Analysis of Linguistic Style Accommodation in Online Debates

Arjun Mukherjee Bing Liu
Department of Computer Science
University of Illinois at Chicago
arjun4787@gmail.com liub@cs.uic.edu

ABSTRACT

Psycholinguistic phenomenon of communication accommodation (Giles et al., 1991) is probably one of the most important contributions in the interdisciplinary field of linguistics, psychology, information, and communication theory. Existing works have applied this theory to various domains like gesture, linguistics, backchannels, and even social media like tweets. In this work, we analyze the psycholinguistic phenomenon of linguistic style accommodation in online debates. First, we present a Joint Topic Expression (JTE) model for modeling debate posts and use it to generate our unique dataset for studying accommodation in debates. Specifically, we analyze the phenomenon across agreeing/disagreeing debating pairs generated using our JTE model. Second, we propose a formal framework for analyzing the linguistic phenomena of accommodation in online debates. Experiments on a large collection of real-life debate posts reveal very interesting insights about the complex phenomenon of psycholinguistic accommodation in online debates.

KEYWORDS : Linguistic style accommodation, linguistic convergence, accommodation in debates, online debate conversations.

Proceedings of COLING 2012: Technical Papers, pages 1831–1846,
COLING 2012, Mumbai, December 2012.

1 Introduction

The psycholinguistic theory of communication accommodation was developed by Howard Giles (Giles et al., 1991). It argues that "when people interact, they adjust their speech, their vocal patterns and their gestures, to accommodate to others". This adjustment or accommodation tends to occur unconsciously, i.e., people tend to instinctively converge to one another's communicative behavior. Over the past five decades, this phenomenon has received a great deal of attention across a myriad of domains: posture (Condon and Ogston, 1967), speech pause length (Jaffe and Feldstein, 1970), head nodding (Hale and Burgoon, 1984), generic linguistic style (Niederhoffer and Pennebaker, 2002), tweets (Danescu-Niculescu-Mizil et al., 2011), etc. This work presents a formal framework to model communication accommodation in online debates. Online debate forums are perhaps the most popular form of debates where people participate in discussions of various issues like politics, religions, society, human rights, etc. It is naturally very interesting to analyze the phenomenon of accommodation in debates.

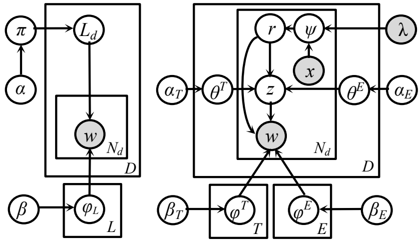
In this work, we focus on linguistic style accommodation in debates. In detail, we will perform the following types of analysis: stylistic cohesion, stylistic accommodation, influence, and accommodation across both agreeing and disagreeing debate posts in online debates. We use the linguistic style markers in LIWC (Pennebaker et al., 2007) to measure the amount of linguistic accommodation exhibited. The underlying hypothesis behind the "measurement" of linguistic accommodation using linguistic style markers is based on the prior works in (Gonzales et al., 2010; Niederhoffer and Pennebaker, 2002; Taylor and Thomas, 2008), which have shown that linguistic accommodation being most pronounced in style dimensions is a good metric for measurement. Linguistic "style" here denotes content independent language constructs, i.e., *how* things are said as opposed *what* is said. Linguistic style has also been shown (Levelt and Kelter, 1982) to be exhibited somewhat unconsciously and hence it is an interesting target for analysis, especially in the domain of online debates. We will explain the meaning of these concepts in detail in the subsequent sections.

To perform these analyses, we need the right data. That is, we need to classify debate posts into those showing agreement and those showing disagreement. Given a large set of debate posts, this problem can be solved using supervised learning. Manually labeling of posts is also possible, but it is too time consuming because we will need to label a huge number of posts in order to ensure that we have enough data to produce statistically reliable results. We take a learning approach. However, the issue is the effective features that should be used for learning. An important characteristic of the debate posts is that they almost always use some specific expressions to express agreement or disagreement, e.g., "I agree," "you're correct," etc., for agreement and "I disagree," "you speak nonsense," etc., for disagreement. Discovering such expressions clearly help improve classification. Accurate classification is essential for our subsequent analysis.

We propose to use generative models for the discovery of such expressions and use them for classification. In fact, such models themselves can be used for classification directly too. In the next section, we propose the models for modeling debate posts, which include the Naïve Bayes model (both supervised and unsupervised) and the Joint Topic Expression (JTE) model. We also report classification results. Section 3 introduces the LIWC framework (Pennebaker et al., 2007). Section 4 presents our probabilistic framework where we analyze linguistic phenomenon like stylistic cohesion, accommodation, influence, and their effect across arguing nature of debating user pairs. Section 5 concludes our work.

2 Modeling debate posts for linguistic style analysis

We employ two generative models (Sections 2.1, 2.2) to accomplish the first task of debate post classification and then generate the data for linguistic style experiments in Section 2.3. However, before proceeding, we briefly review related work on debates. Existing works have two major threads of research. The first thread puts debaters into support and oppose camps. Agrawal et al. (2003) used a graph method to place discussion participants into camps. Murakami and Raymond (2010) used a rule-based method to perform the same task. In (Somasundaran and Wiebe, 2009), opinions/polarities which were correlated with a debate-side were used to classify a post as for or against. However, this thread of research does not model agreements and disagreements in debates.



(a) Naïve Bayes (b) JTE

FIGURE 1: Graphical models in plate notations.

Another thread of research (Galley et al., 2004; Hillard et al., 2003; Thomas et al., 2006, Bansal et al., 2008; Burfoot et al., 2011) studies speaker interaction in the context of discourse and speech act classification of conversational speeches (e.g., U.S. Congress meeting transcripts). The above works mostly use three types of features: *durational* (e.g., time taken by a speaker, time separating two speakers, duration of speaker overlap, speech rate, etc.); *structural* (e.g., no. of speakers per side, no. of spurts with and without time overlap, no. of votes cast by a speaker on a bill, vote labels for and against the bill under discussion); and lexical (e.g., first word, last word, unigrams, *n*-grams, etc.) features to perform classification. While this is related to our approach of modeling agreeing and disagreeing debate posts, online debate forums (e.g., Volconco.com) are textual as opposed to conversational speeches. Thus, *durational* and *structural* features used in the prior works (e.g., time taken by a speaker, speech rate, speaker overlap, votes, etc.) are not directly applicable for our task.

Our approach relies on strong lexical features which we call AD-expressions. AD-expressions refer to Agreement (e.g., “I agree”, “you’re correct”) and Disagreement (e.g., “I disagree”, “you speak nonsense”) expressions. As AD-expressions are an integral part of debates (because while arguing people invariably emit AD-expressions), our approach aims to first mine AD-expressions which serve as strong lexical features and further exploit them to classify debate discussions into agreeing and disagreeing posts. To model debate posts and lexical AD-expressions, we use hierarchical Bayesian generative models. Generative models like LDA (Blei et al., 2003) and PLSA (Hofmann, 1999) have been proved to be very successful in modeling topics and other textual information in an unsupervised manner. For our task of modeling and classifying debate posts we compare performance using two models. The first is the Naïve Bayes model (which serves as a baseline model) and the second is our Joint Topic Expression (JTE) model.

