Do You Feel What I Feel? Social Aspects of Emotions in Twitter Conversations

Suin Kim
Department of Computer Science
KAIST, South Korea
suin.kim@kaist.ac.kr

JinYeong Bak
Department of Computer Science
KAIST, South Korea
jy.bak@kaist.ac.kr

Alice Oh
Department of Computer Science
KAIST, South Korea
alice.oh@kaist.edu

Abstract

We present a computational framework for understanding the social aspects of emotions in Twitter conversations. Using unannotated data and semisupervised machine learning, we look at emotional transitions, emotional influences among the conversation partners, and patterns in the overall emotional exchanges. We find that conversational partners usually express the same emotion, which we name *Emotion accommodation*, but when they do not, one of the conversational partners tends to respond with a positive emotion. We also show that tweets containing sympathy, apology, and complaint are significant emotion influencers. We verify the emotion classification part of our framework by a human-annotated corpus.

Introduction

Popular social network services (SNS), such as Twitter, have become good resources for researchers interested in studying social behaviors. The public nature of Twitter makes it appropriate for various behavioral studies, and it has become especially useful for studying sentiments and emotions (Bollen, Pepe, and Mao 2011; Diakopoulos and Shamma 2010). However, these studies do not look at the inherently social (Parrott 2001) nature of sentiment and emotions. They often arise in social situations, and when expressed in a conversation, the emotions expressed affect the interactions among the participants. The goal of this paper is to study how Twitter conversations can be used for understanding the social aspects of emotions.

In much of existing literature (see (Pang and Lee 2008) for a review), sentiment analysis is usually a simple classification of positive and negative (and sometimes neutral). Emotion analysis has not been explored as much, but (Kamvar and Harris 2011) show that emotions are much more diverse, including for example, *anger*, *surprise*, and *joy*. Both the simple classification of sentiments and the diverse set of emotions are important aspects of conversations.

Twitter is widely used for conversations (Ritter, Cherry, and Dolan 2010), and prior work has studied different aspects of human conversations by using Twitter as a source (Boyd, Golder, and Lotan 2010; Danescu-Niculescu-Mizil, Gamon, and Dumais 2011; Honey and Herring 2009) . Ours is the first paper to look at emotions in Twitter conversations, and we present the computational methods and results for analyses of 1) how different emotions in a tweet lead to the emotions in a tweet responding to it, 2) how certain words or emotions by one interlocutor triggers emotion changes in the conversational partner, and 3) how the disagreement in overall emotion of the conversational partners can reveal interesting conversational topics.

In this paper, we explore in-depth questions of the emotional patterns in conversational interaction. We hypothesize

- that there is a meaningful pattern of emotions in a conversation, and those patterns depend on the topics and words of the conversation,
- that conversational partners can influence each others' emotions and topics,
- that there are patterns in the overall emotions of each conversation, and
- that semantic usage patterns affect the overall emotions of tweets user send and receive.

To find these patterns, we develop a framework to discover the topics and emotions from an unannotated corpus of Twitter conversations. Compared to previous work (Bollen, Mao, and Zeng 2011), our framework is flexible over noisy and highly skewed corpus.

Data Collection and Emotion Detection

In this section, we present an automated computational framework for finding emotions in Twitter conversations. Finding emotions from Twitter data is challenging because Twitter data is noisy and emotions show a highly skewed distribution. For example, in our corpus, 2.6% of emotional tweets express *fear* and 39.7% express *joy*.

