# Loneliness Episodes: A Japanese Dataset for Loneliness Detection and Analysis

**Naoya Fujikawa** and **Quang Toan Nguyen** and **Kazuhiro Ito**
**Shoko Wakamiya** and **Eiji Aramaki**
Nara Institute of Science and Technology, Japan
fujikawa.naoya.fi8@is.naist.jp, nqt.quangtoan@gmail.com
{ito.kazuhiro.ih4, wakamiya, aramaki}@is.naist.jp

## Abstract

Loneliness, a significant public health concern, is closely connected to both physical and mental well-being. Hence, detection and intervention for individuals experiencing loneliness are crucial. Identifying loneliness in text is straightforward when it is explicitly stated but challenging when it is implicit. Detecting implicit loneliness requires a manually annotated dataset because whereas explicit loneliness can be detected using keywords, implicit loneliness cannot be. However, there are no freely available datasets with clear annotation guidelines for implicit loneliness. In this study, we construct a freely accessible Japanese loneliness dataset with annotation guidelines grounded in the psychological definition of loneliness. This dataset covers loneliness intensity and the contributing factors of loneliness. We train two models to classify whether loneliness is expressed and the intensity of loneliness. The model classifying loneliness versus non-loneliness achieves an F1-score of 0.833, but the model for identifying the intensity of loneliness has a low F1-score of 0.400, which is likely due to label imbalance and a shortage of a certain label in the dataset. We validate performance in another domain, specifically X (formerly Twitter), and observe a decrease. In addition, we propose improvement suggestions for domain adaptation.

## 1 Introduction

Loneliness has become a major global concern, affecting both mental and physical well-being. Previous research shows that loneliness constitutes a significant risk factor for both coronary heart disease and stroke (Valtorta et al., 2016). Another study shows that loneliness is a risk factor for morbidity and mortality (Luo et al., 2012). It also increases health-risk behaviors (Shankar et al., 2011). Hence, addressing the issue of loneliness is crucial.

The degree of negative impact varies depending on the intensity of loneliness. Beutel et al. (2017) indicate that the greater the intensity of loneliness, the higher the proportion of individuals experiencing depression and suicidal ideation. Furthermore, it is noteworthy that over half of individuals experiencing strong loneliness report depressive symptoms. Lee et al. (2019) demonstrate that people who feel more loneliness have lower resilience, optimism, and mental well-being. For these reasons, classifying the intensity of loneliness is as important for intervention as discovering individuals experiencing loneliness.

Identifying loneliness and intervening for these people are crucial. People who feel lonely tend to have less contact with supportive family and friends than those who do not, and they often use social media more frequently (Lampraki et al., 2022). Numerous previous studies have created corpora using keyword-based approaches on social media (Andy et al., 2022; Kiritchenko et al., 2020). These corpora only considered loneliness when it was explicitly mentioned, using specific keywords to label posts. This limits finding posts where loneliness is stated implicitly rather than explicitly, possibly leading to underreporting because users hesitate to share loneliness posts due to social stigma around admitting loneliness. An example of loneliness stated explicitly is “I feel lonely and isolated at work,” where the explicit expressions “isolate” and “lonely” are included. In contrast, an example of loneliness stated implicitly is “I am being ignored by various people at school,” where there is no direct expression, but loneliness can be inferred from “being ignored.”

Prior work has created a dataset that could also encompass cases where loneliness was implicitly expressed. A recent study collected text and psychological characteristics (Nakai et al., 2023), but this text is not publicly available. On the contrary, a publicly available dataset has been pub-



*Proceedings of the 14th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis*, pages 280–293
August 15, 2024 ©2024 Association for Computational Linguistics

lished (Jiang et al., 2022). However, their study has two limitations: the annotation process relies on subjectivity due to the absence of a clear definition of loneliness, and they did not differentiate by intensity.

This study aims to provide a publicly available Japanese loneliness dataset with clear annotation guidelines. We created annotation guidelines based on the psychological definition of loneliness and annotated a corpus that is publicly available. Moreover, we labeled whether loneliness is expressed, its intensity in detecting people suffering from severe loneliness, and the contributing factors.

To create the loneliness dataset, we used a Japanese short episode corpus[1], called LIFE STORY. This corpus is composed of episodes recalled from emotions collected through crowdsourcing. In line with previous work, creating datasets using social media would become difficult to utilize if API regulations change. Considering that the LIFE STORY corpus is freely available, it is advantageous for research from the perspective of reproducibility and ease of use.

Our contributions are as follows:

- We built a Japanese loneliness dataset[2] by annotating an episode corpus, which is created for the detection and analysis of loneliness (Section 3);
- We constructed classifiers to determine whether loneliness is expressed and to assess the intensity of loneliness using our dataset (Section 4);
- We indicated the feasibility of domain adaptation, employing posts from X (Section 5).

## 2 Related Work

### 2.1 Definition of Loneliness and Social Needs

Many definitions of loneliness have been proposed (Taylor, 2020; Sullivan, 1953; Ma et al., 2020). However, they all share three important common points (Peplau and Perlman, 1982). First, loneliness arises from inadequacies in an individual's social connections. Second, loneliness is subjective; it does not equate to objective social isolation. It is possible for an individual to be alone without experiencing loneliness, and conversely, one can feel lonely even when surrounded by a crowd. Third, loneliness is uncomfortable and distressing. The differences among various definitions of loneliness are due to the nature of social deficiency. One approach emphasizes social needs, and another approach emphasizes cognitive processes. The social needs approach suggests that individuals may experience loneliness without explicitly identifying themselves as lonely or consciously understanding the nature of their distress. In contrast, the cognitive approach focuses on the perceptions and self-reports of loneliness, paying attention to those who recognize themselves as lonely.

For interventions in loneliness, it is important to identify not only those who recognize their loneliness but also those who are emotionally distressed without recognizing it as loneliness. Therefore, when defining loneliness, we prioritize the social needs aspect. We introduced the definition of loneliness presented by Hawkley and Cacioppo (2010).

Social needs are conceptualized in various ways (Deci and Ryan, 2008; Lindenberg, 1996; Steverink et al., 2020). We have chosen the approach by Ormel et al. (1999), which is one of the most representative, and is composed of *Affection*, *Behavioral Confirmation*, and *Status*.

