Automatic Extraction of Informal Topics from Online Suicidal Ideation

Reilly Grant
Grinnell College
Grinnell, IA, USA
grantrei@grinnell.edu

David Kucher
University of Michigan
Ann Arbor, MI, USA
dkucher@umich.edu

Ana M. León
University Juárez Autónoma
de Tabasco
Tabasco, MX
aleon@tabasco.edu

Jonathan Gemmell[*]
School of Computing,
DePaul University
Chicago, IL, USA
jgemmell@cdm.depaul.edu

Daniela Raicu
School of Computing,
DePaul University
Chicago, IL, USA
draicu@cdm.depaul.edu

Samah Fodeh
Yale Center for
Medical Informatics,
Yale University
New Haven, CT, USA
samah.fodeh@yale.edu

ABSTRACT

Suicide is an alarming public health problem accounting for a considerable number of deaths each year worldwide. Many more individuals contemplate suicide. Understanding the attributes, characteristics, and exposures correlated with suicide remains an urgent and significant problem. As social networking sites have become more common, users have adopted these sites to talk about intensely personal topics, among them their thoughts about suicide. Such data has previously been evaluated by analyzing the language features of social media posts and using factors derived by domain experts to identify at-risk users. In this work, we automatically extract informal latent recurring topics of suicidal ideation found in social media posts. Our evaluation demonstrates that we are able to automatically reproduce many of the expertly determined risk factors for suicide. Moreover, we identify many informal latent topics related to suicide ideation such as concerns over health, work, self-image, and financial issues.

1. INTRODUCTION

Suicide, the act of causing one's own death, is the tenth leading cause of mortality in the United States and is estimated to cost 44.6 billion dollars per year. This understates the severity of the problem, as for every attempted suicide, there are nearly 10 times as many people who contemplate suicide [23]. Suicidal ideation includes a wide range of thoughts from momentary consideration to extensive planning or incomplete attempts. The scope and impact of this mental health issue make understanding it a public health priority.

When discussing their ideations, many individuals often reference common symptoms: feeling helpless, feeling alone, excessive fatigue, low self-esteem, the feeling that one's mind is racing, or excessive focus on dormant goals [1]. Understanding the common themes in suicidal ideation can help us understand the patterns behind suicidal thoughts, ultimately leading to treatment and prevention.

Clinical research toward understanding suicide has identified several risk factors. Mental disorders such as depression, schizophrenia, alcoholism, and substance abuse all play a contributing role. Additionally, the emotional stress caused by bullying, interpersonal relationships, and finances are also important factors [14]. However, these descriptions of suicidal ideation often capture a clinical viewpoint.

With the rise in sophistication and acceptance of online social networks, individuals contemplating suicide have increasingly expressed their suicidal ideation in online forums, tweets, and other online media. The result is a vast collaborative description of the thoughts and motivations associated with suicide. In this paper, we leverage advanced topic modeling techniques to extract informal latent topics from this data.

Topic modeling is a machine learning approach for eliciting abstract topics from a collection of documents. This approach can be leveraged to discover common themes present in online posts such as depression, drug use, or violence. The idea of "depression" might be captured by a collection of related words such as "pain", "feelings", "fear", "stress", and "suffering".

In this paper, we perform topic modeling on over 130,000 submissions to r/SuicideWatch, an online forum described as a place of support for those suffering suicidal thoughts. We begin by learning semantic embeddings for words in the posts via a shallow, two-layer neural network. Then we cluster the words into topics producing informally generated latent topics. Finally, we evaluate these informal topics by comparing them to suicidal risk factors and common themes identified by mental health professionals [14].

Our experimental results reveal that we are able to automatically generate quality embeddings for words and corresponding topic models. Many of these topic models corre-

[*] Corresponding Author

spond to risk factors that domain experts have previously proposed. In some cases, our topic models were more specific, focusing on a narrow interpretation of the risk factor. In other cases, our topics were more broad, encompassing multiple risk factors at once. This suggests that the topics extracted from social media posts created by those experiencing suicidal ideation may have a different focus and specificity than those generated by mental health professionals.

The rest of this paper is organized as follows. In Section 2 we present related work. Section 3 describes how we detect topics from social media posts using Word2Vec and k-means clustering. Our source of data is presented in Section 4 and our experimental results in Section 5. We conclude the paper with a discussion of our findings and direction for future work in Section 6.

2. RELATED WORK

Researchers have previously attempted to use the massive amount of data generated through social media to characterize the mental health of users [5, 17], leading to the development of computational tools [25]. Attempts have been made to predict depression, identify suicidal Twitter posts, and analyze the effect of suicide in the media on suicidal ideation in social platforms [9, 18, 24]. Risk factors of suicide [19, 22] identified by domain experts are often leveraged in such studies.

