Detecting Emotions in Social Media: A Constrained Optimization Approach

Yichen Wang
Georgia Institute of Technology
Atlanta, GA
yichen.wang@gatech.edu

Aditya Pal
IBM Research
San Jose, CA
aditya.pal@gmail.com

Abstract

Emotion detection can considerably enhance our understanding of users' emotional states. Understanding users' emotions especially in a real-time setting can be pivotal in improving user interactions and understanding their preferences. In this paper, we propose a constraint optimization framework to discover emotions from social media content of the users. Our framework employs several novel constraints such as emotion bindings, topic correlations, along with specialized features proposed by prior work and well-established emotion lexicons. We propose an efficient inference algorithm and report promising empirical results on three diverse datasets.

1 Introduction

Social media enables its users to share their views, opinions, events and other personal information online – taking over the role of personal diaries due to their widespread dissemination, ease of use and ubiquitous nature. Such reportage from the daily lives of individuals can enhance our understanding of users' emotional states, which can be useful in several scenarios, such as for detecting emotional crisis in users' life, doctor-patient interactions, customer-care interactions where a real-time feedback (and subsequent adjustments) can lead to a satisfactory experience [Sadilek *et al.*, 2013]. Emotion models can be employed to understand how users feel about a given entity such as a movie, or a live event. Consequently, there is significant value to be derived from analyzing and predicting human emotions.

Researchers have focused on identifying key features that could serve as signals for emotion detection, such as words indicative of different emotions, smiles. These features have been used effectively as side information for building more complex models. However, for different datasets, different models perform better [Kim *et al.*, 2010]. Moreover, there is no standard framework that can leverage the dictionaries and features proposed by the prior research work and address all the key challenges listed below.

Challenges: We highlight the challenges in emotion detection through the sentences: (S1) "*Recent plane crashes makes me sad and angry*" and (S2) "*Yet another plane crashed*".

- *Multiple emotions*: S1 exhibits emotions: *sadness* and *anger*. There are scenarios where multiple emotions are desired. However, majority of prior work that is based on classification models can detect only one emotion and hence they only provide incomplete information.
- *Topic correlation*: The two sentences S1 and S2 are topically correlated. Most prior work would detect the emotions in S1 correctly. However they would fail in detecting any emotion in S2. Incorporating topical interdependencies can be crucial for discovering the base emotion that a topic evokes. There is some prior work recently that aims to address the topical aspect [Bao *et al.*, 2009].
- *Emotion bindings*: Co-existence of emotions differs based on their type. For e.g., it is rare for *joy* and *fear* to co-exist in a sentence. Moreover some emotions can co-occur with a greater flexibility. For e.g. *surprise* can occur with *joy* as well as with *anger*. Most prior work ignores this binding.
- *Noisy labels*: Typically the ground truth for emotions is constructed through surveys. However, emotion perception is very personal and could vary from one person to another. It can be hard for a model to reconcile this variance (or say noise) in the ground truth.

Contributions: To our knowledge, this paper is the first to address all the challenges listed above through a general constraint optimization framework which can leverage new features and refined emotion lexicons learnt by prior researchers.

- We use the vector representation of emotions, which allows ranking and thresholding of emotions in the text. Hence our model can be crafted to get a single emotion or multiple emotions. This is an extremely desirable feature depending on the use-case.
- We propose several novel constraints such as topical, emotional, bias factors, which directly address the above-mentioned challenges and leads to a much-improved performance of our model.
- We propose an efficient inference algorithm based on multiplicative update rule and prove its convergence. We also demonstrate a generic approach to tune model parameters automatically from the training dataset.
- Finally, we evaluate our model through three diverse real world datasets, and show that it outperforms existing state-of-art methods for emotion detection. We also rigorously test each component of our model and show its robustness to noisy ground truth labels.

2 Related Work

We review several key areas that are closely related to the problem of emotion detection.

Sentiment Analysis. Sentiment analysis aims at discovering the contextual polarity of the documents [Pang and Lee, 2008]. [Li *et al.*, 2009] proposed a Non-negative Matrix Factorization (NMF) approach which leverages lexical knowledge for sentiment classification. Recent work [Bollen *et al.*, 2011; Golder and Macy, 2011] has focused on mining temporal and seasonal trends of sentiment. Sentiment analysis is a closely related problem, however emotions are much more expressive than sentiments. Moreover, emotions need not contain a sentiment and vice-versa [Liu, 2012].

Emotion Detection. Emotion models are primarily of two types [Ekkekakis, 2013]: (i) dimensional, and (ii) categorical. Dimensional models represent emotions on three dimensions: *valence*, *arousal* and *dominance*. [De Choudhury *et al.*, 2012a] extended it to circumplex model and studied various aspects of the relationship between mood expression and human behavior in social media. Categorical models represent emotions into finite categories. Ekman's basic emotion set (*anger, disgust, fear, happiness, sadness, and surprise*) is arguably the most popular emotion taxonomy [Ekman, 1992].

