Bauhaus-Universität Weimar
Faculty of Media
Degree Programme Computer Science for Digital Media

Analyzing Debaters' Persuasiveness on Change My View

Master's Thesis

Vishal Khanna
Born Mar. 9, 1995 in New Delhi, India
Matriculation Number 120333

1. Referee: Prof. Dr. Benno Stein
2. Referee: PD. Dr. Andreas Jakoby

Submission date: December 8, 2021

Declaration

Unless otherwise indicated in the text or references, this thesis is entirely the product of my own scholarly work.

Weimar, Germany, December 8, 2021

. . . .
Vishal Khanna

Abstract

Existing work on persuasion analysis in online discussions concentrates on the comments of each discussion individually, disregarding debater-level analysis over the course of multiple discussions. In this thesis, we propose to quantify the persuasion of debaters in the online discussion forum: 'Change My View'. Accordingly, we first group debaters based on their effectiveness in persuasion. Then, we analyze their argumentative text content based on its lexical, syntactical, semantic, and pragmatic attributes. These attributes are exploited to find diverse insights about debaters' effective persuasive strategies. We also model the evolution of the persuasion strategies of debaters as they gain experience. Besides, we extend existing studies on modeling persuasiveness in CMV by proposing a classification task for debater's persuasion effectiveness. Based on various features at the lexical, syntactical, semantic, and pragmatic levels, we implement a new approach and evaluate it against a strong baseline.

Contents

1 Introduction **1**
1.1 Motivation 1
1.2 Contributions 2

2 Background **4**
2.1 Change My View 4
2.2 Related Work 5

3 Debaters in ChangeMyView **8**
3.1 Data Preparation 8
3.2 Analyzing CMV Debaters' Activities 11
3.2.1 Quantifying Debaters' Effectiveness in Persuasion 12
3.2.2 Comment Length and Persuasion 13
3.2.3 Audience Engagement and Persuasion 14
3.2.4 Evolution of Persuasion Strategies Over Time: Impact of Past Experience 15
3.3 Analysis of Debaters' Comments' Content 19
3.3.1 Lexical Analysis of Debaters' Comments 20
3.3.2 Semantic Analysis of Debaters' Comments 23
3.3.3 Syntactical Analysis of Debaters' Comments 25
3.3.4 Pragmatic Analysis of Debaters' Comments 30

4 Modelling CMV Debaters' Persuasiveness **36**
4.1 Prediction Task 36
4.2 Prediction Features 37
4.3 Experiments 39
4.4 Results and Discussion 41

5 Conclusion and Future Work **43**

Bibliography **46**

Chapter 1

Introduction

1.1 Motivation

Persuasion is the ability to convince someone to do a certain action or form a particular belief. Persuasion has always affected the dynamics of communication and social interactions, from engaging with customers while advertising a product to influencing the discourse of political elections.

Since the internet and social media have connected a vast number of people, effective communication and persuasion have become more important than ever. Persuasion has been shown to be employed positively (e.g., raising awareness on critical issues like climate change) and negatively (e.g., changing voter behavior in elections[1] and disseminating propaganda and fake news). Much of the previous work on persuasion is focused around its manifestations in isolated discussions and does not consider the overall persuasiveness of the participants over the course of several discussions. Hence, a more profound understanding of what makes some participants more effective in persuasion than others would yield valuable insights into the social and psychological factors which govern its effect on people.

Existing research studies on persuasion give several insights on its various aspects in online discussions. Multiple related papers successfully model persuasion through a variety of linguistic, argumentative, and user-interaction-based features. Additionally, some papers investigate the role of argumentative units and their semantic types in driving persuasion via manually annotating and analyzing online arguments. Furthermore, few papers study temporal patterns and user-level interaction dynamics of persuasion in online discussions.

[1] `en.wikipedia.org/wiki/Facebook-Cambridge_Analytica_data_scandal`

1.2 Contributions

Similar to the work discussed above, this thesis aims to understand persuasion in the domain of online discussions, specifically in the Reddit-based discussion forum: Change My View(CMV). CMV provides an open and moderated platform for its users to engage in civilized discussions using sound arguments. Additionally, CMV allows users to indicate if they have been persuaded by another user(s)' comment through the delta mechanism, making it suitable for our analysis of persuasion.

Among the various possible directions of analyzing persuasion, this thesis proposes a novel approach which focuses on the *debater-level* analysis of persuasion in CMV discussions. We summarize the goals of this thesis through the following research questions:

1. What aspects of user behavior and their text content on CMV separate effective debaters from ineffective ones? What insights about effective persuasion can be obtained by analyzing effective CMV debaters w.r.t. these attributes?

2. How effective are the lexical, syntactical, semantic, and pragmatic features obtained from CMV users' text content in modeling their effectiveness in persuasion?

To this end, we categorize CMV debaters based on their effectiveness in persuasion and examine various traits and behaviors which distinguish good, average, and poor debaters from each other. Specifically, we find similarities and differences in the debaters' comments' text content based on their syntactical, semantic, lexical and pragmatic attributes. We also examine the evolution of debaters' persuasion strategies as they gain more experience. The analysis methodology and results are detailed in chapter 3.

Besides, we model the persuasion effectiveness of CMV debaters through their argumentative comments on CMV discussions. We investigate the role of semantic, syntactical, lexical, and pragmatic features of debaters' comments on their persuasiveness. In comparison to two commonly used feature groups (vocabulary interplay and bag of words) Tan et al. [2016], we perform a series of persuasiveness classification experiments and analyze the results, highlighting the role of each feature set. Chapter 4 includes the experiments and the results for these classification experiments.

Our analysis of CMV debaters' persuasiveness yields several insights including the following:

- Successfully engaging with the audience by evoking responses from them leads to higher effectiveness in persuasion.

- Effectiveness in persuasion improves over time for mediocre debaters.
- Phrasing arguments such that they are (1) semantically similar to those of the opposing party, and at the same time (2) have high semantic diversity/variability is characteristic of effective persuasion.
- Effective debaters tend to phrase their arguments with relatively less structural centrality than ineffective debaters.
- For modelling CMV debaters' effectiveness in persuasion, bag of words yields a modest baseline with a macro accuracy of 0.60 which was only surpassed by pragmatic features based on the frame types in the debaters' comments which obtained a macro score of 0.74.
- Distribution of argumentative units in the debaters' text content yields below average performance in modelling their persuasiveness. Argumentative features based on presence of certain argument types in the debaters' text content are not indicative of their effectiveness in persuasion.
- Framing of arguments in the comments is highly indicative of debaters' effectiveness in persuasion. Framing strategies that concentrate on 'Cultural Identity' and 'Political' beliefs are particularly effective. The distribution of frames in the debaters' arguments can be exploited to model their persuasiveness.

Chapter 2 provides relevant background and covers past works on persuasion. Chapter 3 introduces the methodology and analysis of various types of debaters in CMV based on their effectiveness in persuasion. Chapter 4 discusses the modeling task for the CMV debaters' effectiveness in persuasion. Finally, Chapter 5 concludes the thesis and presents future work.

Chapter 2

Background

2.1 Change My View

Change My View(CMV) is one of the many 'subreddits' on the aggregation and discussion website Reddit. Founded in 2013, CMV aims to encourage people to question their views by being exposed to those deviating from their own. The official CMV webpage represents itself as follows:

> A place to post an opinion you accept may be flawed, to understand other perspectives on the issue. Enter with a mindset for conversation, not debate.

Over the last years, there has been a rise in information sharing over online platforms. These platforms, utilizing recommendation systems, are known to connect 'similar' users, leading to the establishment of echo chambers where people are largely exposed to views similar to their own(Garimella et al. [2018]).

Such platforms don't encourage users to examine their views and address their individual biases, but rather reaffirm such biases and thereby contribute towards the society's polarization and promote divisiveness.

Platforms like CMV, although a rarity, grant their users the space to have their views criticized and the possibility to augment them.

Discussions on CMV begin with an Original Post by a user referred to as the Original Poster(we use the abbreviation OP for both interchangeably). The post contains OP's opinion on a particular topic along with relevant justifications and explanations. Once posted to CMV, other users whom we refer to as debaters can participate in the discussion by adding comments to it. The OP and the debaters can respond to a debater's comments and attempt to counter, cross-question their arguments; creating multi-layered and complex threads of conversations.

If the OP is successfully convinced by a debater's arguments, they may indicate that by awarding the debater with a *delta* and further explaining the reason for it. Hence, the OP awarding a delta to one or more debater(s) is a sign of successful persuasion. Additionally, other users apart from the OP could express their agreement/disagreement to a debater's arguments by upvoting/downvoting their comment(s). The net score of the comment is simply the difference between the number of its upvotes and downvotes. Hence, CMV discussions consist of debater's arguments in their comments along with an evaluation of their persuasion along two dimensions: (1) OP's evaluation, by awarding/not-awarding delta and (2) Crowdsourced evaluation of other users through the comment's overall score.

We note that CMV is fairly nonrestrictive in allowing any user to participate during the course of a discussion, leading to formation of complex interaction dynamics in a multi-party setting. However, CMV discussions are also actively moderated to maintain quality of argumentation in which all participants must abide by the rules stated on their wiki[1].

2.2 Related Work

CMV has served as a basis for several relevant works on understanding persuasion in online arguments/discussions. In this section, we group the existing work into two broad categories. The first category deals with the role of argumentative units and their semantic types in persuasion. The second considers the role of various feature types in modeling persuasiveness in CMV or similar platforms.

Egawa et al. [2019] annotate CMV discussions with argumentative components called elementary units (EUs) and their five semantic labels (testimony, fact, value, policy, rhetorical statement). They also propose a Bi-LSTM based classification model to detect EUs and their semantic types in argumentative texts. Their annotations on CMV discussions reveal the following:

- Mere presence or absence of certain EU semantic types does not indicate persuasiveness, but their effective use in a comment does.
- Higher presence of fact EU semantic type in the first half of a CMV comment correlates to its persuasiveness.
- Proportional distribution of EU semantic types can be used to distinguish between CMV comments and OPs.

[1] https://www.reddit.com/r/changemyview/wiki/rules

Hidey et al. [2017] annotate CMV discussions with other classes of argumentative features called claims and premises along with their semantic types(interpretation, evaluation, agreement, disagreement and ethos, logos, pathos). They find that the relative positional distribution of argumentative components in a CMV comment can be indicative of its persuasiveness.

Mensah et al. consider the task of predicting an OP's susceptibility towards persuasion by considering LIWC-based linguistic features of their submissions, their confidence in their beliefs gauged by the presence of hedges and boosters in their submissions, and the frequencies of their interactions with debaters over time. Tan et al. [2016] investigate the role of debaters' interaction dynamics with the OP in persuasion and discover crucial behavioral patterns:

- Early responding debaters in a discussion tend to be more successful.
- Up to a threshold, engaging with the OP improves a debater's odds of success.
- Higher debater participation in a discussion improves the odds of the OP's persuasion

They also propose a set of features based on the shared vocabulary between the OP and a debater and show that these features are quite effective in predicting persuasion.

