Self-Disclosure and Relationship Strength in Twitter Conversations

JinYeong Bak, Suin Kim, Alice Oh
Department of Computer Science
Korea Advanced Institute of Science and Technology
Daejeon, South Korea
{jy.bak, suin.kim}@kaist.ac.kr, alice.oh@kaist.edu

Abstract

In social psychology, it is generally accepted that one discloses more of his/her personal information to someone in a strong relationship. We present a computational framework for automatically analyzing such self-disclosure behavior in Twitter conversations. Our framework uses text mining techniques to discover topics, emotions, sentiments, lexical patterns, as well as personally identifiable information (PII) and personally embarrassing information (PEI). Our preliminary results illustrate that in relationships with high relationship strength, Twitter users show significantly more frequent behaviors of self-disclosure.

1 Introduction

We often *self-disclose*, that is, share our emotions, personal information, and secrets, with our friends, family, coworkers, and even strangers. Social psychologists say that the degree of self-disclosure in a relationship depends on the strength of the relationship, and strategic self-disclosure can strengthen the relationship (Duck, 2007). In this paper, we study whether relationship strength has the same effect on self-disclosure of Twitter users.

To do this, we first present a method for computational analysis of self-disclosure in online conversations and show promising results. To accommodate the largely unannotated nature of online conversation data, we take a topic-model based approach (Blei et al., 2003) for discovering latent patterns that reveal self-disclosure. A similar approach was able to discover sentiments (Jo and Oh, 2011) and emotions (Kim et al., 2012) from user contents. Prior

work on self-disclosure for online social networks has been from communications research (Jiang et al., 2011; Humphreys et al., 2010) which relies on human judgements for analyzing self-disclosure. The limitation of such research is that the data is small, so our approach of automatic analysis of self-disclosure will be able to show robust results over a much larger data set.

Analyzing relationship strength in online social networks has been done for Facebook and Twitter in (Gilbert and Karahalios, 2009; Gilbert, 2012) and for enterprise SNS (Wu et al., 2010). In this paper, we estimate relationship strength simply based on the duration and frequency of interaction. We then look at the correlation between self-disclosure and relationship strength and present the preliminary results that show a positive and significant correlation.

2 Data and Methodology

Twitter is widely used for conversations (Ritter et al., 2010), and prior work has looked at Twitter for different aspects of conversations (Boyd et al., 2010; Danescu-Niculescu-Mizil et al., 2011; Ritter et al., 2011). Ours is the first paper to analyze the degree of self-disclosure in conversational tweets. In this section, we describe the details of our Twitter conversation data and our methodology for analyzing relationship strength and self-disclosure.

2.1 Twitter Conversation Data

A Twitter conversation is a *chain* of tweets where two users are consecutively replying to each other's tweets using the Twitter *reply* button. We identified dyads of English-tweeting users who had at least

three conversations from October, 2011 to December, 2011 and collected their tweets for that duration. To protect users' privacy, we anonymized the data to remove all identifying information. This dataset consists of 131,633 users, 2,283,821 chains and 11,196,397 tweets.

2.2 Relationship Strength

Research in social psychology shows that relationship strength is characterized by interaction frequency and closeness of a relationship between two people (Granovetter, 1973; Levin and Cross, 2004). Hence, we suggest measuring the relationship strength of the conversational dyads via the following two metrics. **Chain frequency** (CF) measures the number of conversational chains between the dyad averaged per month. **Chain length** (CL) measures the length of conversational chains between the dyad averaged per month. Intuitively, high CF or CL for a dyad means the relationship is strong.

2.3 Self-Disclosure

Social psychology literature asserts that self-disclosure consists of personal information and open communication composed of the following five elements (Montgomery, 1982).

Negative openness is how much disagreement or negative feeling one expresses about a situation or the communicative partner. In Twitter conversations, we analyze sentiment using the aspect and sentiment unification model (ASUM) (Jo and Oh, 2011), based on LDA (Blei et al., 2003). ASUM uses a set of seed words for an unsupervised discovery of sentiments. We use positive and negative emoticons from Wikipedia.org[1]. **Nonverbal openness** includes facial expressions, vocal tone, bodily postures or movements. Since tweets do not show these, we look at emoticons, 'lol' (laughing out loud) and 'xxx' (kisses) for these nonverbal elements. According to Derks et al. (2007), emoticons are used as substitutes for facial expressions or vocal tones in socio-emotional contexts. We also consider profanity as nonverbal openness. The methodology used for identifying profanity is described in the next section. **Emotional openness** is how much one discloses his/her feelings and moods. To measure this,

we look for tweets that contain words that are identified as the most common expressions of feelings in blogs as found in Harris and Kamvar (2009). **Receptive openness** and **General-style openness** are difficult to get from tweets, and they are not defined precisely in the literature, so we do not consider these here.