### 2.2 Loneliness Dataset

Research on constructing datasets related to implicit loneliness is limited (Nakai et al., 2023; Jiang et al., 2022). Nakai et al. (2023) tried to predict psychological states including loneliness from texts describing eating experiences collected using crowdsourcing with 877 individuals. They collected text data on meal experiences, satisfaction levels with meals, and psychological characteristics, then created a classifier using BERT. The loneliness scores they gathered are not about how lonely the texts express but about how lonely the writers usually feel; thus, loneliness related to the text is not assigned. Additionally, this dataset has not been made publicly available, which motivates this study to construct an openly available dataset.

Jiang et al. (2022) constructed a loneliness dataset from posts on Reddit. To create a corpus, they collected posts from two loneliness-related subreddits (r/loneliness, r/lonely) and two subreddits targeting young adults (r/youngadults, r/college). For each post in the corpus, three anno-

[1] https://sociocom.naist.jp/life-story-data/
[2] https://github.com/sociocom/Japanese-Loneliness-Dataset

tators determined whether it expressed loneliness. Posts judged as expressing loneliness were further annotated with duration, situation, interpersonal relationships, and interactions. However, the criteria for determining whether a post expresses loneliness are not clearly defined. In previous studies, the definition of loneliness for dataset construction has traditionally relied on the subjective judgment of annotators. In contrast, we attempted annotation based on a definition of loneliness that allows for objective interpretation. We believe that this initial attempt is crucial for detecting loneliness to prevent serious conditions such as depression and suicide.

# 3 A Japanese Dataset for Loneliness

## 3.1 Corpus

This study leverages a Japanese short episode corpus, called LIFE STORY[3], for constructing the Japanese dataset. The LIFE STORY corpus, which is freely available, has been continuously collected since 2017, offering age, gender, and open-ended Japanese episodes associated with seven primary emotions: sadness, anxiety, anger, disgust, trust, surprise, and joy.

We focused on sadness and anxiety as emotions related to loneliness (Cacioppo et al., 2010; Mullarkey et al., 2018; Meltzer et al., 2012). In a preliminary study, we evaluated 50 episodes of each emotion to determine if they expressed loneliness based on the criteria specified in the annotation guidelines (details in Section 3.2.2). For sad episodes, out of 50 episodes, one annotator classified 12 episodes as expressing loneliness, while the other annotator classified 16 episodes. For anxious episodes, one annotator classified 1 episode as expressing loneliness, and the other annotator classified 0 episodes. Due to the very low relationship observed between anxiety and loneliness in this sample, we chose to annotate only sad episodes when creating the corpus. We annotated sad episodes extracted from the LIFE STORY corpus. Examples translated from Japanese of such episodes include: "*I couldn't purchase the desired item at the auction*" and "*I had to decline my friend's invitations because I was short on money during Golden Week.*"

[3] https://sociocom.naist.jp/life-story-data/

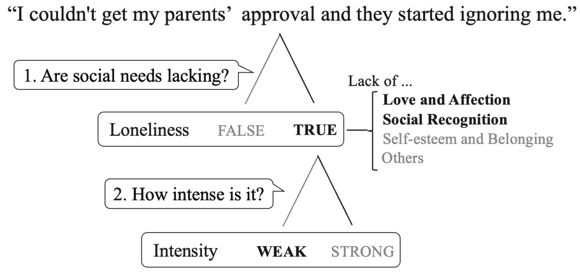


Figure 1: Overview of the annotation process. We annotated the episodes in the LIFE STORY corpus in the order of loneliness or non-loneliness, and the intensity of loneliness. We also classified unsatisfied social needs as loneliness factors within the step of loneliness or non-loneliness. This episode is labeled as *TRUE* due to a lack of *Love and Affection* and *Social Recognition*, and is labeled as weak in terms of the intensity of loneliness.

## 3.2 Label Definitions

### 3.2.1 Overview

We annotated the episodes in the LIFE STORY corpus with the following two steps:

**Step 1** Loneliness or Non-Loneliness: We first labeled episodes to determine whether they express loneliness or not. Loneliness episodes were labeled as *TRUE*, while non-loneliness episodes were labeled as *FALSE*. Additionally, we also labeled unsatisfied social needs as loneliness factors. Social needs include *Love and Affection*, *Social Recognition*, and *Self-esteem and Belonging*.

**Step 2** Intensity of Loneliness: For loneliness episodes, we further categorized the intensity levels of such expressions. The intensity of loneliness is either *STRONG* or *WEAK*.

Figure 1 illustrates the hierarchical labeling structure. For example, consider the episode: "*I couldn't get my parents' approval and they started ignoring me.*" This is labeled as *TRUE* for step 1 and *WEAK* for step 2 based on the criteria (details in Sections 3.2.2 and 3.2.3). Loneliness factors are labeled as unsatisfied social needs for *Love and Affection* and *Social Recognition*.

### 3.2.2 Loneliness or Non-Loneliness

This section details the annotation scheme used to identify loneliness episodes and categorize their loneliness factors. We adopted the definition of loneliness proposed by Hawkley and Cacioppo

(2010): *Loneliness is defined as a distressing feeling that accompanies the perception that one's social needs are not being met by the quantity or especially the quality of one's social relationships.*

Based on the definition, episodes in which any one of the social needs is not satisfied are judged to express loneliness. In other words, a single unsatisfied social need can lead to the judgment of loneliness. In psychology, it has been proposed that social needs are composed of *Affection*, *Behavioral Confirmation*, and *Status* (Ormel et al., 1999). For the purpose of this study, we have slightly adjusted these categories to better relate to loneliness and have classified them into three categories: *Love and Affection*, *Social Recognition*, and *Self-esteem and Belonging*. *Status* can be considered as a factor contributing to loneliness. However, when looking at the definitions, we can see that the definition of *Status* is actually derived from our definition of social needs rather than directly causing loneliness itself. Therefore, in this study we consider that the source of loneliness from *Status* arises from a lack of *Self-esteem and Belonging*, and *Social Recognition*. Below is a description of the social needs employed in this study:

**Love and Affection** This involves receiving affection from people one cares about (e.g., family, friends). It is derived from *Affection* (one of the social needs proposed by Ormel et al. (1999)), with the addition of "love" to enhance understanding. An example of a lack of *Love and Affection* is as follows: "*I had a disagreement with my parents and lost touch with them.*"