Wei et al. [2016] consider relevance ranking of CMV comments by their score in a discussion. They find the comment's score to be influenced by it's temporal entry order as well as the past credibility of its corresponding debater. The credibility is measured by the number of prior deltas received by a debater.The following feature classes are used for the relevance ranking task:

- Linguistic features derived from the comment's text
- Interaction based features obtained by modelling the CMV discussion as a tree
- Argumentative features - proportion of argumentative text, argument relevance and originality

Li et al. [2020] demonstrate the effectiveness of arguments' structural features in persuasiveness prediction. They consider the following features based on proposition types(reference, testimony, fact, value, policy) in the debaters' texts:

- Proposition n-gram frequency
- N-gram frequency of supporting links between propositions

- Features based on graphical representation of argument structure - basic, serial, linked, convergent, divergent

They find the presence of (value, testimony) type bi-grams to be more prevalent in persuasive argumentative texts, indicating that the justification of claims with personal experiences is an effective persuasion strategy. Note that their experiments were performed on the DDO corpus which is based on a different online discussion forum than CMV. This motivates the need to investigate how well their insights generalize to the much larger CMV corpus[2].

Guo et al. [2020] hypothesize that the OP's persuasion in a CMV discussion doesn't happen instantaneously but rather gradually over the course of a multi-party conversation. They perform a prediction task of modeling the cumulative effect of a sequence of comments in a CMV discussion and detect the position where the OP's persuasion occurs. They also conduct a human study to evaluate the persuasiveness of debaters' arguments' and conclude that perception of persuasiveness differs across individuals and is influenced by one's idiosyncrasies i.e. the same argument could be persuasive for one person but non-persuasive for another. Hence, it is crucial to consider the OP's characteristics when evaluating a debater's arguments' persuasiveness. Khatib et al. [2020] build on this idea by modeling debater level characteristics in the form of past beliefs, personality traits, and interests that are obtained from the corresponding user's activity on Reddit. They find the similarity between these user-level attributes of the OP and debater to be indicative of effective persuasion.

Lastly, Atkinson et al. [2019] look at the explanations of successful persuasion provided by the OPs in CMV discussions where they explain their reasoning behind awarding delta(s) to debaters. The authors propose a word-level prediction task to determine which words in a debater's persuasive comment are repeated in the OP's explanation.

[2]https://www.cs.cornell.edu/ esindurmus/ddo.html

Chapter 3

Debaters in ChangeMyView

This chapter covers our analysis of CMV debaters' persuasion strategies. The work detailed in this chapter is motivated by the need to understand the differences in the persuasion strategies of those debaters who experience good success in CMV and those who don't.

We group debaters based on their effectiveness in persuasion and analyze their activities and their comments' text in four aspects - lexical, syntactical, semantic, and pragmatic. Additionally, we model the evolution of debaters' activity on CMV over time and analyze the impact of experience gained on their persuasion strategies. We address the following broad research questions through our analyses:

1. How do effective debaters' persuasion strategies differ from those of ineffective debaters? What insights about persuasion can be gained from analyzing effective debaters' persuasion strategies on CMV?

2. How do CMV debaters' persuasion strategies evolve as they gain experience in persuasion? Is this evolution similar for effective and ineffective debaters?

3.1 Data Preparation

We use the WebisCMV dataset from Khatib et al. [2020] as the data source for our analysis. The dataset comprises the CMV discussions from June 2005 till September 2017. While the dataset covers a diverse set of user and submission level attributes, here we briefly describe the two datasets relevant for our study:

- pairs.jsonl - This dataset consists of pairs of CMV comments by debaters. Each persuasive comment(one which has been awarded a delta) is paired

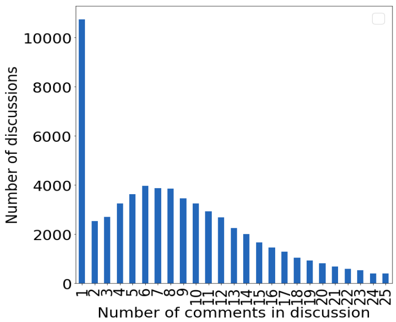

Figure 3.1: Distribution of number of comments in CMV discussions

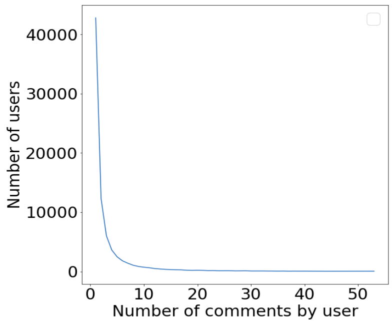

Figure 3.2: Distribution of number of comments by CMV users

with a non-persuasive comment, of matching length, from the same CMV discussion.

- threads.jsonl - This dataset consists of all CMV discussions from the above mentioned time period and contains discussion level attributes along with a structured representation of all comments in each discussion.

A preliminary analysis of the dataset indicates that the highest proportion of CMV discussions fail to garner any attention from the debaters and only receive a single comment which is an automated reply from the Auto Moderator bot. The discussions that witness user activity are distributed normally and peak in the range of $6 - 8$ comments, as shown in figure 3.1.

A similar trend can also be seen in the user-level activity on CMV, shown in figure 3.2. Most users only leave a single comment during the course of their activity and the majority of users have less than 7 comments. The yearly distribution of activity on CMV in figure 3.3 shows a significant increase both in the number of comments and number of discussions from 2013 to 2014 which was followed by a sharp decrease in 2015 post which the activity levels stabilized.

We construct our dataset of CMV debaters from the WebisCMV[1] threads dataset, comprising all CMV discussions during a period of five years, as follows:

1. We maintain a dictionary where each CMV user's unique username is the key and a list of all top-level comments in CMV discussions by that user is stored as the value against the key. Initially, this dictionary is empty.

[1]https://zenodo.org/record/3778298.YZoraHVKjCI

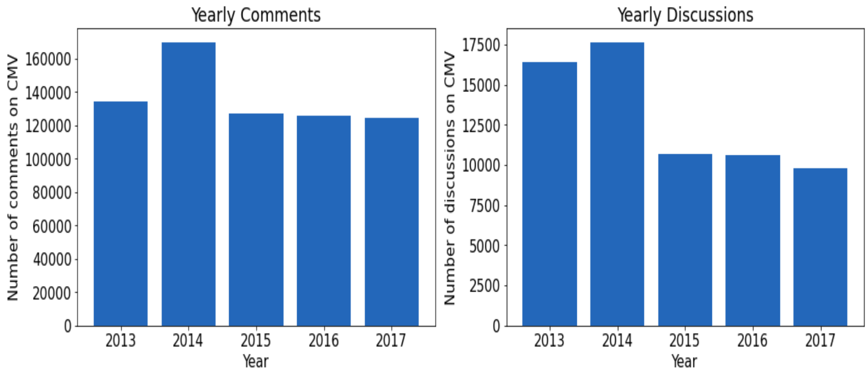

Figure 3.3: Number of comments(left) and discussions(right) on CMV from 2013 to 2017

2. We process each CMV discussion(with at least 10 comments) in the WebisCMV threads dataset as follows:

   (a) For each comment in the CMV discussion, if the comment or its user is not deleted or removed, add it to the list of comments for that user in our dictionary.

3. Once we are finished processing all CMV discussions, we iterate through our dictionary of CMV users and their lists of comments and remove all entries with less than five comments. Hence, we only consider debaters with at least five comments.

Note that the CMV discussions, similar to all discussions on Reddit, can have a number of top-level outer comments as well as inner comments which act as replies to the outer comments. These inner comments can often be non-argumentative such as corrections/clarifications and are thus ignored. We obtain a dataset of 13254 CMV debaters and their top-level comments on various discussions. We refer to this dataset as the CMV debaters dataset.

In figure 3.4, we look at the Pearson correlations between the debater's level of activity in CMV discussions and their level of success achieved. We measure a debater's activity in terms of their number of top-level comments in CMV discussions and their 'activity duration' which is the time elapsed between their first and last comments. We quantify a debater's success in persuasion by their number of delta comments, percentage of their total comments which receive a delta, and their median comment score. We find a significant correlation between the debaters' number of comments and their number of delta comments implying that debaters who comment more are more likely to accumulate a higher number of deltas just by having had more tries towards

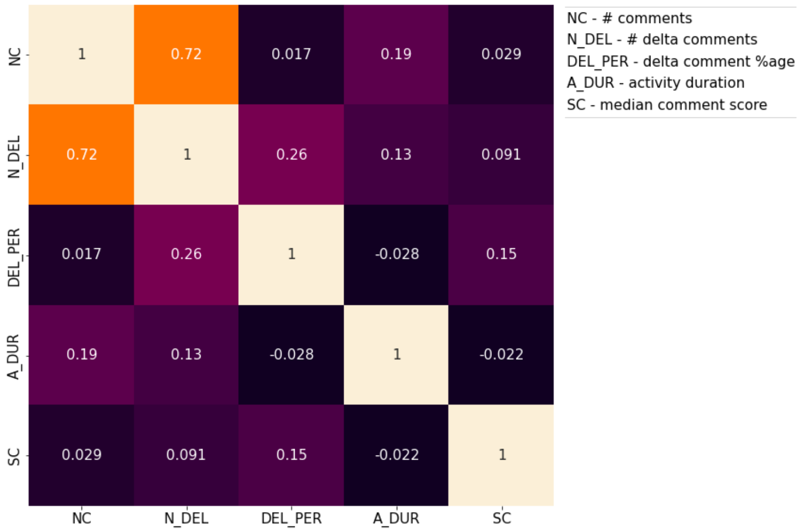

Figure 3.4: Pearson correlation coefficients between debaters' activity levels(# comments, activity duration) and their levels of success achieved(# delta comments, delta comment percentage, median comment score)

persuading the OPs. However, the number of comments doesn't correlate to the percentage of delta comments, hence, just commenting more frequently doesn't necessarily improve your odds of success in relative terms.

We also observe no significant correlation between debaters' number of comments and their duration of activity implying the presence of both the debaters who're active for a long time but don't comment as frequently as well as those who comment very frequently during a short activity duration. Hence, we observe the presence of debaters with varying levels of activity and no correlation of that with their success in persuasion.

3.2 Analyzing CMV Debaters' Activities

We begin our analysis of CMV debaters' persuasion strategies by studying their activities on the forum. We start this section by explaining our methodology for quantifying the debaters' effectiveness and experience in persuasion. We use those notions as the basis for our analysis for the remaining part of this chapter.

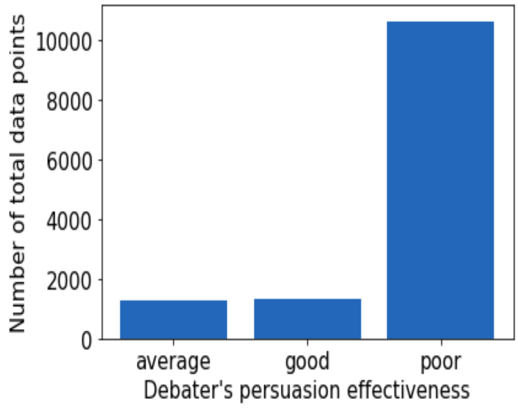

Figure 3.5: Distribution of CMV debaters by their effectiveness in persuasion

3.2.1 Quantifying Debaters' Effectiveness in Persuasion

To differentiate between effective and ineffective debaters, we consider the 'delta comment percentage' as the primary metric for evaluating a debater's effectiveness in persuasion. We choose 'delta comment percentage' since it determines the success rate for a debater normalized w.r.t. the number of comments(different CMV debaters have different number of comments). We refer to a debater's delta comment percentage as their 'persuasion effectiveness'.