A common tool used to analyze social media posts is the Linguistic Inquiry and Word Count (LIWC) [25]. Progress has been made using this tool to analyze text related to suicide and depression, often in social media posts [2, 5, 6, 9, 10, 12, 18].

An early study used the LIWC on Twitter to analyze the impact of depression on social media activity [9]. Twitter data has been used to analyze suicidal ideation [7, 17, 24]. In one study, tweets were filtered by using specific search terms which were associated with 12 suicide risk factors [17]. The twelve risk factors include bullying, depressive feelings, depression symptoms, drug abuse, family violence/discord, gun ownership, impulsively, prior suicide attempts, psychological disorders, self-harm, suicide around the individual, and suicide ideation [19, 22]. We also evaluate the twelve risk factors identified in these studies. These researchers found that the volume of suicide-related tweets correlated to suicide rates by U.S. state, showing that Twitter data could be indicative of a population's mental health.

One study used human "coders" to label tweets according to their level of concern with respect to suicide. Language models were then used to predict the appropriate concern for new tweets [24]. Another study analyzed the social media content of Twitter users prior to their public declaration of a suicide attempt and found that there may be indications of suicidal ideation or intent based on posts leading up to a suicide attempt [7].

There have also been studies which focused on the social media platform Reddit, specifically the subreddit called r/SuicideWatch. One study analyzed changes in suicide content in the wake of celebrity suicides by measuring post volume and modeling topics in the text [18]. Another study observed the propensity of users discussing mental health issues to transition into discussing suicidal [10]. The language that people use in Reddit has been shown to differ between subreddits focused on different mental health concerns [12].

In this work, we leverage computationally generated language models to explore suicidal ideation. Examples of language models include simple bag-of-word models [8] and extend to more robust models such as probabilistic latent semantic analysis [15], latent dirichlet allocation [3], and Word2Vec [20, 21]. Such language models have been used to explore numerous topics such as comparing topics in data [16], recommendation systems [29], and different languages [28]. We focus on the Word2Vec language model developed by Mikolov et al. [20, 21].

Our work extends upon these previous efforts in the following ways. Rather than using pre-defined risk factors or labeled data to identify at-risk users, we automatically discover topics from the users' posts by leveraging Word2Vec language models. We compare the latent topics identified in posts to risk factors proposed by domain experts.

3. METHODOLOGY

In this section, we provide a detailed description of our procedure including how we represent words with Word2Vec models and then use k-means clustering to produce topics in text data. See Figure 1.

3.1 Word Embeddings

We represent words as a vector of real numbers [27]. More formally, each word $\vec{w}$ is represented as:

$$\vec{w} = \langle \phi(i_1), \phi(i_2)...\phi(i_n) \rangle \qquad (1)$$

where $\phi(i_1)$ through $\phi(i_n)$ represent the weight of the ith word in the vector space.

We can think of these word representations as populating a high dimensional space where the relative locations contain semantic information. For example, in previous work, the relationship of a country to its capital city has been represented by their relative position in the vector space [20, 21]. There are several methods for learning these weights; we leverage Word2Vec.

3.2 Word2Vec

Many topic modeling algorithms exist, including latent semantic indexing, latent Dirichlet allocation, and non-negative matrix factorization. In this work, we turn our attention to

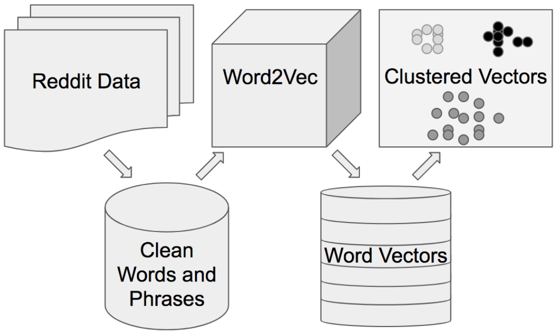

Figure 1: Flowchart of methodology. This describes our process of converting r/SuicideWatch data to a language model, then clustering the vectors provided by Word2Vec to identify the informal latent topics present in the posts.

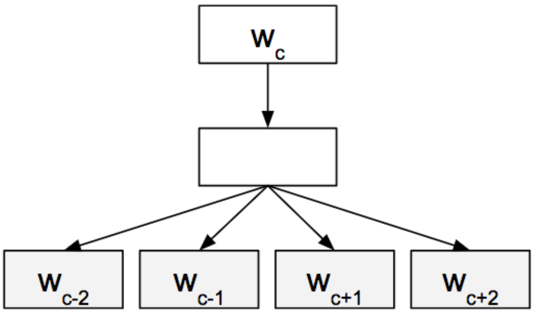

Figure 2: Architecture for the skip-gram model. The skip-gram model predicts the distributed representations of neighbors given a word. In this figure, the representation has a window size of 2, where w_c is the target word being evaluated, and w_{c+i} denotes the surrounding context words.

