

And That’s A Fact: Distinguishing Factual and Emotional Argumentation in Online Dialogue

Shereen Oraby*, Lena Reed*, Ryan Compton*,
Ellen Riloff[†], Marilyn Walker* and Steve Whittaker*

* University of California Santa Cruz

{soraby, lireed, rcompton, mawalker, swhittak}@ucsc.edu

[†] University of Utah

rilloff@cs.utah.edu

Abstract

We investigate the characteristics of factual and emotional argumentation styles observed in online debates. Using an annotated set of FACTUAL and FEELING debate forum posts, we extract patterns that are highly correlated with factual and emotional arguments, and then apply a bootstrapping methodology to find new patterns in a larger pool of unannotated forum posts. This process automatically produces a large set of patterns representing linguistic expressions that are highly correlated with factual and emotional language. Finally, we analyze the most discriminating patterns to better understand the defining characteristics of factual and emotional arguments.

1 Introduction

Human lives are being lived online in transformative ways: people can now ask questions, solve problems, share opinions, or discuss current events with anyone they want, at any time, in any location, on any topic. The purposes of these exchanges are varied, but a significant fraction of them are argumentative, ranging from hot-button political controversies (e.g., national health care) to religious interpretation (e.g., Biblical exegesis). And while the study of the structure of arguments has a long lineage in psychology (Cialdini, 2000) and rhetoric (Hunter, 1987), large shared corpora of natural informal argumentative dialogues have only recently become available.

Natural informal dialogues exhibit a much broader range of argumentative styles than found in traditional work on argumentation (Marwell and

Schmitt, 1967; Cialdini, 2000; McAlister et al., 2014; Reed and Rowe, 2004). Recent work has begun to model different aspects of these natural informal arguments, with tasks including stance classification (Somasundaran and Wiebe, 2010; Walker et al., 2012), argument summarization (Misra et al., 2015), sarcasm detection (Justo et al., 2014), and work on the detailed structure of arguments (Biran and Rambow, 2011; Purpura et al., 2008; Yang and Cardie, 2013). Successful models of these tasks have many possible applications in sentiment detection, automatic summarization, argumentative agents (Zuckerman et al., 2015), and in systems that support human argumentative behavior (Rosenfeld and Kraus, 2015).

Our research examines FACTUAL versus FEELING argument styles, drawing on annotations provided in the Internet Argument Corpus (IAC) (Walker et al., 2012). This corpus includes quote-response pairs that were manually annotated with respect to whether the response is primarily a FACTUAL or FEELING based argument, as Section 2.1 describes in more detail. Figure 1 provides examples of responses in the IAC (paired with preceding quotes to provide context), along with the response’s FACTUAL vs. FEELING label.

FACTUAL responses may try to bolster their argument by providing statistics related to a position, giving historical or scientific background, or presenting specific examples or data. There is clearly a relationship between a proposition being FACTUAL versus OBJECTIVE or VERIDICAL, although each of these different labelling tasks may elicit differences from annotators (Wiebe and Riloff, 2005; Riloff and

Wiebe, 2003; Saurí and Pustejovsky, 2009; Park and Cardie, 2014).

Class	Debate Forum Dialogue
FACT	<p>Quote: Even though our planet is getting warmer, it is still a lot cooler than it was 4000 years ago.</p> <p>Response: <i>The average global temperature follows a sinusoidal pattern, the general consensus is we are supposed to be approaching a peak. Projections show that instead of peaking, there will be continue to be an increase in average global temperature.</i></p>
FACT	<p>Quote: “When you go to war against your enemies...suppose you see a beautiful woman whom you desire...you shall take her..and she shall marry you.” - Deut. 21:10</p> <p>Response: <i>Read to the very end of the verse. “If you are not pleased with her, let her go wherever she wishes. You must not sell her or treat her as a slave, since you have dishonored her.”</i></p>
FEEL	<p>Quote: Talk about begging the question! I don’t want your gun, and if such a law were passed it’s not my job to enforce the law.</p> <p>Response: <i>I see you are willing to violate my constitutional rights yet you expect someone else to do your dirty work.... How typical.</i></p>
FEEL	<p>Quote: “WASHINGTON &#8211; Supreme Court aspirant Sonia Sotomayor said Tuesday that she considers the question of abortion rights is settled precedent and says there is a constitutional right to privacy. The federal appeals court judge was asked at her confirmation...”</p> <p>Response: <i>While I’m still iffy on her with the whole New Haven case, and her off-the-bench comments on race, this is one thing I commend her for and agree completely with.</i></p>

Figure 1: Examples of FACTUAL and FEELING based debate forum Quotes and Responses. Only the responses were labeled for FACT vs. FEEL.

