EMPLOYABILITY OF MULTI-LEVEL SUPPORT VECTOR MACHINE (SVMS) BASED PREDICTIVE MODEL IN THE ENHANCED EFFICACY OF RISK ASSESSMENT FOR MENTAL ILLNESS ON SOCIAL MEDIA PLATFORMS

Bahisht Samar

Amity University, Noida

ABSTRACT

As shown by several previous studies, personal narratives through social media are often indicative of one's psychological state. In particular, mental illnesses such as depression were found to be associated with distinct linguistic patterns. However, many people with mental illness still do not receive full treatment. In this paper, we study mental illnesses through people's choice of words in expressing themselves on two popular social media platforms, Reddit and Twitter. Our goal is to develop an empirical model to detect and diagnose major mental disorders in individuals. We build a substantial dataset of posts made by people suffering from mental illnesses and the control ones, and in order to generate numerical feature from text we apply text cleaning and Word2Vec language modelling, and then for classification we used SVM machine which classifies posts and users with high accuracy. We achieve an accuracy of 95% on Twitter users and an accuracy of 73% on the Reddit challenge.

INTRODUCTION

Nowadays, given the popularity of social media networks such as Flickr, Instagram, Twitter, etc., they have become one of the most important sources for quick and easy access to information on all aspects of life. Social networks allow people to express themselves freely and share their feelings and experiences with others. While people with mental disorders, such as those suffering from depression tend to isolate themselves and withdraw from an active social life, it was found that they easily express their mental illness state on social media, as a way of relaxation. Therefore, the use of social media platforms by mentally related individuals as a powerful tool for communicating with one another, sharing experiences and supporting each other is exponentially growing. The information collected on these platforms is a rich source for sentiment analysis or investigation of people’s mental health problems.

Natural Language Processing (NLP) refers to research that examines ways in which a computer can understand and manipulate natural language in form of text or speech. And with the help of sentiment analysis as a tool for data classification, NLP has shifted from simply focusing on linguistic analysis towards semantic analysis and decision making based on individuals’ writings. Sentiment analysis is used in the following two models. First it examines how social media trends are related to mental illnesses. Second it tries to detect mentally ill users through analysis of the content created by them. In fact, automated analysis of social network has capability of offering methods for early detection of such diseases. Users with mental illness have been identified by what they publicly share on online social media platforms and they can be distinguished from ordinary users by the language patterns in their written texts or other online activities. For example, depressed users tend to use more first-person singular pronouns in spoken language or depressed groups have elevated the use of the word “I” and negative emotions in comparison with control groups. So automated detection methods which are called early risk detection may help to detect depressed or in general mentally ill people through supervising of written text spreading in social medias, and then these individuals could be assessed more thoroughly to have been provided with resources, supports, and full treatments. Our automated detection machine is built by training a classifier using a set of available training data to categorize users into a control group and an affected group. This classification is performed using predictive models that use extracted features from users' written text or corpus.

PROPOSED METHOD

Early risk detection as a subcategory of sentiment analysis is a new field of research that has many applications in a variety of fields specially those applications which are related to health issues and safety. For example, when a predator starts seducing children for abuse intentions, or when an offender starts posting anti-social threats on a social media, or identifying people with mental illnesses such as depression over social media are all some application of early detection system. Many methods based on machine learning techniques have been proposed to be used as early risk detection system. In this project, the main goal is to classify online posts into positive and control groups so that if a post belongs to a person who has a mental illness, we call it positive. Based on the scores of the posts sent by each user, all the users are classified into these two categories. It should be noted that each post of a user is influential in classifying individuals.

The main characteristics of our paper in improving the accuracy of state-of-the-art early detection system for depression on social networks are as follows:

• Use innovative measures which are specifically designed for our early detection system.

• Extraction of text features with higher accuracy than conventional approaches (only based on tf-idf vectorization), by using a two-level feature extraction and classification.

• Exploring the behaviour of the first SVM classifier to better predict positive cases.

As mentioned in the previous section, like most of the proposed methods, our approach has three main parts which are discussed in more detail.

A. Preprocessing

The context of an online post has additional phrases such as hashtags, Image Links, etc. Table I lists some of the most frequent phrases in an online post. In the text version of a post which is used in our processing, these non-informative content does not add any meaning to the subject and also makes the posts look more complicated and therefore harder to process. This step as mentioned in [14] consists of basic text cleaning steps including converting plurals to singular forms, removing verb prefixes and suffixes, removing web address and @user mentions, expanding abbreviations to their full version, removing stop words and normalizing slang forms (e.g., “niiiiiiice” and “goooood” become “nice” and “good”). So by removing these, the posts will be clear and more suitable for sentiment analysis. The first section of the proposed algorithm would be detecting and omitting these additional phrases. This preprocessing is usually done with text cleaner methods such as using regular expressions over post's texts which results in obtaining the clear text of each post as shown in table II.

