# Utilizing Temporal Psycholinguistic Cues for Suicidal Intent Estimation

Puneet Mathur*[1], Ramit Sawhney*[2], Shivang Chopra[3], Maitree Leekha[3], and Rajiv Ratn Shah[4]

[1] University of Maryland, College Park, USA
[2] Netaji Subhas Institute of Technology, Delhi, India
[3] Delhi Technological University, Delhi
[4] MIDAS Labs, IIIT Delhi, India

**Abstract.** Temporal psycholinguistics can play a crucial role in studying expressions of suicidal intent on social media. Current methods are limited in their approach in leveraging contextual psychological cues from online user communities. This work embarks in a novel direction to explore historical activities of users and homophily networks formed between Twitter users for extracting suicidality trends. Empirical evidence proves the advantages of incorporating historical user profiling and temporal graph convolutional modeling for automated detection of suicidal connotations on Twitter.

## 1 Introduction

Suicidal ideation detection is a well studied problem in social media analysis. Various works have tried to identify linguistic patterns correlated with suicidality intent. Despite the sustained efforts from the community, most approaches ignore the psychological relevance of temporal characteristics of suicidal behaviour. Moreover, there has been limited explorations in the space of homophily networks to identify collusive depressive users. We hypothesize that the contextual information embedded in social media engagement and historical activities of users can lead to substantial improvements in automated identification of suicidal ideation. We look beyond linguistic cues into temporal signals throughout this work, with the help of a publicly available dataset given by [14] of 34,306 tweets on suicidality detection.

## 2 Related Work

### 2.1 Challenges on Social Media

The growth of social media websites hosts a number of challenges such as cyberbullying, suicide pacts, and radicalism that motivate suicidal behavior and

* equal contribution

impact the mental health of the users [10]. The associativity of suicide-related verbalizations on social media websites has been found to be strongly related to potential suicidal attempts. Prior studies show how suicidal intent declarations were significantly more assortative than chance, at times connected till 6 degrees of separation [5]. A patient's social media profile can help medical experts gain perspective into their mental health status and identify those at critical risk for suicide attempts [15]. The potential of technological interventions for suicidal risk assessment and mitigation needs to be explored in detail.

### 2.2 Text-based Approaches

Various works have been recently proposed with an objective of automating the detection of social media posts expressing suicide ideation using textual information [17,7,3]. [4] performed a semi-automated content-based analysis on a small number of tweets related to depression in order to derive certain qualitative insights into the behavior of users displaying suicidal behavior. Self-disclosure helps to facilitate psychological well being in individuals with mental illness [2]. Textual descriptions of social media disclosures have been extensively studied in the past [7]. [19] explored deep learning based supervised classifiers for suicidal ideation detection.

### 2.3 Psycho-linguistic Analysis

[13] used social graph based features and gained considerable improvement in the task of abuse detection. [16] performed a psycho-linguistic analysis of online users for a similar task. [1] tried to link users' psychological features such as personality traits including personalities, sentiment and emotion for cyberbulling and trolling. The contributions that we make in this work are different from previous efforts as there has been hardly any attempt to take a combined multi-faceted approach for solving the task of suicidal ideation in Twitter.

### 2.4 Signals from Temporal Data

Temporal graphs can capture the relationships in data with time so as to model new events and comparison to related entities and historical states [18]. [9] detected groups based on interesting features of the time-evolving networks. It studied several clustering frameworks for time-evolving networks for detecting group structure. [6] performed temporal sentiment analysis for early detection of cyberbulling and suicide ideation of a user through graph-based data mining approaches.

## 3 Methodology

The proposed methodology looks beyond text classifiers and leverages tweeting history of users as well as their social network communication patterns. User-based features were extracted from the historical tweeting activity and inter-user

interactions was modeled as a social graph. The methodology is two-fold consisting of historical signal modeling and temporal graph convolutional modeling.

### 3.1 Classification Network

In order to learn from the textual information available in the raw tweets, we trained a **BLSTM + Attention** network [20]. We train a BLSTM model with 100 LSTM units, dropout rate of 0.25 and a recurrent dropout rate of 0.2. The attention layer was followed by another dropout layer of 0.2. This was followed by two dense layers having 256 units and 2 units, respectively.