2.1 Naïve Bayes graphical model

This section introduces the well-known Naïve Bayes model in the light of unsupervised Bayesian graphical models. In generative models for text, words and phrases (*n*-grams) are viewed as random variables, and a document is viewed as a bag of *n*-grams and each *n*-gram takes a value from a predefined vocabulary. In this work, we use up to 4-grams, i.e., n = 1, 2, 3, 4. For simplicity, we use *terms* to denote both *words* (unigrams or 1-grams) and *phrases* (*n*-grams). We denote the entries in our vocabulary by $v_{1 \ldots V}$ where V is the number of unique terms in the vocabulary. The entire corpus contains $d_{1 \ldots D}$ documents. A document (e.g., debate post) d is represented as a vector of terms W_d with N_d entries. W is the set of all observed terms with cardinality, $|W| = \sum_d N_d$. Also, let L_d denote the document class variable (a greeing or $d$isagreeing) we are trying to predict, i.e., $L_d = a$ or $L_d = d$. Lastly, let π denote the prior over document labels and φ_L the label specific distribution over vocabulary terms. Following Bayesian inference, our goal is precisely to choose L_d for W_d that maximizes $P(L_d|W_d)$. Applying Bayes rule, we get $L_d = \text{argmax}_L\, P(L|W_d) = \text{argmax}_L\, P(W_d|L)P(L)$. This lays the foundation for the generative process of the model (Figure 1a) which we detail as follows:

A. Draw $\pi \sim Beta(\alpha)$
B. For each label $L = \{a, d\}$, draw $\varphi_L \sim Dir(\beta)$
C. For each debate post $d \in \{1 \dots D\}$:
   i. Draw $L_d \sim Bernoulli(\pi)$
   ii. For each term $w_{d,j}$, $j \in \{1 \dots N_d\}$:
      a. Emit $w_{d,j} \sim Mult(\varphi_{L_d})$

To learn the model, we employ posterior inference using Monte Carlo Gibbs sampling. The samplers for L and φ_L are given as follows:

$$P(L_d = L|L_{\neg d}, W_{\neg d}, \varphi_L) \propto \frac{n_L+\alpha-1}{D+2\alpha-1} \prod_{v=1}^{V} \left(\varphi_{L,v}\right)^{n_v^d} \quad (1)$$

$$\varphi_L \sim Dirichlet(n_v^L + \beta) \quad (2)$$

where n_L is the number of documents with label L, n_v^d is the number of times term v appears in document d, and n_v^L is the number of times term v appears in all documents with label L. Learning the model according to the Gibbs sampler in (1) and (2) results in a fully unsupervised Naïve Bayes model for document label (agreeing or disagreeing) prediction. However, if we have some labeled data (more details in Section 2.3), we can add supervision into the model using a simple trick. Given a set of labeled documents, D_{Train}, where each post has a document label (i.e., agreeing or disagreeing), we can employ a supervised Naïve Bayes model keeping the label variable, L_d of the training documents fixed to the supplied labels (i.e., we do not samples $L_d, d \in D_{Train}$). Fixing the labels will effectively serve the purpose of "ground truth" evidence for the distributions that created them.

2.2 JTE: A Graphical Model for Debates

We now present the Joint Topic Expression (JTE) model, which was proposed for analyzing debates in (Mukherjee and Liu, 2012). JTE is a hierarchical generative model motivated by the joint occurrence of various topics and AD-expressions in debate posts. A typical debate post mentions a few topics (using semantically related topical terms) and expresses some viewpoints with one or more AD-expression types (using semantically related expressions). This observation motivates the generative process of our model where documents (posts) are represented as random mixtures of latent topics and AD-expression types (Agreement and Disagreement).

Assume we have $t_{1\dots T}$ topics and $e_{1\dots E}$ expression types in our corpus. Note that in our case of Volconvo.com debate posts, based on reading various posts, we hypothesize that E = 2 as in such debates, we mostly find 2 expression types: Agreement and Disagreement[1]. Let $\psi_{d,j}$ denote the distribution over topics and AD-expressions with $r_{d,j} \in \{\hat{t}, \hat{e}\}$ denoting the binary indicator/switch variable (topic or AD-expression) for the j^{th} term of d, $w_{d,j}$. In this work, a document is viewed as a bag of n-grams and we use terms to denote both words (unigrams) and

[1] The hypothesis has been statistically validated using the perplexity metric in (Mukherjee and Liu, 2012). The model is however very general and can be used with any number of expression types, e.g., for modeling review comments in (Mukherjee and Liu, 2012a) with E = 6 expression types: Agreement, Disagreement, Thumbs-up, Thumbs-down, Question, and Answer-acknowledgement.

phrases (n-grams). $z_{d,j} \sim Mult(\theta_d)$ denotes the appropriate topic ($\theta_{d,t}^T$) or AD-expression type ($\theta_{d,e}^E$) index to which $w_{d,j}$ belongs. Also let $\varphi_{t,v}^T$ and $\varphi_{e,v}^E$ denote the topic and expression type specific multinomials over the vocabulary respectively. JTE is a switching graphical model performing a switch between expressions and topics similar to that in (Zhao et al., 2010). The switch is done using a maximum entropy (Max-Ent) model. The idea is due to the observation that topical and AD-expression terms usually play different syntactic roles in a sentence. Topical terms (e.g., "U.S. senate", "marriage", "income tax") tend to be noun and noun phrases while expression terms ("I refute", "how can you say", "probably agree") usually contain pronouns, verbs, wh-determiners, and modals. In order to utilize the part-of-speech (POS) tag information, we place the topic/AD-expression distribution $\psi_{d,j}$ (the prior over the indicator variable $r_{d,j}$) in the term plate (Figure 1) and set it from a Max-Ent model conditioned on the observed feature vector $\overrightarrow{x_{d,j}}$ associated with $w_{d,j}$ and the learned Max-Ent parameters λ. In this work, we encode both lexical and POS features of the previous, current and next POS tags/lexemes of the term $w_{d,j}$. More specifically, the feature vector is $\overrightarrow{x_{d,j}} = [POS_{w_{d,j}-1}, POS_{w_{d,j}}, POS_{w_{d,j}+1}, w_{d,j} - 1, w_{d,j}, w_{d,j} + 1]$. For phrasal terms (*n*-grams), all POS tags and lexemes of $w_{d,j}$ are considered as features. The generative process of JTE (Figure 1b) is given by:

