FrameAxis: Characterizing Framing Bias and Intensity with Word Embedding

Haewoon Kwak
Qatar Computing Research Institute
Hamad Bin Khalifa University
Doha, Qatar
haewoon@acm.org

Jisun An
Qatar Computing Research Institute
Hamad Bin Khalifa University
Doha, Qatar
jisun.an@acm.org

Yong-Yeol Ahn
Indiana University, Bloomington
IN, USA
yyahn@iu.edu

ABSTRACT

We propose FrameAxis, a method of characterizing the framing of a given text by identifying the most relevant semantic axes ("microframes") defined by antonym word pairs. In contrast to the traditional framing analysis, which has been constrained by a small number of manually annotated general frames, our unsupervised approach provides much more detailed insights, by considering a host of semantic axes. Our method is capable of quantitatively teasing out framing *bias*—how biased a text is in each microframe—and framing *intensity*—how much each microframe is used—from the text, offering a nuanced characterization of framing. We evaluate our approach using SemEval datasets as well as three other datasets and human evaluations, demonstrating that FrameAxis can reliably characterize documents with relevant microframes. Our method may allow scalable and nuanced computational analyses of framing across disciplines.

1 INTRODUCTION

Framing is a strategic process of highlighting a certain aspect of an issue to make it salient [8, 12]. By focusing on one particular aspect over another, even without making any biased argument, one can induce a biased understanding of the listeners [12, 17, 25]

Framing has been an active research subject, particularly in political discourse and news media, because framing is considered to be an effective tool for political persuasion [46]. It has been argued that the frames used by politicians and media shape the public understanding of issue salience [8, 26, 28, 50]. Thus, politicians strive to make their framing more prominent among the public [10]. Framing is not, however, confined to politics. It has been extensively studied in marketing [18, 21, 31], public health campaigns [15, 44], and other domains [22, 38]. Yet, the operationalization of framing is inherently vague [45, 48] and remains a challenging open question.

Because the common procedure of the framing research—identifying topics, isolating specific attitudes, building an initial set of frames for an issue, and analyzing the content based on a developed codebook—relies on manual examination [8], it is not only difficult to avoid the issue of subjectivity but also challenging to conduct a large-scale, systematic study that leverages huge online data.

Several computational approaches has been proposed to address these issues; most aim to characterize political discourse, for instance by recognizing political ideology [3, 47] and sentiment [40], or by leveraging established ideas such as the moral foundation theory [14, 23], general media frame [6], and frame-related language [4]. Yet, most studies still rely on small sets of predefined ideas, and it is difficult to generalize to non-political contexts (e.g., marketing messages or product reviews).

To overcome these limitations, we propose FrameAxis, an unsupervised method for characterizing a document with respect to various "microframes." The key ideas are: (i) a 'frame' can be operationalized by a combination of antonym pairs, such as *legal – illegal*, each of which is called a *microframe*; (ii) if a document contains many words that are close to the extreme 'poles' of a microframe (e.g., using a lot of words similar to 'legal' and 'illegal'), it is likely that the document is using that microframe rather than others; (iii) if the words used in the document are biased toward one of the poles, the frame used in the document can be considered as biased toward that pole (e.g., 'illegal' over 'legal'); and finally, (iv) the similarity between words and the poles of microframes can be effectively estimated using word embedding models [2, 27, 32].

To define microframes, FrameAxis considers a comprehensive set of *semantic axes*, each of which is a vector from a word (or a set of similar words) to its antonym(s) in the embedding space. In contrast to the 'general frame' approach [5], FrameAxis captures a set of more nuanced frames if those frames can be mapped into semantic axes. We build a set of 1,621 antonym pairs based on WordNet [33] and GloVe embeddings [39]. The word embedding allows us to quantitatively estimate each word's contribution on a given frame. Specifically, the contribution of a word to a frame is estimated by the alignment with the semantic axis vector that maps into the frame. The continuous nature of word embedding enables us to make use of every word in a document to characterize its framing, although it is possible to apply more sophisticated preprocessing steps. We propose two ways for aggregating word contributions for assessing the existence and bias of the framing. We evaluate our approach using SemEval datasets as well as three other datasets and human evaluations. We then explore the applicability of FrameAxis with a broad range of text from political news to movie reviews.

2 RELATED WORK

Although the concept of framing is intuitive, it is difficult to operationalize and identify. One of the simplest approaches is focusing on words or short phrases. Gentzkow and Shapiro [16] showed that politicians identified as strong Republican or strong Democratic strategically choose words to deliver their views; for instance, one may use the phrase "war on Iraq" instead of "war on terror." Thus, identifying over-represented words and phrases from a certain group (e.g., Republican or Democrat) on a given issue (e.g., guns) may be effective to reveal framing [34, 35]. The framing can be happening at the level of topics as well. Tsur et al. [49] identified a set of words, which is called a "topic cluster", that tends to be used together, such as 'health', 'care', 'law', 'affordable', 'act', 'obama'.

Recently, Boydstun et al. [5] proposed 'general media frames' as a unified topic-agnostic frame analysis framework. They find 15 frames, such as 'Morality' or 'Public opinion', that are commonly used by news media. The annotated dataset, called Media Frames Corpus [6], became an invaluable resource for the following studies on media framing [13, 24]. The 15 general media frames are, however, still restrictive. Liu et al. [29], for instance, introduced five frames that are not in the general media frames in order to analyze gun-violence news.

