Clustering and Feature Analysis of Smartphone Data for Depression Monitoring

Binh Nguyen, Sharadha Kolappan, Venkat Bhat, and Sridhar Krishnan

Presentation for 43rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2021)

Ryerson University
St. Michael's
Inspired Care.
Inspiring Science.

Problem Statement

- Depression is a common mental illness that affects more than 264 million people [1].
- While, there are many known methods and effective treatment options for mental disorders like depression, approximately 76% and 85% of people in low- and middle-income countries do not receive treatment, respectively [2].

Objectives

- Early detection and treatment of depression, promotes remission, reduces relapse and lowers financial burden of the disease [3]
- These benefit can be realized thanks to recent technological advancements which have given opportunity to digital phenotyping
- Analyze active and passive to monitor mental health in users

Ryerson University

St. Michael's
Inspired Care.
Inspiring Science.

Active and passive data

- Passive data
  - Collects data without user input
  - E.g GPS and Accelerometer
- Active
  - Collected data that requires user input
  - E.g Patient Health Questionnaire (PHQ) 9 survey that monitors the severity of depression

State-of-the-art

- Masud et al. [5]:
  - 33 subjects from Bangladesh
  - Collected with smartphone and PHQ-9 information
  - A support vector machine (SVM) classifier was trained using 12 features
  - Overall accuracy: 86.4% for prediction of depression levels
- Farhan et al. [6]:
  - Dartmouth StudentLife dataset
  - Collected with smartphone and PHQ-9 information
  - Wrapper method and 10-fold SVM
  - Multi-view bi-clustering (MVBC) to create behaviour subgroups
  - Overall accuracy: 87.1% for prediction of depression subgroups
- Grunerbl et al. [7]:
  - Psychiatric hospital patients
  - Collected accelerometer, sound, phone calls and GPS data
  - Overall accuracy: 76% for prediction of mental health state change

Dataset

- Dartmouth Studentlife dataset [8]
  - Smartphone data to collect passive and active data
    - Passive: physical activity, audio inferences, conversation inferences, Bluetooth scan, light sensor, global positioning system (GPS), phone charge, phone lock, WiFi, WiFi location
    - Active: PHQ-9
  - 60 college students for 10 weeks

Pre-processing

- Removal of incomplete data
    - Users that did not complete their second questionnaire

Feature extraction

- Composed of three views
  - Average
  - Trend
  - Location

Average view

- This view reflects the participant's overall behavior.
- The average view contains the average of a set of features over all the days that a participant is enrolled.

Sensor data	Feature description
Activity	Act_s, Act_w, and Act_r which represent activities for stationary, walking and running, in a day, respectively.
Conversation	$Convo_d$ and $Convo_c$ which represent conversation of total duration and number of conversations, per day, respectively.
Light	$Dark_d$ $Dark_c$ which represent the total duration and number of times when a user is in a dark environment, per day, respectively.
Audio	$Audio_q$, $Audio_n$, and $Audio_v$ which represent the duration of which user's are classified as quiet, noisy and voice, per day, respectively.
Phone lock	$PhoneLock_d$ and $PhoneLock_c$ which represent the total time for phone lock duration and the number of times phone is locked, per day, respectively.

Ryerson University

St. Michael's
Inspired Care.
Inspiring Science.

Trend view

- The trend view is calibrated to determine the variation of several sensors over the period of the study
- Pseudo steps
  - Wavelet transform
  - Fitted to least squares

$$\min_{c} \sum_{d} (f(\mathbf{c}, d) - \bar{y}_d)^2$$

$$\text{subject to} f(\mathbf{c}, x) = c_1 \sin(\frac{2\pi}{c_2} x + c_3) + c_4$$

Ryerson University

St. Michael's
Inspired Care.
Inspiring Science.

Location view

- The location view was inspired by previous works which showed a significant correlation between location and depressive mood disorder [4], [9].
- The features include
  - Location variance, time in top 3 cluster locations, entropy of (tc1; tc2; tc3) participant location, normalized entropy, percentage of time a participant is spent at their home, percentage of time a participant is spent moving, and total distance covered

Multi-view bi-clustering

- MVBC algorithm is to find subgroups of user behavior, which involves finding the same row clusters from all the view

