Evolving AI for Wellness: Dynamic and Personalized Real-time Loneliness Detection Using Passive Sensing*

Malik Muhammad Qirtas[1][0000-0001-7644-161X], Evi Zafeiridi[2][0000-0001-7986-5442], Eleanor Bantry White[3][0000-0002-7663-6836], and Dirk Pesch[4][0000-0001-9706-5705]

[1] School of Computer Science and Information Technology, University College Cork, Ireland
malik.qirtas@cs.ucc.ie
[2] School of Computer Science and Information Technology, University College Cork, Ireland
EZafeiridi@ucc.ie
[3] School of Applied Social Studies, University College Cork, Ireland
E.BantryWhite@ucc.ie
[4] School of Computer Science and Information Technology, University College Cork, Ireland
dirk.pesch@ucc.ie

Abstract. Loneliness among college students is a growing concern, as it can lead to depression and other associated mental health problems. This paper proposes an innovative approach for loneliness detection using passive sensing, a method that uses smartphone and wearable sensor data to capture daily behavioural patterns. Given the subjective nature of loneliness and the varying daily routines of students, past detection approaches using machine learning models often face challenges with effectively detecting loneliness. The primary objective of our work is the development of a detection system that evolves over time, adapts to new data, and provides real-time loneliness detection among students. The base of our approach is the continuous identification and refinement of behavioural groups among students using an incremental clustering method. As we add new data, the model improves based on changing behavioural patterns. Parallel to this, we create and update classification models to detect loneliness among the evolving groups of students. When unique behavioural patterns are observed, specialized classification models are devised. For predictions of loneliness, a collaborative effort between the generalized and specialized models is employed, treating each prediction as a vote. The final classification is decided through either a majority vote or a weighted approach that takes into account the confidence of each model. This system's flexibility to adapt to students' behaviours improves its accuracy in identifying loneliness, highlighting the value of passive sensing data and the need for personalized and real time mental health detection systems.

Keywords: Grouping · Loneliness · Mobile sensing · Passive sensing · Smartphone

1 Introduction

Loneliness is a growing global issue, with many people reporting it as a primary source of their unhappiness [1]. Loneliness is an experience in which a person perceives a lack of quality social relationships [2]. Loneliness can lead to a variety of health problems, including difficulty in sleeping, increased anxiety, persistent sadness, and a reduced immune response. While many people feel lonely

* This publication has emanated from research conducted with the financial support of Science Foundation Ireland under Grant number 18/CRT/6222

at times, it becomes a concerning issue when it lasts for an extended period of time [3]. Loneliness and mental health are closely linked. People with mental health issues are over twice as likely to feel lonely compared to those without such issues [4]. Given the widespread rise of loneliness following the COVID-19 pandemic and its negative effects, it is crucial to detect loneliness early to reduce its potential harm.

The increasing use of smartphones and wearables has opened up new avenues for continuous and unobtrusive behavioural monitoring of individuals. Their ubiquitous nature and built-in sensors have made them effective tools for tracking user behaviours and daily routines [5]. Passive sensing, an innovative technique where smartphones and wearables collect data without requiring user intervention, is leading this transformation. Such continuous data collection offers a more comprehensive view of an individual's behaviour patterns over time. When combined with established clinical scales for mental health detection, these digital footprints can be processed into biomarkers of mental health [6]. Specifically, by analyzing these digital biomarkers, researchers can extract signs of loneliness in real-time. As an unintrusive and economical method, passive sensing holds immense promise for revolutionizing the early detection of conditions such as loneliness, depression, or anxiety, opening avenues for more tailored and personalized interventions.

In this paper, we propose a dynamic and personalized approach for real-time loneliness detection using passive sensing. The basic idea of this paper is based on our earlier work [7] but with a modified version. At the core of our methodology are two intertwined components: Incremental Clustering and Incremental Classification. The clustering will be used to detect subgroups of students with the same behavioural patterns. This will continuously integrate new data into the existing model, refining behavioural groups over time. The incremental classification part will identify lonely individuals within these continually refining student groups.

In parallel, We will be updating our main classification model. When new behavioural patterns appear, we will create a specific specialised model for them. For predicting loneliness, we will be using a Multi-Model Voting system that consider predictions from both the main and specialized models. Each prediction is like a 'vote'. The final decision will be based on either most votes or a system that weighs the trustworthiness of each model.

Our approach to detecting loneliness in college students through group-based behavioural analysis opens doors to many practical uses. Universities can integrate this method into their mental health applications or support systems. The same technique could be vital in spotting signs of other mental health issues, such as depression, giving professionals a head start in providing help. In a broader context, businesses could use this strategy to better understand and support their employees' mental well-being. For older adults, especially those isolated, it offers a means to detect early signs of loneliness, prompting timely care. This paper presents a new way to understand student loneliness using behavioral groups.

