Knowledge-aware Assessment of Severity of Suicide Risk for Early Intervention

Authors: Manas Gaur, Amanuel Alambo, Joy Prakash Sain, Ugur Kursuncu, Krishnaprasad Thirunarayan, Ramakanth Kavuluru, Amit Sheth, Randy Welton, Jyotishman Pathak

Supervisor:
Dr. Uban Ana-Sabina

Reviewer:
Iordache Adrian-Răzvan

Context & Contributions

In our current society, a problem that grows significantly year by year is the increased number of deaths caused by suicide.

Some would argue that social media played a major role in this situation, but maybe from this possible cause of the problem, we can find a solution.

This paper is meant to represent an analysis of the problem, and the obtained results over a collected dataset composed of posts from various mental conditions subreddits, based on 500 users.

Structure of the paper

In this paper we can distinguish two major stages:

1. The presentation of a "golden standard" methodology for data collection, selection and annotation.

2. The descriprion of various machine learning and deep learning approaches for the proposed problem.

A golden standard dataset of 500 suicidal redditors

01

Data Selection Metodology

01 Statistical Vocabulary Creation

From r/SuicideWatch (93000 users) based on statistical methods we generate a vocabulary of suicide terms which will be later curated by domain experts. This will result in 19000 users.

02 Negation Detection

We use negation detection to reduce possible false positives, probably by removing users that have post with negated sentences that can confound a classifier. After this step, the dataset will contain only 2181 users.

03 User and Content Overlap

We use those two procedures to gain a better understanding of the problem and possible relations with other mental conditions. And also to enrich the context for each user.

User and Content Overlap

This analysis provides deeper insight into how potentially suicidal users communicate on problems including causes, symptoms, and treatment solutions.

User Overlap: Through user overlap we infer the population level similarity between a mental health subreddit and SuicideWatch. We calculated the user overlap through the intersection of the users in SuicideWatch and each other mental health subreddit.

Content Overlap: Content overlap was calculated using a cosine similarity measure through domain-specific lexicon, LDA2Vec, and ConceptNet embeddings. Selecting only posts with a similarity level higher than a threshold of 60%.

User and Content Overlap Results

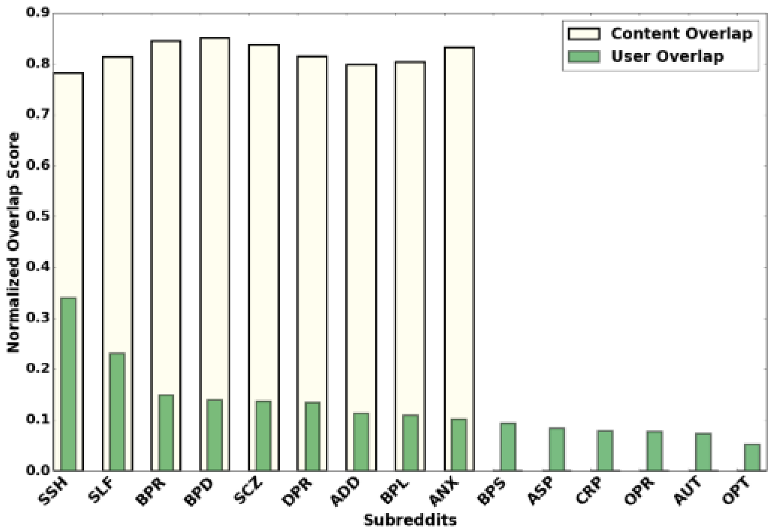

Figure 4: User Overlap and Content Overlap based quantification of influence of other mental health related subreddit to SW. Subreddits *SSH* and *SLF* have the highest content overlap with SW followed by *BPR* and *BPD*.

r/StopSelfHarm SSH),
r/selfharm (SLF),
r/bipolar (BPL),
r/BipolarReddit (BPR),
r/opiates (OPT),
r/Anxiety (ANX),
r/addiction (ADD),
r/BPD (BPD),
r/SuicideWatch (SW),
r/schizophrenia (SCZ),
r/autism (AUT),
r/depression (DPR),
r/cripplingalcoholism (CRP),
r/aspergers (ASP)

Annotation Strategy

The label scheme for this porblem will contain 5 labels:
- **Supportive (SU)**
- **Suicide Indicator (IN)**
- **Suicidal Ideation (ID)**
- **Suicidal Behavior (BR)**
- **Actual Attempt (AT)**

In the final stage of dataset generation, 500 users were selected for annotation and evaluated by 4 practicing clinical psychiatrists using two methodologies for annotator agreement metrics.