Word2Vec, which has been argued to have many advantages over these earlier algorithms [20, 21]

Word2Vec describes two implementations of a shallow neural network, the continuous bag of words (CBOW) model and the skip-gram model. We focus on the skip-gram model in this work, which learns vector representations of words by predicting neighboring words in a text. See Figure 2.

Common words such as "the" add little meaning to the model and add computational time. Instead of using these words, the model often skips over them and goes to the next word when training. Word2Vec does this by using subsampling, a probabilistic approach with the most common words having the greatest chance of being ignored, and the least common words having the least chance of being ignored.

In contrast to many other neural network models, the skip-gram model includes only a single hidden layer, dramatically reducing both training time and complexity [20, 21]. Learning the word representation is achieved by performing back-propagation on our training examples. Instead of updating each of the many neurons used in the neural network, negative sampling [21] updates a small, specified amount of neurons. Since one of the most computationally expensive parts of training a neural network is the act of updating the weights, this technique greatly reduces the training time. Finally, the softmax function normalizes the output of the neural network, so that sum of all outputs is equal to 1.

Word2Vec capitalizes on the fact that similar words should have similar probabilities of appearing in the same context. Therefore the vector representations of similar words are "close" in vector space, often capturing rich semantic characteristics. It has been previously shown that Word2Vec performs accurately on tasks involving word similarity, analogy discovery, and text completion [21].

3.3 Clustering

The word representations are useful in their own right, often containing rich semantic information. However, using these representations as input into other algorithms, such as clustering, can produce meaningful collections of related words. Clustering is a technique wherein items are grouped together based on their similarity. Items in a cluster are "near" one another and "distant" from items in other clusters. In this work, we rely on Euclidean distance because we are interested in the relative positions of the representations in the vector space.

We leverage the k-means clustering algorithm [13] to produce clusters of words. We choose k-means clustering due to its simplicity and ability to create localized, spherical clusters. K-means begins with cluster centers at k random locations in the vector space. The algorithm assigns every item, in this case words to the nearest cluster. For each cluster, the mean of all items is calculated, and the cluster center is moved to that point. The process is repeated until there are no new assignments.

Clusters of words can be viewed as topics. The meaning of a topic is captured by the words in the cluster. For example, a topic containing the words "join", "sports", "team", "joined", "practice", and "won" describes the topic of playing team sports. Thus, we can identify latent topics in a corpus of text by analyzing the clusters of words generated.

4. SUICIDEWATCH

In this section, we first present the data gathered and used in our analysis. Researchers interested in the code and the data are invited to contact the authors.

Reddit is a website which enables users to aggregate, rate and discuss news, entertainment, politics and many other topics. According to Alexa[1], it is the 8th most popular website in the world. It was estimated by the Pew research center that 6% of online adults use Reddit [11]. The site is organized into a collection of "subreddits", each focused on a particular topic and administered by a collection of moderators.

The subreddit, r/SuicideWatch, is a forum in which online users are encouraged to post their thoughts regarding suicide. At the time of our data collection, it had over 58,000 subscribers[2]. Sometimes users express a preoccupation with the thought of suicide. Other times users discuss immediate plans to take their own life. These posts often contain a description of their mental state including depression, reaction to stress, their feelings of being alone and having a low self-esteem.

While most online sources of data are notoriously noisy, this particular subreddit is remarkably clean. Given the serious nature of the subreddit, individuals are less likely to post harassing comments or off-topic remarks. When users post such comments, the moderators of the subreddit quickly remove them.

We collected all posts from its inception in 2008 to 2016. Each post is often commented on by other individuals. In this work, we focused on the original post as it most often represents the suicidal ideation of a user and comments often represent emotional support from other users.

We cleaned this data. First, we removed empty posts in which the content had been deleted. Second, we removed links, and replaced them with the word "link". Third, we concatenated the text of the post to the title, as many users begin their post in the title and continue in the body of the post. Finally, we removed punctuation and other special

[1] www.alexa.com

[2] www.reddit.com/r/SuicideWatch

characters. After cleaning this data, we had 131,728 posts with 27,978,246 words, of which 84,607 words were unique, posted by 63,252 unique users.

5. RESULTS

In this section, we evaluate the models built upon the r/SuicideWatch data. We begin by exploring individual words to subjectively assess whether or not the word representations are effectively capturing semantic information. We analyze the clusters to assess their ability to express latent topics in the data. We then evaluate the clusters by comparing them to the risk factors previously defined by domain experts. Finally, we present a summary analysis of our findings.

5.1 Experimental Parameters

Once we obtained the data we began by creating vector representations using the Word2Vec from the gensim module for python [26]. Each post was processed using the window size of 5, common in the literature, which looks at the previous and next 5 words, along with the current word, looking at a total of 11 words at once.

Negative sampling was set to 20 from the default of 5 on the recommendation of the authors of the Word2Vec model, based on the size of our data [20, 21]. After extensive evaluation, we chose to represent words with a vector of 300 features using the skip-gram model and hierarchical softmax. This value seemed to provide rich semantic descriptions while minimizing computational overhead.

