Toward Macro-Insights for Suicide Prevention: Anaylzing Fine-Grain Distress at Scale

Abstract

This paper is an initial exploratory research

1 Introduction

Suicide is among the leading causes of death for individuals 10–44 years of age in the United States (Heron and Tejada-Vera, 2009). Indeed, while mortality rates for most illnesses decreased between 2008 and 2009, the rate of suicide increased by 2.4% (Heron and Tejada-Vera, 2009). The lifetime prevalence for suicidal ideation is between 5.6 and 14.3 percent in the general population, and as high as 19.8–24.0% among youth (Nock et al., 2008).

The first step toward suicide *prevention* is to identify, ideally in consultation with clinical experts, the risk factors associated with suicide. Due to social stigma and the impact on self-judgement and self-awareness caused by such risk factors as depression or drug use, individuals with suicidal ideation may not always reach out to professionals or, if they do, provide them with accurate information. They may not even realize their own level of suicide risk before it is too late. Self-reporting, then, is not a reliable means of detecting and assessing suicide risk.

Individuals may be more inclined to seek support from informal resources, such as social media, instead of seeking treatment (Crosby et al., 2011; Bruffaerts et al., 2011; Ryan et al., 2010). Evidence suggests that youth and emerging adults usually prefer to seek help from their friends and families; however, higher levels of suicidal ideation are associated with lower levels of help-seeking from both formal or informal resources (Deane et al., 2001).

These trends in help-seeking behavior suggests that social media might be a rich outlet for learning about support seeking. Internet- and telecommunications-driven activity is revolutionizing the social sciences by providing data—much of it publicly available—on human activity in situ, at volumes and a level of time and space granularity never before approached. Can such data improve clinical preventative measures by providing access to at-risk individuals who would otherwise go undetected, and by leading to better science about suicide risk behaviors?

Mann et al. (1999) developed the stress diathesis model for suicidal behavior, using many of the aforementioned risk factors. This model suggests (1) that objective states, such as depression or life events, as well as subjective states and traits, such as family history of depression, suicide, or substance abuse, were among the risk factors that contributed to suicidal ideation and (2) that the presence of these factors could eventually lead to either externalizing (e.g., interpersonal violence) or internalizing aggression (e.g., attempting suicide).

Since the stress-diathesis model was developed using risk factors for suicidal behavior and because it makes a connection between internalized and externalized acts, it is a suitable framework to analyze publicly available linguistic data from social media outlets such as Twitter. Data from social media can be used as a natural experiment to examine depression and suicidal ideation without being constrained by such sample biases as individuals who are willing take part in research and/or seek out formal sources of support. Moreover, this natural experiment method may provide information about individuals who are unlikely to engage in formal help-seeking behaviors and eventually could be used to identify effective methods of natural helping. Hence, this universal approach to screening for suicidal behaviors may have future implications not only for identifying individuals who have a higher prevalence for suicidal behaviors but it could eventually lead to the methods for enhancing protective factors against

suicide.

In this paper, we take steps toward the automatic detection of suicide risk among individuals via social media. We use various lexicon-based methods to retreive microblog posts (tweets) from Twitter and compare the performance of human annotators—some of whom are experts, and some of whom are not—to rate the level of distress of each tweet. Distress is an important risk factor in suicide that is observeable from microblog text, though admittedly observing suicide risk behavior is a highly subjective and noisy venture. Expert annotation, rather than general-purpose tools for content and sentiment analysis such as LIWC (Linguistic Inquiry and Word Count), provides a basis for language-based statistical modeling. We show that keyword based retrieval of training data results in better interannotator aggreement between non-experts and between experts and non-experts when the keywords are tuned toward the task at hand.

We additionally discover social and geographic patterns related to mood that microblogging sites such as Twitter reveal. Social support theory suggests that suicide and related mental health problems are strongly affected by one's physical and social environment (Wellman and Wortley, 1990). We show that otherwise weak correlations in the use of affective language between friends become very strong when one considers only friendships that are strongly *embedded* (informally speaking, in a relatively dense region of a social network) in the social network.

2 Related Work

Data on suicide traditionally comes from healthcare organizations, large-scale studies, or self reporting (Crosby et al., 2011; Horowitz and Ballard, 2009). These sources are limited by sociocultural barriers, such as stigma and shame, among other reasons (Crosby et al., 2011). Moreover, suicide is a fundamentally subjective, complex phenomenon with a low base rate. For these reasons, data on suicide is never particularly reliable and many researchers tend to focus on the relationship between risk factors and suicidal behavior, without relying heavily on theoretical models (Nock et al., 2008).

Approximately one-third of all individuals who reported suicidal ideation in their lifetime made a plan to commit suicide. Nearly three-quarters of those who reported making a suicide plan actually attempted suicide (Kessler et al., 1999). According to Kessler, Borges, and Walters (Kessler et al., 1999), the odds of attempting suicide increased exponentially when individuals endorsed three of more risk factors (e.g., having a mood or substance abused disorder).