A. For each C-expression type e, draw $\varphi_e^E \sim Dir(\beta_E)$
B. For each topic *t*, draw $\varphi_t^T \sim Dir(\beta_T)$
C. For each comment post $d \in \{1 \dots D\}$:
  i. Draw $\psi_d \sim Beta(\gamma \boldsymbol{u})$
  ii. Draw $\theta_d^E \sim Dir(\alpha_E)$
  iii. Draw $\theta_d^T \sim Dir(\alpha_T)$
  iv. For each term $w_{d,j}, j \in \{1 \dots N_d\}$:
    b. Draw $r_{d,j} \sim Bernoulli(\psi_d)$
    c. if ($r_{d,j} = \hat{e}$) // $w_{d,j}$ is a C-expression term
        Draw $z_{d,j} \sim Mult(\theta_d^E)$
      else // $r_{d,j} = \hat{t}$, $w_{d,j}$ is a topical term
        Draw $z_{d,j} \sim Mult(\theta_d^T)$
    d. Emit $w_{d,j} \sim Mult(\varphi_{z_{d,j}}^{r_{d,j}})$

We employ posterior inference using Monte Carlo Gibbs sampling. Denoting the random variables $\{w, z, r\}$ by singular subscripts $\{w_k, z_k, r_k\}$, $k_{1 \dots K}$, where $K = \sum_d N_d$, a single iteration consists of performing the following sampling:

$$p(z_k = t, r_k = \hat{t} | W_{\neg k}, Z_{\neg k}, R_{\neg k}, w_k = v) \propto \frac{\exp(\sum_{i=1}^n \lambda_i f_i(x_{d,j}, \hat{t}))}{\sum_{y \in \{\hat{t}, \hat{e}\}} \exp(\sum_{i=1}^n \lambda_i f_i(x_{d,j}, y))} \times \frac{n_{d,t_{\neg k}}^{DT} + \alpha_T}{n_{d,(\cdot)_{\neg k}}^{DT} + T\alpha_T} \times \frac{n_{t,v_{\neg k}}^{CT} + \beta_T}{n_{t,(\cdot)_{\neg k}}^{CT} + V\beta_T} \quad (3)$$

$$p(z_k = e, r_k = \hat{e} | W_{\neg k}, Z_{\neg k}, R_{\neg k}, w_k = v) \propto \frac{\exp(\sum_{i=1}^n \lambda_i f_i(x_{d,j}, \hat{e}))}{\sum_{y \in \{\hat{t}, \hat{e}\}} \exp(\sum_{i=1}^n \lambda_i f_i(x_{d,j}, y))} \times \frac{n_{d,e_{\neg k}}^{DE} + \alpha_E}{n_{d,(\cdot)_{\neg k}}^{DE} + E\alpha_E} \times \frac{n_{e,v_{\neg k}}^{CE} + \beta_E}{n_{e,(\cdot)_{\neg k}}^{CE} + V\beta_E} \quad (4)$$

where $k = (d, j)$ denotes the j^{th} term of document d and the subscript $\neg k$ denotes assignments excluding the term at (d, j). Counts $n_{t,v}^{CT}$ and $n_{e,v}^{CE}$ denote the number of times term v was assigned to topic t and expression type e respectively. $n_{d,t}^{DT}$ and $n_{d,e}^{DE}$ denote the number of terms in document d that were assigned to topic t and AD-expression type e respectively. $\lambda_{1 \dots n}$ are the parameters of the learned Max-Ent model corresponding to the n binary feature functions $f_{1 \dots n}$ from Max-Ent. Omission of the latter index denoted by $(\cdot)$ represents the marginalized sum over the latter index. We employ a blocked sampler jointly sampling r and z as this improves convergence and reduces autocorrelation of the Gibbs sampler (Rosen-Zvi et al., 2004).

JTE (agreement expressions)	JTE (disagreement expressions)
agree, **I**, correct, yes, true, accept, I agree, indeed correct, **your**, **point**, I concede, is valid, **your claim**, **not really**, *would agree*, **might**, *agree completely*, yes indeed, absolutely, you're correct, *valid point*, **argument**, proves, *do accept*, support, agree with you, *rightly said*, **personally**, well put, *I do support*, *personally agree*, **doesn't necessarily**, exactly, *very well put*, absolutely correct, *kudos*, *point taken*...	**I**, disagree, I don't, I disagree, **argument**, reject, **claim**, I reject, I refute, I refuse, nonsense, *I contest*, dispute, **I think**, completely disagree, don't accept, don't agree, incorrect, *hogwash*, *I don't buy your*, *I really doubt*, your nonsense, **true**, *can you prove*, argument fails, *you fail to*, your assertions, *bullshit*, *sheer nonsense*, *doesn't make sense*, *you have no clue*, *how can you say*, *do you even*, *contradict yourself*, …

TABLE 1: Top terms (comma delimited) of two expression types. **Red (bold)** terms denote possible errors[2]. *Blue (italics)* terms are newly discovered; rest (black) were used in Max-Ent training.

2.3 Dataset Generation using Models

This section uses the models to classify agreeing and disagreeing debate posts which is a pre-requisite for this work. The hyper-parameters for the models were set to the heuristic values α = 1, β = 0.1 for NB and $\alpha_T = 50/T$, $\alpha_E = 50/E$, $\beta_T = \beta_E = 0.1$ for JTE as suggested in (Griffiths and Steyvers, 2004). For both NB and JTE, we estimate model parameters using 5000 Gibbs iterations with a burn-in of 1000. To learn the Max-Ent parameters λ, we randomly sampled 500 terms from our corpus appearing at least 10 times[3] and labeled them as topical (361) or AD-expressions (139) and used the corresponding feature vector of each term (in the context of posts where it occurs) to train the Max-Ent model. Please note that this is term-level labeling which is very different from document labels or "tags" used in LabeledLDA (Ramage et al., 2009). LabeledLDA uses tagged data from del.icio.us setting the number of topics to the number of unique labels in the corpus. It restricts document-topic distributions to be defined only over the topics that correspond to the observed document-labels. For JTE, we induce T = 100 topics and E = 2 (agreement and disagreement) AD-expression types as in debate forums, there are usually two expression types. Values for E > 2 were also tried, but they did not produce any new dominant expression type. Instead, the expression types: *disagreement* and *agreement* became somewhat less specific as the expression-term ($\Phi^E_{E\times V}$) space became sparser. There was also slight increase in the model perplexity showing that values of E > 2 do not fit the data well.

Table 1 lists some top AD-expressions discovered by JTE. We see that JTE can cluster many correct AD-expressions, e.g., "I agree", "you're correct", "agree with you", etc. in agreement and "I disagree", "I refute", "don't accept", etc. in disagreement. In addition, it also discovers and clusters highly specific and more "distinctive" expressions beyond those used in Max-Ent training (marked *blue* in italics), e.g., "valid point", "rightly said", "I do support", and "very well put" in agreement; and phrases like "I don't buy your", "can you prove," "you fail to", and "you have no clue" in disagreement. We will later see that these AD-expressions serve as high quality lexical features for debate post classification. Note that we don't quantitatively evaluate topics, perplexity of the JTE model here as our focus is to classify agreeing and disagreeing posts using discovered AD-expression for our linguistic accommodation experiments on debates.