There is a large body of prior work on emotion classification. [Mishne, 2005; Mishne and De Rijke, 2006] used LiveJournal data to build models which predict levels of various moods and understand their seasonal cycles. [Strapparava and Mihalcea, 2008] proposed several knowledge-based and corpus based methods. [Kozareva *et al.*, 2007] presented a method based on the co-occurrence distribution over content and emotion words. Machine learning techniques such as Support Vector Machines are widely applied [Yang *et al.*, 2007; Aman and Szpakowicz, 2008; Lin *et al.*, 2007]. [Agrawal and An, 2012] proposed an unsupervised model based on semantic relatedness between words and the emotion concepts. However, these approaches do not consider the topical relationship across words. [Bao *et al.*, 2009] overcame this limitation and proposed an emotion-topic model by augmenting Latent Dirichlet Allocation with an intermediate emotion layer.

There has been concerted effort on building dictionaries of words with their corresponding emotion categories, such as LIWC [Pennebaker *et al.*, 2007], WordNet-Affect [Strapparava and Valitutti, 2004], ANEW [Bradley and Lang, 1999], EmoSenticNet [Poria *et al.*, 2014], NRC [Mohammad and Turney, 2013]. There are some approaches based on emotion lexicons. [Alm *et al.*, 2005; Aman and Szpakowicz, 2007] incorporated emotional words as features. [Chaumartin, 2007] combined lexicons with linguistic and rule-based approach. [Mihalcea and Liu, 2006] predicted semantic trends of happiness. [Alm, 2008] used a hierarchical sequential model along with SentiWordNet [Esuli and Sebastiani, 2006]. [De Choudhury *et al.*, 2012b] developed an affect classifier utilizing ANEW and NRC.

Our work differs from prior work in several important ways. First, we propose a model, which improves dictionary-based approaches to classifying text into emotion categories. Second, we endow the model with several novel constraints, such as topic, emotion, and bias factors. Third, we propose an efficient algorithm that is guaranteed to converge to local optimal solution and learns the model parameters automatically. Our framework is quite flexible in the sense that it can incorporate any emotion lexicon and features from prior work and can be used to discover the existence of multiple emotions in input text.

Symbol	*Description*
n	number of documents
k	number of emotion categories
m	number of features
$l \leq m$	number of words in the emotion lexicon
$\mathcal{X} \in \mathbb{R}_+^{m \times n}$	feature document matrix
$\mathcal{D} \in \mathbb{R}_+^{l \times k}$	word-emotion lexicon
$\mathcal{S} \in \mathbb{R}_+^{n \times k}$	document-emotion ground truth
$\mathcal{Q} \in \mathbb{R}_+^{k \times k}$	emotion binding probability
$\mathcal{W} \in R_+^{n \times n}$	topic similarity between all pairs of documents

Table 1: Notations.

3 Approach

Table 1 lists notations used in this paper. $\mathcal{X}$ indicates the sentence features: ngrams, smiles, #exclamation mark, question mark, curse words, greeting words, sentiment polarity. The grams are normalized using tf-idf scheme. $\mathcal{D}$ represents the NRC word-emotion lexicon [Mohammad and Turney, 2013].

Multiple Emotions: Let $D \in \mathbb{R}_+^{m \times k}$ be the latent word-emotion matrix and $S \in \mathbb{R}_+^{n \times k}$ be the latent document-emotion matrix. These two matrices contain scores for each word/document per emotion category. NMF based prior work [Kim *et al.*, 2010; Li *et al.*, 2009] considers the formulation $||\mathcal{X} - DS^T||_F^2$, which penalizes a document unless it contains *all* the words pertaining to the document's emotion - casting an unreasonable expectation on the model. This could lead to a skew in scores towards emotions with smaller lexicon. A more reasonable way to derive a document's emotion is to get each feature's emotion and average over them according to the frequency of their appearance. This doesn't enforce an unrealistic expectation of mentioning all emotionally related words. Hence we consider that S is drawn from the product of $\mathcal{X}^T$ and D with a random i.i.d. Gaussian noise, leading to the following formulation.

$$\min_{D \succeq 0, S \succeq 0} \|S - \mathcal{X}^T D\|_F^2 \tag{1}$$

where $||M||_F$ is the $Frobenius$ norm of matrix M. Next, we describe several novel constraints for our problem.

Topic Constraint: We hypothesize that documents that belong to same topics must exhibit similar emotions. So for two similar documents, their emotion distribution $S_{(i\cdot)}$ and $S_{(j\cdot)}$ must be similar, i.e. $\mathcal{W}_{ij} \sim (SS^T)_{ij}$. This constraint is only applicable to topically related documents. To operationalize it, we consider a binary indicator matrix L, s.t., $L_{ij} = 1$ for $\mathcal{W}_{ij} \geq \tau$, otherwise 0. Here τ is the similarity threshold. The topic constraint is captured through the constraint: $\|L \circ (\mathcal{W} - SS^T)\|_F^2$, where $X \circ Y$ is the Hadamard product or the component wise product of X and Y. There are instances

where one topic might have views of opposing polarity, so for these instances L_{ij} can be set to 0.

Emotion Bindings: Based on our observations, we hypothesize that emotions can be highly correlated (negatively and positively). For e.g., it is highly unlikely that a document exhibiting *joy* would exhibit *sadness* or *anger*. Let $\mathcal{Q}$ indicate the emotion co-existence probability matrix which can be learnt via the word emotion lexicon $\mathcal{D}$ or through domain knowledge. The emotion constraint is then specified by the term: $\|\mathcal{Q} - D^T D\|_F^2$.