Another general framework for framing studies is the Moral Foundation Theory (MFT), which identifies the five basic moral 'axes' using antonyms, such as 'Care/Harm' and 'Fairness/Cheating', as the critical elements for individual judgment [19]. MFT has discovered politicians' stances on issues [23] and political leaning in partisan news [14], demonstrating its high interpretability based on antonymous axes.

Another approach that uses antonymous axes is SemAxis [2]. It characterizes the semantics of a word in different communities or domains (e.g., 'soft' in the context of sports vs. toy) by computing the similarities between the word and a set of predefined antonymous axes ("semantic axes"). The word and a semantic axis are located on the same vector space that is trained using the given text corpora and transfer learning.

FrameAxis was inspired by MFT and SemAxis, employing antonymous axes and the similarity between words and semantic axes. While SemAxis finds the semantic *change* of words, we focus—ignoring the semantic changes—on the *intensity* and *bias* in terms of semantic axes that we call "microframes."

3 FRAMEAXIS: CHARACTERIZING BIAS AND INTENSITY OF FRAMING

FrameAxis involves four steps: (i) building a set of microframes; (ii) computing word contributions to each microframe; (iii) calculating the bias and intensity of framing based on the word contributions; and finally (iv) identifying significant microframes.

3.1 Building a Set of Microframes

FrameAxis considers a semantic axis—a vector from one word to its antonym—as a microframe. We use the two terms microframe and semantic axis interchangeably in the rest of the paper.

Given a pair of antonyms (pole words), p_1 (e.g., 'sad') and p_2 (e.g., 'happy'), the semantic axis vector is $v_a = v_{p_2} - v_{p_1}$, where a is a semantic axis or a microframe (e.g., *sad* – *happy*), and v_{p_1} and v_{p_2} are the corresponding word vectors.

To capture nuanced framing, it is crucial to cover a variety of antonym pairs. We extract 1,828 adjective antonym pairs from WordNet [33] and remove 207 that are not present in the GloVe embeddings (840B tokens, 2.2M vocab, 300d vectors) [39]. As a result, we use 1,621 antonym pairs for the following experiments.

3.2 Contribution of a Word to Microframes

To calculate the contribution of a word to each microframe, we compute the similarity between the word vector v_w and the semantic axis vector v_a:

$$c_a^w = \frac{v_w \cdot v_a}{\| v_w \| \| v_a \|} \tag{1}$$

The absolute value of the similarity between a word vector and a microframe vector captures the relevance of the word to the microframe, while the sign of the similarity captures a bias toward one of the poles in the microframe.

3.3 Framing Bias and Intensity

Let us explain how to capture document-level framing based on the word contributions. First, we define the bias of framing on a microframe as the weighted average of the word contributions on that microframe for all the words in the text. This approach is conceptually rooted in the traditional 'expectancy value model' that explains an individual's attitude to an external entity, such as an object or an issue [36]. In the model, an attitude toward an object is calculated by the weighted sum of the evaluations on attribute a_i, whose weight is the salience of the attribute a_i of the object. An analogous framework based on a weighted average is also proposed in [9] to compute the overall valence score of a text. Similarly, we calculate the bias of the framing, $\mathcal{B}_a^t$, of the text corpus t on a semantic axis a as follows:

$$\mathcal{B}_a^t = \frac{\sum_{w \in W} f_w c_a^w}{\sum_{w \in W} f_w} \tag{2}$$

where f_w is the frequency of word w in document t, and W is a set of words in the document. For instance, if we consider a sentiment axis as a microframe (i.e., *good* – *bad*), the framing bias becomes equivalent to the dictionary-based sentiment score. In other words, the dictionary-based sentiment score is one realization of the framing bias.

Second, we define the intensity of framing on a microframe as the second moment of the word contributions on that microframe. Framing intensity on a microframe captures how *heavily* a microframe is used in a document. For instance, if a document is filled with words about 'legal' or 'illegal' perspectives of the subject, it is reasonable to say that the *legal* – *illegal* microframe is employed no matter what the aggregated bias is. Or, if a document is emotionally charged with many happy words as well as many sad words, we can say that the *happy* – *sad* microframe is actively used even if the overall bias is neutral. More formally, the intensity of framing, $\mathcal{I}_a^t$, of the text corpus t on a semantic axis a is calculated as follows:

$$\mathcal{I}_a^t = \frac{\sum_{w \in W} f_w (c_a^w - \mathcal{B}_a^T)^2}{\sum_{w \in W} f_w} \tag{3}$$

where $\mathcal{B}_a^T$ is the baseline framing bias of the entire text corpus T on a semantic axis a for computing the second moment.

3.4 Statistical Significance of Microframes

The framing bias and intensity should be interpreted with respect to the background distribution. For example, the significance of a positive word in an overall positive text corpus should be lower than that in an overall negative text corpus.

To consider the background distribution, we compute $\mathcal{B}_a^s$ and $\mathcal{I}_a^s$ from a bootstrapped sample s from the entire corpora T. We set the size of the sample s equal to that of the target corpus t. We estimate the statistical significance of the observed value by doing two-tailed tests on the N bootstrap samples. By setting a threshold p-value, we identify the significant semantic axes and rank them by the effect size (η, the difference between the observed value and the sample mean):

$$\eta_{\mathcal{B}} = \left| \mathcal{B}_a^t - \frac{\sum_s \mathcal{B}_a^s}{N} \right| \tag{4}$$

$$\eta_{\mathcal{I}} = \left| \mathcal{I}_a^t - \frac{\sum_s \mathcal{I}_a^s}{N} \right| \tag{5}$$

We then consider the microframe with the largest η as the most significant one. In this work, we use $N = 1,000$ and $p = 0.05$.