Feature ranking

- Followed by clustering the subgroups
- Minimum redundancy maximum relevance (mRMR) was selected

$$\max \Phi(D, R), \Phi = D - R$$

$$\max D(S, c), D = \frac{1}{|S|} \sum_{x_i \in S} I(x_i; c))$$

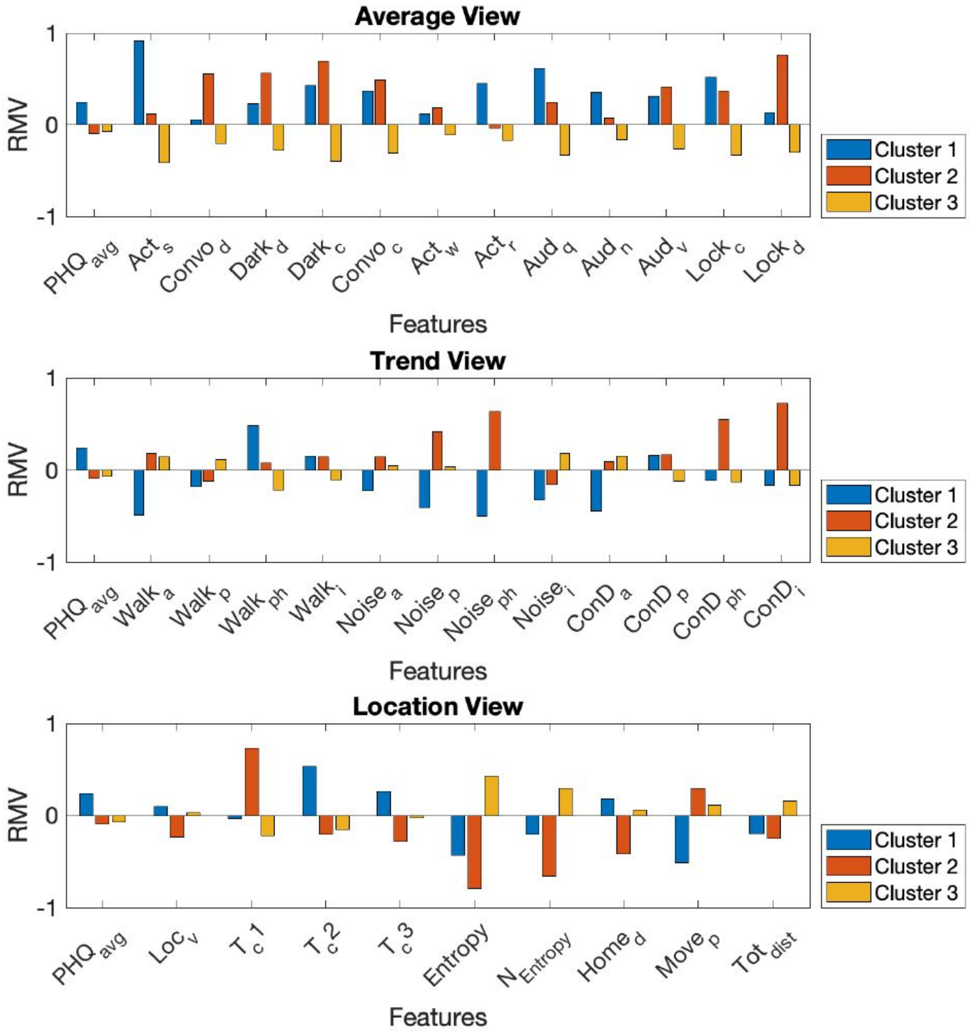
$$\min R(S), R = \frac{1}{|s|^2} \sum_{x_i, x_j \in S} I(x_i; x_j)$$

Ryerson University

St. Michael's
Inspired Care.
Inspiring Science.

