INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT (IJSREM)
VOLUME: 05 ISSUE: 07 | JULY - 2021 ISSN: 2582-3930

TWITTER- SENTIMENT -ANALYSIS -MASTER

DR.VENKATESH,M.E.,PH.D
DEPARTMENT OF COMPUTER SCIENCE
JEPPIAAR ENGINEERING COLLEGE
CHENNAI,INDIA

DINESH KUMAR.S
DEPARTMENT OF COMPUTER SCIENCE
JEPPIAAR ENGINEERING COLLEGE
CHENNAI,INDIA

MAHENDRA VARMA.J
DEPARTMENT OF COMPUTER SCIENCE
JEPPIAAR ENGINEERING COLLEGE
CHENNAI,INDIA

ABSTRACT

The use of Social Network Sites (SNS) is increasing nowadays especially by the younger generations. The availability of SNS allows users to express their interests, feelings and share daily routine. allows users to express their interests, feelings and share daily routine. Many researchers prove that using user-generated content (UGC) in a correct way may help determine people's mental health levels. Mining the UGC could help to predict the mental health levels and depression. Depression is a serious medical illness, which interferes most with the ability to work, study, eat, sleep and having fun.However, from the user profile in SNS, we can collect all the information that relates to person's mood, and negativism.In this research, our aim is to investigate how SNS user's posts can help classify users according to mental health levels.

I INTRODUCTION

In this project the social media tweets (Twitter) are considered. Social Network Sites (SNS) is a source of data and screening tool to classify the users according to user generated contents (UGC). By using machine learning techniquesto find the sentiment of the users.

As depression is very serious problem which is increasing day by day, many people are suffering from this problem. In India, out of total population 7.5% of it facing this problem. It seems to be major issue and that is the reason it motivate us to worked on it .Earlier diagnosis of depressed patient were done on basis of questionnaires and its behaviour reported by his relatives or friends. But the result was not so qualitative and accurate. In contrast with that, social media is powerful tool for predicting depression levels of an individual.

Nowadays, many people are using social media platforms such as Tweeter, twitter. They share their thoughts, feelings, emotions, feelings of guilt, worthlessness, helplessness and egoistic nature of individual etc. Whatever they post is related to their daily activities & happenings. Social media helps to know about individual's thinking, mood, activities & socialization. So by analysing the social media data & applying some algorithms on it, we can able to determine the depression levels of that particular individual. So it will help to diagnose that person before he/she gets more affected to it. This motivates us to do the project, so that this will help not only to psychiatrist but also used by patients who want to do self-diagnosis

II LITERATURE REVIEW

1. Title: A framework for depression dataset to build automatic diagnoses in clinically depressed Saudi patient

   Author: Lubanayusuf

   Year:2016

   Description:

   Depression is a public health problem that has high effects on a person's functional and social relationships. Depression is a growing problem in the society. It causes pain and suffering not only to patients, but also to those who care about them. Depression disorder is hard to diagnose,

because its symptoms could be confused with other disorders and has different cross-cultural symptoms. This paper proposes a framework that would best solve the problem of automatic depression detection in depressed Saudi patients. This paper particularly focuses on designing the collection of Saudi depression dataset using multiple modalities.

2. Title: Detection of Clinical Depression in Adolescents' Speech During Family Interactions

Author: Nammana C madague

Year:2010

Description:

The properties of acoustic speech have previously been investigated as possible cues for depression in adults. However, these studies were restricted to small populations of patients and the speech recordings were made during patients' clinical interviews or fixed-text reading sessions. Symptoms of depression often first appear during adolescence at a time when the voice is changing, in both males and females, suggesting that specific studies of these phenomena in adolescent populations are warranted. This study investigated acoustic correlates of depression in a large sample of 139 adolescents (68 clinically depressed and 71 controls).

3. Title: An improved model for depression detection in micro blog social network

Author: Xingu wang

Year:2013

Description:

Social networks contain a tremendous amount of node and linkage data, providing unprecedented opportunities for a wide variety of fields. As the world's fourth largest disease, depression has become one of the most significant research subjects. Previously, a depression classifier has been proposed to classify the users in online social networks to be depressed or not, however, the classifier takes only node features into account and neglects the influence of linkages. This paper proposes an improved model to calculate the probability of a user being depressed, which is based on both node and linkage features. The linkage features are measured in two aspects: tie strength and interaction content analysis. Moreover, the propagation rule of depression is considered for improving the prediction accuracy