3.5 Per-Word Average Frame Shift

We define the per-word average framing bias and intensity shift as follows:

$$\mathcal{S}^w(\mathcal{B}_a) = \frac{f_w c_a^w}{\sum_{w \in W} f_w} \tag{6}$$

$$\mathcal{S}^w(\mathcal{I}_a) = \frac{f_w (c_a^w - \mathcal{B}_a^T)^2}{\sum_{w \in W} f_w} \tag{7}$$

which shows how a given word (w) brings a shift to framing bias and framing intensity by considering both the word's contribution to the microframe (c_a^w) and its appearances in the document (f_w).

4 EVALUATION

As we propose a new operationalization of framing (microframes), there is no large benchmark dataset yet. Thus, we employ two evaluation procedures: (i) we use related benchmark datasets, which are curated for an aspect-sentiment analysis; and (ii) we conduct a human evaluation with tasks similar to the word intrusion test [7].

4.1 Evaluation using SemEval Datasets

We use SemEval 2014 task 4 [43], SemEval 2015 task 12 [42], and SemEval 2016 task 5 [41] datasets, which are originally built for aspect-based sentiment analysis. We use restaurant reviews that are known to have clearer aspects for opinion targets [20].

We begin with identifying significant microframes based on framing bias for each 'aspect' and 'sentiment' set. We then check whether identified framing biases are aligned with the labeled sentiments. For instance, if the microframe *savory – unsavory* shows the framing bias from the positive reviews on food, then the identified framing bias should be toward 'savory' rather than 'unsavory' because positive reviews are more likely to contain positive perspectives. That is, the framing bias should be more likely to be aligned with the labeled sentiment. Such an alignment can be assessed by comparing the distances between the identified framing bias and each pole of the sentiment axis. We use *good – bad* as the sentiment axis, which performs best in finding domain-specific sentiment lexicons [2].

Table 1 shows the top 5 significant microframes and their biases in the positive reviews from the SemEval 2014 dataset. The significant microframes from the positive reviews are biased toward positive poles, as we expected. Examining the per-word average framing bias shift further confirms that FrameAxis picks up the relevant signals. For instance, the top words that contribute to the bias toward 'tasteful' include great, nice, romantic, excellent, cozy, and perfect.

Aspect	Microframe	Aligned
Ambience	**tasteful** – tasteless	T(rue)
Ambience	**beautiful** – ugly	T
Ambience	**pleasant** – unpleasant	T
Ambience	**elegant** – inelegant	T
Ambience	**comfortable** – uncomfortable	T
Service	**courteous** – discourteous	T
Service	**hospitable** – inhospitable	T
Service	**attentive** – inattentive	T
Service	inferior – **superior**	T
Service	**punctual** – unpunctual	T

Table 1: Top 5 significant microframes based on $\eta_{\mathcal{B}}$ in positive reviews (bold text shows the bias toward that pole).

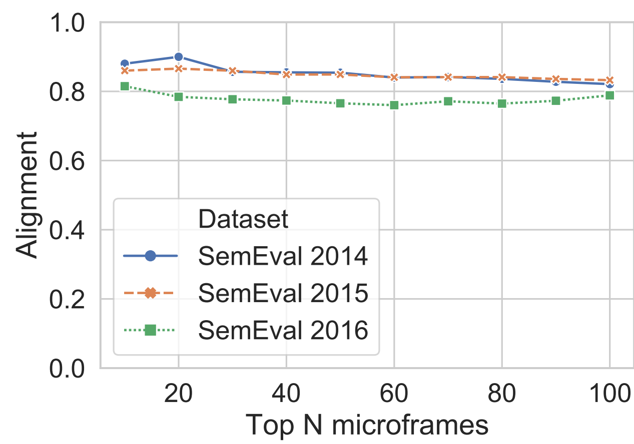

Figure 1: Fraction of aligned framing bias with labeled sentiment.

Figure 1 shows how well the framing bias is aligned with the labeled sentiment when considering the top N significant microframes by $\eta_{\mathcal{B}}$. The alignment is averaged across the aspects. The high levels of alignment with varying N suggests that FrameAxis reliably captures the framing bias.

The evaluation of the framing intensity is inspired by a simple idea that, given an 'aspect', both positive and negative reviews are likely to share similar microframes. For example, given the 'ambience' aspect, we expect to see positive reviews talking about quietness, as well as negative ones about noisiness. In other words, both positive and negative reviews, regardless of their biases, are likely to share a common microframe, such as *quiet – noisy*. Based on this intuition, we compare the overlap of microframes (based on the framing intensity) between pairs of reviews about the same aspect and those about the different aspects. We exclude anecdotes/miscellaneous reviews in this experiment.