Conversational Data Collection We call a Twitter conversation a *chain* and define it as a sequence of replies between two users using the Twitter *reply* button. We identified English Twitter users who *replied* to other people's tweets, and we considered those users as *candidates*. Then, we collected all of the candidates' tweets from May, 2011 to August, 2011. Using those tweets, we identified more candidates by looking at the target users of the *replies* within those tweets. In this way, we identified 136,730 users, 222,024

Anticipation 29	Joy 70	Anger 21	Surprise 14	Fear 5	Sadness 17	Disgust 7	Acceptance 18	Neutral 19
Topic 125	Topic 114	Topic 19	Topic 172	Topic 48	Topic 6	Topic 116	Topic 43	Topic 180
hope	omg	lmao	haha	omg	oh	oh	ok	com
better	love	shit	yeah	oh	sorry	fuck	oh	www
feel	haha	damn	know	know	haha	don't	thank	http
thank	thank	fuck	think	haha	know	ye	cool	check
Topic 26	Topic 107	Topic 4	Topic 89	Topic 27	Topic 59	Topic 22	Topic 102	Topic 184
good	love	ass	know	room	hurt	don't	'll	account
thank	thank	yo	don't	house	got	oh	know	google
luck	follow	lmao	think	don't	bad	think	try	app
haha	wow	nigga	look	come	pain	yeah	let	work

Figure 1: Examples of emotional topics with highest $Corr$ score. Digits next to the heading emotions indicate the number of topics found for each emotion.

Anticipation	hope wait inspire excited bored ready
Joy	awesome amazing wonder excited glad
Anger	shit bitch ass mean damn mad jealous
Surprise	amazing wow wonder weird lucky differ
Fear	scared stress horror nervous terror alarm
Sadness	sorry bad aww sad wrong hurt blue dead
Disgust	sick wrong evil fat ugly horrible gross
Acceptance	okay ok same alright safe relax peace

Table 1: The 8 emotions and the set of expanded seed words.

dyads, and 1,668,308 chains. For running the experiments, we filtered the data by keeping only the chains of four tweets or more. This resulted in 153,054 chains containing 871,544 tweets, 5.69 tweets per chain on average.

Topic Discovery from Unannotated Corpus We use sentence-LDA (Jo and Oh 2011) to discover emotions in Twitter conversational data. We treat each chain as a document composed of a number of sentences, each of which is a single *reply*, or a tweet, in the chain. We set the number of topics as 200 to discover fine-grained topics.

Building Emotional Lexicons We classify topics in the basic 8 emotions from the Plutchik's wheel of emotions (Plutchik 1980): *Anticipation, Joy, Anger, Surprise, Fear, Sadness, Disgust*, and *Acceptance*. We develop a list of emotion *seed words* starting with the highest frequency emotion words from (Kamvar and Harris 2011). We refine and extend the set of words based on the PMI (pointwise mutual information) score. This process expands the emotion lexicon in a flexible way that accommodates the spelling variations and unique lexical patterns of Twitter. Table 1 shows a part of the expanded list of seed words.

Detecting Emotion from Discovered Topics After identifying the seed words, we classify topics into the basic 8 emotions plus a neutral category. We calculate the strength of emotions for each topic based on the PMI score for each emotion given a topic. Let $\mathcal{L}_e$ and N_e be the set of seed words and the number of seed words for emotion e. Then we define the $Corr$ score, correlation metric between an emotion and a topic, for the emotion e and a topic k as

$$Corr(e,k) = \gamma_e \text{expPMI}(e;k) = \gamma_e \frac{P(e|k)}{P(e)}, \text{ where}$$

$$P(e|k) \propto \frac{1}{N_e} \sum_{w \in \mathcal{L}_e} P(w|k) \text{ and } P(e) \propto \sum_k^K P(e|k).$$

γ_e is the lexicon-dependent constant for term frequency normalization. For the every possible combination of $e \in E$ and $k \in K$, we calculate $Corr(e,k)$. Looking from the highest $Corr$ score within a pairs of e and k, if topic k is not tagged as any emotion yet, tag k as emotion e. If $Corr$ score becomes lower than the bound $\mathcal{I}$, we stop tagging topics and tag the remaining topics as neutral.

In this way, we identify 181 emotional and 19 neutral topics from 200 topics. Examples of the discovered topics are described in Figure 1. The figure shows that the numbers of topics tagged by the 8 emotions are not uniformly distributed, indicating the corpus is skewed toward the emotions of *Joy, Acceptance*, and *Anticipation*. Neutral topics consist of the words related to IT and social network.

