Frustrated, Polite or Formal: Quantifying Feelings and Tone in Emails

Niyati Chhaya✉[1], Kushal Chawla[1], Tanya Goyal[2], Projjal Chanda[3], and Jaya Singh[4]

[1]Big Data Experience Lab, Adobe Research, India
[2]University of Texas, Austin, USA
[3]Indian Institute of Technology, Kharagpur, India
[4]Indian Institute of Technology, Guwahati, India
[1]*nchhaya,kchawla@adobe.com*
[2]*tanyagoyal@utexas.edu*
[3,4]*projjalchanda1996,jayasinghiitg@gmail.com*

Abstract

Email conversations are the primary mode of communication in enterprises. The email content expresses an individual's needs, requirements and intentions. Affective information in the email text can be used to get an insight into the sender's mood or emotion. We present a novel approach to model human frustration in text. We identify linguistic features that influence human perception of frustration and model it as a supervised learning task. The paper provides a detailed comparison across traditional regression and word distribution-based models. We report a mean-squared error (MSE) of 0.018 against human-annotated frustration for the best performing model. The approach establishes the importance of affect features in frustration prediction for email data. We further evaluate the efficacy of the proposed feature set and model in predicting other *tone* or *affects* in text, namely formality and politeness; results demonstrate a comparable performance against the state-of-the-art baselines.

1 Introduction

Emails are the primary mode of communication in professional environments. Both formal and informal communication is prevalent through emails. In customer care organizations, email and instant messaging are used for conversations. The content in these communications includes information, conversations, requirements, actions and opinions. Every individual and organization has a style of language, topic of choice as well as patterns in which they communicate. Their personality and at times position (authority, social relationships) drive the choice of words and the tone of their content. Similarly, different recipients react differently to different kinds of content. For example, a professional is more likely to respond to a formal request than a casual request from his subordinate at the workplace. A customer care agent can easily calm down an agitated individual if he is polite. *Tone* in text is defined as this expression of affects or feelings in content. We present a study to measure this *Tone* in text content, specifically in email conversations.

Quantifying text sentiment based on lexical and syntactic features is well studied. Further, measuring ease of read (Kate et al., 2010) as well as coherency of text content has been explored. Sentiment and emotion analysis have been explored for specific affect dimensions (e.g. polarity and Ekman's six Emotion categories). Interpersonal communication illustrates fine–grained affect categories, beyond emotions and sentiments. Frustration is one such dominant affect that is expressed in human interactions (Burgoon and Hale, 1984). We present a study of Frustration.

Expressions, tone of the voice (audio), actions, and physical reactions are easy cues to detect the presence of frustration. In the case of text content, identifying the correct sentence formations, use of words, and lexical content structure for affect detection, specifically frustration, is a hard problem (Calvo and D'Mello, 2010; Munezero et al., 2014). We show how using affect lexica to quantify frustration in text content improves the performance as against using just lexical and syntactic features. Our experiments highlight the importance of using word–level affect features for the prediction task. We show that affect features also contribute to the prediction of formality and politeness, which are *tone* dimensions that have been explored earlier. We compare and contrast a traditional regression model with models based on

Proceedings of the Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media, pages 76–86
New Orleans, Louisiana, June 6, 2018. ©2018 Association for Computational Linguistics

word embeddings. The traditional feature–based models outperform the rest for this dataset and the task.

This paper investigates frustration in online written communication. The contributions are:

- We present a state–of–the–art statistical model for frustration prediction, evaluate it against human judgments, and present an in–depth feature analysis: highlighting the importance of affect features. We also evaluate our model for formality and politeness detection and report comparable accuracy as against the state–of–the–art prior work.

- We present an analysis that studies the relationship of Frustration with Formality and Politeness in text data and report negative correlation across these dimensions. High frustration is observed in content with low formality and low politeness.

- We provide an analysis of what humans tag as frustration in written text across 6 different topics.

2 Related Work

Research around understanding text features and quantifying them has been explored. Methods to measure various lexical, syntactic, and semantic text analysis features have been studied on various datasets and mentioned earlier across different emotion and sentiment dimensions (Das and Kalita, 2017). We are concerned with the specific dimensions of frustration, formality, and politeness in text and hence will not present a detailed review for all other work.

To the best of our knowledge, this is the first work that attempts at a computational model for frustration in text. However, dimensions such as formality and politeness have been explored earlier. The closest work in frustration detection is related to interactions and conversations with intelligent chatbots (Wang et al., 2004; D'Mello et al., 2008). These approaches measure human frustration in either online tutoring systems on chat systems or online game interactions. The features used for affect detection include speech, video, and lexical or syntactic features such as use of emoticons. Ciechanowski et al. (2018) provide an overview of approaches of the current models and algorithms in this space using electromyography as well as other psycho-physiological data and a detailed set of questionnaires focused on assessing interactions and willingness to collaborate with a bot, which is one of the most recent work in the paradigm. The above systems, however, do not try to model and quantify these tone dimensions in long texts such as email or blogs.

