레터논문 (Letter Paper)
방송공학회논문지 제22권 제2호, 2017년 3월 (JBE Vol. 22, No. 2, March 2017)
https://doi.org/10.5909/JBE.2017.22.2.253
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

기계학습 기반의 장애 음성 검출 시스템

정 준 영[a], 김 기 백[a]‡

Machine Learning based Speech Disorder Detection System

Junyoung Jung[a] and Gibak Kim[a]‡

요 약

본 논문에서는 기계학습 기반의 분류 방법을 이용하여 장애 음성을 검출하고자 한다. 음성 장애 중 마비말 장애는 뇌성마비, 파킨슨 질환, 뇌졸중 등 주로 뇌질환에 의해 발생하는 것으로 알려져 있다. 이러한 장애 음성을 검출함으로써 뇌졸중 등의 급성 뇌질환 발생에 대한 조기 처치가 가능하다. 장애 음성 검출은 입력 음성에 대한 특징벡터 추출과 기계학습을 이용한 분류과정을 통해 이루어질 수 있다. 실험을 위해서 장애 음성 DB인 TORGO 데이터를 사용하였으며, 10가지 기계학습 알고리즘과 다양한 특징벡터 스케일링 방법에 대해 장애 음성 검출 성능을 평가하였다.

Abstract

This paper deals with the implementation of speech disorder detection system based on machine learning classification. Problems with speech are a common early symptom of a stroke or other brain injuries. Therefore, detection of speech disorder may lead to correction and fast medical treatment of strokes or cerebrovascular accidents. The speech disorder system can be implemented by extracting features from the input speech and classifying the features using machine learning algorithms. Ten machine learning algorithms with various scaling methods were used to discriminate speech disorder from normal speech. The detection system was evaluated by the TORGO database which contains dysarthric speech collected from speakers with either cerebral palsy or amyotrophic lateral sclerosis.

Keyword : Speech disorder, Machine learning

a) 숭실대학교 전기공학부(School of Electrical Engineering, Soongsil Univ.)
‡ Corresponding Author : 김기백(Gibak Kim)
E-mail: imkgb27@ssu.ac.kr
Tel: +82-828-7266
ORCID: http://orcid.org/0000-0001-5114-4117

※ 이 논문은 한국 산업통상자원부의 로봇산업융합핵심기술사업 프로그램 No.10048474, '고령화 세대에게 개인별 특화된 복지 서비스를 제공하기 위한 빅데이터 기반의 서비스 로봇개발'의 지원으로 수행되었음.
· Manuscript received January 5, 2017; Revised March 2, 2017; Accepted March 2, 2017.

Ⅰ. 서 론

뇌졸중 등 급성 뇌질환이 발생하는 경우, 혈관이 막힘에 따라 신체마비, 언어장애 등의 초기 증상이 발생할 수 있다. 이러한 증상은 일시적으로 나타났다가 다시 회복되는 경우가 있는데, 이 때문에 심각성을 인지하지 못하고 적절한 치료를 받을 시기를 놓쳐서 영구적인 신체장애를 겪게 될 수도 있다. 본 연구에서는 고령화시대 실버용 서비스 로봇 개

발의 일부로서 장애 음성을 검출하여 뇌졸중 등의 급성 뇌질환 가능성에 대처하는 시스템을 개발하고자 한다.

본 논문에서는 기계학습 기반의 분류기를 사용하여 장애음성과 정상 음성을 구분하는 방법을 사용하고자 한다. 기계학습은 영상, 음성처리에 널리 적용되어왔다[1-3]. 입력 음성으로부터 장애음성 검출에 적합한 특징들을 추출하고 추출한 특징들에 대해 스케일링을 적용한 후, 기계학습을 이용하여 장애 음성과 정상 음성을 구분한다.

Ⅱ. 특징 추출 및 기계학습

1. 장애 음성 검출을 위한 특징 추출

음성인식이나 화자인식을 위해서는 일반적으로 MFCC와 같은 입력 음성의 스펙트럼 특성을 나타내는 특징벡터를 사용하나 장애 음성 검출을 위해서는 이와는 다른 특징벡터들을 이용한다. 발성 장애가 있는 경우, 일정한 음높이를 유지하는데 어려움을 겪게 되어 피치주파수의 변화가 심하게 나타나는 편이다. 모음을 길게 발음할 경우 장애가 있으면 일정한 크기의 발성에 어려움 때문에 음성 신호의 진폭 변화가 상대적으로 심하게 나타난다. 구음 장애의 경우 성문 닫힘이 주기적 또는 정확하게 이루어지지 않아 유성음을 발성할 경우에도 무성음과 유사한 발성이 나타난다거나 하모닉스가 제대로 형성되지 않을 수도 있으며, 잡음과 유사한 특성의 신호가 포함되기도 한다.