Emotion Lexicon and Ordering: The mapping of emotions to the columns of the latent emotion matrices D and S is unknown. To fix this mapping, we use the word emotion lexicon $\mathcal{D}$ for which the mapping of its columns to emotions is known. Additional benefit of the emotion lexicon is that it allows our model to encode prior knowledge about the word-emotion categories. The following emotional constraint is added to the model: $\|\mathcal{D} - D(l)\|_F^2$, where $D(l)$ indicates the first l rows of D. We order rows of $\mathcal{X}$ to first represent the l emotion lexicon words and then other features. Typically, emotion lexicons are collected through manual surveys over a limited set of documents from a single domain or a subset of topic. Hence they can be very context specific and our experience suggests that a model based *just* on them doesn't perform well. However they can be quite effective in guiding our model. Similarly, we encode ground truth emotion of the documents ($\mathcal{S}$) through the constraint $\|\mathcal{S} - S\|_F^2$.

Bias Factors: There can be several bias factors that lead to the inclusion of a feature in a document. For e.g., popular words have higher probability of being mentioned than rare words. To handle this, we consider multivariate half-normally distributed random variables $u \sim |\mathcal{N}(\bar{u}, \sigma_u^2 I)| \in \mathbb{R}_+^m$ and $v \sim |\mathcal{N}(\bar{v}, \sigma_v^2 I)| \in \mathbb{R}_+^n$. Adding the bias factors leads to the following formulation.

$$\min_{(D,S,u,v)\geq 0} \|S\text{-}(\mathcal{X}^T\text{-}vu^T)D\|_F^2 + \lambda_u\|\bar{u}\text{-}u\|_2^2 + \lambda_v\|\bar{v}\text{-}v\|_2^2 \quad (2)$$

where $\lambda_u = (2\sigma_u^2)^{-1}$ and $\lambda_v = (2\sigma_v^2)^{-1}$. The bias factors extract the baseline characteristics of the text and helps in addressing the noise by mitigating the effect of rare words.

Based on the above discussion, our model can be described by the following constraint optimization problem.

$$\min_{D\geq 0, S\geq 0, u\geq 0, v\geq 0} \left\{ \mathbf{\Phi}(\mathcal{X}, \mathcal{S}, \mathcal{D}, \mathcal{W}, \mathcal{Q}, \bar{u}, \bar{v}; \Lambda) \right\} \quad (3)$$

where,

$$\begin{aligned}\mathbf{\Phi} = &\|S - (\mathcal{X}^T - vu^T)D\|_F^2 + \lambda_u\|\bar{u} - u\|_2^2 + \lambda_v\|\bar{v} - v\|_2^2 \\ &+ \lambda_d\|\mathcal{D} - D(l)\|_F^2 + \lambda_s\|\mathcal{S} - S\|_F^2 \\ &+ \lambda_w\|L \circ (\mathcal{W} - SS^T)\|_F^2 + \lambda_q\|\mathcal{Q} - D^T D\|_F^2\end{aligned}$$

We note that $\Lambda = \{\lambda_u, \lambda_v, \lambda_d, \lambda_s, \lambda_w, \lambda_q\}$ are the regularization parameters, that control emphasis on the different constraints. For notational simplicity, we omit the parameters of $\mathbf{\Phi}$, unless necessary.

3.1 Convex Sub-Problem

The objective function $\mathbf{\Phi}$ is non-convex in D and S. We derive a convex sub-problem in each variable through a variable substitution technique.

$$\begin{aligned}\mathbf{\Psi} = \mathbf{\Phi} &- \lambda_w\|L \circ (\mathcal{W} - SS^T)\|_F^2 - \lambda_q\|\mathcal{Q} - DD^T\|_F^2 \\ &+ \lambda_w\|L \circ (\mathcal{W} - SA^T)\|_F^2 + \lambda_q\|\mathcal{Q} - DB^T\|_F^2 \\ &+ \lambda_a\|S - A\|_F^2 + \lambda_b\|D - B\|_F^2 \quad (4)\end{aligned}$$

$\mathbf{\Psi}$ is second-order convex function in all its variables though not convex collectively. Note that $\mathbf{\Phi} = \mathbf{\Psi}(A = S, B = D)$.

Lemma 1. *The problem $\mathbf{\Psi}^\star = \min\{\mathbf{\Psi}\}$ provides a lower bound to the problem $\mathbf{\Phi}^\star = \min\{\mathbf{\Phi}\}$.*

Proof. Let $D^\star$, $S^\star$ be the solution to $\mathbf{\Phi}^\star$. As a solution to the minimization problem, we have

$$\begin{aligned}\mathbf{\Psi}^\star &\leq \min\{\mathbf{\Psi}(D{=}D^\star, S{=}S^\star)\} \\ &\leq \min\{\mathbf{\Psi}(D{=}D^\star, S{=}S^\star, A{=}S^\star, B{=}D^\star)\} = \mathbf{\Phi}^\star\end{aligned}$$