Tutoring systems and e-learning systems need to evaluate the quality of the responses as well as the student experience. McQuiggan et al. (2007) model student frustration in their work. Student frustration and boredom along with confusion and concentration is studied by researchers who evaluate the efficiency of online tutoring systems and educational computer games (Conati and Maclaren, 2009; Sabourin et al., 2011). These approaches are based on probabilistic modeling and bayesian reasoning use sensors from multiple physiological and audio–video signals. Our work focuses on modeling similar tone from text. Vizar et al. (2009) study frustration in the process of modeling stress using keystrokes and language information. Their work uses speech data and not written text, which is the focus of this paper. While prior art in frustration and similar tone dimensions exists, it has been modeled only based on multi–modal and multi–sensor data as against the text–based content that we present in this paper.

Two complementary affects along with frustration, are formality and politeness in text. Formality has been defined in different works in varied ways (Brooke et al., 2010; Lahiri, 2015). Pavlick et. al. (2016) assume a user–based definition of formality, we use a similar approach to define frustration in this work. The authors focus on semantic (ngram), lexical and syntactic features to present an empirical study of quantifying formality in various types of text content. Their work ignores the affect–related features. We use their model as a baseline in the experiments for formality prediction. Our approach out performs their model for the email dataset. We study politeness to understand the relationship between politeness and frustration. The state–of–the–art in politeness detection predicts politeness in online requests (Danescu-Niculescu-Mizil et al., 2013). We use that approach as a baseline. Most of the published work in this space of text tone dimensions looks at either social media data or chat related datasets. This paper focuses on an email dataset.

Linguistic analysis of email data has gained popularity since the release of the ENRON dataset (Cohen, 2009). This dataset provides the text, user information as well as the network information. In this work, we use a subset of this publicly available dataset. Enron has been a very popular dataset for social network analysis (Chapanond et al., 2005; Diesner et al., 2005; Shetty and Adibi, 2005; Oselio et al., 2014; Ahmed and Rossi, 2015) and sentiment and authority analysis (Diesner and Evans, 2015; Liu and Lee, 2015; Miller and Charles, 2016; Mohammad and Yang, 2011). Peterson et al. (2011) present an approach to model formality on the ENRON corpus and Kaur et al. (2014) compare emotions across formal and informal emails. Jabbari et al. (2006) analyze business and personal emails as different classes of data. Approaches that study the social relationships in the ENRON dataset (Prabhakaran et al., 2012; Zhou et al., 2010; Miller and Rye, 2012; Cotterill, 2013) refer to formality and politeness as indicative features for such tasks. This vast usage of the ENRON dataset supports our choice of the corpus for modeling frustration in interpersonal text communication.

Human Perceptions and Definitions

Tone or affects such as frustration and formality are highly subjective. As seen in section 2 there are various definitions for these measures. We need to specify our own definitions for frustration before we try to automate the prediction. This work does not attempt to introduce a novel or an accurate measure of frustration (or formality and politeness), but we assume that these are defined by human perception and each individual may differ in their understanding of the metrics. This approach of using untrained human judgments has been used in prior studies of pragmatics in text data (Pavlick and Tetreault, 2016; Danescu-Niculescu-Mizil et al., 2013) and is a suggested way of gathering gold-standard annotations (Sigley, 1997).

We define frustration as the frustration expressed in text for our study. The aim is to answer whether there is any coherence across individual's perception of frustration (3.1.1). If so, what linguistic features, specifically affect features, contribute towards this collective notion? Based on this, we present an automated approach for frustration prediction in text (Section 4).

3 Data and Annotation

Table 1: Dataset Description

Property	**Value**
Total number of emails (Main Experiment)	960
Total number of emails (Pilot Experiment)	90
Min. sentences per email	1
Max. sentences per email	17
Average email size (no. of sentences)	4.22
Average number of words per email	77.5

Table 2: Annotations on Varying Email sizes

Dimension	**Email size (# sentences)**	**# emails**	**mean**	**std. dev.**
Frustration	$0-2$	258	-0.06	0.11
(-2,-1,0)	$3-5$	452	-0.07	0.13
	$6-17$	250	-0.08	0.11
Formality	$0-2$	258	0.11	0.55
(-2,-1,0,1,2)	$3-5$	452	0.37	0.54
	$6-17$	250	0.65	0.46
Politeness	$0-2$	258	0.35	0.33
(-2,-1,0,1,2)	$3-5$	452	0.51	0.34
	$6-17$	250	0.64	0.29

We study the human perception of frustration expressed in text across different topics and message (text) lengths. Prior research on dimensions such as formality and politeness present a similar analysis of how they vary across types of text and genres. Due to the lack of annotated data for frustration, we conducted a crowd sourcing experiment using Amazon's Mechanical Turk. We work off a subset of about 5000 emails from the ENRON email dataset (Cohen, 2009). This contains emails exchanged between over 100 employees and spans across various topics. The analysis presented in this section is based on a subset of about 1050 emails that were tagged across one pilot and one full scale experiment. Table 1 provides an overview of the data statistics of the annotated data.

