Data Processing Pipeline of Short-Term Depression Detection with Large-scale Dataset

Yonggeon Lee, Youngtae Noh
Energy AI, KENTECH
Jeollanam-do, Republic of Korea
{yglee, ytnoh}@kentech.ac.kr

Uichin Lee
School of Computing, KAIST
Daejeon, Republic of Korea
{uclee}@kaist.ac.kr

***Abstract*—Depression is a common, recurring mental disorder that causes significant impairment in people's lives. In recent years, ubiquitous computing using mobile phones can monitor behavioral patterns relevant to depressive symptoms in-the-wild. In this paper, we propose data processing pipeline of short-term depression detection using mobile sensor data. We build a group model classified by depression severity for capturing depressive mood in a short-period time to handle data quality and data imbalance problem in a large-scale dataset. We expect the group model to identify and characterize digital phenotype representing each depressive group as a middle step toward personalization.**

***Index Terms*—Large-scale Data Processing, Short-Term Depression Detection, Positive Computing, Mobile Computing**

I. Introduction

Depression has impacted about 25% of the world's population at least once in their lifetime, and around 7% of the population annually. In 2020, the World Health Organization (WHO) reported around 264 million people worldwide had experience of depression, and depression is expected to rapidly grow as global disease by 2030 [1]–[3]. Depressed people often have psychological difficulties, e.g., guilt, sadness as well as problems of physical symptoms and show higher suicide rates. Therefore, depression has become one of the most critical public health issue in this era.

The most commonly used methods for estimating depression in mental health field are interviews and self-report questionnaires which are time-consuming, expensive, and often require the involvement of professionals. To cope with this, passive sensing is aggressively emerging in pervasive health care. Mobile and wearable devices are capable of capturing human behavior, encompassing physical, mental, and social aspects of well-being by tracking daily activities and routines [4], [5]. An additional benefit is that passive sensing can provide short-term detection of human state and help with the prevention and long-term management of symptoms.

According to recent studies, understanding heterogeneity in passive sensing and depressive mood among individuals has become key approach, toward personalization of depression detection and its further application [6], [7]. However, personalized method is not cost-effective as it encounters problems of data quality and data imbalance in ground-truth. In this case, classifying individuals into different group by similar characteristic, e.g., depressive severity level and perform group analysis can be an alternative and recently, many attempts have been made to segment depressive levels.

Thus, we propose data processing pipeline of short-term depression detection (STDD) using a large-scale mobile sensor data in-the-wild. We perform statistical analysis on 3 different groups by depression level (normal, mild, severe) and demonstrate which set of features is more relevant with depressive mood from each group. Finally we build a machine learning model to detect depression in a short-term and improve the performance by feature selection based on correlated features.

II. Data Processing Pipeline

We recruited total of 767 participants (Android 503, iOS 264) from online social communities. We collected 3 EMAs per day (every 4 h, starting from 10 a.m. or 12 p.m.) for every 3 days and collected passive mobile sensing data from 14, 9 sensors of Android and iOS smartphone each. EMA questions are based Korean version of PHQ-9 [8] and include 4 additional questions related to stress, diet, social activity and spy-question. Each EMA response scale is re-scaled from 0-3 to 1-5, and 4 additional questions are not considered in PHQ-9 score computation. Subjects participated in the data collection for 3 months, and could extend the participation period 2 to 3 month more 1 or 2 times if they wanted. Our proposed STDD pipeline consists of four main steps: 1) Data Pre-processing 2) Feature Extraction, 3) Group Analysis, and 4) Model Training and Evaluation.

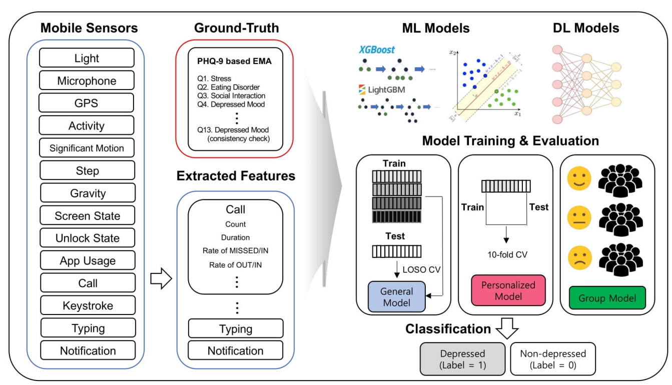

Fig. 1. STDD Data Processing Pipeline

Data Pre-processing This step involves raw sensor data processing and data cleaning. Before detailed analysis of the data, it is important to remove data points that are clearly erroneous and confirm the data quality. For evaluating data quality, first we focus on the missing data. We filter out participants by sufficient amount of data samples with continuous sensing data, e.g., ambient light and EMA based on their sampling rate. We visualize the data distribution in time series and set the cut-off for number of samples for filtering.

Next, we remove data points that are clearly erroneous or can be regarded as unreliable. We define three criteria for EMA data cleaning: multiple submissions of same EMA, response duration for EMA submission, and consistency check by spy question. We exclude EMA samples that are applicable to second response, short duration, and large PHQ-9 score gap between spy and corresponding question. For multi-dimensional sensor data, we implement outlier removal and data imputation to handle data redundancy and data loss problem. Then we refine the data and quantify event-based data e.g., screen on/off by computing counts and duration.