We now turn our attention to debate post classification. In this work, we use debate forum posts from Volconvo.com. We extracted 309376 debate posts from various domains like Politics, Religion, Society, Science, etc. To evaluate model performance, we construct a validation set. We randomly sampled 2000 posts from the corpus and asked two judges (CS grad students) to

[2] Clustering errors is a known issue with unsupervised generative models for text because the objective function of the model does not always correlate well with human judgments (Chang et al., 2009).
[3] A minimum frequency count of 10 ensures that the training data is representative of the corpus.

Feature Setting	Agreement			Disagreement		
	P	R	F_1	P	R	F_1
NB-unsupervised	0.69	0.65	0.67	0.71	0.69	0.70
NB-supervised	0.72	0.73	0.72	0.75	0.76	0.75
JTE-unsupervised	0.70	0.71	0.70	0.73	0.73	0.73
W+POS 1-4 grams + SVM (all terms)	0.75	0.76	0.75	0.80	0.81	0.80
W+POS 1-4 grams + SVM + χ^2 (top 1%)	0.79	0.77	0.78	0.84	0.84	0.84
W+POS 1-4 grams + SVM + χ^2 (top 2%)	0.80	0.78	0.79	0.85	0.85	0.85
AD-Expressions, Φ^E (top 1000) + SVM	0.84	0.81	0.82	0.88	0.86	0.87
AD-Expressions, Φ^E (top 2000) + SVM	0.86	0.83	0.84	0.88	0.87	0.87

TABLE 2: Precision (P), Recall (R) and F_1 scores of different models. Improvements in F_1 using AD-expression as features (Φ^E) are statistically significant (p<0.001) using paired t-test across 5-fold cross validation.

label the overall arguing nature of each post as agreeing or disagreeing or none[4]. We obtained strong agreement using κ_{Cohen} = 0.87. This is not surprising, as debate post classification is a fairly easy task and one can almost certainly make out whether a post expresses agreement or disagreement. Finally, we deemed a post as agreeing or disagreeing if both judges deemed it so. This yielded 1268 disagreement, 621 agreement posts. Out of the rest 111 posts, 39 are labeled "none" while 72 had no consensus among judges. We evaluate our models on the validation set, D_V, of 1268 disagreement and 621 agreement (1889) posts.

We consider the following classifiers:

i) NB-unsupervised, i.e., estimating the model (document labels) directly from D_V.
ii) NB-supervised, which performs 5-fold cross validation (CV) on D_V.
iii) JTE-unsupervised, which estimates the posterior on θ_d^E over D_V and classifies a post as agreeing if $\theta_{d,e=Agreement}^E > \theta_{d,e=Disagreement}^E$ else disagreeing. We call this unsupervised because although JTE uses Max-Ent term-level supervision for switching between topics and AD-expressions, it does not use the document-labels produced by judges.
iv) SVM + W+POS n-gram. We train a SVM classifier with the linear kernel[5] using standard word and POS n-gram features and 5-fold CV.
v) SVM + W+POS n-gram + χ^2. We extend (iv) by employing feature selection using Chi-Squared test[6].
vi) SVM+AD-expressions. We induce a SVM classifier using AD-expressions as features over 5-fold CV.

For unsupervised learners (no learning), we compute precision and recall on the corresponding bin of testing for 5-fold CV. For feature selection using χ^2, and AD-expressions (as they are basically rankings from φ_e^E), we try two settings: top 1% and 2% features. Results across agreement and disagreement posts are summarized in Table 2. For SVM, we used SVM^{light} (Joachims, 1999). We see that AD-expressions+SVM performs the best. This shows that AD-expressions discovered by JTE are of high quality. Next in order is SVM + χ^2. This shows that feature selection (FS) is useful. AD-expressions can be thought of as an FS scheme where a set

[4] First posts of thread who start a topic, ambiguous, vague, partly agreeing/disagreeing posts, etc. belong to the "none" category.
[5] Polynomial, RBF, and sigmoid kernels were tired but yielded poorer results hence not reported. Linear kernel has been shown very effective for text classification problems by many researchers, e.g., (Joachims, 1998).
[6] We also tried other feature selection schemes like Information Gain, Mutual information. However, they yielded poorer results than Chi-Squared test and hence not reported.

Dimension	Examples	Size
Article	a, an , the	3
Certainty	always, never	83
Conjunction	and, but, whereas	28
Discrepancy	should, could, would	76
Exclusive	but, without, exclude	17
Inclusive	and, with, include	18
Indefinite Pronoun (Indef-Pron.)	it, those, it's	46
Negation	no, not, never	57
Preposition	to, with, above	60
Quantifier	few, many, several	89
Tentative	maybe, guess, perhaps	155
1st Person Singular Pronoun (1st-S-Pron.)	I, me, mine	12
1st Person Plural Pronoun (1st-P-Pron.)	we, our, us	12
2nd Person Pronoun (2nd-Pron.)	you, your, thou	20

TABLE 3: LIWC Style Dimensions.

of highly discriminative lexical features are selected using JTE. It is understandable that the unsupervised methods are inferior to the supervised baselines. But JTE does attain a respectable F_1 of 0.70 for agreement and 0.73 for disagreement and is better than NB-unsupervised.

We now turn to our task of generating the debate dataset (agreeing and disagreeing posts) for linguistic accommodation study. While the ideal situation would involve manually labeling all 309376 debate posts under study, it is impractical. Hence we resort to SVM+AD-expression as our classifier. Since the labeled data contains three categories, we train a multiclass SVM using our labeled data: agreement (621), disagreement (1268), and none (39) with AD-expressions. Classification on our debate corpus resulted in 123751 agreement, 177087 disagreement, and 8538 none (e.g., first posts of thread that start a topic, ambiguous, vague, partly agreeing/disagreeing posts, etc.) posts. While this classification is not perfect and may have some noise, labels on our unlabeled debate posts are sufficiently reliable as the confidence of the classifier (SVM+AD-expression) is reasonably high on the validation set. Our database consists of 7973 authors and 4387 author pairs who have debated/interacted with each other [7] and 6828 discussion threads. We now proceed to linguistic accommodation experiments.

3 LIWC: A metric for Linguistic Style

To study the general phenomenon of linguistic accommodation in debates, we need a metric for linguistic style. Following prior work on linguistic accommodation in stylometry and psycholinguistics (Niederhoffer and Pennebaker, 2002; Taylor and Thomas; 2008), we use the psycholinguistic framework Linguistic Inquiry and Word Count (LIWC) (Pennebaker et al., 2007). LIWC measures word usage in psychologically meaningful style dimensions (e.g., articles, pronouns, emotion words, etc.) and has been proven useful in the analysis of personality (Yee et al., 2010); gender, age (Mukherjee and Liu, 2010; Argamon et al., 2007); deceptive opinions (Ott et al., 2011); social relations (Scholand et al., 2010), etc. In this work, we focus on 14 strictly non-topical style dimensions detailed in Table 3 [8], i.e., we study the linguistic phenomenon of accommodation in debates over those 14 style dimensions. Please refer to (Pennebaker et al., 2007) for full list of terms. A debate post is said to exhibit a style dimension if it contains at least one word form that respective LIWC category.