We follow the annotation protocol of the Likert Scale (Allen and Seaman, 2007) for three affect dimensions: Frustration, Formality, and Politeness. Each email is considered as a single data point and only the text in the email body is provided for tagging. Frustration is tagged on a 3 point scale with neutral being equated to 'not frustrated'; 'frustrated' and 'very frustrated' are marked with -1 and -2 respectively. Formality and politeness follow a 5 point scale from -2 to $+2$ where both extremes mark the higher degree of presence and absence of the respective dimension. We use a mean of 10 annotators score for each input email.[1]

[1] Dataset can be accessed at https://goo.gl/WFkDnS

3.1 Analysis

The data has been tagged by 69 individuals, where the average time spent per email is 28.2 seconds. The average number of emails annotated by an individual are approximately 139.

3.1.1 Inter-annotator Agreement

To measure whether the individuals intuition of the affect dimensions is consistent with other annotators' judgment, we use interclass correlation[2] to quantify the ordinal ratings. This measure accounts for the fact that we may have a different group of annotators for each data point. Agreements reported for 3 class and 5 class annotations are 0.506 ± 0.05, 0.73 ± 0.02, and 0.64 ± 0.03 for frustration, formality, and politeness respectively. These numbers are comparable to any other psycholinguistic task. Example emails with their corresponding annotations are provided in Table 3.

3.1.2 Email size and Tone dimensions

Table 2 shows the variance in frustration, formality and politeness in comparison to the email size. We observe that while formality and politeness vary with content size, frustration does not have a significant variance.

3.1.3 Comparison with Readability

We observe that the Readability of the content does not impact the tagged frustration values as against the case with formality and politeness. Figure 1 shows how frustration varies across different readability scores. Prediction experiments (see Table 5) support this observation.

3.1.4 Affective Content

One purpose of this study is to understand the words that are associated with emotions and whether affect plays a role in understanding frustration in this data. Figure 2 shows this analysis. The graphs show the variance in frustration with respect to three psycholinguistic features. As seen in the figure, PERMA relationship(POS) has a very different behavior with the positive and the negative frustration class. This analysis helps in confirming the hypothesis on relationship between frustration in text and psycholinguistic features.

4 Modeling Frustration

We analyze whether an algorithm can distinguish between existence and non-existence of the expression of frustration in text and which linguistic features are important for this task.

4.1 Setup

The data described in section 3 is used for training, using the mean of the annotators' scores as the gold-standard label. We model the problem as a regression task. The task is to predict frustration in given text. We also report results for formality and politeness prediction and compare against baselines for both these dimensions. The model is implemented using the Scikit[3] package in Python.

4.2 Features

Table 4 provides a summary of the features considered. Ngrams and other semantic features are ignored as they introduce domain-specific biases. Word-embeddings are treated separately and considered as raw features to train a supervised model. 55 features are divided into 4 sub-groups: Lexical, Syntactic, Derived(e.g. readability) and Affect-based features. The lexical and syntactic features are defined based on standard definitions. These include features such as 'averageNumberofWords per sentence' and 'number of capitalizations'. The derived features focus on features that can help quantify the readability of text. Hedges, Contractions, and Readability scores are included in this set of features. The fourth group of features are the Affect–related features. These features are lexica–based and quantify the amount of affective content present in the input text. We use Stanford Corenlp[4] and TextBlob[5] for our linguistic processing and feature extraction. All features used by Pavlick et. al. (2016) for formality detection and by Danescu et al. (2013) for politeness detection have been included in our analysis for a comparison against baselines. To the best of our knowledge, this is not only the first of its kind work for quantifying frustration in text using linguistic features but also the first attempt at explicitly using affect features for such an affect detection task.

[2] We report the average raters absolute agreement (ICC1k) using the psych package in R.

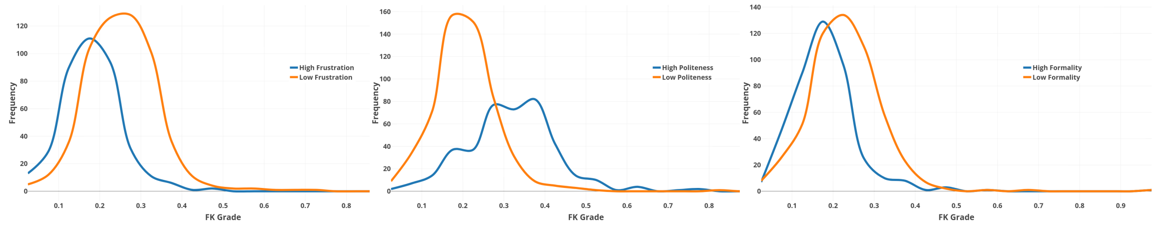
[3] http://scikit-learn.org/

[4] https://stanfordnlp.github.io/CoreNLP/

[5] https://textblob.readthedocs.io/en/dev/

Table 3: Example emails with high and low inter-annotator agreements.