2 Methodology

2.1 Dataset

We will be using a dataset from a U.S. university, collected during the Spring quarter of 2019 over 10 weeks from March to June [10]. This period was chosen to consider any seasonal influences. The dataset, named DS-2, includes information from 218 undergraduate students, majority females, who joined through email and social media invitations. Please refer to Table 1 for detailed information about participants' demographics.

Table 1. Study Information and Participant Demographics. Gender acronyms: F: Female, M: Male, NB: Non-binary. Racial acronyms: A: Asian, B: Black or African American, H: Hispanic, N: American Indian/Alaska Native, PI: Pacific Islander, W: White, NA: Did not report. The & symbol denotes participants who identified with multiple races.

Category	Data
Participants	Total: 218
	Gender: F 111, M 107
	Ethnicity: A 102, B 6, H 10, N 2, PI 1, W 70,
	A&B 1, A&W 16, H&W 2, B&W 2,
	A&H&W 1, B&H&W 1, H&N&W 1, NA 3
Ground Truth	Pre-study 10-items UCLA scale
	Post-study 10-items UCLA scale
Sensor Data	Bluetooth, WIFI, Call Logs, Location, Phone Usage, Physical Activity, Sleep

For the collection process, the AWARE smartphone application was employed, available for both iOS and Android platforms. This app operates continuously in the background without requiring active user interaction [8]. Additionally, a Fitbit was used to gather sleep and physical activity data. To offer an in-depth understanding of these patterns, the study integrated readings from various sensors, including accelerometer, gyroscope, microphone, and light level sensors.

On the ethical side, the study prioritized participant privacy. The data collection initiative received ethical approval from the University of Washington's IRB (IRB number:STUDY00003244), and every participant gave informed consent. Adhering strictly to anonymization protocols ensured data confidentiality, with direct identifiers limited to the core data team. Moreover, any data from participants choosing to withdraw was immediately removed from the database.

The study used the revised 10-item UCLA loneliness scale to assess loneliness [11] and to label the data. Participants completed this questionnaire at the beginning and end of the study, rating questions on a scale of 1 ("never") to 4 ("always"). After reverse scoring five items, the scores were combined, resulting in a total score ranging between 10 to 40. Higher scores suggest increased feelings of loneliness.

2.2 Data Preprocessing

During preprocessing, we refined our dataset to include only the data from 205/218 students who completed the post-study loneliness questionnaire, covering multiple sensors over a ten-week period. By converting UNIX timestamps using participants' timezone data, we categorized a day into three sessions: day (9am-6pm), evening (6pm-12am), and night (12am-9am).

Our post-processed dataset organized participants into two categories based on the UCLA loneliness scale: those scoring above 20, termed 'lonely' (87 students), and the rest (118 students).

For data cleaning, missing continuous values will be imputed using median values specific to the session. Categorical data missing values will be replaced by the mode for that session. Using the Reproducible Analysis Pipeline for Data Streams (RAPIDS) [9], we will derive digital biomarkers to quantify behavioural patterns in students, capturing routines, irregularities, and variabilities. To optimize clustering, we will apply principal component analysis (PCA) for dimensionality reduction.

2.3 Incremental Clustering

To prepare our data for clustering, we will normalize each feature to ensure uniform value ranges, enabling each feature to equally influence distance calculations and facilitating improved cluster formation.

Among various clustering algorithms, we will use IDBSCAN algorithm [12], given its relevance in our context of incremental clustering. IDBSCAN, an extension of the traditional DBSCAN, is tailored for incremental clustering. Its strength is in seamlessly adjusting to new data, adapting to changes without needing to fully retrain the model. IDBSCAN works by maintaining a structure that reflects the concentration of the data. When new data comes in, it checks it against current clusters. Based on closeness and density, the data either joins an existing cluster, gets labelled as noise, or starts a new cluster.

2.4 Incremental Classification

Our primary objective with incremental classification is to dynamically detect individuals experiencing loneliness within subgroups identified during the incremental clustering step, adapting to their shifting behavioural patterns over time. Our main classification model will need incremental updates with new data. As we will analyze behaviour more deeply, spotting unique patterns, especially from our clustering process, will lead to the development of specialized models for those specific patterns.

For prediction tasks, we will not rely solely on a singular model but will be using insights from both the primary and the specialized models. This combined method will use a broad understanding of our main model and the specific accuracy of our specialized ones. By blending these techniques, we aim to improve the accuracy of our loneliness detection, allowing it to effectively address the diverse aspects of human behaviour.

2.5 Multi-Model Voting

One of the key components in our proposed approach is the incorporation of Multi-Model Voting. This approach is rooted in the belief that combining insights from multiple models increases the reliability and consistency of our classification outcomes.