Figure 2 shows that the reviews about the same aspect share a significantly higher number of microframes than those about the different aspects. Additionally, the per-word average framing intensity shift helps to explain why the significant microframes are

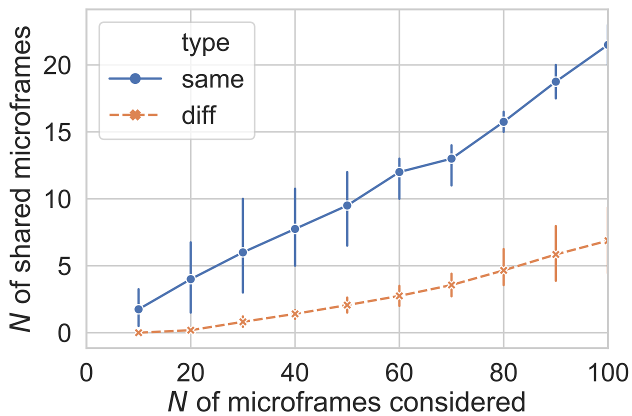

Figure 2: The overlap of microframes between same aspects but different sentiments and between different aspects (SemEval14)

identified. For instance, the top words for the intensity of the *cheap – expensive* microframe in the reviews on price include expensive, pricey, overpriced, and cheap.

4.2 Human Evaluation

Next, we perform human evaluations through Amazon Mechanical Turk (MTurk). Similar with the word intrusion test in evaluating topic modeling [7], we assess the quality of identified microframes by human raters. We evaluate framing bias and intensity separately, although both share most of the same procedures.

For framing bias, we prepare the top 10 significant microframes based on framing bias (i.e., answer set) and randomly selected 10 semantic axes with a random highlight of either pole (i.e., random set) for each pair of aspect and sentiment (e.g., positive reviews on ambience). As a unit of question-and-answer tasks in MTurk (Human Intelligence Task [HIT]), we ask "Which set of antonym pairs do better characterize a *positive* restaurant review on *ambience*? (A word on the right side of each pair (in bold) is associated with a *positive* restaurant review on *ambience*.)" The italic text is adapted to every aspect and sentiment. We note that, for every HIT, the order of semantic axes in both sets is shuffled and the location (i.e., top or bottom) of the answer set is also randomly chosen. Then, a worker chooses either the answer set or random set.

For framing intensity, we prepare the top 10 significant microframes (i.e., answer set) and randomly selected 10 semantic axes (i.e., random set) for each pair of aspect and sentiment. We then ask "Which set of antonym pairs do better characterize a *positive* restaurant review on *service*?" The rest of the procedure is the same.

For the quality control, we recruit workers who (1) live in the U.S., (2) have more than 1,000 approved HITs, and (3) achieve 95% of approval rates. Also, we allow a worker to answer up to 10 HITs. We recruit 15 workers for each (aspect, sentiment) pair. We pay 0.02 USD for each HIT. For human evaluation in the following sections with other datasets, we use the same process with refined questions.

(Sentiment) Aspect	Acc. for $\mathcal{B}$	Acc. for $\mathcal{I}$
(+) Service	1.000	0.867
(+) Price	0.867	0.733
(+) Food	0.933	0.800
(+) Ambience	1.000	0.600
(-) Service	0.867	0.867
(-) Price	0.667	0.667
(-) Food	0.867	0.733
(-) Ambience	0.800	0.733
Average	0.875	0.750

Table 2: Human evaluation for significant microframes based on framing bias and framing intensity.

Table 2 shows the fraction of the correct choices of workers (i.e., choosing the answer set). The overall average accuracy is 87.5% and 75.0% for significant microframes based on framing bias and framing intensity, respectively. For framing bias, (+) Service and (+) Ambience, which shows high relevance to context in Table 1, human raters chose the answer sets correctly without errors. In contrast, for framing intensity, some groups show a relatively lower performance. We manually check the (+) Ambience case. Interestingly, we find that many significant microframes actually are relevant (e.g., *tasteful – tastless*, *active – quiet*, *beautiful – ugly*, *restful – restless*, *hurried – unhurried*), but some are not (e.g., *debilitating – invigorating*, *feminine – neuter*). While such noisy microframes are a few, they might mislead users. Also, when generic axes, such as *good – bad*, appeared in a random set, workers tended to choose the random set.

5 MICROFRAMES IN MEDIA FRAMES CORPUS

Here we use the 15 predefined general media frames [5] and the Media Frames Corpus [6] to conduct further evaluations.

5.1 Significant Microframes Based on Framing Bias

We begin with framing bias in the general frames. Table 3 shows the top 10 significant microframes based on framing bias from the news on immigration with 'Capacity and resources' and 'Crime and punishment' frames. Bold text shows the bias toward that pole on the corresponding semantic axis.

We see that most microframes in Table 3 are relevant to the corresponding general frames. While news of Capacity and resources frame tend to have the bias toward positive perspectives (e.g., capable, eligible, good, reputable, adjusted), news of Crime and punishment frame is more likely to have the bias toward negative perspectives (e.g., intoxicated, false, unlawful, suspected, negligent) rather than positive perspectives. The top words of per-word average framing bias shift on the *lawful – unlawful* axis, include illegal, arrested, charges, officials, police, criminal, smuggling, and guilty. They are relevant for describing the illegal perspectives of immigration news. These biases toward either positive or negative

Capacity and resources	Crime and punishment
capable – incapable	**intoxicated** – sober
eligible – ineligible	blue-collar – **white-collar**
just – unjust	**false** – true
evil – **good**	lawful – **unlawful**
disreputable – **reputable**	**suspected** – unsuspected
adjusted – maladjusted	diligent – **negligent**
immodest – **modest**	monogamous – **polygamous**
compliant – defiant	**illegal** – legal
regenerate – unregenerate	better – **worse**
maternal – paternal	**dead** – live

Table 3: Top 10 significant microframes based on $\eta_{\mathcal{B}}$. Bold text represents the bias toward that pole.

perspectives indicate that a general frame intentionally or unintentionally tends to convey particular connotations.