Solving $\mathbf{\Psi}$ ensures a parsimonious fit to $\mathbf{\Phi}$. □

3.2 Inference Algorithm

We propose the multiplicative update rule, which provides a good compromise between speed and ease of implementation[1]. The update rule is constructed by placing the negative part of the derivative ($\nabla\mathbf{\Psi}$) in the numerator and the positive part in the denominator. For e.g., update rule for S is as follows.

$$S \leftarrow S \circ \frac{\mathcal{X}^T D + \lambda_w L \circ \mathcal{W}A + \lambda_a A + \lambda_s \mathcal{S} + \epsilon}{vu^T D + \lambda_w L \circ (SA^T)A + (\lambda_a{+}\lambda_s{+}1)S{+}\epsilon} \quad (5)$$

Note that here division is also component wise. As suggested by prior work [Pauca *et al.*, 2006], we add a small positive constant $\epsilon = 10^{-5}$ to prevent division by zero. Since all variables are updated in this multiplication form, non-negativity is always satisfied. Algorithm 1 presents our inference algorithm to obtain a locally optimal solution.

Algorithm 1 MINIMIZE $\mathbf{\Psi}$

1: Initialize $S^{(0)}, D^{(0)}, u^{(0)}, v^{(0)}, A^{(0)}, B^{(0)}$ randomly
2: $t = 0$
3: **repeat**
4: $\quad t = t + 1$
5: $\quad$ Compute $S^{(t)}$ using multiplicative rule (Eq 5).
6: $\quad$ Similarly compute $D^{(t)}, u^{(t)}, v^{(t)}, A^{(t)}, B^{(t)}$.
7: **until** $\mathbf{\Psi}^{(t-1)} - \mathbf{\Psi}^{(t)} \leq \epsilon$ or $t \geq$ maxIteration
8: **return** $S^{(t)}, D^{(t)}, u^{(t)}, v^{(t)}, A^{(t)}, B^{(t)}$

Theorem 1. *Algorithm 1 is guaranteed to converge to a locally-optimal solution.*

We first define an auxiliary function which is similar to that used in the EM algorithm [Dempster *et al.*, 1977].

Definition 1. *$G(x, y)$ is an auxiliary function for $F(x)$ if the following conditions are satisfied.*

$$G(x, y) \geq F(x), \qquad G(x, x) = F(x)$$

[1]The original algorithm from [Lee and Seung, 2001] is specific to NMF without constraints. Our alogorithm improves over NMF through the constraints.

Lemma 2. *If G is an auxiliary function, then F is non-increasing under the update*

$$x^{t+1} = \arg\min_x G(x, x^t) \quad (6)$$

Proof. $F(x^{t+1}) \leq G(x^{t+1}, x^t) \leq G(x^t, x^t) = F(x^t)$. □

Clearly, if the update rule confirms to Eq. 6 then F would converge to a local minima. Since all the variables are sequentially updated, it is sufficient to show that update rule for S confirms to Eq. 6 for an appropriate auxiliary function.

Lemma 3. *Let $s = S_{ij} > 0$. The function $G(x, s)$ is an auxiliary function for $F(s)$=$\mathbf{\Psi}(S_{ij}$=$s)$.*

$$G(x,s) = F(s) + \frac{\partial F(s)}{\partial s}(x-s) + \frac{(vu^T D + \lambda_w L \circ SA^T A + (\lambda_a + \lambda_s + 1)S + \epsilon)_{ij}}{s}(x-s)^2$$

Proof. Clearly, $G(s, s) = F(s)$. Taylor expansion of F is:

$$F(x) = F(s) + \frac{\partial F(s)}{\partial s}(x-s) + \frac{1}{2}\frac{\partial^2 F(s)}{\partial s^2}(x-s)^2$$

Note that the higher order terms in the Taylor expansion vanishes because F is a second order polynomial in s. In order for G to be an auxiliary function, $G(x, s) \geq F(x)$. Alternately, we need to show that

$$\frac{(vu^T D + \lambda_w L \circ SA^T A + (\lambda_a + \lambda_s + 1)S + \epsilon)_{ij}}{s} \geq \frac{1}{2}\frac{\partial^2 F}{\partial s^2}$$

It is easy to see above using the following inequalities.

$$\frac{(SM^T M)_{ij}}{s} = \frac{1}{s}\sum_k S_{ik}(M^T M)_{kj} \geq (M^T M)_{jj}$$

$$\frac{(uv^T D)_{ij} + (\lambda_a + \lambda_s + 1)S_{ij} + \epsilon}{s} \geq \lambda_a + \lambda_s + 1$$

This establishes that G is an auxiliary function for F. □

Proof of Theorem 1. To show that Algorithm 1 converges, we need to show that update rule for S follows Eq. 6. Solving $dG(x, s^{(t)})/dx = 0$ for x, we get the next update as mentioned in Eq. 5. Since G is the auxiliary function for F, so F is non-increasing under this update rule. Thus Algorithm 1 converges to a locally optimal solution. □

Parameter Learning: We consider EM style approach to estimate the parameters of our model. For e.g., λ is estimated through the minimization: $\min_\lambda\{\min \mathbf{\Psi} - \log Pr(\lambda)\}$, where $Pr(\lambda) \sim Beta(\alpha_\lambda, \beta_\lambda)$ is the prior. We alternate between Algorithm 1 and parameter learning until convergence, which is typically 3-4 iterations.