Other established risk factors include demographics, previous suicide attempts, mental health concerns (i.e., depression, substance abuse, suicidal ideation, self-harm, and impulsivity), family history of suicide, interpersonal conflicts (i.e., family violence and bullying), mechanism or means for suicidal behavior (e.g., firearms) are commonly cited risk factors for suicidal behavior (Nock et al., 2008; Crosby et al., 2011; Gaynes et al., 2004; Harriss and Hawton, 2005; Shaffer et al., 2004; Shaffer et al., 2004; Brown et al., 2000).

Evidence suggests that when it comes to judgments that involve clinical phenomena, experts and novices behave differently. For example, in a medical image inspection task, Li et al. (2012) identified differences in perceptual expertise patterns between novices (students) and clinically trained physicians. Similarly, Womack et al. (2012) identified differences linguistic behaviors between experienced, attending dermatologists vs. resident dermatologists-in-training based on diagnostic verbal narratives. Such distinctions intuitively make sense, as the learning of medical domain knowledge requires advanced education in conjunction with substantial practical field experience. In a task such as medical image inspection, the subtle cues that point an observer to evidence that allow them to identify a clinical condition, while accessible to experts with training and perceptual expertise to guide their exploration, are likely to be missed by novices who lack that background and clinical understanding. Such expertise can then be integrated into human-centered health-IT systems (Guo et al., 2014), in order to introduce novel ways to retrieve medical images and take advantage of an understanding of which information is useful. It is reasonable to assume that this knowledge gap also applies to other knowledge-intensive clinical domains such as mental health. In this study, we explore this question and study if novice vs. expert annotation makes a difference for identifying distress in social media texts, as well as what the impact of expert vs. novice annotation is for subsequent computational modeling

with the annotated data.

Affect in language is a phenomenon that has been studied both in speech and in the text analysis domain, as well as in many other modalities (Calvo and D'Mello, 2010). Clearly, emotion is a key element in the human experience, but it is notoriously difficult to pin down and scholars in the affective sciences lack a single agreed-upon definition for emotion. Accordingly, different theoretical constructs have been proposed to describe affect and affect-related behaviors (Picard, 1997). In addition, research on affect in language has shown that such phenomena tend to be subjective, lack real ground truth (often resulting in moderate kappa scores), and have particularly fuzzy semantics in the gray zone where neutrality and emotion meet (Alm, 2008). These kinds of problem characteristics bring with them their own set of demanding challenges from a computational perspective (Alm, 2011). Yet, the nature of such problems make them incredibly important to study, despite the challenges involved.

Level of distress is a key element to consider when evaluating at-risk behaviors with respect to suicidality or depression. Lehrman et al. (2012) conducted a first study on the computational modeling of distress based on short forum texts, yet left many areas wide open for continued study. For example, analysis at scale is one such open issue. More specifically, Pestian and colleagues (Matykiewicz et al., 2009; Pestian et al., 2008) used computational methods to understand suicide notes. However, when it comes to preventive contexts, such data are less insightful. For preventive health, access to real time health-related data that dynamically evolves can allow us to address macro-level analysis, and social media texts provide the additional opportunity to model the phenomena of interest at scale.

Sentiment analysis has been widely studied in a number of computational settings, including on various social networking sites. Relatively little of this work has focused on suicide or related psychological conditions. Masuda et al. (2013) study suicide on mixi. Cheng et al. (2012) consider the ethical and political implications of online data collection for suicide prevention. Jashinsky et al. (2013) show correlations between frequency in tweets related to suicide and actual suicide in the 50 United States of America. Sadilek et al. (2014) study depression on Twitter. De Choudhury and collaborators studied depression—in general and postpartem—in Twitter (De Choudhury et al., 2012a; De Choudhury et al., 2012b; De Choudhury et al., 2013; De Choudhury and Counts, 2013) and Facebook (De Choudhury et al., 2014). Homan et al. (2014) investigate depression in TrevorSpace. A number of social theories of suicide have been proposed (Wray et al., 2011). Most of this work was with respect to offline social systems.

A rather substantial body of work already exists on the use of Twitter to study emotion (Bollen et al., 2011b; Dodds et al., 2011; Wang et al., 2012; Pfitzner et al., 2012; Kim et al., 2012; Bollen et al., 2011a; Pfitzner et al., 2012; Bollen et al., 2011c; Mohammad, 2012; Golder and Macy, 2011; De Choudhury et al., 2012a; De Choudhury et al., 2012b; De Choudhury et al., 2013; De Choudhury and Counts, 2013; Hannak et al., 2012; Thelwall et al., 2011; Pak and Paroubek, 2010). For instance, Golder and and Macy study aggregate global trends in "mood," and show, among other , that people wake up in a relatively good mood that decays as the day progresses (Golder and Macy, 2011), Bollen et al. (Bollen et al., 2011c) show that tweets from users who took a standard diagnostic instrument for mood are often tied to current events, such as elections and holidays.