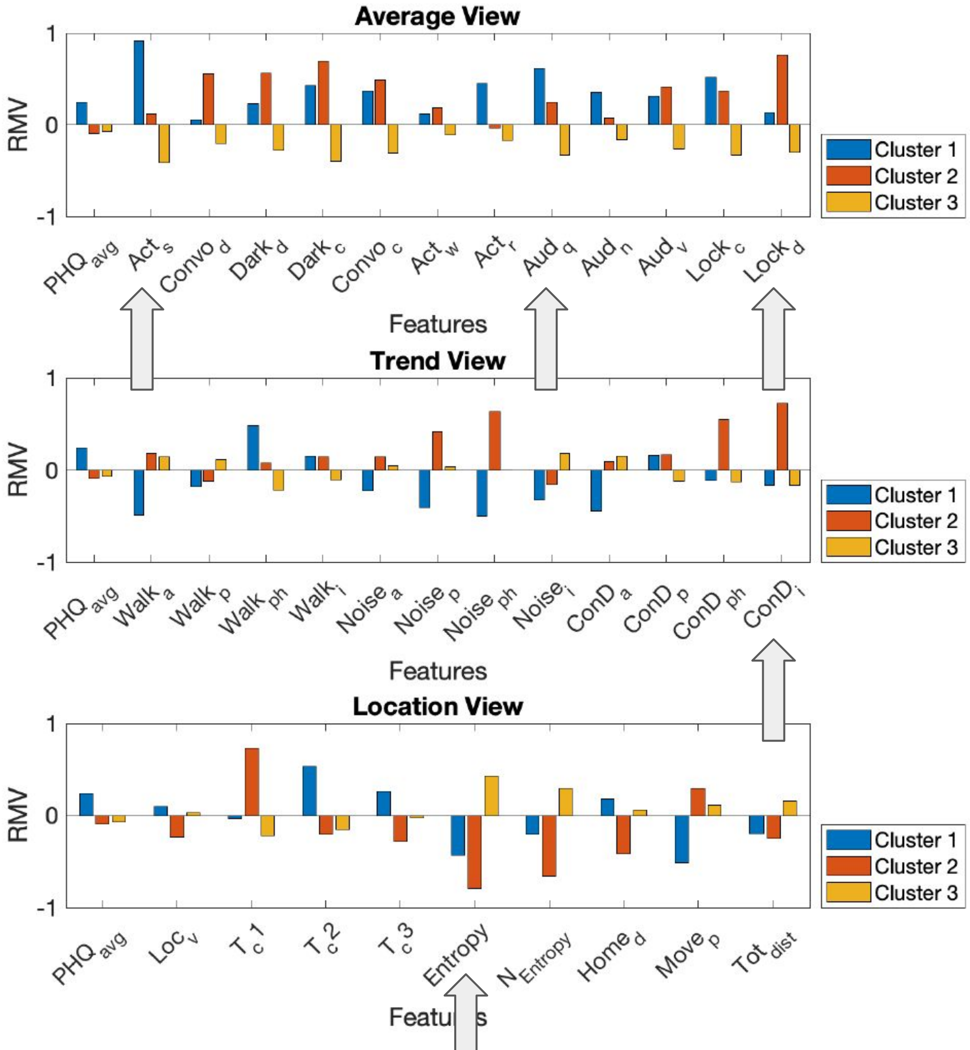
Results

- After pre-processing, 38 users were used
- We visualize the bar plot that present the relative mean values (RMV) of the PHQ-9 scores (average of pre and post) and the features for each of the three clusters



