Analysis and Metaphysics
ISSN: 1584-8574, e-ISSN: 2471-0849
Vol 20 (1), 2021
pp. 01–07

A Deep Attentive Multimoda Learning Approach for Diaster Indetification From Social Media Post

Arekatla Madhava Reddy, Motupalli Mallikarjuna Rao, Aremandla Sai Pujitha, Lankala Mounika
[1,2,3,4,] Assistant Professor
amreddy2008@gmail.com, motupalli mallikarjun 286@gmail.com, sapuja502@gmail.com, anji.amrexamcell@gmail.com
Department of CSE, A M REDY MEMORIAL COLLEGE OF ENGINEERING AND TECHNOLOGY, PETLUVARI PALEM, ANDHRA PRADESH-522601

ABSTRACT

In order to detect suicidal inclinations and kinds, this research details content analysis of text. The article goes on to detail the steps necessary to build a sentence classifier using a neural network trained using a variety of Python tools dedicated to machine learning. Teen suicide and online "groups of death" are major issues, and people are looking for solutions to put an end to suicide propaganda targeted at young people. Evaluation of previously published data pertaining to so-called "groups of death" and their online dissemination.

I. INTRODUCTION

When people talk about suicide on social media, the words they use change. There are a lot of people who are vulnerable who utilize online discussion groups to talk about their issues or find resources for things like this. The main purpose of our research is to showcase current efforts in the field of automated suicide post detection. We take a look at how to use classification techniques based on deep learning and machine learning to identify suicidal thoughts on Reddit early on. This is why we use a hybrid LSTM-CNN model to assess and contrast various categorization algorithms. Based on our experimental findings, the most effective approach for relevance classification is to combine neural network design with word embedding methods. In addition, our findings provide credence to the power and capability of deep learning architectures in developing a reliable model for evaluating the risk of suicide in different text categorization tasks. Suicide claims the lives of over 800,000 people annually. With an aggregate incidence of 10.5 per 100,000 persons, suicide continues to be the second worst killer of young people. By 2020, experts anticipate that the mortality rate will reach one every twenty seconds [1].

As a result of inadequate and frequently scant resources for detection and management, about 79% of suicides take place in low- and middle-income nations. An obsessive fixation with self-destruction, suicidal thoughts, or suicidal ideation may manifest in a variety of ways, including despair, suicidal planning, or both [2]. The behaviors of suicidal ideation, suicidal planning, and suicidal attempt completion are indicators of an individual's vulnerability [3]. On a regular basis, researchers debate the nature of the connection between these two groups. Some research suggests that the vast majority of people who contemplate suicide do not really try to end their lives in this way. Take Klonsky et al. [4] as an example. They argue that the often mentioned risk factors for suicide, such as melancholy, hopelessness, and frustration, actually predict suicidal thought rather than the actual act of attempting suicide. But research by Pompili et al. [5] shows that "several variables assumed to be risk factors for suicidal behavior" might be very comparable to suicidal ideators and those who have attempted suicide. Worldwide, efforts to cut suicide rates by 10% by 2020 have focused on developing and implementing national suicide prevention strategies, one of which is early diagnosis of suicidal thinking [1]. A lot of people's mental health, especially young people's, can be seen via the "window" that is social media these days. It allows users to join various online groups in an anonymous manner, opening up a forum for open dialogue on taboo subjects. Suicide notes are left by over 50% of those who accomplish suicide attempts and over 20% of those who attempt suicide in general [6].

Consequently, it is important to interrogate a person about the presence of their ideas whenever there is a written suicide indicator. Choudhury et al. [7] states that online notes, forum posts, blog entries, and tweets are often recorded in the present and are well maintained. It may reduce the likelihood of incorrect text interpretations caused by looking backwards, as compared to an offline text.

