Cross Validation of Cepstral Coefficients in Classifying Suicidal Speech from Depressed Speech

Thaweesak Yingthawornsuk

Abstract—This paper describes a way to distinguish high risk suicidal patients from depressed patients using mel-frequency cepstrum coefficients. Distinguishing high risk suicidal patients from the depressed patients is an important problem. A solution to this problem may help to prevent loss of many lives. In this study, the vocal characteristics of female and male patients’ speech samples were analyzed and ten mel-frequency cepstrum coefficients were used to classify high risk suicidal patients and depressed patients. The cross validation scheme was used to observe the classification performance. There were two different types of speech samples from both male and female patients. One of them was the speech samples during the interview and the other one is speech samples during the reading session.

Keywords—Pairwise Classification, Speech, MFCC, Depression

I. INTRODUCTION

It is reported in [1] that mental disorders are very common in the United States and internationally. %26.2 percent of Americans whose age is 18 or older suffers from a diagnosable mental disorder in 2005. In the same year, major depressive disorder affects just about % 6.7 percent of the U.S. population. [1] Another statistics shows that, major depressive disorder is more common in women than in men. [2] In addition, more than %90 percent of the people who committed suicide had a diagnosable mental disorder, most commonly a depressive disorder. [3] So, the depression and suicide has an important relationship. An interesting statistic shows that men die because of a suicide four times more than women die. [4] However, men attempt suicide two to three times as seldom as women. [5]

As it can be seen from these statistics, suicide is an important public health problem and suicide has a strong relationship with depression. Therefore, it is very important to evaluate a depressed patient’s risk of committing suicide. The psychiatrists evaluate this risk using some techniques, like Hamilton depression rating scale. [6] On the other hand, it is widely known that the psychological state has an effect on the person’s speech production system. It is proposed by S. E. Silverman that vocal parameters of human speech could be used for making a decision if a patient is suicidal or not. [7]

Some researchers have studied the relationship between vocal tract characteristics and the suicidal risk. Tolkmitt et al. compared the patients' speech's formant information of vowels that occurs in the identical phonetic context during the patient's recovery period. [8] France et al. observed long term averages of the formant information and checked the difference to distinguish the high risk suicidal patients from depressed and controlled patient groups. [9] Yingthawornsuk et al. used the percentages of the total power, its highest peak value and its frequency location to distinguish between high risk suicidal, depressed and remitted (had been depressed previously but recovered) groups. [10] Ozdas used lower order mel-cepstral coefficients to distinguish high risk suicidal patients from non-suicidal ones using Gaussian mixture models and unimodal Gaussian models. [11]

Mel-frequency cepstral coefficients are useful parameters that had been used in many speech processing systems, like [11], Logan proposed using mel-frequency cepstral coefficients for modeling music. [12] Godino-Llorente et al. used short term mel-cepstral parameters for pathological voice quality assessment. [13] Choi worked on compensating the mel-frequency cepstral coefficients for speech recognition in noisy environments. [14]

This paper presents the work on distinguishing high risk suicidal patients from depressed patients using the first ten mel-frequency cepstrum coefficients for female and male patients. The cross validation scheme was used to calculate the classification performance. The optimal mel-frequency cepstrum coefficients are found for female and male patients and for both the reading and interview sessions of each gender.

II. METHODOLOGY

A. Information about the Database

The database of this research is obtained from a continuing American Foundation for Suicide Prevention supported study of Psychiatry Department of the Vanderbilt University School of Medicine. The database is composed of male and female subjects whose ages are between 25 and 65 years of old. A psychiatrist categorized these patients as either depressed or at high risk of suicide patients. The number of the female patients that are used in this study is shown in Table 1 and the number of male patients is shown in Table 2.

Thaweesak Yingthawornsuk is with King Mongkut’s University of Technology Thonburi (KMUTT), Thailand.
The database consists of two different types of speech samples. One of the types is recorded while the patient was interviewing with a therapist. This type of speech sample is named as "Interview Session". The other one is named as "Reading Session". This type of speech sample was gathered during the patient was reading a predetermined part of a book. The reading passage was "Rainbow Passage" Error! Reference source not found. which is phonetically balanced and has all of the normal sounds in spoken language. The same recording environment was used for all patients.

