.

[2] E. Behm-Morawitz and D. E. Mastro. Mean girls? the influence of gender portrayals in teen movies on emerging adults' gender-based attitudes and beliefs. *Journalism & Mass Communication Quarterly*, 85(1):131–146, 2008.

[3] C. Danescu-Niculescu-Mizil and L. Lee. Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs. In *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics, ACL 2011*, 2011.

[4] M. Honnibal and I. Montani. spacy 2: Natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing. *To appear*, 7(1), 2017.

[5] M. Iyyer, A. Guha, S. Chaturvedi, J. Boyd-Graber, and H. Daumé III. Feuding families and former friends: Unsupervised learning for dynamic fictional relationships. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1534–1544, 2016.

[6] N. Lee, Y. Bang, J. Shin, and P. Fung. Understanding the shades of sexism in popular tv series. In *Proceedings of the 2019 Workshop on Widening NLP*, pages 122–125, 2019.

[7] J. Pennington, R. Socher, and C. D. Manning. Glove: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, 2014.

[8] H. Rashkin, S. Singh, and Y. Choi. Connotation frames: A data-driven investigation. *arXiv preprint arXiv:1506.02739*, 2015.

[9] M. Sap, M. C. Prasettio, A. Holtzman, H. Rashkin, and Y. Choi. Connotation frames of power and agency in modern films. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2329–2334, 2017.