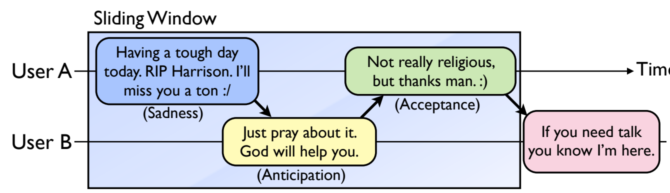


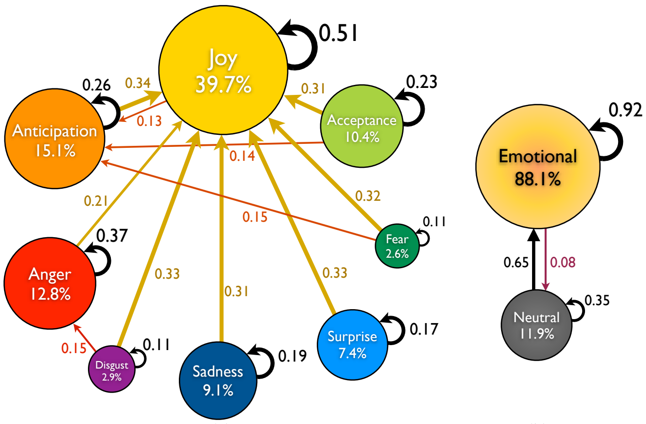
Figure 2: The sliding window method for finding influence. Second tweet (yellow) *influences* A's emotion from *Sadness* to *Acceptance*. We can also see an emotion *transition* from *Sadness* to *Anticipation*.

Social Aspects of Emotions

We describe the method and the results of discovering patterns in social aspects of emotions. We define two ways of emotional effect, emotion transition and emotion influence by using a sliding window in Figure 2.

Emotion Transitions

We find what emotion a user is likely to express after receiving a message containing a certain emotion. We call this emotional pattern as an emotion *transition*. Figure 3 shows the emotions spread over Twitter conversation data and the transition probabilities among the emotions. The transition probability from emotion e_a to e_b indicates the probability that a Twitter user writes a tweet containing emotion e_b in a reply to a tweet expressing e_a. We discovered that a twitter user is likely to express a positive emotion regardless of the emotion in the previous tweet. The next highest transitions are the self-transitions, indicating that twitter



(a) (b)

Figure 3: (a) Emotion transitions and the distribution of emotions over emotional tweets. Percentage indicates the ratio of each emotion over the tweets expressing emotion. Numbers next to the arrow show the transition probabilities among the emotions. (b) Transitions between emotional and neutral tweets.

users share common feelings with their conversational partners. We call this phenomenon an *emotion accommodation*. Also, the average transition probabilities across the opposing emotion pairs of Plutchik's wheel of emotions are quite low (0.101), compared to the average transition probabilities for all other pairs of emotions (0.141). Figure 3 (b) also shows that a twitter user tends to respond with an emotional tweet in response to a neutral tweet.

Emotion Influences

In a conversation, an interlocutor may alter the emotion of the conversational partner. We hypothesize that this is done mainly through the topics and the emotions of the utterances.

Emotion as an Emotion Influencer Using the sliding window method with size 3, we look for the emotions that have the highest success rate to change the partner's emotion from emotion e_a to e_b. We compute that success rate by

$$P(e_a \underset{e_k}{\longrightarrow} e_b) = \frac{N(e_a \underset{e_k}{\longrightarrow} e_b)}{\sum_{b' \in E} N(e_a \underset{e_k}{\longrightarrow} e_{b'})},$$

where $N(e_a \underset{e_k}{\longrightarrow} e_b)$ indicates the number of emotional influences from e_a to e_b by tweet expressing emotion e_k in the conversational corpus. We discovered that $P(e_a \underset{e_b}{\longrightarrow} e_b)$, representing that the influencing emotion is mirrored by the partner, is the highest regardless of e_a, without any exception. This result strongly implies that twitter users tend to show *Emotional accommodation* in a conversation.