### 3.2 Temporal Modeling of Suicidal Tendency

**Motivation:** The idea of temporal modeling of suicidal tendencies is inspired by [11] with additions. According to [11], a representation for the historical activity can be formulated as a temporal weighting scheme $\phi_i$ which is a sum of two independent time varying functions of suicidality - ideation build-up $\lambda_i(t)$ and sinusoidal episodes $\mu_i(t)$. Extrapolating from this, we add a third independent time-varying function - white Gaussian noise $z_i(t)$. Let $\Delta t_i$ be the time offset from the original tweet and the temporal representation function $z$ be given by Equation 1.

$$z(u, H) = \sum_{h_i \in H} \phi_i(\Delta t) f(h_i) \tag{1}$$

**Suicidal Ideation Build-up**: Each user's historical tweets can be modeled as an exponential function in time given by Equation 2 where $\alpha$ and $\beta$ are hyper parameters tuned over training data.

$$\lambda_i(\Delta t) = \alpha e^{\beta \Delta t_i} \tag{2}$$

**Suicidal Episodes**: Phased changes in suicidal intent are mathematically represented by Equation 3. As per [12], the hyper parameters for the same are given by Table 1.

$$\mu_i(\Delta t) = \sum_{1}^{Q} (a_q cos(\frac{2\pi q \Delta t_i}{U}) + b_q sin(\frac{2\pi q \Delta t_i}{U})) \tag{3}$$

| **Hyperparameters** | **Value** |
|---|---|
| Q | 3 |
| U | {1,2,3,4,5,6,7} |
| $a_q$ , $b_q$ | $\approx \eta(0, \sigma^2)$ |

Table 1: Hyper parameters for Equation 3 [11]

**Temporal surprise**: Similar to any channel medium, social media platforms are prone to noise that adds randomness to the temporal suicidal patterns. The white Gaussian noise is modeled as being derived from a normal distribution with the expectation value of the noise term $\zeta_(t)$ equal to 1.

$$\phi_i(\Delta t) = \lambda_i(\Delta(t) + \mu_i(\Delta t) + \zeta_i(\Delta t) \tag{4}$$

For each of the tweet samples, the historical activity representation was an input to logistic regression model to learn temporal embeddings from these features which was used as an input to the final model.

### 3.3 Graph Convolutional Networks for User Profiling

Learning user representations can be significantly enriched by leveraging information derived from the inter-user interactions in social media channels. For this purpose, Graph Convolutional Networks (GCN) [8] can be effectively utilized that are capable of modeling social interactions in the form of features of nodes in the graph and allow contextual learning of information with respect to a node's neighbourhood.

**Temporal GCN**: We tried to incorporate the historical views into the extended graph by constructing time weighted TF-IDF vectors of the historical tweets. The author nodes were modified to consist of temporal weighting of TF-IDF representation of tweets. Let the TF-IDF vector $f_k^t$ of tweet at timestamp $t$ for $k^{th}$ author be defined by Equation 5, where $C_k$ is the global noise parameter, $\hbar$ controls the margin of influence of a user on its neighbours social activity and $\omega$ is the rate of decay of the suicidal sentiment. The external parameters $C_k$, $\hbar_k$ and $\omega_k$ are learnt from the training portion of the dataset in an unsupervised fashion.

$$f^t = \hbar_k \exp^{\omega_k \Delta t} \tag{5}$$

## 4 Experiments

### 4.1 Data Description and Setup

To gauge the effectiveness of our proposed approach, we use the dataset from SNAP-BATNET[14] which consists of $34,306$ tweets with $3,984$ of them suicidal ideations. For each of these users, the tweet timelines were also collected to create the set of historical tweets. 10-fold stratified cross-validation was employed to evaluate models on each of the 10 train-val splits. The hyper parameters for the temporal weighted combination were tuned using a grid search over the grid $\alpha = \{0.1, 0.5, 1.0\}$, $\beta = \{0, 0.01, 0.1, 1\}$, $U = \{1, 2..., 7\}$ yielding $\alpha = 0.5, \beta = 1, U = 7$. $t_0$ was assigned to time series points with values equal to $argmax(|\mu|)_i$.

| Model | F1 | P | R |
|---|---|---|---|
| **SNAP-BATNET [14]** | 92.60 | 72.20 | 93.52 |
| **BiLSTM + Attention (Text) [11]** | 91.26 | 70.02 | 91.23 |
| **Text + Temporal Modeling** | 92.75* | 91.98* | 93.70* |
| **Temporal GCN** | **93.89*** | 88.73 | **94.54*** |