Computational Complexity: The document feature matrix $\mathcal{X}$ is typically very sparse with $p \ll mn$ non-zero entries and $k < 10$. Using sparse matrix multiplications, the updates can be very efficiently and easily implemented. Specifically, the cost of one iteration is $O(k^2(m+n) + kp)$. Empirically, the number of iterations required for convergence is (~ 100). Hence our approach scales linearly to dataset size.

Prediction. For predicting the emotion categories of a test document, we can use the locally-optimal D, S, u, v to solve the following minimization.

$$\min_{Z \succeq 0} \left\{ \|Z - (\mathcal{X}_{test}^T - v_0 u^T)D\|_2^2 + \lambda_w \|\mathcal{W}_{test} - SZ^T\|_2^2 \right\}$$

where v_0 is the mean of v and $\mathcal{W}_{test} \in \mathbb{R}_+^n$ is the topic similarity of the test document with the similar training documents. Only documents with similarity of τ or greater with the test documents are retained in $\mathcal{W}_{test}$. One clear advantage of our setup is that predictions can be computed extremely fast – making our approach desirable in an online setting.

4 Dataset

We consider 3 diverse datasets to evaluate our approach.

SemEval[2]: This dataset consists of 1250 news headlines annotated by human coders on six emotions: *anger, disgust, fear, joy, sadness* and *surprise* [Strapparava and Mihalcea, 2007]. The news headlines are typically short and written to evoke emotions and attract readers' attention. The dataset contains score of each emotion for each headline on a 100-point scale. We tag each headline with emotions that have a score greater than the average emotion score of that headline.

ISEAR[3]: This dataset consists of 7666 sentences annotated by 1096 participants with different cultural backgrounds [Scherer and Wallbott, 1994]. They completed questionnaires about experiences and reactions for seven emotions: *anger, disgust, fear, joy, sadness, shame and guilt*. In contrast to *SemEval*, each experience is tagged with a single emotion.

Twitter Dataset: We collected a sample of 1800 tweets using the Twitter API. In order to ensure that the tweets express some emotion, we first randomly sampled 30 words per emotion category from NRC lexicon and then randomly picked 10 tweets that mentioned the selected emotion words in a hashtag format (e.g. #sad). The tweets were coded on 6 emotion categories (same as *SemEval*) and 1 no-emotion category by 3 mechanical turkers. The inter-rater agreement is 0.62, indicating a moderate agreement. We took the majority vote to pick the emotion categories of tweets.

5 Results and Evaluation

We use 10-fold cross validation to run our experiments and report precision (*P*), recall (*R*), and F-score (*F*). All significance tests are done through one-sided test with 95% confidence interval. To operationalize our model, we first filtered the rare words (i.e. words that appeared in less than 5 sentences). We use cosine similarity between the documents' features to compute their topic similarity and set threshold τ=0.8. We compare our model with several baselines as described below.

b1: This model uses a non-Linear SVM over n-gram features for classifying emotions.

b2: This is the best prior state-of-art model. For *SemEval*, we report the results from [Kim *et al.*, 2010]. For *ISEAR*, we report the results from [Agrawal and An, 2012].

[2]http://web.eecs.umich.edu/∼mihalcea/downloads.html

[3]http://www.affective-sciences.org/researchmaterial

		SemEval			*ISEAR*		
	Model	P	R	F	P	R	F
Anger	*our*	**0.57**	0.44	**0.50**	**0.71**	**0.68**	**0.69**
	b1	0.21	0.20	0.20	0.46	0.58	0.51
	b2	0.29	0.26	0.28	*na*	*na*	0.63
	b3	0.22	**0.85**	0.35	0.19	0.59	0.29
	b4	0.50	0.52	0.51	0.58	0.64	0.61
Fear	*our*	**0.59**	0.66	**0.62**	**0.78**	**0.78**	**0.78**
	b1	0.32	0.24	0.27	0.68	0.58	0.63
	b2	0.53	0.75	**0.62**	*na*	*na*	0.59
	b3	0.37	**0.95**	0.54	0.22	0.09	0.12
	b4	0.55	0.65	0.60	0.69	0.69	0.69
Joy	*our*	0.58	0.71	0.64	**0.71**	**0.81**	**0.76**
	b1	0.30	0.18	0.23	0.61	0.63	0.62
	b2	**0.77**	0.56	**0.65**	*na*	*na*	0.56
	b3	0.51	**0.94**	0.64	0.20	0.21	0.20
	b4	0.56	0.62	0.59	0.66	0.70	0.68
Sad	*our*	**0.57**	**0.80**	**0.65**	**0.77**	**0.67**	**0.72**
	b1	0.24	0.14	0.16	0.64	0.51	0.57
	b2	0.50	0.45	0.47	*na*	*na*	0.41
	b3	0.45	**0.80**	0.57	0.20	0.06	0.09
	b4	0.56	0.54	0.55	0.71	0.54	0.65
Avg.	*our*	**0.61**	0.65	**0.63**	**0.74**	**0.73**	**0.74**
	b1	0.26	0.19	0.22	0.60	0.57	0.58
	b2	0.52	0.50	0.51	*na*	*na*	0.54
	b3	0.39	**0.90**	0.53	0.20	0.21	0.20
	b4	0.57	0.59	0.58	0.69	0.64	0.66

Table 2: Performance of models over *SemEval* and *ISEAR*.

b3: This is the lexicon based model. It tags each document with the weighted average of lexicon words' emotions.

