The Journal of Studies in Language

The Journal of Studies in Language 39(1), 055-071 (2023)

Detecting Suicide Notes with the Probability of Positive Sentiment and Violin Plot

Lee, Yong-hun
Chungnam National University

ABSTRACT

The Journal of Studies in Language **39.1, 055-071**. This paper tries to propose a new technique for identifying suicide notes that utilizes sentiment analysis. Because suicide notes are short in length, detecting signals of suicide can be challenging. The algorithm for sentiment analysis which is proposed is based on the probability of positive sentiments (PPS) rather than traditional categorical classifications. The original BERTLARGE model is modified so that we can calculate the PPS value of each sentence. In the analysis, 8 corpora will be used, 4 of which are suicide notes, and the others are ordinary texts. The PPS values are calculated for each sentence in the corpus using the BERTLARGE model. Ordinary texts have convexed parts around the score 50, whereas suicide notes demonstrate no such tendency. **(Chungnam National University)**

Keywords: suicide notes, BERTLARGE, sentiment analysis, probability of positive sentiment, violin plot

https://doi.org/10.18627/jslg.39.1.202305.55

pISSN : 1225-4770
eISSN : 2671-6151

Received: April 01, 2023

Revised: May 06, 2023

Accepted: May 18, 2023

1. Introduction

As suicide rates rise nowadays, it's critical to identify the warning symptoms of suicide before someone actually commits the deed. Because suicide notes are short, however, automatic detection of suicide indications is usually difficult. Suicide notes are difficult to be identified using only the conventional techniques in forensic linguistics. Also, many people who choose to end their lives leave a suicide note saying their motivations as well as their psychological and emotional conditions.

This study tries to provide a novel approach that combines deep learning architecture and sentiment analysis to efficiently detect suicide notes. Regarding the dataset, 8 corpora in total are used, 4 of which are suicide notes and the others are ordinary texts. Virginia Woolf's four literary works are also included in the dataset so that it is possible to compare them to two of her suicide notes. Virginia Woolf's literary works are included in this study to examine if the proposed algorithm can distinguish clearly between suicide notes and other sorts of ordinary texts, even when the same person wrote both.

The Bidirectional Encoder Representations from Transformers (BERT; Devlin et al., 2019) model in deep learning is used in this paper. Among many BERT models, a $\text{BERT}_{\text{LARGE}}$ model was downloaded from the Hugging Face website and trained with the IMDB dataset.[1)] The majority of earlier sentiment analysis studies classify sentiment using a binary (*positive* or *negative*) or ternary classification (*positive*, *neutral*, or *negative*). In this study, however, the probability of the positive sentiment (PPS) is used for each sentence, which can be derived during the process of determining the sentiment label (*positive* or *negative*). In other words, the modified $\text{BERT}_{\text{LARGE}}$ model is used to determine the PPS value for each sentence in the corpora. Then, the distributions of PPS values are then represented graphically with a box plot and a violin plot.

Four different types of analysis results will be presented along with their characteristics to illustrate how the best results can be obtained when the distributions of PPS values and a violin plot are combined: (i) ternary classification with a box plot; (ii) PPS values with a box plot; (iii) ternary classification with a violin plot; and (iv) PPS values with a violin plot. An experiment is conducted in order to assess the proposed model. 200 files are created in the experiment from two different types of corpora (suicide notes and ordinary texts), and the developed model is evaluated with precision, recall, *F*-score, and accuracy.

The contributions of this study are as follows:

(1) Contributions of this paper

a. It suggests a more effective method for identifying suicide notes that makes use of the Transformer model (Vaswani et al., 2017).
b. Compared to previous studies, the approach had a higher accuracy rate (about 85.5%) in identifying suicide notes.
c. The approach can be used with both small and big size of corpora.
d. The technique visually displayed the emotional state of suicide notes.
e. The technique makes clear the fundamental emotional characteristics of suicide notes.

The algorithm proposed in this paper has the advantage of practical application to numerous sorts of texts including Twitter and SNS.

2. Previous Studies

2.1 Forensic Linguistics

Forensic linguistics is the foundation of the majority of the earlier investigations on suicide notes. It is a branch of applied linguistics which uses linguistic data, analysis methods, and insights of linguistics in the context of things like law, criminal investigations, court cases, judicial procedure, etc. It is specifically a branch of corpus linguistics, and its scope encompasses various perspectives of criminal investigations and judicial procedures in addition to authorship identification (such as authorship verification, authorship profiling, and authorship attribution).

1) https://huggingface.co/

In 1968, Professor Jan Svartvik first uses the term *forensic linguistics* in his analysis of a prominent murder suspect's (Timothy John Evans') publications (Svartvik, 1968). He examines four books with different linguistic properties and finds big disparities between them. It is hypothesized that the authors of the writings might not be the same people. In addition to the International Journal of Law, Language, and Discourse which began publication in 1994, the International Association of Forensic Linguists (IAFL) was founded in 1993.

According to Olsson (2004), forensic linguistics is characterized as being placed at the nexus of language, crime, and law which apply linguistic knowledge to a particular (social) situation (i.e., a legal scenario). According to Olsson (2008), any spoken or written words that are cited in court or during an investigation may be regarded as forensic texts. A suicide note generally contains lines that propose a way to kill oneself, thus the study also notes that an analysis of the suicide notes must be included in the investigation of forensic linguistics. Applying linguistic expertise to the investigation of suicide notes is necessary in order to determine the truth and intention of the suicide.

2.2 Studies on Suicide Notes Using Forensic Linguistics

As forensic linguistics develops, numerous examinations of suicide notes have been carried out. According to the literature in forensic linguistics, suicide notes exhibit typical (linguistic) traits. To examine suicide notes in forensic linguistics, we have to consult all of theoretical linguistics (including phonetics, phonology, morphology, syntax, semantics, pragmatics, and discourse analysis). An assembled corpus served as the foundation for the majority of the early linguistic study on suicide notes (Sheidman and Faberow, 1963). The corpus contains 66 writings, a mixture of actual and fictitious suicide notes. Discourse analysis methods, the use of various auxiliary verbs (including *modals*), or verbs that distinguish between the real and the fake have been used to study the messages of suicide notes.