A new field of computational linguistics is developing around social media and the many online forums that deal with issues of mental health. In order to identify suicides and prevent them in the future, it offers a great platform for study into new technical techniques and enhancements [8]. As an intervention point, it may be useful. Researchers Kumar et al. [9] looked into how people who follow celebrity death news on Reddit's SuicideWatch subreddit post. To effectively reduce the number of suicides among prominent figures, he proposed a new strategy. The transition from discussions on mental health to those about suicidal thoughts on Reddit was investigated by Choudhury et al. [7]. To identify these changes, he created a statistical method based on propensity score matching. For the purpose of early suicide ideation identification, Ji et al. [10] have recently created a unique data protection system and sophisticated optimization approach (AvgDiffLDP). In addition to the more conventional techniques of text categorization, deep learning algorithms have achieved remarkable strides in computer vision and pattern recognition. Neural networks trained on dense vector representations outperform their more conventional machine learning counterparts on a range of natural language processing (NLP) tasks, which may be laborious and feature-poor [11]. Outperforming more conventional machine learning methods in suicide risk assessments, deep neural networks and word embedding [12,13] continue to gain traction. Sharing what we've learned about suicidal thoughts on Reddit forums via data analysis using effective deep learning architectures is the main goal of our work. We primarily want to investigate the feasibility of using LSTM, CNN, and a hybrid model for various classification tasks in the context of suicidal thoughts. We want to determine if language modeling and text classification performance may be enhanced by integrating CNN and LSTM classifiers into a single model. Our goal is to show that when it comes to themes connected to suicide, the LSTM-CNN model performs better than both the CNN and LSTM classifiers alone, as well as older, more conventional machine learning systems. Data sets from any online forum or blog might potentially have it incorporated. Prior to conducting the experiment, we establish the data source, formulate the model we want to use, and examine the baseline variables. We then find instances of suicidal ideation by calculating the dataset's gram frequency, which includes both unigrams and bigrams. Using the baseline and the model we developed, we assess the experimental strategy. Lastly, we find the optimal hyper-parameter selection for suicide ideation detection by training our LSTM-CNN model with 10-fold cross-validation. Our dataset is based on information retrieved from the social media site Reddit, where users are able to compose more extensive articles.

There are three main benefits to our study: Our evaluation of the n-gram analysis reveals that suicide-related forums often address the manifestations of suicidal inclinations and diminished social contacts. Various psychological phases, including increased inward emphasis as a sign of despair, frustration, anxiety, or isolation, are linked to the shift towards social ideation. Traditional feature analysis: we compare the efficacy of statistical features, bag-of-words, and TF-IDF with that of word embedding utilizing CNN, LSTM, and LSTM-CNN combined model analysis. In order to enhance the current state-of-the-art technique, we conduct a comparative assessment and look at how well the LSTM-CNN combined class of deep neural networks perform on tasks involving the identification of suicidal thoughts. We assess its capabilities and strengths on a real-world dataset by comparing it with CNN and LSTM deep learning methods as well as four conventional ML classifiers (SVM, NB, RF, and XGBoost).

II. LITERATURE SURVEY

1) Addressing Suicidal Thoughts and Behaviors in Substance Abuse Treatment: Information You Need To Know

Bryabrina T.V., Gibert A.I., and Shtrahova A.V. are the authors. Any drug addiction counselor working directly with clients should be prepared to talk about the possibility of suicide. Those who work as frontline counselors in drug abuse treatment programs, with clients who suffer from co-occurring mental health and substance misuse disorders, or who provide consulting or supervision to frontline counselors may find this chapter to be of particular interest. Counselors working with clients who display suicidal thoughts or actions might find useful information in this TIP, even if it is tailored to clients with a diagnosis of a drug use problem. There is a high rate of suicide ideation and attempt among people undergoing treatment for substance abuse (Ilgen, Harris, Moos, & Tiet, 2007) and a significantly higher rate of death by suicide among people who have previously been in treatment for substance abuse compared to those without a diagnosis (Wilcox et al., 2004). So, those who work in the field of drug misuse treatment need to be ready to regularly assess their clients, send them to appropriate resources, and even take part in their treatment if they are thinking about or have attempted suicide. In order to enhance the results of drug addiction therapy, it is important to investigate, identify, and treat any co-occurring illnesses that are manifest in suicidal thoughts and actions. These disorders include major depressive disorder, bipolar disorder, post-traumatic stress disorder (PTSD), schizophrenia, and certain personality disorders.

