# Emotional Expressions in Online Settings: How I Spun Down a Data Mining Hole of Emotion

Ryan Compton

## 1. Project Rationale and Framing

Online communities are prevalent and important (Kraut and Resnick 2012). Millions of people use them each day in their personal and working lives. Communities are often used to solicit advice and receive social support. Recent work has begun to explore community interaction examining the content of members' conversation (Nguyen and Rose 2011). A key question concerns the impact of emotional versus factual language. In some communities, emotional support for other members is critical (Matthews et al. 2015). In these communities it may be vital to respond to the affective content of a post rather than the factual information that the poster is nominally requesting. This interpersonal focus may then promote long-term relations between members (Muller et al. 2012). In contrast, other communities serve short-term informational needs, where the most common interaction is a simple information request from a first-time poster where a factual response is optimal (Nguyen and Rose 2011; Kraut and Resnick 2012). The key takeaway from this previous work is that context is key when expressing emotional content and whether that emotional content is going to be considered beneficial to the community. What remains is a comparison across communities in how emotional expression is conducted, as well as examining how emotional expressions affects discussion.

In this work, I will explore emotional expression within and across two online communities. I intend to do a cross-domain analysis of two different Conflict and Collaborative focused communities. The first being gathered from the Internet Argumentation Corpus (Walker, M.A. and et al. 2012) and the second being an intranet online enterprise community.

*IAC:*

> This online community is where people go to discuss and argue for a popular political issue. This online community typically discusses societal issues such as abortion, religion, immigration, gay marriage and other issues of a similar nature. The societal significance of these issues leads to engaged debate in which both factual and emotional language are overt and prevalent. The corpus holds responses to forums, particularly responses to previous posts within the forum. These responses are unique in that they contain the previous post they are responding to within a quote. The corpus annotates Factual vs Emotional language for each post response on a scale ranging from -5 to +5, shown within Fig 1. Each forum response was annotated by 5-7 annotators.

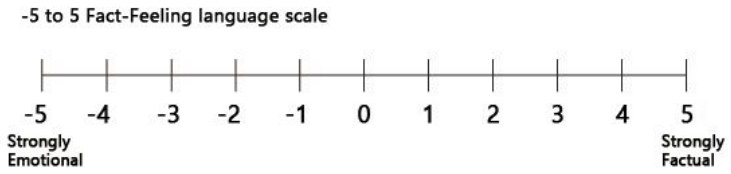


*Figure 1. Scale for Human Annotations for Factual and Emotional Content*

*Enterprise Communities:*

This online community database collected the online community activity for roughly 5 years. It contains the posts of users within separate communities they were able to join. Each community has a set of social tools they can use: Forums, Blogs, Wikis. Similar to the IAC corpus described above, a sample of posts within these tools were annotated for Factual vs. Emotional language on the same scale and same criteria. This was part of a previous study examining how emotional expression is similar across the two databases. Another unique aspect of this database is that each user has been given a label that corresponds to their social role within the online community. So a user is labeled as an owner or a member. Owners are given more community access and allowed to lead the community in activities and moderation, so it is expected to see large differences in owner and member behaviors.

These databases allow for some interesting exploration into how people interact in different online settings. The IAC corpus gives a sample of conflicting online community interactions, while the Enterprise Communities database mainly shows cooperation within online communities. Furthermore, the aspect of examining forums and their responses can allow for naturalistic experimentation on how people react to other's posts and how emotional of a response they give, since these databases have the emotional annotation. Therefore my goal for this project is:

**Goal:** Explore the differences that arise in emotional expression between conflict and cooperation online communities.

This goal has multiple steps of evaluation:

**1.** A comparison of emotional markers is needed, which can be done through Natural Language Processing. This will entail examining how the text differs between emotional expressions in each type of community (Cooperative or Conflicting). The text can be examined based on lexical and syntactical features. Differences can be magnified by overfitting a model that is intending to predict the annotated emotional level that exists in each separate database and examine how the linguistic features are affecting each model.

**2.** Each dataset provides an interesting avenue of exploration into emotional expression. Due to the role labeling within the Enterprise Communities database, I can examine differences in

discussions based on role distributions. For instance, how does a member respond to another member's post compared to that of a leader's post? Role status may exhibit a difference in emotionality that is expressed within the response from a member. This setup gives a naturalistic set-up to experiment.

**3.** The IAC corpus contains not only Factual v. Emotional annotations, but also Agreement vs Disagreement, Negotiate vs. Attack, Nice vs Nasty, and Personal vs Audience annotations to understand different aspects of the response. These difference annotation allow for more complex explorations into how people are responding to previous posts. First, the differences in what prompts this type of reaction can be explored due to the presence of the previous post in this database. Second, the level of emotion can be combined with other annotations (Nice vs Nasty) to enhance more than the level of emotion but potentially the type of emotional expression.

## 2. Related Work

Affective Computing is computing that relates to, arises from, or deliberately influence emotion or other affective phenomena. This involves the understanding of affectual expression through a digital medium. The majority of human to human interaction online is through text. Not only does text communicate information, but also emotional states (Alm et al., 2005).

Understanding emotional expression of text has been a challenge that researchers have been tackling for some time, each with a different level of emotions they are trying to model. Early work on sentiment detection made bimodal judgments about whether a given text expresses a positive versus negative evaluation. Sentiment detection uses a mixture of machine learning, lexicon based, and hybrid approaches. There is a vast amount of work focusing on various levels of granularity of Sentiment Analysis including Document, Word, Aspect, Sentence, Phrase, to Concept and Clause (Ravi and Ravi, 2015). It has been successfully applied in evaluating reviews for movies and consumer products as well as examining public mood and social media. Areas of research within sentiment detection have focused on Polarity determination, Vagueness Resolution , Multi-lingual and cross-lingual analysis, and cross-domain classification (Hai et al., 2014).