In order to preserve the meaning of phrases, we turned common phrases into single tokens, called n-grams. This allows phrases such as “new york” to be separate from “new” and “york” alone, which have very different meanings. This resulted in an increase in the size of vocabulary to 97,368 unique words and phrases. To avoid noise in the data, we set the minimum count for a word to be included as 10 occurrences. This removed noise in terms of misspelled words and unrecognized characters among other things. After filtering our vocabulary, we preserved 99.41% of all of the words in our data, which decreased our vocabulary to its final size of 28,663 unique words.

Next, we clustered the vector representations of words by using k-means clustering implemented by scikit-learn [4]. An important input to the algorithm is the selection of k. After extensive evaluation, we chose a value of 100 because it offered a sufficient number of clusters to capture the topics of the posts without being too large to manually evaluate. Regarding an error as the distance of each vector to its cluster center, we calculated the sum of the squared errors (SSE) for clusterings of size 5 through 400. The knee of the SSE curve was approximately 100 clusters.

To evaluate the clusters, we took the ten most common words from each cluster and attempted to assign the clusters to one of the twelve risk factors previously identified by experts in suicide ideation. The risk factors with some prominent clusters are shown in Table 2.

5.2 Analysis of Word Representations

To visually inspect the effectiveness of our word representations, we first subjectively evaluated the representations of “heartbreak”, “pills”, and “knife”. Table 1 contains these query words along with the most similar words from the corpus. For example, the most similar token to “knife” is “kitchen knife”.

In all three cases, the related words share meaningful semantic information. In the case of “knife”, the related words are synonyms. In the case of “pills”, the related words are specific types of pills such as “painkillers” or “tylenol”. In the case of “heartbreak”, the word representations appear to capture this emotional concept.

heartbreak	**pills**	**knife**
loneliness	sleeping pills	kitchen knife
betrayal	painkillers	blade
heartache	tylenol	razor
sadness	pain killers	razor blade

Table 1: Nearest words in vector space to test word.

Now, after looking at semantic similarity, we attempted to see if our word vectors could be used for analogical reasoning in the same way they were used in [21]. Since words are represented as vectors, it is possible to add and subtract them from each other. We first consider the vector resulting from “[father] - [man] + [woman]”. We found that the vector representation most similar to the vector created by the preceding arithmetic is the vector representation of “mother”.

In addition to capturing general semantic meanings, our model also captured semantic information relevant to suicidal ideation. For example, when considering vector representations, we found that “[abusive] - [physical] + [words]” is most similar to ”emotionally abusive”, and “[suicide] + [self]” is most similar to ”killing myself”. This indicated to us that the word embeddings have captured semantic information relevant to the topic of suicidal ideation.

Finally, we observed that our model captured subtle distinctions between some similar words. This is demonstrated by the relation “[family] - [love] + [obligation]” being most similar to the word ”relatives”. This example shows that even though ”family” and ”relatives” have many similar semantic components, our model is able to capture subtle distinctions in their meaning.

As the previous examples matched our intuitions, we believed that our model has effectively extracted significant semantic information from the corpus and is suitable for clustering to extract informal, latent topics.

5.3 Analysis of Informal Topics

To evaluate the clusters, we visually inspected the most common words in 100 clusters to see if they are related. For example one cluster contains the following terms: “since”, “past”, “suicidal”, “havent”, “times”, “attempt’, “attempted suicide”, “suicide attempt”, “almost killed”, and “failed attempt”. The words in this cluster discuss suicide attempts. Note that there are n-grams appearing in our clusters, indicating that the words constituting the phrases “attempted suicide”, “suicide attempt’, “almost killed”, and “failed attempt” were often used together in their respective phrases. These words when clustered together appear to capture the topic of past suicide attempts.

In another example, a cluster contains “physically”, “emotionally”, “bullied”, “treated”, “mentally”, “raped”, “ignored”, “rejected”, “abused”, and “abandoned”. These terms are mostly verbs describing some sort of abuse, both mental and physical. We observe that users often use these words when talking about physical abuse.

Finally, one cluster contains the terms “school”, “college”, “failed”, “class”, “university”, “grades”, “classes”, “failing”, “degree” and “major”. These terms are all used to describe education, especially higher education, with vocabulary ranging from matriculation to graduation. While this cluster does not represent a risk factor for suicide, it does indicate that people often talk about college in the context of suicide, perhaps as a stressor that can lead to suicidal tendencies.

Some clusters capture concrete topics such as those containing “drugs” or “guns”. Still, others capture emotional topics such as those containing “anxiety” or “sadness”. Some clusters appear immediately relevant to the study of suicide such as those containing “cut” or “pain”, while others represent cohesive clusters but do not clearly represent topics related to suicide such as those containing “clothes” or “week”. While it is not possible to present all clusters here, a curated selection can be found in Table 2.