**Social Recognition** This involves receiving external validation and acceptance through one's behavior in a social environment. It is derived from *Behavioral Confirmation* (one of the social needs proposed by Ormel et al. (1999)), which is the sense of approval by others. It depends on external evaluations. An example of a lack of *Social Recognition* is as follows: "*Many of my friends at school are fashionable, but I don't have the money to buy a variety of clothes, so I can't be fashionable and join in the conversation.*"

**Self-esteem and Belonging** This involves the internal sense of being accepted and valued within a group or society. It is also rooted from *Behavioral Confirmation*. It depends on internal evaluations. An example of a lack of *Self-esteem and Belonging* is as follows: "*I feel out of place because I am the only one without a Ph.D.*"

Furthermore, we introduced an *Others* label to address situations where none of the specified unsatisfied social needs were identified, yet loneliness was perceived. The example is "*It's sad to spend every day feeling lonely and in a depressed state.*" If an episode does not lack any of the social needs, we assigned only a *FALSE* label for loneliness.

To ensure consistent labeling within the dataset, we added two criteria for episodes labeled as *FALSE*. First, episodes where loneliness arises from an external source rather than the writer themselves are classified as *FALSE*. The example is "*My wife said she doesn't fit in at work and feels lonely.*" Second, episodes lacking explicit vocabulary related to social connection are also labeled as *FALSE*. On this criterion, the example of *TRUE* is "*I have a disagreement with my parents and have lost touch with them*" because it involves both explicit social connection vocabulary ("parent" and "lost touch with") and expresses loneliness. In contrast, the example of *FALSE* is "*I had a dream about a sad event from the past, and the sadness came back to me when I remembered it*".

#### 3.2.3 Intensity of Loneliness

We also assigned the intensity of loneliness to episodes classified as *TRUE*, distinguishing between two levels: *STRONG* and *WEAK*. *STRONG* denotes situations in which loneliness markedly disrupts daily life or is explicitly accompanied by the expression of negative emotions. In contrast, all episodes that are not labeled *STRONG* are labeled *WEAK*. *STRONG* episodes include sentences such as "*I can't work since my parents passed away*" and "*I can't sleep at night because my beloved dog died.*" Conversely, *WEAK* episodes include sentences such as "*My family is busy with work and I'm lonely*" and "*I was sad when I had to transfer schools and say goodbye to my friends.*"

### 3.3 Dataset Construction

Two annotators independently labeled the data. We preprocessed the sad episodes excerpted from the LIFE STORY corpus (May and August 2023 surveys) by removing noises (# and * symbols, which are used in preprocessing the LIFE STORY corpus) and excluding texts shorter than or equal to 10 characters. We also converted text to lowercase, normalized Unicode using NFKC, and replaced numbers with 0. The magnitude of numbers can be important for understanding loneliness; however, NLP models often struggle to handle these numbers effectively. Therefore, numbers were not used as a feature in this model.

Each annotator labeled a total of 600 episodes including 200 common episodes for inter-annotator agreement calculation. In cases where annotators encountered difficulty determining loneliness or non-loneliness, they labeled it as *TRUE* to prioritize recall. Moreover, when faced with uncertainty regarding the intensity, annotators assigned a label of *WEAK* to prioritize the precision of *STRONG*.

Our dataset contains 800 annotated episodes, along with an additional 200 common episodes for inter-annotator agreement calculation. We randomly sampled 100 episodes each from these 200, ensuring an equal proportion of labels from both annotators. These were combined with the remaining 800 episodes for a total of 1,000 episodes. The breakdown of labels is as follows: there are 350 episodes labeled as *TRUE* and 650 episodes labeled as *FALSE*. Among those labeled as *TRUE*, 25 are classified as *STRONG* and 325 as *WEAK*. In the original Japanese dataset, the average length of the texts was 28.6 characters, with a standard deviation of 25.3 characters and a median of 21.0 characters.

### 3.4 Inter-Annotator Agreement

To calculate the agreement between annotators, we used the agreement rate and Cohen's $\kappa$ coefficient (Cohen, 1960). The agreement rates for the labels are notably high, with 0.935 (187 out of 200) for determining loneliness or non-loneliness, 0.905 for social needs (181 out of 200), 0.984 (62 out of 63) for intensity among episodes determined as *TRUE* by the two annotators, and 0.905 (181 out of 200) encompassing all labels (loneliness or non-loneliness, social needs, intensity). Moreover, Cohen's $\kappa$ coefficient also indicates substantial agreement, measuring at 0.857 for loneliness or non-loneliness and 0.849 for intensity among episodes determined as *TRUE* by the two annotators. This suggests that we have constructed a consistent dataset and that our definition is clear enough to understand, enabling researchers to easily expand the dataset.

We conducted a qualitative analysis to understand the limitations of the guidelines. Examples translated from Japanese where the annotations do not match are shown in Table 1. Note that original examples and their transliterations are listed in Appendix A. Examples (1)-(5) do not match in terms of loneliness. The social needs shown in the table are labeled by an annotator who classified them as *TRUE*. The cause of the disagreement in annotations would arise from differences in the emotions held by annotators when they encounter the same situation as the episode. For example, in example (3), the difference arises depending on whether annotators believe they are socially accepted when they cannot communicate. Similarly, in example (5), the difference arises depending on whether annotators lack self-esteem or a sense of belonging when their thoughts are not understood.

## 4 Experiments

### 4.1 Settings

To validate the applicability of our dataset, we created two classification models: $M_{lonely}$, which classifies loneliness or non-loneliness, and $M_{inten}$, which assesses the intensity of loneliness. To train the models, we split the data into 70% for training, 15% for validation, and 15% for testing while maintaining class balance. We used the Japanese pre-trained BERT model[4]. We inserted an affine layer into the final layer of the pre-trained BERT model for classification. We set the learning rate of the pre-trained layers to $5.0 \times 10^{-5}$, and the learning rate of the final layer to $1.0 \times 10^{-4}$. We used the Adam optimizer with 20 epochs. For early stopping criteria, if the maximum validation accuracy of $M_{lonely}$ and the maximum F1 score of $M_{inten}$ did not change continuously for three epochs, the learning process was finished.