[7] As it may not be interesting to study linguistic accommodation across pairs who interacted only a few times, we only consider pairs who interacted at least 20 times.
[8] Other dimensions like Family, Sexuality, Religion, etc. do not convey any style information.

4 Probabilistic Framework

This section introduces a probabilistic framework to model the linguistic phenomenon of accommodation in a principled manner.

4.1 Stylistic Cohesion

Stylistic cohesion is the general phenomenon which is grounded on the following hypothesis: Related conversations tend to be stylistically closer (hence the nomenclature, cohesion) than unrelated conversations. In the context of online debates, this transforms as follows: Related debate posts (i.e., post pairs comprising of the original post, say d and another post, say r which quotes or replies to d. Related debate posts are denoted by $d \leftrightarrow r$ from now on) exhibit significantly higher stylistic cohesion than unrelated posts. Formally, for a given style dimension, s, we can measure stylistic cohesion on s using the following probabilistic expression:

$$Coh(s) \triangleq P(d^s \wedge r^s | d \leftrightarrow r) - P(d^s \wedge r^s | d \not\leftrightarrow r) \quad (5)$$

where d^s , r^s denote the event that debate posts d, r respectively exhibit style dimension s. Thus, statistically, if the former probability expression in Eq. (5) tends to be greater than the latter, we say that related debate posts $d \leftrightarrow r$ tend to "agree" on the style dimension s. $d \not\leftrightarrow r$ denotes that d and r do not form a conversation pair. Before proceeding, it is worthwhile to test the hypothesis on our debate domain. Establishing that stylistic cohesion is exhibited in online debates corresponds to rejecting the null hypothesis that the two probabilities in Eq. (5) are equal. A two tailed t-test rejects the null hypothesis with p-value < 0.001 for all 14 style dimensions in Table 4. Table 4 (a) shows the differences of expected probabilities[9]across each style dimension over all posts in our debate database.

Having established that stylistic cohesion is exhibited in online debates, we now turn our

s	$P(d^s \wedge r^s \mid d \leftrightarrow r)$	$P(d^s \wedge r^s \mid d \not\leftrightarrow r)$	$Coh(s)$
Article	0.295	0.271	0.024*
Certainty	0.042	0.034	0.008**
Conjunction	0.212	0.176	0.036*
Discrepancy	0.069	0.062	0.007**
Exclusive	0.074	0.068	0.006**
Inclusive	0.238	0.223	0.015*
Indef-Pron.	0.278	0.261	0.017*
Negation	0.157	0.134	0.023*
Preposition	0.342	0.315	0.027*
Quantifier	0.076	0.067	0.009**
Tentative	0.097	0.091	0.006**
1st-S-Pron.	0.221	0.201	0.02*
1st-P-Pron.	0.019	0.009	0.01*
2nd-Pron.	0.124	0.120	0.004**

TABLE 4 (a)

$Coh_{Agree}(s)$	$Coh_{Diagree}(s)$
0.021	0.016
0.007	0.002
0.034	0.028
0.005	0.002
0.005	0.003
0.012	0.007
0.014	0.010
0.017	**0.019**
0.026	0.022
0.005	0.003
0.003	0.002
0.018	0.015
0.007	0.003
0.003	0.002

TABLE 4 (b)

Table 4: (a)**:** Effect of stylistic cohesion across each style dimension. The differences are statistically significant (*: p<0.0001 **: p<0.001) over two-tailed t-test. (b): Cohesion across agreeing and disagreeing debate discussions. Differences are significant p<0.001.

[9] The expectation was taken over discussion threads, i.e., the probabilities in Eq. (5) were computed for each thread and averaged over all threads in our database. We do so because $|d \not\leftrightarrow r|$ is very large $\approx \binom{309376}{2}$, for our database.

attention to the analysis of stylistic cohesion across agreeing and disagreeing debate posts. Let $d \xleftarrow{Agree} r$ denote the post pair where the post r is an agreement post and it quotes/replies to d. Analogously, we have $d \xleftarrow{Disagree} r$ when post r is a disagreement post. Extending the definition of (5), stylistic cohesion in related posts expressing agreement, $Coh_{Agree}$is given by:

$$Coh_{Agree}(s) \triangleq P\left(d^s \wedge r^s \middle| d \xleftarrow{Agree} r\right) - P(d^s \wedge r^s | d \nleftrightarrow r) \quad (6)$$

and $Coh_{Diagree}$ is given by:

$$Coh_{Disagree}(s) \triangleq P\left(d^s \wedge r^s \middle| d \xleftarrow{Disagree} r\right) - P(d^s \wedge r^s | d \nleftrightarrow r) \quad (7)$$

Table 4 (b) compares Coh_{Agree} and $Coh_{Diagree}$ across 14 style dimensions. We find that cohesion across agreeing conversations is significantly ($p<0.001$) more than cohesion in disagreeing conversations except for the style dimension Negation. This gives a very interesting insight to the phenomenon. While it is intuitive that agreeing discussions tend to have more cohesion for most style dimensions, the situation becomes reversed for negation. We believe this is so because owing to the very nature of debates, when people disagree over conversations (i.e., chain of post interactions debating via arguing and disagreeing), they try to negate the views of the other partner resulting in more cohesion.

4.2 Stylistic Accommodation

We now throw light on our key objective: linguistic style accommodation in debates. The theory of communication accommodation in linguistics (Giles et al., 1991) hinges on the general observation that during conversations/communications both textual and spoken, people try to unconsciously converge to one another's communicative behavior. In other words, there exists some coordination among conversers across a variety of dimensions like words, syntax, style, etc. Extending our probabilistic framework, we measure accommodation of a user b to another user a (where a and b are conversing pairs, i.e., they interact via reply/quote relations) as whether the stylistic dimension s in the initial post (of user a) increases the probability of s in the reply (of user b) beyond what is normally expected from user b. Formally:

$$Acc_{(a \leftarrow b)}(s) \triangleq P(d_b^s | d_a^s, d_a \hookleftarrow d_b) - P(d_b^s |\, d_a \hookleftarrow d_b) \quad (8)$$

where d_a^s and d_b^s denote the event that posts d_a and $d_b$exhibit style dimension s respectively by users a and b. $d_a \hookleftarrow d_b$ denotes the reply/quote relation from b to a. We want to emphasize the temporal aspect ($\hookleftarrow$ relation) of our modeling which offers us two crucial advantages. First, accounting the temporal aspect of accommodation (i.e., a user can accommodate to his/her conversation partner only after receiving his/her input), we minimize the effects of background style similarity like homophily. Secondly, encoding the temporal aspect gives a richer formulation than prior works (Niederhoffer and Pennebaker, 2002; Taylor and Thomas; 2008) which used correlation based measures[10]. Eq. (8) defines pairwise accommodation and can be extended to compute global accommodation for a given style dimension, s by taking the expectation over all conversing pairs, i.e.,$Acc(s) = E[Acc_{(a \leftarrow b)}(s)]$.