2)Scikit-learn: Machine Learning in Python

Pedregosa F., Varoquaux G., Gramfort A., and Michel V. are the authors. For both supervised and unsupervised problems of medium size, the scikit-learn Python package incorporates a diverse set of cutting-edge machine learning methods. Using a high-level, general-purpose language, this program aims to make machine learning accessible to those without specialized knowledge. Consistency in the API, documentation, performance, and simplicity of use are priorities. Academic and business institutions are encouraged to utilize it due to its minimum dependencies and distribution under the simplified BSD license. Python is quickly becoming a top choice for scientific computing among computer programmers. Dubois (2007) and Milmann and Avaizis (2011) both note that it is an attractive option for algorithmic creation and exploratory data analysis because to its high-level interactive character and its expanding ecosystem of scientific libraries. However, because to its versatility, it is finding more and more applications outside of academia, particularly in business. Using this robust environment, scikit-learn offers state-of-the-art implementations of several popular machine learning algorithms with a user-friendly interface that is deeply integrated with Python. This satisfies the increasing need for non-experts to analyze statistical data in domains outside of computer science, such biology and physics, and in the software and online sectors. In comparison to other Python machine learning toolboxes, scikit-learn has a number of notable differences: 1. It is distributed under the BSD license. 2. Unlike MDP and pybrain, it incorporates compiled code for efficiency. 3. Unlike pymvpa, which has optional dependencies like R and shogun, it only depends on numpy and scipy for easy distribution. 4. Unlike pybrain, which uses a data-flow framework, it focuses on imperative programming. The package is mostly developed in Python, but it includes the C++ libraries LibSVM and LibLinear, which provide reference implementations of support vector machines and generalized linear models, respectively, with comparable licensing (Chang and Lin, 2001; Fan et al., 2008). Windows and any POSIX platform are among the many possible platforms for binary packages.

3) National suicide prevention strategies: progress, examples and indicators

WHO is listed as one of the authors. For suicide prevention to be a political priority, there must be a national strategy. To further suicide prevention efforts, a national strategy and corresponding action plan are required. Suicide prevention will continue to get little attention until these are in place. The purpose of this publication is to provide policymakers and government officials with information they may use to create a national plan to prevent suicide. There is a wide range of methodologies and indicators used, as shown by examples from each WHO region. Presented below are steps to overcome typical obstacles and a description of the components needed to create, execute, and assess a nationwide suicide prevention plan.

III. EXISTING SYSTEM:

The current setup uses data collected from mixi, a popular SNS, to investigate the link between social networks and suicidal thoughts. Very few restrictions are addressed in this technique. At the outset, there is a fully functional social network of users; a connection between any two users signifies an open and mutually supported relationship. Like other social networks, some individuals have a quite big number of friends. Secondly, we are able to precisely determine the amount of triangles for every user due to the same rationale. Because registration is free for everyone, another advantage of this data collection is the reasonably diversified sample it contains.

DISADVANTAGES OF EXISTING SYSTEM:

User-defined community is a mixi function that is pertinent to this research.

Having your facts presented correctly is crucial. Therefore, you should not apply logistic regression until you have already determined all of the relevant independent variables.

Multivariate logistic regression is the algorithm in question.

IV. PROPOSED SYSTEM:

Data for the research comes from the websites of individuals who have either committed suicide or are contemplating suicide. The study's authors analyzed the text to determine the degree of suicidal ideation and expression. Using the data gathered, a tool called TextAnalyst investigates the reasons behind suicide thoughts and actions. Using a neural network, this research aims to categorize statements as either non-suicidal or suicidal. Finding out whether the text is suicidal, or how to overcome the challenge of its binary categorization, is essential in our system based on random text. Data distribution according to parameters is known as classification.

ADVANTAGES OF PROPOSED SYSTEM:

The dataset that was gathered has two columns: sentence and label. The former contains sentences, while the latter has the values suicidal and non-suicidal.

Words and symbols without any semantic weight may be solved using the NLTK package.

Sentences were converted into arrays using the Tokenizer function.

Method: Convolutional Neural Network

SYSTEM ARCHITECTURE:

V. IMPLEMENTATION

User:

As a first step, the user may register. In order to proceed with the registration, he needed to provide a valid email and cell phone number. The customer may be activated by the admin after they have registered. User access to our system is granted after the customer has been activated by the admin. The user may test a twitter message in this project once the CNN model is loaded. An first tweet will be sent by the user. We can foretell the outcomes using our model. Whether or whether the message was about suicide. We may also get its scores. We will construct the word-to-vector graph later. The database keeps track of each user's search results.

User: The user may log in using their credentials. He may activate users once he logs in. Our programs only allow the activated user to log in. There is data in the media folder.Tweets and labels are stored in the csv file. We may construct the neural network model using this. Which users tested which kind of twitter messages may be seen by the admin. All of those tweets will show up in your browser.