이러한 장애 음성의 특성을 이용하기 위해 피치의 변화율을 계산하는 Jitter, 음성 신호의 최대 진폭 변화율을 측정하는 Shimmer, 유성음 발성에 따른 고조파 성분과 잡음의 비율을 계산하여 평균 및 표준편차를 계산하는 HNR/NHR (Harmonics to Noise Ratio/Noise to Harmonics Ratio) 등을 기반으로 하는 전통적인 특징을 사용한다[4]. 이와 더불어 피치의 안정도 측정을 위한 엔트로피 측정 (PPE), 비선형적인 성대 바이브레이션과 비슷한 저주파 바이브레이션을 표현하는 non-stationary 신호에서의 자기상관도를 측정 (DFA), 바이브레이션의 주기 변이를 측정하여 성대의 지속적 진동 유지 능력을 측정 (RPDE) 하여 특징벡터로 사용한다[5]. 본 논문에서는 전통적인 특징들과 최근 제안된 특징들을 사용하였으며, 그 특징들은 다음과 같다.

1) Jitter: $\frac{1}{N}\sum_{i=1}^{N-1}|F_{0,i}-F_{0,i+1}|$
2) Jitter를 F_0의 평균값으로 나눈 값
3) Jitter의 K싸이클에 대한 perturbation quotient (K=5)
4) F_0 편차에 대한 평균값
5) Shimmer: $\frac{1}{N}\sum_{i=1}^{N-1}|A_{0,i}-A_{0,i+1}|$
6) Shimmer를 A_0의 평균값으로 나눈 값
7) Shimmer의 K싸이클에 대한 perturbation quotient (K=3)
8) Shimmer의 K싸이클에 대한 perturbation quotient (K=5)
9) A_0 편차에 대한 평균값
10) HNR: $10\log\left[\frac{R_{xx}(l_{\max})}{1-R_{xx}(l_{\max})}\right]$
$R_{xx}(l_{\max})$: 최대 자기 상관도
11) HNR의 표준편차
12) NHR: $10\log\left[\frac{1-R_{xx}(l_{\max})}{R_{xx}(l_{\max})}\right]$
13) NHR의 표준편차
14) PPE (Pitch Period Entropy)
15) DFA (Detrended Fluctuation Analysis)
16) RPDE (Recurrence Period Density Entropy)

1) ~ 4)는 Jitter 기반의 특징들이고 5) ~ 9)은 Shimmer 기반의 특징들이다. 아래 식에서 $F_{0,i}$는 i번째 프레임의 피치 주파수이고, $A_{0,i}$는 i번째 프레임의 진폭 (프레임 내 샘플 절대값의 합)을 나타낸다.

2. 특징벡터 스케일링

특징벡터는 다양한 알고리즘을 통해 추출되었고, 따라서 각 특징들은 다양한 범위를 갖게 된다. (예를 들어, Jitter는 0.16~130, Shimmer는 0.01~0.15) 상당수의 기계학습 알고리즘에서는 이런 특징벡터 내의 특징들에 대한 스케일링과정을 전처리과정으로 거쳐야 하는 경우가 많다. 특히 두 특징벡터 간의 거리를 유클리디안 거리로 계산하는 기계학습의 경우는 특징벡터 중 특정 특징이 분포하는 범위가 넓다면, 그 특징에 의해서 전체 유클리디안 거리가 결정될 수 있으므로, 특징벡터 스케일링과정이 성능에 많은 영향을

미치게 된다. 또한 경사 하강법(Gradient descent)을 사용하여 학습하는 기계학습의 경우, 특징벡터 내의 각 특징들의 스케일링 여부가 파라미터 학습의 수렴과정에 영향을 미치게 된다. 경사 하강법으로 파라미터를 학습할 때는 특징이 파라미터를 업데이트하는 식에 포함되므로 만일, 특징 간 범위가 다르다면 각 특징마다 파라미터의 업데이트 속도가 달라지게 된다.