Establishing that linguistic phenomenon of stylistic accommodation is observable for a given dimension s is reduced to showing that $Acc(s) > 0$. Denoting $\bar{P}(d_b^s | d_a^s, d_a \hookleftarrow d_b)$ as the expected value of the subtrahend and $\bar{P}(d_b^s |\, d_a \hookleftarrow d_b)$ as the expected value of the minuend in Eq. (8),

[10] This is so because, correlation based measures do not distinguish between the case when the initial post exhibits a style dimension s but the reply/quote does not, and the reverse case when the initial post does not exhibit s but the reply does.

Table 6 (a) shows the differences in these means. The expectation is taken over all ordered pairs in our database. Differences between these means are statistically significant using a two-tailed *t*-test for all style dimensions (except 2nd person pronoun and 1st person singular pronoun. See caption of Table 6 (a)). It is clear from Table 6 (a) that accommodation exhibits significantly across major style dimensions. In fact, we find the highest accommodation in the *negation* style dimension. This is intuitive as debates usually involve a lot of negation (especially while disagreeing, more details follow in the next section).

Validating the existence of accommodation in online debates paves the way for comparative analysis of accommodation across agreeing and disagreeing pairs in debates. We define the notion of mutual accommodation $Acc_{(a,b)}(s)$ of a pair (a, b)as the expected accommodation of each with the other:

$$Acc_{(a,b)}(s) = \frac{1}{2}\left[Acc_{(b\leftarrow a)}(s) + Acc_{(a\leftarrow b)}(s)\right] \quad (9)$$

With mutual accommodation of a pair defined, we now analyze mutual accommodation across agreeing and disagreeing pairs. To perform this experiment, it is required to classify the 4387 pairs in our database into agreeing and disagreeing pairs. Before proceeding, we note the following important aspect of agreement and disagreement in debates. It reflects the intuition that when a user (say a) mostly agrees with the views of his conversing partner b (i.e., (a, b) form a conversing pair), b also agrees (or at least does not completely disagree) with a. Similarly, if a mostly disagrees with the views of b, it is highly unlikely that b completely agrees with a, i.e., b also inherently disagrees or at least does not completely agree with a. This hypothesis is grounded on the very human psychological nature of debating and arguing with others. Building on this intuition of human arguing/debating nature, it is reasonable to deem a pair of users as:

1. agreeing, if more than *k*% of their interactions (posts) exhibit agreement.
2. disagreeing, if more than *k*% of their interactions (posts) exhibit disagreement.
3. mixed, (i.e., partly agreeing and partly disagreeing), otherwise.

Choosing the threshold k is somewhat subjective as the definition of agreeing and disagreeing pairs may have different degrees of strictness according to end users. In this work, we experiment with two thresholds *k* = 65% and *k* = 75%. The thresholds are reasonable because a threshold of 50% says that the pair is mixed while if any one of the arguing nature (agreeing or disagreeing) is more pronounced then that nature is dominant. As we have the labels (agreeing or disagreeing) for each post (using our classifier in Section 2), each pair in our database was classified as agreeing, disagreeing or mixed according to the above scheme. It resulted in the following split:

k	Agreeing	Disagreeing	Mixed
0.65	1360 (31%)	2588 (59%)	439 (10%)
0.75	1141 (26%)	2676 (61%)	570 (13%)

TABLE 5: Distribution of split according to two thresholds.

As it may not be interesting to study accommodation over pairs with mixed nature (they also comprise a relatively small percentage), we focus on agreeing and disagreeing pairs. Table 6(b) shows the difference in average mutual accommodation over agreeing ($\overline{Acc}^{Agree}_{(a,b)}(s)$) and disagreeing ($\overline{Acc}^{Disagree}_{(a,b)}(s)$) pairs across each style dimension using the threshold *k* = 0.75. Table 6 (c) reports the corresponding results using the threshold *k* = 0.65. From Tables 6 (b), we note the following. For most style dimensions, mutual accommodation across agreeing pairs is more than that for disagreeing pairs. However, for 4 style dimensions, *negation*, *exclusive*, *discrepancy*, and *2nd person pronoun*, we find the trend reversed (values marked in bold).

s	$\bar{P}(d_b^s \mid d_a^s, d_a \hookleftarrow d_b)$	$\bar{P}(d_b^s \mid d_a \hookleftarrow d_b)$	$Acc(s)$
Article	0.376	0.343	0.033*
Certainty	0.145	0.113	0.032*
Conjunction	0.283	0.247	0.036*
Discrepancy	0.261	0.202	0.059*
Exclusive	0.243	0.198	0.045*
Inclusive	0.309	0.301	0.008**
Indef-Pron.	0.278	0.261	0.017*
Negation	0.257	0.178	**0.079***
Preposition	0.365	0.332	0.033*
Quantifier	0.213	0.201	0.012**
Tentative	0.176	0.169	0.007**
1st-S-Pron.	0.322	0.320	0.002
1st-P-Pron.	0.336	0.318	0.018**
2nd-Pron.	0.251	0.247	0.004

(a)

$\overline{Acc}_{(a,b)}^{Agree}(s)$	$\overline{Acc}_{(a,b)}^{Disagree}(s)$
0.019	0.015
0.021	0.019
0.027	0.025
0.017	**0.021**
0.021	**0.026**
0.005	0.003
0.011	0.007
0.035	**0.042**
0.021	0.018
0.007	0.004
0.004	0.003
0.0009	0.0006
0.011	0.008
0.001	**0.003**

(b)

$\overline{Acc}_{(a,b)}^{Agree}(s)$	$\overline{Acc}_{(a,b)}^{Disagree}(s)$
0.021	0.018
0.024	0.023
0.026	0.025
0.013	**0.016**
0.018	**0.021**
0.004	0.003
0.015	0.012
0.017	**0.023**
0.023	0.021
0.006	0.005
0.0059	0.0050
0.0008	0.0006
0.013	0.011
0.0013	**0.00018**

(c)

TABLE 6: (a): Effect of accommodation across each style dimension. The differences are statistically significant (*: p<0.0001 **: p<0.001) over two-tailed *t*-test. Table 6 (b, c): Average mutual accommodation over agreeing and disagreeing pairs using two different thresholds, *k*: Table 6 (b): *k* = 0.75 Table 6 (c): *k* = 0.65.

Disagreeing pairs happen to accommodate more. The effect is most pronounced for *negation* style dimension. We believe this is because disagreeing pairs in debates invariably emit the above 4 style dimensions to other partners who in turn also emit the same style dimensions to counter/debate. To get an intuitive feeling, we list some of the frequent expressions among disagreeing posts as follows: "**your** claim **should**", "I **would** disagree", "**you cannot exclude**", "**without** knowing", "**you** do **not**", "**I don't** accept **your**", etc. We mark the words in **red** (**bold**) which appear in the above mentioned 4 style dimensions. Lastly, we note that the results follow a similar trend for the threshold k = 0.65 used to split agree/disagree pairs (Table 6(c)). This renders confidence in our results. Also interesting to note is that the mutual accommodation differences between agreeing and disagreeing pairs have reduced when k = 0.65 (Table 6(c)) than the results using k = 0.75 (Table 6 (b)). This is not surprising as the agree/disagree pair split using k = 0.75 creates a better demarcation of pairs based on their arguing nature.