The FEELING responses may seem to lack argumentative merit, but previous work on argumentation describes situations in which such arguments can be effective, such as the use of emotive arguments to draw attention away from the facts, or to frame a discussion in a particular way (Walton, 2010; Macagno and Walton, 2014). Further-

more, work on persuasion suggest that FEELING based arguments can be more persuasive in particular circumstances, such as when the hearer shares a basis for social identity with the source (speaker) (Chaiken, 1980; Petty and Cacioppo, 1986; Benoit, 1987; Cacioppo et al., 1983; Petty et al., 1981). However none of this work has documented the linguistic patterns that characterize the differences in these argument types, which is a necessary first step to their automatic recognition or classification. Thus the goal of this paper is to use computational methods for pattern-learning on conversational arguments to catalog linguistic expressions and stylistic properties that distinguish Factual from Emotional arguments in these on-line debate forums.

Section 2.1 describes the manual annotations for FACTUAL and FEELING in the IAC corpus. Section 2.2 then describes how we generate lexico-syntactic patterns that occur in both types of argument styles. We use a weakly supervised pattern learner in a bootstrapping framework to automatically generate lexico-syntactic patterns from both annotated and unannotated debate posts. Section 3 evaluates the precision and recall of the FACTUAL and FEELING patterns learned from the annotated texts and after bootstrapping on the unannotated texts. We also present results for a supervised learner with bag-of-word features to assess the difficulty of this task. Finally, Section 4 presents analyses of the linguistic expressions found by the pattern learner and presents several observations about the different types of linguistic structures found in FACTUAL and FEELING based argument styles. Section 5 discusses related research, and Section 6 sums up and proposes possible avenues for future work.

2 Pattern Learning for Factual and Emotional Arguments

We first describe the corpus of online debate posts used for our research, and then present a bootstrapping method to identify linguistic expressions associated with FACTUAL and FEELING arguments.

2.1 Data

The IAC corpus is a freely available annotated collection of 109,553 forum posts (11,216 discussion

threads).¹ In such forums, conversations are started by posting a topic or a question in a particular category, such as society, politics, or religion (Walker et al., 2012). Forum participants can then post their opinions, choosing whether to respond directly to a previous post or to the top level topic (start a new thread). These discussions are essentially dialogic; however the affordances of the forum such as asynchrony, and the ability to start a new thread rather than continue an existing one, leads to dialogic structures that are different than other multi-party informal conversations (Fox Tree, 2010). An additional source of dialogic structure in these discussions, above and beyond the thread structure, is the use of the quote mechanism, which is an interface feature that allows participants to optionally break down a previous post into the components of its argument and respond to each component in turn.

The IAC includes 10,003 Quote-Response (Q-R) pairs with annotations for FACTUAL vs. FEELING argument style, across a range of topics. Figure 2 shows the wording of the survey question used to collect the annotations. Fact vs. Feeling was measured as a scalar ranging from -5 to +5, because previous work suggested that taking the means of scalar annotations reduces noise in Mechanical Turk annotations (Snow et al., 2008). Each of the pairs was annotated by 5-7 annotators.

For our experiments, we use only the response texts and assign a binary FACT or FEEL label to each response: texts with score > 1 are assigned to the FACT class and texts with score < -1 are assigned to the FEELING class. We did not use the responses with scores between -1 and 1 because they had a very weak Fact/Feeling assessment, which could be attributed to responses either containing aspects of *both* factual and feeling expression, or neither. The resulting set contains 3,466 FACT and 2,382 FEELING posts. We randomly partitioned the FACT/FEEL responses into three subsets: a training set with 70% of the data (2,426 FACT and 1,667 FEELING posts), a development (tuning) set with 20% of the data (693 FACT and 476 FEELING posts), and a test set with 10% of the data (347 FACT and 239 FEELING posts). For the bootstrapping method, we also used 11,560 responses from the unannotated data.

¹<https://nlds.soe.ucsc.edu/iac>

Slider Scale -5,5: Survey Question
Fact/Emotion: Is the respondent attempting to make a fact based argument or appealing to feelings and emotions?

Figure 2: Mechanical Turk Survey Question used for Fact/Feeling annotation.

2.2 Bootstrapped Pattern Learning

The goal of our research is to gain insights into the types of linguistic expressions and properties that are distinctive and common in factual and feeling based argumentation. We also explore whether it is possible to develop a high-precision FACT vs. FEELING classifier that can be applied to unannotated data to find new linguistic expressions that did not occur in our original labeled corpus.

To accomplish this, we use the AutoSlog-TS system (Riloff, 1996) to extract linguistic expressions from the annotated texts. Since the IAC also contains a large collection of unannotated texts, we then embed AutoSlog-TS in a bootstrapping framework to learn additional linguistic expressions from the unannotated texts. First, we briefly describe the AutoSlog-TS pattern learner and the set of pattern templates that we used. Then, we present the bootstrapping process to learn more Fact/Feeling patterns from unannotated texts.

2.2.1 Pattern Learning with AutoSlog-TS

To learn patterns from texts labeled as FACT or FEELING arguments, we use the AutoSlog-TS (Riloff, 1996) extraction pattern learner, which is freely available for research. AutoSlog-TS is a weakly supervised pattern learner that requires training data consisting of documents that have been labeled with respect to different categories. For our purposes, we provide AutoSlog-TS with responses that have been labeled as either FACT or FEELING.