Affect Dimension	Example	Annotations
Frustration: Low Agreement	See highlighted portion. We should throw this back at Davis next time he points the finger.	(-1, -1, 0, 0, -2, -2, 0, 0, -2, 0)
Frustration: High Agreement	Please see announcement below. Pilar, Linda, India and Deb, please forward to all of your people. Thanks in advance, adr	(0, 0, 0, 0, 0, 0, 0, 0, 0, 0)
Formality: Low Agreement	I talked with the same reporters yesterday (with Palmer and Shapro). Any other information that you can supply Gary would be appreciated. Steve, did Gary A. get your original as the CAISO turns email? GAC	(0, 0, -1, 1, 1, 1, 0, -1, -2, -1)
Politeness: High Agreement	John, This looks fine from a legal perspective. Everything in it is either already in the public domain or otherwise non-proprietary. Kind regards, Dan	(1, 1, 1, 1, 1, 1, 1, 1, 2, 1)



(a) Frustration (b) Politeness (c) Formality

Figure 1: Comparing Readability Index (Flesh–Kincaid readability score-FKGrade- with *Tone* Dimensions: Graphs show the distribution of readability score for the positive and negative class for each of the dimensions. The two classes correspond to the presence or absence of the respective tone.

Lexical and Syntactic Features: The lexical features capture various counts associated with the content. Prior art in formality and politeness prediction extensively relies on such features for their analysis and hence we hypothesize that the lexical properties will contribute to our task. Syntactic features include NER–based features, Number of blank lines, and text density. Text density is defined as follows:

$$\rho = \frac{\#(sentences)}{1 + \#(lines)}$$

where ρ is the text density, $\#(sentences)$ denotes number of sentences in the text content and $\#(lines)$ number of lines including blank lines in the text message.

Derived: Readability Features: The derived features capture information such as readability of text, existence of hedges, subjectivity, contractions, and sign–offs. Subjectivity, contractions, and hedges are based on the TextBlob implementation.

Readability is measured based on Flesh–Kincaid readability score which is given by the following equation:

$$FKGrade = 0.39\frac{words}{sentences} + 11.8\frac{syllables}{words} + 15.59$$

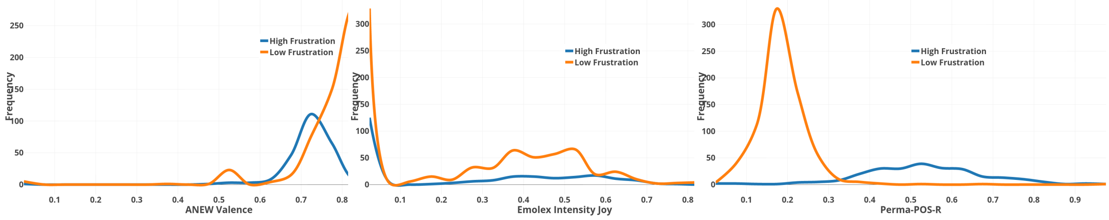
This score is a measure of ease of reading of given piece of text. We use the textstat package[6] in Python for the implementation.

Affect Features: The affect features used in our analysis include:

1. **Valence-Arousal-Dominance (PAD) Model** (Mehrabian, 1980): This three dimensional model quantifies the valence which is the happy-unhappy scale, arousal: the excited–calm scale, and dominance, which indicates the forcefulness of the expressed affect. We use the Warriner's lexicon (Warriner et al., 2013) for the feature extraction.

2. **Ekman's Emotions (Ekman, 1992):** Ekman's model provides the 6 basic human emotions: anger, disgust, admiration, surprise, anticipation, and sadness. We use the

[6] https://pypi.python.org/pypi/textstat/0.1.6



(a) ANEW Valence (b) Emolex Joy (c) PERMA Relationship

Figure 2: Frustration and Affect: Graphs show how specific affect dimension (based on lexica) varies for the positive and negative class of Frustration. PERMA relationship (POS) has a sharp peaked distribution for one class where as a flat distribution for the other. ANEW-Valence and EmolexIntensity-JOY also vary across classes dimensions

Table 4: Summary of feature groups used in our model. To the best of our knowledge, those marked with (*) have not been previously studied to model any of the three affects: Frustration, Formality, and Politeness. This list is the set of features that were finally used in our model. A larger list of explored features is provided as supplementary material.

