Detecting Depression Using Auditory and Linguistic Indications

Poojan Patel[1], Navin Agarwal[2], Prof. Rakhee Kundu[3]

Computer Engineering Deaprtment, Mukesh Patel School of Technology and Management- NMIMS University

[1]poojan17.nmims@gmail.com

[2]official.navin.agarwal@gmail.com

[3]rakheekundu@rocketmail.com

***Abstract*— Depression is known to result in neuro-physiological and neurocognitive changes that affect control of motor, semantic, and cognitive functions. In this paper, biomarkers are derived from all of these techniques, drawing first from previously developed neurophysiological motivated auditory and outer coordination and timing features. In addition, a unique indicator of lower vocal tract constriction in articulation is assimilated that relates to vocal projection. Linguistic features are analyzed for content using a skimpy coded rhetorical embedding space, and for circumstantial clues related to the individuals current or previous depression condition.**

Keywords— ***Depression classification, Multimodal, decision tree, speech synthesis, Task functional magnetic resonance imaging, Public Heath Questionnaire.***

1. Introduction

According to world health organization [1] depression affects more than 300 million people every year and it is one of the most common mental disorder.Depression affects between 5% and 10% of individuals in primary care but is only diagnosed in around 50% of cases. Similar type of problems was also found in general hospital settings, where there is substantial under identification and unmet need for psychological health services**.**

Depression is different with respect to the usual mood fluctuations and short-lived emotions to challenges in everyday life. When the condition of emotional imbalance last for a longer or a moderate period of time also it needs special attention as it might be depression. A person suffering from depression can affect his personal and professional life adversely and leave a negative impact everywhere. The worst thing that a person suffering from depression can do is harm an individual or even attempt suicide. Nearly 800000 people attempt suicide yearly and major suicidal attempts are done by age group 15-29 years.

In the current paper we are speaking about the major technique used to detect depression with the dataset provided by the Audio/Visual Emotion Challenge (AVEC) 2016 and Audio/Visual Emotion Challenge (AVEC) 2017 along with the data developed during the Distress Analysis Interview Corpus (DAIC)

2. Background

The Distress Analysis Interview Corpus (DAIC) is a collection of sub structured clinical interviews. Designed to simulate usual protocols for identifying people at risk for post-traumatic stress disorder (PTSD) and depression, these interviews were generated as part of a bigger effort to create a computer that interviews people and identifies verbal and nonverbal indicators of mental illness (De Vault et al., 2014). The corpus contains four types of interviews:

Face-to-face interviews between participants and a human interviewer

Teleconference interviews, conducted by a human interviewer over a teleconferencing system;

Wizard-of-Oz interviews, conducted by an animated virtual interviewer called Ellie, controlled by a human interviewer in another room;

Automated interviews, where participants are interviewed by Ellie operating as an agent in a fully automated mode.

3. METHODOLOGY

A. META-ANALYSIS AND PATIENT HEALTH QUESTIONNAIRE:

The paper uses the method of questionnaire for the purpose of identifying depression ion patients at various medical centers. It identifies 17 integral validation studies conducted in medical care; medical outpatients; and specialist medical services (cardiology, stroke, dermatology, head injury, and otolaryngology). It uses arbitrary repercussions bivariate meta-analysis at suggested cut points to generate summary receiver-operator characteristic (sROC) curve.

The aim was to establish the psychometric characteristics of the PHQ2 and PHQ9 as screening instruments for depression. The diagnostic validity of the PHQ2 was only checked in 3 months and showed high variability in sensitivity. The pre-conditions for this screening and case finding is that the instruments must be valid, reliable, simple, self-administrable and easy to use. The particular items on the instruments are obtained from the PRIME-MD interview schedules and are engineered to establish DSM-IV criteria based psychiatric diagnoses. All paragraphs must be indented. All paragraphs must be justified, i.e. both left-justified and right-justified.

A.1) Study Design and Setting.

Includes all cross-sectional validation research of the PHQ in primary care, community and hospital places amongst adults.

A.2) Condition and Reference Test.