More recent work has extended bimodal positive vs negative distinctions, instead aiming to identify the presence vs absence of emotions in text. For example Aman et al. (2007) classifies whether a sentence is emotional or not, at 73% accuracy compared with human judges. Similarly Alm et al. (2005) developed and used both semantic and lexical machine learning features to obtain an accuracy of 69% in classifying neutral vs. emotional sentences.

Other work develops methods to recognize different types of emotions in text. One approach is to use expert judges to annotate texts for different emotions building a machine learning classifier for different affective states in social media, e.g. Joviality, Fear, Fatigue, Guilt, Hostility, Surprise, Sadness, Shyness, Serenity, Attentiveness, and Self-assurance. Some previous work of mine attempted to build a model that predicted a subset of these types of emotions and compared this to work done by Balahur et al. (2012).

| | Balahur Precision / Current Precision | Balahur Recall / Current Recall | Balahur F-Measure / Current F-Measure |
|---|---|---|---|
| Anger | 0.610 / 0.500 | 0.284 / 0.091 | 0.154 / 0.388 |
| Fear | 0.712 / 0.857 | 0.33 / 0.261 | 0.451 / 0.400 |
| Disgust | 0.692 / 0.000 | 0.202 / 0.000 | 0.313 / 0.000 |
| Happiness | 0.895 / 0.750 | 0.218 / 0.200 | 0.351 / 0.316 |
| Sadness | 0.336 / 0.750 | 0.895 / 0.222 | 0.489 / 0.343 |

*Table 1. Results Previous Emotional Modeling using Word based features compared to Balahur et al. (2012) results.*

While certain emotions had some decent levels of precision (Fear, Disgust, and Happiness), many of them had very low levels of recall. This indicates that for specific emotions, the model would signal that emotion when it was very clear that it was that emotion being expressed, but it left many text excerpts with an incorrect negative classification. This may be an indicator of the model only able to detect strong expressions of emotion while ignoring weaker expression but this avenue was not explored due to such low performance and lacking feature bases as this work was mainly using word based measures such as those used within previous work on sentiment detection.

Word based methods are often criticised for their limit capability to capture context, since they are ignoring how the word is used at a higher level. Modern methods have taken the approach of utilizing Appraisal Theory which was theorized by social scientists. Appraisal theory has the notion that “emotions are elicited and differentiated on the basis of the cognitive evaluation of the personal significance of a situation, object or event” (Ellsworth and Scherer, 2003). Work by Balahur et al. (2012) used this theory to argue that methods that didn’t account for the implicit expression of emotion, and only focused on the explicit expression. Therefore they focused on the interpretation of emotion through situations and built EmotiNet. Poria et al. (2014) iterated on that work by incorporating the six emotions within WordNet Affect into SenticNet concepts. They tested this system by building a model from these implicit emotional situations and found the model to be outperforming previous models of explicit emotional expression.

| Actual labels | Predicted labels | | | | | | Precision (%) | Recall (%) |
|---|---|---|---|---|---|---|---|---|
| | Surprise | Joy | Sadness | Anger | Fear | Disgust | | |
| Surprise | 39 | 4 | – | – | – | 1 | 85 | 89 |
| Joy | 3 | 91 | 2 | 3 | 1 | 1 | 91 | 90 |
| Sadness | – | 3 | 52 | 1 | – | – | 85 | 93 |
| Anger | 2 | 2 | 3 | 72 | 4 | 2 | 90 | 85 |
| Fear | – | – | 3 | 1 | 63 | 2 | 91 | 91 |
| Disgust | 2 | – | 1 | 3 | 1 | 34 | 85 | 85 |

*Table 2. Results of Emotion Modeling from Poria et al. 2014.*

## 3. Progress and Feedback

### Data Description

Initial exploration into the differences of each dataset is shown within table 3.

| | IAC Data | Enterprise Data |
|---|---|---|
| N | 8205 | 552 |
| Avg. Word Count | 36.12 | 41.87 |
| Avg. Words Per Sentence | 19.42 | 18.34 |
| Six Letter Words | 20.29 | 19.81 |

*Table 3. Differences across IAC and Enterprise datasets.*

While the IAC has a larger amount of annotated posts, there aren't too many differences in the quantity or length of each post across the datasets, with the main difference being Enterprise posts having more words in a post.

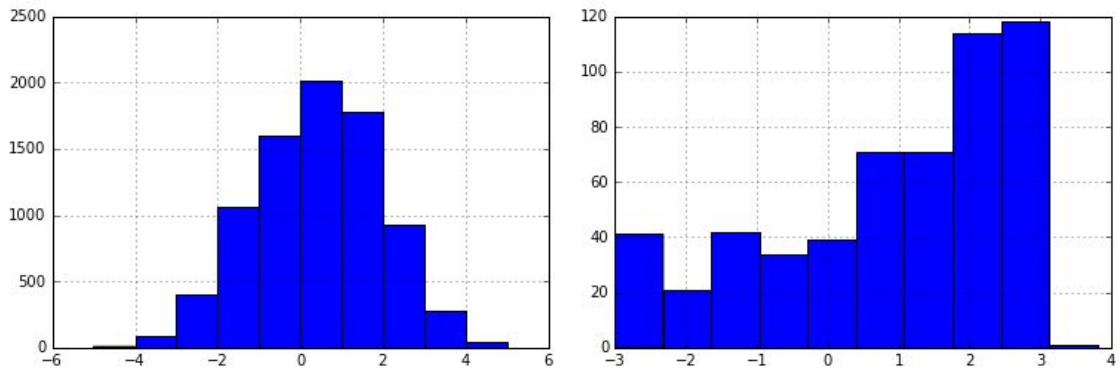


*Figure 2. Distribution of Fact-Feeling Annotation across both the IAC(Left) and Enterprise Communities(Right)*

Figure 2 shows the differences in the distribution of the Fact-Feeling annotation of posts. The IAC has a much more normalized distribution centered around the neutral value of 0, while within Enterprise Communities there is a strong skew to being more factual, which makes sense given the context differences.