4. Title: Toward the development of cost effective e-depression effective system

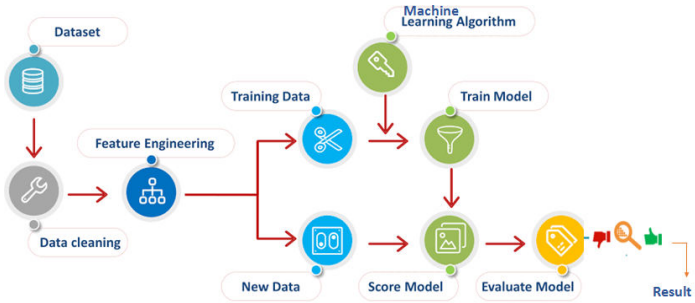
Author: Taun D Pham

Year:0212

Description:

Diagnosis and prevention of depressive disorders at any scale have been attracting considerable attention of the public healthcare in Japan because depression is one of the most rapidly pervasive mental disorders in the country. A major issue that hinders the feasibility of depression screening for its prevention is the availability of some simple and cost-effective methods for depression detection and monitoring. Here in this paper, we present the development of a computerized tool for depression detection. The tool utilizes the theory of chaos and systems complexity to extract robust dynamically statistical features of physiological signals provided by the low-cost technology of photoplethysmography.

III IMPLEMENTATION



System architecture is used to design and develop an web application, which provide an easy and convenient way to get information about depression levels of user by using machine learning algorithms and according to the location of user, the information about doctor is provided. The extraction class will performed the extraction of textual data from tweeter through tweeter graph API.Preprocessing class

is used to preprocess the extracted data. Data must be clear,right and it is preprocessed for taking care of missing or repetitive attributes. The data ought to be complete and reliable data to deliver the best result from the data mining methodology. Preprocessing of data takes place by using techniques such as tokenization, lower case conversion, word stemming and words removal.Term frequency (tf) has been computed to measure term occurrence. In proposed system user is on Tweeter, according to his Tweeter post system can find out user in stressed or not as well as different quaternaries which is provided by the system. If user's are not on Tweeter they can attempt only quaternaries which is provided by the system according to that we can find out user's in stressed or not.

SYSTEM IMPLEMENTATION

Modules :

- Data Collection
- Pre-processing

MODULE DESCRIPTION:

Data Collection and Pre-processing

- Data Collection is one of the most important tasks in building a machine learning model.
- It is the gathering of task related information based on some targeted variables to analyse and produce some valuable outcome.
- However, some of the data may be noisy, i.e. may contain inaccurate values, incomplete values or incorrect values.
- Hence, it is must to process the data before analysing it and coming to the results.
- Data pre-processing can be done by data cleaning, data transformation, data selection.

Data Collection and Pre-processing

- Data cleaning: Fill in missing values, smooth noisy data, identify or remove outliers, and resolve inconsistencies.
- Data transformation may include smoothing, aggregation, generalization, transformation which improves the quality of the data.
- Data selection includes some methods or functions which allow us to select the useful data for our system.

IV.SYSTEM REQUIREMENTS

The software requirements specification is produced at the culmination of the analysis task. The function and performance allocated to software as part of system engineering are refined by establishing a complete information description as functional representation of system behavior, an indication of performance requirements and design constraints, appropriate validation criteria.

HARDWARE REQUIREMENTS

System	: Pentium IV 2.4 GHz
Hard Disk	: 40 GB
Floppy Drive	: 1.44 Mb
Monitor	: 15 VGA Colour
Mouse	: Logitech
Ram	:4GB

SOFTWARE REQUIREMENTS

Operating system	: Windows 10
IDE	: Anaconda
Coding Language	: python

V. CONCLUSION

+

The experiments results show that the use of behavioural information on Tweeter, both in forms of messages and activities, could predict depression. However, the sample of this research is relatively small because Tweeter has limited their permission to collect personal information and the process of gaining approval has become more complicated. Thus, the results getting from this study might not cover all relevant factors. Moreover, as the language-related features had to be translated from Thai to English for analyzing the process, there might be some errors due to this process because some important sentiment polar words might have been eliminated during the translation process.

FUTURE ENHANCEMENT

For the future research, we intend to collect more data to get more relevant and valid features. Manual annotating all complex attributes using crowdsourcing and deeper dimensions should also be analysed in order to be able to create a better depression detection algorithm.

REFERENCES

[1]Wikipedia, "Twitter, " http://en.wikipedia.org/wiki/Twitter.2014.

[2] Xinhuanet, "Sina Microblog Achieves over 500 Million Users," http://news.xinhuanet.com/tech/2012-02/29/c_122769084.htm. 2014.