Several studies actively analyze suicide notes using corpus linguistics and machine learning techniques. Numerous linguistic components, including personal pronouns, past tense verbs, nouns, and various types of semantic categories, are statistically analyzed in these investigations (Olsson, 2008). Automated corpus analysis methods have been increasingly popular in recent years as a method of identifying suicide notes. To identify and classify typical suicide notes, academics employ automated machine-learning approaches (Pestian et al., 2010) or corpus-analysis techniques (Shapero, 2011).

The Linguistic Inquiry and Word Count (LIWC) program was created as a result of these investigations (Pennebaker et al., 2001; Tausczik and Pennebaker, 2010). The application has been extensively used in several academic disciplines, including computer science and psycholinguistics. The software analyzes writings by ordinary people and compiles data on specific semantic word patterns using 72 language features (Pennebaker and King 1999). The variables (i.e., linguistic components) of the LIWC can be divided into four categories: Standard Linguistic Dimension, Psychological Process, Relativity, and Personal Issues. Based on their frequency, the algorithm separates the 72 variables into the aforementioned four categories. The ratios of words in each category show the author's psychological states. The proportions of words in each category reveal the writer's mental state and writing procedures. According to Pennebaker et al. (2001), LIWC contains more than 3,000 frequently used content words in addition to a variety of word types, word lengths, and function words (such as articles, prepositions, and first-, second-, and third-person pronouns). It is possible to compile the analysis results in several individual psychological factors and modifications, which cannot be obtained from the prior studies, by counting function words and pronouns.

2.3 Machine Learning/Deep Learning Approaches to Suicide Notes

According to Samuel (1959), machine learning can be defined as "fields of study that gives computers the ability to learn without being explicitly programmed." Mitchell (1997) also defines it as follows: "A computer program is said to learn from experience *E* with respect to some class of tasks *T* and performance measure *P* if its performance at tasks in *T*, as measured by *P*, improves with experience *E*." Combining these two well-known definitions, we can define machine learning as a branch of computer science, specifically within artificial intelligence, which enables a machine (a computer) to *automatically* learn from training data (which can be referred to as *E*) and perform a class of tasks *T* (either classification or regression) with the performance metric *P*.

Previous research has looked into suicide notes using machine learning and deep learning techniques. First of all, Pestian et al. (2012a) describe how a corpus of suicide notes was created which can be utilized for machine learning. There are 1,278 notes in the corpus that were written by suicide victims.

Pestian et al. (2012b) describe a collaborative task where a few different emotions were assigned to suicide notes. The early analysis of the results, an explanation of the data creation method, and evaluation criteria make this study distinct from past joint efforts in the biomedical sector.

Yang et al. (2012) mention the creation of automatic algorithms that can recognize the affective text of 15 specific emotions at the sentence level. This paper also shows their efforts to provide automatic algorithms which can recognize the emotive language from suicide notes at the sentence level using 15 different emotions. With a micro-averaged *F*-measure score of 61.39% in textual emotion recognition, the automated system is proven to do well compared to the manually annotated gold standard and produced positive results.

McCart et al. (2012) focus on sentiment analysis in order to determine if 15 emotions (labels) were present simultaneously in a collection of suicide letters spanning more than 70 years. A combination of rules and statistical text mining (STM) models were used by the winning entry to provide a micro-averaged *F*-score of 0.5023, which was slightly higher than the average of the 26 participating teams.

Wang et al. (2012) present the answer to the i2b2 sentiment classification problem. Their hybrid strategy combines machine learning and rule-based classifiers. By combining the machine learning classifier and the rule-based classifier, the hybrid system obtains the highest micro-averaged *F*-score 0.5038, which is more than the mean (0.4875) and median (0.5027) micro-average *F*-score among all participating teams.

Desmet and Hoste (2013) investigate whether current advances in sentiment analysis and natural language processing may be used to precisely detect 15 different emotions that could be an indicator of suicide notes. The best feature combination for each of the different emotions is found via bootstrap resampling, and the classification's accuracy can reach 68.86% of the *F*-score.

Ghosh et al. (2020) establish a fine-grained emotion annotated corpus (CEASE) of English suicide notes in order to perform emotion detection on the well-curated dataset. The corpus (version 1) consists of 2,393 sentences from about 205 suicide notes that were collected from various sources. A set of 15 different fine-grained emotion labels (forgiveness, happiness, peacefulness, love, pride, hopefulness, thanksgiving, blame, anger, fear, abuse, sadness, hopelessness, guilt, information, and instructions), are used to annotated each sentence with emotion classes. They develop an ensemble architecture for the evaluation, and they achieve the maximum cross-validation accuracy of 60.32% and test accuracy of 60.17%.

3. Materials and Methods

3.1 Corpus Compilation

There are two different kinds of corpora required in order to detect suicide notes using sentiment analysis. The first is a corpus of suicide notes, and the second is a corpus of ordinary texts that suicide notes are contrasted with.

The following 8 corpora are used in this paper for this purpose. The number of sentences in each corpus is listed in Table 1. It is important to look at how many sentences are in each corpus because sentiment analysis is done on every sentence.

Table 1. Number of Sentences per Corpus

Corpus	# of Sentences	Corpus	# of Sentences
01.Noh	14	05.TheVoyageOut	8,088
02.Cobain	34	06.NightAndDay	7,531
03.Woolf-S	13	07.Jacob'Room	3,413
04.Woolf-H	22	08.MondayOrTuesday	1,142

There are 8 corpora in all; the first 4 corpora in the table (left column) are corpora for suicide notes, while the remaining 4 (right column) are corpora for ordinary writings. The corpus files 01~04 and 06~09 were obtained from Lee and Joh (2019). These corpora also include materials produced by the same author, Virginia Woolf, in corpora 03~04 and 05~08. They are included in the study because it is vital to determine whether the suicide notes differ from other texts written by the same author in terms of their characteristics.

The detailed information for each corpus is listed below. 01 is a suicide note written by a Korean politician. Although it was originally published in Korean, an English translation was done for the investigation. Kurt Cobain, a musician who took his own life in 1994, left a suicide note 02. Virginia Woolf wrote suicide notes 03 and 04. In the first letter, she addresses her sister; in the second, her husband. Virginia Woolf also wrote the novels 05~08 as literary works.

We can make the following hypothesis from this table.

(2) Hypothesis

a. The corpora 01~04 will exhibit different characteristics from the corpora 05~08.

b. Though they were written by the same author, the corpora 03~04 and 05~08 will exhibit different characteristics.

These hypotheses will be examined in Section 6.