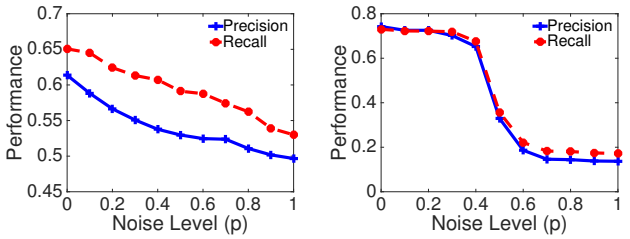
b4: This is the NMF based model [Kim *et al.*, 2010; Li *et al.*, 2009]. We append this model with all the constraints of our model in order to make it at par with our approach.

Table 2 shows the performance of the different models. We observe that the average performance of our model is statistically significantly better than the best reported accuracy (*b2*) across the two datasets, using one-sided *ttest* with $p \sim 0$. We note that the bag-of-word (*b1*) model performs poorly for the *SemEval* dataset although it has good performance on *ISEAR*. The improved performance of our model across these different datasets highlights its superiority and its general applicability for emotion classification.

Model Analysis. We investigate the sensitivity of our model with regard to all various constraints it employs. We turn the constraint off and measure the performance of the model (see Table 3). The result shows that all the constraints are significantly important in improving the model performance. Among the constraints, topic constraint provides the largest boost to the performance.

	SemEval			*ISEAR*		
	P	R	F	P	R	F
$w/o\ u, v$	0.57	0.62	0.59	0.72	0.72	0.72
$\lambda_q = 0$	0.56	0.57	0.57	0.70	0.69	0.70
$\lambda_d = 0$	0.56	0.61	0.58	0.71	0.71	0.71
$\lambda_w = 0$	0.51	0.54	0.53	0.65	0.66	0.66

Table 3: Model performance without constraints.



(a) *SemEval* (b) *ISEAR*

Figure 1: Model performance with different noise levels.

5.1 Noisy Ground Truth

The emotion ground truth is typically gathered by asking the human coders. Since emotion classification can be tricky, we see large disagreements between the coders. As a result ground truth can be noisy (high false positives and false negatives). In order to emulate the noisy setting, we add noise to the training dataset. For *SemEval*, we set $\mathcal{S}_{ij} = \min\{1, \max\{0, \mathcal{S}_{ij} + Y\}\}$, where $Y \sim \mathcal{N}(p, 0.1)$ is a Gaussian random variable. $\mathcal{S}$ is a binary for *ISEAR*, so we flip its cells with probability p. *Noise is not added to the test data.*

Figure 1 shows that our model performs well even when there is 40% noise in the training data – indicating that our model is robust to large levels of noise. One intuitive reason for such high tolerance is that the model is not solely dependent on the training labels. It has several other constraints that ensure that model doesn't over fit.

5.2 Evaluation on Microblog Dataset

We perform a qualitative evaluation of our model on the Twitter dataset. This evaluation serves multiple purposes. It shows how our model performs over short snippets of user-generated content that is often mired with sarcasm, and conflicting emotions. Additionally, It enables us to closely examine the scenarios where our model misclassified and scenarios where the ground truth appeared to be incorrect.

Model Performance. Table 4 shows the performance of our model in comparison to the baselines. We note that our model is statistically significantly better. It improves recall substantially by 42% over the NMF based prior state-of-art and precision by 5%.

our		*b1*		*b3*		*b4*	
P	R	P	R	P	R	P	R
0.43	**0.67**	0.16	0.14	0.33	0.27	0.40	0.47

Table 4: Performance of models over *Twitter* dataset.

Qualitative Analysis. We present several interesting examples where our model surprisingly fared well and some of those instances in which we intuitively see that the ground truth appears to be either incorrect or non-specific.

Sentence 1. *Update AmberAlert #Abduction of 2yo Edwin-Vargas LosAngeles CA: Car found abt 1.5hrs ago in East L.A.*

The ground truth label for the above sentence is *no emotion*. However, our model labeled it as *fear* and *sad*. This happened because the model associated the words *amberalert*

Anger	bitch, assault, choke, crazy, bomb, challenge, bear, contempt, broken, banish
Disgust	coward, bitch, bereavement, blight, abortion, alien, banish, contempt, beach, abduction
Fear	concerned, abduction, alien, alarm, birth, abortion, challenge, crazy, bear, assault
Joy	abundance, art, calf, birth, cheer, adore, choir, award, good, challenge
Sad	broken, bereavement, crash, choke, concerned, banish, blight, abortion, crazy, bitch
Surprise	shock, alarm, crash, alien, bomb, climax, abduction, good, cheer, award

Table 5: Top 10 words per emotion category sorted in decreasing order of their importance.

and *abduction* with the emotion *fear* and *sad*. The sentence clearly reflects that human coders made a mistake.