**Features Used**

To accomplish the first goal of this work, a model was developed to predict the Fact-Feeling annotation for a post. Two models were developed, one for the IAC and one for Enterprise Communities. Each model had the same feature set, which was a combination of Lexical and Syntactical measures of the post content. A list of sample features is described below in Table 4.

*Lexical Features:*

LIWC v2007 (Tausczik and Pennebaker 2010) is a lexicon that provides frequency counts of words that signify important psychological constructs, as well as some relevant topics (e.g. Leisure, Work). LIWC is widely used and reliable compared with human judges (Tausczik and Pennebaker 2010; Wang et al. 2012; Matthews et al. 2015). The LIWC dictionary defines 81

word categories, each containing multiple words. It indexes categories such as pronouns ('I', 'you'), as well as words with psychological relevance, e.g. that express positive and negative emotion or verbs of cognition. Categories are not exclusive; so words can belong to multiple categories.

The Emotion lexicon (EmoLex) (Mohammad and Turney 2010) is specifically focused on emotional terms. It contains 14182 words classified into 10 emotional categories: Anger, Anticipation, Disgust, Fear, Joy, Negative, Positive, Sadness, Surprise, and Trust.

Our final lexicon was less directly concerned with emotions. Instead it was focused on whether words expressed positive or negative sentiment. The Subjectivity Lexicon is part of OpinionFinder (Wilson et al. 2005). It consists of 8222 stemmed and un-stemmed words annotated by a group of trained annotators as either strongly or weakly subjective. Subjectivity has been found as a useful lexicon for analyzing sentiment (Ravi and Ravi 2015).

*Syntactic Features:*
Lexicon based approaches have limitations. They use a simple "bag of words" which assumes that social and psychological meaning can be derived from individual words alone. This ignores syntax, punctuation, conversational structure, and other relational features of text. We therefore also included structural features of language use, such as use of questions and grammatical tense that might also signal emotional or factual expression. Syntactic choices show an emphasis on concepts (nouns) versus actions (verbs) (Pennebaker 2013). Syntax also indicates a focus on past, present or future. We therefore used a part of speech (POS) tagger to count the relative frequencies of nouns, verbs, adjectives and adverbs, use of questions as well as tense and aspect information (Toutanova et al. 2003).

**LEXICAL FEATURES:**

**LIWC**:

- **Pronoun**: I, we, you, she/he, they, impersonal pronoun
- **Tense**: auxiliary verb, past, present, future
- **Affect:** positive emotion, negative emotion (anxiety, anger, sadness)
- **Topic**: cognitive mechanism, biological processes, time, religion, death

**Emotion Lexicon**:

- **Anger**: contraband, hate, idiot
- **Anticipation**: convergence, labor, journey
- **Disgust**: abject, decompose, lagging
- **Fear**: hysteria, intimidate, kerosene
- **Joy**: kudos, kiss, jubilee
- **Sadness**: abandon, abortion, intolerant
- **Surprise:** abrupt, decoy, invade
- **Trust**: convince, deed, kindred

**Subjective**:

- **Weak Subjective**: achieve, discreet, involuntary
- **Strong Subjective**: accusation, dread, oppression

**SYNTACTIC FEATURES**

**Part of Speech Features**:

- **Adjective, Noun, Adverb, Verb (Past, Present participle)**: Frequencies

*Table 4. Examples of Lexical and Linguistic Features*

**Modeling**

A gradient boosted random forest (Parameters: 1000 trees with a max depth of 7) was used for a regression model which outputs a scalar emotionality evaluation, which better represents the given input feature representation for each post. Previous work has examined conducting a simple binary classification of Factual or Emotional class by splitting the data at the 0 mark, however this lead to more signal loss and less representation of the data at hand. Each dataset was split into a Train-Test split of 66% Training data and 33% Testing data.

| | IAC Data | Enterprise Data |
|---|---|---|
| $R^2$ | 0.163 | 0.621 |
| Mean Squared Error | 1.939 | 1.233 |

*Table 5. Evaluation Metrics for Regression Models of Fact-Feeling between Datasets*

Table 5 shows the results of each model and immediately it is clear how well the model of Enterprise Communities is over the IAC model, as it has an $R^2$ of 0.621 and the IAC is only around 0.163. This shows that perhaps the level of emotional evaluation from the annotators within the IAC have a more complex level of emotional expression than that of the Enterprise annotations. Secondly, mean squared error is given to compare the variance of the model with the variance of the annotations. The worst model (IAC model) is varying around 17% in its predictions (given there were 11 possible values for the annotators to choose). In comparison, human annotators had a standard deviation of 2.08 for all posts, thus the model is varying in a way that is comparable with the overall judgements of a group of annotators.

**Feature Exploration**

Next involved exploring how each feature is influencing the models. This will answer the question of how each dataset differs in which features are important in emotional expression. Figure 3 below shows one plot of all the feature importances within the Enterprise model, this mainly shows that the level of feature importance is rather skewed to just a few features, with the top feature being the number of words within a post. Table 6 shows the comparison between the top 10 features of the two models. The first 5 are similar (Word Count, Words per Sentence, 6 letter words, Number of LIWC dictionary hits, and Function words) which show that the complexity of the post gives an indication that the level of emotional expression vs factual expression and lies outside the context of the conversation. However there are some differences between the list of features, as LIWC-Articles and Syntactic Determiners were found to be more important within the IAC and Subjectivity and LIWC-Affect were found to be important within Enterprise Communities. As Subjectivity and LIWC-Affect are derived from a set lexicon, it appears that formally understood word usage may be more important within the Enterprise context as opposed to the more open internet types of discussions that exist within the IAC. This may be due to the nature of the more formal business oriented discussions taking place.