A common theme in social network analysis is that actors who share ties generally share similar properties. A widely-used (Bliss et al., 2012; Coviello et al., 2014; Bollen et al., 2011a) metric for testing the overall similarity between actors in a network for some property X is *assortativity*, defined as the Pearson correlation coefficient of X over all pairs of actors who share a tie (Newman, 2002). One line of research seeks to discover the mechanism through which such correlations occur (Newman, 2002). At the most fundamental level, this is a matter of whether like individuals seek each other out (called selection, or—confusingly enough—homophily) or whether related individuals influence one another. Teasing apart which of these two processes can be rather challenging and generally requires some level of experimental design (Centola, 2010; Centola, 2011) For instance, Coviello et al. (2014) study the spread of mood in Twitter. They notice a very small—but statistically significant—spreading of mood over Facebook.

From the clinical perspective of detecting in-

dividuals who exhibit a high risk for committing suicide, determining causality remains a challenge to this multidimensional problem; however, finding patterns in the social interactions of individuals who exhibit distress and/or talk about suicide or suicide risk factors may provide additional insight. For our purposes, then a more relevant theory is perhaps that of social support (Wellman and Wortley, 1990) which seeks to clarify the social forces—protective, preventative, persuasive, or coercive—that affect behavior.

At the most basic level, one can distinguish between *weak* and *strong* social ties and observe different behavior and effects between them.

Following in the work of Bliss et al. (Coviello et al., 2014) and Bollen et al. (Bollen et al., 2011a) we show that mood is assortative. We additionally consider the predictive power of various measureable notions of tie strength. We study suicide risk levels here but we would expect our methods would apply to other domains.

3 Methods

In this section, we describe the methods we use to label and detect distress in Twitter data. Our process involves four main phases: (1) We filter a corpus, obtained from Sadilek et al. (2012), of approximately 2.5 million tweets from 6,237 unique users in the New York City area that were sent during a 1-month period between May and June, 2010, into a set of 2,000 tweets that are relatively likely to be centered around suicide risk factors. (2) We annotated each of these 2,000 tweets with their level of distress. (3) We then train support vector machines and topic models with the annotated data, except for a held-out subset of 200 tweets. (4) Finally, we assess the effectiveness of these methods on the held-out set.

3.1 Filtering tweets

We first preprocessed each tweet in the corpus by (a) converting all text to lower case; (b) stripping out punctuation and special characters; and (c) building a dictionary of more than 5,400 terms that captured informal Twitter registers, such as abbreviations and netspeak, based on http://www.noslang.com/dictionary.

In order to test the effectiveness of various methods of capturing training data, we used two different methods to filter for tweets that are relatively likely to center on suicide risk factors. The first method, we used the Linguistic Inquiry and Word Count (LIWC) to capture 1,370 tweets by sampling randomly from the all tweets with at least the $2,000$th-highest LIWC sad score. LIWC has been widely used to estimate emotion in online social networks, and specifically to mood on Twitter. This slight amount of randomness in filtering tweets this way was intended to avoid selecting obvious false positives, such as the use of "sad" in nicknames.

Next, we adopted a collection of of inclusive search terms/phrases from (Jashinsky et al., 2013), which was designed specifically for capturing tweets related to suicide risk factors, and applied them to our source corpus. These terms yielded 630 tweets.

Source tweets	Number of tweets	2,535,706
	Unique geo-active users	6,237
	"Follows" relationships	102,739
	"Friends" relationships	31,874
Filtered tweets	Number of tweets	2000
	Unique users	1467
	Unique tokens	1714167
	Unique bigrams	9246715
	Unique trigrams	13061142
Categories distribution	LIWC sad	1370
	Depressive feeling	283
	Suicide ideation	123
	Depression symptoms	72
	Self harm	67
	Family violence/discord	47
	Bullying	10
	Gun ownership	10
	Drug abuse	6
	Impulsivity	6
	Prior suicide attempts	2
	Suicide around individual	2
	Psychological disorders	2

Table 1: Summary statistics of the and thematic categories distributions of the collected dataset. The data was collected from NYC. Geo-active users are those who geo-tag (i.e., automatically post the GPS location of) their tweets relatively frequently (more than 100 times per month).

3.2 Novice and Expert Tweet Annotation

We then divided the resulting set of 2,000 filtered tweets (1,370 from the LIWC sad dimension and 630 from suicide-specific search terms), into two sets of 1,000 tweets each. Both sets had the same proportion of LIWC-filtered and suicide-specific-filtered tweets. Two non-experts annotated the first set and a clinical psychologist specializing in suicide prevent annotated the second set. Each tweet in each set was rated on a four-point scale (H, ND,

```
978: Date: fri, 04 jun 2010 13:46:21 +00
    -3: dat man on maury is overreacting
    -2: @XXXX cedes!!! [-0:21:25]
    -1: yesssss! da weatherman was wronq
>>> @tragedytm812 awwww thanks trae-trae
     1: rt @XXXX: abt 2 hop in a kab to
     2: @XXXX yeaa [+0:03:59]
     3: @XXXX wassup? [+0:05:28]
Msg_id: 15416569951  [Distress: ND, LIWC
```

Figure 1: Example input for annotator. Each line is one tweet. The tweet being annotated is indicated by >>>.