Sentence 2. *I want to believe! UFO Spotted Next To Plane In Australia #UFO #Alien #space #Computer.*

Sentence 3. *SHOCKING Statements from The #Tyrant of the USA! #obama Declares Absolute Authority of the #NewWorldOrder Over Us.*

The ground truth and the model agreed on the above two sentences. For sentence 2, labels were *surprise* and *joy* and for sentence 3 they were *surprise* and *anger*. Clearly model could infer that the emotion *surprise* could co-occur with the two contradicting emotions: *joy* and *anger*.

Sentence 4. *We ran amazing #bereavement training yesterday for our #caregivers. This is another way we support our amazing team!*

Our model labeled this sentence as *joy* and *surprise*, which is the same as ground truth. Although the word *bereavement* (which has negative connotations in our dictionary) is in the sentence, the model could automatically place an emphasis on the positive word *amazing* and make correct prediction.

Sentence 5. *Give me and @AAA more Jim Beam you bitch #iloveyou #bemine #bitch.*

Our model labeled this sentence as *anger*, while the ground truth label is *joy*. A possible explanation for the incorrect labeling by the model is that it associated the word *bitch* with *anger* (see Table 5). Moreover, *bitch* occurs twice in the tweet while *love* is mentioned only once.

Frequently Occurring Words. Table 5 shows the top words that the model associates with each emotion category. Intuitively, we see that the top words are quite indicative of their emotion category. We also observe that some emotion categories share words. For instance, *contempt* appears both in *anger* and *disgust*. Similarly *bitch* appears both in *anger*, *disgust* and *sadness*. Moreover, we note that *surprise* appears to exhibit positive as well as negative sentiment polarity. It contains positive polarity words like *climax* and *award* and negative polarity words like *bomb* and *abduction*.

We also note that a word can be used to indicate conflicting emotions. For example, the word *challenge* is utilized to express both *fear* and *joy*. We list some tweets that highlight the conflicting situations in which words can be used.

Anger	deadbeat, asshole, mean, salvage, authority, mandate, charge, old, hammer, snowden
Disgust	trouble, narcissism, bill, leprosy, snowden, pouring, stop, referee, blogger, unreal
Fear	raven, dwindle, hacking, strange, parent, capitalism, salvage, diet, misinterpret, desperate
Joy	congratulation, bestie, christmas, haha, support, spa, new, weekend, dream, gangam_style
Sad	hospice, sleepy, drop, extreme, miss, unreal, stumble, chill, dose, old
Surprise	ufo, magic, rocket, planet, nasa, computer, discount, tarot, higher, believe

Table 6: Top 10 non-emotion dictionary words per emotion category sorted in decreasing order of their importance.

Sentence 6 (*FEAR*). *So afraid of the finals, a huge **#challenge** for me.*

Sentence 7 (*JOY*). *LoL, I **#challenge** you to do a striptease on GangnamStyle !!*

In summary, the effective meaning of a word is not simply propagated from the lexicon, but is inferred through its structural bindings. The above two sentences demonstrate that our model can capture this high level information.

Lexicon Expansion. One of the salient features of our model is that it can automatically learn the emotion score of the words not in the emotion lexicon. These words embed emotion information that can be utilized to expand the lexicon further. For our collection of tweets, only a small portion of words is in the emotional lexicon. Table 6 shows the top non-dictionary words that model associated with each emotion category.

For *anger*, we discover the word *Snowden*, which is the last name of Edward Snowden. Since he leaked numerous confidential documents to the press, the tweets regarding him exhibits *anger*. For emotion *joy*, the word *bestie* can not be found in any dictionary, but it is a popular slang used to refer to one's best friend. Another term *Gangnam_Style* refers to a popular music video which has funny dance steps.

6 Conclusion

In this paper, we proposed a constraint optimization framework for detecting emotion from text. There is a likely sweet spot between the different constraints and our model can automatically find it through an efficient parameter-tuning algorithm. Our model is linear in the input size and hence quite suitable for large datasets. One key aspect of our framework is that it can be easily configured to add new features as well as incorporate refined emotion lexicons. There has been plenty of prior work that could guide the model in improving its performance further. Another distinguishing feature of our model is that it solves multi-label classification problem and allows a document to have multiple emotions. It can also be used to pick one single most dominant emotion category for a document. In conjunction with assigning categories to documents, it also assigns categories to the words. Hence can be used for expansion of the emotion lexicon.

References

[Agrawal and An, 2012] Ameeta Agrawal and Aijun An. Unsupervised emotion detection from text using semantic and syntactic relations. In *WI-IAT*, 2012.

[Alm *et al.*, 2005] Cecilia Ovesdotter Alm, Dan Roth, and Richard Sproat. Emotions from text: Machine learning for text-based emotion prediction. In *HLT/EMNLP*, 2005.

[Alm, 2008] Ebba Cecilia Ovesdotter Alm. *Affect in text and speech*. ProQuest, 2008.

[Aman and Szpakowicz, 2007] Saima Aman and Stan Szpakowicz. Identifying expressions of emotion in text. In *Text, Speech and Dialogue*. Springer, 2007.

[Aman and Szpakowicz, 2008] Saima Aman and Stan Szpakowicz. Using roget's thesaurus for fine-grained emotion recognition. In *IJCNLP*, 2008.