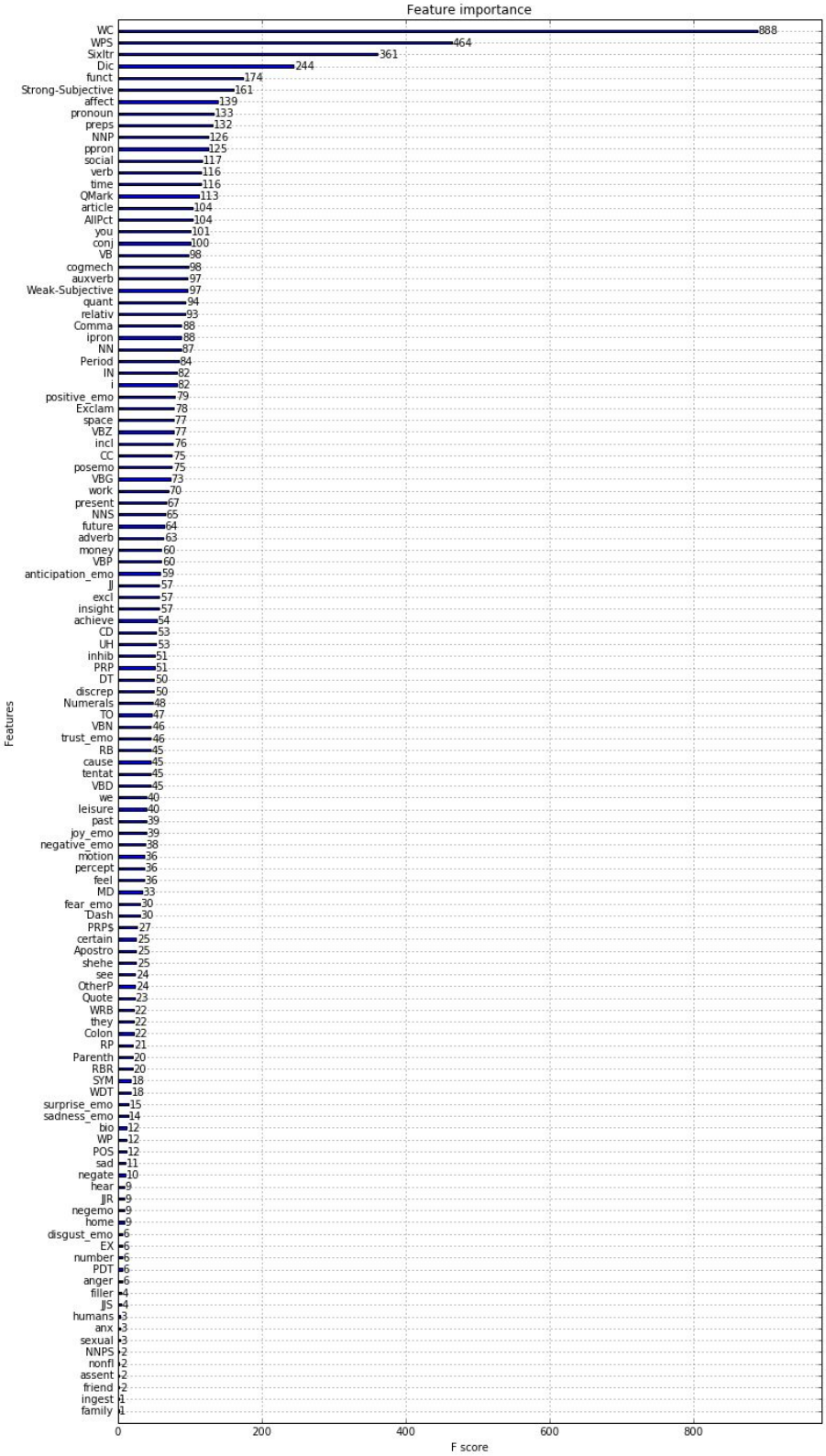


*Figure 3. Relative Importance metrics for each feature in the Enterprise Communities Model*

| Forums Features | RI | Enterprise Features | RI |
|---|---|---|---|
| Word Count | 0.0794 | Word Count | 0.1133 |
| Words Per Sentence | 0.0394 | Words Per Sentence | 0.0592 |
| Six Letter Words | 0.0331 | Six Letter Words | 0.0460 |
| LIWC Dictionary Hit | 0.0308 | LIWC Dictionary Hit | 0.0311 |
| Function Words | 0.0220 | Function Words | 0.0222 |
| Articles | 0.0183 | Strong-Subjective | 0.0205 |
| Pronouns | 0.0172 | Affect | 0.0177 |
| Impressional Pronouns | 0.0154 | Pronouns | 0.0169 |
| Determiners | 0.0153 | Prepositions | 0.0168 |
| Noun, singular or mass | 0.0150 | Proper noun, singular | 0.0160 |

*Table 6. Top 10 features for each model which were ranked by Relative Importance(RI)*

To further explore each features importance, an experiment was run to find how many features are needed to create the best model. Figure 4 shows how the performance of both models as a feature importance threshold is used to limit the number of features used. Both models have a slight drop in performance when the threshold is around 0.01, however the main drop occurs when the threshold reaches around 0.02, which removes the majority of the features (95%). This indicates that the best model in terms of performance and reduced complexity is needed to have around a threshold of 0.005-0.01.