Features	Feature list
Lexical	Average Word Length, Average Words per Sentence, # of Upper Case Words, # Ellipses, # Exclamation marks, # Question Mark, # Multiple Question Marks, # Words, # Lower Case words, First word upper case, # NonAlphaChars, # Punctuation Chars
Syntactic	# BlankLines, NER-Person, NER-Location, NER-PersonLength, NER-Organization, TextDensity
Derived	# Contractions, ReadabilityScore- FKgrade, FirstPerson, Hedge, Subjectivity, Sentiment, ThirdPerson, SignOff
Affect*	ANEW-arousal, ANEW-dominance, ANEW-valence, EmolexIntensity-anger, EmolexIntensity-fear, EmolexIntensity-joy, EmolexIntensity-sadness, Emolex-anger, Emolex-anticipation, Emolex-disgust, Emolex-fear, Emolex-joy, Emolex-negative, Emolex-positive, Emolex-sadness, Emolex-surprise, Emolex-trust, Perma-NEG-A, Perma-NEG-E, Perma-NEG-M, Perma-NEG-P, Perma-NEG-R, Perma-POS-A, Perma-POS-E, Perma-POS-M, Perma-POS-P, Perma-POS-R
Formal Words	formal-words, informal-words (Brooke et al., 2010)

NRC lexicon (EMOLEX) (Mohammad et al., 2013) which provides a measure for the existence of the emotion as well as the intensity of the detected emotion.

3. **PERMA Model (Seligman, 2011):** The PERMA model is a scale to measure positivity and well–being in humans (Seligman, 2012). This model defines the 5 dimensions: Positive Emotions, Engagement, Relationships, Meaning, and Accomplishments as quantifiers and indicators of positivity and well–being. Schwartz et al. (Schwartz et al., 2013) published a PERMA lexicon. We use this lexicon in our work. Frustration is considered as an important measure in the study of Positive Psychology. Hence, we leverage the PERMA model for our features.

4. **Formality Lists**: Brooke et al. (Brooke et al., 2010) provide a list of words that usually indicate formality or informality in text. We use these lists for our experiments.

4.2.1 ENRON–embeddings

We train a Word2Vec CBOW model (Mikolov et al., 2013) on raw 517, 400 emails from the ENRON email dataset to obtain the word embeddings. We keep the embedding size as 50 and a window of 5, taking a mean of all the context words to obtain the context representation. For optimization, we use negative sampling, drawing 5 noisy samples at each instance. An aggregate of these embeddings (see ENRON–trained embeddings in table 5) is considered as a feature set for one of our experiments.

5 Experiments

This section describes experiments associated with this work. All experiments report the accuracy against the ground-truth dataset described earlier.

Tone Prediction - Can you predict Frustration? This section reports the results for predicting frustration on a held out test dataset. Table 5 reports the mean squared error for different regression models with varying feature sets. We also report results for formality and politeness against the same settings. **Ridge regression** with lexical, syntactic, and affect features is the best performing model for frustration. The politeness baseline is the best performing model for both formality and politeness prediction. We also report MSE values using the 50-dimensional ENRON–trained embeddings as features. Even though these features are trained on the large ENRON dataset(500, 000 emails), they underperform as against the affect features. We conclude that the psycholinguistic features(i.e. affect features) are more predictive for such subjective tasks.

Classification: To understand whether one can differentiate between the positive and the negative class for *tone* dimensions such as frustration, we also model the problem as a 2–class classification problem. Neutral tags are considered a part of the negative class. Hence, the classification model predicts where the text has frustration (or formality, or politeness) or not. Table 6 shows the performance of different classification models across different feature groups where the positive class is oversampled to compensate for the class imbalance [Frustration: 249 (positive class), 731 (negative class); Formality: 455 (positive class), 525 (negative class); Politeness: 423 (positive class), 557 (negative class)]. Note that this experiment is done on the same dataset with 3 annotation/email as against 10 annotations. **Random Forest** (10 trees) is the best performing model with an accuracy of 0.86. Random Forest is the best predictor for Frustration while Logistic Regression has the highest accuracies for Formality and Politeness prediction.

Feature Importance: Which features help to predict Frustration? Figure 3 shows the relative feature importances of top few features across the three affect dimensions. PERMA-positivity has very negative correlation with frustration but is moderately negatively correlated with politeness. This confirms the hypothesis of contribution from affect features. Frustration is best predicted with affect features, formality and politeness are not.

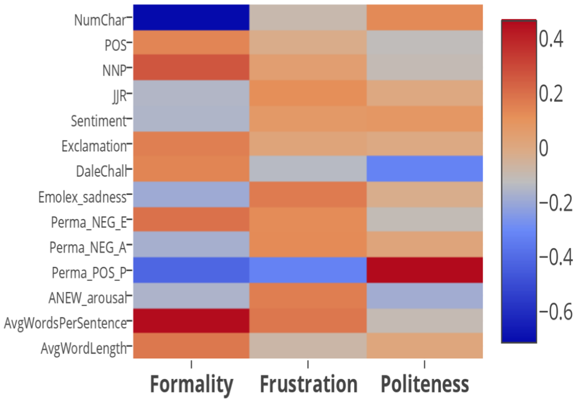


Figure 3: Figure shows the relative feature importance of top few features across all three dimensions. The importance is calculated based on results of logistic regression. PERMA-positivity has very negative correlation with frustration but is negatively correlated with politeness.