4.3 Stylistic Influence

Linguistic accommodation has a unique characteristic of asymmetry, i.e., the accommodation of a user b to another user a over a style dimension s, $Acc_{(a \leftarrow b)}(s)$ is potentially different from the accommodation of a to b over the dimension s, $Acc_{(b \leftarrow a)}(s)$. This gives rise to the notion of stylistic influence. Theoretically, considering the following probabilistic expression:

$$I_{a,b}(s) = |Acc_{(a \leftarrow b)}(s) - Acc_{(b \leftarrow a)}(s)| \quad (10)$$

when $I_{a,b}(s) > 0$, we have the following two *exclusive* cases:

Case 1: $Acc_{(a \leftarrow b)}(s) > Acc_{(b \leftarrow a)}(s)$
Case 2: $Acc_{(b \leftarrow a)}(s) > Acc_{(a \leftarrow b)}(s)$.

However, both cannot happen simultaneously. In either case, it implies that there is a difference in the amount one user accommodates to the other. Put in other words, one of the users has an "influence" over the other. If case 1 holds, then b accommodates to a more than a does to b on style dimension s, i.e., a is influencing b to emit style dimension s. If case 2 holds, then b is influencing a to emit style dimension s.

Before proceeding to analyze this interesting and fine-grained linguistic phenomenon of stylistic influence in debates, we need to statistically validate the existence of stylistic influence in online debates. Precisely, we are interested in answering the following question: In general, across various conversing pairs, is one of the users (former) forming the pair stylistically influencing (or causing the other user (latter) to accommodate) more than the extent to which he/she (the former user) is accommodating to him/her (the latter user)?

s	SA (%)	AS (%)	DA (%)	NA (%)	SA (%)	AS (%)	DA (%)	NA (%)
	(a): Agreeing				(b) Disagreeing			
Article	32	32	35	1	29	21	48	2
Certainty	40	24	34	2	33	22	41	4
Conjunction	47	22	29	2	42	23	33	2
Discrepancy	52	25	22	1	49	19	29	3
Exclusive	46	21	31	2	38	23	35	4
Inclusive	38	32	29	1	32	32	33	3
Indef-Pron.	57	13	28	2	54	15	29	2
Negation	**42**	30	26	2	**53**	7	38	2
Preposition	40	33	26	1	35	32	30	3
Quantifier	27	44	28	1	25	39	34	2
Tentative	41	27	30	2	39	27	32	2
1st-S-Pron.	35	26	38	1	33	25	41	1
1st-P-Pron.	62	14	22	2	58	8	31	3
2nd-Pron.	**33**	40	26	1	**39**	41	18	2

(a): Agreeing (b) Disagreeing

TABLE 7: Percentage of Agreeing (TABLE 7 (a)) and Disagreeing (TABLE 7(b)) pairs exhibiting different types of accommodation.

Answering the above question is reduced to the following statistical test: can we reject the null hypothesis that $E\left[I_{a,b}(s)\right] = 0$? Put in simple language, whether in expectation there is an imbalance (in either way[11]) between the amounts of accommodation among users in a conversing pair (this is the alternate hypothesis) or there is balance and each user accommodates to the other in almost the same extent in expectation (this is the null hypothesis). The key term here is "in expectation". Establishing that stylistic influence is exhibited in debates corresponds to rejecting the null hypothesis, H_1: $E\left[I_{a,b}(s)\right] = 0$. As $I_{a,b}(s)$ is an absolute value as defined in Eq. (10), we also need to test the hypothesis, H_2:

$$E\left[\max\left(Acc_{(a\leftarrow b)}(s), Acc_{(b\leftarrow a)}(s)\right)\right] = E\left[\min\left(Acc_{(a\leftarrow b)}(s), Acc_{(b\leftarrow a)}(s)\right)\right]$$

We subject H_1 and H_2 to a paired t-test over all pairs in our database. Paired t-test rejects both H_1 (with p-value < 0.0001) and H_2 (with p-value < 0.002) for all style dimensions. This empirically validates that stylistic influence is exhibited in online debates.

4.4 Accommodation across Arguing Nature

Having established that stylistic influence is exhibited in online debates, we now further investigate this intriguing phenomenon across arguing nature.

In psycholinguistic literature (Giles et al., 1991), it has been observed that when accommodation is exhibited, it can occur symmetrically (when both partners accommodate to each other) or asymmetrically (when only one of the conversers accommodates). Using our probabilistic framework, given a pair of interacting/debating user pair (a, b), the following cases arise for a style dimension s:

Case 1: Symmetry: When both $Acc_{(a\leftarrow b)}(s) > 0$ and $Acc_{(b\leftarrow a)}(s) > 0$, i.e., both of the conversers accommodate to each other.

Case 2: Asymmetry: When only one of $Acc_{(a\leftarrow b)}(s)$ or $Acc_{(b\leftarrow a)}(s)$ is > 0, i.e., only one

[11] Hence we need to employ a two-tailed t-test for testing the hypothesis.

accommodates. This further gives rise to the following two subcases. Say $Acc_{(a \leftarrow b)}(s) > 0$, i.e., b accommodates to a, then we can have:

Case 2 (a): Default asymmetry: The other non-accommodating converser maintains his "default" behavior, i.e., $Acc_{(b \leftarrow a)}(s) = 0$.

Case 2 (b): Divergent asymmetry: The non-accommodating converser accentuates his communication behavior in the opposite direction, i.e., *diverges* and $Acc_{(b \leftarrow a)}(s) < 0$

Case 3 No accommodation: None of the conversers accommodates, i.e., both $Acc_{(a \leftarrow b)}(s)$ and $Acc_{(b \leftarrow a)}(s)$ are ≤ 0.

To investigate the above cases, we compute the percentage of various forms of accommodation mentioned above across agreeing and disagreeing debating pairs in Table 7. For nomenclature, we use the following acronyms: Symmetric accommodation (SA), Default asymmetry (AS), Divergent asymmetry (DA), No accommodation (NA). However, we report results for agree/disagree pair split using threshold k = 0.75 (see Table 5) only as split using a higher threshold ensures better demarcation of agreeing/disagreeing pairs. We note the following interesting observations from Table 7 (a, b):

i) From column SA, we find that among agreeing pairs, percentage of pairs exhibiting symmetric accommodation (i.e., both members of a pair accommodating to each other) is more than that for disagreeing pairs. However, for style dimensions *negation* and *2nd person pronoun*, percentage of symmetric accommodating pairs among disagreeing pairs is more than that in agreeing pairs (shown in bold in SA column). The reason can be linked to the similar phenomenon in Section 4.2, i.e., disagreeing pairs in debates invariably emit style dimensions like *negation* and *2nd person pronoun* to other partners who in turn also emit the same style dimensions in order to counter/debate eventually resulting in somewhat symmetric accommodation.

iii) From column DA, we find that percentage of pairs exhibiting divergent asymmetry is more in disagreeing posts than agreeing posts. This is intuitive as divergent asymmetry calls for the non-accommodating converser to accentuate his communication behavior in the opposite direction so as to signal a stylistic "disagreement" along with a disagreement of views.

iv) Percentage of non-accommodating pairs among disagreeing pairs is in general more than that in agreeing pairs (See column NA in Table 7 (a, b)). This is reasonable and a plausible reason for such phenomenon is that pairs express "disagreement" in linguistic style by not accommodating at all. However, it is important here to note the following point. Earlier in Section 4.2, we showed that $Acc(s) > 0$ and accommodation is expressed in online debates. But in Table 7 we find that there are some pairs with $Acc(s) \leq 0$. It should not be considered as a contradiction to our results in Section 4.2. The key point is that we are interested in the "expected" accommodation over pairs and $E[Acc_{(a \leftarrow b)}(s)] > 0$ for all style dimensions.