Topic as an Emotion Influencer The subject of an utterance, topic, is the significant cause of change in emotion in human conversation. We are interested in what topics effectively influence the conversational partner's emotion from e_a to e_b. We show the topics having the highest probability of emotion influence for some pair of emotions in Figure 4. We

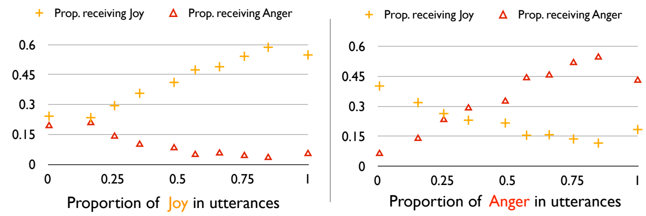


Figure 5: Proportion of receiving tweets in Joy and Anger, in respect to proportion of Joy and Anger in utterances.

discovered that topics that influence the conversational partner's emotion from a negative emotion to a positive emotion include *greeting*, *sympathy* and *recommendation*. Also, topics effectively influence conversational partner's sentiment from positive emotion to negative emotion include *worry*, *teasing*, and *complaint*. We also revalidate the prior finding that tweets expressing a certain emotion tend to cause the conversational partner to mirror the emotion.

Emotion Patterns over Conversations

We propose a way of finding interesting conversations by looking at the overall emotion patterns of the interlocutors. Our data shows that in most conversations, the interlocutors will share a common emotion. It would be interesting when one interlocutor shows a strong overall emotion, while the other shows a strong opposite overall emotion. We first look for conversations with the interlocutors expressing constant emotions, and check if they have the opposite emotions. In our data, about 4% of the chains show that pattern, and the topics in that pattern include *complaining*, *sympathy*, and *apology*. These chains show that one interlocutor is feeling upset about something, and the partner shows sympathy, or one is complaining to the partner, and that partner is making an apology.

A Case Study: Semantic Patterns In this study, we verify the traditional proverb-"One ill word asks another" or "Nice words for nice words" by statistical method. Will this adage still hold on Twitter? We explore the emotional usage and receiving patterns of each interlocutor. We concentrate on two strong emotions: *Joy* and *Anger*. Figure 5 shows that as proportion of Joy in total tweet increases, the proportion of Joy in receiving tweet increases and Anger decreases, and vice versa.

Evaluation of the Framework

We measure the performance of our framework, comparing with the traditional *word counting* method and random baseline. Word counting method classifies the instance if it contains a word that is known to be of a certain category a priori. We set the ground truth of emotional classification of tweets by using human annotation with *Amazon Mechanical Turk*. For randomly selected 1,000 tweets in the Twitter conversations, we ask 5 different workers for each tweet to select one of the 8 basic emotions, or *Neutral* if they think the given tweet is purely informational.

From 5 responses for each of 1,000 tweets, we measure the agreement as the number of workers who submitted the

Anger → Anticipation	Disgust → Joy	Sadness → Joy	Acceptance → Anger	Anticipation → Surprise	Joy → Acceptance	Anger → Fear	Anger → Acceptance	Fear → Joy	Joy → Sadness
Topic 57	**Topic 188**	**Topic 101**	**Topic 31**	**Topic 56**	**Topic 87**	**Topic 130**	**Topic 8**	**Topic 25**	**Topic 77**
morning	smile	watch	i'm	twitter	haha	sleep	night	hair	game
good	look	love	got	app	want	night	love	look	win
day	kiss	episode	lmao	new	time	hour	good	blond	play
thank	laugh	season	shit	link	okay	bed	sleep	really	team
Topic 26	**Topic 61**	**Topic 99**	**Topic 13**	**Topic 96**	**Topic 8**	**Topic 140**	**Topic 123**	**Topic 80**	**Topic 117**
good	watch	happy	lmao	music	night	time	dm	look	tweet
thank	new	birthday	shit	listen	love	got	check	thank	people
luck	live	day	nigga	play	good	work	got	forward	don't
haha	tv	thank	smh	song	sleep	think	number	great	read