For a human evaluation, we ask "Which set of antonym pairs do you think better align with this *Morality* frame? The bold text shows the perspective with emphasis." with the description of *Morality*. The italic text is changed according to the general frame. Workers are likely to choose the microframes identified by FrameAxis with an average accuracy of 0.768. The full evaluation results are presented in Appendix.

5.2 Significant Microframes Based on Framing Intensity

Morality	Public Opinion
dishonest – honest	anti – pro
evil – good	surprised – unsurprised
misused – used	disenchanted – enchanted
unworthy – worthy	easy – uneasy
professional – unprofessional	kind – unkind
ethical – unethical	happy – unhappy
improper – proper	interested – uninterested
false – true	displeased – pleased
trustworthy – untrustworthy	afraid – unafraid
disreputable – reputable	moved – unmoved

Table 4: Top 10 significant microframes based on $\eta_{\mathcal{I}}$.

Table 4 shows the top 10 significant microframes based on framing intensity in the news on smoking with Morality and Public opinion frames. The *dishonest – honest* microframe for the Morality frame and the *anti – pro* microframe for the Public Opinion frame are the most evident examples that are directly connected to both frames. More interestingly, microframes reveal rich details beyond the general frames. News on smoking with the Morality frame deal with worthiness, professionalism, ethics, and trustworthiness as well as honesty. It becomes clearer when we see the top words that contribute to each frame, which include deceptive, misleading, conspired, and manipulated for the *dishonest – honest* frame, and moral, ethical, lied, responsibility, and truth for the *ethical – unethical* frame. Similarly, news on smoking with the Public opinion frame contain kindness, happiness, and interest as well as support or opposition. Top words include anti, ban, antismoking, prevention, stop, and advocates for the *anti – pro* frame.

The identified microframes from other general frames, which are not included in Table 4, also show finer-grained dimensions embedded in general frames, such as the *fair – unfair* microframe in news of Fairness and equality frame, in which the top words include unfair, fair, discrimination, ban, and rights, and the *illegal – legal* microframe in news of Legality constitutionality frame, in which the top words include legal, attorney, lawyers, court, litigation, and judge.

Our human evaluation shows an average accuracy of 0.792. The full evaluation results are presented in Appendix.

6 MICROFRAMES IN PARTISAN NEWS

News headlines concisely convey the key message of the article [37] and thus have been extensively studied [11]. In this work, we examine framing in news headlines published by liberal and conservative media collected from AllSides.com [1]. Our full corpus contains 50,073 headlines from 572 news media.

6.1 Significant Microframes Based on Framing Bias

Healthcare	Immigration
essential - **inessential**	**illegal** - legal
compassionate - **uncompassionate**	ethical - **unethical**
healthful - **unhealthful**	constitutional - **unconstitutional**
geared - **ungeared**	**immoral** - moral
social - **unsocial**	**concealing** - revealing
domestic - **undomestic**	critical - **noncritical**
surprising - **unsurprising**	**intoxicated** - sober
functional - **nonfunctional**	**false** - true
critical - **noncritical**	diligent - **negligent**
aboral - oral	competitive - **noncompetitive**

Table 5: Top 10 significant microframes based on the absolute differences of $\eta_{\mathcal{B}}$ between liberal and conservative media. Bold text shows the bias of the conservative media toward that pole.

We examine how political leaning is represented in the news on healthcare and immigration. Table 5 shows the top 10 significant microframes based on the absolute difference in the framing bias between liberal and conservative media for the news on healthcare and immigration, which is computed by ($\mathcal{B}_a^{c_{\text{con}}} - \mathcal{B}_a^{c_{\text{lib}}}$) where c_{con} is the news from conservative media and c_{lib} is the news from liberal media. Bold text shows the bias of the conservative media toward that pole.

The negative stance of conservative media toward healthcare reform (the Affordable Care Act) concurs with the set of significant microframes and the bias of conservative media. The conservative media convey inessential, uncompassionate, unhealthful, and nonfunctional perspectives. Top words contributing to the *essential – inessential* microframe, for instance, include defund, flub, defunding, weaseling, cancelations, freeloading, and overpromising. Regarding the news on immigration, conservative media emphasize illegal, negligent, and noncompetitive perspectives. In contrast, liberal media stress ethical, constitutional, moral, and diligent perspectives when reporting news on immigration and immigrants. Top words contributing to the *illegal – legal* microframe include border, illegal, legal, reform, ban, and undocumented.

For human evaluation, we ask workers "Which set of antonym pairs do you think better distinguish between what liberal and conservative news media report on *Immigration*?" The italic text changes according to the topic. The average accuracy across the topics is 0.747. The full evaluation results are presented in Appendix.

6.2 Significant Microframes Based on Framing Intensity

Liberal media	Conservative media
maternal - paternal	nonsignificant - significant
dramatic - lyric	surprised - unsurprised
bisexual - homosexual	surprising - unsurprising
adjusted - unadjusted	gathered - ungathered
hateful - lovable	evitable - inevitable
center - right	influential - uninfluential
thrifty - wasteful	excited - unexcited
disingenuous - ingenuous	obvious - unobvious
reconstructed - unreconstructed	nonrepresentative - representative
disenchanted - enchanted	ready - unready

Table 6: Top 10 significant microframes based on $\eta_{\mathcal{I}}$.