AutoSlog-TS uses a set of syntactic templates to define different types of linguistic expressions. The left-hand side of Figure 3 shows the set of syntactic templates defined in the AutoSlog-TS software package. PassVP refers to passive voice verb phrases (VPs), ActVP refers to active voice VPs, InfVP refers to infinitive VPs, and AuxVP refers to VPs where the main verb is a form of “to be” or “to have”. Subjects (subj), direct objects (dobj), noun phrases (np), and possessives (genitives) can be ex-

tracted by the patterns. AutoSlog-TS applies the Sundance shallow parser (Riloff and Phillips, 2004) to each sentence and finds every possible match for each pattern template. For each match, the template is instantiated with the corresponding words in the sentence to produce a specific lexico-syntactic expression. The right-hand side of Figure 3 shows an example of a specific lexico-syntactic pattern that corresponds to each general pattern template.²

Pattern Template	Example Pattern
<subj> PassVP	<subj> was observed
<subj> ActVP	<subj> observed
<subj> ActVP Dobj	<subj> want explanation
<subj> ActInfVP	<subj> expected to find
<subj> PassInfVP	<subj> was used to measure
<subj> AuxVP Dobj	<subj> was success
<subj> AuxVP Adj	<subj> is religious
ActVP <dobj>	create <dobj>
InfVP <dobj>	to limit <dobj>
ActInfVP <dobj>	like to see <dobj>
PassInfVP <dobj>	was interested to see <dobj>
Subj AuxVP <dobj>	question is <dobj>
NP Prep <np>	origins of <np>
ActVP Prep <np>	evolved from <np>
PassVP Prep <np>	was replaced by <np>
InfVP Prep <np>	to use as <np>
<possessive> NP	<possessive> son

Figure 3: The Pattern Templates of AutoSlog-TS with Example Instantiations

In addition to the original 17 pattern templates in AutoSlog-TS (shown in Figure 3), we defined 7 new pattern templates for the following bigrams and trigrams: Adj Noun, Adj Conj Adj, Adv Adv, Adv Adv Adv, Adj Adj, Adv Adj, Adv Adv Adj. We added these n-gram patterns to provide coverage for adjective and adverb expressions because the original templates were primarily designed to capture noun phrase and verb phrase expressions.

The learning process in AutoSlog-TS has two phases. In the first phase, the pattern templates are applied to the texts exhaustively, so that lexico-syntactic patterns are generated for (literally) every instantiation of the templates that appear in the corpus. In the second phase, AutoSlog-TS uses the la-

²The examples are shown as general expressions for readability, but the actual patterns must match the syntactic constraints associated with the pattern template.

bels associated with the texts to compute statistics for how often each pattern occurs in each class of texts. For each pattern p , we collect $P(\text{FACTUAL} | p)$ and $P(\text{FEELING} | p)$, as well as the pattern’s overall frequency in the corpus.

2.2.2 Bootstrapping Procedure

Since the IAC data set contains a large number of unannotated debate forum posts, we embed AutoSlog-TS in a bootstrapping framework to learn additional patterns. The flow diagram for the bootstrapping system is shown in Figure 4.

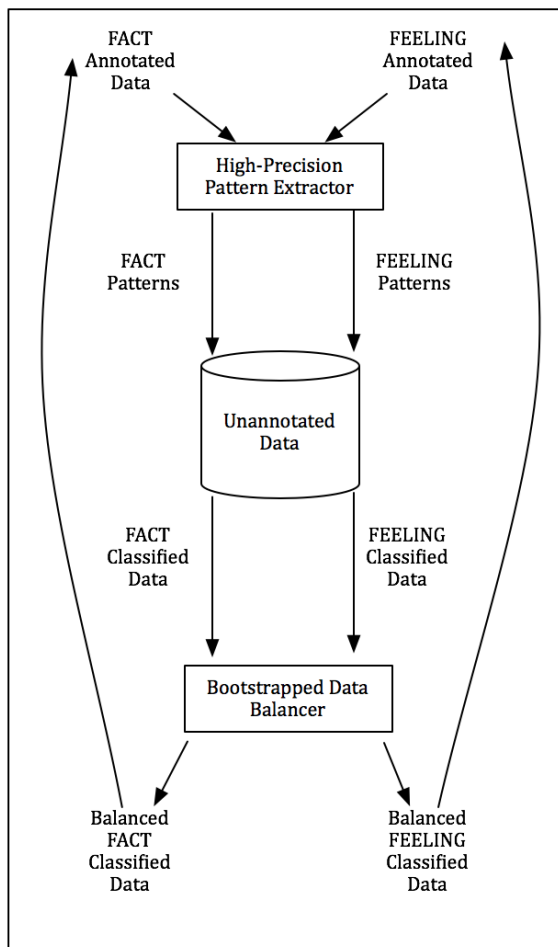


Figure 4: Flow Diagram for Bootstrapping Process

Initially, we give the labeled training data to AutoSlog-TS, which generates patterns and associated statistics. The next step identifies high-precision patterns that can be used to label some of the unannotated texts as FACTUAL or FEELING. We define two thresholds: θ_f to represent a mini-

imum frequency value, and θ_p to represent a minimum probability value. We found that using only a small set of patterns (when θ_p is set to a high value) achieves extremely high precision, yet results in a very low recall. Instead, we adopt a strategy of setting a moderate probability threshold to identify reasonably reliable patterns, but labeling a text as FACTUAL or FEELING only if it contains at least a certain number different patterns for that category, θ_n . In order to calibrate the thresholds, we experimented with a range of threshold values on the development (tuning) data and identified $\theta_f=3$, $\theta_p=.70$, and $\theta_n=3$ for the FACTUAL class, and $\theta_f=3$, $\theta_p=.55$, and $\theta_n=3$ for the FEELING class as having the highest classification precision (with non-trivial recall).