6 Discussion

- **Comparing Frustration with Formality and Politeness:** Table 7 shows the pairwise Pearson's correlation coefficient across the *tone* dimensions. Both politeness and formality are negatively correlated with frustration. Hence, more formal you are, less frustration might be detected in the text. While the correlations are negative, no concrete relationship across these dimensions can be stated due to the subjectivity.

- **Analysis of Affect Features:** Three types of affect features used in our model follow different properties. To understand the contribution of each of them, we further study the feature importance of these features. To identify the most predictive features, we report the p–values calculated for the F-scores reported against the F-regression tests for each of the *tone* dimensions. F–test reports the p–values indicating the importance of the regression. As seen in the table 8 PERMA and ANEW

Table 5: MSE for Prediction of Frustration, Formality, and Politeness in a Regression setting. Ridge Regression out performs all other models. The Feature set with Lexical, Syntactic, and Affect features performs best for all dimensions. Values denote the MSE across 10–fold cross validation.

Model	**Lex+ Syn**	**Lex + Syn + Derived**	**Lex + Syn + Affect**	**All**	**ENRON–trained Embeddings**	**Baseline-Formality**[7] **(Pavlick and Tetreault, 2016)**	**Baseline- Politeness**[8] **(Danescu-Niculescu-Mizil et al., 2013)**
Frustration							
Linear Regression	0.02954	0.02823	0.02935	0.02872	0.02653	1.5356e+13	0.0655
Lasso Regression	0.02433	0.02433	0.02433	0.02433	0.02433	0.0245	0.0253
Ridge Regression	**0.02283**	**0.02231**	**0.02157**	**0.02121**	0.0265	0.0249	0.0373
SVR Regression	0.02958	0.02887	0.02633	0.0263	0.02483	–	0.0219
Formality							
Linear Regression	0.0289	0.02847	0.02803	0.02805	0.03542	2.0708e+14	0.0808
Lasso Regression	0.02807	0.02807	0.02807	0.02807	0.03756	**0.0279**	0.0429
Ridge Regression	**0.01817**	**0.01794**	**0.0176**	**0.01745**	0.0354	0.0232	0.0372
SVR Regression	0.02375	0.0242	0.02288	0.02296	0.03247	–	0.0182
Politeness							
Linear Regression	0.02082	0.01934	0.01966	0.0189	0.01922	1.6484e+14	0.0575
Lasso Regression	0.02041	0.02041	0.02041	0.0204	0.02062	0.0202	0.0218
Ridge Regression	**0.01771**	**0.01671**	**0.0161**	**0.01556**	0.01921	0.01561	0.0266
SVR Regression	0.02119	0.02035	0.02007	0.02058	0.01909	–	**0.0130**

Table 6: Accuracy for Frustration prediction when modeled as a 2-class classification problem. The positive class is oversampled to correct for class imbalance. Random forest is the best performing classifier with a precision= 0.88, recall= 0.85, and F1-Score= 0.85. The Affect features contribute more to the accuracy as compared to the derived features. All values are reported for the experimental setup with a 80–20 train-test split with 10 fold cross validation.

Model	**Lex + Syn**	**Lex + Syn + Derived**	**Lex + Syn + Affect**	**All**	**Baseline-Formality**	**Baseline-Politeness**
Logistic Regression	0.62	0.61	0.67	0.66	0.72	**0.72**
SVC	0.66	0.69	0.62	0.64	–	–
Linear SVC	0.56	0.55	0.58	0.57	0.68	0.68
Random Forest	**0.85**	**0.85**	**0.86**	**0.86**	0.73	0.70
Nearest Neighbors	0.64	0.62	0.65	0.62	**0.72**	0.71

Table 7: Pearson's Coefficients for pair–wise affects. Interestingly, the affects are negatively correlated. Being formal may make individuals less frustrated at the cost of politeness!

Affects	**Formality**	**Politeness**	**Frustration**
Formality	1	-0.129	-0.183
Politeness	-0.129	1	-0.252
Frustration	-0.183	-0.129	1

Table 8: p–values for top affect features using a F–Regression Test. Low values show high predictability.

