IJAMEC
e-ISSN: 2147-8228

INTERNATIONAL JOURNAL OF APPLIED MATHEMATICS ELECTRONICS AND COMPUTERS

www.dergipark.org.tr/ijamec

International Open Access

Volume 08
Issue 04

December, 2020

Research Article

Gender Determination Using Voice Data

Yavuz Selim Taspinar [a*], ***Mucahit Mustafa Saritas*** [b], ***İlkay Cinar*** [c] ***Murat Koklu*** [c]

[a]*Doganhisar Vocational School, Selcuk University, Doganhisar, 43930, Konya, Turkey*
[b]*Biomedical Engineering Department, Faculty of Technology, Selcuk University, 42250, Konya, Turkey*
[c]*Computer Engineering Department, Faculty of Technology, Selcuk University, 42250, Konya, Turkey*

ARTICLE INFO

Article history:
Received 12 October 2020
Accepted 24 November 2020

Keywords:
Voice
Gender recognition
Regression
Machine learning

ABSTRACT

The rapid advancement of today's technologies, it is tried to facilitate whichever system will be used by using voice features such as person recognition and speech recognition by making use of the voices of the users. Organizations serving in these systems need less manpower and facilitate the operation by helping users faster. The decision-making process using sound features is a very challenging process. With gender recognition, which is one of these steps, it is possible to address the user by gender. In this study, it is aimed to define the genders according to the voices in terms of both forensic informatics and the rapid and accurate progress of the processes. In this study, 3168 male and female voice samples were taken as a dataset. Sound samples were first analyzed by acoustic analysis in R using seewave and tuneR packages. Artificial neural networks were used in the classification stage. In order to increase the classification accuracy, the dataset was divided into 10 parts and each part was excluded from training for testing and used for retesting. Average classification success was found by taking the arithmetic mean of the results. In the classification made with artificial neural networks, male and female voices could be distinguished from each other with a success of 97.9%.

1. Introduction

Speech recognition, which has an important role in human machine interaction, has been used frequently recently. Factors such as environmental conditions, accent and diction can affect the success of voice recognition systems. However, the sound signal samples in the dataset to be created cannot be taken from people with the same environmental conditions, accents and diction. This situation can aggravate the burden of the voice recognition system. In order to overcome this problem, large-volume datasets should be used and the number of attributes of sound samples should be selected at the optimum level [1]. Accent recognition and gender recognition from voices is easy for humans, but not easy to identify gender by computer. There are many studies in literature to find a solution to this issue. Studies have been conducted on effective feature extraction and high accuracy classification architectures to determine the speaker gender from voice signals [2]. Gender recognition by voice gives people the opportunity to help people more by being used in health information systems and education [3]. In a voice recognition study on telephone applications, the vibration, noise ratio, sparkle and frequency properties of the sound were used, and voice recognition was performed with different techniques such as bayesian networks [4]. Sound data, including information about the age of the speaker, were tried to be estimated using artificial neural networks, but could not be successful because the sound samples showed similar characteristics [5]. With the changes made on the Random Forest algorithm, a classification success rate of 96.7% in gender recognition from voice was achieved [6].

Gender recognition studies have been made not only in the field of sound but also from the movements on the screen of touch screen phones and a success of 93.65% has been achieved [7]. In a study on gender and age estimation with fully-connected and convolutional neural networks

* Corresponding author. E-mail address: ytaspinar@selcuk.edu.tr
DOI: 10.18100/ijamec.809476

using voice data collected from German speakers, age recognition rate was found to be 57.53%, and gender recognition rate was 88.8% [8]. Estimating the emotional state of the speakers is a very challenging task as it is influenced by many factors such as thought, mood, behavior and personality. Gender determination in emotion recognition is a factor that increases prediction success. There are studies focusing on gender recognition with gradient enhancing machines and a different version of the Random Forest algorithm [9].

In a study where classification and regression algorithms were combined and used in gender recognition from voice, an ensemble method was created using the Support Vector Machine (SVM), Neural Network and Random Forest methods. This structure was more successful than the methods used singularly in the study [10]. Different methods used in feature extraction from sound data and the selection of effective features among the extracted features are among the factors affecting the success of classification. PCA (Principal component analysis) is a method that enables the reduction of the number of features and obtaining new effective features by making use of the similarities between features. In a study using PCA and SVM algorithm, gender was estimated from voice data and 98.42% success was achieved [11]. Deep neural networks have recently attracted considerable attention as a method that increases the success of classification by detecting hidden features in data. A success of 96.74% was achieved in a study of gender recognition using the deep neural network method [12].

In this study, a dataset containing 20 features and 1 label obtained from the audio data was used. Classification process has been carried out with the Neural network. The material and method used in the second part of the article, the experimental results obtained in the third part, the discussion and the results in the fourth chapter are given.