Table 6 shows the top 10 significant microframes based on framing intensity to characterize news on abortion from liberal and conservative media. In addition to framing bias, the result of framing intensity presents how liberal and conservative news media deal with the same topic differently. Overall, liberal news media underscore individual aspects. The semantic axes related to individual aspects rather than systemic issues, such as the *maternal – paternal* microframe of which top words include women, reproductive, birth, pregnancy, and pill, and the *hateful – lovable* microframe of which top words include demonizing, extremist, hypocrisy, and religious, strongly appear in liberal media rather than conservative media. Given that 'pro-choice' emphasizes the right of women, it makes sense that liberal media uses the *maternal – paternal* microframe more prominently than conservative media.

For human evaluation, we ask "Which set of antonym pairs do you think better align with how *liberal* news media report on *Immigration*?" The italic text is changed according to the political leaning of the news media and topics. The average accuracy is 0.613 for both liberal and conservative media. Through manual inspection, we find that some microframes that capture the characteristics of headlines to attract readers, such as *surprised – unsurprised*, might mislead workers. The full evaluation results are presented in Appendix.

Documentary	Animation
forgettable – **unforgettable**	**happy** – unhappy
inspiring – uninspiring	humorless – **humorous**
valuable – worthless	**lucky** – unlucky
dead – **live**	**colorful** – colorless
beautiful – ugly	**playable** – unplayable
professional – unprofessional	**friendly** – unfriendly
creative – uncreative	**popular** – unpopular
interesting – uninteresting	**helpful** – unhelpful
meaningful – meaningless	**animated** – unanimated
strong – weak	misused – **used**

Table 7: Top 10 significant microframes based on $\eta_{\mathcal{B}}$. Bold text shows the bias toward that pole.

7 MICROFRAMES IN IMDB MOVIE REVIEWS

Finally, we explore the applicability of FrameAxis to non-political text, IMDb movie reviews [30]. We look into how movie reviews of different genres use different microframes.

7.1 Significant Microframes Based on Framing Bias

We begin with identifying significant microframes based on framing bias in movie reviews of different genres. Table 7 shows the top 10 significant microframes based on framing bias from Documentary and Animation movie reviews.

As the average ratings of reviews for both genres are quite high (7.19 for Documentary and 6.90 for Animation), it is reasonable that the positivity bias is observed from the framing bias. Also, the framing bias in Documentary movie reviews shows which perspectives are considered important compared to other genres. That is, people like beautiful, meaningful, and creative perspectives of documentaries. Similarly, people like happy, humorous, colorful, and friendly aspects of Animation movies.

For human evaluation, we ask "Which set of antonym pairs do you think better align with *Comedy* movies?" The italic text is changed according to the movie genre. The average accuracy across all genres is 0.733. The full evaluation results are presented in Appendix.

7.2 Significant Microframes Based on Frame Intensity

Table 8 shows the top 10 significant microframes based on framing intensity to characterize Comedy and Mystery movie reviews. It shows that people actively talk about irreverentness, profaneness, inappropriateness, and unconventional in Comedy movie reviews. For example, top words include silly, bad, satire, ridiculous, quirky, annoying, spoof, awful, and lame as well as funny, hilarious, humor, fun, clever, story, brilliant, and interesting for the *irreverent – reverent* microframe. In other words, the line between humor and off-limit jokes is what people intensively discuss in Comedy reviews. In Mystery movie reviews, the discovered microframes fit in the unique characteristics of the Mystery genre, which usually involves solving a crime or exploring an unexplained phenomena.

Comedy	Mystery
irreverent – reverent	normal – paranormal
profane – sacred	solved – unsolved
appropriate – inappropriate	better – worse
humorless – humorous	natural – supernatural
incoming – outgoing	lucky – unlucky
even – odd	settled – unsettled
conventional – unconventional	best – worst
expected – unexpected	unwilling – willing
tasteless – tasty	misused – used
original – unoriginal	easy – uneasy

Table 8: Top 10 significant microframes based on $\eta_{\mathcal{I}}$.

For instance, top words for the *normal – paranormal* microframe, which include thriller, suspense, and horror, show the reasonable connection of paranormal elements with specific moods. The average accuracy across the genres from a human evaluation is 0.827 (see Appendix for full details).

8 DISCUSSION AND CONCLUDING REMARKS

In this work, we proposed FrameAxis, a method to characterize the framing bias and intensity of a given text based on the semantic axes embedded in a vector space. Using multiple datasets, we showed that FrameAxis successfully identified meaningful microframes from political text (partisan news) to non-political text (movie reviews).

Our quantitative evaluation using SemEval datasets showed that identified bias are well aligned with labeled sentiments and successfully discovered finer-grained microframes. Our human evaluation showed that FrameAxis can reliably characterize framing of the text. Also, the semantic axes identified by the framing bias are more relevant to some corpus, and those identified by the framing intensity are better reflections of another corpus. They capture different characteristics from the text, like the sentiment (positive or negative) and the polarity (polarized or non-polarized) analyses reflect different properties of the text.