Dataset Samples Examples

Table 5: Paraphrased posts from candidate suicidal redditors and associated suicide risk severity level

Always time for you to write your happy ending .... doesnt need to be spelled out with alcohol and Xanax.... keep an open mind	SU
Ive never really had a regular sleep schedule....no energy to hold a conversation....no focus on study....barely eat and sleep....fluffy puppy dog face	IN
Sometimes I literally cant bear to move....my depression....since I was 14....suffering rest of my life....only Death is reserved for me.	ID
Driving a sharp thing over my nerve. Extreme depression and loneliness.... worthless excuse for a life....used everything from wiring to knife blades	BR
I am going to off myself today...loaded gun to my head..determined....huge disappointment....screwed family life....breaks my heart everyday.	AT

Machine Learning and Deep Learning Experiments

02

Baseline and Experiments

This baseline is a rule-based model for classifying a user based on a strict and soft match criteria according to presence of a concept in the user's content and the suicide risk severity lexicon.

We categorize our experiments into three schemes:

Experiment 1: evaluates the performance of the models over 5 labels (supportive, indicator, ideation, behavior, and attempt)

Experiment 2: evaluates models performances over 4 labels in which supportive (or negative) samples are removed

Experiment 3: comprises labels defined according to 4-label categorization (where supportive and indicator classes are merged into one class : no-risk)

Preprocessing of the input

1. Further, for each experiment, the input data is of two forms:

2. (I1) Only textual features (TF) represented as vectors of 300 dimensions generated using ConceptNet embeddings

3. (I2) having Characteristics features (CF) and textual features (CF+TF)

Machine Learning and Deep Learning Models

5.3 Convolutional Neural Network

We have implemented a convolutional neural network (CNN) as proposed in [30] for our contextual classification task [53].

The model takes embeddings of user posts as input and classifies into one of the suicide risk severity levels. We combine embeddings of posts for each user through concatenation, and pass into the model.

$$posts_u = post_{u,1} \bigoplus post_{u,2} \bigoplus ..post_{u,p}.. \bigoplus post_{u,P} \tag{6}$$

$$post_{u,p} = \vec{v}_{u,p,1} \bigoplus \vec{v}_{u,p,2} \bigoplus ..\vec{v}_{u,p,w}.. \bigoplus \vec{v}_{u,p,W} \tag{7}$$

Here $\bigoplus$ represents the concatenation operation of P posts of user u, where each post p of user u ($post_{u,p}$) is the concatenation of vectors of each word w ($\vec{v}_{u,p,w}$) where W is the total number of words in a post. Embeddings of the posts for each user ($posts_u$) have variable length. Hence, we use minimum length padding to make the dimensions of the representations uniform. The model has a convolution layer with filter window {3, 4, 5} and 100 filters for each. After getting the convoluted features, we apply max-pooling and concatenate the representative pooled features. We pass the pooled features through a dropout layer with dropout probability of 0.3, followed by an output softmax layer. The learning rate was set to 0.001 with adam optimizer [31]. While training the model, we have used mini batch of size 4 and trained for 50 epochs. CNN's performance is compared and evaluated in Section 6.

For Machine Learning models, we consider five learning models:

- SVM with Radial Basis Function (SVM-RBF)
- SVM with Linear Kernel (SVM-L)
- Random Forest (RF)
- Feed-Forward Neural Network (FFNN)

On the Deep Learning side, the authors proposed a solution based Convolutional Neural Networks. A summarized description of the implementation can be seen in the left image.

5-Label Classification Results

Table 7: Experiment with 5-label Classification

Approach	Input	With Supportive Class			
		Graded Precision	Graded Recall	F-Score	OE
Baseline	text	0.56	0.36	0.44	0.38
SVM-RBF	I1	0.53	0.51	0.52	0.12
	I2	0.57	0.62	0.61	0.12
SVM-L	I1	0.60	0.45	0.52	0.12
	I2	0.77	0.40	0.53	0.09
RF	I1	0.68	0.49	0.57	0.19
	I2	0.62	0.45	0.52	0.11
FFNN	I1	0.45	0.59	0.51	0.15
	I2	0.52	0.63	0.57	0.12
CNN	I1	0.71	**0.60**	0.65	0.10
	I2	0.70	**0.59**	0.64	0.09

Figure 6: Confusion Matrix of 5-label scheme. (left) CNN, and (right) SVM-L. Y-Axis: True Level, X-Axis: Predicted Level