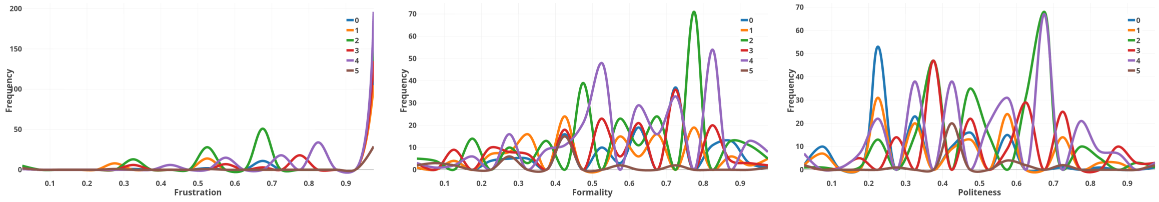
Features	**Formality**	**Politeness**	**Frustration**
Perma-POS-R	2.43e-08	1.22e-22	0.61
Perma-NEG-M	4.31e-13	2.26e-06	2.63e-15
Perma-NEG-A	5.75e-19	0.03	4.09e-14
ANEW-arousal	4.07e-05	0.01	0.08
ANEW-dominance	0.09	5.14e-10	0.17
Emolex Intensity Sadness	0.02	0.25	5.24e-11

features report a very low p–value showing the significance of the corresponding features for regression.

- **Does the *Tone* in text change with topics?** Figure 4 shows the affect distribution across different topics. These topics are derived based on topic modeling using Latent Dirichlet Allocation followed by KMeans clustering. A given email is tagged with a single topic and the distributions are computed over these disjoint clusters. While the affect values for all topics have a similar range, they follow a different distribution. For topic 2 which denotes content about sports-related conversations.

7 Conclusion

We present a novel approach for Frustration detection in text data. Our approach proves the importance of affect based features for this task and our traditional regression as well as classification models outperform the baselines and the word-



(a) Frustration

(b) Formality

(c) Politeness

Figure 4: Topics and Tone: Graph shows how the different text tone dimensions vary for different topics. Topic 2 which is content about sport has a very different frustration distribution as compared to other topics.

embeddings-based method for frustration prediction. We also show our model does comparable to baselines for formality and politeness prediction. We plan to extend this work towards defining linguistic aspects of frustration in text. We believe, this is the very first attempt at modeling a hard dimension such as frustration.

References

Nesreen Kamel Ahmed and Ryan Anthony Rossi. 2015. Interactive visual graph analytics on the web. In *ICWSM*, pages 566–569.

I Elaine Allen and Christopher A Seaman. 2007. Likert scales and data analyses. *Quality progress*, 40(7):64.

Julian Brooke, Tong Wang, and Graeme Hirst. 2010. Automatic acquisition of lexical formality. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, pages 90–98. Association for Computational Linguistics.

Judee K. Burgoon and Jerold L. Hale. 1984. The fundamental topoi of relational communication. *Communication Monographs*, 51(3):193–214.

Rafael A Calvo and Sidney D'Mello. 2010. Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on affective computing*, 1(1):18–37.

Anurat Chapanond, Mukkai S Krishnamoorthy, and Bülent Yener. 2005. Graph theoretic and spectral analysis of enron email data. *Computational & Mathematical Organization Theory*, 11(3):265–281.

Leon Ciechanowski, Aleksandra Przegalinska, and Krzysztof Wegner. 2018. *The Necessity of New Paradigms in Measuring Human-Chatbot Interaction*. Springer International Publishing, Cham.

William W Cohen. 2009. Enron email dataset.

Cristina Conati and Heather Maclaren. 2009. Empirically building and evaluating a probabilistic model of user affect. *User Modeling and User-Adapted Interaction*, 19(3):267–303.

Rachel Cotterill. 2013. Using stylistic features for social power modeling. *Computación y Sistemas*, 17(2).

Cristian Danescu-Niculescu-Mizil, Moritz Sudhof, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013. A computational approach to politeness with application to social factors. *arXiv preprint arXiv:1306.6078*.

Sufal Das and Hemanta Kumar Kalita. 2017. Sentiment analysis for web-based big data: A survey. *International Journal*, 8(5).

Jana Diesner and Craig S Evans. 2015. Little bad concerns: Using sentiment analysis to assess structural balance in communication networks. In *Advances in Social Networks Analysis and Mining (ASONAM), 2015 IEEE/ACM International Conference on*, pages 342–348. IEEE.

Jana Diesner, Terrill L Frantz, and Kathleen M Carley. 2005. Communication networks from the enron email corpus it's always about the people. enron is no different. *Computational & Mathematical Organization Theory*, 11(3):201–228.

Sidney K. D'Mello, Scotty D. Craig, Amy Witherspoon, Bethany McDaniel, and Arthur Graesser. 2008. Automatic detection of learner's affect from conversational cues. *User Modeling and User-Adapted Interaction*, 18(1):45–80.

Paul Ekman. 1992. An argument for basic emotions. *Cognition & Emotion*, 6(3-4):169–200.

Sanaz Jabbari, Ben Allison, David Guthrie, and Louise Guthrie. 2006. Towards the orwellian nightmare: separation of business and personal emails. In *Proceedings of the COLING/ACL on Main conference poster sessions*, pages 407–411. Association for Computational Linguistics.

Rohit J Kate, Xiaoqiang Luo, Siddharth Patwardhan, Martin Franz, Radu Florian, Raymond J Mooney, Salim Roukos, and Chris Welty. 2010. Learning to predict readability using diverse linguistic features.