Based on the persuasion effectiveness values, we group debaters into three groups as follows:

1. Good debaters - Effective debaters with delta comment percentage of 5% or above.

2. Average debaters - Somewhat effective debaters with delta comment percentage less than 5% but greater than 0%.

3. Poor debaters - Ineffective debaters with delta comment percentage 0%. These debaters don't receive any deltas during their activity duration on CMV.

In figure 3.5, we observe that the distribution of debaters in CMV is heavily skewed towards poor debaters with over eighty percent of the total samples belonging to that group. Hence, most debaters on CMV are ineffective in persuasion and only about twenty percent of all debaters achieve any success during their activity duration on the forum. We balance our CMV debaters dataset for persuasion effectiveness as follows:

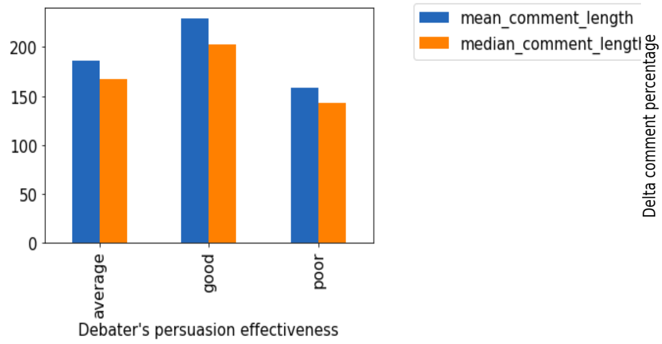

Figure 3.6: Aggregated comment lengths for the three debater types

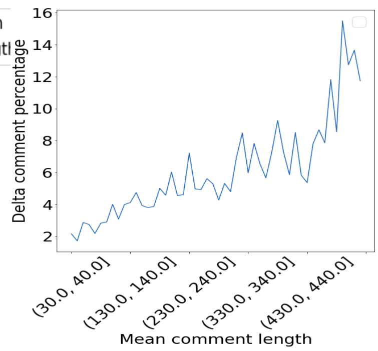

Figure 3.7: Change in delta comment percentage w.r.t. debater's mean comment length

1. For each good debater, we find an average and a poor debater such that the absolute difference between the number of comments between the three debaters is minimized. If multiple matching debaters exist of either type, we further choose the debater which also minimizes the absolute difference between their mean comment lengths.

Through our re-sampling of good, average, and poor debaters, we attempt to mitigate the difference in the number of comments and the mean comment length of the debaters in order to limit the possible influence of those attributes in our analysis. Note that we re-sample without replacement so each debater entry in our dataset is unique. We end up with a dataset of 3801 entries, evenly distributed across the three debater types.

3.2.2 Comment Length and Persuasion

Prior works have established a direct relationship between the argument's length and its persuasiveness. We look at the aggregated lengths of CMV debaters' comments and their relationship with the debaters' persuasion effectiveness. We compute the mean and median comment lengths for each debater and then further aggregate those values for each group of debaters(good, average, poor). We find a correlation between both mean and median comment lengths and persuasion effectiveness, concluding that effective debaters tend to write longer arguments on average.

Having reaffirmed the influence of arguments' length on their persuasiveness, we further investigate the nature of this correlation. Specifically, we attempt to establish whether the influence of a debater's comments' length towards their persuasion effectiveness is relevant up to a threshold or not. Additionally, we attempt to find out whether there exists a point beyond which

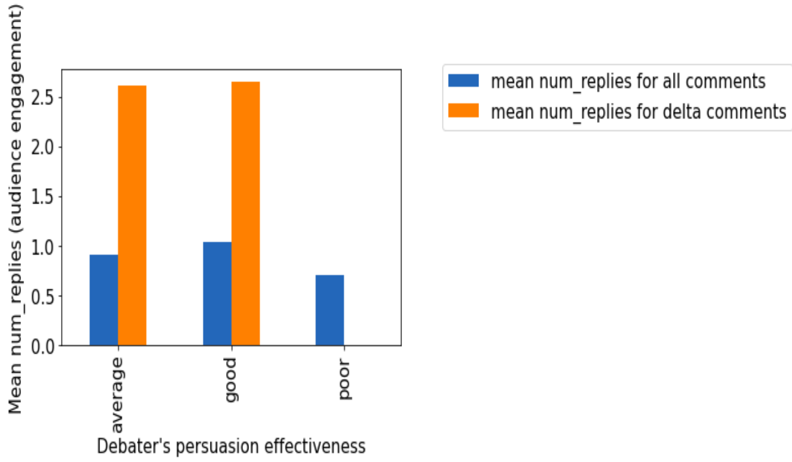

Figure 3.8: Audience engagement for CMV debaters measured as mean number of replies to their comments

makes the arguments longer doesn't increase the odds of persuasion any further. We do this by grouping debaters by their mean comment length into buckets of size 10 and computing the mean delta comment percentage for each such bucket.

Figure 3.7 shows that the aggregated length of the debaters' comments and their average delta comment percentage values approximate a linear correlation without the presence of a threshold beyond which the influence of length weakens. Hence, we conclude that debaters' effectiveness in persuasion consistently increases with increasing their comments' lengths. This calls for further investigation into how this increase in length leads to an increase in the debaters' effectiveness in persuasion, whether this increase in comment's length corresponds to inclusion of additional information or mere repetition/re-emphasis of the same information. We address these questions in our analysis of the debaters' comments in later sections.

3.2.3 Audience Engagement and Persuasion

In this section, we investigate the impact of effectively engaging with other CMV users on debaters' effectiveness in persuasion. We quantify a debater's 'degree of audience engagement' as the average number of replies to their top-level comments in CMV discussions. Hence, debaters with a higher number of replies to their comments engage with the audience to a greater degree.

Figure 3.8 shows that the mean number of replies to debaters' CMV comments decreases between good, average, and poor debaters. Hence, effective debaters tend to engage with the audience more effectively. Additionally, for both good and average debaters, the mean number of comment replies for the delta comments is much higher than the overall mean value for all comments implying that effectively engaging with the audience often preempts successful persuasion of the OP and can be a strong indicator of the debater's persuasion effectiveness.

3.2.4 Evolution of Persuasion Strategies Over Time: Impact of Past Experience

We model the evolution of debaters' persuasion strategies during their activity's duration on CMV to understand the impact of 'experience gained' on persuasion effectiveness. In this regard, we address the following broad research questions:

1. How do debaters' persuasion strategies change over time? What impact does the experience in persuasion accumulated over time have on debaters' effectiveness in persuasion?

2. Are these changes in persuasion strategies similar for effective and ineffective debaters?

We operationalize the 'experience gained' of a debater on CMV through a parameter called 'percentage of comments elapsed' defined as:

Percentage of comments elapsed For a comment C made at time t by debater D, the percentage of comments elapsed value is the percentage of the debater's total comments, when ordered temporally, that occur before C

The percentage of comments elapsed value for a debater at any time quantifies their percentage of total experience that they've accumulated by that time. For a debater D, with temporally ordered comments $C_0, C_1, C_2...C_n$, we calculate the percentage of comments elapsed value for comment Ci at index i as:

$$percentage\_comments\_elapsed(Ci, D) = \frac{i}{|C_0, C_1...C_n|} \times 100 \qquad (3.1)$$

For a debater, the percentage of comments elapsed value for their first comment will be 0 (implying that they havenât gained any experience at that

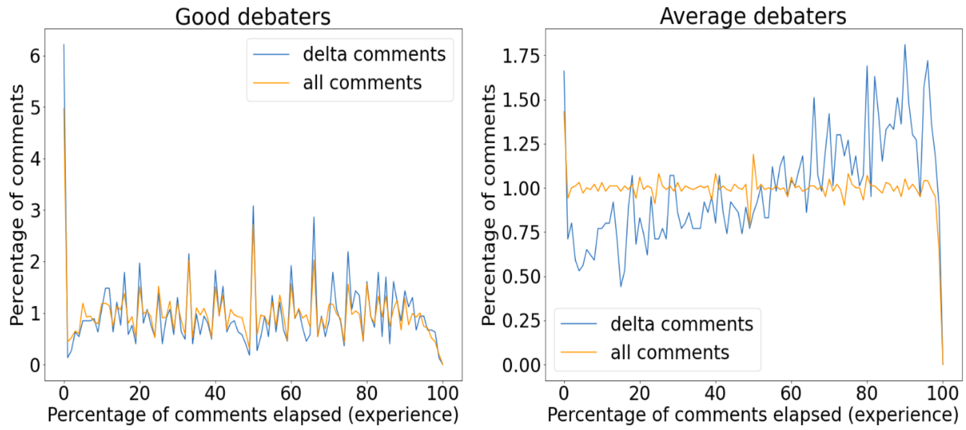

Figure 3.9: Evolution of debaters' level of activity and their rate of success with experience

time) and that for their last comment will be 100 (implying that they have gained all of their total experience by then).

Evolution of Debater's Activity and Success Rates with Experience
In this section, we analyze the evolution of debaters' activity levels as well as their rates of success as they accumulate experience in persuasion on CMV. We assess a debater's activity level at a time t by the percentage of their total comments that are made at that time. Their success rate at a time is measured as the percentage of their total number of deltas that are awarded at the time. We only consider good and average debaters for this analysis, discarding poor debaters since they don't receive any delta during their activity duration on CMV.

Our observations from figure 3.9 indicate that good debaters don't show a clear evolution in their activity levels and success rates as they gain experience. Average debaters however exhibit a gradual increase in their success rates with experience, while maintaining consistent activity levels. Hence, persuasion is a skill that can be acquired and improved upon with experience. This increased success in persuasion by average debaters motivates the need to detect changes in their persuasion strategies which lead to this improvement. We analyze their comments' lexical, syntactical, semantic and pragmatic attributes in the upcoming sections to find indicators of this improvement in persuasion over time.

Evolution of Amount of Experience Between Consecutive Successes
Here, we define the amount of experience gained by a debater between two comments. For a debater D with comments $C1$ and $C2$ made at times $t1$ and

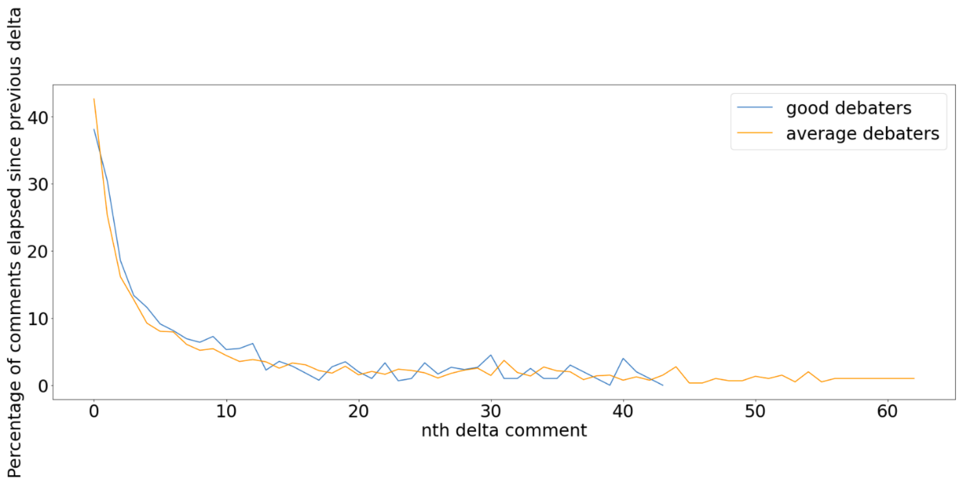

Figure 3.10: Evolution of amount of experience between successive deltas for CMV debaters

$t2$ respectively, the amount of experience gained between the two comments is the difference between the percentage of comments elapsed values for the two times. Similarly, if debater D is awarded their n^{th} delta for comment C_i at time t_i and their $(n+1)^{th}$ delta at time tj, then the experience gained by them between the two delta comments can be calculated as:
$percentage\_comments\_elapsed(t_j) - percentage\_comments\_elapsed(t_i)$

The experience gained by a debater between their two consecutive deltas can be seen as the amount of experience required by them to achieve their $(n+1)^{th}$ delta having already achieved n deltas.