Figure 4: List of topics with high influences on conversational partner's emotions.

same response. We have discovered that the agreement of submitted responses are quite low (shown in table 2). Every workers submitted all the same response for only 3% of the test tweets. We measure the performance of our framework using the tweets that satisfy at least 3 out of 5 agreement.

Agreement	2/5	3/5	4/5	5/5
Ratio	95.1%	50.8%	17.3%	3.02%

Table 2: Ratio of agreements from the responses from 5 distinct workers for each tweet.

Our framework correctly classified 31.3% of test tweets, which is about twice more accurate than the naïve word counting method, and about 3 times more accurate than random baseline, which scored 15.7% and 11.1% of accuracy, respectively. From the test results, the expected accuracy of emotion classification by a human is 54.4%, which implies that emotion classification task is a difficult task even for a human. We conclude this accuracy is a acceptable improvement over the baselines.

Conclusion

We have presented a novel computational framework for analyzing the social aspects of emotions in Twitter conversations. First, in emotion transitions, we found that the answer to *do you feel what I feel?* is a *yes*, in that self-transitions account for the large proportions. However, there are significant transitions to other positive sentiments and emotions, implying Twitter users tend to feel good even when the conversational partners do not. We discovered a significant pattern that expressing a desired emotion is the best strategy to alter conversational partner's emotion. We also found that *greeting*, *sympathy*, *worry*, and *complaining* play significant roles in emotion influences. Using our framework, a traditional proverb, "One ill word asks another.", is proved effective in Twitter conversations. Finally, we showed that examining sentiment patterns in conversations leads us to discover interesting conversations. To verify our findings, we measured the accuracy of our framework, which is reasonable. There are many future promising directions stemming from this work including temporal patterns, profiling patterns, and comparisons with other communication platforms.

References

Bollen, J.; Mao, H.; and Zeng, X. 2011. Twitter mood predicts the stock market. *Journal of Computational Science*.

Bollen, J.; Pepe, A.; and Mao, H. 2011. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*.

Boyd, D.; Golder, S.; and Lotan, G. 2010. Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In *Proceedings of the 2010 43rd Hawaii International Conference on System Sciences*, 1–10. IEEE Computer Society.

Danescu-Niculescu-Mizil, C.; Gamon, M.; and Dumais, S. 2011. Mark my words!: linguistic style accommodation in social media. In *Proceedings of the 20th international conference on World wide web*, 745–754. ACM.

Diakopoulos, N. A., and Shamma, D. A. 2010. Characterizing debate performance via aggregated twitter sentiment. In *Proceedings of the 28th international conference on Human factors in computing systems*, CHI '10, 1195–1198. New York, NY, USA: ACM.

Honey, C., and Herring, S. 2009. Beyond microblogging: Conversation and collaboration via twitter. In *System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference on*, 1–10. IEEE.

Jo, Y., and Oh, A. 2011. Aspect and sentiment unification model for online review analysis. In *Proceedings of the fourth ACM international conference on Web search and data mining*, 815–824. ACM.

Kamvar, S., and Harris, J. 2011. We feel fine and searching the emotional web. In *Proceedings of the fourth ACM international conference on Web search and data mining*, 117–126. ACM.

Pang, B., and Lee, L. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval* 2(1-2):1–135.

Parrott, W. 2001. *Emotions in social psychology: essential readings*. Psychology Pr.

Plutchik, R. 1980. *Emotion: A psychoevolutionary synthesis*. Harper & Row New York.

Ritter, A.; Cherry, C.; and Dolan, B. 2010. Unsupervised modeling of twitter conversations. In *Proceedings of HLT-NAACL*.