LD, HD) according to the level of distress evident (Table 2).

Code	Distress Level
H	happy
ND	no distress
LD	low distress
HD	high distress

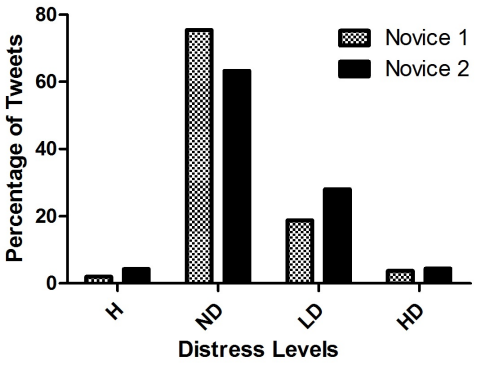
Table 2: Distress-related categories used to annotate the tweets.

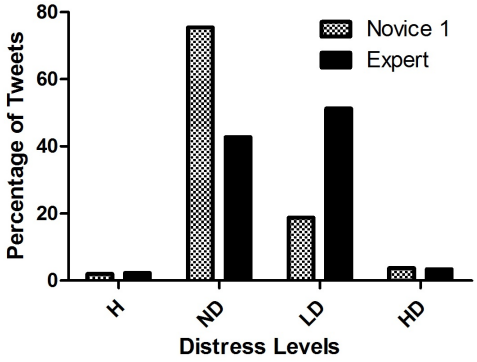
For the annotation process itself each tweet was provided with a context, i.e., three tweets before and after the tweet to be annotated, along with the timestamp of these tweets and thematic category to which the tweet belonged (Figure 1).

3.3 Modeling

We represent each tweet as a collection of all unigram, bigram and trigram in the message. For example, a simple tweet "I am so happy" is represented as the following *feature vector*: {I, am, so, happy, I am, am so, so happy, I am so, am so happy}. This method allows one to construct prior probabilities on pairs and triples of consecutive words and thus model the probability spaces of arbitrarily long utterances, in a way that is natural and often effective in representing linguistic data for the purpose of classification or topic modeling

We perform topic modeling on our dataset to compare the topics. Topic modeling is often used to analyze text data by finding topics within a corpus of documents. A topic is characterized by lexical items that are likely to occur with the topic. These models are capable of connecting words with similar meanings and distinguish words with multiple meanings. We utilize Latent Dirichlet Algorithm (LDA) (Blei et al., 2003) to create these topics, in this method the documents (in our case





tweets) are represented as random mixtures over latent topic where each topic is characterized by a distribution over words. Before performing the topic modeling, the stop words and words that occur only once in the dataset are removed. The LDA algorithm then establishes three topics using 100 iterations.

We use Support Vector Machines (SVM) to evaluate the statistical power of our annotations. SVMs treat each tweet as a point in an extremely high dimensional space (one dimension per uni-, bi-, and tri- gram in the corpus). SVMs are a form of *linear separator*. They have proven to be an extremely effect tool in classifying text in numerous settings, including Twitter.

4 Results

4.1 Annotation

4.2 Geography and Emotional Language

4.3 Strength of Ties

One of the fundamental properties of social networks appears to be *tie strength*, or how "close" socially two people are. A large body of literature suggests that people are more likely to share personal information with stronger ties, and that weak ties play an important role in providing new information. Measuring tie strength is problem-

-3:	i wish i had an older brother...	[-0:06:08]
-2:	i'll be alright though. one day...	[-0:02:38]
-1:	@XXXX yeah i'll be alright, it must be a phase...	[-0:02:00]
>>>	i'm not committing suicide or anything of that nature so don't panic. maybe i'm going through a phase, maybe not. just tired of loneliness.	<<<
1:	@XXXX that's easier said then done though, you don't just break out of it overnight bro. especially if you've been a loner for a while..	[+0:07:07]
2:	@silentbx i hear you.	[+0:07:37]
3:	N/A	[+0:07:56]

Table 3: Example for High Distress

-3:	highed up right now watching bloopers	[-11:33:54]
-2:	rt @XXXX: good morning to everybody !!!!	[-8:25:21]
-1:	hate in my heart	[-0:38:44]
>>>	i need to leave this world for good	<<<
1:	i need a break from life	[+0:00:33]
2:	all you hypocrites of the world will perish in hell one day	[+1:29:39]
3:	watching every body hates chris	[+2:02:56]

Table 4: Example for High Distress

atic, as there is no gold standard here. Social networking services seem to exacerbate the disparity between strong and weak ties, as many have "friends" or "followers" whom they may not even know personally, and also create their own problems and opportunities for estimating tie strength. In large-scale network analysis, researchers have sometimes characterized tie strength by the *embeddedness* of an edge, which is the number of friends in common that two actors sharing a tie have. Highly embedded links are part of a strong social fabric, and represent strong ties. Another method of estimating tie strength is to measure the amount of activity between users. In this study, we investigate both the role that embeddness and activity play in corrolating suicide-related language.