5.4 Comparison to Risk Factors

In our previous section, we showed how the clusters we found extract meaningful topics from the r/SuicideWatch data. In this section, we compare these informally extracted topics to risk factors proposed by domain experts. In this work, we draw from the risk factors used in Jashinsky et al., where Twitter data was analyzed according to risk factors identified by the National Institute of Mental Health and by Lewinsohn et al. [17, 19, 22]. The twelve risk factors can be seen in the left-hand column of Table 2.

While analyzing our clusters, we identified many topics that matched very closely with the proposed risk factors. For example, the notion of “Suicide Ideation” is captured by several clusters. For convenience, we have labeled the columns “Cluster 1” through “Cluster 5”, but there is no natural order to the clusters. On the row labeled “Suicide Ideation” we find that the first cluster expresses thoughts about committing suicide. The second, third, and fourth clusters discuss methods of committing suicide, and the fifth cluster portrays the user’s thoughts about planning suicide. Additionally, the risk factor “Self-Harm” also aligns well with our informal latent topics. Cluster one captures the notion of cutting oneself while cluster two focuses more clearly on damage to body parts such as “body”, “blood”, “burn” and many other words describing harm to one’s body. These topics both fit within the risk factor “Self-Harm”, showing agreement between our automatically generated topics and expert opinions. The informal topics captured by these clusters seem to embody the notion on suicidal ideation and suggest that our topics agree with the proposed risk factor.

We found that some clusters were not squarely matched with risk factors. For example, we assigned the cluster containing “mom”, “dad”, “kill herself” and “kill himself” to the risk factor “Suicide Around Individual”. This cluster also includes “friend”, “dog”, ”gf”, “boyfriend” and a long list of other types of individuals in the user’s life. This informal latent topic seems to capture not only the occurrence of suicide but also examples of strong personal relationships, the loss of which could be particularly traumatic. Thus, this cluster relates to both of the risk factors “Family Violence and Discord” and “Suicide Around Individual”. In fact, there were conceptual overlaps in many clusters, especially those pertaining to depression, suicide ideation, psychological disorders, and self-harm.

In addition to finding more general topics, in some cases, the informal latent topics are more specific than the expertly derived risk factors. A good example of this is the risk factor “Drug Abuse” and the related informal topics. The first cluster represents the notion of “pills” and “sleeping pills”. The second cluster represents the notion of “medication” and “meds”. The third cluster represents the notion of “alcohol”, ‘drinking”, and recreational drugs such as “weed”. All of these clusters fit well under the heading of “Drug Abuse”, but vary significantly in their focus. The nuances in the discussions of drug abuse in online social media appear to result in topics capturing differing dimensions of this risk factor.

We also occasionally didn’t find clusters associated with risk factors. Despite “Gun Ownership” previously being identified as an important risk factor[22], we were unable to find a cluster which explicitly represented the idea of owning a gun. We did find the word, “gun”, in our clusters as well as many related words such as “shoot”. However, these words are clustered with terms related to suicidal thoughts rather than ownership. This example highlights one of the main differences between the expertly derived risk factors and the informal latent topics extracted from social media. While it may be true that those who have access to a gun are at greater risk to commit suicide, it does not appear that those who express suicide ideation online reference their ownership of a gun with as much clarity as they discuss other topics.

Some clusters were particularly difficult to classify. The clusters corresponding to “Depressive Feelings” and “Depressive Symptoms” were difficult to differentiate. The Anxiety and Depression Association of America[3] lists symptoms of depression as, among other things, irritability, insomnia, fatigue, difficulty making decisions, persistent physical symptoms, and feelings of hopelessness, worthlessness, pessimism, and helplessness.

Many users discuss their depression, not as a dichotomy between feelings and symptoms, but instead use the words more casually. When assigning clusters to risk factors, we attempt to make a distinction between feelings and symptoms. Symptoms can be identified as physical ailments or development of disorders and conditions such as anxiety, sadness, and stress. Feelings tend to be a more nuanced description of one’s self and experience.

One cluster contains the words “no”, “any”, “real”, “future”, “real”, “experience”, “motivation”, “social”, “dreams”, “purpose”, and “plans”. We classified this cluster as depressive feelings because the words seem to indicate a lack of purpose and a sense of uselessness. On the other hand, the cluster which contains “these”, “thoughts”, “feelings”, “suicidal thoughts”, “emotions’ and “panic attacks” is more focused on symptoms of depression that one may face.