### 4.2 Results

Table 2 shows the evaluation results of the models. $M_{lonely}$ achieves an accuracy of 0.880 and an F1-score of 0.833. $M_{inten}$ achieved an accuracy of 0.943 whereas exhibits a low F1-score of 0.400. This discrepancy will be due to the scarcity of *STRONG* labels and the imbalance of labels in the dataset. Specifically, there are only 25 episodes

[4] https://huggingface.co/tohoku-nlp/bert-base-japanese-whole-word-masking

| | Episode | Social needs |
|---|---|---|
| (1) | A friend I have known for a long time has become too weak to go out after undergoing a coronary artery bypass surgery. | *Love & Affection* |
| (2) | They make it sound as if I am spreading things that I am not saying.. | *Social Recognition* |
| (3) | It's sad to think that I cannot communicate due to the language barrier. | *Social Recognition* |
| (4) | I'm troubled that I'm not being understood even doing the right thing and fail at what I want to do. | *Self-esteem & Belonging* |
| (5) | It's sad that I'm not being appreciated enough. | *Self-esteem & Belonging* |

Table 1: Examples translated from Japanese of annotation disagreement. The social needs are labeled by an annotator who categorized them as *TRUE*. Discrepancies in annotations stem from variations in the emotions experienced by annotators when confronted with the same scenario as the episode.

| | Acc | F1 | Prec | Rec |
|---|---|---|---|---|
| $M_{lonely}$ | 0.880 | 0.833 | 0.804 | 0.865 |
| $M_{inten}$ | 0.943 | 0.400 | 1.00 | 0.250 |

Table 2: Evaluation metrics. $M_{lonely}$ achieves a high accuracy and F1-score. $M_{inten}$ achieved a high accuracy whereas exhibits a low F1-score.

labeled as *STRONG*. This shortage results in a deflated F1-score, despite achieving a precision of 1.0 for the *STRONG* class.

### 4.3 Discussion

We conducted an error analysis on the test data. Table 3 shows examples translated from Japanese of episodes, ground truth labels, and predicted labels by $M_{lonely}$. Note that original examples and their transliterations are seen in in Appendix A. Examples (6)-(10) represent correct predictions by $M_{lonely}$, whereas examples (11)-(15) represent incorrect ones. $M_{lonely}$ accurately classified the loneliness experienced when the relationship with the person who you loved drifts away, as exemplified in (6), as well as the loneliness resulting from the loss of family members or pets, as depicted in (7) and (8). However, $M_{lonely}$ tended to misclassify instances where *Social Recognition* or *Self-esteem and Belonging* were unsatisfied, as seen in (11) and (12). When annotators label episodes involving *Social Recognition* and *Self-esteem and Belonging*, there is often inconsistency, which is considered difficult for $M_{lonely}$ to predict. Furthermore, $M_{lonely}$ occasionally misclassified the episode where someone else, not the author themselves, felt lonely as *TRUE*, as observed in (13). Moreover, there were several episodes, such as (14) and (15), where *FALSE* was the ground truth label but could potentially be labeled as *TRUE* by annotators. Episodes like these are challenging to label even manually and are also difficult for machine learning models to predict.

Table 4 shows examples translated from Japanese of episodes, ground truth labels, and predicted labels by $M_{inten}$. Note that original examples and their transliterations are seen in Appendix A. Examples (16)-(20) represent correct predictions by $M_{inten}$, whereas examples (21)-(23) represent incorrect ones. $M_{inten}$ correctly classified the episode with a strong negative expression indicating a loss of trust in people as *STRONG*, as exemplified in (16), which is only episode classified as *STRONG*. $M_{inten}$ also correctly classified the episode mentioning only the death of a relative or a pet as *WEAK* in (17) and (18). In addition, $M_{inten}$ correctly classified episodes where there were no strong negative expressions and no interference with daily life as *WEAK* in (19) and (20). In contrast, $M_{inten}$ did not correctly classify some episodes with interference in daily life or strong negative expressions as *STRONG*, as seen in (21), (22), and (23). Annotators classify these episodes as *STRONG* using these words as clues, such as "sad every day" in (22) and "really sad" in (23), following the guidelines. $M_{inten}$ does not predict them as *STRONG* and is not learning in accordance with the guidelines. There are only 17 episodes of the label *STRONG* in the training data, which is insufficient for achieving consistent classification for $M_{inten}$.

## 5 Prediction of Social Media Posts

To assess the feasibility of domain application, we evaluated the created models using X data.

### 5.1 Social Media Posts

We collected posts from Japan from July 1 to 31, 2022 by using the X (formerly Twitter) API[5].

[5] https://developer.x.com/en/docs/twitter-api

| | Episode | Gold | Pred |
|---|---|---|---|
| (6) | It's sad that my daughter, who used to spend most of her holidays with her parents, has recently prioritized her boyfriend and go out more often. | T | T |
| (7) | My father died of cancer. | T | T |
| (8) | The loss of a beloved pet. | T | T |
| (9) | I'm sad because I lost the key to my motorcycle. | F | F |
| (10) | The team I support couldn't win the championship. | F | F |
| (11) | I didn't pass the part-time job interview. | T | F |
| (12) | I was sad because my boss did not appreciate my work. | T | F |
| (13) | My daughter might be getting divorced. | F | T |
| (14) | I experienced power harassment in workplace from a senior colleague in the same department this spring. | F | T |
| (15) | My thought is not understood. | F | T |

Table 3: Examples translated from Japanese of episodes, ground truth labels, and predicted labels by $M_{lonely}$. Examples (6)-(10) represent correct predictions by $M_{lonely}$, whereas examples (11)-(15) represent incorrect ones. T and F mean *TRUE* and *FALSE*, respectively.

| | Episode | Gold | Pred |
|---|---|---|---|
| (16) | I am a self employed business person. A contractor I have been working well with for 5 years defaulted on 1.25m yen in debt and did a moonlight flit. The lawyer I consulted told me to give up. I lost faith in humanity. | S | S |
| (17) | My parent passed away. | W | W |
| (18) | My pet has passed away. | W | W |
| (19) | I was not allowed to attend the dinner party with my husband and children. | W | W |
| (20) | The intimidation from the colleague and what I want to convey do not come across. | W | W |
| (21) | I am as saddened as my wife is by the passing of her mother. I will never forget the hospitality she extended to me the first time we stayed at her parents' house, though I believe she lived out her natural life at the age of 95. I feel like I lost my parent too. | S | W |
| (22) | I was harassed by the former company's president and felt sad every day. | S | W |
| (23) | I was really sad when my wife secretly borrowed money and ran away because she couldn't pay it back. | S | W |

