

Utilizing Neural Network and Linguistic Metadata for Early Detection of Depression Indication in Text Sequence

Ms. M. Bavithra¹, Ms. A. Ramya², Ms. J. Sushma³, Ms. S. Vanitha⁴

Professor, Department of Master of Computer Applications¹

UG Students, Department of Information Technology^{2,3,4}

Dhanalakshmi Srinivasan Engineering College (Autonomous), Perambalur, Tamil Nadu

Abstract: *Depression is ranked as the largest contributor to global disability and is also a major reason for suicide. Still, many individuals suffering from forms of depression are not treated for various reasons. Previous studies have shown that depression also has an effect on language usage and that many depressed individuals use social media platforms or the internet in general to get information or discuss their problems. This paper addresses the early detection of depression using machine learning models based on messages on a social platform. In particular, a convolutional neural network based on different word embeddings is evaluated and compared to a classification based on user-level linguistic metadata. An ensemble of both approaches is shown to achieve state-of-the-art results in a current early detection task. Furthermore, the currently popular ERDE score as metric for early detection systems is examined in detail and its drawbacks in the context of shared tasks are illustrated. A slightly modified metric is proposed and compared to the original score. Finally, a new word embedding was trained on a large corpus of the same domain as the described task and is evaluated as well. Medical professionals diagnose depression by interpreting the responses of individuals to a variety of questions, probing lifestyle changes and ongoing thoughts. Like professionals, an effective automated agent must understand that responses to queries have varying prognostic value. In this study we demonstrate an automated depression-detection algorithm that models interviews between an individual and agent and learns from sequences of questions and answers without the need to perform explicit topic modeling of the content.*

I. INTRODUCTION

Depression is a common illness worldwide, with more than million people affected. Depression is different from usual mood fluctuations and short-lived emotional responses to challenges in everyday life. Especially when long-lasting and with moderate or severe intensity, depression may become a serious health condition. It can cause the affected person to suffer greatly and function poorly at work, at school and in the family. At its worst, depression can lead to suicide. Close to 800 000 people die due to suicide every year. Suicide is the second leading cause of death in 15-29-year-olds. Although there are known, effective treatments for mental disorders, between 76% and 85% of people in low- and middle- income countries receive no treatment for their disorder(2). Barriers to effective care include a lack of resources, lack of trained health-care providers and social stigma associated with mental disorders. Another barrier to effective care is inaccurate assessment. In countries of all income levels, people who are depressed are often not correctly diagnosed, and others who do not have the disorder are too often misdiagnosed and prescribed antidepressants. Drug Use and Health in the United States report that, during the year 2016, 12.8% of adolescents between 12 and 17 years old and 6.7% of adults had suffered a major depressive episode (MDE) depression can be described to lead to an altered mood and may also be accompanied, for example, by a negative self-image, wishes to escape or hide, vegetative changes, and a lowered overall activity level. In addition to possible personal reasons for avoiding treatment, this is often due to a limited availability of mental health care, for example in conflict regions. Via a telephone survey in Germany researchers found out that shame and self-stigmatization seem to be much stronger

reasons to not seek psychiatric help than actual perceived stigma and negative reactions of others. The burden of depression and other mental health conditions is on the rise globally.

National security concern is the primary goal of any nation. Criminology studies focus on identifying criminal characteristics. The application of data mining techniques can help with this identification. Crime analysis, a part of criminology, is a law enforcement function that involves the systematic analysis of identifying and analyzing both patterns and trends in crime and disorder. In the current world, the criminals are becoming technologically sophisticated, often expressing their emotions on the web.

The World Wide Web's phenomenal growth has resulted in more users expressing their opinions online. Customers use these opinions to buy a product, conduct market analysis, and so forth. Twitter is one of the most popular online social networks to date, where users post their opinions in short text called "tweets". These tweets are typically limited to 140 characters. Twitter has approximately 500 million users; approximately 340 million tweets are sent every day. Twitter is used, primarily, for the following four reasons .

- Daily Chatter (e.g., status messages on what the user is doing)
- Conversations (e.g., tweeting to either a user or a group of users within a community)
- Sharing information (e.g., posting links to web pages)
- Reporting new (e.g., status updates on current affairs).

According to Bollen, "a tweet is a microscopic, temporally-authentic instantiation of sentiment". Since tweets are crispy and brief, the public sentiment can be easily explored. Twitter also provides the feature of Retweet (RT), which allows users to share content posted by another user.

Sentiment analysis (also known as opinion mining) refers to the use of natural language processing, text analysis, and computational linguistics to identify and extract subjective information in source materials. It is used to determine an author's attitude, with respect to a particular topic or the overall contextual polarity in the text. The rapid growth of Social media has spurred interest in sentiment. Various forms of online expressions (e.g., opinions-like reviews, ratings, and recommendations) have become major sources of information for businesses looking to market their products and manage their reputations. The challenge of detecting crime patterns lies in geographically analyzing crime-related tweets and then performing sentiment analysis to identify crime prone zones in nearly real-time. Most of the studies that focused on crime pattern detection [8, 9] used data mining techniques to better understand historic data. This study used online social media to detect crime prone areas in almost real-time.