Figure 4.3 shows how the use of sad language (measure by the LIWC sad feature) correlates among users sharing different tie strength and between personal and broadcast messages.

5 Discussion

As previously mentioned, many of the risk factors for suicidal behavior may be linked to other expressions of distress such as aggression and inter-

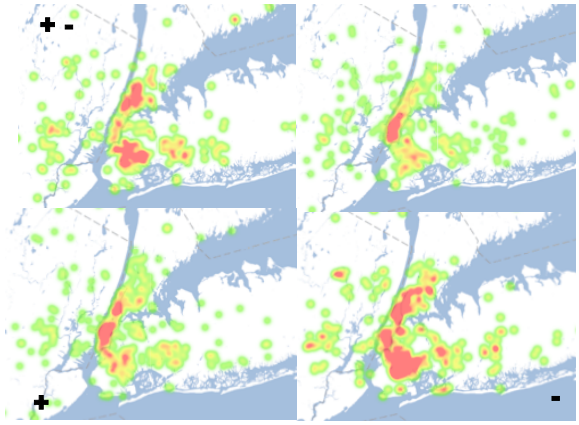


Figure 2: Heatmaps showing the most common tweet location for those individuals who (clockwise from upper left) are among the top 25% in using negative and positive emotional language, in the bottom 25% of both categories, in the top 25% of negative language and bottom 25% of positive language, and in the top 25% of positive language and bottom 25% of negative language.

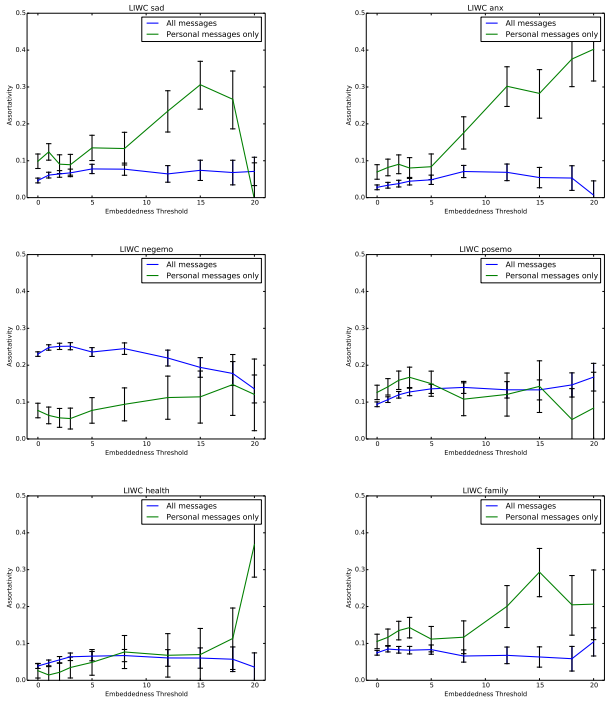


Figure 3: Correlations between twitter friends of various LIWC scores increase as the strength of ties increase, while others behave very differently. Interestingly, two of the LIWC scores most closely associated with distress, sadness and anxiety, are among those most strongly affected by ignoring those friendships having fewer than a certain number mutal friends (i.e., embeddedness).

Group	Kappa
LIWC	0.3562
Thematic Category	0.5567
250 Tweets	0.4925

Table 5: Inter Annotator Agreement for Novice 1 and 2

	H	ND	LD	HD
H	0	2	0	0
ND	1	85	2	1
LD	0	22	9	0
HD	0	1	0	2

Table 6: Confusion Matrix for LIWC for Novice 1 and 2

	H	ND	LD	HD
H	4	6	0	0
ND	0	55	12	1
LD	0	12	22	5
HD	0	1	0	4

Table 7: Confusion Matrix for Thematic Category for Novice 1 and 2

	H	ND	LD	HD
H	4	8	0	0
ND	0	150	14	2
LD	0	34	31	5
HD	0	2	3	4

Table 8: Confusion Matrix for 250 tweets for Novice 1 and 2

personal violence (Mann et al., 1999). The goal of this study is to classify whether or not tweets were related to distress in order to determine the feasibility of classifying suicidal behaviors. However, due to the overlap between internal and external expressions of anger, it is difficult to classify suicidal behavior without more contextual information. Consistent with the stress diathesis model for suicidal behavior, aggression was an emerging theme that arose from the data. A number of individuals tweeted about feeling empty, hopeless, angry, frustrated, and alone. While these are risk factors for internalizing aggression (i.e., suicidal behavior); these states are also associated with externalizing aggression. In addition to overt expressions of anger and violence, many of the humorous tweets had an aggressive undertone. Individuals often exert dominance by using pejorative and derogatory language, and the content of these tweets was often suggestive of past or current distress.