The high-precision patterns are then used in the bootstrapping framework to identify more FACTUAL and FEELING texts from the 11,561 unannotated posts, also from 4forums.com. For each round of bootstrapping, the current set of FACTUAL and FEELING patterns are matched against the unannotated texts, and posts that match at least 3 patterns associated with a given class are assigned to that class. As shown in Figure 4, the Bootstrapped Data Balancer then randomly selects a balanced subset of the newly classified posts to maintain the same proportion of FACTUAL vs. FEELING documents throughout the bootstrapping process. These new documents are added to the set of labeled documents, and the bootstrapping process repeats. We use the same threshold values to select new high-precision patterns for all iterations.

3 Evaluation

We evaluate the effectiveness of the learned patterns by applying them to the test set of 586 posts (347 FACT and 239 FEELING posts, maintaining the original ratio of FACT to FEEL data in train). We classify each post as FACTUAL or FEELING using the same procedure as during bootstrapping: a post is labeled as FACTUAL or FEELING if it matches at least three high-precision patterns for that category. If a document contains three patterns for both categories, then we leave it unlabeled. We ran the bootstrapping algorithm for four iterations.

The upper section of Table 1 shows the Precision and Recall results for the patterns learned dur-

ing bootstrapping. The Iter 0 row shows the performance of the patterns learned only from the original, annotated training data. The remaining rows show the results for the patterns learned from the unannotated texts during bootstrapping, added cumulatively. We show the results after each iteration of bootstrapping.

Table 1 shows that recall increases after each bootstrapping iteration, demonstrating that the patterns learned from the unannotated texts yield substantial gains in coverage over those learned only from the annotated texts. Recall increases from 22.8% to 40.9% for FACT, and from 8.0% to 18.8% for FEEL.³ The precision for the FACTUAL class is reasonably good, but the precision for the FEELING class is only moderate. However, although precision typically decreases during bootstrapping due to the addition of imperfectly labeled data, the precision drop during bootstrapping is relatively small.

We also evaluated the performance of a Naive Bayes (NB) classifier to assess the difficulty of this task with a traditional supervised learning algorithm. We trained a Naive Bayes classifier with unigram features and binary values on the training data, and identified the best Laplace smoothing parameter using the development data. The bottom row of Table 1 shows the results for the NB classifier on the test data. These results show that the NB classifier yields substantially higher recall for both categories, undoubtedly due to the fact that the classifier uses

³The decrease from 19.2% to 18.8% recall is probably due to more posts being labeled as relevant by *both* categories, in which case they are ultimately left unlabeled to avoid overlap.

Table 1: Evaluation Results

	Fact		Feel	
	Prec	Rec	Prec	Rec
Pattern-based Classification				
Iter 0	77.5	22.8	65.5	8.0
Iter 1	80.0	34.6	60.0	16.3
Iter 2	80.0	38.0	64.3	18.8
Iter 3	79.9	40.1	63.0	19.2
Iter 4	78.0	40.9	62.5	18.8
Naive Bayes Classifier				
NB	73.0	67.0	57.0	65.0

Table 2: Examples of Characteristic Argumentation Style Patterns for Each Class

Patt ID#	Probability	Frequency	Pattern	Text Match
FACT Selected Patterns				
FC1	1.00	18	NP Prep <np>	SPECIES OF
FC2	1.00	21	<subj> PassVP	EXPLANATION OF
FC3	1.00	20	<subj> AuxVP Dobj	BE EVIDENCE
FC4	1.00	14	<subj> PassVP	OBSERVED
FC5	0.97	39	NP Prep <np>	RESULT OF
FC6	0.90	10	<subj> ActVP Dobj	MAKE POINT
FC7	0.84	32	Adj Noun	SCIENTIFIC THEORY
FC8	0.75	4	NP Prep <np>	MISUNDERSTANDING OF
FC9	0.67	3	Adj Noun	FUNDAMENTAL RIGHTS
FC10	0.50	2	NP Prep <np>	MEASURABLE AMOUNT
FEEL Selected Patterns				
FE1	1.00	14	Adj Noun	MY ARGUMENT
FE2	1.00	7	<subj> AuxVP Adj	BE ABSURD
FE3	1.00	9	Adv Adj	MORALLY WRONG
FE4	0.91	11	<subj> AuxVP Adj	BE SAD
FE5	0.89	9	<subj> AuxVP Adj	BE DUMB
FE6	0.89	9	Adj Noun	NO BRAIN
FE7	0.81	37	Adj Noun	COMMON SENSE
FE8	0.75	8	InfVP Prep <np>	BELIEVE IN
FE9	0.87	3	Adj Noun	ANY CREDIBILITY
FE10	0.53	17	Adj Noun	YOUR OPINION

all unigram information available in the text. Our pattern learner, however, was restricted to learning linguistic expressions in specific syntactic constructions, usually requiring more than one word, because our goal was to study *specific* expressions associated with FACTUAL and FEELING argument styles. Table 1 shows that the lexico-syntactic patterns did obtain higher precision than the NB classifier, but with lower recall.