The main goal is to use insights from several models to get reliable classification results, even when faced with irregular data or outliers. By relying on more than one model, we aim to overcome the limitations of individual models, making sure our findings aren't overly affected by any single perspective.

The methodology unfolds in two main stages:

Collecting Predictions: When new data comes in, it is crucial to get insights not only from the primary model but also from specialized models designed for specific behavioral patterns noticed during the clustering phase. For instance, consider two models: one trained on data from all students and another specifically trained on data from students who frequently use their phones at night. For a new student who has high evening phone usage, the specialized model might predict a higher likelihood of loneliness than the primary model. So each model will be providing a different prediction, based on its distinct training background and emphasis.

Vote Aggregation: After gathering predictions, they act like 'votes'. We will determine the final classification by combining these votes. Although a straightforward majority might work in many situations, sometimes a more detailed method is needed. For example, a weighted voting system might give more importance to votes based on how confident a prediction is or the past accuracy of the model. This recognizes that some models may offer more reliable predictions than others.

3 Expected Results

Through our detailed methodology, we aim to uncover deep-rooted behavioural patterns of the study participants. The choice of IDBSCAN should allow us to dynamically group data, offering insights into distinct behavioural subgroups related to feelings of loneliness. With the integration of our main and specialized classification models, we hope to achieve enhanced accuracy in identifying lonely individuals across diverse behavioural groups. By employing a multi-model voting system, we anticipate achieving more consistent and reliable classifications, minimizing the risk of errors that might arise from data anomalies or biases.

References

1. J. T. Cacioppo and W. Patrick, *Loneliness: Human nature and the need for social connection*, WW Norton & Company, 2008.
2. L. A. Peplau and S. E. Goldston, *Preventing the Harmful Consequences of Severe and Persistent Loneliness: Proceedings of a Research Planning Workshop Held in Cooperation with the Department of Psychology, University of California, Los Angeles, February 10-12, 1982*, vol. 5, US Department of Health and Human Services, Public Health Service, Alcohol..., 1985.
3. M. Campbell, *Loneliness, social anxiety and bullying victimization in young people: A literature review*, Psychology and Education, vol. 50, no. 3-4, pp. 1–10, 2013, Springer Netherlands.
4. M. E. Beutel, E. M. Klein, E. Brähler, I. Reiner, C. Jünger, M. Michal, J. Wiltink, P. S. Wild, T. Münzel, K. J. Lackner, et al., *Loneliness in the general population: prevalence, determinants and relations to mental health*, BMC Psychiatry, vol. 17, no. 1, pp. 1–7, 2017, BioMed Central.
5. V. P. Cornet and R. J. Holden, *Systematic review of smartphone-based passive sensing for health and wellbeing*, Journal of Biomedical Informatics, vol. 77, pp. 120–132, 2018, Elsevier.
6. M. M. Qirtas, E. Zafeiridi, D. Pesch, and E. B. White, *Loneliness and social isolation detection using passive sensing techniques: scoping review*, JMIR mHealth and uHealth, vol. 10, no. 4, pp. e34638, 2022, JMIR Publications Inc., Toronto, Canada.
7. M. M. Qirtas, E. B. White, E. Zafeiridi, and D. Pesch, *Personalising Loneliness Detection through Behavioural Grouping of Passive Sensing Data from College Students*, IEEE Access, 2023, IEEE.
8. D. Ferreira, V. Kostakos, and A. K. Dey, *AWARE: mobile context instrumentation framework*, Frontiers in ICT, vol. 2, pp. 6, 2015, Frontiers Media SA.
9. J. Vega, M. Li, K. Aguillera, N. Goel, E. Joshi, KC. Durica, AR. Kunta, and CA. Low, *Rapids: Reproducible analysis pipeline for data streams collected with mobile devices*, J Med Internet Res Preprints. URL: https://preprints.jmir.org/preprint/23246 [accessed 2020-08-18], 2020.
10. X. Xu, H. Zhang, Y. Sefidgar, Y. Ren, X. Liu, W. Seo, J. Brown, K. Kuehn, M. Merrill, P. Nurius, et al., *GLOBEM Dataset: Multi-Year Datasets for Longitudinal Human Behavior Modeling Generalization*, Advances in Neural Information Processing Systems, vol. 35, pp. 24655–24692, 2022.
11. R. G. Knight, B. J. Chisholm, N. V. Marsh, H. P. D. Godfrey, *Some normative, reliability, and factor analytic data for the revised UCLA Loneliness Scale*, Journal of Clinical Psychology, vol. 44, no. 2, pp. 203–206, 1988.
12. N. Goyal, P. Goyal, K. Venkatramaiah, P. C. Deepak, P. S. Sanoop, *An efficient density based incremental clustering algorithm in data warehousing environment*, In Proceedings of the 2009 International Conference on Computer Engineering and Applications, 2009.