Regardless of whether or not a word is labeled as a symptom or as a feeling, our informal latent topics often capture very specific depressive language. One cluster contains “depressed”, “angry” and “upset” capturing common emotional keywords. Another cluster contains “chest”, “stomach” and “heavy” describing the physical reaction to stress. A third contains “into”, “fall” and “down” using the familiar imagery of downward movement when describing depression. Indeed, we found a total of nineteen clusters relevant to depressive feelings and symptoms, a few of which are presented in Table 2. The diversity and specificity of our informal latent

[3] www.adaa.org

Risk Factor	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Bullying (8)	hit (5372) turned (5016) broke (4541) throw (2145) beat (2132)*	against (3612) abuse (2218) involved (1606) behavior (923) rape (767)*	mad (2645) threatened (1054) fights (740) yelling (560)* bully (455)*	physically (3458) emotionally (2451) bullied (2375) treated (2183) raped (1625)*	started (21893)* high school (8846) hated (3034)* dropped out (1830)* bullying (1005)*
Depressive Feelings (10)	very (33446) depressed (20645) become (9027) angry (5752) extremely (4807)	feel (125439) am (114263) feeling (26390) alone (25203) sad (10933)	no (109179) any (50721) real (12766) future (10001) experience (5321)	happiness (5251) sense (5040) sort (4433) lack (3606) desire (2629)	into (32556) fall (22057) down (3249) slowly (3151) deep (2733)
Depressive Symptoms (9)	these (16523) thoughts (13084) feelings (7829) suicidal thoughts (5999) emotions (3258)	day (36329) days (14379) sleep (13630) hours (10181) cry (7077)	room (8007) crying (7023) cried (2405) tears (2356)* screaming (1551)*	pain (21604) fear (6492) constant (3805) sadness (2810) guilt (2749)	chest (2843) stomach (1543) heavy (1047) panic attack (989) panic (983)
Drug Abuse (3)	pills (6482) bottle (2204) overdose (1451) sleeping pills (1036) xanax (729)*	medication (5829) meds (4811) medicine (1537) antidepressants (1432) mg (1211)	drugs (5581) drunk (4560) drinking (3660) alcohol (3006) weed (1423)*		
Family Violence and Discord (1)	father (9207) kids (6564) child (5125) abusive (2063) divorce (1508)*				
Gun Ownership (0)					
Impulsivity (0)					
Prior Suicide Attempt (3)	since (32577) past (17510) suicidal (15001) times (10523)* attempt (4238)*	hospital (8014) admitted (1339) er (1121) hospitalized (1049)* committed (792)*			
Psychological Disorders (6)	depression (26860) anxiety (10311) diagnosed (3905) bipolar (1964) social anxiety (1810)	therapist (6517) doctor (5205) psychiatrist (3084) treatment (1861) mental health (1680)	problems (12848) due (7429) issues (5561) stress (4894) emotional (2733)	tried (30016) therapy (6619) doctors (2947) several (2791) medications (1264)	results (836) combination (513) ect (461)‡ levels (455) hormones (422)
Self-Harm (2)	cut (7379) cutting (3095) knife (2143) wrists (1392) scars (1016)*	body (7210) heart (5707) blood (1618)* burn (697)* tear (562)*			
Suicide Around Individual (1)	family (41145)† parents (30890) kill herself (1327)* kill himself (1064)* committed suicide (971)*				
Suicide Ideation (8)	thought (30326) suicide (25416) thinking (20582) mind (15745) killing myself (8223)	hanging (3298) hang (2736) rope (1836) neck (1598) noose (811)	head (13150) gun (4240) hand (3929) pull (2406) trigger (1577)	edge (2508) near (2191) jump (2177) bridge (1764) building (1684)	plan (6527) easy (5018) option (2907) method (1576) quick (1559)
Other Important (23)	stupid (11266) such (11259) man (6788) failure (6169) selfish (5821)	money (18076) pay (6742) debt (3360) cant afford (2526) rent (2136)	died (5854) cancer (1974) killed (1449) disease (995) brain damage (216)*	nice (7526) perfect (3430) beautiful (3388) strong (3303)* smart (2577)*	girl (15934) relationship (11081) guy (10570) loved (8486)* broke up (4120)*
Accessory (26)	eat (4649) food (3724) buy (2794) water (1440) pack (522)*	internet (3720) music (3034) watch (2706) computer (2481) game (2426)	bought (1799) clothes (1348) bag (687) table (683) laptop (646)	years (41839) year (22996) two (17720) months (13798) week (13076)	phone (4797) online (4081) text (3848) contact (3480) message (2318)

Table 2: Notable word clusters representing informal latent topics extracted from posts to r/SuicideWatch. Clusters are manually labeled according to risk factors proposed by experts. Gun Ownership and Impulsivity intentionally left blank, as none of the discovered clusters matched those risk factors. Additionally, other clusters were identified that held relevant semantic meaning, but didn't fit into the risk factors. These were included under the label Accessory.

Number next to risk factor indicates number of clusters with its label
Number next to word indicates frequency of occurrence in corpus
* : Term is not in top 5 terms of cluster
† : This cluster includes terms which may be labeled under 'Family Violence and Discord'
‡ : 'ect' is an acronym for 'electroconvulsive therapy'

topics seem to capture subtle differences in how users discuss suicide in online posts.