2.4 PII, PEI, and Profanity

PII and PEI are also important elements of self-disclosure. Automatically identifying these is quite difficult, but there are certain topics that are indicative of PII and PEI, such as *family*, *money*, *sickness* and *location*, so we can use a widely-used topic model, LDA (Blei et al., 2003) to discover topics and annotate them using MTurk[2] for PII and PEI, and profanity. We asked the Turkers to read the conversation chains representing the topics discovered by LDA and have them mark the conversations that contain PII and PEI. From this annotation, we identified five topics for profanity, ten topics for PII, and eight topics for PEI. Fleiss kappa of MTurk result is 0.07 for PEI, and 0.10 for PII, and those numbers signify *slight agreement* (Landis and Koch, 1977). Table 1 shows some of the PII and PEI topics. The profanity words identified this way include *nigga, lmao, shit, fuck, lmfao, ass, bitch*.

PII 1	PII 2	PEI 1	PEI 2	PEI 3
san	tonight	pants	teeth	family
live	time	wear	doctor	brother
state	tomorrow	boobs	dr	sister
texas	good	naked	dentist	uncle
south	ill	wearing	tooth	cousin

Table 1: PII and PEI topics represented by the high-ranked words in each topic.

To verify the topic-model based approach to discovering PII and PEI, we tried supervised classification using SVM on document-topic proportions. Precision and recall are 0.23 and 0.21 for PII, and 0.30 and 0.23 for PEI. These results are not quite good, but this is a difficult task even for humans, and we had a low agreement among the Turkers. So our current work is in improving this.

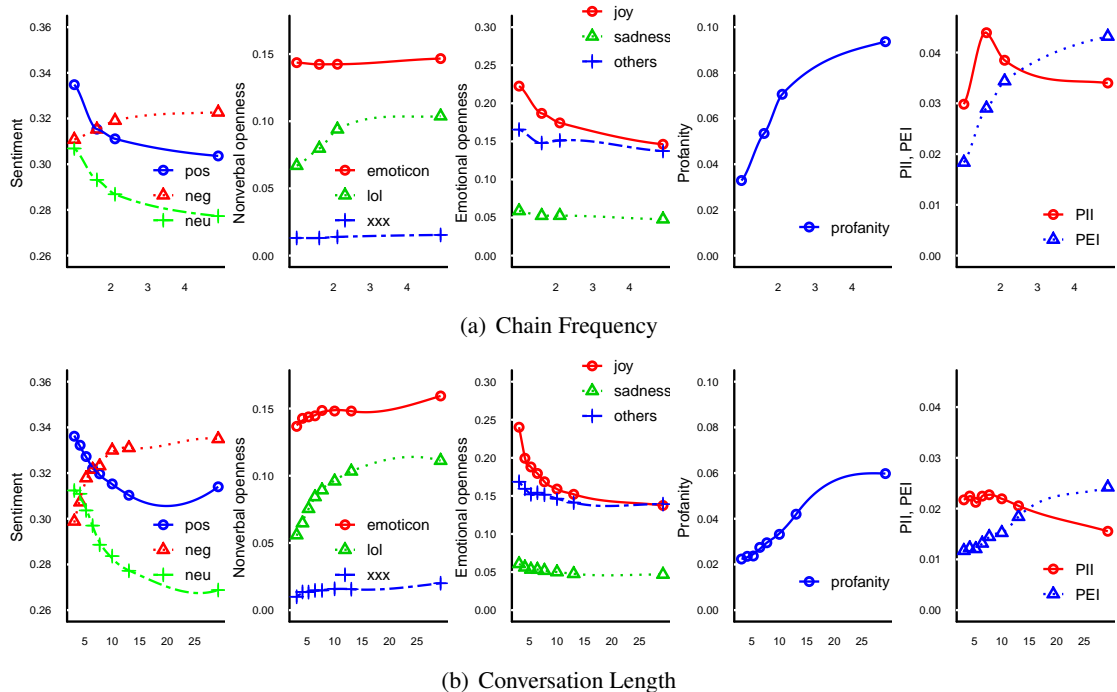---

[1] http://en.wikipedia.org/wiki/List of emoticons

[2] https://www.mturk.com

(a) Chain Frequency

(b) Conversation Length

Figure 1: Degree of self-disclosure depending on various relationship strength metrics. The x axis shows relationship strength according to tweeting behavior (chain frequency and chain length), and the y axis shows proportion of self-disclosure in terms of negative openness, emotional openness, profanity, and PII and PEI.