5.1 Challenges with Annotation

There are unique challenges in annotating data from Twitter. Aside from having to become familiar with different types of slang and abbreviations that could have multiple meanings, this format provides limited background context to inform the annotation process. Outside of the theoretical informed annotation process, there were emerging themes of aggression, privilege and oppression, and daily struggles, among other. As a result of the aggressive context, personal bias may have impacted annotation decisions. For instance, numerous tweets contained sarcasm and dark humor which may result in annotators underestimating or overlooking actual distress. In addition, by pulling data from Twitter, critical information such as pictures and the context behind information that has been retweeted. Specifically, a few individuals retweeted in a humorous manner about what to say to someone who considering suicide; however, without any knowing the circumstances of the original message it was difficult to classify this tweet. It could represent dark humor or it could be a form of bullying.

5.2 Limitations

As ground truth, we rely on tweets hand-annotated by experts and non-experts. However, the mental state of another individual, observed from a line or two of text often written in an informal register is necessarily hard to discern and, even under less noisy conditions, extremely subjective; even the observers' personal understandings of such concepts as "distress" may differ drastically. This makes inter-annotator agreement quite a challenge, to say nothing of observation in some objective fashion of the true mental state.

Higher levels of suicidal ideation have an inverse relationship with all types of help-seeking and a positive correlation with the decision to not seek support (Deane et al., 2001). Thus we would expect suicidal individuals to generally be less active on social media than those who are not. (One ray of sunshine is that a number of studies have shown a positive correlation between online social network use and negative mood. Perhaps this means in part that individuals who are depressed

	ND	D
ND	153	16
D	36	150

Table 9: Confusion Matrix for 250 tweets for Novice 1 and 2

	ND	D
ND	88	3
D	23	45

Table 10: Confusion Matrix for LIWC tweets for Novice 1 and 2

are slower to disengage on- rather than off-line.) Part of the problem in assessing the effectiveness of self-reporting is the relative rareness by which suicide occurs, and by the inherent subjectivity of the act, which makes any data on suicide fuzzy.

5.3 Tools Used

6 Conclusion and Future Work

Acknowledgments

References

Cecilia Ovesdotter Alm. 2008. Affect in text and speech.

Cecilia Ovesdotter Alm. 2011. Subjective natural language problems: Motivations, applications, characterizations, and implications. In *Proceedings of 49th Annual Meeting of the Assoc. for Computational Linguistics: Human Language Technologies, Portland, OR*, pages 107–112.

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022, March.

Catherine A Bliss, Isabel M Kloumann, Kameron Decker Harris, Christopher M Danforth, and Peter Sheridan Dodds. 2012. Twitter reciprocal reply networks exhibit assortativity with respect to happiness. *Journal of Computational Science*, 3(5):388–397.

Johan Bollen, Bruno Gonçalves, Guangchen Ruan, and Huina Mao. 2011a. Happiness is assortative in online social networks. *Artificial life*, 17(3):237–251.

Johan Bollen, Huina Mao, and Xiaojun Zeng. 2011b. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8.

Johan Bollen, Alberto Pepe, and Huina Mao. 2011c. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, pages 450–453.

	ND	D
ND	65	13
D	23	34

Table 11: Confusion Matrix for Thematic Category tweets for Novice 1 and 2

Topic No	**Words**	**random**
Topic 1	miss u, leave alone, sleep forever, win lose, gon lose, left alone, #iconfess hate, lost best, best friend, think insomnia	let-know, don't-want, bout-2, even-tho, right-now, jus-got, gotta-go, don't-wanna, wit-da
Topic 2	hate job, feel sad, don't wanna, feel helpless, bed lonely, feel better, miss you, sad =/, tired everything, miss love	feel-like, look-like, let's-go, last-night, looks-like, don't-get, fuck-wit, show-love, smile-face, gonna-start
Topic 3	miss you!, wanna cry, committing suicide, tired living, miss 2, one person, broke bitches, worst feeling, leave world, bout go	time-get, go-sleep, know-(cont), can't-wait, even-though, hip-hop, big-baby, lil-wayne, listen-2, don't-think
Topic 4	commit suicide, get hurt, miss baby, feel empty, :(miss, lost phone, don't let, drug overdose, can't wait, work	don't-know, make-sure, dont-see, wats-good, hell-yea, r-u?, need-new, yall-niggas, can't-get, don't-care
Topic 5	feel like, tummy hurts, lost friend, ima miss, deserve die, right now..., hurts :(, get fat, every day, like crying	good-morning, happy-birthday, bout-go, what's-good, jus-w, chris-brown, right-now!, 2-da, don't-feel, don't-understand

Table 12: Topic Analysis on bigrams of High Distress and Random Tweets

Gregory K Brown, Aaron T Beck, Robert A Steer, and Jessica R Grisham. 2000. Risk factors for suicide in psychiatric outpatients: a 20-year prospective study. *Journal of consulting and clinical psychology*, 68(3):371.