Dataset	N=10	N=50	N=100
MFC	0.20	0.18	0.19
AllSides	0.09	0.12	0.14
IMDb	0.08	0.07	0.09

Table 9: Fraction of common microframes based on framing bias and intensity

We quantitatively show the difference of microframes based on framing bias and intensity. We find that 19 (MFC), 14 (AllSides), and 9 (IMDb) microframes are common between the top 100 microframes based on framing bias and intensity. Table 9 shows the fraction of common microframes when considering the top N microframes based on framing bias and intensity. For example, the fraction of 0.18 (MFC, N=50) means that, 9.0 microframes (= 50 × 0.18) are common when comparing the top 50 microframes identified by framing bias with ones by framing intensity and being averaged across the topics and general frames. The overall low number of the common microframes prove that microframes based on framing bias and intensity capture different dimensions of the text. In addition, human evaluations explained in earlier sections showed that their capturing dimensions are relevant to context.

Some limitations should be noted. First, we use static embedding for efficient computation, but contextual embedding has many advantages compared to static embedding. Second, while the current version of FrameAxis shows reasonable performance, assigning different weights to words based on their importance or sophisticated preprocessing of data might capture better microframes.

We will release FrameAxis as an easy-to-use tool with appropriate supporting visualization of per-word average frame shift for a broader audience. We believe that such efforts would incur computational analyses of framing across disciplines.

REFERENCES

[1] AllSides. 2012. AllSides – Topics & Issues. (2012). https://www.allsides.com/topics-issues.
[2] Jisun An, Haewoon Kwak, and Yong-Yeol Ahn. 2018. SemAxis: A Lightweight Framework to Characterize Domain-Specific Word Semantics Beyond Sentiment. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics.
[3] David Bamman and Noah A Smith. 2015. Open extraction of fine-grained political statements. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. 76–85.
[4] Eric Baumer, Elisha Elovic, Ying Qin, Francesca Polletta, and Geri Gay. 2015. Testing and comparing computational approaches for identifying the language of framing in political news. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 1472–1482.
[5] Amber E Boydstun, Justin H Gross, Philip Resnik, and Noah A Smith. 2013. Identifying media frames and frame dynamics within and across policy issues. In *New Directions in Analyzing Text as Data Workshop, London.*
[6] Dallas Card, Amber E. Boydstun, Justin H. Gross, Philip Resnik, and Noah A. Smith. 2015. 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)*. Association for Computational Linguistics, Beijing, China, 438–444. https://doi.org/10.3115/v1/P15-2072
[7] Jonathan Chang, Sean Gerrish, Chong Wang, Jordan L Boyd-Graber, and David M Blei. 2009. Reading tea leaves: How humans interpret topic models. In *Advances in neural information processing systems*. 288–296.
[8] Dennis Chong and James N Druckman. 2007. Framing theory. *Annu. Rev. Polit. Sci.* 10 (2007), 103–126.
[9] Peter Sheridan Dodds and Christopher M Danforth. 2010. Measuring the happiness of large-scale written expression: Songs, blogs, and presidents. *Journal of happiness studies* 11, 4 (2010), 441–456.
[10] James N Druckman and Kjersten R Nelson. 2003. Framing and deliberation: How citizens' conversations limit elite influence. *American Journal of Political Science* 47, 4 (2003), 729–745.
[11] Stephanie Edgerly, Rachel R Mourão, Esther Thorson, and Samuel M Tham. 2019. When Do Audiences Verify? How Perceptions About Message and Source Influence Audience Verification of News Headlines. *Journalism & Mass Communication Quarterly* (2019), 1077699019864680.
[12] Robert M Entman. 1993. Framing: Toward clarification of a fractured paradigm. *Journal of communication* 43, 4 (1993), 51–58.
[13] Anjalie Field, Doron Kliger, Shuly Wintner, Jennifer Pan, Dan Jurafsky, and Yulia Tsvetkov. 2018. Framing and Agenda-setting in Russian News: a Computational Analysis of Intricate Political Strategies. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Brussels, Belgium, 3570–3580. https://www.aclweb.org/anthology/D18-1393
[14] Dean Fulgoni, Jordan Carpenter, Lyle H Ungar, and Daniel Preotiuc-Pietro. 2016. An Empirical Exploration of Moral Foundations Theory in Partisan News Sources.. In *LREC*.
[15] Kristel M Gallagher and John A Updegraff. 2011. Health message framing effects on attitudes, intentions, and behavior: a meta-analytic review. *Annals of behavioral medicine* 43, 1 (2011), 101–116.