Various studies have been undertaken for reporting of the ability of the PHQ to detect depression. Disorders had to be stated as per the classificatory systems such as, The International Classification of Diseases (ICD) or Diagnostic and Statistical Manual of Mental Disorders (DSM).

A.3) Diagnosis and Screening Instrument Cut Points.

The PHQ measures were introduced within the PRIME-MD set of instruments and scales and were made for use in primary care and 20 non-psychiatric settings. The nine-item PHQ contains items derived from the DSM-IV classification system pertain to:

(1) anhedonia, (2) depressed mood, (3) trouble sleeping, (4) feeling tired, (5) change in appetite, (6) guilt or worthlessness, (7) trouble concentrating, (8) feeling slowed down or restless, (9) suicidal thoughts.

The two-item PHQ includes only 2 questions pertaining to:

(1) anhedonia, and (2) low mood.

A.4) Search Strategy.

It developed search keywords from a series of depression-specific terms developed by the Cochrane Depression and Anxiety group and implemented a series of free-text terms to identify any publications that mentioned the PHQ and PRIME-MD instruments by keyword in their abstracts.

A.5) Data Abstraction and Quality Assessment.

It initially generated 2×2 tables for all studies at suggested cut off points. From these it calculated sensitivity, likelihood ratios (positive and negative), specificity. The likelihood ratio indicates a solution of the predictive ability of a test, which, unparallel to positive predictive value, is a fundamental predictive feature of the instrument that does not change according to the baseline prevalence of the disorder in consideration.

We also processed the diagnostic odds ratio (DOR): the ratio of the odds of a positive test among those with the disorder to the odds of a positive result among those without the disorder.

A.6) Data Synthesis and Meta-Analysis.

It used a bivariate diagnostic meta-analysis to gain control of pooled estimates of sensitivity and specificity; positive and negative likelihood ratios; and a summary diagnostic odds ratio. In short, this method fits a 2-level model, with independent binomial distributions for the true positives and true negatives conditional on the sensitivity and specificity in each study, and a bivariate normal model for the logit transforms of specificity between studies and sensitivity.

Receiver Operator Characteristic (ROC) curve is the most informative method of demonstrating the inherent trade-offs between sensitivity and specificity for the test or diagnostic instrument. Thus, it created a single graph of sensitivity and specificity in the ROC space. Summary Receiver Operator Characteristic Curve (sROC) had then been developed using the bivariate model to generate a 95% confidence ellipse within ROC Space.

This stepwise overview of the diagnostic attributes of the PHQ2 and PHQ9 instruments is the initial step to summarize current validation work beyond the initial population studies. For major good depressive disorder, the PHQ9 has diagnostic properties, and was ready to accurately identify major depression (sensitivity 92%) while being able to dodge this condition with some certainty (specificity 80%). Despite the clinical heterogeneity of studies in terms of settings (community, primary care, and a range of hospital specialties), the characteristics of the PHQ for major depression were relatively stable between a range of settings and specialties. The diagnostic properties of the PHQ9 were relatively good using either the ≥ 10 or "diagnostic algorithm" process, and the meta-analysis was not able to find a significant difference in performance according to which method was used.

Clinicians and researchers may therefore vary in their choice of cut point according to the clinical population, and the data in the present meta-analysis provide some empirical justification for this. There is enough under-recognition of depressive disorders in primary care and hospital settings, and the PHQ has now emerged a reasonable instrument, which can be adopted in routine practice, within a variety of settings. It compares well with longer or clinician-administered instruments. Even highly sensitive and specific screening tests are likely to be wrong more often than they are right in non-specialist settings with a low base rate of depression. Candidate depression enhancements include collaborative care, and successful programs often include case finding instruments such as the PHQ.

B. MultiModal Depression Detection

The study [15] suggests that men are likely to develop MDD at risk of 10-20% whereas women have chances of 5-12% of developing the same. The study [16] indicates to the fact that median period of depression is said to be three months, and subsequently, prediction of the depression stage of any individual should be liable to analysis and observation for longer durations.