Table 3: Number of New Patterns Added after Each Round of Bootstrapping

	FACT	FEEL	Total
Iter 0	1,212	662	1,874
Iter 1	2,170	1,609	3,779
Iter 2	2,522	1,728	4,250
Iter 3	3,147	2,037	5,184
Iter 4	3,696	2,134	5,830

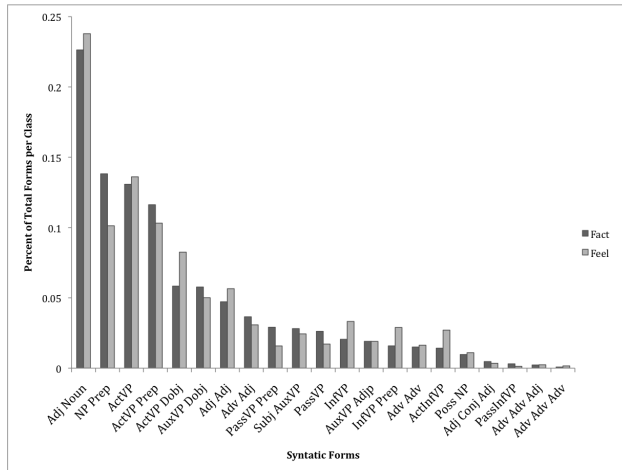
Table 3 shows the number of patterns learned from the annotated data (Iter 0) and the number of new patterns added after each bootstrapping iteration. The first iteration dramatically increases the

set of patterns, and more patterns are steadily added throughout the rest of bootstrapping process.

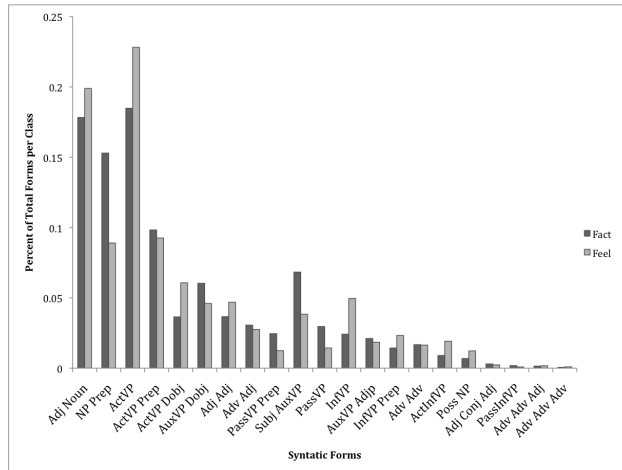
The key take-away from this set of experiments is that distinguishing FACTUAL and FEELING arguments is clearly a challenging task. There is substantial room for improvement for both precision and recall, and surprisingly, the FEELING class seems to be harder to accurately recognize than the FACTUAL class. In the next section, we examine the learned patterns and their syntactic forms to better understand the language used in the debate forums.

4 Analysis

Table 2 provides examples of patterns learned for each class that are characteristic of that class. We observe that patterns associated with factual arguments often include topic-specific terminology, explanatory language, and argument phrases. In contrast, the patterns associated with feeling based arguments are often based on the speaker’s own beliefs or claims, perhaps assuming that they themselves are credible (Chaiken, 1980; Petty et al., 1981), or they involve assessment or evaluations of the arguments



(a) Percentage of Each Unique Syntactic Form



(b) Percentage of Each Syntactic Form, by Instance Counts

Figure 5: Histograms of Syntactic Forms by Percentage of Total

of the other speaker (Hassan et al., 2010). They are typically also very creative and diverse, which may be why it is hard to get higher accuracies for FEELING classification, as shown by Table 1.

Figure 5 shows the distribution of syntactic forms (templates) among all of the high-precision patterns identified for each class during bootstrapping. The x-axes show the syntactic templates⁴ and the y-axes show the percentage of all patterns that had a specific syntactic form. Figure 5a counts each lexico-syntactic pattern only once, regardless of how many times it occurred in the data set. Figure 5b counts the number of instances of each lexico-syntactic pattern. For example, Figure 5a shows that the *Adj Noun* syntactic form produced 1,400 different patterns, which comprise 22.6% of the distinct patterns learned. Figure 5b captures the fact that there are 7,170 instances of the *Adj Noun* patterns, which comprise 17.8% of all patterns instances in the data set.