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

Damon Centola. 2010. The spread of behavior in an online social network experiment. *science*, 329(5996):1194–1197.

Damon Centola. 2011. An experimental study of homophily in the adoption of health behavior. *Science*, 334(6060):1269–1272.

Qijin Cheng, Shu-Sen Chang, and Paul SF Yip. 2012. Opportunities and challenges of online data collection for suicide prevention. *The Lancet*, 379(9830):e53–e54.

Lorenzo Coviello, Yunkyu Sohn, Adam D. I. Kramer, Cameron Marlow, Massimo Franceschetti, Nicholas A. Christakis, and James H. Fowler. 2014. Detecting emotional contagion in massive social networks. *PloS one*, 9(3):e90315.

Alex E Crosby, LaVonne Ortega, and Cindi Melanson. 2011. *Self-directed violence surveillance: Uniform definitions and recommended data elements*. Centers for Disease Control and Prevention, National Center for Injury Prevention and Control, Division of Violence Prevention.

Train Set	Test Set	Precision	Recall	F-Measure
N1	N1	0.53	0.63	0.58
N1	E	.58	0.27	.37
E	E	0.59	0.71	0.64
E	N1	0.34	0.85	0.48
N1 + E	N1 + E	0.33	0.41	0.37

Table 13: Perform of SVN-based classification when the training and testing sets are alternately Novice 1 (N1) or the Expert (E). In cases where the training and test source coincide, the test set is a held-out set of 100 randomly selected (or 200 when N1 and E both used as in the last row) tweets. Otherwise, the training and test sets have 1000 tweets each and are disjoint.

Munmun De Choudhury and Scott Counts. 2013. Understanding affect in the workplace via social media. In *16th ACM Conference on Computer supported cooperative work and Social Media (CSCW 2013)*, pages 303–316. ACM.

Munmun De Choudhury, Scott Counts, and Michael Gamon. 2012a. Not all moods are created equal! exploring human emotional states in social media. In *Sixth International AAAI Conference on Weblogs and Social Media*.

Munmun De Choudhury, Michael Gamon, and Scott Counts. 2012b. Happy, nervous or surprised? classification of human affective states in social media. In *Sixth International AAAI Conference on Weblogs and Social Media*.

Munmun De Choudhury, Scott Counts, and Eric Horvitz. 2013. Major life changes and behavioral markers in social media: Case of childbirth. In *Proc. CSCW*.

Munmun De Choudhury, Scott Counts, Eric J Horvitz, and Aaron Hoff. 2014. Characterizing and predicting postpartum depression from shared facebook data. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pages 626–638. ACM.

Frank P Deane, Coralie J Wilson, and Joseph Ciarrochi. 2001. Suicidal ideation and help-negation: Not just hopelessness or prior help. *Journal of clinical psychology*, 57:901–914.

Peter Sheridan Dodds, Kameron Decker Harris, Isabel M Kloumann, Catherine A Bliss, and Christopher M Danforth. 2011. Temporal patterns of happiness and information in a global social network: Hedonometrics and twitter. *PloS one*, 6(12):e26752.

Bradley N Gaynes, Suzanne L West, Carol A Ford, Paul Frame, Jonathan Klein, and Kathleen N Lohr. 2004. Screening for suicide risk in adults: a summary of the evidence for the us preventive services task force. *Annals of Internal Medicine*, 140(10):822–835.

S.A. Golder and M.W. Macy. 2011. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science*, 333(6051):1878–1881.

Xuan Guo, Rui Li, Cecilia Ovesdotter Alm, Qi Yu, Jeff Pelz, Pengcheng Shi, and Anne Haake. 2014. Infusing perceptual expertise and domain knowledge into a human-centered image retrieval system: A prototype application. In *Proceedings of the Symposium on Eye Tracking Research and Applications*. ACM.

Aniko Hannak, Eric Anderson, Lisa Feldman Barrett, Sune Lehmann, Alan Mislove, and Mirek Riedewald. 2012. Tweetinin the rain: Exploring societal-scale effects of weather on mood. In *Proceedings of the 6th International AAAI Conference on Weblogs and Social Media (ICWSM12)Dublin2012*.

Louise Harriss and Keith Hawton. 2005. Suicidal intent in deliberate self-harm and the risk of suicide: the predictive power of the suicide intent scale. *Journal of Affective Disorders*, 86(2):225–233.

Melonie Heron and Betzaida Tejada-Vera. 2009. Deaths: leading causes for 2005. *National vital statistics reports: from the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System*, 58(8):1–97.

Christopher M Homan, Naiji Lu, Xin Tu, Megan C Lytle, and Vincent Silenzio. 2014. Social structure and depression in trevorspace. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pages 615–625. ACM.