The challenges involved in making such a depression detection model are:

1) **Large Decision Unit**. The size of the data must be enormous for the accuracy but processing such huge volumes of data is an arduous task. During the processing of huge audio or video data, using statistical operations such as e.g. min, max, mean, quartiles to smaller duration features over the entire period of observation can lead to misplacement of highly detailed and minute temporal information such as short-term sighs in despair, anger or laughing.
2) **Restricted number of samples**. With a diminutive sample size, the amount of features should also be small to evade the challenges of dimensionality and overfitting.

For solving the aforementioned problems, we use modeling based multi modal feature. The interview [3] is first segmented according to topics. Then, audio, video, and semantic features are generated for every segment individually and subsequently placed in a different slot of the topic in the feature vector. After the features for all topics have been placed, a two-step feature selection algorithm is performed to reduce the feature vector and only retain the most differential features. Following results were expected:

1) **To Logically Arrange short-term characteristic depending on context.**
2) **Increase in flexibility and accurate unearthing of important feature.** The paper includes only those features which are in useful context. The suggested feature building method supplies any permutation of features and context like smiling (friends) and smiling (family).

The different components are explained here:

B.1) Topic Modeling.

Topic Modeling [4] needs algorithm like latent Dirichlet allocation [17] (LDA). First, we form a preparative sentence dictionary by coursing through all the Ellie's speech and store all non-repetitive sentences. Then, we conduct manual cleansing of the introductory dictionary, where sentences that do not initiate substitutive topics (e.g., "that's better") are disposed. Later, we start clustering of the dictionary, in which the sentences which start the similar topic are combined together. This is done in two moves. First, very akin sentences with up to 3 or 4 symbols in dissimilitude are assembled involuntarily. Secondly, added manual assembling and detection check are conducted. Then, we retrospect each sentence group, link each group to the corresponding group, and place it into the topic dictionary. Thus, the topic dictionary is reframed as [topic name, corresponding Ellie sentences]. For conducting context-aware analysis, the attribute vector needs to store the features of each topic individually.

B.2) Features.

Audio features taken from C——OVAREP [18] toolkit which develops a 74- dimensional attribute (feature) vector for permanent features. Video features use action units (AUs) from OpenFace Toolkit including 20 key AUs. Also has

Semantic features such as Linguistic Inquiry and Word Count (LIWC) having 93 divisions. Only a minimal amount of features are useful and we assume the number of features to be puny enough to dodge potential overfitting.

B.3) Regression Model Building.

Data Balancing has been hugely reported that imbalanced classes of data will greatly affect the ability of machine learning and its algorithms. Regressors conduct a grid search for the below regression models: random forest regression (number of trees: 1, 10, 20, 30, 40, 50, 100, and 200), stochastic gradient descent (SGD) regression, and support vector regression (SVR) (kernel: linear, polynomial, and radial basis function (RBF)).

The Experimental setup is as follows:

1) **Test Strategy:** performing both utilization and testing on the generating set will lead to significant overfitting on the development set.

10-fold stratified cross-validation (CV): In this testing method, the training set and development set are combines and then distributed into 10 folds in a stratified manner. Every time, one-fold is used for testing and rest 9 folds are utilized for training. The model-independent attribute selection is conducted. Instead, we find the most useful model, hyper-parameters, and feature numbers in the CV test and utilize them to construct the model on the observant set.

2) Metrics: Root mean square error (RMSE) is the target; thus, all optimizations, combining model selection and feature selection, are conducted according to this metric. Mean absolute error (MAE) [19] is another metric used with RMSE to understand the dissimilarities between ground truth and prediction of test set results. Pearson correlation coefficient (CC) is an important metric to evaluate the regression performance, which can reflect the linear correlation between ground truth and prediction. F1-score computes the version of binary depression categorization.