Table 2: Performance Analysis

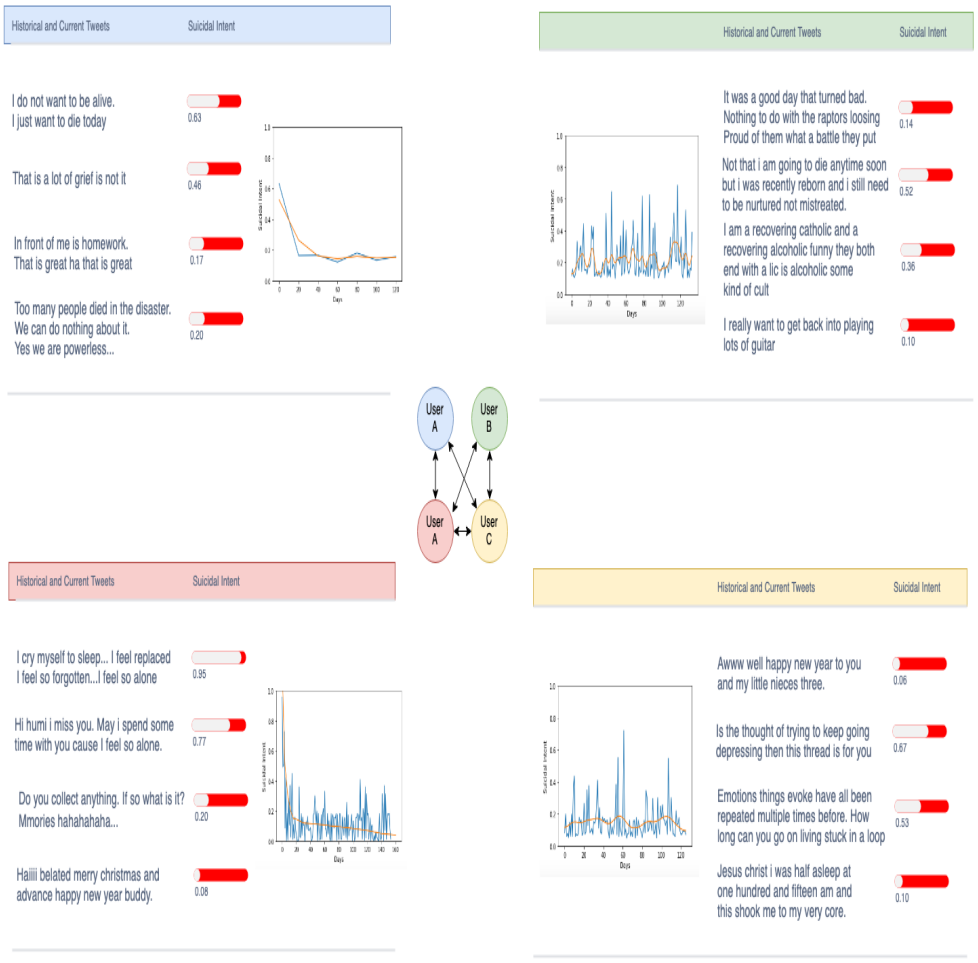


Fig. 1: Analysis of historical behaviour of users in a community over time

## 5 Results and Ablation Analysis

The ablation study of experimented features presented in Table 2 highlights the significance of temporal features extracted from social media in suicide ideation risk assessment. Temporal GCN provides a substantial gain over text in prediction confidence due to the user interactions. Additionally, it is interesting to observe the ability of the GCN model to better represent historical suicidal signals in comparison to naive historical and textual features to a sufficient degree. Empirically, temporal features help suppress false positives induced by text classifiers that try to overfit on the presence of anecdotal suicidal phrases such as "*kill me...hahaha !!*" that may be considered as noise in non-suicidal text. The most optimal weights for temporal signal modeling **Text + Builtup + Episodic + Surprise** were derived to be 0.52, 0.04, 0.04 and 0.32 through cross-validation experiments.

Figure 1 elucidates the impact of including psychological contextual cues on a small sample of connected users from the test dataset. It is evident from the historic trends of Users B and C that they follow a nearly episodic nature with scattered surprises. Analysing the trend plots for Users A and D reveals an inverse build-up thereby demonstrating that there can be either a positive or negative build-up in the suicidal intent of users. All these aspects when captured by our model has led to a statistically significant increase in the model's performance.