3.2 Ternary Classifications of Sentiment Analysis

The early approach to sentiment analysis produces the outputs with either binary or ternary classifications. The sentiment is categorized as either *positive* or *negative* using binary classifiers. Sentiment analysis divides sentiments into three categories: *positive*, *neutral*, and *negative*. Several studies categorize the sentiment into multiple categories.

For the comparisons with the analysis method in Section 3.3, a ternary classification of sentiment analysis is used in this research. All the words in the corpus are divided into three categories in order to achieve this goal. A pre-trained BERT model is obtained from the Hugging Face website (*positive*, *neutral*, and *negative*).[2] Each label is then changed into a number: *positive* becomes 1, *neutral* becomes 0, and *negative* becomes -1 in order to be represented in a box plot and a violin plot. Next, a box plot and a violin plot is used to compare the results.

3.3 Sentiment Analysis with the PPS Values

The approach that this study adopts uses the probability of positive sentiment (PPS) rather than sentiment analysis utilizing ternary classifications. This section explains the process of this analysis technique.

First, a pre-trained $\mathrm{BERT}_{\mathrm{LARGE}}$ model is obtained from the Hugging Face website.[3]

Second, this deep learning model was trained using the IMDB dataset, which will be called as the $\mathrm{BERT}_{\mathrm{LARGE}}$-IMDB model.

Third, the procedure in Section 3.4 determines the PPS value for each sentence in the IMDB dataset.

Fourth, the $\mathrm{BERT}_{\mathrm{LARGE}}$-IMDB model was assessed in the following ways. The given sentence is marked as *negative* if the PPS value is less than 50. If not, a *positive* label is placed next to the given statement. The labels received are then compared with the original labels in the IMDB dataset. We obtain over 98% of accuracy in this step.

Fifth, the PPS values are calculated for all the sentences in the corpora in Table 1.

Finally, the distributions of PPS values are depicted with either a box plot or a violin plot.

This paper makes use of a violin plot for visual representation. A violin plot, which depicts the peaks in data, is a combination of a box plot and a kernel density plot. It is used to show how numeric data is distributed. Violin plots provide summary statistics as well as the density of each variable, unlike box plots, which can only show summary statistics. A violin plot is superior to a simple box plot because it shows the whole distribution of the data, whereas a box plot just shows summary statistics like mean, median, and interquartile ranges. The difference is especially helpful when the data distribution is multimodal (more than one peak). A box plot with the additional density information does not adequately reveal the structure of the data, but a violin plot does. For this reason, the violin plot is preferred to the box plot when there is sufficient data to determine the density.

3.4 Algorithm for Obtaining the Probabilities of Positive Sentiment

The starting point of the algorithm is the basic architecture of the BERT model in Figure 1. After the BERT model has processed the input sentence, it produces a class label that is either TRUE or FALSE (Devlin et al., 2019: 15).

2) https://finiteautomata/bertweet-base-sentiment-analysis
3) https://huggingface.co/bert-large-uncased

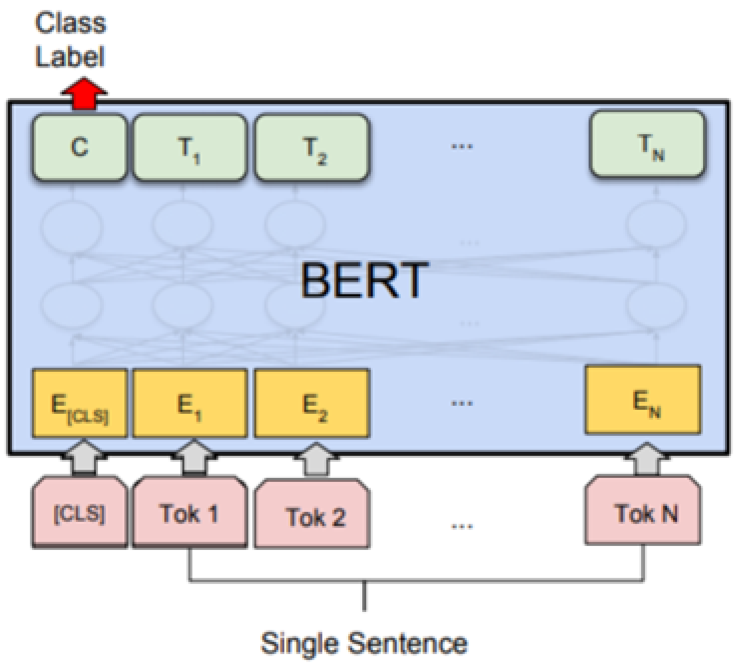

Fig. 1. BERT Model with Single Sentence

In this study, the final output portion has been altered so that the model can now return the probability that the supplied sentence would be TRUE rather than a class label. The $BERT_{LARGE}$-IMDB model returns a tensor containing the probability of FALSE (0) and TRUE (1) during the determination of the class label. They are typically *logit values*. The *inverse logit function* is used to turn them into probability after that. The probability of TRUE is then calculated and converted to a percentage, with values ranging from 0 to 100.[4)]

4. Results

4.1 Ternary Classifications plus Box Plot

The below box plot displays the sentimental distributions for the ternary classification. The first 4 are corpora for suicide notes, while the others are corpora for ordinary texts. Here, the medians are represented by the thick black lines, while the mean values are represented by the black dots. Remember that the values 1, 0, and -1 represent the sentiment labels *positive*, *neutral*, and *negative* respectively (Section 3.2).

4) Lee (2021) and Lee (2022) use similar algorithm, but they use it for calculating acceptability scores of sentences.