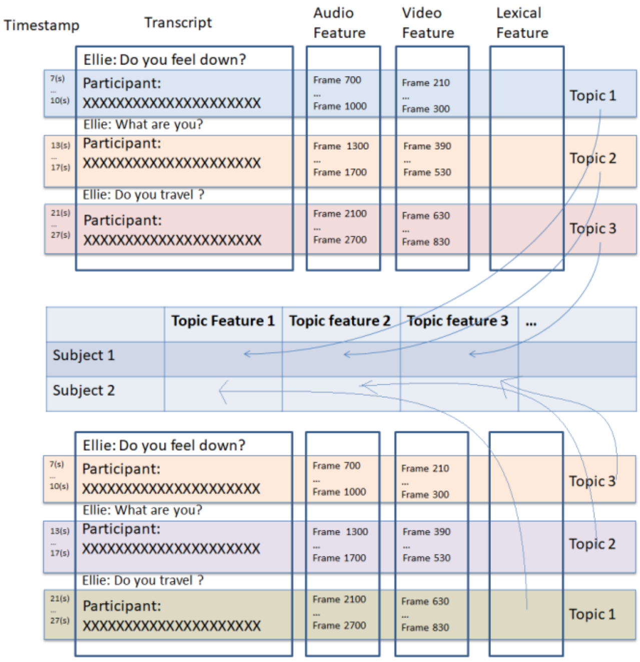

Figure 1: Representation of the proposed topic modeling based multi-modal feature vector generation scheme.

4. CONCLUSION

Thus the paper shares several techniques and sugests to conduct a questionnare based survey in detail with reference to the subjects decided. In AVEC 2016 text is evaluated on a topic level and audio/video features are distinguished and extracted and then mixed with semantic features, i.e., topic modeling is not used in these approaches However, the question/answer distillation is only applied to text analysis. Audio and video analysis is still conducted separately.

On the other hand, topic modeling is a technique to discover topics from documents that has been widely picked up in applications such as text mining and recommendation systems. *In addition, to check the effectiveness of the proposed two-step* *feature selection algorithm, we compare it with a baseline feature selection algorithm that only consists of step 2 of the proposed method, which only considers the score of each feature individually*

Therefore, we propose a novel topic modeling based approach to perform context aware analysis.

5. REFERENCES

[1] http://www.who.int/news-room/fact-heets/detail/depression-3/10/2018 [14:08]

[2] Williamson, J. R., Godoy, E., Cha, M., Schwarzentruber, A., Khorrami, P., Gwon, Y., ... & Quatieri, T. F. (2016, October). Detecting depression using vocal, facial and semantic communication cues. In *Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge* (pp. 11-18). ACM.DOI:http://dx.doi.org/10.1145/2988257.2988263

[3] Gratch, J., Artstein, R., Lucas, G. M., Stratou, G., Scherer, S., Nazarian, A., ... & Traum, D. R. (2014, May). The Distress Analysis Interview Corpus of human and computer interviews. In *LREC* (pp. 3123-3128).

[4] Gong, Y., & Poellabauer, C. (2017, October). Topic Modeling Based Multi-modal Depression Detection. In *Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge*(pp. 69-76). ACM.

[5] Lv, H. R., Lin, Z. L., Yin, W. J., & Dong, J. (2008, June). Emotion recognition based on pressure sensor keyboards. In *Multimedia and Expo, 2008 IEEE International Conference on*(pp. 1089-1092). IEEE.
DOI:10.1109/ICME.2008.4607628

[6] Yang, L., Jiang, D., He, L., Pei, E., Oveneke, M. C., & Sahli, H. (2016, October). Decision tree based depression classification from audio video and language information. In *Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge* (pp. 89-96). ACM.

[7] Yamamoto, K., Takahashi, K., Kishiro, K., Sasaki, S., & Hayashi, H. (2018, January). Analysis of emotional expression by visualization of the human and synthesized speech signal sets—A consideration of audio-visual advantage. In *Advanced Image Technology (IWAIT), 2018 International Workshop on*(pp. 1-4). IEEE.

[8] Youssofzadeh, V., Akhtar, Z., Amiri, A. M., & Falk, T. H. (2017, November). An automated framework for emotional fMRI data analysis using covariance matrix. In *Signal and Information Processing (GlobalSIP), 2017 IEEE Global Conference on*(pp. 760-763). IEEE.