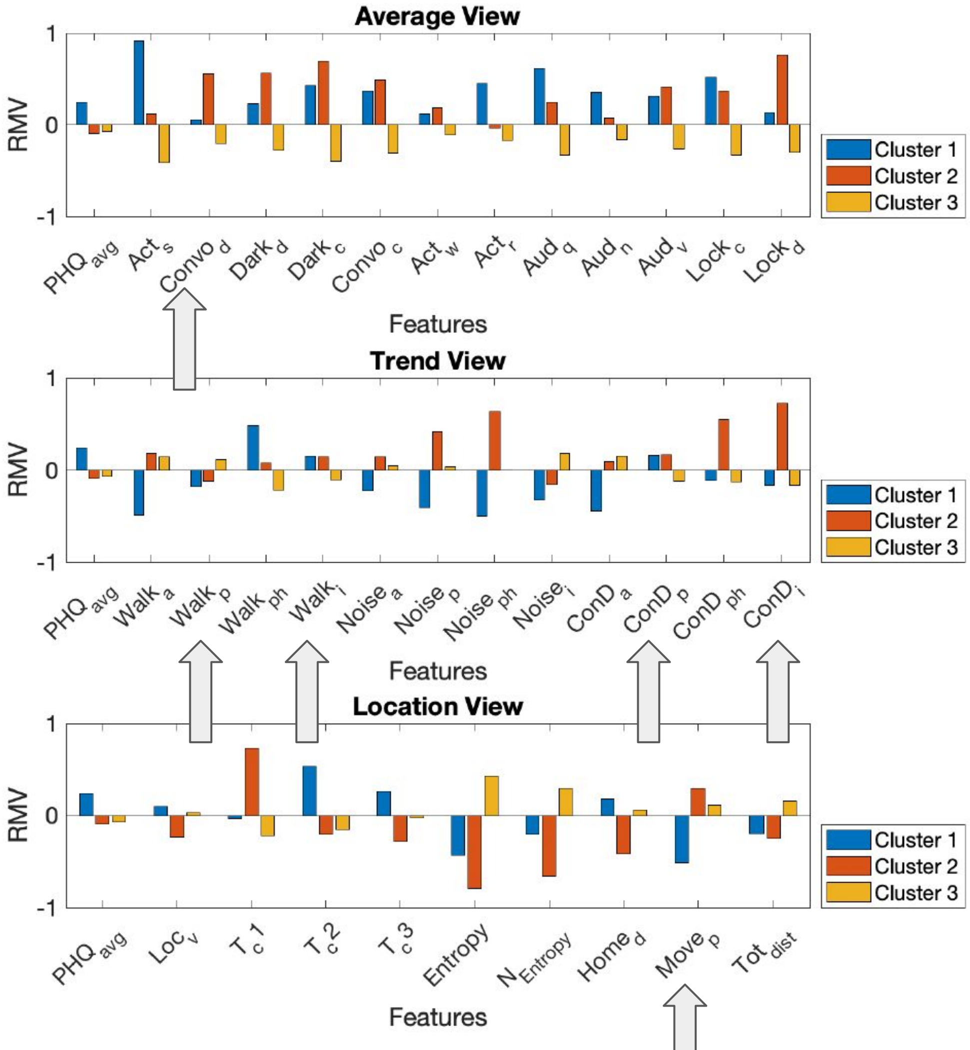
PHQ-9

- Cluster 1 (+)
- Cluster 2 (-)
- Cluster 3 (/)



PHQ-9

- Cluster 1 (+)
- Cluster 2 (-)
- Cluster 3 (/)



PHQ-9

- Cluster 1 (+)
- Cluster 2 (-)
- Cluster 3 (/)

Results continue

- Reduced feature set (n=16) achieved 94.7±1.62%

Predicted Classes

Actual Classes	C1	C2	C3
C1	9	0	0
C2	0	6	1
C3	0	1	21

Discussion and Conclusion

- The MVBC algorithm was applied to the sensing data from the smartphone followed by mRMR feature ranking was used to obtain the rank of the key features
- Of the reduced feature set, trend view seems to be the most significant as 8 of the 16 features pertain to this view.
  - Noise_i, Walk_a, Walk_ph, Noise_a, ConD_i, Walk_p, ConD_p, and ConD_a
- Due to the high accuracy with the use of a reduced feature set, detection of depression may be done locally on smartphones with future opportunities for low power secured connected healthcare and continuous human activity monitoring applications [10], [11].

References

[1] S. L. James, D. Abate, K. H. Abate, S. M. Abay, C. Abbafati, N. Abbasi, and et al., “Global, regional, and national incidence, prevalence, and years lived with disability for 354 Diseases and Injuries for 195 countries and territories, 1990-2017: A systematic analysis for the Global Burden of Disease Study 2017,” The Lancet, vol. 392, pp. 1789–1858, nov 2018.
[2] P. S. Wang, S. Aguilar-Gaxiola, J. Alonso, M. C. Angermeyer, G. Borges, E. J. Bromet, and et al., “Use of MH services for anxiety, mood, and substance disorders in 17 countries in the WHO world MH surveys,” Lancet, vol. 370, pp. 841–850, sep 2007.
[3] “Depression: the importance of diagnosis and prevention - ProQuest.”
[4] L. Canzian and M. Musolesi, “Trajectories of depression: Unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis,” in UbiComp 2015 - Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, (New York, New York, USA), pp. 1293–1304, Association for Computing Machinery, Inc, sep 2015
[5] M. T. Masud, M. A. Mamun, K. Thapa, D. H. Lee, M. D. Griffiths, and S. H. Yang, “Unobtrusive monitoring of behavior and movement patterns to detect clinical depression severity level via smartphone,” Journal of Biomedical Informatics, vol. 103, mar 2020.
[6] A. A. Farhan, J. Lu, J. Bi, A. Russell, B. Wang, and A. Bamis, “Multiview Bi-clustering to Identify Smartphone Sensing Features Indicative of Depression,” in Proceedings - 2016 IEEE 1st International Conference on Connected Health: Applications, Systems and Engineering Technologies, CHASE 2016, pp. 264–273, Institute of Electrical and Electronics Engineers Inc., aug 2016.
[7] R. Wang, F. Chen, Z. Chen, T. Li, G. Harari, S. Tignor, X. Zhou, D. Ben-Zeev, and A. T. Campbell, “Studentlife: Assessing MH, academic performance and behavioral trends of college students using smartphones,” UbiComp 2014 - Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 3–14, 2014.
[8] A. Gru¨nerbl, A. Muaremi, V. Osmani, G. Bahle, S. O¨ hler, G. Tro¨ster, O. Mayora, C. Haring, and P. Lukowicz, “Smartphone-based recognition of states and state changes in BPD patients,” IEEE Journal of Biomedical and Health Informatics, vol. 19, pp. 140–148, jan 2015.
[9] S. Saeb, M. Zhang, C. J. Karr, S. M. Schueller, M. E. Corden, K. P. Kording, and D. C. Mohr, “Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: An exploratory study,” Journal of Medical Internet Research, vol. 17, p. e175, jul 2015.
[10] B. Nguyen, Y. Coelho, T. Bastos, and S. Krishnan, “Trends in HAR with focus on machine learning and power requirements,” Machine Learning with Applications, vol. 5, p. 100072, sep 2021.
[11] S. Krishnan, “Biomedical signal analysis for connected healthcare,” Elsevier.