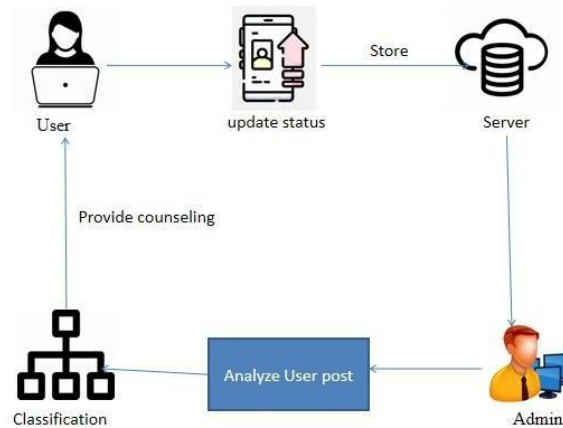
II. PROPOSED SYSTEM

Opinion Mining is also used for market trend analysis and for "mood prediction". Researchers have conducted in Sentiment Analysis regarding different Tools and their pros, cons and attributes Various classifiers comparison has done in this paper to identify the most efficient one. In this project we detect the depression level of the user. We collect that people social media activity and using machine learning algorithm to predict the mode. Personality assessment framework is proposed using DISC and Twitter Data. conducts analysis of user's occupation through his twitter content. Major and minor Groups are the result of this analysis.

III. SENTIMENT ANALYSIS

Sentiment analysis was used to determine a writer's/speaker's attitude with respect to either a topic or the overall contextual polarity of a text. Researchers use this analysis to measure emotions in online texts. The rise of social media fueled interest in using sentiment analysis to identify public opinions and interests. Several open source software tools utilize machine learning, statistics, and natural language processing techniques to automate sentiment analysis on a large collection of texts that have been gathered from various sources Sentiment analysis is a two-step process that includes both subjectivity classification and sentiment classification. The term subjectivity classification is defined as distinguishing factual sentences from those used to present opinions, before analyzing sentiments. Paragraphs that present facts are typically removed so the researcher can focus on those paragraphs in which the author expresses opinions. Both naive Bayes classification and Cut Based classification are used for subjectivity classification.

IV. SYSTEM ARCHITECTURE



V. SYSTEM TESTING

Testing is the stage of implementation of which aimed at ensuring that the system works accurately and efficiently before live operation commences. Testing is vital to the success of the system. System testing makes a logical assumption that if all the parts of the system are correct the goal will be achieved. The candidates system subject to a variety of tests. Online response, volume, stress, recovery, security and usability tests. A series of testing are performed for the proposed system before the system is ready for user acceptance testing.

5.1 Unit Testing

The procedure level testing is made first. By giving improper inputs, the errors occurred are noted and eliminated .Then the web form level is made.

5.2 Integration Testing

Testing is done for each module. After testing all the modules, the modules are integrated and testing of the final system is done with the test data, specially designed to show that the system will operate successfully in all its aspects conditions. Thus the system testing is a confirmation that all its correct and an opportunity to show the user that the system works.

5.3 Validation Testing

Software Testing - Validation Testing. The process of evaluating software during the development process or at the end of the development process to determine whether it satisfies specified business requirements. Validation Testing ensures that the product actually meets the client's needs.

5.4 Regression Testing

Regression Testing is defined as a type of software testing to confirm that a recent program or code change has not adversely affected existing features. Regression Testing is nothing but a full or partial selection of already executed test cases which are re- executed to ensure existing functionalities work fine.

VI. CONCLUSION

A crime pattern can be detected, nearly in real-time, when online social media is monitored. Crime can occur anywhere at any time. Previous statistics do not accurately identify the crime intensity of a specific location. More accurate results can be drawn from social media. Results from geographic data analysis conducted on various tweets provided a clear picture of the criminal trends in several different cities. The crime intensity day-wise positively

correlated with crime statistics from cops, which ultimately prove the hypothesis. The Ferguson shooting case study clearly differentiates the city's safe and dangerous pattern. To be more precise, we analyzed the specific twitter accounts which tweet only about the crime scenarios happened in the city based on sheriff data and visualized.

REFERENCES

- [1]. D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," in Proc. 9th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2003.
- [2]. J. Goldenberg, B. Libai, and E. Muller, "Talk of the network: A complex systems look at the underlying process of word-of-mouth," Marketing letters, vol. 12, no. 3, pp. 211–223, 2001
- [3]. H. W. Hethcote, "The mathematics of infectious diseases," SIAM review, vol. 42, no. 4, pp. 599–653, 2000.
- [4]. M. E. J. Newman, "The structure and function of complex networks," SIAM review, vol. 45, no. 2, pp. 167–256, 2003.
- [5]. S. Wang, X. Hu, P. S. Yu, and Z. Li, "MMRate: inferring multi-aspect diffusion networks with multi-pattern cascades," in Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2014, pp. 1246–1255.
- [6]. Y. Tang, Y. Shi, and X. Xiao, "Influence maximization in near-linear time: a martingale approach," in Proc. ACM Int. Conf. Special Interest Group Manage. Data, 2015.
- [7]. L. Weng, A. Flammini, A. Vespignani, and F. Menczer, "Competition among memes in a world with limited attention," Scientific Reports, vol. 2, 2012.
- [8]. S. A. Myers and J. Leskovec, "Clash of the contagions: Cooperation and competition in information diffusion," in Proc. 12th IEEE Int. Conf. Data Mining, 2012.
- [9]. X. Rong and Q. Mei, "Diffusion of innovations revisited: from social network to innovation network," in Proc. 22th ACM Int. Conf. Inf. Knowl. Manage., 2013.
- [10]. Y. Bi, W. Wu, and Y. Zhu, "Csi: Charged system influence model for human behavior prediction," in Proc. 13th IEEE Int. Conf. Data Mining, 2013