[Bao *et al.*, 2009] Shenghua Bao, Shengliang Xu, Li Zhang, Rong Yan, Zhong Su, Dingyi Han, and Yong Yu. Joint emotion-topic modeling for social affective text mining. In *ICDM*, 2009.

[Bollen *et al.*, 2011] Johan Bollen, Huina Mao, and Alberto Pepe. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In *ICWSM*, 2011.

[Bradley and Lang, 1999] Margaret M Bradley and Peter J Lang. Affective norms for english words (anew): Instruction manual and affective ratings. Technical report, Citeseer, 1999.

[Chaumartin, 2007] François-Régis Chaumartin. UPAR7: A knowledge-based system for headline sentiment tagging. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, 2007.

[De Choudhury *et al.*, 2012a] Munmun De Choudhury, Scott Counts, and Michael Gamon. Not all moods are created equal! exploring human emotional states in social media. In *ICWSM*, 2012.

[De Choudhury *et al.*, 2012b] Munmun De Choudhury, Michael Gamon, and Scott Counts. Happy, nervous or surprised? classification of human affective states in social media. In *ICWSM*, 2012.

[Dempster *et al.*, 1977] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–38, 1977.

[Ekkekakis, 2013] Panteleimon Ekkekakis. *The Measurement of Affect, Mood, and Emotion: A Guide for Health-behavioral Research*. Cambridge University Press, 2013.

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

[Esuli and Sebastiani, 2006] Andrea Esuli and Fabrizio Sebastiani. Sentiwordnet: A publicly available lexical resource for opinion mining. In *LREC*, 2006.

[Golder and Macy, 2011] Scott A Golder and Michael W Macy. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science*, 2011.

[Kim *et al.*, 2010] Sunghwan Mac Kim, Alessandro Valitutti, and Rafael A Calvo. Evaluation of unsupervised emotion models to textual affect recognition. In *NAACL HLT Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, 2010.

[Kozareva *et al.*, 2007] Zornitsa Kozareva, Borja Navarro, Sonia Vázquez, and Andrés Montoyo. UA-ZBSA: a headline emotion classification through web information. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, 2007.

[Lee and Seung, 2001] Daniel D Lee and H Sebastian Seung. Algorithms for non-negative matrix factorization. In *NIPS*, 2001.

[Li *et al.*, 2009] Tao Li, Yi Zhang, and Vikas Sindhwani. A non-negative matrix tri-factorization approach to sentiment classification with lexical prior knowledge. In *ACL-IJCNLP*, 2009.

[Lin *et al.*, 2007] Kevin Hsin-Yih Lin, Changhua Yang, and Hsin-Hsi Chen. What emotions do news articles trigger in their readers? In *SIGIR*, 2007.

[Liu, 2012] Bing Liu. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 2012.

[Mihalcea and Liu, 2006] Rada Mihalcea and Hugo Liu. A corpus-based approach to finding happiness. In *AAAI*, 2006.

[Mishne and De Rijke, 2006] Gilad Mishne and Maarten De Rijke. Capturing global mood levels using blog posts. In *AAAI CAAAW*, 2006.

[Mishne, 2005] Gilad Mishne. Experiments with mood classification in blog posts. In *SIGIR Workshop on Stylistic Analysis Of Text For Information Access*, 2005.

[Mohammad and Turney, 2013] Saif M Mohammad and Peter D Turney. Nrc emotion lexicon. 2013.

[Pang and Lee, 2008] Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2):1–135, 2008.

[Pauca *et al.*, 2006] V Paul Pauca, Jon Piper, and Robert J Plemmons. Nonnegative matrix factorization for spectral data analysis. *Linear algebra and its applications*, 416(1):29–47, 2006.

[Pennebaker *et al.*, 2007] James W Pennebaker, RJ Booth, and ME Francis. Linguistic inquiry and word count: Liwc [computer software]. *Austin, TX: liwc. net*, 2007.

[Poria *et al.*, 2014] Soujanya Poria, Alexander Gelbukh, Erik Cambria, Amir Hussain, and Guang-Bin Huang. Emosenticspace: A novel framework for affective common-sense reasoning. *Knowledge-Based Systems*, 2014.

[Sadilek *et al.*, 2013] Adam Sadilek, Christopher Homan, Walter Lasecki, Vincent Silenzio, and Henry Kautz. Modeling fine-grained dynamics of mood at scale. 2013.

[Scherer and Wallbott, 1994] Klaus R Scherer and Harald G Wallbott. Evidence for universality and cultural variation of differential emotion response patterning. *Journal of personality and social psychology*, 66(2):310, 1994.

[Strapparava and Mihalcea, 2007] Carlo Strapparava and Rada Mihalcea. Semeval-2007 task 14: Affective text. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, 2007.

[Strapparava and Mihalcea, 2008] Carlo Strapparava and Rada Mihalcea. Learning to identify emotions in text. In *SAC*, 2008.

[Strapparava and Valitutti, 2004] Carlo Strapparava and Alessandro Valitutti. Wordnet affect: an affective extension of wordnet. In *LREC*, 2004.

[Yang *et al.*, 2007] Changhua Yang, K.H. Lin, and Hsin-Hsi Chen. Emotion classification using web blog corpora. In *WI*, 2007.