For FACTUAL arguments, we see that patterns with prepositional phrases (especially *NP Prep*) and passive voice verb phrases are more common. Instantiations of *NP Prep* are illustrated by **FC1**, **FC5**, **FC8**, **FC10** in Table 2. Instantiations of *PassVP* are illustrated by **FC2** and **FC4** in Table 2. For FEELING arguments, expressions with adjectives and active voice verb phrases are more common. Almost every high probability pattern for FEELING includes

⁴We grouped a few of the comparable syntactic forms together for the purposes of this graph.

an adjective, as illustrated by every pattern **except FE8** in Table 2. Figure 5b shows that three syntactic forms account for a large proportion of the instances of high-precision patterns in the data: *Adj Noun*, *NP Prep*, and *ActVP*.

Next, we further examine the *NP Prep* patterns since they are so prevalent. Figure 6 shows the percentages of the most frequently occurring prepositions found in the *NP Prep* patterns learned for each class. Patterns containing the preposition “of” make up the vast majority of prepositional phrases for both the FACT and FEEL classes, but is more common in the FACT class. In contrast, we observe that

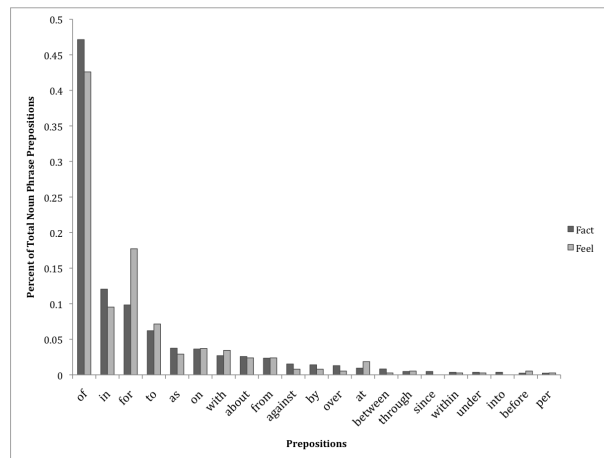


Figure 6: Percentage of Preposition Types in the *NP Prep* Patterns

patterns with the preposition “for” are substantially more common in the FEEL class than the FACT class.

Table 4 shows examples of learned *NP Prep* patterns with the preposition “of” in the FACT class and “for” in the FEEL class. The “of” preposition in the factual arguments often attaches to objective terminology. The “for” preposition in the feeling-based arguments is commonly used to express advocacy (e.g., *demand for*) or refer to affected population groups (e.g., *treatment for*). Interestingly, these phrases are subtle indicators of feeling-based arguments rather than explicit expressions of emotion or sentiment.

Table 4: High-Probability FACT Phrases with “OF” and FEEL Phrases with “FOR”

FACT “OF” Phrases	FEEL “FOR” Phrases
RESULT OF	MARRIAGE FOR
ORIGIN OF	STANDING FOR
THEORY OF	SAME FOR
EVIDENCE OF	TREATMENT FOR
PARTS OF	DEMAND FOR
EVOLUTION OF	ATTENTION FOR
PERCENT OF	ADVOCATE FOR
THOUSANDS OF	NO EVIDENCE FOR
EXAMPLE OF	JUSTIFICATION FOR
LAW OF	EXCUSE FOR

5 Related Work

Related research on argumentation has primarily worked with different genres of argument than found in IAC, such as news articles, weblogs, legal briefs, supreme court summaries, and congressional debates (Marwell and Schmitt, 1967; Thomas et al., 2006; Burfoot, 2008; Cialdini, 2000; McAlister et al., 2014; Reed and Rowe, 2004). The examples from IAC in Figure 1 illustrate that natural informal dialogues such as those found in online forums exhibit a much broader range of argumentative styles. Other work has on models of natural informal arguments have focused on stance classification (Somasundaran and Wiebe, 2009; Somasundaran and Wiebe, 2010; Walker et al., 2012), argument summarization (Misra et al., 2015), sarcasm detection (Justo et al., 2014), and identifying the structure of arguments such as main claims and their justifications (Biran and Rambow, 2011; Purpura et al.,

2008; Yang and Cardie, 2013).

Other types of language data also typically contains a mixture of subjective and objective sentences, e.g. Wiebe et al. (2001; 2004) found that 44% of sentences in a news corpus were subjective. Our work is also related to research on distinguishing subjective and objective text (Yu and Hatzivassiloglou, 2003; Riloff et al., 2005; Wiebe and Riloff, 2005), including bootstrapped pattern learning for subjective/objective sentence classification (Riloff and Wiebe, 2003). However, prior work has primarily focused on news texts, not argumentation, and the notion of objective language is not exactly the same as factual. Our work also aims to recognize emotional language specifically, rather than all forms of subjective language. There has been substantial work on sentiment and opinion analysis (e.g., (Pang et al., 2002; Kim and Hovy, 2004; Wilson et al., 2005; Bethard et al., 2005; Wilson et al., 2006; Yang and Cardie, 2014)) and recognition of specific emotions in text (Mohammad, 2012a; Mohammad, 2012b; Roberts et al., 2012; Qadir and Riloff, 2013), which could be incorporated in future extensions of our work. We also hope to examine more closely the relationship of this work to previous work aimed at the identification of nasty vs. nice arguments in the IAC (Lukin and Walker, 2013; Justo et al., 2014).