We found other clusters which we could not label according to the twelve suicide risk factors, and which we accordingly labeled "Other Important". These clusters were identified as possible contributors to suicide ideation, and contain information which we determined may be valuable to identify and assess suicide risk in social media posts. For example, one cluster includes "stupid", "failure" and "selfish". Authors of the posts often use these words to describe their self-image. Another cluster includes "died", "cancer", and "disease" presenting the occurrence of a serious medical condition in the user's life.

In all, we identified 22 important clusters that did not fit well into the 12 previously proposed risk factors, many representing stressors that might lead to suicidal ideation. Other topics include poor performance in school, trouble with money, and disgust with one's physical appearance.

The complexity of natural language often made it challenging to categorize the informal topics. For example, positive words are sometimes used to express negative feelings. A cluster containing mainly positive tokens such as "nice", "beautiful", "perfect", "strong", and "smart" may be referencing legitimately positive characteristics. On the other hand, a user might be posting about how good the life of other people seems to be while their life is lacking. Examples of these sentiments from posts are, "My family acts so perfect and seems so perfect from the outside" and "Why is everyone else so beautiful?".

Finally, many of the latent topics did not seem immediately relevant to suicidal ideation but were often present in the online posts. Five of these clusters are shown in the last row of Table 2. For example, one cluster represents the notion of food while another represents clothes.

5.5 Discussion

To evaluate our models, we first subjectively evaluated the latent topics represented by clusters of words. We then compared these topics to risk factors generated by domain experts. Our in-depth analysis revealed several key findings.

First, we found that the topics discovered by our analysis had a large scope. Topics ranged from crying to clothing to the calendar. This illustrates that our model was able to identify different latent topics within the corpus and separate them into meaningful clusters. It also shows that there are topics that people discuss which are not directly related to suicide, as not every word is on the topic of suicide.

When comparing our automatically generated topics to previously identified risk factors, we found that there were some differences in the focus of the topics compared to that of the risk factors. In the case of "Drug Abuse", people tended to discuss recreational drugs, specifically alcohol, separately from medications and pills. This difference in focus shows how the public view of these two topics may fit under the umbrella term provided by experts, but differ enough to be separate topics. On the other hand, in the case of "Family Violence and Discord" and "Suicide Around Individual", the topics generated by our model seemed to indicate a broader topic, rather than topics as specific as these risk factors.

A result of collecting data from public users with presumably no professional medical experience is the difference in precision of language between users and medical professionals. An indicator of this difference is in discussing depression. While professionals made a difference between "Depressive Feelings" and "Depressive Symptoms", the topics identified from users' posts overlapped these ideas. This may be partly due to the fact that depressive feelings are a symptom of depression, but also to a lack in precise use of language to describe specific experiences and symptoms.

Our contribution to this field is the discovery of latent topics within textual data known to contain suicidal ideation. A common method for identifying suicidal ideation in social media is to use a filter designed by medical professionals to extract data. Such a technique may impose a structure on the data by medical professionals that does not reflect the actual language used by those experiencing suicidal ideation. Our method uses topic modeling to uncover informal, latent topics directly from social media posts, which captures the ideas deemed important by those who are sharing their experiences with an online community. This information will inform the medical community which informal topics are important to monitor in informal contexts, such as social media, to effectively identify suicidal ideation.

6. CONCLUSION

In this work, we automatically extracted informal latent topics from online social media expressing suicidal ideations. We first subjectively evaluated the latent topics and then exhaustively compared them to risk factors proposed by domain experts. In general, we found that our informal topics are similar to the expert's risk factors; however, our topics differ in several important ways. Our topics can be more specific or more general. Some of our topics express meaningful ideas not contained in the risk factors and some risk factors do not have complimentary latent topics. In short, our analysis of the latent topics extracted from social media containing suicidal ideations suggests that users of these systems express ideas that are complementary to the topics defined by experts but differ in their scope, focus, and precision of their language.

This effort opens up many possibilities for future work. First, we will build models leveraging the informal topics to predict the urgency of the posts. Second, we plan to compare these results to other topic modeling algorithms such as latent Dirichlet analysis and latent semantic analysis. Finally, we will extend our analysis to other mental health issues such as post-traumatic stress disorder and depression.

Acknowledgment

This work is made possible in part by the National Science Foundation under the Research Experience for Undergraduates grant number 1659836. We would like to thank Jason Baumgartner of www.pushshift.io for maintaining and sharing archives of Reddit, which helped make this analysis possible. Finally, we also wish to express our sincerest gratitude for the moderators of r/SuicideWatch who perform an extraordinary service for their community.

7. REFERENCES

[1] A. P. Association et al. *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub, 2013.