Table 4: Examples translated from Japanese of episodes, ground truth labels, and predicted labels by $M_{inten}$. Examples (16)-(20) represent correct predictions by $M_{inten}$, whereas examples (21)-(23) represent incorrect ones. S and W mean *STRONG* and *WEAK*, respectively.

| | Episode | Gold | Pred |
|---|---|---|---|
| (24) | I can't meet the person I want to see due to various obstacles. | T | T |
| (25) | Jealous...death. | F | T |
| (26) | Good night. | F | T |

Table 5: Examples translated from Japanese of posts predicted as *TRUE* by $M_{lonely}$, along with their ground truth labels. T means *TRUE*, F means *FALSE*.

We preprocessed the collected posts by removing emojis, URLs, mentions, RT, tweets from users with 'bot' in their username, and duplicate posts, in addition to the preprocessing steps performed on the LIFE STORY corpus. The number of posts after preprocessing is 750,240. To align with the sad category of the annotated LIFE STORY corpus used to fine-tune our BERT models ($M_{lonely}$ and $M_{inten}$), we conducted emotion analysis on X posts by creating a new model using the LIFE STORY corpus with Naive Bayes (MultinomialNB [6] from scikit-learn) to extract posts expressing sadness. Through this process, we can better ensure that the sadness posts extracted from X, used as input, will be of the same nature as the data used for fine-tuning our models. We used the emotion categories of the corpus as ground truth labels and calculated the probabilities for classification into each emotion category. Note that for the preliminary experiment, we also constructed BERT trained on LIFE STORY corpus for emotion analysis. However, since the performance did not differ significantly from Naive Bayes, we used Naive Bayes due to its high interpretability. Subsequently, we extracted the posts classified as sad, which is 57,648 posts, and inputted them into $M_{lonely}$. We fed the posts classified as *TRUE* by $M_{lonely}$ into $M_{inten}$. Finally, 6,902 and 496 posts

[6] https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.MultinomialNB.html

| | Episode | Gold | Pred |
|---|---|---|---|
| (27) | Please let me take a break from Twitter for a while. This morning, my mother passed away. Over the past two years, she was bedridden in a facility, going in and out of the hospital, and back to the facility repeatedly. I knew this moment would inevitably come, but it's still extremely painful. It was so hard not being able to see her for two years due to the pandemic. | S | S |
| (28) | I woke up and now I can't sleep. Prime Minister Suga's comment that former Prime Minister Abe was lonely resonates deeply in my heart, and it's too painful. I wonder if Mrs. Akie Abe is also lonely and spending sleepless nights. I'm worried about her. | W | S |

Table 6: Examples translated from Japanese of posts predicted as *STRONG* by $M_{inten}$, along with their ground truth labels. S means *STRONG*, W means *WEAK*.

were classified as *TRUE* and *STRONG*, respectively.

### 5.2 Results and Discussion

To evaluate the performance of the two models ($M_{lonely}$ and $M_{inten}$), we randomly sampled and manually annotated 300 posts: 150 posts classified as *TRUE* and 150 posts classified as *STRONG*. We calculated precision from these annotated posts. The reason for using precision as the evaluation metric instead of F1-score or accuracy is due to the inability to annotate on a large scale. To use F1-score or accuracy, posts need to be annotated before prediction. Due to the extremely low number of posts classified as *TRUE* and *STRONG* within the data from X, a large number of annotations are necessary to properly evaluate it. Therefore, this time annotations were made after predictions, on posts labeled by the model as *TRUE* and *STRONG* to calculate precision. Calculating accuracy and F1-score is planned for future work. The resulting precision for $M_{lonely}$ was 0.113, and for $M_{inten}$ it was 0.02, which is significantly lower compared to its performance on the LIFE STORY corpus. From those results, we can conclude that these models lacked the ability for domain adaptation for X posts. These inferior results can be attributed to the significant differences in syntax, vocabulary, and word usage between social media texts and the LIFE STORY episodes.

Table 5 lists examples translated from Japanese of posts predicted as *TRUE* by $M_{lonely}$, along with their ground truth labels. Note that original examples and their transliterations are listed in Appendix A. As seen in (24), $M_{lonely}$ can detect loneliness when someone wants to meet but cannot. However, $M_{lonely}$ often mistakenly predicts *TRUE* when influenced by expressions related to "death" because those words frequently occur in loneliness episodes in our dataset, as seen in (25). Many of the posts predicted as *TRUE* included greetings, as seen in (26). As LIFE STORY corpus does not contain greetings, it appears that the model cannot predict accurately when the text contains only greetings.

Table 6 shows examples translated from Japanese of posts predicted as *STRONG* by $M_{inten}$, along with their ground truth labels. Note that original examples and their transliterations are listed in Appendix A. As shown in (27), $M_{inten}$ correctly classified the posts expressing hardship over the loss of a mother with strong negative expressions as *STRONG*. Similar to the evaluation on the dataset, consistency in $M_{inten}$ was not observed.

## 6 Limitations

Based on the definition of loneliness, we classified whether the text expressed loneliness, but the emotions perceived by the readers and the writers may be different (Kajiwara et al., 2021; Ramos et al., 2022). Accordingly, we plan to collect texts and loneliness scores of their writers through crowdsourcing. In terms of completeness, we annotated episodes evoked by sadness; hence, loneliness that occurs alongside other emotions or loneliness that occurs independently of other emotions may not be captured. In addition, the loneliness dataset we created lacks a sufficient number of *STRONG* labels for learning. We plan to expand our dataset to secure an ample number of *STRONG* labels, thereby addressing the low recall issue in the model for classification of intensity. Regarding the guidelines, the disagreement among annotators often arises from the fact that different people perceive the same situation differently. This is evident in texts involving *Social Recognition* and *Self-esteem and Belonging*.