[9] Dinculescu, A., Vizitiu, C., Nistorescu, A., Marin, M., & Vizitiu, A. (2015, November). Novel approach to face expression analysis in determining emotional valence and intensity with benefit for human space flight studies. In *E-Health and Bioengineering Conference (EHB), 2015* (pp. 1-4). IEEE.

[10] Hamada, R., Ohno, K., Matsubara, S., Hoshi, T., Nagasawa, M., Kikusui, T., ... & Yamakawa, T. (2017, October). Real-time emotional state estimation system for Canines based on heart rate variability. In *Cyborg and Bionic Systems (CBS), 2017 IEEE International Conference on* (pp. 298-303). IEEE.

[11] Liu, C., Xu, L., Wang, Z., & Wang, N. (2016, August). Sentiment analysis method based on context view. In Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), 2016 12th International Conference on (pp. 1577-1581). IEEE.

[12] Zhiping, L., Yu, S., & Jili, X. (2015, July). The research of user models with emotional analysis. In *Computer Science & Education (ICCSE), 2015 10th International Conference on*(pp. 455-459). IEEE.

[13] Yamamoto, K., Takahashi, K., Kishiro, K., Sasaki, S., & Hayashi, H. (2018, January). Analysis of emotional expression by visualization of the human and synthesized speech signal sets—A consideration of audio-visual advantage. In *Advanced Image Technology (IWAIT), 2018 International Workshop on*(pp. 1-4). IEEE.

[14] Byun, K. S., Park, C. H., & Sim, K. B. (2004, August). Emotion recognition from facial expression using hybrid-feature extraction. In *SICE 2004 Annual Conference* (Vol. 3, pp. 2483-2487). IEEE.

[15] James W Pennebaker, development and psychometric properties of LIWC2015. Technical report.

[16] JAN Spijker, Ron De Graaf, Rob V Bijl, Aartjan TF Beekman, Johan Ormel, and Willem A Nolen. 2002. Duration of major depressive episodes in the general population: results from.e Netherlands Mental Health Survey and Incidence Study (NEMESIS). .e British journal of psychiatry 181, 3(2002), 208–213.

[17] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation.Journal of machine Learning research 3, Jan (2003), 993–1022.

[18] Gilles Degottex, John Kane, Thomas Drugman, Tuomo Raitio, and Stefan Scherer. 2014. COVAREP: A collaborative voice analysis repository for speech technologies. In Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on. IEEE, 960–964.

[19] Fabien Ringeval, Bjo¨rn Schuller, Michel Valstar, Jonathan Gratch, Roddy Cowie, Stefan Scherer, Sharon Mozgai, Nicholas Cummins, Maximilian Schmitt, and Maja Pantic. 2017. AVEC 2017: Real-life Depression, and Affect Recognition Workshop and Challenge. In Proceedings of the 7th International Workshop on Audio/Visual Emotion Challenge. ACM, 1–8.

[20] Gilbody, S., Richards, D., Brealey, S., & Hewitt, C. (2007). Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): a diagnostic meta-analysis. *Journal of general internal medicine*, *22*(11), 1596-1602

[21] Katon, W., & Ciechanowski, P. (2002). Impact of major depression on chronic medical illness. *Journal of psychosomatic research*.

[22] Simon, G. E., & VonKorff, M. (1995). Recognition, management, and outcomes of depression in primary care. *Archives of family medicine*, *4*(2), 99.

[23] Streiner, D. L., Norman, G. R., & Cairney, J. (2015). *Health measurement scales: a practical guide to their development and use*. Oxford University Press, USA.

[24] Williams Jr, J. W., Noël, P. H., Cordes, J. A., Ramirez, G., & Pignone, M. (2002). Is this patient clinically depressed?. *Jama*, *287*(9), 1160-1170.

[25] Reitsma, J. B., Glas, A. S., Rutjes, A. W., Scholten, R. J., Bossuyt, P. M., & Zwinderman, A. H. (2005). Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of clinical epidemiology*, *58*(10), 982-990.