Figure 3.10 shows the change in the 'experience gained between two successive delta comments' for good and average debaters as they achieve more number of deltas. We observe that the amount of experience between two consecutive deltas declines as the debaters achieve more deltas. This implies that a higher number of past successes makes it easier for a debater to achieve more successes in the future. Additionally, once a debater achieves a threshold number of successes, achieving further successes becomes significantly easier. This is illustrated by a sharp negative peak around the 5$^{\text{th}}$ delta mark on the X-axis. Hence, experience in persuasion compounds over time for both good and average debaters.

Evolution of Debaters' Comments' Lengths with Experience In previous sections, we have discussed the correlation between the length of a debater's comments and their success in persuasion. Additionally, we have ob-

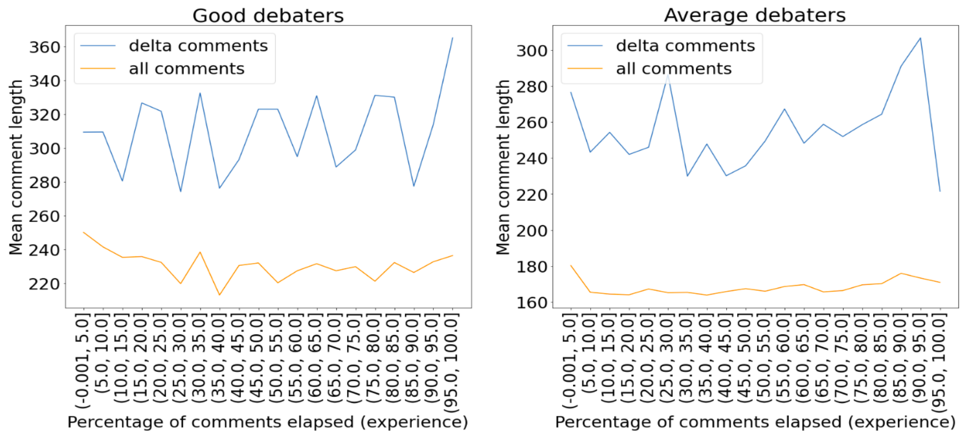

Figure 3.11: Evolution of debaters' comment lengths over their experience in persuasion

served that average debaters improve their effectiveness in persuasion and obtain better success as they gain experience. In this section, we analyze whether that improvement in persuasion effectiveness also brings an increment in their comments' lengths.

We don't observe a significant increment in average debaters' comments' lengths with experience in figure 3.11. However, for both good and average debaters, the mean comment length of their delta comments is significantly higher than the overall mean for all their comment lengths on CMV. It appears that, with respect to the length of their comments, all debaters deviate significantly from their usual behaviour in order to be persuasive. There could be two reasons behind this behaviour:

1. The debaters intentionally write longer comments when they are trying to be persuasive. This possibility can be discarded since we assume that every debater is trying to be persuasive when posting a top-level comment on a CMV discussion. Hence, all comments in our dataset, regardless of their lengths, are intended to be persuasive by their respective debaters.

2. Whenever a debater is persuasive, their comments tend to be longer than what they otherwise are. This motivates the need to investigate what in a debaterâs persuasive comments leads to them being longer than their non-persuasive comments, Possible reasons could be re-emphasizing key points through repetition (elaboration) or including more information content (informativeness).

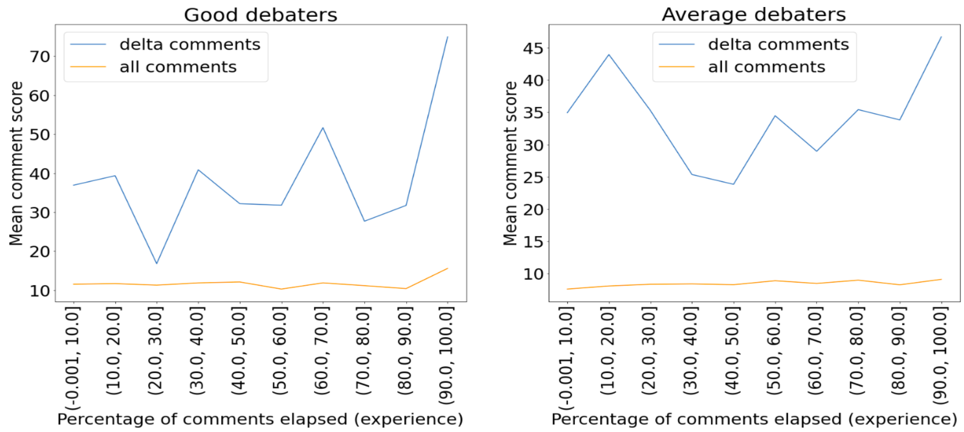

Figure 3.12: Evolution of debaters' comment scores over their experience in persuasion

Evolution of Debaters' Comments' Scores with Experience In our background on persuasion in CMV, we briefly touch upon two metrics of gauging a comment's persuasiveness. While the delta is a single user's evaluation of the comment's persuasiveness, the overall score of the comment is the audience's perception of the same. We study the evolution of the relationship between these two metrics of persuasiveness as the debater gains experience in persuasion. We compute the mean comment score for debaters' all comments as well as just their delta comments during the course of their activity periods.

While we don't observe a clear trend as to how the comments' scores evolve w.r.t. the debater's experience in persuasion in figure 3.12, we find the delta comments' mean score to be much higher than the overall mean for all comments. Hence, there is an overall agreement in the OP's evaluation of the comments' persuasiveness and the audience's. This correlation between the comments' score and their likelihood of being awarded a delta raises the need to also investigate their causation.

3.3 Analysis of Debaters' Comments' Content

Using our definitions of debaters' success/effectiveness in persuasion as well as their experience, we now move towards a deeper analysis of the text content of their comments. In this part of the thesis, we analyze the debaters' comments along four dimensions - syntactical, semantic, lexical and pragmatic; looking for indicators of successful persuasion as well as evolution w.r.t. experience.

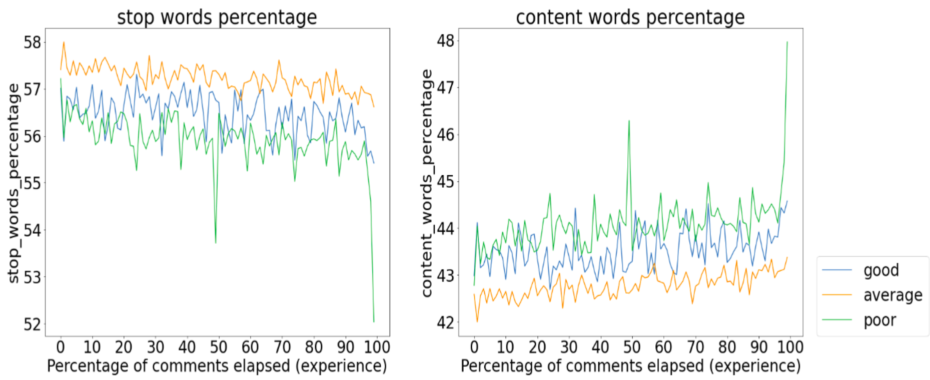

Figure 3.13: Percentage distribution of debaters' comments stop and content words over their experience in persuasion

3.3.1 Lexical Analysis of Debaters' Comments

Having established the role of comments' length in indicating a debater's effectiveness in persuasion, we now look at more fine-grained linguistic attributes of the comments' text to find indicators of effectiveness in persuasion as well as improvement in persuasion effectiveness with experience.

In this section, we focus on the lexical attributes based on the two constituent word groups for all text - stop and content words. We look at the relative distribution of both the word-groups as well as the size of the vocabulary and their impact on the debater's persuasiveness through the following research questions:

1. How does the relative distribution of stop and content words in a debater's comments correlate to their effectiveness in persuasion? Additionally, how does this distribution evolve as the debater gains experience in persuasion?

2. What role does the size of the debater's vocabulary play in their persuasion effectiveness? How does the debater's vocabulary evolve with experience in persuasion?

3. How do the lexical attributes considered here depend on the length of the debater's comments?

Role of Stop and Content Words in Debater's Effectiveness in Persuasion We compute the percentage of stop and content words in each debater's comments during the course of their activity duration on CMV. We then average these percentage values over the 'percentage of comments elapsed' values for each of our three groups of debaters (good, average, poor). Note that we use 'percentage of comments elapsed' as an indicator of the debater's experience in persuasion. Hence, we obtain a plot that shows the evolution of the relative distribution of stop and content words in the debaters' comments as they gain experience in persuasion.

Figure 3.13 shows no significant difference between the relative distribution of stop and content words for effective and ineffective debaters. Additionally, we don't observe a clear evolution of these distributions w.r.t. the debaters' experience in persuasion. It appears that the relative distribution of stop and content words has no impact on a debater's effectiveness in persuasion. Additionally, both effective and ineffective debaters maintain a consistent ratio between the stop and content words as they accumulate experience in persuasion.

Role of Vocabulary Size in Debater's Effectiveness in Persuasion For a debater, we consider the type-token percentage for the content words in their comments as an indicator of the size of their vocabulary relative to the overall size of their comments. A type-token percentage value of a hundred percent indicates no repetition of content words and the lower the percentage value, the higher is the repetition of content words in the debater's comments.

As discussed previously, figure 3.14 shows the length (number of tokens) of a debater's comments to correlate directly with their effectiveness in persuasion. Content type-token percentage correlates inversely with debater's effectiveness in persuasion since the values decrease in the order - poor > average > good. Hence, effective debaters tend to have smaller vocabularies relative to the overall size of their comments and repeat their content words more frequently. While this would imply that effective debaters tend to re-emphasize their arguments (potentially leading to higher word repetition), we suspect that this effect is simply because more persuasive debaters tend to write longer comments which increases the likelihood of them repeating the same words (especially since a debater's comment on a CMV discussion is centered around a specific topic). We elaborate on this in section 3.3.1.

Relationship between Comment Length and Lexical Features We group debaters' comments by their lengths (measured as the number of tokens) and compute the mean values for our lexical features at each length. Figure 3.15 shows that beyond a threshold value of length, comments of all lengths have a consistent ratio between the stop and content words implying that the distribution of the two fundamental token types in the text is independent of the text size.