## 6 Conclusion

In spite of high importance of suicidal ideation identification on social media, little research has focused on looking beyond linguistic patterns. Through our work, we demonstrate that user interactions and past user behaviour are strong indicators of a potentially concerning mental state of online users. In this study, employing both qualitative and quantitative methods, we address this gap by investigating the impact of augmenting text based suicidal ideation detection models with contextual cues based on historical tweeting behavior and social media engagement.

## References

1. Balakrishnan, V., Khan, S., Arabnia, H.R.: Improving cyberbullying detection using twitter users' psychological features and machine learning. Computers & Security p. 101710 (2020)
2. Balani, S., De Choudhury, M.: Detecting and characterizing mental health related self-disclosure in social media. In: Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems. pp. 1373–1378. ACM (2015)
3. Benton, A., Mitchell, M., Hovy, D.: Multi-task learning for mental health using social media text. arXiv preprint arXiv:1712.03538 (2017)

4. Cavazos-Rehg, P.A., Krauss, M.J., Sowles, S., Connolly, S., Rosas, C., Bharadwaj, M., Bierut, L.J.: A content analysis of depression-related tweets. Computers in human behavior **54**, 351–357 (2016)
5. Cero, I., Witte, T.K.: Assortativity of suicide-related posting on social media. American Psychologist (2019)
6. Chatterjee, A., Das, A.: Temporal sentiment analysis of the data from social media to early detection of cyberbullicide ideation of a victim by using graph-based approach and data mining tools. In: Intelligence Enabled Research, pp. 107–112. Springer (2020)
7. De Choudhury, M., Gamon, M., Counts, S., Horvitz, E.: Predicting depression via social media. In: Seventh international AAAI conference on weblogs and social media (2013)
8. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)
9. Lee, K.H., Xue, L., Hunter, D.R.: Model-based clustering of time-evolving networks through temporal exponential-family random graph models. Journal of Multivariate Analysis **175**, 104540 (2020)
10. Lopez-Castroman, J., Moulahi, B., Azé, J., Bringay, S., Deninotti, J., Guillaume, S., Baca-Garcia, E.: Mining social networks to improve suicide prevention: A scoping review. Journal of neuroscience research (2019)
11. Mathur, P., Sawhney, R., Shah, R.R.: Suicide risk assessment via temporal psycholinguistic modeling (student abstract). In: Proceedings of the 34th AAAI Conference on Artificial Intelligence 2020. AAAI (2020)
12. Mathur, P., Shah, R., Sawhney, R., Mahata, D.: Detecting offensive tweets in hindi-english code-switched language. In: Proceedings of the Sixth International Workshop on Natural Language Processing for Social Media. pp. 18–26 (2018)
13. Mishra, P., Del Tredici, M., Yannakoudakis, H., Shutova, E.: Author profiling for abuse detection. In: Proceedings of the 27th International Conference on Computational Linguistics. pp. 1088–1098 (2018)
14. Mishra, R., S.P.S.R.M.D.M.P., Shah, R.: Snap-batnet: Cascading author profiling and social network graphs for suicide ideation detection on social media. Proceedings of the 2019 NAACL Student Research Workshop (pp. 147-156). (2019)
15. Pourmand, A., Roberson, J., Caggiula, A., Monsalve, N., Rahimi, M., Torres-Llenza, V.: Social media and suicide: a review of technology-based epidemiology and risk assessment. Telemedicine and e-Health **25**(10), 880–888 (2019)
16. Qian, J., ElSherief, M., Belding, E.M., Wang, W.Y.: Leveraging intra-user and inter-user representation learning for automated hate speech detection. arXiv preprint arXiv:1804.03124 (2018)
17. Sawhney, R., Manchanda, P., Mathur, P., Shah, R., Singh, R.: Exploring and learning suicidal ideation connotations on social media with deep learning. In: Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. pp. 167–175 (2018)
18. Steer, B., Cuadrado, F., Clegg, R.: Raphtory: Streaming analysis of distributed temporal graphs. Future Generation Computer Systems **102**, 453–464 (2020)
19. Tadesse, M.M., Lin, H., Xu, B., Yang, L.: Detection of suicide ideation in social media forums using deep learning. Algorithms **13**(1), 7 (2020)
20. Zhou, P., Qi, Z., Zheng, S., Xu, J., Bao, H., Xu, B.: Text classification improved by integrating bidirectional lstm with two-dimensional max pooling. arXiv preprint arXiv:1611.06639 (2016)