In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 546–554. Association for Computational Linguistics.

Jasleen Kaur and Jatinderkumar R Saini. 2014. Emotion detection and sentiment analysis in text corpus: a differential study with informal and formal writing styles. *International Journal of Computer Applications*, 101(9).

Shibamouli Lahiri. 2015. SQUINKY! A Corpus of Sentence-level Formality, Informativeness, and Implicature. *CoRR*, abs/1506.02306.

Sisi Liu and Ickjai Lee. 2015. A hybrid sentiment analysis framework for large email data. In *Intelligent Systems and Knowledge Engineering (ISKE), 2015 10th International Conference on*, pages 324–330. IEEE.

Scott W. McQuiggan, Sunyoung Lee, and James C. Lester. 2007. *Early Prediction of Student Frustration*. Springer Berlin Heidelberg, Berlin, Heidelberg.

Albert Mehrabian. 1980. Basic dimensions for a general psychological theory implications for personality, social, environmental, and developmental studies.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Christopher A Miller and Jeffrey Rye. 2012. Power and politeness in interactions: Admire-a tool for deriving the former from the latter. In *Social Informatics (SocialInformatics), 2012 International Conference on*, pages 177–184. IEEE.

R Miller and EYA Charles. 2016. A psychological based analysis of marketing email subject lines. In *Advances in ICT for Emerging Regions (ICTer), 2016 Sixteenth International Conference on*, pages 58–65. IEEE.

Saif M Mohammad, Svetlana Kiritchenko, and Xiaodan Zhu. 2013. Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets. *arXiv preprint arXiv:1308.6242*.

Saif M Mohammad and Tony Wenda Yang. 2011. Tracking sentiment in mail: How genders differ on emotional axes. In *Proceedings of the 2nd workshop on computational approaches to subjectivity and sentiment analysis*, pages 70–79. Association for Computational Linguistics.

Myriam D Munezero, Calkin Suero Montero, Erkki Sutinen, and John Pajunen. 2014. Are they different? affect, feeling, emotion, sentiment, and opinion detection in text. *IEEE transactions on affective computing*, 5(2):101–111.

Brandon Oselio, Alex Kulesza, and Alfred O Hero. 2014. Multi-layer graph analysis for dynamic social networks. *IEEE Journal of Selected Topics in Signal Processing*, 8(4):514–523.

Ellie Pavlick and Joel Tetreault. 2016. An empirical analysis of formality in online communication. *Transactions of the Association for Computational Linguistics*, 4:61–74.

Kelly Peterson, Matt Hohensee, and Fei Xia. 2011. Email formality in the workplace: A case study on the enron corpus. In *Proceedings of the Workshop on Languages in Social Media*, pages 86–95. Association for Computational Linguistics.

Vinodkumar Prabhakaran, Huzaifa Neralwala, Owen Rambow, and Mona T Diab. 2012. Annotations for power relations on email threads. In *LREC*, pages 806–811.

Jennifer Sabourin, Bradford Mott, and James C. Lester. 2011. *Modeling Learner Affect with Theoretically Grounded Dynamic Bayesian Networks*. Springer Berlin Heidelberg, Berlin, Heidelberg.

H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Lukasz Dziurzynski, Stephanie M Ramones, Megha Agrawal, Achal Shah, Michal Kosinski, David Stillwell, Martin EP Seligman, et al. 2013. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one*, 8(9):e73791.

Martin EP Seligman. 2011. Flourish: a visionary new understanding of happiness and well-being. *Policy*, 27(3):60–1.

Martin EP Seligman. 2012. *Flourish: A visionary new understanding of happiness and well-being*. Simon and Schuster.

Jitesh Shetty and Jafar Adibi. 2005. Discovering important nodes through graph entropy the case of enron email database. In *Proceedings of the 3rd international workshop on Link discovery*, pages 74–81. ACM.

Robert J Sigley. 1997. Text categories and where you can stick them: a crude formality index. *International Journal of Corpus Linguistics*, 2(2):199–237.

Lisa M Vizer, Lina Zhou, and Andrew Sears. 2009. Automated stress detection using keystroke and linguistic features: An exploratory study. *International Journal of Human-Computer Studies*, 67(10):870–886.

Hua Wang, Helmut Prendinger, and Takeo Igarashi. 2004. Communicating emotions in online chat using physiological sensors and animated text. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '04, pages 1171–1174, New York, NY, USA. ACM.

Amy Beth Warriner, Victor Kuperman, and Marc Brysbaert. 2013. Norms of valence, arousal, and dominance for 13,915 english lemmas. *Behavior Research Methods*, 45(4):1191–1207.

Yingjie Zhou, Kenneth R Fleischmann, and William A Wallace. 2010. Automatic text analysis of values in the enron email dataset: Clustering a social network using the value patterns of actors. In *System Sciences (HICSS), 2010 43rd Hawaii International Conference on*, pages 1–10. IEEE.