[3] D. Perito, C. Castelluccia, M.A. Kaafar, and P. Manils, "How unique and traceable are usernames?," Privacy Enhancing Technologies (PETS'11), pp. 1-17, 2011.

[4] J. Liu, F. Zhang, X. Song, Y.I. Song, C.Y. Lin, and H.W. Hon,"What's in a name?: an unsupervised approach to link users across communities," Proc. of the 6th ACM international conference on Web search and data mining(WDM'13), pp. 495-504, 2013.

[5] R. Zafarani and H. Liu, "Connecting corresponding identities across communities," Proc. of the 3rd International ICWSM Conference, pp. 354-357, 2009.

[6] R. Zafarani and H. Liu, "Connecting users across social media sites: a behavioral-modeling approach, " Proc. of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'13), pp.41-49, 2013.

[7] A. Acquisti, R. Gross and F. Stutzman, "Privacy in the age of augmented reality," Proc. National Academy of Sciences, 2011.

[8] T. Iofciu, P. Fankhauser, F. Abel, and K. Bischoff, "Identifying users across social tagging systems," Proc. of the 5th International AAAI Conference on Weblogs and Social Media, pp. 522-525, 2011.

[9] M. Motoyama and G. Varghese, "I seek you: searching and matching individuals in social networks," Proc. of the 11th international workshop on Web Information and Data Management (WIDM'09), pp. 67-75, 2009.

[10] O. Goga, D. Perito, H. Lei, R. Teixeira, and R. Sommer, "Largescale Correlation of Accounts across Social Networks," Technical report, 2013.

[11] K. Cortis, S. Scerri, I. Rivera, and S. Handschuh, "An ontology based technique for online profile resolution," Social Informatics,
Berlin: Springer, pp. 284-298, 2013.

[12] F. Abel, E. Herder, G.J. Houben, N. Henze, and D. Krause, "Cross-system user modeling and personalization on the social web," User Modeling and User-Adapted Interaction, vol. 23, pp. 169-209, 2013.

[13] O. De Vel, A. Anderson, M. Corney, and G. Mohay, "Mining email content for author identification forensics," ACM Sigmod Record, vol. 30, no. 4, pp. 55-64, 2001.

[14] E. Raad, R. Chbeir, and A. Dipanda, "User profile matching in social networks," Proc. Of the 13th International Conference on Network-Based Information Systems (NBiS'10), pp.297-304, 2010.

[15] J. Vosecky, D. Hong, and V.Y. Shen, "User identification across multiple social networks," Proc. Of the 1st International Conference on Networked Digital Technologies, pp.360-365, 2009.

[16] P. Jain, P. Kumaraguru, and A. Joshi, "@ i seek 'fb. me': identifying users across multiple online social networks," Proc. of the 22nd International Conference on World Wide Web Companion, pp. 1259-1268, 2013.

[17]P. Jain and P. Kumaraguru, "Finding Nemo: searching and resolving identities of users across online social networks," arXiv preprint arXiv:1212.6147, 2012.

[18] R. Zheng, J. Li, H. Chen, and Z. Huang, "A framework for authorship identification of online messages: writing-style features and classification techniques," J. of the American Society for Information Science and Technology, vol. 57, no. 3, pp. 378-393,2006.

[19] M. Almishari and G. Tsudik, "Exploring linkability of user reviews," Computer Security–ESORICS 2012 (ESORICS'12), pp. 307- 324, 2012.

[20] X. Kong, J. Zhang, and P.S. Yu, "inferring anchor links across multiple heterogeneous social networks," Proc. of the 22nd ACM International Conf. on Information and Knowledge Management (CIKM'13), pp. 179-188, 2013.

[21] O. Goga, H. Lei, S.H.K. Parthasarathi, G. Friedland, R. Sommer, and R. Teixeira, "Exploiting innocuous activity for correlating users across sites," Proc. 22nd international conference on World WideWeb (WWW'13),pp. 447-458, 2013.

[22] A. Narayanan and V. Shmatikov, "De-anonymizing social networks," Proc. Of the 30th IEEE Symposium on Security and Privacy (SSP'09), pp. 173-187, 2009.

[23] S. Bartunov, A. Korshunov, S. Park, W. Ryu, and H. Lee, "Joint link-attribute user identity resolution in online social networks," The 6th SNA-KDD Workshop '12, 2012.

[24] N. Korula and S. Lattanzi, "An efficient reconciliation algorithm for social networks," arXiv preprint arXiv:1307.1690, 2013.

[25] P. Erdős and A. Rényi, "On random graphs I," Publ. Math. Debrecen, vol. 6, pp. 290-297, 2010.