We created a BERT-based classifier, but using other models may result in higher performance. Considering the rapid development in recent years, it is also necessary to consider using generative models represented by GPT-4. The performance of our models decreased when predicting on X data, which is distinct from our dataset. Research has shown significant differences in syntax, vocabulary, and word usage between normal conversations and social media text, such as that found on X (Bryden et al., 2013). Social media platforms often feature unique language patterns influenced by their community structures. For example, X users may adopt specific terminologies, abbreviations, and stylistic choices that reflect the norms and culture of the online community they engage with. Therefore, our current model is not appropriate for predicting loneliness and its intensity in social media texts, and it is quite challenging to achieve good performance with such inconsistent data. To address this problem, Arefyev et al. (2021) proposed a technique for more efficient domain and task adaptation of pre-trained masked language models such as BERT before fine-tuning them on a specific task. This technique forces the model to predict words that are highly indicative of the target task classes (e.g., sentiment words for sentiment analysis), allowing it to learn better task-relevant representations during adaptation. We will use this method for domain adaptation in future work.

## 7 Conclusion

We present a freely available Japanese loneliness dataset[7], which is created by annotating a short episode corpus, with clear guidelines. Our annotation guidelines are based on the psychological definition of loneliness. Using this guideline for annotation, the results showed a Cohen's $\kappa$ coefficient of 0.857 for loneliness or non-loneliness and 0.849 for intensity among episodes determined as loneliness by two annotators, indicating consistency. We also construct classifiers to identify whether loneliness is expressed and to assess the intensity of loneliness using our dataset. The model classifying loneliness or non-loneliness achieved an F1-score of 0.833. However, the model identifying loneliness intensity had a low F1-score of 0.400, which is likely due to insufficient learning of a specific label and the imbalance of labels in the dataset. In addition, these models show low performance in a domain distinct from the texts used for training.

[7] https://github.com/sociocom/Japanese-Loneliness-Dataset

In the future, we plan to expand our dataset to alleviate data imbalance and address the shortage of a specific label, as well as to improve domain adaptation for social media. We also plan to create a classification model for social needs to identify factors contributing to loneliness, which will provide valuable insights for detection and intervention methods.

## Acknowledgments

This work was supported by JST-Mirai Program Grant Number JPMJMI21J2, Japan.

## References

Anietie Andy, Garrick Sherman, and Sharath Chandra Guntuku. 2022. Understanding the expression of loneliness on twitter across age groups and genders. *PLOS ONE*, 17(9):1–19.

Nikolay Arefyev, Dmitrii Kharchev, and Artem Shelmanov. 2021. NB-MLM: Efficient domain adaptation of masked language models for sentiment analysis. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9114–9124, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Manfred E Beutel, Eva M Klein, Elmar Brähler, Iris Reiner, Claus Jünger, Matthias Michal, Jörg Wiltink, Philipp S Wild, Thomas Münzel, Karl J Lackner, et al. 2017. Loneliness in the general population: prevalence, determinants and relations to mental health. *BMC psychiatry*, 17(1):1–7.

John Bryden, Sebastian Funk, and Vincent Jansen. 2013. Word usage mirrors community structure in the online social network twitter. *EPJ Data Science*, 2.

John T. Cacioppo, Louise C. Hawkley, and Ronald A. Thisted. 2010. Perceived social isolation makes me sad: 5-year cross-lagged analyses of loneliness and depressive symptomatology in the chicago health, aging, and social relations study. *Psychology and aging*, 25 2:453–63.

Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20:37 – 46.

Edward L Deci and Richard M Ryan. 2008. Self-determination theory: A macrotheory of human motivation, development, and health. *Canadian psychology/Psychologie canadienne*, 49(3):182.

Louise C Hawkley and John T Cacioppo. 2010. Loneliness matters: A theoretical and empirical review of consequences and mechanisms. *Annals of behavioral medicine*, 40(2):218–227.

Yueyi Jiang, Yunfan Jiang, Liu Leqi, and Piotr Winkielman. 2022. Many ways to be lonely: Fine-Grained characterization of loneliness and its potential changes in COVID-19. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 16, pages 405–416.

Tomoyuki Kajiwara, Chenhui Chu, Noriko Takemura, Yuta Nakashima, and Hajime Nagahara. 2021. WRIME: A new dataset for emotional intensity estimation with subjective and objective annotations. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2095–2104, Online. Association for Computational Linguistics.

Svetlana Kiritchenko, Will E Hipson, Robert J Coplan, and Saif M Mohammad. 2020. SOLO: A corpus of tweets for examining the state of being alone. *arXiv preprint arXiv:2006.03096*.

Charikleia Lampraki, Adar Hoffman, Angélique Roquet, and Daniela S Jopp. 2022. Loneliness during COVID-19: Development and influencing factors. *PLOS ONE*, 17(3):e0265900.

Ellen E. Lee, Colin Depp, Barton W. Palmer, Danielle Glorioso, Rebecca Daly, Jinyuan Liu, Xin M. Tu, Ho-Cheol Kim, Peri Tarr, Yasunori Yamada, and et al. 2019. High prevalence and adverse health effects of loneliness in community-dwelling adults across the lifespan: role of wisdom as a protective factor. *International Psychogeriatrics*, 31(10):1447–1462.

Siegwart Lindenberg. 1996. Continuities in the theory 10 of social production functions. *Verklarende sociologie: Opstellen voor Reinhard Wippler, Thesis Publishers, Amsterdam*, pages 169–184.

Ye Luo, Louise C. Hawkley, Linda J. Waite, and John T. Cacioppo. 2012. Loneliness, health, and mortality in old age: A national longitudinal study. *Social Science Medicine*, 74(6):907–914. Part Special Issue: Migration, 'illegality', and health: Mapping embodied vulnerability and debating health-related deservingness.

Ruimin Ma, Farhana Mann, Jingyi Wang, Brynmor Lloyd-Evans, James Terhune, Ahmed Al-Shihabi, and Sonia Johnson. 2020. The effectiveness of interventions for reducing subjective and objective social isolation among people with mental health problems: a systematic review. *Social Psychiatry and Psychiatric Epidemiology*, 55.

Howard Meltzer, Paul Bebbington, Michael Dennis, Rachel Jenkins, Sally McManus, and Traolach Brugha. 2012. Feelings of loneliness among adults with mental disorder. *Social psychiatry and psychiatric epidemiology*, 50.