3 Results and Discussions

Chain frequency (CF) and chain length (CL) reflect the dyad's tweeting behaviors. In figure 1, we can see that the two metrics show similar patterns of self-disclosure. When two users have stronger relationships, they show more negative openness, nonverbal openness, profanity, and PEI. These patterns are expected. However, weaker relationships tend to show more PII and emotions. A closer look at the data reveals that PII topics are related to cities where they live, time of day, and birthday. This shows that the weaker relationships, usually new acquaintances, use PII to introduce themselves or send trivial greetings for birthdays. Higher emotional openness in weaker relationships looks strange at first, but similar to PII, emotion in weak relationships is usually expressed as greetings, reactions to baby or pet photos, or other shallow expressions.

It is interesting to look at outliers, dyads with very strong and very weak relationship groups. Table 3 summarizes the self-disclosure behaviors of these outliers. There is a clear pattern that stronger relationships show more nonverbal openness, nega-

str1	str2	weak1	weak2	weak3
lmao	sleep	following	ill	love
lmfao	bed	thanks	sure	thanks
shit	night	followers	soon	cute
ass	tired	welcome	better	aww
smh	awake	follow	want	pretty

Table 2: Topics that are most prominent in strong ('str') and weak relationships.

tive openness, profanity use, and PEI. In figure 1, emotional openness does not differ for the strong and weak relationship groups. We can see why this is when we look at the topics for the strong and weak groups. Table 2 shows the topics that are most prominent in the strong relationships, and they include daily greetings, plans, nonverbal emotions such as 'lol', 'omg', and profanity. In weak relationships, the prominent topics illustrate the prevalence of initial getting-to-know conversations in Twitter. They welcome and greet each other about kids and pets, and offer sympathies about feeling bad.

One interesting way to use our analysis is in iden-

	strong	weak
# relation	5,640	226,116
CF	14.56	1.00
CL	97.74	3.00
Emotion	0.21	0.22
Emoticon	0.162	0.134
lol	0.105	0.060
xxx	0.021	0.006
Pos Sent	0.31	0.33
Neg Sent	0.32	0.29
Neut Sent	0.27	0.29
Profanity	0.0615	0.0085
PII	0.016	0.019
PEI	0.022	0.013

Table 3: Comparing the top 1% and the bottom 1% relationships as measured by the combination of CF and CL. From 'Emotion' to PEI, all values are average proportions of tweets containing each self-disclosure behavior. Strong relationships show more negative sentiment, profanity, and PEI, and weak relationships show more positive sentiment and PII. 'Emotion' is the sum of all emotion categories and shows little difference.

tifying a rare situation that deviates from the general pattern, such as a dyad linked weakly but shows high self-disclosure. We find several such examples, most of which are benign, but some do show signs of risk for one of the parties. In figure 2, we show an example of a conversation with a high degree of self-disclosure by a dyad who shares only one conversation in our dataset spanning two months.

4 Conclusion and Future Work

We looked at the relationship strength in Twitter conversational partners and how much they self-disclose to each other. We found that people disclose more to closer friends, confirming the social psychology studies, but people show more positive sentiment to weak relationships rather than strong relationships. This reflects the social norm toward first-time acquaintances on Twitter. Also, emotional openness does not change significantly with relationship strength. We think this may be due to the inherent difficulty in truly identifying the emotions on Twitter. Identifying emotion merely based on keywords captures mostly shallow emotions, and deeper emotional openness either does not occur much on
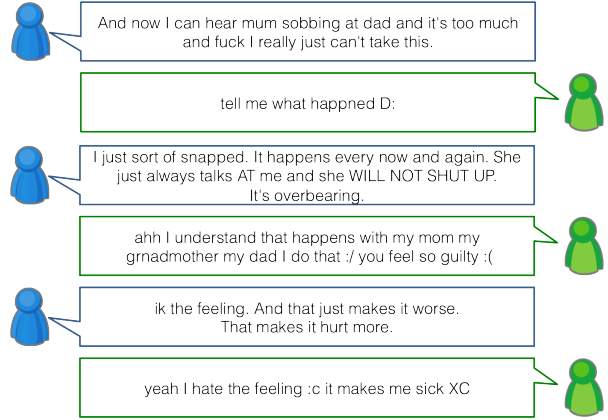

Figure 2: Example of Twitter conversation in a weak relationship that shows a high degree of self-disclosure.