B. Preprocessing

All speech signals were digitized by using a 16 – bit analog to digital converter at a sample rate of 10 kHz with an anti – aliasing filter. GoldWave v.5.08 audio editor was used to remove the silences which are longer than 0.5 seconds and the voices that is not belong to the patient. In this study, 76 seconds of female patient's continuous speech were stored for the mel-frequency cepstrum study both for interview and reading sessions. For male patients, 66 seconds of continuous speech were stored for the same purpose. For all patients' speech data, the voiced and unvoiced detection was performed and after that only voiced segments were used for subsequent analysis.

C. Information about the System's Hardware

For the recording purpose, a portable audio data acquisition system was used. This system composed of a Sony VAIO laptop computer with a 2GHz Pentium IV CPU, 60 GB hard drive, 512 Mb memory, 20X CD/DVD read/write unit, 250 GB external hard drive, Windows XP Operating System, ProTools LE digital audio editor, a Digital Audio MBox for audio signal acquisition; and an Audix SCX – one cardioid microphone.

D. Feature Extraction

In this study, mel – frequency cepstrum coefficients were extracted as the feature. Ten mel-frequency cepstrum coefficients were computed for each patient’s speech sample. Each speech signal was divided into 512 points of voiced segments. For each voiced speech segment the log – magnitude spectrum was computed. This achieved by taking the logarithm of the discrete Fourier transform (DFT). The spectrum was then filtered by a series of 16 triangular band-pass filters. The filter bank that is used in this work is similar to that was employed by Davis and Mermelstein [16] which simulates the critical band filtering by a set of triangular band-pass filters. The bandwidths and center frequencies of these filters are chosen according to the mel - scale.

Human ear is more sensitive to changes in the low frequency portion of the frequency spectrum. [17] Mel – scale was formulated for the sampling of the frequency spectrum based on this property of human auditory perception. The linear frequency spectrum was mapped based on the human auditory perception with mapping approximately linear on the 0 – 1 kHz range and logarithmic above 1 kHz. This mapping is shown with Fig. 1. The following formula is the suggested formula that models this relationship in which Fmel is the perceived frequency and FHz is the real frequency.

\[
F_{\text{mel}} = 2595 \log\left(1 + \frac{F_{\text{Hz}}}{700}\right)
\]

(1)

Fig. 1 Mapping between linearly spaced frequency and mel – scale frequency

The vocal tract length normalization was performed to each patient. The bandwidths and center frequencies of the filters in the mel – scale filter bank were then adjusted according to this normalization factor. [18] Then the last step is to calculate the inverse discrete Fourier transform (IDFT) to obtain the mel-frequency cepstrum coefficients. The procedure for extraction of mel-frequency cepstrum coefficients is shown in Fig. 2 below. After ten mel-frequency cepstrum coefficients were calculated for each frame, the values in all frames are averaged to have one value for each mel-frequency cepstrum coefficient.
E. Cross-Validation Classification

The k – fold cross validation technique [19] with quadratic discriminant function was performed on the mel – frequency cepstrum coefficients data. The data files were split randomly into two subsets. One of them is for training the data and the other one is for testing the data. % 65 of the data was used to train the data for estimating the quadratic classification function. Then using this quadratic classification function, %35 of the data was tested to perform the classification. The variance of the performance estimates was reduced by averaging the results from 10 different runs of cross validation. The cross validation classification is performed for each mel – frequency cepstrum coefficient separately. The cepstral coefficient that gave the maximum classification result is determined. Next, this cepstral coefficient is paired with all the other cepstral coefficients and cross validation classification is performed again. Two cepstral coefficients that gave the maximum classification result are determined. The same process is repeated for three cepstral coefficients that gave the maximum classification. Three classification performances (one coefficient performance, two coefficients performance, and three coefficients performance) are then compared and then the largest one is assigned as the optimal coefficients.

This performance testing is performed for determining only the maximum depressed classification, and then only for the maximum high risk suicidal classification, and finally for the maximum total classification of depressed – high risk suicidal classification.

III. RESULTS

The depressed- high risk suicidal pairwise classification using k-fold cross validation technique was performed for finding the optimal coefficient(s) that gave the maximum classification performance. The table that is shown in Table 3 shows the results for male interview session. The result for male reading session is shown in Table 4.