2. Material and Methods

2.1. Dataset

Acoustic analysis of the dataset used in the study [13] has been done before and 20 sound features and 1 classification label have been added to the dataset [14]. The dataset consists of 3168 rows and 21 columns. There are 1584 male and 1584 female sound samples. The sound characteristics and descriptions obtained as a result of pre-processing are shown in Table 1. The effect of these features in the dataset on the success of the classification may differ. Classification success may increase when some of these features are removed [15].

Table 1. Sound features in the dataset

Properties	**Description**
duration	signal lenght
meanfreq	frequency
sd	frequency deviation
median	avarage frequency
Q25	first quantile
Q75	third quantile
IQR	interquantile range
skew	skewness
kurt	kurtosis
sp.ent	spectral entropy
sfm	spectral flatness
mode	mode frequency
centroid	frequency centroid
peakf	peak frequency
meanfun	avarage frequency
minfun	minimum frequency
maxfun	maximum frequency
meandom	avarage dominant frequency
mindom	minimum dominant frequency
maxdom	maximum dominant frequency
dfrange	range of dominant frequency
modindx	modulation index

2.2. Confusion matrix

The evaluation of a classifier model is not only based on success rate. There are different parameters required for this process. The table required to calculate these parameters is called a confusion matrix. It shows in which category each data in the confusion matrix dataset is classified. Various parameters are obtained by calculating the values in this table and information about the performance of the classifier can be obtained. Table 2 shows the description of the confisuon matrix and the values it contains.

Table 2. Confusion matrix

		Predicted	
		Positive	**Negative**
Actual	**Positive**	True positive (TP) : Positive samples predicted correctly	False negative (FN) : Incorrectly predicted negative samples
	Negative	False positive (FP) : Incorrectly predicted positive samples	True negative (TN) : Negative samples predicted correctly

By using these values, accuracy, precision, recall and F-1 score values can be obtained. The purpose and formulation of these values are shown in Table 3.

Table 3. Accuracy, recall, precision and f1 score parameters

Accuracy	*(TP+TN)/Total*	Returns the proportion of correctly predicted samples from all samples
Recall (r)	*TP/(TP+FN)*	Returns the correct classification rate of positive samples
Precision (p)	*TP/(TP+FP)*	Returns the ratio of correctly classified positive samples to total positive samples
F1-Score	*(2*p*r)/(p+r)*	It is the harmonic mean of sensitivity and evaluates recall with precision.

2.3. Artificial Neural Network

They are systems that can provide learning in a way that can make inferences from input and output data by imitating the cells and functioning principle in the human brain. Different approaches can be preferred in the artificial neural networks (ANN) classification model due to the data required to make predictions in voice recognition systems. In ANN, the training process is carried out first. During this training process, neurons communicate with each other and all neurons have weights that enable the network to learn. It takes time to create these weights during training. However, in the second stage, the test stage, the information about which output an input will give is faster [16]. The training of the model was carried out by determining the activation function ReLu (Rectified Linear Unit), the optimization function Adam, the learning rate 0.0001, and the number of iterations as 200. The artificial neural network model, which consists of 3 layers, 20 inputs, 100 hidden and 2 output neurons, is shown in Fig. 1.

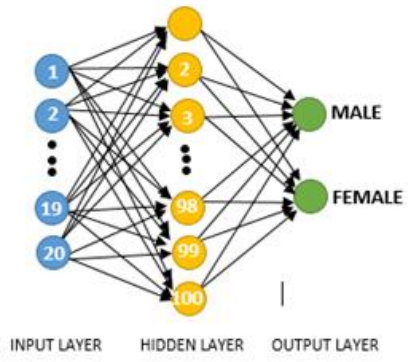

Figure 1. Artificial neural network model used in the study

3. Experimental results

In the study, artificial neural networks were used as a classifier in gender recognition processes by using voice signals properties. Activation function ReLu (Rectified linear unit) is used in artificial neural networks. The iteration number was set at 200. 1584 of 3168 voice data belong to male and 1584 female speakers. The cross validation technique was used for the reason that the classification result is considered to be a reliable value. In this technique, the dataset is divided into k parts. In each training process, the k-1 part of the dataset is reserved for training. The remaining piece is used for testing. This process continues until k different parts are used for testing. In other words, training and testing is done as much. In this study, k value was determined as 10. The average success rate obtained from the classification made by using sound features was found to be 97.9%. The confusion matrix obtained as a result of the classification is shown in Table 4.

Table 4. Confusion matrix resulting from classification

		PREDICTED	
		FEMALE	MALE
ACTUAL	FEMALE	1554	30
	MALE	37	1547

True positive value was 1554, true negative value was 1547, false negative value was 30, and false positive value was 37. The values obtained for classification performance using these values are shown in Table 5.

Table 5. Performance Metrics

Measure	Formula	Results
Accuracy	*(TP+TN)/Total*	97.9%
Recall (r)	*TP/(TP+FN)*	98%
Precision (p)	*TP/(TP+FP)*	97.7%
F1-Score	*(2*p*r)/(p+r)*	97.9%

The fact that the values in Table 5 are very close to each other is due to the fact that the numbers of data classified correctly and incorrectly are very close to each other. In addition, the high rates show that the classifier is successful in training and testing.