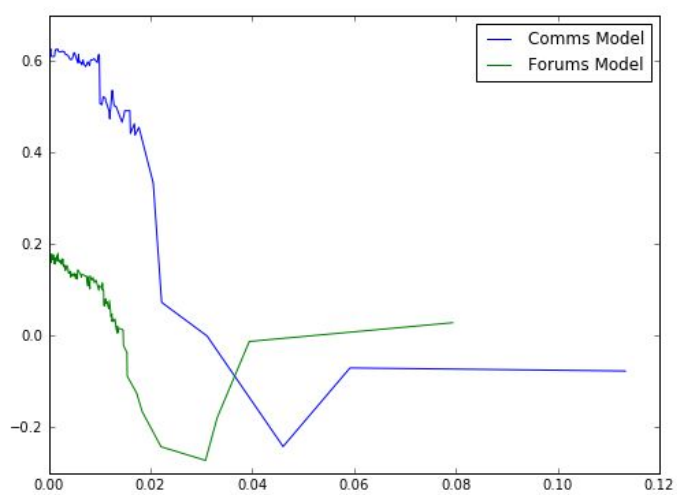


*Figure 4. Performance of Models($R^2$) over Relative Importance Threshold.*

**Implicit Emotions**

It has been argued that only examining explicit emotional expression is a reduction in how emotion express actually takes place (Balahur et al. 2012). Emotions are expressed through subtext or understanding a situation which don't typically have explicit descriptions of the emotional response. In order to test this, I used the EmoSenticNet tool developed by Poria et al. (2012). EmoSenticNet itself is like a lexicon that has a semantic frame that is annotated with an emotional response (6 Emotional Responses: Anger, Fear, Sad, Joy, Surprise, Disgust). I found the emotional output of EmoSenticNet on some posts within the IAC.

> ***Highest Joy***
> *The link between the current understanding of subsurface hydrology and evolution concerns the age of the water at a given point in a confined aquifer. The water in some confined aquifers is in the order of millions of years old.*
>
> ***Highest Anger***
> *"Accidents for one. Just last week a man who was drinking and cleaning his firearm accidentally shot and killed his daughter.*
> *Run a google search for ""man shoots self"" the amount of accidental shootings is rather disturbing yet funny. Most of these how"*
>
> ***Highest Sadness***
> *If abortion is made illegal I predict that you will see across the border abortion clinics, in canada and states that still premit abortion. Are we to fill the prisons with those seeking abortions in places where it is legal if they live in states that do*
>
> ***Miss (Lots of Joy, little Sadness and Fear)***
> *So they (pro-life peeps) say abortion is murder. Of course, you can only murder people. So their rule is that fetuses (feti?) are people also. But then they can't explain why fetuses aren't listed on the US Census, or why there isn't a funeral every time*

When reading through these posts that are considered a high value of a certain emotional expression, it seems like the tool is missing in its judgements. For one, examining the highest joy post above doesn't seem to be actually conveying any joy and is actually highly descriptive, furthermore when the tool misses a classification it seems to miss it by a lot (shown in the Miss example above).

**Feedback:**

This work received some good feedback into how I can explore additional avenues within text measurements as well as outside the realm of machine learning or data science. First it was recommended that I look at the fictional literature community in order to see how they teach expressing emotional content within a fictional setting. This lead to many sources of thought but the best summary is the rule "Show Don't Tell", which was present within many sources (one for reference:
http://thewritepractice.com/how-to-show-not-tell-paranoia-hope-and-other-moods/ ).
The main points I found from this rule were that it is best to express emotion through the following:

1. Motivation: Make the motivation of the character known to the reader. This allows for the impact of events to be more clear to how it affects the character.
2. Action: Describe the actions that a character would due under a particular emotional state. An example given is about the descriptions of what a person who is paranoid would be doing when they are paranoid, something a reader can pick up on and understand their emotional state.
3. Description: This is not the description of the character but the description of what details the character is noticing when they are exploring or interacting with their setting.

This exploration of writing procedures helped me identify that there may be an importance to looking at how people are describing things within their posts. Maybe the level of adjectives (lexically and structurally) could be a good indicator of emotional language, which can be easily examined by going back to the feature exploration done above. Adjectives and adverbs, both structurally and lexically, were found to be within the top 20 features and around the 0.02 threshold found within the feature threshold experiment. These features were within the set of features that were removed right around the drop-off in performance that we see back within Figure 4 (X-Axis = 0.02 to find drop), which suggests that even though they aren't the strongest in feature importance, they still provide a strong amount of signal within both contexts.

The next piece of feedback I received was on another type of textual metric, the complexity of the post through measuring readability. There are many measures on readability but a commonly used metric is the Flesch Reading Ease:

$$206.835 - 1.015\left(\frac{\text{total words}}{\text{total sentences}}\right) - 84.6\left(\frac{\text{total syllables}}{\text{total words}}\right)$$

*Formula for Flesch Reading Ease*

This can be measured for each post and then examined for its relationship with the Fact-Feeling annotation. This can serve well in trying to understand differences across roles within the Enterprise dataset.

## 4. Final Results

**Social Role Effects on Emotional Language Use**

The second goal of this work was to examine if social roles have a different effect on the level of emotional expression. Since social roles are defined within the Enterprise dataset, this was the best context to explore this further. There are two social roles within Enterprise Communities, which are simply Members and Leaders. There is a third Unknown category which was the case when users were not classified either as a member or a leader. Figure 5 is a violin plot that shows the initial look into the average level of Fact-Feeling for each post made by a Social Role by plotting each role's distribution of Fact-Feeling posts. It is quite clear that there are no noticeable differences across the amount of factual or emotional expression by each role. This is very much against expectations as it is was expected to see leaders taking a more to the point posting style that avoided emotional expression as it may be seen as off topic with what the goal of the community is, as the majority of enterprise communities have a goal related to an output for their work. It appears that all roles are equally more Factually focused.