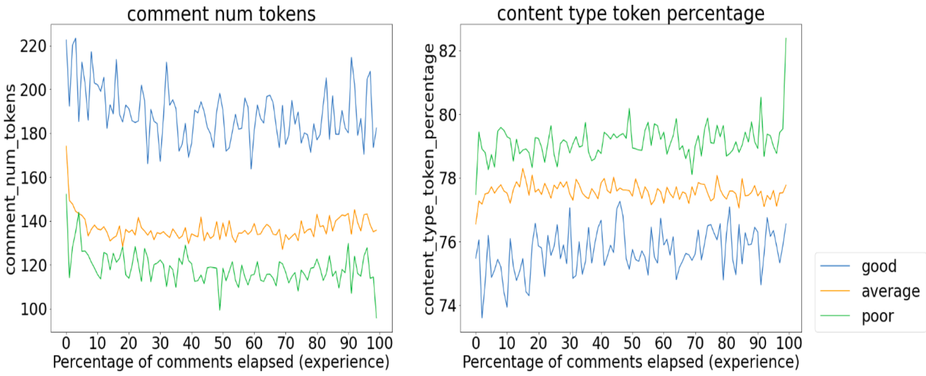

Figure 3.14: Number of tokens and type-token percentage of debaters' comments over their experience in persuasion

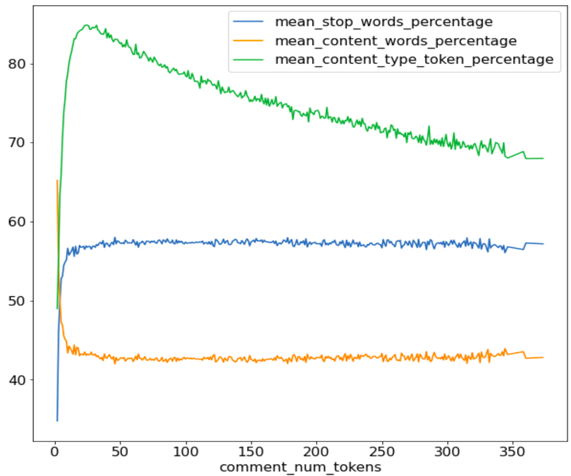

Figure 3.15: Debaters' comments' lexical features over the comment length(number of tokens)

The type-token percentage for content words gradually decreases with an increase in length. Hence, as the text gets longer, repetition of content words becomes more frequent. While we observe in the previous section that debaters effective in persuasion tend to repeat content words in their comments more frequently, the findings here indicate that this is primarily due to their comments being longer. Perhaps in an experimental setting where the debaters' comments could be perfectly normalized for their lengths, the content words' type-token percentage could also be indicative of effectiveness in persuasion.

3.3.2 Semantic Analysis of Debaters' Comments

Lexical analysis of debaters' comments suggests that while relative distribution of stop and content words remains consistent for all argumentative texts and has no impact on their persuasiveness, repetition of content words might indicate successful persuasion to some degree. Through further investigation of CMV debaters' comments semantic attributes, we attempt to address the following two research questions:

1. What influence does the semantic similarity between a debater's comments and their respective OPs have on their effectiveness in persuasion? Additionally, how does this semantic similarity between the comments and the OPs evolve as the debater gains experience in persuasion?

2. How can we quantify the amount of semantic information present in a debater's comments? What impact does it have on the debater's effectiveness in persuasion?

We use Word Mover's Distance (WMD) to measure the overall semantic difference between two texts. WMD allows for a semantic comparison between texts which don't have any words in common and has been shown to outperform previously used word similarity metrics like k-nearest neighbors (Kusner et al. [2015]). WMD computes the semantic differences between two texts by leveraging word vector representations of the constituent words in the texts and factoring in their token frequencies. We compute the following WMD based metrics for our semantic analysis of the debaters' persuasiveness:

- WMD between debatersâ comments and the corresponding discussionâs original post - This statistic is a measure of the semantic similarity between the OP and the debaterâs text in a CMV discussion. A lower value for this statistic corresponds to a higher semantic similarity between the OP and the debaterâs comment.

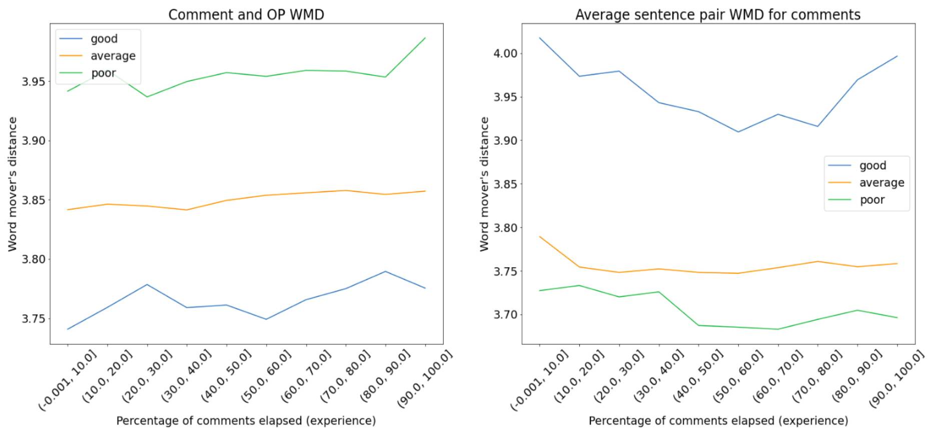

Figure 3.16: WMD based semantic attributes of comments over debaters' experience in persuasion

- Average WMD between all sentence pairs of a comment - For a comment C, containing n sentences $S1, S2...Sn$, this statistic is computed by dividing the total WMD between each pair of sentences in C and averaging over the total number of sentence pairs in C as:

$$avg\_sentence\_pair\_wmd(C) = \frac{\sum_{i=1}^{i=n-1} \sum_{j=i+1}^{j=n} wmd(Si, Sj)}{n(n-1)/2} \tag{3.2}$$

  Hence, we compute the average semantic diversity between the commentâs sentences which is indicative of its overall semantic content.

Note that WMD is an anti-similarity measure, hence a lower WMD value corresponds to higher semantic similarity between the texts and vice versa. We use the Gensim[2] library and fastText's word embeddings[3] which are pre-trained on the Common Crawl corpus for the analysis.

Role of Semantic Difference between Debater's Comments and OPs
Figure 3.16 shows that the semantic similarity between a debater's comments and the OPs is observed to be directly proportional to the debater's effectiveness in persuasion since the values of 'Comment and OP WMD' decrease in the order - poor $>$ average $>$ good. Hence, debaters effective in persuasion tend to capture the semantics of their OPs more closely.

[2]https://radimrehurek.com/gensim/index.html
[3]https://fasttext.cc/docs/en/english-vectors.html

We note that while the absolute value differences between our three groups is marginal, WMD based metrics are often evaluated in relative terms and the presence of a clear differentiating trend between good, average and poor debaters highlights the usefulness of semantics in the context of understanding persuasion.

Role of Debater's Comments' Semantic Information Content Figure 3.16 shows that the overall semantic diversity in a debater's comments directly correlates to their effectiveness in persuasion since the 'Average sentence pair WMD' values decrease in the order - good > average> poor. Hence, persuasive debaters manage to include a higher amount of semantic information content in their comments leading to a higher overall sentence pair semantic difference value.

The observations on the semantic attributes of CMV debaters' comments reveal that including a higher information content in the arguments while staying semantically closer to the OP is an effective persuasive strategy, characteristic of highly effectiveness debaters.

3.3.3 Syntactical Analysis of Debaters' Comments

Having observed the role of semantics in effective persuasion, we now move to a closely related class of features which deal with CMV debaters' comments' syntactical attributes. We attempt to find structural aspects of text which are indicative of effective persuasion through the following research question:

- What syntactical properties of CMV debaters' comments are characteristic of effective persuasion? In what ways do effective debaters structure their arguments differently than ineffective ones?

We consider the comments' text complexity as well as the distribution of different parts of speech tag to asses the role of syntactical properties towards persuasion.

Text Complexity and Effectiveness in Persuasion We assess a CMV debater text's complexity through several linguistic and stylistic complexity measures. We rely on the text complexity computation library by Proisl et al.[4] to obtain the required complexity measures. Since the library requires a larger volume of text to compute some of the metrics, we obtain the overall value for each of these features per debater instead of computing them separately for each comment. We do so by concatenating all top-level comments by a

[4]https://github.com/tsproisl/textcomplexity

Table 3.1: Text complexity features with highest correlations with debaters' delta comment percentage

Text Complexity Feature	**Correlation**
Outdegree Centrality	-0.17
Dependents per Word	0.17
Closeness Centrality	-0.16

CMV debater into a singular unit of text representative of the debater's entire argumentative text on CMV. Having obtained the text complexity feature values for each CMV debater, we compute the Pearson statistical correlation for each feature with the debater's effectiveness in persuasion (delta comment percentage). Table 3.1 lists the features with the highest absolute value of the Pearson correlation coefficient.

We observe that dependency parse-based complexity metrics listed in table 3.1 have the highest influence on CMV debaters' effectiveness in persuasion. These dependency parse-based features are computed by first building a dependency parse graph for the text through the recursive application of constituency grammar rules. Successful application of this step gives us a graphical representation of the debater's text where each directed edge represents a dependency relationship between words and/or phrases in the text. We then use this graphical representation to evaluate the structural complexity of the debater's text using different graph-based features. We briefly describe the most important graph-based complexity metrics concerning persuasion below:

- Outdegree centrality - For a node in a directed graph, outdegree centrality is the number of outgoing nodes adjacent to it. Hence, the outdegree centrality value is a measure of how central/influential a node is to the graph. It is a commonly used feature in the study of social networks to detect influential users in the nextwork. For a graph as a whole, outdegree centrality is calculated as the sum of maximum outdegrees of the graph minus the outdegrees of the remaining nodes divided by the largest possible sum of the maximum outdegrees of a similar-sized graph minus the outdegrees of the remaining nodes. Complicated definitions aside, we intuitively define outdegree centrality as a measure of how centralized a graph's structure is. A star graph with one central node and multiple other dependent nodes has an outdegree centrality value of 1. Hence, we can also view outdegree centrality as a measure of how closely a graph's structure matches that of a star graph.

  A negative correlation between the debater's text's outdegree centrality value and their effectiveness in persuasion signifies that persuasive debaters tend to follow sentence structures that are less centralized than those of their ineffective counterparts. Hence, syntactical arrangements where dependencies between the constituents are more evenly distributed and less centralized lead to higher overall effectiveness in persuasion.

- Number of dependents per word - The dependency graph uses directed edges to represent the dependencies between the constituent grammatical units of the text. The number of dependents per word is simply the average number of outgoing edges from each node in the dependency graph. A higher value of this metric implies a higher degree of inter-dependencies between the words in the text. A positive correlation with effectiveness in persuasion implies that persuasive debaters tend to incorporate textual structures in their text with higher inter-dependencies as compared to ineffective debaters.

- Closeness centrality - For a pair of nodes in a graph, closeness centrality is the reciprocal of the shortest distance between them in the graph. We consider the number of edges in the path between two nodes as their distance. For dependency graphs, we usually consider the average shortest distance between the root and all the other nodes of the graph. This is a measure of the graph's average dependency depth. Closeness centrality value is 1 for a graph with two nodes - a root and a dependent node, and decreases as the average shortest distance between the root and the other nodes increases. Outdegree and closeness centrality go hand in hand since both are graph centrality metrics, one concerning the number of adjacent nodes and the other with respect to the average shortest distance.