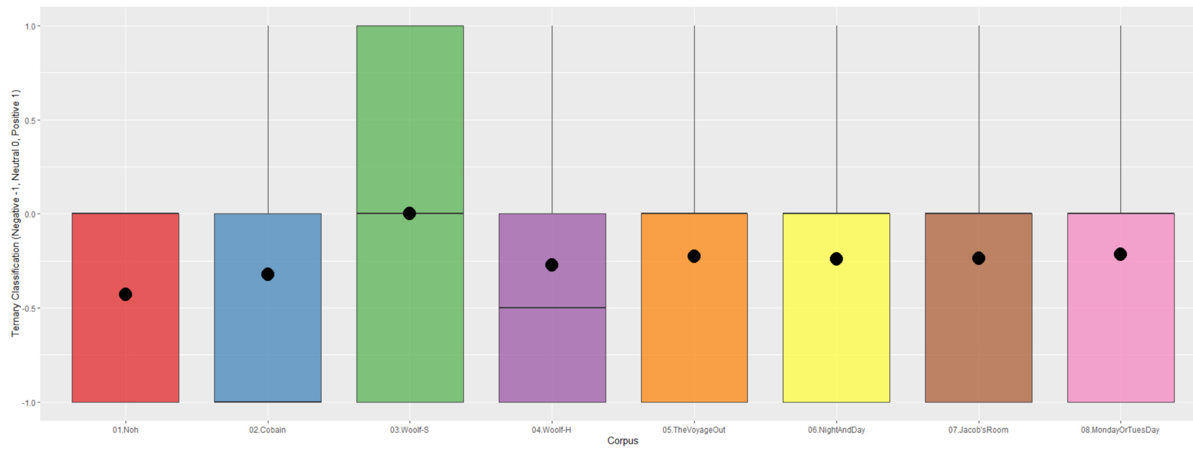

Fig. 2. Box Plot for Ternary Classifications

Although the distributions of sentiment labels in suicide notes (01~04) and ordinary texts (05~08) differ, it is difficult to distinguish between the two. In this box plot, there are no common characteristics that can be utilized to recognize suicide notes. Since there is no discernible tendency in the box plot for suicide notes, it is also challenging to utilize mean values or medians. Particularly, look at the suicide note 03. The sentiments are evenly distributed around zero, as illustrated in the box plot, and the mean and median are similarly nearby.

4.2 PPS Distributions plus Box Plot

The sentiment distributions of the PPS values are displayed in the following box plot.

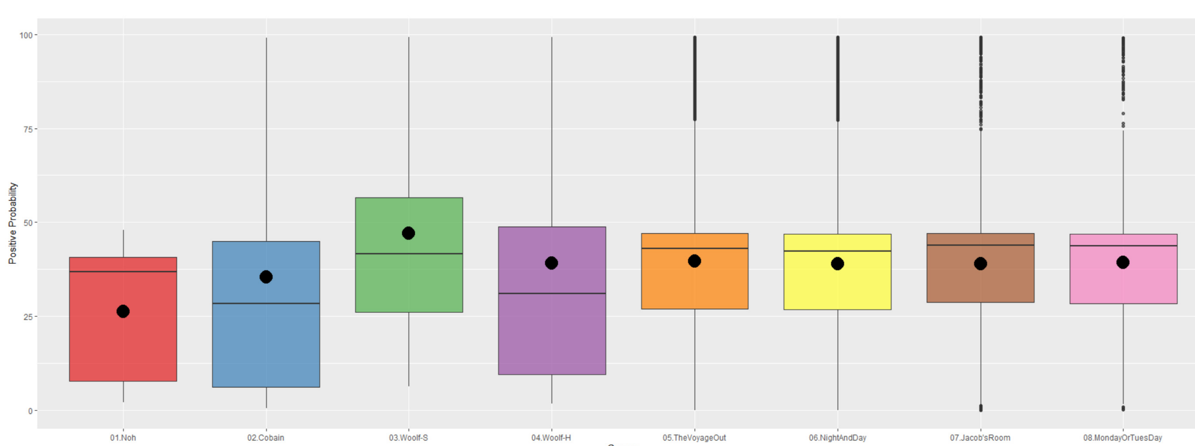

Fig. 3. Box Plot for PPS Values

Here, two intriguing characteristics in the suicide notes (01~04) are shown. First, suicide notes typically have lower medians than ordinary texts (05~08). It suggests that suicide notes employ a lot more negative language. Second, compared to ordinary texts, the box sizes (from the first to third quantile) are larger in the suicide notes. It means that suicide notes have greater sentimental ups and downs than ordinary texts. Also observe that the mean and median values for the suicide note 03 are a little lower in Figure 3 than those in Figure 2. This fact suggests that the overall sentimental distribution is slightly positively skewed in the box plot of Figure 3.

However, it is equally challenging to identify suicide notes using mean values or medians. The mean values of the ordinary texts (05~08) are similar to one another, whereas the mean values of the suicide notes are close to those of the ordinary texts. While some values (suicide note 03) appear greater than the values of the ordinary texts, others (such as suicide note 01) are located below the means of ordinary texts.

4.3 Ternary Classifications plus Violin Plot

The sentiment distributions for the ternary classification are shown in the violin plot below.

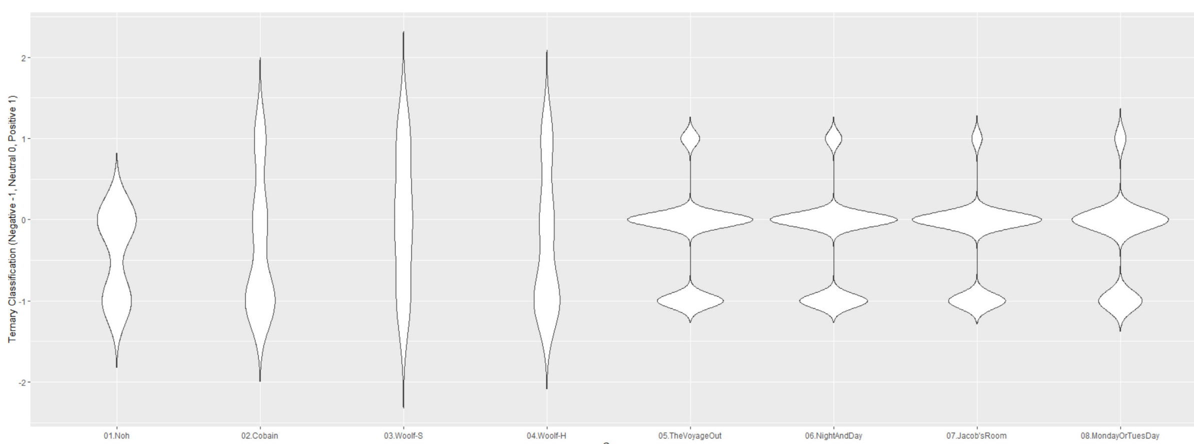

Fig. 4. Violin Plot for Ternary Classifications

The height of each violin plot is proportional to the confidence intervals, and the width of each curve is to the estimated frequency of data points in each region. The short height of the corpora 05~08 shows that they are the huge size of corpora since the confidence intervals are inversely proportional to the number of data points.