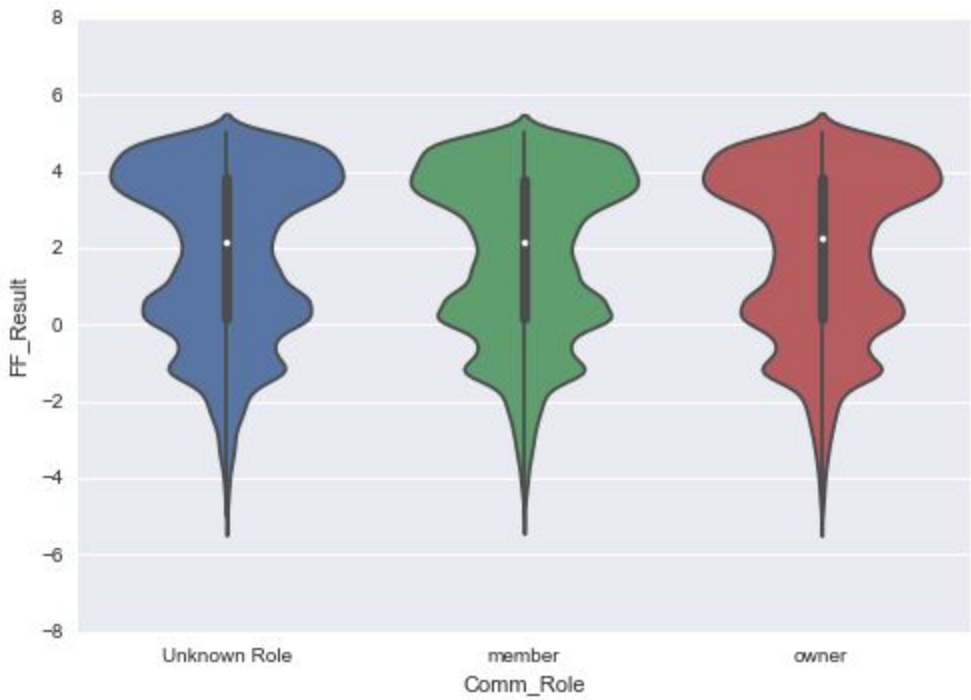


*Figure 5. Distribution of Fact-Feeling Measure across Social Roles in Enterprise Communities*

Next, the relationship between readability and the level of emotional language within a post was examined and within Table 7 you can see these results. Social role was given as a category to see if there are any differences across roles, however no role showed to have any relationship between readability and the emotional annotation.

| Role | Correlation of Readability and Fact Feeling |
|---|---|
| Overall | -0.04 |
| Member | -0.03 |
| Leader | -0.06 |

*Table 7. Correlation of Readability and Fact Feeling of Posts by Social Roles*

**Social Roles and Emotion within Discussion Threads**

I next wanted to look into the types of discussions that are sparked by each social role. I wanted to test whether there is a specific amount of emotional expression that occurs from either a leader or a member. As leaders are expected to be more formal, it is also expected that members will be more formal to match the culture of workplaces. Table 8 shows some descriptive statistics on the initial post that starts a discussion, which is compared across two types of tool sets that users can use within an Enterprise Community (Blogs and Forums).

| | Blogs | | Forums | |
|---|---|---|---|---|
| | Readability | Replies | Readability | Replies |
| Member | 47.070632 | 0.30824 | 63.22984 | 0.857861 |
| Leader | 27.289757 | 0.449678 | 60.4802 | 4.330299 |

*Table 8. Readability of Initial Blog and Forum post by Social Role and Average Number of Replies for each Initial Post.*

Within Blogs we see that Leaders are creating more complex initial posts than Members (lower the readability measure the more complex the post). This is also the case for Forums, however the effect is not nearly as strong. Looking at Tools, it found that Blogs tend to have more complex posts than Forums, which is the same regardless of which social role is posting. This is most likely due to the intention of a Blog post compared to a Forum post. A Blog post is a more opinion piece or personal narrative about a subject of interest, while within Forums the posts are meant to be more discussive, which we can see as Forum posts generally had more replies with Leaders producing the most replies.

| | Blogs | | | | Forums | | | |
|---|---|---|---|---|---|---|---|---|
| | LIWC Affective Processes | Fact Feeling Model | Emotion Lexicon | Readability | LIWC Affective Processes | Fact Feeling Model | Emotion Lexicon | Readability |
| Member Started Thread | 8.453 | 7.349 | 19.127 | 63.359 | 5.189 | 7.371 | 13.210 | 64.057 |
| Leader Started Thread | 8.249 | 7.394 | 18.946 | 66.071 | 5.747 | 7.108 | 14.544 | 62.215 |

*Table 9. Average Affect, Fact-Feeling Measure, and Readability of Blog and Forum Replies. Row Labels indicate if the Reply is to a Member Started Thread or a Leader Started Thread*

Table 9 shows the average amount of affectual language and complexity of Blog and Forum replies as opposed to initial posts. When compared to initial posts, replies are generally easier to read (high level of Readability measure) and this is across whichever role has started the thread. Across which role has started the thread, we find that there is more emotional replies to Leader started threads within Forums, which is not expected. This is seen through the higher frequency of LIWC Affect Processes and Emotion Lexicon hits and the Lower level of Fact Feeling Modeling Output (Lower means more emotional content). However this isn't the case

within Blogs.

| | Blogs | | | | Forums | | | |
|---|---|---|---|---|---|---|---|---|
| | LIWC Negative Emotion | LIWC Positive Emotion | EmoLex Negative | EmoLex Positive | LIWC Negative Emotion | LIWC Positive Emotion | EmoLex Negative | EmoLex Positive |
| Member Started Thread | 0.4842 | 7.9399 | 0.9638 | 6.6655 | 0.7112 | 4.4461 | 1.1792 | 4.1985 |
| Leader Started Thread | 0.4490 | 7.7655 | 0.9118 | 6.4430 | 0.5170 | 5.1959 | 0.9851 | 4.9891 |

*Table 10. Amount of Emotional Words (LIWC and Emotion Lexicon) within Replies to Member or Leader Started Thread*

To go deeper, I looked into the valence of emotions that are being expressed within the Blog and Forum replies, I looked into the amount of Positive and Negative words being used (Table 10). You can see that there is more Positive Emotion words being used within the Leader Started Forum Threads and enough of a difference that this may account for the total Forum emotional difference seen within Table 9. The differences in Blog emotional word use matches the overall trends shown in Table 9.