  A negative correlation between the debater's text's closeness centrality value and their effectiveness in persuasion signifies that persuasive debaters tend to follow sentence structures where the dependencies between the constituent units of the text are spread out across multiple levels instead of being centralized around the root node. Such sentence structures are less embedded and more evenly distributed, leading to deeper dependency graphs.

Parts of Speech Tags and Effectiveness in Persuasion Parts of speech (POS) are the fundamental units upon which a language's grammar is built. POS tagging is the process of assigning a unique parts of speech label to each token in the text based on its grammatical role. POS tagging takes into account the token as well as the context it is used in a sentence. POS tags are

Table 3.2: POS n-grams with highest positive correlations with debaters' delta comment percentage
IN - preposition, JJ - adjective, NN - noun, VBG - verb, present participle/gerund, DT - determiner

POS n-gram	Correlation with Delta Comment Percentage	Correlation with Comment Length
IN JJ	0.11	0.05
NN IN JJ	0.10	0.03
JJ NN IN	0.09	0.07
VBG DT JJ	0.08	0.02

Table 3.3: POS n-grams with highest negative correlations with debaters' delta comment percentage
PRP - possessive pronoun, VPB - verb present tense, WRB - Wh adverb

POS n-gram	Correlation with Delta Comment Percentage	Correlation with Comment Length
PRP VBP	-0.13	-0.05
PRP	-0.12	-0.10
WRB VBP	-0.11	-0.13
NN WRB	-0.11	-0.05

widely used for a variety of NLP tasks due to their simplicity in computation through the use of automated tagging libraries. In the context of persuasion in CMV, Tan et al. [2016] achieved a modest baseline for their persuasiveness prediction experiments using POS based n-gram features derived from the debater's comments. For our analysis of CMV debaters' persuasiveness, we too focus on the distribution of different POS tags and their evolution with the debaters' experience in persuasion.

For our analysis of the use of POS tags by CMV debaters, we annotate each debater's comments using the NLTK library. We rely on the Universal Tagset for our annotations. At the end of this step, we obtain a string representing the POS tags for each comment in our dataset. We use these 'POS strings' to further obtain TF-IDF distributions for unigrams, bigrams and trigrams of the POS tags. We limit our analysis to the 1000 most frequently occurring POS based n-gram features due to memory constraints.

For each POS n-gram, we compute the Pearson correlation between its mean TF-IDF value for the CMV debaters and their respective delta comment percentage values. To account for the influence of comment length on the distribution of POS tags, we also compute the correlation of the TF-IDF

Table 3.4: POS n-grams with highest absolute difference in correlation values with debaters' experience in persuasion between good and average debaters
RB - adverb, VBZ - verb present tense, third person singular

POS n-gram	Correlation with Good Debaters' Experience in Persuasion	Correlation with Average Debaters' Experience in Persuasion
VBG IN JJ	-0.26	0.36
VBZ RB	-0.07	0.53
VBZ RB DT	-0.22	0.36
JJ IN PRP	0.11	-0.40
VBZ	0.10	0.58

values for the POS n-grams and the comment length. Tables 3.2 and 3.3 show the POS n-grams with the highest positive and negative correlation coefficient values with CMV debaters' delta comment percentage. We do not observe significant positive or negative correlations between the distribution of POS n-grams and debaters' effectiveness in persuasion. Nouns (NN), prepositions (IN) and adjectives (JJ) occur most frequently for effective debaters since they have the highest positive correlations. Possessive pronouns (PRP), verbs in present tense (VPB) and Wh-adverbs (WRB) appear to be used more by ineffective debaters. It is important to note that the comment length also influences the distribution of the POS tags since the POS n-grams with positive correlations with delta comment percentage also correlate positively with comments' lengths and vice versa.

Next, we investigate the changes in the usage of POS tags for CMV debaters as they accumulate experience in persuasion and how this evolution of the POS tags' usage is different for good and average debaters. We compute the Pearson correlations for each POS n-gram with the debaters' experience in persuasion. Recall that we represent this experience for a CMV debater at any time as the percentage of their total comments elapsed at that time. The POS n-grams with positive correlations show an increase in their usage while those with negative correlations show a decline in their usage w.r.t. debaters' experience in persuasion. Table 3.4 lists the POS n-grams which show the highest change in their usage over experience in persuasion between good and average debaters. Verbs (VBG, VBZ) used in combination with other POS tags like nouns (NN), prepositions (IN), and adverbs (RB) show a decline in their usage by good debaters but an increase in the same by average debaters. Adjectives (JJ) combined with possessive pronouns (PRP) using prepositions(IN) show a slight increase in their occurrence for good debaters but a sharp decline for average debaters. Common examples of such phrases include 'many of them' and 'good for you'. Verbs in the present tense used for the third person show an increase for both good and average debaters.

3.3.4 Pragmatic Analysis of Debaters' Comments

Pragmatics is a sub-field of linguistics which deals with the understanding of implications of speech and text by analyzing factors like situational context, mental states of the parties involved and previous history of communication. For our analysis of CMV debaters' persuasiveness, we look at the distribution of various argumentation and framing strategies in their comments and try to assess their impact on the debaters' effectiveness in persuasion.

Argumentative Features We briefly introduced the argumentative features used in previous work on annotating CMV discussions and using them to understand persuasiveness in CMV. In this section, we look at the distribution of these argumentative features in CMV debaters' comments and attempt to find argumentation strategies which indicate effective persuasion of the OP. We consider two classes of argumentative features for our experiments:

- Elementary units(EUs) - Previous works have established EUs to be relevant for argument mining and retrieval. Khatib et al. [2016] proposed an Argumentative Discourse Unit based approach to model persuasiveness in news editorials. For our experiments, we consider the study of EUs in CMV discussions byEgawa et al. [2019]. They annotated CMV discussions with the following EU semantic types at the token level:
  - Testimony - An objective proposition based on the debater's personal experience. For example - I have studied science in school.
  - Fact - Objective fact that can be verified via existing evidence. For example - Germany is part of the European Union.
  - Value - Subjective evaluation/judgement without specifying a course of action. For example - Winter in Germany is quite harsh.
  - Policy - Specification of a course of action to be performed. For example - One should supplement Vitamin D during winters.
  - Rhetorical Statement - Statement where a subjective judgement is not stated directly but implied. For example - Sure, it makes sense to not wear a mask and just get Covid, right?
- Claims - They are used to expresses the debater's stance on a topic. We use the two-tier annotation scheme by Hidey et al. [2017] using which they annotated CMV discussions with following claim semantic types:

  - Interpretation - They express prediction or explanations. For example - I think the Greens will win this election.
  - Evaluation - They express positive or negative judgements. Evaluations are further classified into rational(For example - He is a fast runner) or emotional(For example - Waking up early in the morning is hard).
  - Agreement/Disagreement - They are used to concur or differ with someone's opinion. For example - I agree that this is not easy, I don't think this is easy.

- Premises - They are used to justify/validate a claim. Hidey et al. [2017] considered the following premise semantic types:
  - Ethos - Appeal to the credibility established by reputation or expertise. For example - As a doctor, I can agree that wearing masks helps to stop infection from spreading.
  - Logos - Appeal to reason through examples or evidence. For example - Covid will be over by next year as many people are getting vaccinated.
  - Pathos - Appeal to emotion by connecting with the audience over shared experiences. For example - We should take precautions against Covid, it is a dangerous disease to contract.

For annotating our CMV debaters' comments with Elementary Units and their semantic types, we consider Egawa et al. [2019]'s dataset with token-level annotations of CMV discussions. We resort to using coarser sentence-level annotations of the debaters' comments for our analysis of persuasion by discarding sentences with multiple EUs and considering the EU semantic type as the label for that sentence. We use this dataset with sentence-level annotations to train a BERT classifier to label a sentence with an EU semantic type. The train-test split for this classification was done such that comments and OPs from the same CMV discussion are included in the same split. In order to account for the skewed class distribution between sentence-level EU semantic type annotations in Egawa et al's dataset, we performed random oversampling on the training set thus obtaining the best macro and micro accuracy scores of 0.55 and 0.75 respectively. We then retrained the BERT classifier on the entire Egawa et al's dataset (train + test) and used that classifier to annotate each comment's sentences in our CMV debaters' dataset. Using the sentence level EU annotations, we compute unigram, bigram, and trigram TF-IDF distributions for the comments in our CMV debaters' dataset.

We obtain annotations for Claims and Premises and their semantic types using the dataset by Hidey et al. [2017] similarly. We train five BERT classifiers for different annotations as follows:

1. Claim premise classifier - Given a sentence, label it as claim/premise/none
2. Claim semantic type classifier - Given a sentence labeled as a claim, further label it with one of the claim semantic types - agreement, disagreement, evaluation emotional, evaluation ration, interpretation
3. Premise ethos classifier - Given a sentence labeled as a premise, further classify it as ethos or not ethos
4. Premise logos classifier - Given a sentence labeled as a premise, further classify it as logos or not logos
5. Premise pathos classifier - Given a sentence labeled as a premise, further classify it as pathos or not pathos

Note that while a claim can only have a single semantic type, a premise can have one or more semantic types. Hence, we train three binary classifiers for each premise semantic type separately. We use sentence-level annotations from each of the classifiers to build the corresponding unigram, bigram, and trigram-based TF-IDF features for the CMV debaters' comments. Once we have the TF-IDF value for the argumentative units' n-grams for all CMV debaters, we compute their Pearson correlation with the debaters' delta comment percentage values to find the argumentative units which correlate with debaters' effectiveness in persuasion.

Table 3.5 lists the argumentative n-grams with the highest absolute value of the Pearson correlation coefficient. Most argumentative n-grams do not appear to correlate significantly with CMV debaters' effectiveness in persuasion and we observe only three EU semantic type n-grams having correlation values over 0.10. Based on the relatively higher correlations of EU n-grams based on rhetorical statement and value, we conclude that the use of isolated rhetorics and subjective statements, as well as their combination, leads to slightly higher debater effectiveness on CMV.

Frames Framing is a persuasion strategy where arguments are formulated in a way such that some aspects of the overall theme are highlighted more than others, in an attempt to manipulate the discourse of the debate. Framing can be seen as the process of directing an argument towards certain directions and away from certain other directions. A frame is simply a dimension where a debater can potentially direct a discussion towards. In this section, we study

Table 3.5: Argumentative features with largest absolute values of Pearson correlation coefficient with CMV debaters' effectiveness in persuasion

Argumentative n-gram	Correlation with Debaters' Effectiveness in Persuasion
(EU) rhetorical_statement	0.19
(EU) value	0.13
(EU) rhetorical_statement value	0.11

the use of frames by CMV debaters and the influence of their framing strategies on their persuasion effectiveness in CMV.

Our broad research question for this section is as follows:

- How does the choice of frames in CMV debaters' comments influence their effectiveness in persuasion?

To annotate our CMV debaters' comments with the different frames, we use the Media Frames corpus by Card et al. [2015]. This Media Frames corpus consists of news articles on three broad themes manually annotated with fifteen frames labels. The frame labels used along with their brief descriptions are listed in figure 3.17. We use this corpus to train a BERT model for the multi-label classification task of predicting the constituent frames for a CMV debater's comment. The best performing model achieved macro and micro accuracy scores of 0.51 and 0.68 respectively. We re-train this model on the entire Media Frames corpus and then annotate each comment in our CMV debaters dataset using it.