There are studies in the literature made with the same dataset. The comparison of these studies is given in Table 6.

Table 6. Comparison with literature studies

Methods	Acc. (%)	Studies
Sequential Minimum Optimization	98.0	Liviersi et al. [17]
kNearest Neighbor	97.5	Liviersi et al. [17]
Decision Tree	96.2	Liviersi et al. [17]
Ensemble-(iCST Voting)	98.4	Liviersi et al. [17]
Deep Neural Network	96.8	Buyulyilmaz et al. [12]
Logistic Regression	97.7	Pondhu et al. [18]
kNearest Neighbor	97.7	Pondhu et al. [18]
Naive Bayes	89.4	Pondhu et al. [18]
Decision Tree	96.7	Pondhu et al. [18]
Random Forest	97.6	Pondhu et al. [18]
Support Vector Machine	97.9	Pondhu et al. [18]
Deep Neural Network	99.8	Pondhu et al. [18]
Artificial Neural Network	97.7	This Method

4. Conclusions

In this study, artificial neural networks, a traditional method, are used for classifying voice data. It was examined that emotion recognition, gender recognition and age prediction can be made by using voice data. The artificial neural network method used was able to predict gender with 97.9% success. Recall, precision and F-1 score values were found affecting the classifier performance. These values were 98%, 97.7% and 97.9%, respectively. It is thought that these values can be increased with different classifiers and hybrid classifiers. In addition, it is thought that higher classification success can be achieved by selecting effective features from among 20 features and removing those that do not contribute to classification or have negative effects. For this reason, studies on different classification methods and feature selection are planned in our future studies.

Authors Note

Abstract version of this paper was presented at 9th International Conference on Advanced Technologies (ICAT'20), 10-12 August 2020, Istanbul, Turkey with the title of "Gender Determination Using Voice Data".

References

[1] Amodei, D., et al. Deep speech 2: End-to-end speech recognition in english and mandarin. in International conference on machine learning. 2016.
[2] Chen, C., et al., A bilevel framework for joint optimization of session compensation and classification for speaker identification. Digital Signal Processing, 2019. 89: p. 104-115.
[3] Black, M., et al. Automatic classification of married couples' behavior using audio features. in Eleventh annual conference of the international speech communication association. 2010.
[4] Metze, F., et al. Comparison of four approaches to age and gender recognition for telephone applications. in 2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07. 2007. IEEE.
[5] Konig, Y., N. Morgan, and C. Chandra, GDNN: a gender-dependent neural network for continuous speech recognition. 1991: International Computer Science Institute.
[6] RAMADHAN, M.M., et al., Parameter Tuning in Random Forest Based on Grid Search Method for Gender Classification Based on Voice Frequency. DEStech Transactions on Computer Science and Engineering, 2017(cece).
[7] Jain, A. and V. Kanhangad, Gender recognition in smartphones using touchscreen gestures. Pattern Recognition Letters, 2019. 125: p. 604-611.
[8] Markitantov, M. and O. Verkholyak. Automatic Recognition of Speaker Age and Gender Based on Deep Neural Networks. in International Conference on Speech and Computer. 2019. Springer.
[9] Zvarevashe, K. and O.O. Olugbara. Gender voice recognition using random forest recursive feature elimination with gradient boosting machines. in 2018 International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD). 2018. IEEE.
[10] Gupta, P., S. Goel, and A. Purwar. A stacked technique for gender recognition through voice. in 2018 Eleventh International Conference on Contemporary Computing (IC3). 2018. IEEE.
[11] Sharma, G. and S. Mala. Framework for gender recognition using voice. in 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence). 2020. IEEE.
[12] Buyukyilmaz, M. and A.O. Cibikdiken. Voice gender recognition using deep learning. in 2016 International Conference on Modeling, Simulation and Optimization Technologies and Applications (MSOTA2016). 2016. Atlantis Press.
[13] Dataset. [08.07.2020]; Available from: https://raw.githubusercontent.com/primaryobjects/voice-gender/master/voice.csv.
[14] Araya‐Salas, M. and G. Smith‐Vidaurre, warbleR: an R package to streamline analysis of animal acoustic signals. Methods in Ecology and Evolution, 2017. 8(2): p. 184-191.
[15] Ertam, F., An effective gender recognition approach using voice data via deeper LSTM networks. Applied Acoustics, 2019. 156: p. 351-358.
[16] Akçayol, M., Bir anahtarlamalı relüktans motorun sinirsel-bulanık denetimi. 2001, Doktora Tezi, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
[17] Livieris I, Pintelas E, Pintelas P. Gender Recognition by Voice using an Improved Self-Labeled Algorithm. Mach Learn Knowl Extr 2019;1:492–503.
[18] Pondhu LN, Kummari G. Performance Analysis of Machine Learning Algorithms for Gender Classification. Proc Int Conf Inven Commun Comput Technol ICICCT 2018 2018.