Twitter or cannot be captures very well.

With our automatic analysis, we showed that when Twitter users have conversations, they control self-disclosure depending on the relationship strength. We showed the results of measuring the relationship strength of a Twitter conversational dyad with chain frequency and length. We also showed the results of automatically analyzing self-disclosure behaviors using topic modeling.

This is ongoing work, and we are looking to improve methods for analyzing relationship strength and self-disclosure, especially emotions, PII and PEI. For relationship strength, we will consider not only interaction frequency, but also network distance and relationship duration. For finding emotions, first we will adapt existing models (Vaassen and Daelemans, 2011; Tokuhisa et al., 2008) and suggest a new semi-supervised model. For finding PII and PEI, we will not only consider the topics, but also time, place and the structure of questions and answers. This paper is a starting point that has shown some promising research directions for an important problem.

5 Acknowledgment

References

D.M. Blei, A.Y. Ng, and M.I. Jordan. 2003. Latent dirichlet allocation. *The Journal of Machine Learning Research*, 3:993–1022.

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

C. Danescu-Niculescu-Mizil, M. Gamon, and S. Dumais. 2011. Mark my words!: linguistic style accommodation in social media. In *Proceedings of the 20th International World Wide Web Conference*.

D. Derks, A.E.R. Bos, and J. Grumbkow. 2007. Emoticons and social interaction on the internet: the importance of social context. *Computers in Human Behavior*, 23(1):842–849.

S. Duck. 2007. *Human Relationships*. Sage Publications Ltd.

E. Gilbert and K. Karahalios. 2009. Predicting tie strength with social media. In *Proceedings of the 27th International Conference on Human Factors in Computing Systems*, pages 211–220.

E. Gilbert. 2012. Predicting tie strength in a new medium. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*.

M.S. Granovetter. 1973. The strength of weak ties. *American Journal of Sociology*, pages 1360–1380.

J. Harris and S. Kamvar. 2009. *We Feel Fine: An Almanac of Human Emotion*. Scribner Book Company.

L. Humphreys, P. Gill, and B. Krishnamurthy. 2010. How much is too much? privacy issues on twitter. In *Conference of International Communication Association, Singapore*.

L. Jiang, N.N. Bazarova, and J.T. Hancock. 2011. From perception to behavior: Disclosure reciprocity and the intensification of intimacy in computer-mediated communication. *Communication Research*.

Y. Jo and A.H. Oh. 2011. Aspect and sentiment unification model for online review analysis. In *Proceedings of International Conference on Web Search and Data Mining*.

S. Kim, J. Bak, and A. Oh. 2012. Do you feel what i feel? social aspects of emotions in twitter conversations. In *Proceedings of the AAAI International Conference on Weblogs and Social Media*.

J.R. Landis and G.G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics*, pages 159–174.

D.Z. Levin and R. Cross. 2004. The strength of weak ties you can trust: The mediating role of trust in effective knowledge transfer. *Management science*, pages 1477–1490.

B.M. Montgomery. 1982. Verbal immediacy as a behavioral indicator of open communication content. *Communication Quarterly*, 30(1):28–34.

A. Ritter, C. Cherry, and B. Dolan. 2010. Unsupervised modeling of twitter conversations. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 172–180.

A. Ritter, C. Cherry, and W.B. Dolan. 2011. Data-driven response generation in social media. In *Proceedings of EMNLP*.

R. Tokuhisa, K. Inui, and Y. Matsumoto. 2008. Emotion classification using massive examples extracted from the web. In *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*, pages 881–888.

F. Vaassen and W. Daelemans. 2011. Automatic emotion classification for interpersonal communication. *ACL HLT 2011*, page 104.

A. Wu, J.M. DiMicco, and D.R. Millen. 2010. Detecting professional versus personal closeness using an enterprise social network site. In *Proceedings of the 28th International Conference on Human Factors in Computing Systems*.