Framing and CMV Debaters' Effectiveness in Persuasion We begin our study of the influence of a debater's choice of frames on their effectiveness in persuasion by computing the average number of frames in the debaters' comments. We do not observe any significant correlation between the average number of frames in a debater's comments and their effectiveness in persuasion. Hence, the number of frames included in a comment has little impact on its persuasiveness. We do observe a slight correlation of 0.14 between the average number of frames in a debater's comments and their average comment length since including additional frames in a comment would require adding more text, making the comment longer. We also observe that the average number of frames per comment values increase for debater's effectiveness in persuasion in the order - poor debaters $<$ average debaters $<$ good debaters. Additionally, we observe that for all debaters the average number of frames value for their delta

Economic: costs, benefits, or other financial implications
Capacity and resources: availability of physical, human or financial resources, and capacity of current systems
Morality: religious or ethical implications
Fairness and equality: balance or distribution of rights, responsibilities, and resources
Legality, constitutionality and jurisprudence: rights, freedoms, and authority of individuals, corporations, and government
Policy prescription and evaluation: discussion of specific policies aimed at addressing problems
Crime and punishment: effectiveness and implications of laws and their enforcement
Security and defense: threats to welfare of the individual, community, or nation
Health and safety: health care, sanitation, public safety
Quality of life: threats and opportunities for the individual's wealth, happiness, and well-being
Cultural identity: traditions, customs, or values of a social group in relation to a policy issue
Public opinion: attitudes and opinions of the general public, including polling and demographics
Political: considerations related to politics and politicians, including lobbying, elections, and attempts to sway voters
External regulation and reputation: international reputation or foreign policy of the U.S.
Other: any coherent group of frames not covered by the above categories

Figure 3.17: Frame types included in the Media Frames corpus

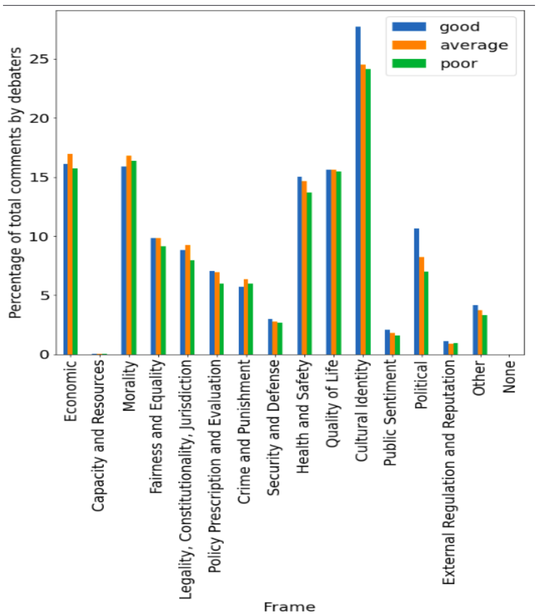

Figure 3.18: Distribution of frame types in CMV debaters' comments

comments is slightly higher than that for their non-delta comments. However, these difference in the average number of frames values is not significant enough to conclude.

Next, we analyze the distribution of the debaters' comments between the fifteen frame classes. Fig 3.18 shows that most frame types are distributed roughly equally between different debater types. Hence, all three debater types (good, average, poor) use most of the frames to a similar degree. However, we observe a slightly higher representation of good debaters (represented by the blue bar) for the frame types 'Cultural Identity' and 'Political'. Cultural Identity frame type appeals to the traditions/customs and social values whereas Political frame type appeals to the political ideology of the audience. Both of these are highly polarizing themes and the observations suggest that debaters who can resonate with the socio-cultural and/or political inclinations of their respective OPs are more effective in persuading them to change their minds.

Chapter 4

Modelling CMV Debaters' Persuasiveness

Inspired by the extensive previous work on modeling CMV comments' persuasiveness using a variety of lexical, syntactical, pragmatic, and users' interaction dynamics-based features; we propose the new task of identifying 'persuasive debaters' based on a set of features derived from the text content of their comments on CMV. In this regard, we perform a comparison of various feature types that we have explored in this thesis - lexical, syntactical, semantic, and pragmatic. Additionally, we also consider the vocabulary interplay based features proposed by Tan et al. [2016] for comparison. Our broad research question for this chapter is:

- Previous works have achieved success in modeling persuasive comments in CMV. How effective are the features used in those tasks in modeling debaters' persuasion effectiveness in CMV?

4.1 Prediction Task

We perform the classification task of modeling CMV debaters' effectiveness in persuasion from all of their top-level argumentative comments on CMV discussions. Recall that we operationalize effectiveness in persuasion through the debaters' delta comment percentage values. We divide debaters into groups of good, average, and poor based on these delta comment percentage values. Debaters with a value of 5 percent or above are considered to be highly effective in persuasion. We formally define our classification task as:

Given a debater D with comments $C_1, C_2 ... C_n$ predict whether this debater is highly effective in persuasion or not.

4.2 Prediction Features

Based on our review of past works in persuasiveness prediction in CMV, we group the features used for the task into four broad categories:

- Surface level text features - These features are extracted from the text of the debater's comment and exploit the corresponding text's linguistic (lexical, syntactic, semantic) properties to model its persuasiveness.
- Interaction-based features - These features examine the nature of the interactions of the debater with the OP and other debaters/users in the CMV discussion.
- Pragmatic features - These features also rely on the text content of the debater's comment but explore its higher-level pragmatic properties.
- User-based features - These features exploit user-level attributes such as past activity, credibility(prior evidence of success) as well as compatibility with the OP(such as through shared beliefs) to estimate the likelihood of successful persuasion.

For our experiments in persuasiveness modeling of CMV debaters, we evaluate the effectiveness of five feature classes - lexical, syntactical, semantic, pragmatic, and vocabulary interplay on the CMV debaters dataset. Linguistic features based on the lexical and/or syntactical properties of CMV comments have been used extensively in previous works and provide a modest baseline for persuasiveness prediction. Vocabulary interplay-based features proposed by Tan et al. [2016], while being quite simple to compute, have been surprisingly effective in their experiments. Pragmatic features based on the distribution of various argumentative units and frame types have been relatively unexplored for persuasiveness prediction and, to the best of our knowledge, have not been previously used to model debaters' persuasiveness in CMV. Hence, a comparison of their effectiveness against features already proven to be successful in the task could aid in understanding their role in modeling CMV debaters' persuasiveness.

A brief recap of the features used for each of the five feature classes is as follows:

1. Semantic features - These features capture the semantic similarity between the debater's comment and the respective OP as well as the semantic content of the debater's comment. We use Word-Mover's Distance based semantic difference metrics discussed previously in chapter 3 as features for classification.

2. Syntactical features - These features capture the grammatical structure of the debater's comment. We consider various text complexity metrics and n-grams based on part-of-speech tags as features for classification.

3. Pragmatic features - These features capture the audience's/OP's interpretation of the debater's arguments by analyzing relevant contextual information. We consider the two types of argumentative features(claims, premises, and EUs) and frame types as features for classification.

4. Lexical features - These features capture the writing style of the debater and operate under the assumption that the debater's writing style indicated by the presence of certain word groups in the text correlates with its persuasiveness. We consider the following types of linguistic features for our experiments:

    - Word category-based features - These features look for the presence of certain word groups in the text. Examples of such features include hedges, boosters, first-person pronouns, negative and positive words. We use an assortment of freely available corpora for each word category to compute these features for our dataset. For each of these features, we compute their absolute counts as well as their proportions in the text.
    - Text attributes - These include commonly used text-based features like the number of sentences and paragraphs and readability metrics like Flesch Kincaid.

    The list of features used for our experiments is given in table 4.1. We admit some overlap with syntactical features. However, lexical and syntactical attributes of text often go hand in hand and a clear separation between the two might not be entirely possible. Hence, we choose to proceed with this division of the feature categories for our experiments.

5. Vocabulary interplay features - Vocabulary interplay features are based on the shared vocabulary of the debater and OP. Tan et al. [2016] hypothesize that the interplay between the vocabulary of the two parties is indicative of the OP's successful persuasion. For a debater's comment with vocabulary C and its corresponding OP with vocabulary O, they considered the following features:

    - Number of common words between the two - Computed as: $|C \cap O|$
    - Debater fraction - Proportion of the debater's vocabulary that they share with the OP. Computed as: $\frac{|C \cap O|}{|C|}$

- OP fraction - Proportion of OP's vocabulary that they share with the debater. Computed as: $\frac{|C \cap O|}{|O|}$
- Jaccard similarity - Computed as: $\frac{|C \cap O|}{|C \cup O|}$

These features are computed for three sets of vocabularies - stop words, content words, all words combined. The authors found these vocabulary interplay features to be highly effective towards indicating the OP's persuasion. Additionally, their findings revealed that persuasive comments on CMV have a higher proportion of common stop words with the OP, but a lower proportion of common content words than non-persuasive comments. The usage of stop words is an indicator of the author's writing style and content words of the comment's information content. Hence, persuasive comments in CMV tend to match their corresponding OPs in terms of their writing style(through a similar use of stop words) while adding information through the inclusion of new content words.

Apart from the above-mentioned feature classes, we also consider a bag of words as a baseline for our comparison of the features for debaters' effectiveness prediction.

4.3 Experiments

For each of the 3801 debaters in our CMV debaters dataset, we compute the values for each feature type for each comment. The values for argumentative, semantic, and syntactical features are computed similar to how they were obtained in our analysis of CMV debaters in chapter 3. For frames, we compute the absolute and relative number of comments by a debater for each of the fifteen frame types as features for classification. Vocabulary interplay features are computed based on the formulas proposed by Tan et al. [2016] and lexical features are computed through our implementations. We then obtain the feature vectors for each debater by aggregating the feature vectors for each of their comments as follows:

For a debater D with comments $C_1, C_2 ... C_n$ with feature vectors $V_1, V_2 ... V_n$, we obtain the overall feature vector V for the debater by computing the mean vector for $V_1, V_2 ... V_n$.

We consider the following experimental settings for our classification task:

1. Good vs (Poor + Average) - Predicting effective debaters from an evenly distributed dataset of good(effective), average(partially effective), and poor(ineffective) debaters. This dataset has a 1:2 ratio of positive(good

Table 4.1: Lexical features used in the experiments

Feature Name
words
definite articles
indefinite articles
positive words
2nd person pronoun
links
negative words
hedges
1st person pronoun
1st person plural pronoun
.com links
examples
question marks
PDF links
.edu links
quotations
arousal
valence
word entropy
sentences
type-token ratio
paragraphs
Flesch-Kincaid grade levels

debaters) and negative(average and poor debaters) samples and is therefore unbalanced.

2. Good vs average - Predicting effective debaters from an evenly distributed dataset of good(effective) and average(partially effective) debaters. This dataset is balanced between positive(good debaters) and negative(average debaters) samples.

3. Good vs poor - Predicting effective debaters from an evenly distributed dataset of good(effective) and poor(ineffective) debaters. This dataset is balanced between positive (good debaters) and negative (poor debaters) samples.

The motive behind the different experimental settings is to assess whether our features' effectiveness varies based on the distribution of the dataset. For instance, to get an idea of whether certain features are more effective in identifying effective debaters from a population of effective and partially effective debaters than from a population of effective and ineffective debaters or vice versa.