Michael Mullarkey, Igor Marchetti, and Christopher Beevers. 2018. Using network analysis to identify central symptoms of adolescent depression. *Journal of Clinical Child Adolescent Psychology*, 48.

Kentaro Nakai, Ritsuko Iwai, and Takatsune Kumada. 2023. An examination of eating experiences in relation to psychological states, loneliness, and depression using bert. *New Frontiers in Artificial Intelligence*, pages 196–214.

Johan Ormel, Siegwart Lindenberg, Nardi Steverink, and Lois M Verbrugge. 1999. Subjective well-being and social production functions. *Social indicators research*, 46:61–90.

Letitia Anne Peplau and Daniel Perlman. 1982. *Loneliness: A Sourcebook of Current Theory, Research and Therapy*. Wiley Series on Personality Processes. Wiley.

Patrick John Ramos, Kiki Ferawati, Kongmeng Liew, Eiji Aramaki, and Shoko Wakamiya. 2022. Emotion analysis of writers and readers of Japanese tweets on vaccinations. In *Proceedings of the 12th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis*, pages 95–103, Dublin, Ireland. Association for Computational Linguistics.

Aparna Shankar, Anne Mcmunn, James Banks, and Andrew Steptoe. 2011. Loneliness, social isolation, and behavioral and biological health indicators in older adults. *Health psychology : official journal of the Division of Health Psychology, American Psychological Association*, 30:377–85.

Nardi Steverink, Siegwart Lindenberg, Tali Spiegel, and Anna P. Nieboer. 2020. The associations of different social needs with psychological strengths and subjective well-being: An empirical investigation based on social production function theory. *Journal of Happiness Studies*, 21(3):799–824.

Harry Stack Sullivan. 1953. *The Interpersonal Theory of Psychiatry*. Routledge.

Harry Taylor. 2020. Social isolation's influence on loneliness among older adults. *Clinical Social Work Journal*, 48.

Nicole K Valtorta, Mona Kanaan, Simon Gilbody, Sara Ronzi, and Barbara Hanratty. 2016. Loneliness and social isolation as risk factors for coronary heart disease and stroke: systematic review and meta-analysis of longitudinal observational studies. *Heart*, 102(13):1009–1016.

## A Appendix

| | Episode | Social needs |
|---|---|---|
| (1) | 長らく付き合いがあって、仲良くしていた友人が、冠動脈置換手術によって、弱ってしまって出かけられなくなった。<br>Nagaraku tsukiai ga atte, nakayoku shiteita yûjin ga, kandômyakushujutsu ni yotte, yowatte shimatte dekake rarenaku natta.<br>A friend I have known for a long time has become too weak to go out after undergoing a coronary artery bypass surgery. | *Love & Affection* |
| (2) | 私が言いふらしていないことをあたかも私が言っているようにされていること。<br>Watashi ga iifurashite inai koto wo atakamo watashi ga itteiru yôni sareteirukoto.<br>They make it sound as if I am spreading things that I am not saying. | *Social Recognition* |
| (3) | 言葉が通じないなと思うと悲しい。<br>Kotoba ga tûjinai na to omou to kanashii.<br>It's sad to think that I cannot communicate due to the language barrier. | *Social Recognition* |
| (4) | 正しいことをしても理解されていないことがあり、何とかしたいがなんともならない悩み。<br>Tadashii koto wo shite mo rikaisa reteinaikoto ga ari, nantoka shitai ga nantomo naranai nayami.<br>I'm troubled that I'm not being understood even when I'm doing the right thing. I want to do something about it but I can't. | *Self-esteem & Belonging* |
| (5) | 自分自身の評価があまり良くなくて悲しい。<br>Jibunjishin no hyôka ga amari yokuna kute kanashii.<br>It's sad that I'm not being appreciated enough. | *Self-esteem & Belonging* |

Table 7: Original examples of annotation disagreement and their transliterations. The social needs are labeled by an annotator who categorized them as *TRUE*. Discrepancies in annotations stem from variations in the emotions experienced by annotators when confronted with the same scenario as the episode. Note that The original text in example (2) has been corrected due to a typographical error (the particle ' が' was duplicated).

| | Episode | Gold | Pred |
|---|---|---|---|
| (6) | これまで親と休日を過ごすことが多かった娘が、最近は彼氏を優先するようになり、外出が増えたことが悲しい。<br>Koremade oya to kyûjitsu wo sugosu koto ga ôkatta musume ga, saikin wa kareshi wo yûsensuru yô ni nari, gaishutsu ga fueta koto ga kanashii.<br>It's sad that my daughter, who used to spend most of her holidays with her parents, has recently prioritized her boyfriend and go out more often. | T | T |
| (7) | 父ががんで亡くなったこと<br>Chichi ga gan de nakunatta koto.<br>My father died of cancer. | T | T |
| (8) | 可愛がっていたペットが亡くなったこと<br>Kawaigatte ita petto ga nakunatta koto.<br>The loss of a beloved pet. | T | T |
| (9) | バイクの鍵をなくしてしまって悲しい<br>Baiku no kagi wo nakushite shimatte kanashii.<br>I'm sad because I lost the key to my motorcycle. | F | F |
| (10) | 応援しているチームが優勝できなかったこと<br>Ôenshite iru tîmu ga yûshô dekinakatta koto.<br>The team I support couldn't win the championship. | F | F |
| (11) | パートの面接に受からなかった。<br>Pâto no mensetsu ni ukara nakatta.<br>I didn't pass the part-time job interview. | T | F |
| (12) | 職場で仕事の成果物が上司に理解されず悲しかった。<br>Shokuba de shigoto no seikabutsu ga jôshi ni rikaisare zu kanashi katta.<br>I was sad because my boss did not appreciate my work. | T | F |
| (13) | 娘が離婚するかも知れなくなった。<br>Musume ga rikonsuru kamo shirenaku natta.<br>My daughter might be getting divorced. | F | T |
| (14) | 今年春、職場で、同じ部の先輩からパワハラを受けた。<br>Kotoshi haru, shokuba de, onaji bu no sempai kara pawahara wo uketa.<br>I experienced power harassment in workplace from a senior colleague in the same department this spring. | F | T |
| (15) | 思いを分かってもらえない<br>Omoi wo wakatte moraenai.<br>My thought is not understood. | F | T |