Feature Extraction As the goal of our research is to capture the temporary depressive mood in our daily lives, we extract features from the window length of previous 4 to 24 hours for each EMA in increments of 4 h. Larger length of window indicates that more past information is used whereas small window window focuses on recent information. For each different data source, features correlated with depression are considered based on the related studies.

In general, statistical parameters, e.g., mean, standard deviation are used as features in the case of continuous sensing data. For event-based data, number of data samples and time different between each consecutive sample are calculated for quantification. However, it is important to represent each different sensor data as features based on their distinctive characteristics. For example, features extracted from location data must consider mobility trace of an individual. These include distance travelled as well as radius of gyration, standard deviation of displacement, number of visited places etc [9].

Group Analysis Although personalized model shows best performance in depression detection, training a personalized model requires a large amount of data, especially ground truth label from each person to establish a good individual profile. For most of the passive mobile sensing studies obtaining ground-truth label is expensive and data quality problem e.g., data completeness matters due technical problems and participants engagement. In addition, there is a high chance of data imbalance on depressive mood within individuals.

In this study, we classify participants by normal, mild, and severe group based on pre-test of PHQ-9, and compare following variables: EMAs (number of EMAs, avg transformed PHQ-9 scores, label imbalance ratio), CFD (Correlated feature dynamics) defined as avg difference of each extracted feature between label, and demographics (gender, age). We conduct non-parametric statistics to find out statistically significant differences among groups using Wilcoxon ranksum test and chi-square test, since distributions of each dependent variable do not satisfy the normality.

Model Training and Evaluation Methodology The task of the study is to detect if the participant is feeling depressive mood or not. For data labeling, we sum up the scores of each 13 EMA questions and label the data with 2 classes. Here, summation larger than 15 is depressed (label 1) and summation smaller than 15 is non-depressed (label 0).

We develop group model to detect the daily depressive mood, using diverse machine learning (ML) algorithms. Total of 6 ML models including Gradient-Boosting Decision Tree (GBDT) algorithms such as XGBoost (XGB), and support vector machine (SVM) are utilized. XGB is a high-performance model combining tree-based weak prediction models, and sequentially reduces errors by using a boosting technique. From numerous studies, XGBoost has verified to be highly efficient and flexible algorithm for classification tasks. SVM uses kernels for smooth calculations by transforming the data into higher dimensions. Furthermore, we build black box model using deep learning (DL) techniques to improve detection accuracy. For the baseline, generalized model and personalized model are trained and evaluated to compare detection performance with group model.

For group model and personalized model, k-fold cross-validation is used for each depressive mood group or individual model. For generalized model, Leave-One-Subject-Out (LOSO) cross-validation is used for hyperparamter tuning and maximizing performance metric. The performance of classifiers is estimated by following metrics: *F1-score, AUC*, to handle label data imbalance.

Furthermore, we compute correlation between each feature and PHQ-9 scores using Pearson correlation coefficient and search on different relevant feature sets regarding generalized model, personalized model, and each group model. We select correlated features based on correlation analysis and train models with them to improve the detection performance.

III. Future work

From this work, we will address the key makers of short-term depression detection for 3 groups based on EMA statistics, features, and demographics. We are envisioned that the lessons learned from this study can be utilized in other mental disorder studies (e.g., bipolar disorder, stress, and so on) and applications related to depressive mood in our daily lives.

References

[1] Kessler, R.C.; Chiu, W.T.; Demler, O.; Walters, E.E. Prevalence, Severity, and Comorbidity of 12-Month DSM-IV Disorders in the National Comorbidity Survey Replication. JAMA Psychiatry 2005, 62, 617–627. [CrossRef] [PubMed] Depression.

[2] World Health Organisation (2020). URL: https://www.who.int/news-room/fact-sheets/detail/depression [accessed 2021-06-19]

[3] Lépine J, Briley M. The increasing burden of depression. Neuropsychiatr Dis Treat 2011 May;7(Suppl 1):3-7 [FREE Full text] [doi: 10.2147/NDT.S19617] [Medline: 21750622]

[4] Kasckow, J.; Zickmund, S.; Rotondi, A.; Mrkva, A.; Gurklis, J.; Chinman, M.; Fox, L.; Loganathan, M.; Hanusa, B.; Haas, G. Development of Telehealth Dialogues for Monitoring Suicidal Patients with Schizophrenia: Consumer Feedback. Community Ment. Health J. 2013, 50, 339–342. [CrossRef] [PubMed]

[5] Burns, M.; Begale, M.; Duffecy, J.; Gergle, D.; Karr, C.J.; Giangrande, E.; Mohr, D.C.; Proudfoot, J.; Dear, B. Harnessing Context Sensing to Develop a Mobile Intervention for Depression. J. Med. Internet Res. 2011, 13, e55. [CrossRef] [PubMed]

[6] Wenjun Jiang, Qi Li, Lu Su, Chenglin Miao, Quanquan Gu, and Wenyao Xu. 2018. Towards personalized learning in mobile sensing systems. In 2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS). IEEE, 321–333.

[7] Xu, Xuhai, et al. "Leveraging Collaborative-Filtering for Personalized Behavior Modeling: A Case Study of Depression Detection among College Students." Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 5.1 (2021): 1-27.

[8] An, Je Yong, et al. "Standardization of the Korean version of screening tool for depression (Patient Health Questionnaire-9, PHQ-9)." J Korean Soc Biol Ther Psychiatry 19.1 (2013): 47-56

[9] Canzian, Luca, and Mirco Musolesi. "Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis." Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing. 2015