Lastly, we note that the above experiments reveal no specific trend for percentage of pair exhibiting default asymmetry (DA) among agreeing and disagreeing posts based on our dataset of debate posts from Volconvo.com.

5 Conclusion

This paper studied the sociolinguistic phenomenon of accommodation in online debates. It first discussed a graphical model to perform debate post analysis to generate the required data for linguistic experiments. It then carried out a comprehensive analysis of various complex linguistic phenomena like stylistic cohesion, stylistic accommodation, influence, and accommodation across both agreeing and disagreeing debate posts. Several interesting results were obtained which dovetail with the intuitive psychology of online debaters, i.e., agreement and disagreement are also exhibited in the "style" dimension (beyond mere content) using symmetric and divergent asymmetric accommodation respectively. To our knowledge, this is the first study to report such fine-grained analysis of the linguistic phenomenon of accommodation in online debates. All experimental results were empirically validated using a large number of real-life debate posts.

References

Agrawal, R.; Rajagopalan, S.; Srikant, R.; and Xu. Y. (2003). Mining newsgroups using networks arising from social behavior. Proceedings of the International World-Wide Web Conference (WWW-2003).

Argamon, S., Koppel, M., Pennebaker, J. W., Schler, J. (2007). Mining the Blogosphere: Age, Gender and the varieties of self-expression, First Monday, 2007 - firstmonday.org

Bansal, M., Cardie, C., and Lee, L. (2008). The power of negative thinking: Exploiting label disagreement in the min-cut classification framework. In Proceedings of the International Conference on Computational Linguistics (Coling-2008): Companion volume: Posters.

Blei, D.; Ng, A.; and Jordan, M.; (2003). Latent Dirichlet Allocation. Journal of Machine Learning Research (JMLR).

Burfoot, C.; Bird, S.; and Baldwin, T. (2011). Collective Classification of Congressional Floor-Debate Transcripts. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL-2011).

Chang, J., Boyd-Graber, J., Wang, C. Gerrish, S. Blei, D. (2009). Reading tea leaves: How humans interpret topic models. Proceedings of the Neural Information Processing Systems (NIPS-2009).

Cristian Danescu-Niculescu-Mizil, Michael Gamon, and Susan Dumais. (2011). Mark my words! Linguistic style accommodation in social media. Proceedings of the International World-Wide Web Conference (WWW-2011).

Condon and Ogston. (1973). A segmentation of behavior. Journal of psychiatric research.

Galley, M.; McKeown, K.; Hirschberg, J.; and Shriberg, E. (2004). Identifying agreement and disagreement in conversational speech: Use of Bayesian networks to model pragmatic dependencies. Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL-2004).

Gonzales, A. L., J. T. Hancock, and J. W. Pennebaker. (2010). Language style matching as a predictor of social dynamics in small groups. Communication Research, 37(1):3.

Giles, H. J. Coupland, and N. Coupland. (1991). Accommodation theory: Communication, context, and consequences. In Contexts of accommodation: developments in applied sociolinguistics. Cambridge University Press, 1991.

Hale, J. and J. Burgoon. (1984). Models of reactions to changes in nonverbal immediacy. Journal of Nonverbal Behavior, 8(4):287.

Hillard, D., Ostendorf, M., and Shriberg, E. (2003). Detection of agreement vs. disagreement in meetings: Training with unlabeled data. Proceedings of Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT-2003).

Hofmann, T. (1999). Probabilistic latent semantic analysis. Proceedings of Conference on Uncertainty in Artificial Intelligence (UAI-1999).

Jaffe, J. and S. Feldstein. (1970). Rhythms of dialogue. Academic Press.

Joachims, T. (1998). Text Categorization with Support Vector Machines: Learning with Many Relevant Features. European Conference on Machine Learning (ECML-1998).

Joachims, T. Making large-Scale SVM Learning Practical. (1999). Advances in Kernel Methods - Support Vector Learning, B. Schölkopf and C. Burges and A. Smola (ed.), MIT-Press, 1999.

Levelt, W. and S. Kelter. (1982). Surface form and memory in question answering. Cognitive Psychology, 14(1):78.

Murakami A.; and Raymond, R. (2010). Support or Oppose? Classifying Positions in Online Debates from Reply Activities and Opinion Expressions. In Proceedings of the International Conference on Computational Linguistics (Coling-2010).

Mukherjee, A. and Liu, B. (2010). Improving gender classification of blog authors. Empirical Methods in Natural Language Processing (EMNLP-2010).

Mukherjee, A. and Liu, B. (2012). Mining Contentions from Discussions and Debates. KDD-2012.

Mukherjee, A. and Liu, B. (2012a). Modeling Review Comments. ACL-2012.

Niederhofer, K. and J. Pennebaker. (2002). Linguistic style matching in social interaction. Journal of Language and Social Psychology.

Ott M., Choi Y., Cardie C., Hancock J. T. (2011). Finding deceptive opinion spam by any stretch of the imagination, Proc. of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT-2011).

Pennebaker, J. W., R. J. Booth, and M. E. Francis. (2007). Linguistic Inquiry and Word Count (LIWC): A computerized text analysis program. LIWC.net, 2007.

Ramage, D.; Hall, D.; Nallapati, R; Manning, C. (2009). Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora. Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP-2009).

Rosen-Zvi, M.; Griffiths, T.; Steyvers, M.; and Smith, P. (2004). The author-topic model for authors and documents. Proceedings of Conference on Uncertainty in Artificial Intelligence (UAI-2004).

Somasundaran, S. and Wiebe, J. (2009). Recognizing stances in online debates. Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL-2009).

Scholand, J. A., Tausczik, Y. R. and Pennebaker, J. W. (2010). Social language network analysis. In Proceedings of CSCW, pages 23–26, 2010.

Taylor, P. and S. Thomas. (2008). Linguistic style matching and negotiation outcome. Negotiation and Conict Management Research, 1(3):263.

Yee, N., Harris, H., Jabon, M. and Bailenson, J. (2010). The Expression of Personality in Virtual Worlds. Social Psychological & Personality Science (in press).

Zhao, X., Jiang, J. Yan, H., Li, X. (2010). Jointly modeling aspects and opinions with a MaxEnt-LDA hybrid. EMNLP. 2010.