For the sake of this model, let's say that we're working on a system that can determine if a given tweet contains suicidal or non-suicidal language. Thus, we will have to forecast fresh textual tweets once the model is trained. To do this, the fresh tweets will need to undergo the same data preparation procedures as the model's training data.Before we do any data preparation, we will divide the dataset into a training set and a test set to make sure this restriction is part of our model assessment. This implies that we couldn't utilize any information from the test set to improve the data preparation (such as the terms used) for training the model.

Use white space to split tokens.

- Take words and remove any punctuation.
- Take out any words that aren't made up entirely of letters.
- Eliminate any stops words that may be present.
- Take out any term that is less than or equal to one character in length.

Results of the Prediction:

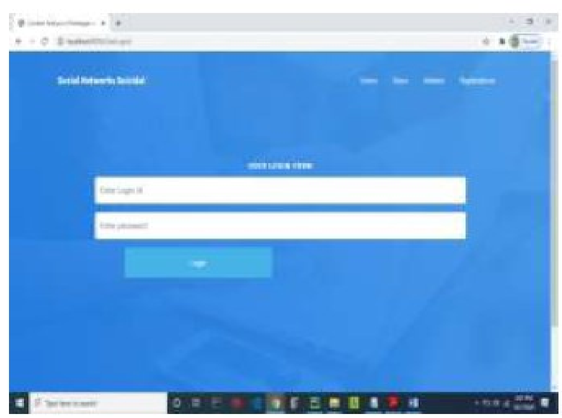
Because of their track record of effectiveness with document categorization challenges, we use a Convolutional Neural Network (CNN). We utilize a 32-filter, 8-kernel, rectified linear ('relu') activation function CNN setup, which is conservative and uses parallel fields to process words. After the convolutional layer, there is a pooling layer that halves the output. To represent the 'features' recovered by the CNN, the 2D output from the CNN section of the model is then flattened to one long 2D vector. In order to decipher the CNN characteristics, the model's back-end employs conventional Multilayer Perceptron layers. The review's positive and negative sentiments are produced as integers between 0 and 1 via the output layer's sigmoid activation function. Since this is an issue with binary classifications, we apply a binary cross entropy loss function. Ten iterations of the training data, or epochs, are used to train the model.

VI. SCREEN SHOTS

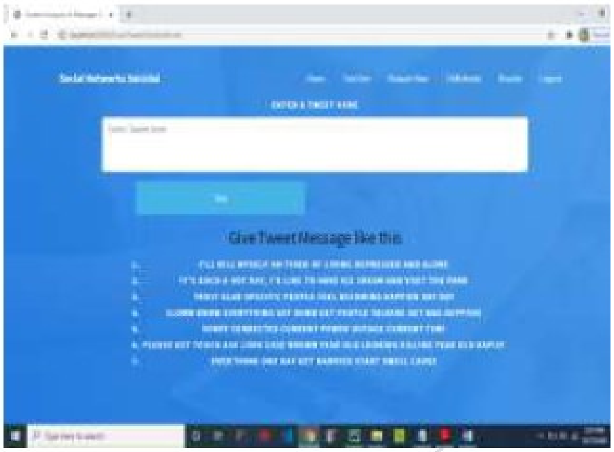
Home page:

User register form

User Login Form:



Test the message:



Result:

VII. CONCLUSION

The suicide rate is higher than the rates of murder, military activities, and traffic accidents combined, making this model unique among textual content analysis methods in its ability to categorize suicidal statements. It is worth mentioning that «death groups» were seen on several messengers and social networks, not only «Vkontakte». To stop the spread of suicide among young people who can't envision life without social media, this neural network can scan the language of social media postings for suicidal connotations and ban them. If a parent is concerned about their kid but does not want to pry into his private chats, they may utilize this application as a kind of parental control by simply collecting data and providing it to the program. It is also possible to utilize this neural network to filter out groups where there is an unusually high concentration of suicide-related messages.

Further Enhancement

Our study has certain limitations, such as a lack of data and potential bias in the annotations. One of the most pressing problems in modern research, which relies heavily on supervised learning methods, is data scarcity [86]. Typically, they need human annotation. Unfortunately, more study cannot be supported due to the lack of annotated data. Annotation bias, which arises from using established annotation criteria in hand labeling, is another concern. Occasionally, the annotation may cause labels to be biased, which can lead to false evidence that supports the authors' suicide act. Our research has the potential to inform future machine learning efforts aimed at developing a social media-based suicide detection and reporting system that can effectively connect people in crisis with mental health services.