6 Conclusion

In this paper, we use observed differences in argumentation styles in online debate forums to extract patterns that are highly correlated with factual and emotional argumentation. From an annotated set of forum post responses, we are able extract high-precision patterns that are associated with the argumentation style classes, and we are then able to use these patterns to get a larger set of indicative patterns using a bootstrapping methodology on a set of unannotated posts.

From the learned patterns, we derive some characteristic syntactic forms associated with the FACT and FEEL that we use to discriminate between the classes. We observe distinctions between the way that different arguments are expressed, with respect to the technical and more opinionated terminologies used, which we analyze on the basis of grammatical

forms and more direct syntactic patterns, such as the use of different prepositional phrases. Overall, we demonstrate how the learned patterns can be used to more precisely gather similarly-styled argument responses from a pool of unannotated responses, carrying the characteristics of factual and emotional argumentation style.

In future work we aim to use these insights about argument structure to produce higher performing classifiers for identifying FACTUAL vs. FEELING argument styles. We also hope to understand in more detail the relationship between these argument styles and the heuristic routes to persuasion and associated strategies that have been identified in previous work on argumentation and persuasion (Marwell and Schmitt, 1967; Cialdini, 2000; Reed and Rowe, 2004).

Acknowledgments

This work was funded by NSF Grant IIS-1302668-002 under the Robust Intelligence Program. The collection and annotation of the IAC corpus was supported by an award from NPS-BAA-03 to UCSC and an IARPA Grant under the Social Constructs in Language Program to UCSC by subcontract from the University of Maryland.

References

- W.L. Benoit. 1987. Argument and credibility appeals in persuasion. *Southern Speech Communication Journal*, 42(2):181–97.
- S. Bethard, H. Yu, A. Thornton, V. Hatzivassiloglou, and D. Jurafsky. 2005. Automatic Extraction of Opinion Propositions and their Holders. In *Computing Attitude and Affect in Text: Theory and Applications*. Springer.
- O. Biran and O. Rambow. 2011. Identifying justifications in written dialogs. In *2011 Fifth IEEE International Conference on Semantic Computing (ICSC)*, pages 162–168. IEEE.
- C. Burfoot. 2008. Using multiple sources of agreement information for sentiment classification of political transcripts. In *Australasian Language Technology Association Workshop 2008*, volume 6, pages 11–18.
- J.T. Cacioppo, R.E. Petty, and K.J. Morris. 1983. Effects of need for cognition on message evaluation, recall, and persuasion. *Journal of Personality and Social Psychology*, 45(4):805.
- S. Chaiken. 1980. Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of personality and social psychology*, 39(5):752.
- Robert B. Cialdini. 2000. *Influence: Science and Practice (4th Edition)*. Allyn & Bacon.
- J. E. Fox Tree. 2010. Discourse markers across speakers and settings. *Language and Linguistics Compass*, 3(1):1–13.
- A. Hassan, V. Qazvinian, and D. Radev. 2010. What’s with the attitude?: identifying sentences with attitude in online discussions. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 1245–1255. Association for Computational Linguistics.
- John E. Hunter. 1987. A model of compliance-gaining message selection. *Communication Monographs*, 54(1):54–63.
- Raquel Justo, Thomas Corcoran, Stephanie M Lukin, Marilyn Walker, and M Inés Torres. 2014. Extracting relevant knowledge for the detection of sarcasm and nastiness in the social web. *Knowledge-Based Systems*, 69:124–133.
- Soo-Min Kim and Eduard Hovy. 2004. Determining the sentiment of opinions. In *Proceedings of the 20th International Conference on Computational Linguistics (COLING 2004)*, pages 1267–1373, Geneva, Switzerland.
- Stephanie Lukin and Marilyn Walker. 2013. Really? well. apparently bootstrapping improves the performance of sarcasm and nastiness classifiers for online dialogue. *NAACL 2013*, page 30.
- Fabrizio Macagno and Douglas Walton. 2014. *Emotive language in argumentation*. Cambridge University Press.
- G. Marwell and D. Schmitt. 1967. Dimensions of compliance-gaining behavior: An empirical analysis. *sociometry*, 30:350–364.
- Simon McAlister, Colin Allen, Andrew Ravenscroft, Chris Reed, David Bourget, John Lawrence, Katy Börner, and Robert Light. 2014. From big data to argument analysis. *Intelligence*, page 27.
- Amita Misra, Pranav Anand, Jean E. Fox Tree, and Marilyn Walker. 2015. Using summarization to discover argument facets in dialog. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.
- Saif Mohammad. 2012a. #emotional tweets. In **SEM 2012: The First Joint Conference on Lexical and Computational Semantics*.
- Saif Mohammad. 2012b. Portable features for classifying emotional text. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.

- Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 79–86.
- Joonsuk Park and Claire Cardie. 2014. Identifying appropriate support for propositions in online user comments. *ACL 2014*, page 29.
- R.E. Petty and J.T. Cacioppo. 1986. The elaboration likelihood model of persuasion. *Advances in experimental social psychology*, 19(1):123–205.
- R.E. Petty, J.T. Cacioppo, and R. Goldman. 1981. Personal involvement as a determinant of argument-based persuasion. *Journal of Personality and Social Psychology*, 41(5):847.
- S. Purpura, C. Cardie, and J. Simons. 2008. Active learning for e-rulemaking: Public comment categorization. In *Proceedings of the 2008 international conference on Digital government research*, pages 234–243. Digital Government Society of North America.
- Ashequl Qadir and Ellen Riloff. 2013. Bootstrapped learning of emotion hashtags# hashtags4you. In *Proceedings of the 4th workshop on computational approaches to subjectivity, sentiment and social media analysis*, pages 2–11.
- Chris Reed and Glenn Rowe. 2004. Araucaria: Software for argument analysis, diagramming and representation. *International Journal on Artificial Intelligence Tools*, 13(04):961–979.
- Ellen Riloff and William Phillips. 2004. An introduction to the sundance and autoslog systems. Technical report, Technical Report UUCS-04-015, School of Computing, University of Utah.
- E. Riloff and J. Wiebe. 2003. Learning Extraction Patterns for Subjective Expressions. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*.
- E. Riloff, J. Wiebe, and W. Phillips. 2005. Exploiting Subjectivity Classification to Improve Information Extraction. In *Proceedings of the 20th National Conference on Artificial Intelligence*.
- Ellen Riloff. 1996. Automatically generating extraction patterns from untagged text. In *AAAI/IAAI, Vol. 2*, pages 1044–1049.
- Kirk Roberts, Michael A. Roach, Joseph Johnson, Josh Guthrie, and Sanda M. Harabagiu. 2012. Empatweet: Annotating and detecting emotions on twitter. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012)*. ACL Anthology Identifier: L12-1059.
- Ariel Rosenfeld and Sarit Kraus. 2015. Providing arguments in discussions based on the prediction of human argumentative behavior. *AAAI*.
- Roser Saurí and James Pustejovsky. 2009. Factbank: A corpus annotated with event factuality. *Language resources and evaluation*, 43(3):227–268.
- R. Snow, B. O’Connor, D. Jurafsky, and A.Y. Ng. 2008. Cheap and fast—but is it good?: evaluating non-expert annotations for natural language tasks. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 254–263. Association for Computational Linguistics.
- S. Somasundaran and J. Wiebe. 2009. Recognizing stances in online debates. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1-Volume 1*, pages 226–234. Association for Computational Linguistics.
- S. Somasundaran and J. Wiebe. 2010. Recognizing stances in ideological on-line debates. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 116–124. Association for Computational Linguistics.
- M. Thomas, B. Pang, and L. Lee. 2006. Get out the vote: Determining support or opposition from Congressional floor-debate transcripts. In *Proceedings of the 2006 conference on empirical methods in natural language processing*, pages 327–335. Association for Computational Linguistics.
- Marilyn Walker, Pranav Anand, , Robert Abbott, and Jean E. Fox Tree. 2012. A corpus for research on deliberation and debate. In *Language Resources and Evaluation Conference, LREC2012*.
- Douglas Walton. 2010. *The place of emotion in argument*. Penn State Press.
- J. Wiebe and E. Riloff. 2005. Creating Subjective and Objective Sentence Classifiers from Unannotated Texts. In *Proceedings of the 6th International Conference on Computational Linguistics and Intelligent Text Processing*, pages 486–497, Mexico City, Mexico, February.
- Janyce Wiebe, Theresa Wilson, and Matthew Bell. 2001. Identifying collocations for recognizing opinions. In *Proceedings of the ACL-01 Workshop on Collocation: Computational Extraction, Analysis, and Exploitation*, pages 24–31, Toulouse, France.
- Janyce Wiebe, Theresa Wilson, Rebecca Bruce, Matthew Bell, and Melanie Martin. 2004. Learning subjective language. *Computational Linguistics*, 30(3):277–308.
- Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the 2005 Human Language Technology Conference / Conference on Empirical Methods in Natural Language Processing*.

- T. Wilson, J. Wiebe, and R. Hwa. 2006. Recognizing strong and weak opinion clauses. *Computational Intelligence*, 22(2):73–99.
- Bishan Yang and Claire Cardie. 2013. Joint inference for fine-grained opinion extraction. In *ACL (1)*, pages 1640–1649.
- B. Yang and C. Cardie. 2014. Context-aware learning for sentence-level sentiment analysis with posterior regularization. In *Proceedings of the Association for Computational Linguistics (ACL)*.
- Hong Yu and Vasileios Hatzivassiloglou. 2003. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*, pages 129–136, Sapporo, Japan.
- Inon Zuckerman, Erel Segal-Halevi, Avi Rosenfeld, and Sarit Kraus. 2015. First steps in chat-based negotiating agents. In *Next Frontier in Agent-based Complex Automated Negotiation*, pages 89–109. Springer.