4-Label Classification Results

Table 8: Experiment with 4-label Classification

Approach	Input	Without Supportive Class			
		Graded Precision	Graded Recall	F-Score	OE
Baseline	text	0.43	0.57	0.49	0.20
SVM-RBF	I1	0.63	0.47	0.54	0.12
	I2	0.66	0.59	0.62	0.12
SVM-L	I1	0.62	0.53	0.57	0.12
	I2	0.68	0.57	0.61	0.09
RF	I1	0.67	0.41	0.51	0.22
	I2	0.64	0.47	0.54	0.18
FFNN	I1	0.63	0.58	0.60	0.15
	I2	0.67	0.62	0.64	0.12
CNN	I1	0.72	**0.59**	0.65	0.11
	I2	0.70	**0.57**	0.62	0.1

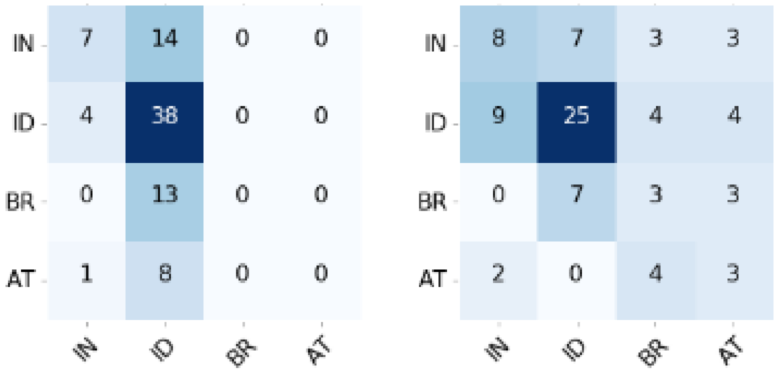

Figure 7: Confusion Matrix of 4-label scheme. (left) CNN, and (right) SVM-L

3+1 Label Classification Results

Table 9: Experiment with 3+1-label Classification

Approach	Input	Collapsed Supportive and Indicator Class			
		Graded Precision	Graded Recall	F-Score	OE
SVM-L	I1	0.81	0.54	0.65	0.12
	I2	0.74	0.54	0.63	**0.09**
CNN	I1	0.83	**0.57**	0.676	0.07
	I2	0.85	**0.57**	0.68	**0.06**

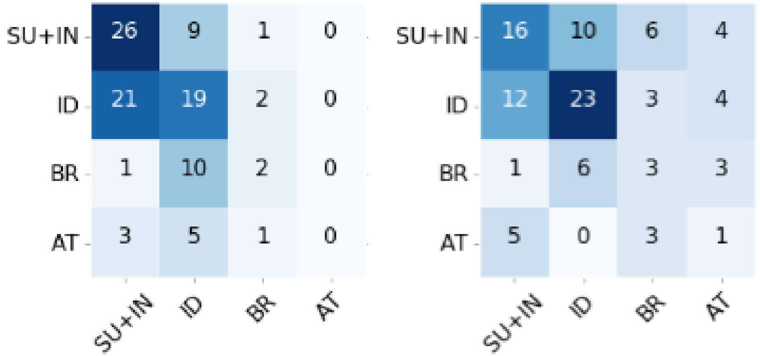

Figure 8: Confusion Matrix of (3+1)-label scheme. (left) CNN, and (right) SVM-L

Authors Conclusion

In this study, we presented an approach to predict severity of suicide risk of an individual using Reddit posts, which will allow medical health professionals to make more informed and timely decisions on diagnosis and treatment. A gold standard dataset of 500 suicidal redditors with varying severity of suicidal risk was developed using suicide risk severity lexicon. We then devised a 5-label classification scheme to differentiate non-suicidal users from suicidal ones, as well as suicidal users at different severity levels of suicide risk (e.g., ideation, behavior, attempt). Our 5-label classification scheme outperformed the two baselines. We specifically noted that CNN provided best performance among others including SVM and Random Forest. We make both the gold standard dataset and the suicide risk severity lexicon publicly available to the research community for further suicide-related research.

Strengths and Weaknesses

- Generating a gold standard dataset in a not so much explored domain
- Using Convolutional Neural Networks in an unconventional manner
- Various experiments based on multiple label schemes

- Using mostly standard machine learning algorithms
- Details about evaluation metrics not specified
- Missing samples in dataset
- Failed to reproduce

Personal Opinion

My suggestion for the authors would be to focus more on the details for reproducible results. To present more details about experiments, complete metrics details, types of stratification, more error analysis at word level and visualization level instead of simple metrics, etc.

In the first part of the paper, for the dataset generation part, the details are much well-written, but unfortunately, the uploaded dataset is corrupt and it’s such a shame because the current academic society could gain so much from it.

As a final opinion would be interesting using much compare transformer models with more advanced machine learning techniques

Thank you for your attention!

Any Questions?