We perform an $80-20$ train-test split on the dataset for each of the experimental settings. We use the logistic regression classifier from sklearn for our experiments. We consider the classifier's macro accuracy as the primary metric to assess the features' effectiveness in the different experimental settings.

Table 4.2: Dataset distribution for the three experimental settings. Note that the CMV debaters dataset contains 1267 samples for each of the three debater types: good, average, poor

Experimental Setting	**# Positive Samples**	**# Negative Samples**
Good vs Poor	1267	1267
Good vs Average	1267	1267
Good vs (Poor + Average)	1267	2534

4.4 Results and Discussion

Table 4.3 lists the macro accuracy scores for the different features across the three experimental settings. Bag of words gives a strong baseline for modeling CMV debaters' effectiveness in persuasion and none of the feature classes apart from frames appear to surpass it. Bag of words also performs better at modeling CMV debaters' persuasiveness than at modeling that for CMV comments(based on the macro accuracy scores observed by Khatib et al. [2020] for their classification task for CMV comments' persuasiveness). We suspect two possible reasons behind this stronger baseline. First, smaller size for our CMV debaters dataset since the number of comments in CMV far exceeds the number of users/debaters. Second, more data per record for the classifier to find text n-grams indicative of persuasiveness since we combine all the top-level comments by a debater to generate the features for classification.

The general order of effectiveness for the various feature classes is as follows - Frames > Bag of Words > Vocabulary Interplay, Lexical > Argumentative, Semantic > Parts of Speech, Text Complexity.

A lackluster performance by argumentative features reaffirms the observations made by Egawa et al. [2019] that mere distribution of argumentative units in the text is insufficient in indicating its persuasiveness. N-gram features based on just the information about claims and premises without their semantic type labels perform worst out of all argumentative features.

Table 4.3: Classification macro accuracy scores for the three experimental settings

Feature Type	**Feature**	**Good vs Average + Poor**	**Good vs Average**	**Good vs Poor**
-	Bag of Words	0.64	0.60	0.68
-	Vocabulary Interplay	0.61	0.58	0.67
Lexical	Lexical	0.61	0.62	0.67
Pragmatic	Elementary Units	0.57	0.51	0.59
Pragmatic	Claim and Premise	0.52	0.47	0.55
Pragmatic	Claim Semantic Type	0.57	0.48	0.58
Pragmatic	Premise Semantic Type	0.56	0.48	0.58
Pragmatic	Claim and Premise with Semantic Types	0.56	0.48	0.58
Pragmatic	**Frames**	**0.74**	**0.70**	**0.72**
Semantic	Word Mover's Distance features	0.57	0.59	0.63
Syntactical	Parts of Speech	0.52	0.57	0.51
Syntactical	Text Complexity	0.53	0.65	0.61

The distribution of CMV debaters' comments between the fifteen frame types results in the best macro scores across the three experimental settings. The most useful features for determining a debater's persuasion are their number of comments for frame types 'Quality of Life', 'Morality', and 'Health and Safety'. Note that these features have negative weights hence ineffective debaters have a higher number of comments for the corresponding frame types.

For most of the features, the best macro score is observed in the 'Good vs Poor' experimental setting and the worse in the 'Good vs Average' experimental setting. Frames and text complexity features are the only exceptions to this trend. Hence, most features are better at separating effective debaters from ineffective ones than from partially effective ones. Hence, larger the difference between the debaters' effectiveness in persuasion, the easier it is to separate effective ones from their ineffective counterparts.

Chapter 5

Conclusion and Future Work

In this thesis, we have described an analytical approach towards understanding debater-level persuasion strategies in the online discussion forum Change My View (CMV). As part of the work done here, we have studied CMV debaters' activities on the forum as well as the text content of their comments along four dimensions: syntactical, lexical, semantic, and pragmatic. We also have extended previous work on modeling persuasion in online discussions by proposing a debater-level persuasion effectiveness modeling task and evaluating the performance of various features sets for that task.

In chapter 1, particularly, we discuss our motivation behind studying persuasion strategies of CMV debaters and present the following broad research questions for the thesis:

1. What aspects of user behavior and their text content on CMV separate effective debaters from ineffective ones? What insights about effective persuasion can be obtained by analyzing effective CMV debaters w.r.t. these attributes?

2. How effective are the lexical, syntactical, semantic, and pragmatic features obtained from CMV users' text content in modeling their effectiveness in persuasion?

We layout the essential background information about CMV and its workings in chapter 2. Additionally, we briefly describe the related past works grouped into the following two categories:

1. Modelling comment-level persuasiveness in CMV and similar online forums using a diverse set of features - text-based, user-based, user interactions based.

2. Annotating CMV discussions with argumentative units and their semantic types, and analyzing their impact on persuasiveness.

Chapter 3 covers our analysis of CMV debaters' persuasion strategies. We begin by curating a dataset of CMV debaters using the WebisCMV dataset as the basis. We then propose notions to quantify debaters' effectiveness in persuasion as well as their experience in persuasion accumulated over time. We use those notions to detect behaviors as well as attributes of debaters' text content that are indicative of effective persuasion.

In chapter 4, we describe the classification task aimed towards modeling CMV debaters' effectiveness in persuasion. We evaluate the effectiveness of syntactical, semantic, lexical, and pragmatic features proposed in chapter 3, for the prediction task. We also use vocabulary interplay based features proposed by Tan et al. [2016] for comparison. We find the features based on the distribution of framing strategies in debaters' comments to be the most effective. Argumentative features fail to surpass the bag of words baseline, indicating that mere distribution of argumentative units in the debaters' comments has limited utility in determining their effectiveness in persuasion.

Future Work Through the experiments and the analyses conducted in this thesis, we establish several useful persuasion strategies which could lead to an improvement in a debater's odds of success in persuasion. However, there remains scope for further analysis both in terms of depth and breadth. Some possible future extensions of this thesis are listed as follows:

- While our notion of persuasiveness in CMV is limited to a debater's comment being awarded a delta, or in some cases the overall score for the comment; a more holistic notion would be one which accounts for a degree of subjectivity in the evaluation of a comment's persuasiveness based on the evaluating user's idiosyncrasies. Guo et al. [2020] briefly touch upon this in their human study of persuasiveness where they find that despite a general agreement about what's persuasive, there are variations in evaluations of the persuasiveness of a comment based on the evaluating party's individual attitudes that could influence their judgment

- For both the analysis of debaters in CMV as well as modeling their effectiveness in persuasion, we limit the scope to features derived from the text content of their comments. A similar analysis of other classes of features such as those based on the debaters' activities on Reddit as well as their interaction dynamics with other users could reveal useful insights about their persuasiveness.

- For both the analysis as well as our modeling experiments, length acted as a confounding variable and had some influence on the findings. While previous works have noted the observation that longer text often corresponds to higher effectiveness in persuasion, further work needs to be done to quantify the influence of length in persuasion and come up with ways to mitigate its influence. This would provide insights into factors which govern persuasion in the absence of length's confounding influence.

- Our classification experiments for modeling persuasion use logistic regression as the model for the binary classification. A possible improvement in the performance could be achieved by using sophisticated classification models. Guo et al. [2020] propose a Conditional Random Fields(CRF) based model for modeling the cumulative effect of persuasion in CMV discussions. Li et al. [2020] use a bi-LSTM and BERT-based model for their persuasiveness modeling task.

- While argumentative features based on the distribution of argumentative units achieve lackluster performance on our modeling task, possible improvements could be achieved through the use of features that capture effective use of argumentative units. Possible ideas could include features based on the inter-dependencies between different argumentative units in the text(proposed by Li et al. [2020]) as well as the relative ordering between them(proposed by Hidey et al. [2017]).

Bibliography

David Atkinson, Kumar Bhargav Srinivasan, and Chenhao Tan. What gets echoed? understanding the "pointers" in explanations of persuasive arguments. In Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan, editors, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 2911–2921. Association for Computational Linguistics, 2019. URL `https://doi.org/10.18653/v1/D19-1289`.

Dallas Card, Amber E. Boydstun, Justin H. Gross, Philip Resnik, and Noah A. Smith. The media frames corpus: Annotations of frames across issues. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, July 2015.

Ryo Egawa, Gaku Morio, and Katsuhide Fujita. Annotating and analyzing semantic role of elementary units and relations in online persuasive arguments. In Fernando Alva-Manchego, Eunsol Choi, and Daniel Khashabi, editors, *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28 - August 2, 2019, Volume 2: Student Research Workshop*, pages 422–428. Association for Computational Linguistics, 2019. URL `https://doi.org/10.18653/v1/p19-2059`.

Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. Political discourse on social media: Echo chambers, gatekeepers, and the price of bipartisanship. *CoRR*, abs/1801.01665, 2018. URL `http://arxiv.org/abs/1801.01665`.

Zhen Guo, Zhe Zhang, and Munindar Singh. In opinion holdersâ shoes: Modeling cumulative influence for view change in online argumentation. In *Proceedings of The Web Conference 2020*. Association for Computing Machinery, 2020. URL `https://doi.org/10.1145/3366423.3380302`.

Christopher Hidey, Elena Musi, Alyssa Hwang, Smaranda Muresan, and Kathy McKeown. Analyzing the semantic types of claims and premises in an online persuasive forum. In Ivan Habernal, Iryna Gurevych, Kevin D. Ashley, Claire Cardie, Nancy Green, Diane J. Litman, Georgios Petasis, Chris Reed, Noam Slonim, and Vern R. Walker, editors, *Proceedings of the 4th Workshop on Argument Mining, ArgMining@EMNLP 2017, Copenhagen, Denmark, September 8, 2017*, pages 11–21. Association for Computational Linguistics, 2017. URL `https://doi.org/10.18653/v1/w17-5102`.

Khalid Al Khatib, Henning Wachsmuth, Johannes Kiesel, Matthias Hagen, and Benno Stein. A news editorial corpus for mining argumentation strategies. In Nicoletta Calzolari, Yuji Matsumoto, and Rashmi Prasad, editors, *COLING 2016, 26th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, December 11-16, 2016, Osaka, Japan*, pages 3433–3443. ACL, 2016. URL `https://aclanthology.org/C16-1324/`.

Khalid Al Khatib, Michael Völske, Shahbaz Syed, Nikolay Kolyada, and Benno Stein. Exploiting personal characteristics of debaters for predicting persuasiveness. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel R. Tetreault, editors, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 7067–7072. Association for Computational Linguistics, 2020. URL `https://doi.org/10.18653/v1/2020.acl-main.632`.

Matt J. Kusner, Yu Sun, Nicholas I. Kolkin, and Kilian Q. Weinberger. From word embeddings to document distances. JMLR.org, 2015.

Jialu Li, Esin Durmus, and Claire Cardie. Exploring the role of argument structure in online debate persuasion. *CoRR*, abs/2010.03538, 2020. URL `https://arxiv.org/abs/2010.03538`.

Humphrey Mensah, Lu Xiao, and Sucheta Soundarajan. Characterizing susceptible users on reddit's changemyview.

Chenhao Tan, Vlad Niculae, Cristian Danescu-Niculescu-Mizil, and Lillian Lee. Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions. In *Proceedings of WWW*, 2016.

Zhongyu Wei, Yang Liu, and Yi Li. Is this post persuasive? ranking argumentative comments in online forum. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 2: Short Papers*. The Association for Computer Linguistics, 2016. URL `https://doi.org/10.18653/v1/p16-2032`.