B. Feature Extraction

If you want to feed words into machine learning models, you need to map them onto some set of numeric vectors. In the word2vec technique, we train a two-layer neural network and then the trained network will be sable to predict the surrounding words in a sentence [4]. After the training, the network produces a numerical representation for each input word, in a way that each word would be mapped to a point in the corresponding space. Typically, this space has several hundreds of dimensions. In our case, it has one hundred dimensions. This representation of words in vector space is such that words that have a close meaning are also close in vector space. The resulting learnt vector, also known as the embeddings, is a numeric representation of each word. In this section, we map each word of a post into a vector and to get the numeric vector of each post, for each word of a post, we calculate the arithmetic mean of all corresponding vectors. Therefore, we have a numeric vector for each post with a size equal to 100 which is the size of the numeric vectors of words in the word2vec model.

C. Classifier

The classical SVM attempts to find a hyperplane that can divides the data into two groups, so that the hyperplane is at the most distant from both of data sets. In our case, we use a two-level SVM. In the first SVM, instead of labeling data, the machine gives each post a score which defines its probability of belonging to a certain class. It means that the higher the score, the higher the likelihood that the person who sent the post has a mental illness. The second classifier considers the scores of all posts of an individual user, and classifies people into positive and control groups.

1) First Classifier: The first SVM uses word vectors of each tokenized post as the dataset, trained into the train and test subsets. We randomly choose about half of the posts (half of the positive and half of the negative ones) as the training and testing samples. SVMs treat each post as a point in an extremely high dimensional space (one dimension for each value of word vector). If a post belongs to a person who has a mental illness, SVM labels the post as 1 and otherwise as 0. As mentioned before, unlike the usual way of using an SVM in classification, where only the predicted labels are used, we use a continuous result of SVM which gives each post a score.

This score is the distance between that post and the imaginary separator hyper plane between the negative and the positive data. Therefore, higher distance shows a higher certainty of the post belonging to a certain group.

2) Second Classifier: The second SVM uses the score of the posts for each user from the first SVM and classifies each user into either the mentally ill group or the controlled group. The output of the first SVM for each user is a one-dimensional vector of scores of posts, which naively can feed into the second SVM directly. But this results in a high complexity for the SVM and thus, a poor performance. Therefore, we use 36 statistical measures, such as average, variance, standard deviation and so on, to extract valuable information. These measures are defined in detail in appendix A. Based on our result, not all features are as descriptive. So we used fisher ratio feature selection method [6] to choose the most descriptive statistical of all 36 measures. Fisher ratio algorithm select 8 most expressive measures. The selected measures are listed in table III. Thus, the second SVM is trained based on these 8 measures to classify each user. If a user has a mental illness, they get a label equaled to 1 and otherwise a label of 0. Below you can see the algorithm for the classification.

Inputs: Numeric vector representation of tweets of all users (V_j^i).
Output: predicted class of users (L^i). (1=mental illness, 0=control group)

$For\ i = 1..N$
 $For\ j = 1..M_i$ (M_i: Number of tweets per user)
  S_j^i = FirstSVM.DecisionScore(V_j^i)
 $FR_i[1..8] = Fisher\_Ratio(S^i[1..M_i])$
 $L^i = SecSVM.Predict(FR_i[1..8])$

EXPERIMENTS

A. Experiment

The first experiment was done using the twitter dataset built by Olteanu et al.[14]. Their purpose was to gain an appreciation of power of social network as a data resource for data mining the people's experiences for decision-making systems. They provided a various set of topics in which people will normally seek information when try to make decisions. It consists of 3 main domains cataloged as business, health, and society and each domain include experiences and events people faced. In our experiment, we use random people suffering from mental illnesses which is a subcategory of the health domain as the positive group, and random users of other subcategories and domains as a control group. The dataset of each group including tweets and user information are downloaded via Twitter API by tweet IDS. Because of performance issues only users with more than 90 tweets are qualified to be assessed in our test. Details of the number of case studies are in table IV. As mentioned earlier in section 3, we have a 2 level SVM classifier. The first SVM classifies each tweet and gives a score that measures the distance of the tweet from each of the two groups. Based on this distance we can categorize tweets into positive and control groups. Because of unbalanced data, in order to have the same false positive and false negative ratio, we used Receiver Operating Characteristic

(ROC) curve analysis. The results are presented in terms of F-measure (F1), Precision (P), and Recall (R). The second SVM categorizes each person based on scores of all tweets they have sent.

