DA-IICT, Gandhinagar, India

CAMS: An Annotated Corpus for Causal Analysis of Mental Health Issues in Social Media Posts

Kushal Barot[1], Aayush Vithalani[2], Harshil Shah[3] and Sobhan Behuria[4]

Abstract

The research community has witnessed substantial growth in the detection of mental health issues and their associated reasons from analysis of social media. We introduce a new dataset for Causal Analysis of Mental Health Issues in Social Media Posts (CAMS). Our contributions for causal analysis are two-fold: causal interpretation and causal categorization. We introduce an annotation schema for this task of causal analysis and demonstrate the efficacy of our schema on two datasets: (i) crawled and annotated 3155 Reddit posts and (ii) re-annotated publicly available SDCNL dataset of 1896 instances for interpretable causal analysis.

Keywords

Causal Analysis, Mental Health, Social Media, Machine Learning, Natural Language Processing, Clinical Depression, Clinical Psychology, Intent Classification, Suicidal Tendency

1. Introduction

This research project addresses the rising impact of social media on mental health disorders, including anxiety, depression, and suicide. Unlike prior studies focusing on sentiment analysis, it explores causal links between social media activities and mental health issues. Using advanced causal analysis, the project identifies key factors in social media content linked to disorders like suicidal ideation. The goal is to develop a computational model that assesses suicide risk severity from unstructured textual data, such as social media posts and narratives. This automated system aims to enable real-time risk assessment, facilitating timely interventions to reduce suicide rates.

2. Dataset

The dataset includes social media posts sourced from platforms such as Reddit, annotated with mental health-related labels. Data preprocessing involves noise removal, tokenization, and language normalization. A subset of posts with temporal and contextual metadata supports causal inference.

- **Reddit Corpus:** 3,362 posts were collected from the subreddit *r/depression* using the Python Reddit API Wrapper (PRAW). After expert cleaning, 3,155 valid samples remained.
- **SDCNL Dataset:** Contains 1,896 posts: 1,517 for training and 378 for testing, assumed to be pre-cleaned.
- **Combined Dataset:** The two corpora were merged to create the *CAMS* dataset, annotated by trained student annotators.
- Verification was conducted by consulting mental health practitioners, including a clinical psychologist and a rehabilitation counselor.

202318006@daiict.ac.in (K. Barot); 202318023@daiict.ac.in (A. Vithalani); 202318033@daiict.ac.in (H. Shah); 202318040@daiict.ac.in (S. Behuria)
202318006 (K. Barot); 202318023 (A. Vithalani); 202318033 (H. Shah); 202318040 (S. Behuria)

Class	Crawled corpus			SDCNL Training data			SDCNL Test data		
	Min	Max	Avg	Min	Max	Avg	Min	Max	Avg
No reason	1	508	59.78	1	1785	68.58	1	1562	84.85
Bias or Abuse	6	2109	347.48	5	4378	227.24	6	578	149.80
Jobs and career	13	2258	228.28	17	2771	255.70	20	1481	206.95
Medication	5	1552	213.83	3	3127	205.86	11	1124	165.60
Relationship	2	3877	229.35	14	2739	240.08	9	756	202.56
Alienation	3	1592	153.86	1	899	147.01	12	683	145.67

Figure 1: Dataset distribution across categories.

2.1. Imbalanced Data

The dataset shows an imbalance, with the largest number of samples in the *Relationship* and *Alienation* categories, reflecting societal challenges like loneliness and worthlessness. The *No reason* category has fewer posts due to dataset cleaning. Posts in the *Bias or abuse* category are notably fewer, with less than half the samples of the *Jobs and careers* and *Medication* categories.

Cause	CC	Train_S	Test_S	CAMS
No reason	292	332	70	694
Bias or abuse	122	194	35	351
Jobs/careers	399	181	48	628
Medication	410	170	43	623
Relationship	956	297	91	1344
Alienation	976	340	92	1408
Total	**3155**	**1517**	**379**	**5051**

3. Models

This project employs machine learning and deep learning models to classify causes of mental health issues in social media posts, leveraging pre-trained embeddings and advanced architectures for optimization.

3.1. Machine Learning Models

The machine learning models include Logistic Regression and Support Vector Machines (SVM). Logistic Regression, trained on TF-IDF features with regularization, performs multi-class classification, while SVM with an RBF kernel captures non-linear patterns using TF-IDF features.

3.2. Deep Learning Models

The deep learning models include Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Convolutional Neural Network (CNN). LSTM and GRU, designed for sequence learning, utilize GloVe embeddings, dropout regularization, and softmax outputs. CNN extracts local text patterns through convolutional and pooling layers, followed by a dense output layer.

3.3. Hybrid Models

Combines strengths of different architectures:

- **CNN+LSTM:** Merges CNN's feature extraction with LSTM's sequential learning.
- **CNN+GRU:** Combines CNN's pattern recognition with GRU's sequence modeling.

3.4. Bidirectional Models

Enhance context understanding by processing text in both directions:

- **Bi-LSTM:** Captures dependencies from both past and future contexts.
- **Bi-GRU:** Extends GRU with bidirectional processing.

4. Evaluation Metrics

Model performance was evaluated using Accuracy to measure overall correctness, Precision and Recall to assess positive classifications, and the F1-Score to balance Precision and Recall. The Confusion Matrix was also used to provide a breakdown of true positives, true negatives, false positives, and false negatives, helping to identify areas of misclassification.

5. Results

Model performance was evaluated using accuracy as the primary metric. A summary is provided in Table 1.

Table 1
Model Performance Summary

Model	Accuracy
Logistic Regression (LR)	0.4881
Support Vector Machine (SVM)	0.4826
Long Short-Term Memory (LSTM)	0.4405
Convolutional Neural Network (CNN)	0.4730
Gated Recurrent Unit (GRU)	0.3865
Bidirectional LSTM (Bi-LSTM)	0.4946
Bidirectional GRU (Bi-GRU)	0.4243
CNN + LSTM	0.4946
CNN + GRU	0.4054

6. Conclusion

This study explored machine learning and deep learning models to classify causes of mental health issues in social media posts. Results highlight the potential of combining traditional and advanced methods, with models like Bi-LSTM and CNN+LSTM delivering strong performance. The findings emphasize the role of automated causal analysis in mental health research and provide a foundation for future work on advanced architectures and real-time applications.

References

- CAMS: An Annotated Corpus for Causal Analysis of Mental Health Issues in Social Media Posts[1].
- CNN-GRU model based on attention mechanism[2].
- CNN and LSTM Networks[3].
- Bidirectional GRU on Papers with Code[4].
- Understanding BiLSTM on Medium[5].
- Handling Imbalanced Data: Challenges and Solutions[6].
- A Study on the Suicide vs Depression Dataset for Classification Tasks[7].

[1]Available at: http://www.lrec-conf.org/proceedings/lrec2022/pdf/2022.lrec-1.686.pdf
[2]Available at: https://www.frontiersin.org/journals/energy-research/articles/10.3389/fenrg.2023.1228256/full
[3]Available at: https://www.machinelearningmastery.com/cnn-long-short-term-memory-networks/
[4]Available at: https://paperswithcode.com/method/bigru
[5]Available at: https://medium.com/@anishnama20/understanding-bidirectional-lstm-for-sequential-data-processing-b83d6283befc
[6]Available at: https://www.turintech.ai/what-is-imbalanced-data-and-how-to-handle-it/
[7]Available at: https://paperswithcode.com/dataset/sdcnl-suicide-vs-depression-classification