Lisa M Horowitz and Elizabeth D Ballard. 2009. Suicide screening in schools, primary care and emergency departments. *Current opinion in pediatrics*, 21(5):620–627.

Jared Jashinsky, Scott H Burton, Carl L Hanson, Josh West, Christophe Giraud-Carrier, Michael D Barnes, and Trenton Argyle. 2013. Tracking suicide risk factors through twitter in the US. *Crisis*, pages 1–9.

Ronald C Kessler, Guilherme Borges, and Ellen E Walters. 1999. Prevalence of and risk factors for lifetime suicide attempts in the national comorbidity survey. *Archives of general psychiatry*, 56(7):617–626.

Suin Kim, J Bak, and Alice 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*.

Michael Lehrman, Cecilia Ovesdotter Alm, and Ruben Proano. 2012. Detecting distressed vs. non-distressed affect state in short forum texts. In *Proceedings of the Workshop on Language in Social Media (LSM 2012) at the Conference of the North Am. Chapter of the Assoc. for Comp. Linguistics-Human Language Technologies, Montreal, Canada*, pages 9–18.

Rui Li, Jeff Pelz, Pengcheng Shi, and Anne Haake. 2012. Learning image-derived eye movement patterns to characterize perceptual expertise. In *CogSci*, pages 1900–1905.

J John Mann, Christine Waternaux, Gretchen L Haas, and Kevin M Malone. 1999. Toward a clinical model of suicidal behavior in psychiatric patients. *American Journal of Psychiatry*, 156(2):181–189.

Naoki Masuda, Issei Kurahashi, and Hiroko Onari. 2013. Suicide ideation of individuals in online social networks. *PloS one*, 8(4):e62262.

Pawel Matykiewicz, Wlodzislav Duch, and John P. Pestian. 2009. Clustering semantic spaces of suicide notes and newsgroup articles. In *Proceedings of the Workshop on BioNLP, Boulder, Colorado*, pages 179–184.

Saif M Mohammad. 2012. # emotional tweets. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation*, pages 246–255. Association for Computational Linguistics.

Mark EJ Newman. 2002. Assortative mixing in networks. *Physical review letters*, 89(20):208701.

Matthew K Nock, Guilherme Borges, Evelyn J Bromet, Christine B Cha, Ronald C Kessler, and Sing Lee. 2008. Suicide and suicidal behavior. *Epidemiologic reviews*, 30(1):133–154.

Alexander Pak and Patrick Paroubek. 2010. Twitter as a corpus for sentiment analysis and opinion mining. In *Proceedings of LREC*, volume 2010.

John P. Pestian, Pawel Matykiewicz, and Jacqueline Grupp-Phelan. 2008. Using natural language processing to classify suicide notes. In *BioNLP 2008: Current Trends in Biomedical Natural Language Processing, Columbus, Ohio*, pages 96–97.

René Pfitzner, Antonios Garas, and Frank Schweitzer. 2012. Emotional divergence influences information spreading in twitter. *AAAI ICWSM*, 2012:2–5.

Rosalind W. Picard. 1997. *Affective Computing*. MIT Press, Cambridge, MA, USA.

Adam Sadilek, Henry A Kautz, and Vincent Silenzio. 2012. Predicting disease transmission from geotagged micro-blog data. In *AAAI*.

Adam Sadilek, Christopher Homan, Walter S. Lasecki, Vincent Silenzio, and Henry Kautz. 2014. Modeling fine-grained dynamics of mood at scale. In *WSDM 2014 Workshop on Diffusion Networks and Cascade Analytics*.

David Shaffer, Michelle Scott, Holly Wilcox, Carey Maslow, Roger Hicks, Christopher P Lucas, Robin Garfinkel, and Steven Greenwald. 2004. The columbia suicidescreen: Validity and reliability of a screen for youth suicide and depression. *Journal of the American Academy of Child & Adolescent Psychiatry*, 43(1):71–79.

Mike Thelwall, Kevan Buckley, and Georgios Paltoglou. 2011. Sentiment in twitter events. *Journal of the American Society for Information Science and Technology*, 62(2):406–418.

Wenbo Wang, Lu Chen, Krishnaprasad Thirunarayan, and Amit P Sheth. 2012. Harnessing twitter 'big data' for automatic emotion identification. In *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)*, pages 587–592. IEEE.

Barry Wellman and Scot Wortley. 1990. Different strokes from different folks: Community ties and social support. *American journal of Sociology*, pages 558–588.

Kathryn Womack, Wilson McCoy, Cecilia Ovesdotter Alm, Cara Calvelli, Jeff B. Pelz, Pengcheng Shi, and Anne Haake. 2012. Disfluencies as extra-propositional indicators of cognitive processing. In *Proceedings of the Workshop on Extra-Propositional Aspects of Meaning in Computational Linguistics*, pages 1–9. Association for Computational Linguistics.

Matt Wray, Cynthia Colen, and Bernice Pescosolido. 2011. The sociology of suicide. *Annual Review of Sociology*, 37:505–528.