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

Presented at: The World Wide Web Conference, May 2019, Pages 514–525, Research Article Link

Paper Review

Reviewer Name:	Iordache Adrian-Răzvan
Supervisor:	Dr. Uban Ana-Sabina
Faculty/University Centre:	Faculty of Mathematics and Computer Science, University of Bucharest
Specialization:	Artificial Intelligence, Group 507
Reviewed Paper Domain:	Biomedical Natural Language Processing
Current Date:	*March 30, 2022*

1 Core Review

Paper Summary: This paper represents an analysis of a very important problem concerning the future of society, the risk of suicide based on assumed mental conditions. Also, this paper provides obtained results over a collected dataset composed of posts from various mental conditions subreddits, based on 500 users.

The paper can be distinguished into two main stages. The first part of the paper is represented by the methodology for data collection, selection, and annotation. In the second stage, the authors of the paper describe experiments, evaluation metrics, and results obtained by various machine learning and deep learning models on the previously collected dataset.

Strengths: This topic represents a major interest and can bring a robust solution for possible cases of early intervention of suicide attempts. The main contribution of this paper is introducing a well-structured annotation strategy for dataset generation. Current literature lacks such a type of filtered datasets by a complex selection method. The dataset starts by using 270000 Reddit users with 8 million posts from 15 mental health subreddits, such as: r/autism, r/depression, r/schizophrenia, r/Anxiety, r/SuicideWatch, this represents a great approach (instead of using just users from r/SuicideWatch) because this allows us to study various relations between different type of mental health assumed conditions. To focus attention only on Reddit users from r/SuicideWatch (93000 users) based on statistical methods we generate a vocabulary of suicide terms which will be later curated by domain experts, resulting in 19000 users. After using negation detection for a reduction of possible false positives, we select the final 2181 users from the r/SuicideWatch users. For those final selected users was performed user and content overlap to enrich the context. Based on a similarity measure, their posts from other mental health subreddits were added to the post from the r/SuicideWatch subreddit. 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. This represents a selective method for dataset generation. Another strong plus of this paper is based on applying Convolutional Neural Networks on concatenated embeddings for each word in a post, this represents a less seen method for this kind of problem.

Weaknesses: Unfortunately, this paper seems to approach mainly standard Machine Learning algorithms, without trying to explore deeper and more complex possible solutions like voting weighted ensembles, gradient boosting, or stacking. On the Deep Learning side of the paper as well we can observe not a lot of experimental architectures, not considering any type of recurrence between posts or words, or some kind of positional embeddings added to the input embeddings. It could have been interesting to see how transformer models react to this kind o problem. Another issue here is the lack of information in the Evaluation Metrics subsection, we get the definition for the False Positives and False Negatives, but not for the graded precision or graded recall. And taking into consideration the fact that the standard definition for false positives/negatives is represented by a number and not a percentage (as in the current paper), we cannot substitute the false positives/negatives in the default definition of precision/recall. The final issue with this paper is that the presented dataset uploaded was structured in a format that caused some invalid posts in the dataset. So the dataset presented in this paper instead of having 15755 posts as specified, only has 9099 posts, representing only approx. 58% of the actual dataset.

Comments and Suggestions: 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.

---

2 Reasons to Accept

- Generating a gold standard dataset in a not so much explored domain
- Using Convolutional Neural Networks in an unconventional manner
- Creating an innovative evaluation metric for dealing difficult to unambiguously annotated datasets
- Various experiments based on multiple label schemes

---

3 Reasons to Reject

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

---

4 Recommendation

☐ Strong Accept: I would argue strongly for accepting this paper; 5.0

☐ Between possibly accept and strong accept; 4.5

☑ Possibly Accept: I would argue for accepting this paper; 4.0

☐ Between neutral and possibly accept; 3.5

☐ Neutral: I am unable to argue for accepting or rejecting this paper; 3.0

☐ Between possibly reject and neutral; 2.5

☐ Possibly Reject: The submission is weak and probably shouldn't be accepted, but there is some chance it should get in; 2.0

☐ Between reject and possibly reject; 1.5

☐ Reject: I would argue for rejecting this paper; 1.0

---

5 Expertise

Provide your expertise in the topic area of this paper.

☐ Expert

☑ Knowledgeable

☐ Passing Knowledge

☐ No Knowledge