Note that the curves in the ordinary texts in this violin plot are substantially thicker (or convexed) around 0 than those in suicide notes. The thickness surrounding the values 1, 0, and -1 represent the distributions of *positive*, *neutral*, and *negative* sentiment respectively, because violin plots show us the density of each variable. The thick or convexed areas around 0 in ordinary texts suggest that people tended to use sentimentally *neutral* statements more frequently than those with *positive* or *negative* feelings. On the other hand, in the suicide note, the area around 0 is not as thick as it is in ordinary texts. Sentimentally neutral sentences are therefore significantly less common in suicide notes.

The shapes of the suicide notes 01~04 are very different from those of the ordinary texts 05~08 in terms of the overall shape. In other words, the suicide notes are not distinguishable visually from the ordinary texts. Accordingly, it is not easy to say that this corpus is a suicide note in this violin plot.

4.4 PPS Distributions plus Violin Plot

The following violin plot displays the sentiment distributions of PPS values.

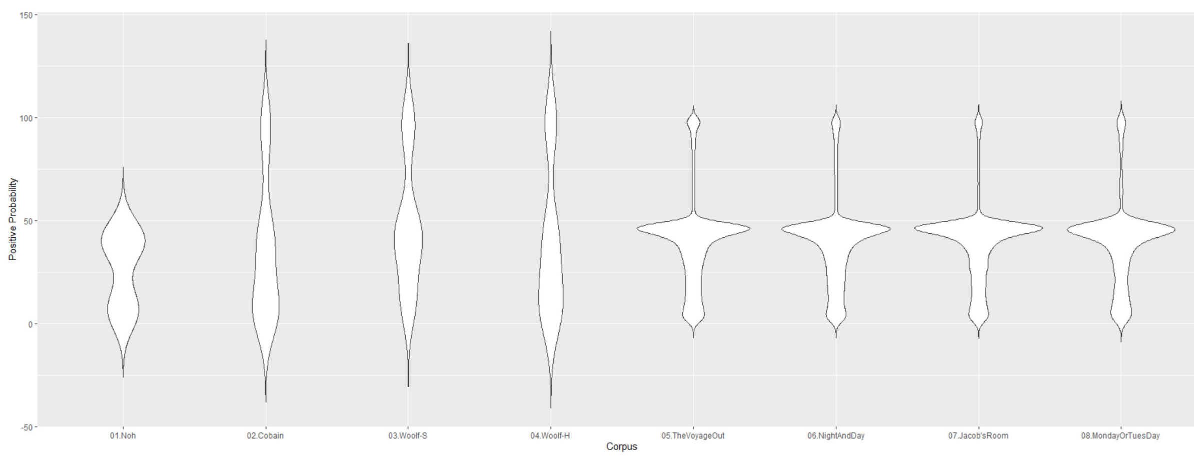

Fig. 5. Violin Plot for PPS Values

Note that while the general shapes of the ordinary texts are somewhat similar, the shapes of the suicide notes (numbered 01~04) are different from those of the ordinary texts (numbered 05~08). As a result, it can be concluded that identifying suicide notes using PPS values and a violin plot is far more accurate than doing so using ternary classifications and a violin plot.

5. Experiment

5.1 Dataset

The method for identifying suicide notes using PPS values clearly distinguishes them from ordinary texts. However, comparing the analysis results of small-size texts (01~04) with those of large-size corpora (05~08) presents a substantial challenge. As seen in Table 1, the former categories of corpora have fewer than 50 sentences, whereas the latter have thousands or more than 10 thousand. Comparing the analysis results in 01~04 or 05~08 might be doable, but is it also possible to compare the analysis results of 01~04 with those of 05~08? We create an experiment to find a solution to this challenge.

First, two different kinds of corpora are constructed. The first is a collection of suicide notes, and the second is a collection of ordinary texts.

Second, 20 sentences are randomly chosen from each type of corpus, and the extracted sentences are combined into a single file. The process is repeated until 100 files of each type of corpus are constructed. This process yields two groups of files, each file of which has exactly 20 sentences (100 files for suicide notes and another 100 files for ordinary texts).

Third, using the algorithm in Section 3.4, the PPS values are determined for each sentence in each file.

Finally, a violin plot is used to depict the distributions of PPS values. The analysis results for suicide notes are then contrasted with those for ordinary texts. Four basic statistics (precision, recall, *F*-score, and accuracy) are used for evaluation.

5.2 Evaluation with Violin Plot

The violin plots for suicide notes are listed below. Keeping this in mind that each type of corpus has 100 files and that each file has exactly 20 sentences.