Table 8: Original examples of episodes, their transliterations, ground truth labels, and predicted labels by $M_{lonely}$. Examples (6)-(10) represent correct predictions by $M_{lonely}$, whereas examples (11)-(15) represent incorrect ones. T means *TRUE*, F means *FALSE*.

| | Episode | Gold | Pred |
|---|---|---|---|
| (16) | 仕事上自営業をしており 5 年ぐらいの付き合いで仲良く仕事を貰ってた業者さんに 125 万円踏んだおされました。その業者は夜逃げしてしまい弁護士に相談したら勉強代だと思って諦めた方が良いと言われました。本当に人間不信になります。<br>Shigoto jô jieigyô wo shite ori 5 nen gurai no tsukiai de nakayoku shigoto wo moratte ita gyôsha san ni 125 manen fundaosare mashita. Sono gyôsha wa yonigeshite shimai bengoshi ni sôdan shitara benkyôdai da to omotte akirameta hô ga ii to iware mashita. Hontô ni ningenfushin ni narimasu.<br>I am a self employed business person. A contractor I have been working well with for 5 years defaulted on 1.25m yen in debt and did a moonlight flit. The lawyer I consulted told me to give up. I lost faith in humanity. | S | S |
| (17) | 親が亡くなったことです<br>Oya ga nakunatta koto desu.<br>My parent passed away. | W | W |
| (18) | ペットがしんでしまった。<br>Petto ga shinde shimatta.<br>My pet has passed away. | W | W |
| (19) | 夫と子供のお食事会に参加させてもらえなかった事<br>Otto to kodomo no oshokujikai ni sankasasete moraenakatta koto.<br>I was not allowed to attend the dinner party with my husband and children. | W | W |
| (20) | 職場の同僚からの恫喝や伝えたい事が伝わらない事<br>Shokuba no dôryô kara no dôkatsu ya tsutaetai koto ga tsutawaranai koto.<br>The intimidation from the colleague and what I want to convey do not come across. | W | W |
| (21) | 妻の母親が亡くなって妻同様に悲しみに暮れている。初めて妻の実家に泊まったときによくもてなしをしていただいたことが忘れられない。95 歳で天寿を全うしたと思うが。これで私には親は居なくなってしまった。<br>Tsuma no hahaoya ga nakunatte tsuma dôyô ni kanashimi ni kureteiru. Hajimete tsuma no jikka ni tomatta toki ni yoku motenashi wo shite itadaita koto ga wasurerarenai. 95 sai de tenju wo mattoushita to omouga. Kore de watashi ni wa oya wa inakunatte shimatta.<br>I am as saddened as my wife is by the passing of her mother. I will never forget the hospitality she extended to me the first time we stayed at her parents' house, though I believe she lived out her natural life at the age of 95. I feel like I lost my parent too. | S | W |
| (22) | 前の会社の社長からパワハラを、受けて毎日悲しかった。<br>Mae no kaisha no shachô kara pawahara wo, ukete mainichi kanashi katta.<br>I was harassed by the former company's president and felt sad every day. | S | W |
| (23) | 妻が私に内緒で借金をし、支払えなくなって家出した時は本当に悲しかった。<br>Tsuma ga watashi ni naisho de shakkin wo shi, shiharaenaku natte iedeshita toki wa hontô ni kanashi katta.<br>I was really sad when my wife secretly borrowed money and ran away because she couldn't pay it back. | S | W |

Table 9: Original examples of episodes, their transliterations, ground truth labels, and predicted labels by $M_{inten}$. Examples (16)-(20) represent correct predictions by $M_{inten}$, whereas examples (21)-(23) represent incorrect ones. S means *STRONG*, W means *WEAK*.

| | Episode | Gold | Pred |
|---|---|---|---|
| (24) | 会いたい人に会えない。。。いろいろ障壁があって<br>Aitai hito ni aenai... Iroiro shôheki ga atte.<br>I can't meet the person I want to see due to various obstacles. | T | T |
| (25) | 羨ましい... 死亡。<br>Urayamashii...shibô.<br>Jealous...death. | F | T |
| (26) | おやすみなさいまし。<br>Oyasuminasai mashi.<br>Good night. | F | T |

Table 10: Original examples of posts and their transliterations predicted as *TRUE* by $M_{lonely}$, along with their ground truth labels. T means *TRUE*, F means *FALSE*.

|  | Episode | Gold | Pred |
|---|---|---|---|
| (27) | Twitter しばらく休ませて下さい。今朝、おふくろが他界しました。 ここ 2 年間施設で寝たきりで病院入退院、施設と何回も繰り返しでこの時は必ずは来るとは思って覚悟はしてたけど、かなり辛いです。コロナ禍で 2 年も会えなく辛すぎでした。<br>Twitter shibaraku yasumasete kudasai. Kesa, ofukuro ga takaishi mashita. Koko 2 nenkan shisetsu de netakiri de byôin nyûtaiin, shisetsu to nankai mo kurikaeshi de kono toki wa kanarazu wa kuru to wa omotte kakugo wa shiteta kedo, kanari tsurai desu. Koronaka de 2 nen mo aenaku tsura sugi deshita.<br>Please let me take a break from Twitter for a while. This morning, my mother passed away. Over the past two years, she was bedridden in a facility, going in and out of the hospital, and back to the facility repeatedly. I knew this moment would inevitably come, but it's still extremely painful. It was so hard not being able to see her for two years due to the pandemic. | S | S |
| (28) | 目が覚めて、寝れなくなってしまった。安倍さんは寂しがりやだったという菅さんのコメントが、胸に響いてつらすぎます。昭恵夫人も寂しくて、寝られない日々を送られているのだろうか。心配です。<br>Me ga samete, nerenaku natte shimatta. Abe san wa samishigariya datta to iu Suga san no komento ga, mune ni hibiite turasugi masu. Akie fujin mo samishikute, nerarenai hibi wo okurarete iruno darôka. Shimpai desu.<br>I woke up and now I can't sleep. Prime Minister Suga's comment that former Prime Minister Abe was lonely resonates deeply in my heart, and it's too painful. I wonder if Mrs. Akie Abe is also lonely and spending sleepless nights. I'm worried about her. | W | S |

Table 11: Original examples of posts and their transliterations predicted as *STRONG* by $M_{inten}$, along with their ground truth labels. S means *STRONG*, W means *WEAK*.