어떤 특징벡터 $\boldsymbol{x}=\{x_1,x_2,\cdots,x_i,\cdots,x_N\}$에 대해, 기계학습의 전처리로 사용될 수 있는 스케일링 기법들은 다음과 같다[6].

1) Standardization: 각 특징의 분포가 평균은 0, 표준편차는 1이 되도록 정규화하는 과정으로서 학습데이터의 특징 평균과 표준편차를 이용하여 아래와 같이 변환한다.

$$\overline{x_i}=\frac{x_i-\mu_{x_i}}{\sigma_{x_i}} \tag{1}$$

2) Min-max scaling: 특징 분포의 범위를 제한하는 것으로서 보통 최소값은 0, 최대값은 1로 제한하도록 변환한다.

$$\overline{x_i}=\frac{x_i-x_{i,\min}}{x_{i,\max}-x_{i,\min}} \tag{2}$$

3) Abs-max scaling: 특징의 절대값 분포를 제한하는 것으로서 보통 절대값의 최대값이 1이 되도록 한다. 즉, 특징의 최대절대값이 1이 넘지 않도록 정규화한다. 앞의 두 가지 스케일링 기법과 달리 특징으로부터 상수를 차감하지 않으므로 스케일링 후에도 특징의 부호가 바뀌지 않는다.

$$\overline{x_i}=\frac{x_i}{|x_i|_{\max}} \tag{3}$$

4) Normalization: 특징벡터의 norm이 1이 되도록 정규화한다. 이러한 정규화는 두 특징벡터 간의 거리 측정을 위해 내적을 사용하는 경우는 L2 norm을, 히스토그램을 나타내는 특징에서 맨하탄거리 로 거리를 측정할 때는 L1 norm을 사용할 수 있다.

$$\overline{x_i}=\frac{x_i}{\|\boldsymbol{x}\|} \tag{4}$$

위에서 열거한 스케일링 기법 중 어떤 것을 사용해야 하는가는 기계학습 방법에 의해서만 결정되지는 않는다. 기계학습의 종류와 함께 특징의 속성에 따라 스케일링 기법들의 효과가 달라지게 되므로, 일반적으로 스케일링 기법의 선택은 실험적으로 결정된다.

3. 기계학습 방법

KNN (K-Nearest Neighbors): 수학적인 모델을 사용하지 않는 분류방법으로서, 테스트 데이터의 특징벡터와 가장 가까운 학습 데이터의 특징벡터 K개를 선택한 후, 가장 많은 수의 데이터가 속한 클래스로 할당한다.

SVM (Support Vector Machine): Vapnik에 의해 제안된 방법으로서 클래스 간의 마진을 최대로 하는 선형 함수를 찾아 두 클래스를 분류한다. 커널법을 도입하여 비선형 문제를 해결할 수 있다.

Decision Tree: 질문에 대한 답변에 따라 다른 가지로 분기하여 최종 노드에서 클래스를 분류하는 방법이다. 통계적 방법을 이용하는 CART (Classification and Regression Trees) 학습을 적용하였다.

Random Forest: decision tree의 확장 개념으로서 전체 학습데이터에서 임의로 데이터 셋을 추출하여 여러 개의 tree를 생성한 후, bootstrap aggregating (bagging) 방법을 이용하여 분류 결과를 결정한다.

Adaboost: 단순한 분류기들을 결합하여 높은 성능의 분류기를 만드는 방법으로서 분류 성능에 따라 가중치를 학습하게 된다.

Naive Bayes: 특징들이 서로 독립이라는 가정하에 각 특징에 대해 Bayes 분류기를 만들어 결합한 분류기이다.

Neural Networks: 인공 신경망 분류기로서 본 논문에서는 입출력층 외에 은닉층을 갖는 구조를 적용하는 Multi-layer Perceptron을 사용한다.

LDA (Linear Discriminant Analysis): 두 개의 클래스가 각각 가우시안분포를 따르고 분산행렬이 대각행렬이라고 가정하여 선형 경계면을 결정한다.

QDA (Qudratic Discriminant Analsysis): LDA와 달리 분산행렬이 대각행렬이라는 가정을 적용하지 않는다.