[16] Matthew Gentzkow and Jesse M Shapiro. 2010. What drives media slant? Evidence from US daily newspapers. *Econometrica* 78, 1 (2010), 35–71.
[17] Erving Goffman. 1974. *Frame analysis: An essay on the organization of experience.* Harvard University Press.
[18] Dhruv Grewal, Jerry Gotlieb, and Howard Marmorstein. 1994. The moderating effects of message framing and source credibility on the price-perceived risk relationship. *Journal of consumer research* 21, 1 (1994), 145–153.
[19] Jonathan Haidt and Jesse Graham. 2007. When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research* 20, 1 (2007), 98–116.
[20] Ruidan He, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier. 2018. Effective attention modeling for aspect-level sentiment classification. In *Proceedings of the 27th International Conference on Computational Linguistics*. 1121–1131.
[21] Pamela M Homer and Sun-Gil Yoon. 1992. Message framing and the interrelationships among ad-based feelings, affect, and cognition. *Journal of Advertising* 21, 1 (1992), 19–33.
[22] Albert H Huang, Kuanchin Chen, David C Yen, and Trang P Tran. 2015. A study of factors that contribute to online review helpfulness. *Computers in Human Behavior* 48 (2015), 17–27.
[23] Kristen Johnson and Dan Goldwasser. 2018. Classification of moral foundations in microblog political discourse. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 720–730.
[24] Kristen Johnson, Di Jin, and Dan Goldwasser. 2017. Leveraging behavioral and social information for weakly supervised collective classification of political discourse on Twitter. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 741–752.
[25] Daniel Kahneman and Amos Tversky. 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica* 47, 2 (1979), 263–292.
[26] Donald R Kinder. 1998. Communication and opinion. *Annual review of political science* 1, 1 (1998), 167–197.
[27] Austin C. Kozlowski, Matt Taddy, and James A. Evans. 2019. The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings. *American Sociological Review* 84, 5 (2019), 905–949. https://doi.org/10.1177/0003122419877135
[28] George Lakoff. 2014. *The all new don't think of an elephant!: Know your values and frame the debate.* Chelsea Green Publishing.
[29] Siyi Liu, Lei Guo, Kate Mays, Margrit Betke, and Derry Tanti Wijaya. 2019. Detecting Frames in News Headlines and Its Application to Analyzing News Framing Trends Surrounding U.S. Gun Violence. In *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*. Association for Computational Linguistics, Hong Kong, China, 504–514. https://doi.org/10.18653/v1/K19-1047
[30] Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. 2011. Learning Word Vectors for Sentiment Analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, Portland, Oregon, USA, 142–150. http://www.aclweb.org/anthology/P11-1015
[31] Durairaj Maheswaran and Joan Meyers-Levy. 1990. The influence of message framing and issue involvement. *Journal of Marketing research* 27, 3 (1990), 361–367.
[32] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*. 3111–3119.
[33] George A Miller. 1995. WordNet: a lexical database for English. *Commun. ACM* 38, 11 (1995), 39–41.
[34] Burt L Monroe, Michael P Colaresi, and Kevin M Quinn. 2008. Fightin'words: Lexical feature selection and evaluation for identifying the content of political conflict. *Political Analysis* 16, 4 (2008), 372–403.
[35] Supun Chathuranga Nakandala, Giovanni Luca Ciampaglia, Norman Makoto Su, and Yong-Yeol Ahn. 2017. Gendered conversation in a social game-streaming platform. In *Eleventh International AAAI Conference on Web and Social Media.*
[36] Thomas E Nelson, Zoe M Oxley, and Rosalee A Clawson. 1997. Toward a psychology of framing effects. *Political behavior* 19, 3 (1997), 221–246.
[37] Raphael Nir. 1993. A discourse analysis of news headlines. *Hebrew Linguistics* 37 (1993), 23–31.
[38] Luc G Pelletier and Elizabeth Sharp. 2008. Persuasive communication and proenvironmental behaviours: How message tailoring and message framing can improve the integration of behaviours through self-determined motivation. *Canadian Psychology/Psychologie Canadienne* 49, 3 (2008), 210.
[39] Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. 1532–1543.
[40] Ferran Pla and Lluís-F Hurtado. 2014. Political tendency identification in Twitter using sentiment analysis techniques. In *Proceedings of COLING 2014, the 25th international conference on computational linguistics: Technical Papers*. 183–192.
[41] Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad AL-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, Véronique Hoste, Marianna Apidianaki, Xavier Tannier, Natalia Loukachevitch, Evgeniy Kotelnikov, Nuria Bel, Salud María Jiménez-Zafra, and Gülşen Eryiğit. 2016. SemEval-2016 Task 5: Aspect Based Sentiment Analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*. Association for Computational Linguistics, San Diego, California, 19–30. https://doi.org/10.18653/v1/S16-1002
[42] Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Suresh Manandhar, and Ion Androutsopoulos. 2015. SemEval-2015 Task 12: Aspect Based Sentiment Analysis. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*. Association for Computational Linguistics, Denver, Colorado, 486–495. https://doi.org/10.18653/v1/S15-2082
[43] Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. SemEval-2014 Task 4: Aspect Based Sentiment Analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*. Association for Computational Linguistics, Dublin, Ireland, 27–35. https://doi.org/10.3115/v1/S14-2004
[44] Alexander J Rothman and Peter Salovey. 1997. Shaping perceptions to motivate healthy behavior: the role of message framing. *Psychological bulletin* 121, 1 (1997), 3.
[45] Dietram A Scheufele. 1999. Framing as a theory of media effects. *Journal of communication* 49, 1 (1999), 103–122.
[46] Dietram A Scheufele and David Tewksbury. 2006. Framing, agenda setting, and priming: The evolution of three media effects models. *Journal of communication* 57, 1 (2006), 9–20.
[47] Yanchuan Sim, Brice DL Acree, Justin H Gross, and Noah A Smith. 2013. Measuring ideological proportions in political speeches. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. 91–101.
[48] Paul M Sniderman and Sean M Theriault. 2004. The structure of political argument and the logic of issue framing. *Studies in public opinion: Attitudes, nonattitudes, measurement error, and change* (2004), 133–65.
[49] Oren Tsur, Dan Calacci, and David Lazer. 2015. A frame of mind: Using statistical models for detection of framing and agenda setting campaigns. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Vol. 1. 1629–1638.
[50] John R Zaller et al. 1992. *The nature and origins of mass opinion.* Cambridge university press.