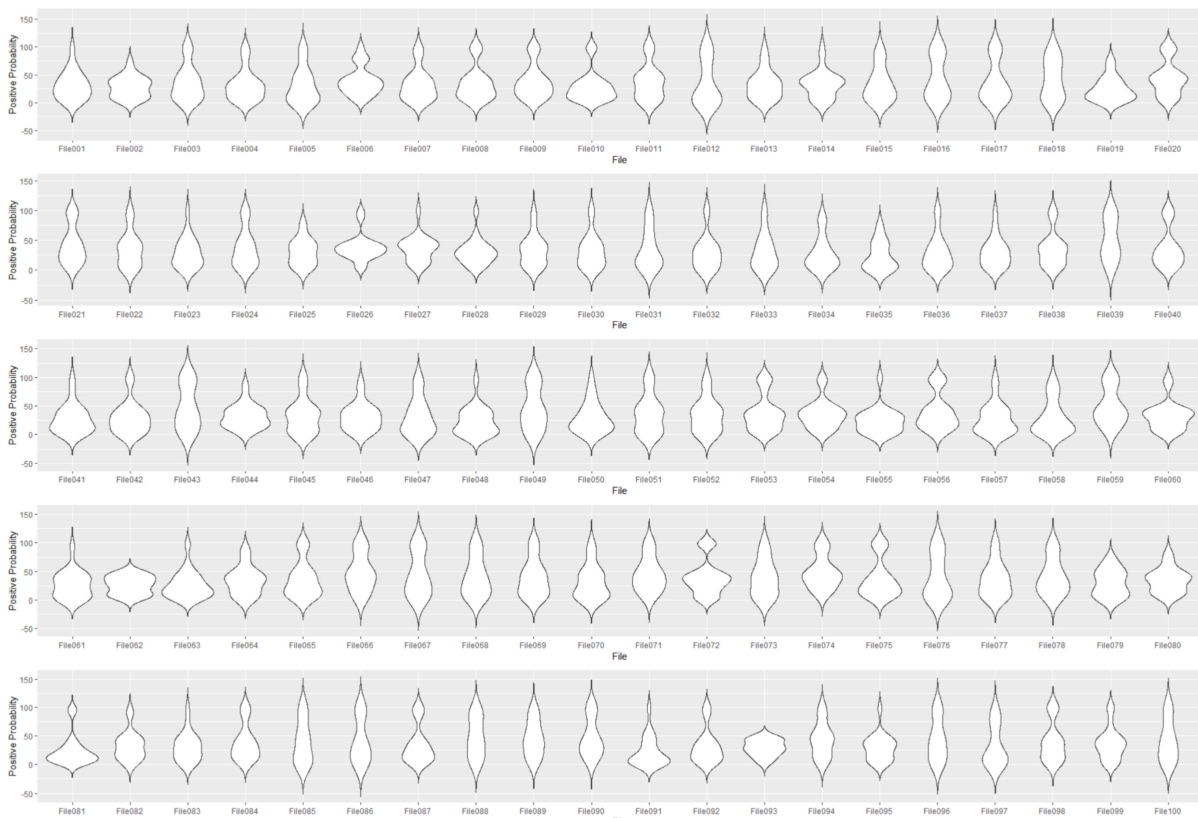

Fig. 6. Violin Plots for Suicide Notes

Very often, the thickest/convexed areas can be found below the line of 50 based on the shape of violin plots. By having more sentimentally negative sentences, this form of violin plot demonstrates one of the characteristics of suicide notes.

The violin plots for ordinary texts are listed here.

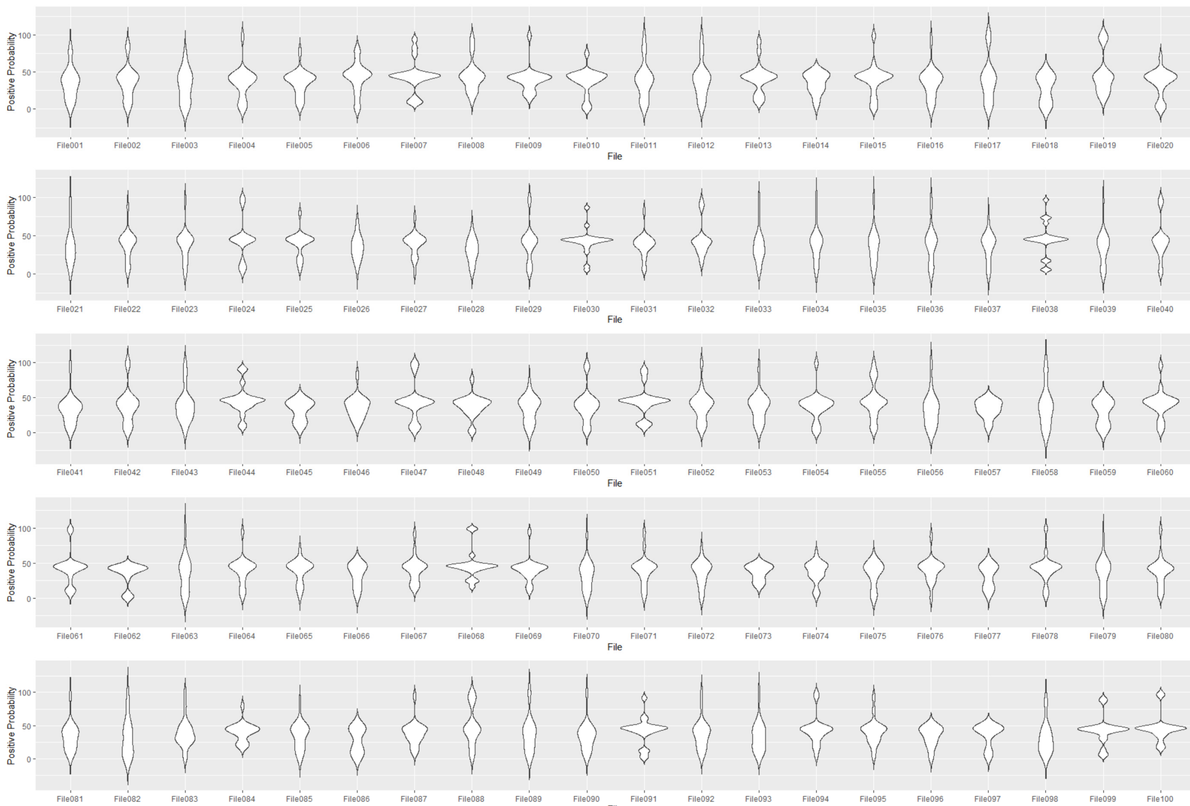

Fig. 7. Violin Plots for Ordinary Texts

The overall shapes of the violin plots show that, in most cases, the most thickest/convexed areas are located near the line of 50.

The fact that this kind of text has so many emotionally neutral sentences demonstrates a characteristic of ordinary texts. The following 18 files are selected during the process in Figure 6: File002, File011, File014, File020, File025, File026, File027, File029, File038, File045, File053, File060, File061, File062, File064, File072, File074, and File080. On the other hand, for the violin plot in Figure 7, all the plots are extracted whose most thickest part is NOT located around 50. The following 11 files are chosen through this process: File026, File028, File033, File058, File081, File082, File083, File086, File089, File093, and File098. The following table shows the evaluation results.

Table 2. Evaluation with Violin Plots

		Predicted	
		Suicide Notes	Ordinary Texts
Actual	Suicide Notes	82	18
	Ordinary Texts	11	89

This table is used to calculate the following evaluation metrics: precision 0.820, recall 0.882, *F*-score 0.850, and accuracy 0.855. Along with the small amount of sentences, the proposed algorithm using PPS values and violin plots correctly divides the texts into two groups (suicide notes and ordinary texts) with 85.5% accuracy.

5.3 Comparison with Previous Studies

The performance of the proposed algorithm is contrasted with that of earlier investigations in the table below.

Table 3. Evaluation Summary

Model	*F1*-score	Accuracy
Yang et al. (2012)	61.39	
McCart et al. (2012)	50.23	
Wang et al. (2012)	50.38	
Desmet and Hoste (2013)	68.86	
Ghosh et al. (2020)		60.71
PPS Values + Violin Plot	**85.00**	**85.50**

The proposed detecting algorithm with PPS values plus violin plots has substantially higher *F*-scores and accuracies, as shown in this table. It suggests that the proposed algorithm performs admirably in identifying suicide notes from ordinary texts.