**Implicit Emotional Expression**

The next goal of this work is to explore how implicit emotional expression effects reply type posts. Due to the nature of the Quote and Response pairs within the IAC, this seemed the best avenue to explore how people respond to posts that exhibit a specific kind of implicit emotion. In order to do this, I measured the amount of implicit emotional expression from each Quote and then looked at how well that implicit emotion predicts the level of factual or emotional expression in the response. This is a direct test to see if emotional posts prompt an emotional response. Some expectations were that higher emotional posts would prompt higher emotional responses and each emotional valence would match.

| Adjusted $R^2$ | P-value |
|---|---|
| 0.002436 | 0.02591 |

| | Beta Weight | P-value |
|---|---|---|
| Anger | -0.0166 | 0.3788 |
| Disgust | 0.0033 | 0.8564 |
| Joy | 0.0178 | 0.3220 |
| Sad | -0.0429 | 0.0132 * |
| Surprise | 0.0260 | 0.1405 |
| Fear | -0.0298 | 0.1060 |

*Table 11. Regression results for predicting the Fact-Feeling annotation on IAC responses using the Implicit Emotions from quotes*

Table 11 shows the results of a regression analysis on the different types of implicit emotions predicting the Fact-Feeling annotation within the responses. These results were significant, however they show an extremely weak relationship indicating that implicit emotions don't prompt either factual or emotional responses. The only significant implicit emotion was Sadness and to further examine why this might be, a violin plot of the Fact-Feeling level in responses over the Count of Implicit Sad words or phrases (Figure 6). This figure actually shows that the reason for this relationship is most likely caused by only the very few posts that had 5 Sad words or phrases, which showed higher levels of Factual responses. Since there are so few number of posts that have this many Sad words or phrases, this seems to only be an effect found by a potential outlier therefore nullifying the significant effect in the regression results.

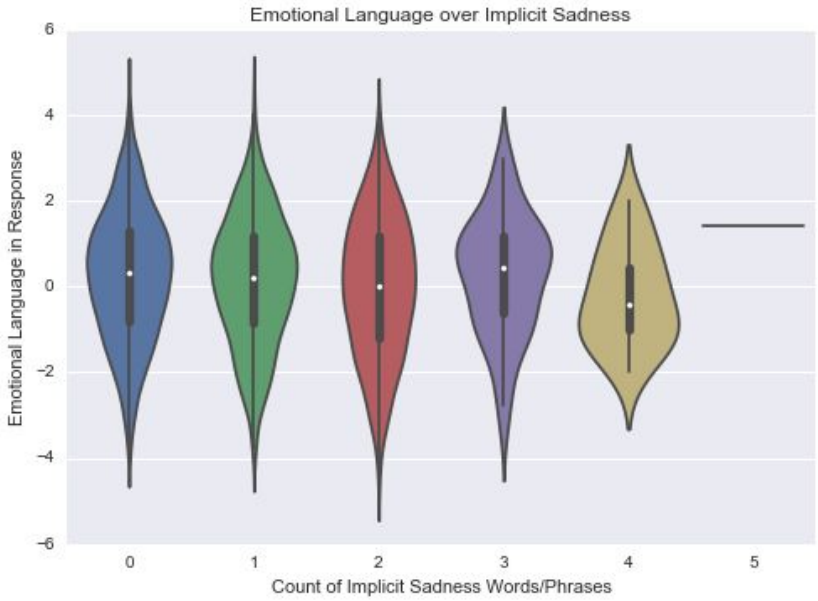


*Figure 6. Plot of the Distribution of Factual and Emotional Language within Responses over the Number of Implicit Sad Words or Phrases in Quotes*

Since the IAC also contains additional annotations for each response, it seemed best to examine them to see how they potentially relate to Emotional responses. Another regression analysis was made to see how much each annotation related the Fact-Feeling annotation. Below is a list of each additional response annotation along with the survey question asked of annotators.

- Agree-Disagree:
    - Does the respondent agree or disagree with the prior post?
- Defeater-Undercutter:
    - Is the argument of the respondent targeted at the entirety of the original poster's argument OR is the argument of the respondent targeted at a more specific idea within the post?
- Negotiate-Attack:
    - Does the respondent seem to have an argument of their own OR is the respondent simply attacking the original poster's argument?
- Nice-Nasty:
    - Is the respondent attempting to be nice or is their attitude fairly nasty?
- Audience-Personal:
    - Is the respondent's arguments intended more to be interacting directly with the original poster OR with a wider audience?
- Questioning-Asserting:
    - Is the respondent questioning the original poster OR is the respondent asserting their own ideas?

| Adjusted $R^2$ | P-value |
|---|---|
| 0.2517 | < 2.2e-16 |

| | Beta Weight | P-value |
|---|---|---|
| Agree Disagree | -0.0801 | 6.35e-07 *** |
| Defeater Undercutter | -0.0043 | 0.7854 |
| Negotiate Attack | -0.1499 | < 2e-16 *** |
| Nice Nasty | 0.2518 | 2.36e-11 *** |
| Audience Personal | 0.03427 | 0.0419 * |
| Questioning Asserting | 0.0201 | 0.2223 |

*Table 12. Regression results of predicting Fact-Feeling from Additional Response Annotations*

4 annotations were found to have strong relationships within Fact-Feeling. Agree-Disagree is showing a negative relationship meaning that when responses are emotional they tend to agree with the quote. Negotiate-Attack also shows a negative relationship which means that responses that are emotional tend to be an argument of their own, while factual posts are more so seen as attacking the quote. Nice-Nasty has a positive relationship which means that more emotional posts are viewed as nasty, while factual posts are viewed as nice. Lastly, Audience-Personal shows a positive relationship, so factual posts are viewed as talking toward an audience as opposed to being more personal.