**TABLE III**

<table>
<thead>
<tr>
<th></th>
<th>Optimal Coefficient(s)</th>
<th>Classification Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max. Depressed</td>
<td>Coefficients 1 and 4</td>
<td>%78.60</td>
</tr>
<tr>
<td>Max. High Risk Suicidal</td>
<td>Coefficient 3</td>
<td>%86.00</td>
</tr>
<tr>
<td>Max. Total Classification</td>
<td>Coefficient 3</td>
<td>%77.20</td>
</tr>
</tbody>
</table>

**TABLE IV**

<table>
<thead>
<tr>
<th></th>
<th>Optimal Coefficient(s)</th>
<th>Classification Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max. Depressed</td>
<td>Coefficients 2, 9, and 1</td>
<td>%89.80</td>
</tr>
<tr>
<td>Max. High Risk Suicidal</td>
<td>Coefficient 2</td>
<td>%93.00</td>
</tr>
<tr>
<td>Max. Total Classification</td>
<td>Coefficient 2</td>
<td>%78.00</td>
</tr>
</tbody>
</table>

The optimal features for depressed classification are coefficients 1 and 4 for the interview session with a classification performance of %78.60; on the other hand optimal features are coefficients 2, 9 and 1 for the reading session with a classification performance of %89.80. Coefficient 3 is the optimal feature for both high risk suicidal classification and total classification of depressed – high risk suicidal classification with a classification performance of %86 and %77.20 respectively.

The optimal feature for both high risk suicidal classification and total classification of depressed – high risk suicidal classification of the reading session is coefficient 2. The classification performance was %93 for the high risk suicidal classification and %78 for the total classification of depressed – high risk suicidal classification.

**TABLE V**

<table>
<thead>
<tr>
<th></th>
<th>Optimal Coefficient(s)</th>
<th>Classification Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max. Depressed</td>
<td>Coefficients 1, 5, and 7</td>
<td>%78.90</td>
</tr>
<tr>
<td>Max. High Risk Suicidal</td>
<td>Coefficient 9</td>
<td>%70.10</td>
</tr>
<tr>
<td>Max. Total Classification</td>
<td>Coefficient 9</td>
<td>%66.40</td>
</tr>
</tbody>
</table>

The Table 5, shows the results for female interview session. The optimal features for depressed classification are coefficient 1, 5 and 7 with a classification performance of %78.90; on the other hand the optimal feature is coefficient 9 for both high risk suicidal classification and total classification of depressed – high risk suicidal classification with a classification performance of %70.10 and %66.40 respectively.

**TABLE VI**

<table>
<thead>
<tr>
<th></th>
<th>Optimal Coefficient(s)</th>
<th>Classification Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max. Depressed</td>
<td>Coefficients 3, and 2</td>
<td>%70.10</td>
</tr>
<tr>
<td>Max. High Risk Suicidal</td>
<td>Coefficient 8</td>
<td>%71.10</td>
</tr>
<tr>
<td>Max. Total Classification</td>
<td>Coefficient 9</td>
<td>%63.90</td>
</tr>
</tbody>
</table>

Table 6 demonstrates the female reading session classification results and optimal features. The optimal features for depressed classification are coefficients 3 and 2 with a classification performance of %70.10. For high risk suicidal classification, the optimal feature is coefficient 8 with a classification performance of %71.10. Coefficient 9 is the optimal coefficient for the total classification of depressed – high risk suicidal classification with a classification performance of %63.90.

IV. DISCUSSION AND CONCLUSION

In this paper, it was demonstrated that mel-frequency cepstrum coefficients is a good indicator for discriminating the high risk suicidal patients and depressed patients. The results
showed that for the male and female patients, the mel-frequency cepstrum coefficients are a good indicator for discriminating the high risk suicidal patients and discriminating depressed patients with a performance more than %70. The controlled reading tended to give better results for male subjects especially for high risk suicidal classification and depressed classification. The total classification was about the same for both reading session and interview sessions. The maximum classification results that are obtained from the male subjects are much better than the female subjects' results. This result is an important and useful result since classifying suicidal risk for men is much more important than classifying women, since men die because of a suicide four times more than women die [4] as mentioned in the introduction part of the paper.

REFERENCES