6. Discussion

6.1 Implications

In this work, a new approach which is based on PPS values and violin plots is suggested. The PPS values are obtained by converting the *logit values* into the PPS values after the $\mathrm{BERT}_{\mathrm{LARGE}}$ model has been trained using the IMDB dataset. A violin plot is then used to graphically show the distributions of PPS values. The distributions of PPS values in the suicide notes are clearly different from those in the ordinary texts (Figure 6 vs. Figure 7). In ordinary texts, the neutral sentiment (around 50) has the most thickest/convexed curves, while the *positive* sentiment parts (around 100) and the *negative* sentiment parts (around 0) have more slender curves. It suggests that sentences with *neutral* sentiment take up far more parts in ordinary texts.

The proposed algorithms for detecting suicide notes (in violin plot) are not only applicable to large size of corpora but also to the small size of corpora. The datasets in Section 5 are built from two separate corpora. One is from the ordinary texts, and the other is from the suicide notes. Twenty sentences are randomly chosen from each of these corpora and constitute one file. In the experiment, a total of 100 files are created for each corpus. Following that, the composed files are examined using PPS values and violin plots. The criterion is whether the most thickest/convexed region of the violin plot is around 50. The given file is declared to be an ordinary text if this requirement is met. Otherwise, it is regarded as a suicide note. In the experiment, we achieved 85.50% of accuracy. It suggests that this criterion for identifying suicide notes (the PPS values + violin plot) applies to both small and large corpora of suicide notes.

The strategy suggested in this study makes use of or actually exposes certain significant properties of suicide notes. The first is that suicide notes generally express more *negative* sentiment than do other ordinary texts. This propensity is expected. When someone decides to commit a suicide, he or she has a profoundly gloomy attitude on their surroundings, and this pessimistic propensity is reflected in the emotional distributions of suicide notes. The box plot in Figure 3 shows this *negative* sentiment. The boxes of suicide notes (01~04) are located a little lower than those of ordinary texts (05~08). The most thickest/convexed areas of Figure 5 are found below line 50.

The second is that the sentiment in suicide notes don't fit neatly into categories; instead, they appear as a spectrum. Even though the thickness is not as pronounced, several sentences with PPS values between 0 and 50 are also visible in the violin plot of Figure 5. This shows that there are some feelings that people experience which aren't quite *positive* or *neutral*. This trait is captured by the method suggested in this study, which helps in the identification of suicide notes. The violin plot in Figure 4 shows that sentiment analysis using ternary classifications falls short in capturing this kind of fine-grained sentiment.

The third characteristic of suicide notes is that they can display a broad spectrum of sentiments. In suicide notes, the majority of the sentences are extremely depressing, yet some of them are really encouraging. These sentiment may come to mind for suicide victims when they reflect on their former joyful times. As a result, the sentiment distributions may cover a large range of spectrum. The box plot in Figure 3, in which the boxes for suicide notes have greater height, visually displays this tendency.

The fact that the suicide notes display different qualities from ordinary texts is another intriguing characteristic, despite being written by the same author. Virginia Woolf is the same author of the works 03~04 and 05~08, as was stated in Section 3.1. Figure 5 illustrates how significantly different the sentiment patterns of 05~08 are from those of 03~04. The less convex curve of around 50 is seen in the suicide notes, whereas the most convex part of around 50 is seen in the ordinary texts. It demonstrates that the sentiment patterns of suicide notes are quite different from those of ordinary texts despite the fact that these two different types of texts are generated by the same person.

As mentioned in Section 2, forensic linguistics is used to examine suicide notes (Coulthard and Johnson, 2016; Olsson, 2004; Olsson, 2008; Svartvik, 1968). Expressions like apologies, affection, rage, complaints, or psychological shock are used in some research, such as Chaski (2012). Certain language units are used by Olsson (2008), and some studies use discourse analysis (Edelman and Renshaw, 1982) or semantic space (Leenaars, 1988). The LIWC, which incorporates a variety of linguistic traits, is used by Pestian et al. (2008) and Pestian et al. (2010). However, because suicide notes typically have short length, these kinds of methods might not be available when looking into suicide notes. Some linguistic features of forensic linguistics may not appear in such a brief text due to the length of the texts. The detection algorithm with sentiment analysis, however, may be able to extract more accurate cues from such a short text because sentiment can be measured for every sentence regardless of the length of texts. Also, more research is required to determine how the classifications in Durkheim (1951) or Shneidman (1996) might be related to the sentiment distributions of suicide notes (the distributions of PPS values).

Now go back to the hypotheses in Section 1. The first hypothesis is that the suicide notes will exhibit different characteristics from the ordinary texts. The violin plot in Figure 5 lends support to this hypothesis. The suicide notes have a less thickest/convexed curve of around 50. On the other hand, ordinary texts have the most thickest/convexed part around 50. The second hypothesis states that suicide notes will exhibit unique characteristics from ordinary texts despite being written by the same author. The evidence also supports this hypothesis. Figure 5 shows a clear distinction between emotional patterns between 03~04 and 05~08. It shows that even though these two very different sorts of writings are written by the same author, the sentiment patterns of suicide notes are very different from those of ordinary texts. As a result, it can be argued that all of the hypotheses in Section 1 are supported by the analysis results in Section 4 and Section 5.

6.2 Applications and Limitations

The proposed algorithm in this study can be used to analyze both small size of suicide notes and huge size of corpora. The developed algorithm can therefore be used to identify suicide indicators in online texts like Twitter or SNS, in addition to detecting suicide signs in off-line texts. As was indicated in Section 1, most suicide notes are short. It is difficult to use some linguistic aspects of forensic linguistics because of this short length (Section 2.1 and Section 2.2). Applying the algorithm to detect suicide notes is very helpful because the developed algorithm in this paper is also available for the small size of online texts.

Next, take note that the first corpus in our study (01) is a suicide note of a Korean politician. While being originally published in the Korean language, the English translation was made for research purposes. This corpus illustrates the usual characteristics of suicide notes, as shown in Figure 5. This suggests that the proposed algorithm is also accessible to English translations from other languages. It also provides a useful detection algorithm to English translations.

The developed algorithms in this paper do, however, suffer from two significant shortcomings. First, there is a big discrepancy between precision and recall. Precision and recall in the evaluation using violin plots are 0.820 and 0.882, respectively. Hence, low precision suggests a high number of false positives (the texts which are actually suicide notes but wrongly classified as ordinary texts). The reason appears to have its roots in the short length of texts. Therefore, it is essential to develop more advanced approaches that can solve this problem. Second, we have to manually examine all of the violin plot. It would be better if this process is automated.