To further examine how implicit emotions may have a more complex picture than simply a direct effect, I did another regression analysis that involved the interaction of implicit emotions with these additional annotations. There were only a couple of interactions that were significant: Implicit Sadness with Negotiate-Attack and Implicit Sadness with Nice-Nasty. These relationships were found to show that if the quote is more sad and the respondent is attacking, then even more emotional language is present and if the quote has implicit sadness and the respondent is nice, then there is also more emotional language present. These results however have a few caveats as there is a lot to unpack when it comes to interactions and it isn't clear without having a strong experimental design that accounts for potential interactions. So this only promotes the need for future work.

## 5. Conclusions

Overall this work has found that it is easier to detect emotion algorithmically within a more cooperative and work formal community. As it is a fact of exploring highly dynamic and varying contexts of online communities this doesn't go without limitations and potential for future work.

For one, these Enterprise communities have many different community types, ones that are very large and broad in topic and contrasting with those are ones small and very focused on specific tasks. Future work is needed to understand of community type comes into play for these results. Furthermore it needs to be explored as to why the more open internet style of posts within the IAC are harder to predict for the Fact-Feeling annotation. This may be that there is an underlying subtext that was not grabbed by any lexical, syntactical, or implicit measure used here.

Another unexpected result found was that people of authority (Leaders) created more positive emotional expression within discussion based tools (Forums). This was unexpected as you would imagine that leaders want to stay on a specific task or organize the community more so, which emotional language may be viewed more so a off-topic or a distraction from the community goals.

Some additional issues with this work is that the definition of a Member may be too vague and maybe we don't see such a similar promotion of emotional expression because Members can be very different. Potential future work can account for this by using some measures like community tenure or number of posts by member to help differentiate members from newcomers.

Lastly the work here on implicit emotions showed little. This can be an a potential lacking of the tool used to measure implicit emotions or it can be that within the IAC context implicit emotion does not charge people emotionally, as the IAC is capturing argumentation so implicit emotion may just have no relationship with people in an argumentation setting regardless if they are agreeing or disagreeing. The interactions do suggests that there can be a more complex relationship with implicit emotion, however future work is needed to verify this as these results were weak.

## 5. References


Alm, C. O., Roth, D., & Sproat, R. (2005, October). Emotions from text: machine learning for text-based emotion prediction. In Proc.of EMNLP (pp. 579-586). ACL.

Aman, S. and Szpakowicz, S. 2007. Identifying expressions of emotion in text. TSD. (2007), 196–205.

Balahur, A. et al. 2012. Building and exploiting emotinet, a knowledge base for emotion detection based on the appraisal theory model. Affective Computing, IEEE Transactions on. 3, 1 (2012), 88–101.

Bollegala, D. et al. 2013. Cross-domain sentiment classification using a sentiment sensitive thesaurus. Knowledge and Data Engineering, IEEE Transactions on. 25, 8 (2013), 1719–1731.

Chen, L. et al. 2012. Comparison of feature-level learning methods for mining online consumer reviews. Expert Systems with Applications. 39, 10 (2012), 9588–9601.

Ellsworth, P. C., & Scherer, K. R. (2003). Appraisal processes in emotion. Handbook of affective sciences, 572, V595.

Hai, Z. et al. 2014. Identifying features in opinion mining via intrinsic and extrinsic domain relevance. Knowledge and Data Engineering, IEEE Transactions on. 26, 3 (2014), 623–634.

Kraut, R.E. and Resnick, P. 2012. Building successful online communities: Evidence-based social design. MIT. (2012).

Matthews, T. and et al. 2015. They Said What? Exploring the Relationship Between Language Use and Member Satisfaction in Communities. Proc. of CSCW. (2015).

Mohammad, S.M. and Turney, P.D. 2010. Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotion lexicon. Proc. of NAACL. (2010), 26–34.

Muller, M. et al. 2012. Diversity among enterprise online communities: collaborating, teaming, and innovating through social media. Proc. of CHI (New York, NY, USA, 2012), 2815–2824.

Nguyen, D. and Rose, C.P. 2011. Language use as a reflection of socialization in online communities. Proc. of Languages in Social Media. (2011), 76–85.

Poria, S. et al. 2014. EmoSenticSpace: A novel framework for affective common-sense reasoning. Knowledge-Based Systems. 69, (2014), 108–123.

Soujanya Poria, Alexander Gelbukh, Amir Hussain, Dipankar Das, Sivaji Bandyopadhyay. Enhanced SenticNet with Affective Labels for Concept-based Opinion Mining. IEEE Intelligent Systems, vol. 28, issue 2, 2013, pp. 31–38, doi:10.1109/MIS.2013.4.

Ravi, K. and Ravi, V. 2015. A survey on opinion mining and sentiment analysis: tasks, approaches and applications. Knowledge-Based Systems. 89 (2015), 14–46.

Tausczik, Y.. and Pennebaker, J.W. 2010. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. Journal of Language and Social Psychology. 29, 1 (2010), 24–54.
Toutanova, K. and et al. 2003. Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network. Proc. of NAACL. (2003), 252–259.

Walker, M.A. and et al. 2012. A Corpus for Research on Deliberation and Debate. LREC. (2012).

Wang, Y. et al. 2012. To Stay or Leave? The Relationship of Emotional and Informational Support to Commitment in Online Health Support Groups. Proc. of CSCW. (2012).

Wilson, T. et al. 2005. Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. Proc. of HLT/EMNLP. (2005).