[2] P. Bamidis, G. Coppersmith, B. Spitzberg, L. LItman, M.-H. Tsou, S. Konstantinidis, S. R. Braithwaite,

C. Giraud-Carrier, J. West, M. D. Barnes, and C. L. Hanson. Validating machine learning algorithms for twitter data against established measures of suicidality. *Journal of Medical Internet Research Mental Health*, 3(2), 2016.

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

[4] L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. VanderPlas, A. Joly, B. Holt, and G. Varoquaux. API design for machine learning software: experiences from the scikit-learn project. In *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, pages 108–122, 2013.

[5] G. Coppersmith, M. Dredze, and C. Harman. Quantifying mental health signals in twitter. In *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pages 51–60, Baltimore, Maryland, USA, June 2014. Association for Computational Linguistics.

[6] G. Coppersmith, R. Leary, E. Whyne, and T. Wood. Quantifying suicidal ideation via language usage on social media. In *Joint Statistics Meetings Proceedings, Statistical Computing Section*. JSM, 2015.

[7] G. Coppersmith, K. Ngo, R. Leary, and A. Wood. Exploratory analysis of social media prior to a suicide attempt. In *Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology*, pages 106–117, San Diego, CA, USA, June 2016. Association for Computational Linguistics.

[8] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray. Visual categorization with bags of keypoints. In *In Workshop on Statistical Learning in Computer Vision, ECCV*, pages 1–22, 2004.

[9] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz. Predicting depression via social media. In *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media*. AAAI, July 2013.

[10] M. De Choudhury, E. Kiciman, M. Dredze, G. Coppersmith, and M. Kumar. Discovering shifts to suicidal ideation from mental health content in social media. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 2098–2110, New York, NY, USA, 2016. ACM.

[11] M. Duggan and A. Smith. 6% of online adults are reddit users. *Pew Internet & American Life Project*, 3:1–10, 2013.

[12] G. Gkotsis, A. Oellrich, T. Hubbard, R. Dobson, M. Liakata, S. Velupillai, and R. Dutta. The language of mental health problems in social media. In *The Third Computational Linguistics and Clinical Psychology Workshop (CLPsych)*, pages 63–73, 6 2016.

[13] J. A. Hartigan and M. A. Wong. Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1):100–108, 1979.

[14] K. Hawton and K. van Heeringen. Suicide. *The Lancet*, 373(9672):1372 – 1381, 2009.

[15] T. Hofmann. Probabilistic latent semantic analysis. In *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, pages 289–296. Morgan Kaufmann Publishers Inc., 1999.

[16] L. Hong and B. D. Davison. Empirical study of topic modeling in twitter. In *Proceedings of the first workshop on social media analytics*, pages 80–88. ACM, 2010.

[17] J. Jashinsky, S. H. Burton, C. Hanson, and T. Argyle. Tracking suicide risk factors through twitter in the US. *Crisis The Journal of Crisis Intervention and Suicide Prevention*, 35(1):1–9, 2013.

[18] M. Kumar, M. Dredze, G. Coppersmith, and M. De Choudhury. Detecting changes in suicide content manifested in social media following celebrity suicides. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media*, pages 85–94. ACM, 2015.

[19] P. M. Lewinsohn, P. Rohde, and J. R. Seeley. Psychosocial risk factors for future adolescent suicide attempts. *Journal of Consulting and Clinical Psychology*, 62(2):297–305, 1994.

[20] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

[21] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.

[22] Suicide in the U.S.: Statistics and prevention, 2012.

[23] M. K. Nock, G. Borges, E. J. Bromet, J. Alonso, M. Angermeyer, A. Beautrais, R. Bruffaerts, W. T. Chiu, G. De Girolamo, S. Gluzman, et al. Cross-national prevalence and risk factors for suicidal ideation, plans and attempts. *The British Journal of Psychiatry*, 192(2):98–105, 2008.

[24] B. O'Dea, S. Wan, P. Batterham, A. Calear, C. Paris, and H. Christensen. Detecting suicidality on twitter. *Internet Interventions*, 2(2):183–188, 2015.

[25] J. W. Pennebaker, M. E. Francis, and R. J. Booth. Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates*, 71(2001):2001, 2001.

[26] R. Řehůřek and P. Sojka. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50, Valletta, Malta, May 2010. ELRA. http://is.muni.cz/publication/884893/en.

[27] G. Salton, A. Wong, and C. S. Yang. A vector space model for automatic indexing. *Commun. ACM*, 18(11):613–620, Nov. 1975.

[28] L. Wolf, Y. Hanani, K. Bar, and N. Dershowitz. Joint word2vec networks for bilingual semantic representations. *Int. J. Comput. Linguistics Appl.*, 5(1):27–42, 2014.

[29] G. Zanotti, M. Horvath, L. N. Barbosa, V. T. K. G. Immedisetty, and J. Gemmell. Infusing collaborative recommenders with distributed representations. In *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems*, pages 35–42. ACM, 2016.