7. Conclusion

In this paper, a new technique was proposed for identifying suicide notes which makes use of sentiment analysis. 8 corpora are built for this purpose, of which 4 were the suicide note and the others are ordinary texts. The $\text{BERT}_{\text{LARGE}}$ model was used to determine the probability of positive sentiment (PPS) for each sentence in the corpora. Following the analysis, the PPS values are visually depicted with a box plot and a violin plot.

Analysis results show that suicide notes have a less convexed curve of around 50, whereas ordinary texts have the most thickest/convexed parts around 50. Additionally, it was found that the algorithm suggested in this study is not limited to large size of corpora but can also be used to analyze small size of texts. Another fascinating finding is that, despite being written by the same person, suicide notes have a distinctive style from the ordinary texts. We believe the suggested algorithm can be used to identify suicide warning signs in a variety of online and offline suicide notes. Additionally, we anticipate that more sophisticated deep learning technology will be used to save more lives.

References

Chaski, C. 2012. Author Identification in the Forensic Setting. In L. Solan and P. Tiermsa (eds.), *The Oxford handbook of forensic linguistics*. Oxford, UK: Oxford University Press, 333-372.

Coulthard, M. and Johnson, A. 2016. *An introduction to forensic linguistics*. Cambridge, MA: Cambridge University Press.

Desmet, M. and Hoste, V. 2013. Emotion Detection in Suicide Notes. *Expert Systems with Applications 40*.16, 6351-6358.

Devlin, J., Chang, M., Lee, K., and Toutanova, K. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint 2019, arXiv:1810.04805.

Durkheim, E. 1951. *Suicide*. New York: The Free Press.

Edelman, A. and Renshaw, L. 1982. Genuine versus Simulated Suicide Notes: An Issue Revisited through Discourse Analysis. *Suicide and Life-Threatening Behavior 12*.2, 103-113.

Ghosh, S., Ekbal, A., and Bhattacharyya, P. 2020. CEASE, a Corpus of Emotion-Annotated Suicide Notes in English. In *Proceedings of the 12th language resources and evaluation conference*. Marseille, France, 11-16 May 2020, 1618-1626.

Lee, Y. 2021. English Island Constraints Revisited: Experimental vs. Deep Learning Approach. *English language and linguistics 27*.3, 21-45.

Lee, Y. 2022. Negative Polarity Items in English: A Deep Learning Model and Statistical Analysis. *Korean journal of linguistics 47*.1, 29-56.

Lee, Y. and Joh, G. 2019. Identifying Suicide Notes Using Forensic Linguistics and Machine Learning. *The Linguistic Association of Korean Journal 27*.2, 171-191.

Leenaars, A. 1988. *Suicide notes*. New York: Human Sciences Press.

McCart, J., Finch, D., Jarman, J., Hickling, E., Lind, J., Richardson, M., Berndt, D., and Luther, S. 2012. Using Ensemble Models to Classify the Sentiment Expressed in Suicide Notes. *Biomedical Informatics Insights 5*.1, 77-85.

Mitchell, T. 1997. *Machine learning*. New York: McGraw Hill.

Olsson, J. 2004. *Forensic linguistics: An introduction to language, crime, and the law*. London, UK: Continuum.

Olsson, J. 2008. *Forensic linguistics: An introduction to language, crime, and the law*, 2nd edition. London, UK: Continuum.

Pennebaker, J. and King, L. 1999. Linguistic Styles: Language Use as an Individual Difference. *Journal of Personality and Social Psychology 77*.6, 1296-1312.

Pennebaker, W., Francis, E., and Booth, J. 2001. *Linguistic inquiry and word count (LIWC): LIWC2001*. Mahwah, NJ: Lawrence Erlbaum Associates.

Pestian, J., Matykiewicz, P., Grupp-Phelan, J., Lavanier, A., Combs, J., and Kowatch, R. 2008. Using Natural Language Processing to Classify Suicide Notes. In *Proceedings of the workshop on current trends in biomedical natural language processing (BioNLP'08)*. Columbus, Ohio, USA, 19 June 2008, 96-99.

Pestian, J., Matykiewicz, P., and Linn-Gust, M. 2012a. What's in a Note: Construction of a Suicide Note Corpus. *Biomedical Informatics Insights 5*.5, 1-6.

Pestian, J., Matykiewicz, P., Linn-Gust, M., South, B., Uzuner, O., Wiebe, J., Cohen, B., Hurdle, J., and Brew, C. 2012b. Sentiment Analysis of Suicide Notes: A Shared Task. *Biomedical Informatics Insights 5*.1, 3-16.

Pestian, J., Nasrallah, H., Matykiewicz, P., Bennett, A., and Leenaars, A. 2010. Suicide Note Classification Using Natural Language Processing: A Cntent Aalysis. *Biomedical Informatics Insights 3*.3, 19-28.

Samuel, A. 1959. Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal 3*, 210-229.

Shapero, J. 2011. *The language of suicide notes*. Unpublished doctoral dissertation, University of Birmingham, Birmingham, UK.

Sheidman, E. and Faberow, N. 1963. *Clues to suicide*. New York: McGraw-Hill.

Shneidman, S. 1996. *The suicidal mind*. Oxford, UK: Oxford University Press.

Svartvik, J. 1968. *The Evans statements: A case for forensic linguistics*. Gothenburg, Sweden: University of Gothenburg Press.

Tausczik, Y. and Pennebaker, J. 2010. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology 29*.1, 24-54.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, L., and Polosukhin, I. 2017. Attention is all you need. arXiv preprint arXiv:1706.03762.

Wang, W., Chen, L., Tan, M., Wang, S., and Sheth, A. 2012. Discovering Fine-Grained Sentiment in Suicide Notes. *Biomedical Informatics Insights 5*.1, 137-145.

Yang, H., Willis, A., De Roeck, A., and Nuseibeh, B. 2012. A Hybrid Model for Automatic Emotion Recognition in Suicide Notes. *Biomedical Informatics Insights 5*.1, 17-30.

Lee, Yong-hun, Instructor
99 Daehak-ro, Yuseong-gu, Daejeon 34134, Korea
Department of Linguistics, Chungnam National University
E-mail: yleeuiuc@hanmail.net