REFERENCES

[1] World Health Organization. National Suicide Prevention Strategies: Progress, Examples and Indicators; World Health Organization: Geneva, Switzerland, 2018. [Google Scholar]

[2] Beck, A.T.; Kovacs, M.; Weissman, A. Hopelessness and suicidal behavior: An overview. JAMA 1975, 234, 1146–1149. [Google Scholar] [CrossRef] [PubMed]

[3] Silver, M.A.; Bohnert, M.; Beck, A.T.; Marcus, D. Relation of depression of attempted suicide and seriousness of intent. Arch. Gen. Psychiatry 1971, 25, 573–576. [Google Scholar] [CrossRef] [PubMed]

[4] Klonsky, E.D.; May, A.M. Differentiating suicide attempters from suicide ideators: A critical frontier for suicidology research. Suicide Life-Threat. Behav. 2014, 44, 1–5. [Google Scholar] [CrossRef] [PubMed]

[5] Pompili, M.; Innamorati, M.; Di Vittorio, C.; Sher, L.; Girardi, P.; Amore, M. Sociodemographic and clinical differences between suicide ideators and attempters: A study of mood disordered patients 50 years and older. Suicide Life-Threat. Behav. 2014, 44, 34–45. [Google Scholar] [CrossRef] [PubMed]

[6] DeJong, T.M.; Overholser, J.C.; Stockmeier, C.A. Apples to oranges?: A direct comparison between suicide attempters and suicide completers. J. Affect. Disord. 2010, 124, 90–97. [Google Scholar] [CrossRef] [PubMed]

[7] De Choudhury, M.; Kiciman, E.; Dredze, M.; Coppersmith, G.; Kumar, M. 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, San José, CA, USA, 9–12 December 2016; ACM: New York, NY, USA, 2016; pp. 2098–2110. [Google Scholar]

[8] Marks, M. Artificial Intelligence Based Suicide Prediction. Yale J. Health Policy Law Ethics 2019. Forthcoming. [Google Scholar]

[9] Kumar, M.; Dredze, M.; Coppersmith, G.; De Choudhury, M. Detecting changes in suicide content manifested in social media following celebrity suicides. In Proceedings of the 26th ACM conference on Hypertext & Social Media, Prague, Czech Republic, 4–7 July 2015; ACM: New York, NY, USA, 2015; pp. 85–94. [Google Scholar]

[10] Ji, S.; Long, G.; Pan, S.; Zhu, T.; Jiang, J.; Wang, S. Detecting Suicidal Ideation with Data Protection in Online Communities. In Proceedings of the International Conference on Database Systems for Advanced Applications, Chiang Mai, Thailand, 22–25 April 2019; Springer: Berlin, Germany, 2019; pp. 225–229. [Google Scholar]

[11] Yang, Y.; Zheng, L.; Zhang, J.; Cui, Q.; Li, Z.; Yu, P.S. TI-CNN: Convolutional neural networks for fake news detection. arXiv 2018, arXiv:1806.00749. [Google Scholar]

[12] Mikolov, T.; Karafiát, M.; Burget, L.; Černockỳ, J.; Khudanpur, S. Recurrent neural network based language model. In Proceedings of the Eleventh Annual Conference of the International Speech Communication Association, Makuhari, Chiba, Japan, 26–30 September 2010. [Google Scholar]

[13] Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.S.; Dean, J. Distributed representations of words and phrases and their compositionality. In Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, CA, USA, 5–10 December 2013; pp. 3111–3119. [Google Scholar]

[14] Coppersmith, G.; Ngo, K.; Leary, R.; Wood, A. Exploratory analysis of social media prior to a suicide attempt. In Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology, San Diego, CA, USA, 16 June 2016; pp. 106–117. [Google Scholar]

[15] Hsiung, R.C. A suicide in an online mental health support group: Reactions of the group members, administrative responses, and recommendations. CyberPsychol. Behav. 2007, 10, 495–500. [Google Scholar] [CrossRef]

[16] Jashinsky, J.; Burton, S.H.; Hanson, C.L.; West, J.; Giraud-Carrier, C.; Barnes, M.D.; Argyle, T. Tracking suicide risk factors through Twitter in the US. Crisis 2014, 35, 51–59. [Google Scholar] [CrossRef]

[17] Colombo, G.B.; Burnap, P.; Hodorog, A.; Scourfield, J. Analysing the connectivity and communication of suicidal users on twitter. Comput. Commun. 2016, 73, 291–300. [Google Scholar] [CrossRef]