CONCLUSION AND FUTURE WORKS

A fair number of studies have been made and their results are published on how social network data can help and improve the prediction of sentiment analysis. This study sets out to assess whether the text analysing of online user posts such as Twitter or Reddit can be of use for finding users who suffer from mental illnesses. The main obstacle we had to overcome was finding a way to conveniently extract useful information from such a large number of messages for each user. To this purpose, we have applied Word2Vec model and some feature selection techniques, namely Fisher Ratio and multi-level classifiers that allow us to tell mentally ill users from the control ones with great precision. Advances in natural language processing and machine learning have made it possible to monitor social medias at large-scale for sentiment analysis purposes such as early risk detection. We showed the ability of our system as a tool to detect major mental illnesses in individuals. One problem, though, is the lack of sufficient labeled data for the training phase. Fortunately, as time goes by and more data is collected and interest in this research area increases, this may lead to larger datasets in combination with explicitly associated diagnostic criteria, assessments, or health records, to emphasize validity. Instead of just using one model for predicting people's mental state, we could use many of them and decide the final prediction by weighted majority vote and in the end, this could lead to the development of systems for automated public health tracking at large scale.

Table I: Comparison of our method with participant of Task 1 ERisk challenge 2019

	F1	**P**	**R**
Our Method	65.30	52.02	**87.67**
lirmm	68	77	60
UppsalaNLP	34	39	30
CLaC	**71**	64	79
BiTeM	19	73	11
UDE	61	51	74

Table II: Comparison of our method with participant of Task 2 ERisk challenge 2019

	F1	**P**	**R**
Our Method	47.75	34.40	**78.04**
BiTeM	46	52	41
UNSL	**52**	**71**	41
UDE	30	45	22
Lirmm	46	52	41
CAMH	22	12	98

REFERENCES

1. G. Shen et al., “Depression Detection via Harvesting Social Media: A Multimodal Dictionary Learning Solution,” in Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, {IJCAI-17}, 2017, pp. 3838–3844.
2. T. Mikolov, G. . Corrado, K. Chen, and J. Dean, “Efficient Estimation of Word Representations in Vector Space,” 2013, pp. 1–12.
3. C. M. Homan, N. Lu, X. Tu, M. C. Lytle, and V. M. B. Silenzio, “Social Structure and Depression in TrevorSpace,” in Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing, 2014, pp. 615–625.
4. R. Kshirsagar, R. Morris, and S. Bowman, “Detecting and Explaining Crisis,” 2017.
5. J. Serrà, I. Leontiadis, D. Spathis, G. Stringhini, J. Blackburn, and A. Vakali, “Class-based Prediction Errors to Detect Hate Speech with Out-of-vocabulary Words,” in Proceedings of the First Workshop on Abusive Language Online, 2017, pp. 36–40.
6. A. G. Reece, A. J. Reagan, K. L. M. Lix, P. S. Dodds, C. M. Danforth, and E. J. Langer, “Forecasting the onset and course of mental illness with Twitter data,” Sci. Rep., vol. 7, no. 1, p. 13006, 2017.
7. J. H. Shen and F. Rudzicz, “Detecting Anxiety through {R}eddit,” in Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology {---} From Linguistic Signal to Clinical Reality, 2017, pp. 58–65.
8. K. Niederhoffer, J. Schler, P. Crutchley, K. Loveys, and G. Coppersmith, “In your wildest dreams: the language and psychological features of dreams,” in Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology {---} From Linguistic Signal to Clinical Reality, 2017, pp. 13–25.
9. J. Pennebaker, R. Boyd, K. Jordan, and K. Blackburn, “The Development and Psychometric Properties of LIWC2015.” 2015.
10. R. C. Hawkins II and R. L. Boyd, “Such stuff as dreams are made on: Dream language, LIWC norms, and personality correlates.,” Dreaming, vol. 27, no. 2, pp. 102–121, 2017.
11. A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, “Bag of Tricks for Efficient Text Classification,” 2017, pp. 427–431.
12. C. Pool and M. Nissim, “Distant supervision for emotion detection using {F}acebook reactions,” in Proceedings of the Workshop on Computational Modeling of People{’}s Opinions, Personality, and Emotions in Social Media ({PEOPLES}), 2016, pp. 30–39.
13. S. Paul, J. Sree Kalyani, and T. Basu, “Early Detection of Signs of Anorexia and Depression Over Social Media using Effective Machine Learning Frameworks,” 2018.
14. A. Olteanu, O. Varol, and E. Kiciman, “Distilling the Outcomes of Personal Experiences: A Propensityscored Analysis of Social Media,” in Proceedings of The 20th ACM Conference on Computer-Supported Cooperative Work and Social Computing, 2017, Computer-S.