III. 실험 결과

성능 검증을 위한 실험을 위해 TORGO 데이터베이스를

사용하였다. TORGO 데이터베이스는 캐나다 토론토대학과 Holland-Bloorview Kids Rehabilitation 병원에 의해 수집된 데이터로서 뇌성마비, 근위축성 측삭경화증 환자로부터 수집되었다[7]. 단어, 문장, 음절의 반복, sustained vowel 등의 데이터를 포함하고 있는데, 본 연구에서는 sustained vowel을 이용하여 실험하였다. sustained vowel은 모음을 지속적으로 발성한 데이터로서 음성 장애가 있을 경우 진폭 및 피치가 일정하지 않을 가능성이 높다. TORGO 데이터베이스에서는 /a/ 발음을 지속적으로 발성하도록 하였으며, 8명의 환자(남자 5명, 여자 3명)와 7명의 정상발음 (남자 4명, 여자 3명)을 포함하고 있다. 본 실험에서 사용한 발음은 총 96개의 발성이다.

II-3에서 나타낸 10가지 기계학습 방법을 이용하여 실험한 결과를 <표 1>에 나타내었다. <표 1>에서 표시한 기계학습 방법에서 SVM-L, SVM-R은 각각 linear, RBF 커널을 사용한 SVM을 나타내고, D.T.은 decision tree, R.F.은 random forest, N.B.은 naive Bayes, N.N.은 neural networks를 나타낸다. 특징벡터는 II-1에서 열거한 16차원이고, KNN에서 k=3, N.N.에서 은닉층은 50으로 하였다. P1~P5는 특징벡터 정규화 방법을 나타내는 것으로서 P1은 학습, 테스트 데이터들을 각각 standardization변환한 것이다. 그러나 테스트 과정에서는 테스트 데이터를 전체를 모아서 테스트하는 것이 아니라 각각의 발성에 대해 테스트하는 것이므로 테스트 데이터의 전체 분포를 알 수 없다. P2는 테스트 데이터 standardization을 학습 데이터 변환식을 이용한 것이다. P3~P5는 각각 min-max scaling, abs-max scaling, normalization을 적용한 결과이다.

표 1. 검출정확도 (%)
Table 1. Accuracy (%)

	None	P1	P2	P3	P4	P5
KNN	86.1	86.1	86.6	**88.7**	85.8	85.3
SVM-L	85.3	86.6	**87.9**	57.4	56.3	56.1
SVM-R	58.2	58.9	65.3	**86.8**	**86.8**	83.4
D.T.	83.2	79.2	**84.7**	83.9	82.1	82.9
R.F.	78.2	69.2	80.5	79.7	80.3	**84.7**
Adaboost	**83.7**	80.3	83.4	**83.7**	83.4	81.6
N.B.	82.1	80.5	82.1	82.1	82.1	**84.2**
N.N.	82.4	85.3	**88.9**	88.2	86.6	58.9
LDA	**89.5**	87.4	**89.5**	**89.5**	**89.5**	88.2
QDA	**89.5**	69.7	**89.5**	**89.5**	**89.5**	85.6
평균치	81.8	78.3	83.8	83.0	82.2	79.0

본 실험에서는 LDA, QDA결과가 89.5%로 가장 높게 나타났다. LDA, QDA는 특징에 대한 변환이 내재되어 있어 정규화에 대한 효과가 없는 것으로 확인되었다. 나머지 방법들에서는 알고리즘에 따라 차이가 있으나 정규화를 적용한 것이 정규화를 적용하기 전보다 향상된 결과를 보임을 확인하였다. 각 알고리즘에서 가장 좋은 결과를 보인 값을 진하게 표시하였다.

IV. 결 론

장애 음성 판별은 공학적으로 분석된 특징들을 이용하여 음성학 전문가가 개별적으로 판단하는 시스템에 의존해왔으며, 국내에서는 자동판별에 관한 연구가 이루어지지 않았다. 본 논문에서는 다양한 특징벡터 추출과 기계학습 알고리즘을 이용하여 자동으로 장애 음성을 검출하는 시스템을 구현하였다. 16개의 특징들로 이루어진 특징벡터를 이용하여 기계학습 분류기를 학습하였다. 10개의 기계학습 알고리즘을 다양한 특징벡터 스케일링 방법에 대해 테스트하여 최고 89.5%의 검출정확도를 얻었다.

참 고 문 헌 (References)

[1] H. Yun et, al., “On-Line Audio Genre Classification using Spectrogram and Deep Neural Network,” Journal of Broadcast Engineering, Vol.21, No.6, pp. 977-984, November 2016.

[2] H. Kim and J. Park, “Vocal Separation Using Selective Frequency Subtraction Considering with Energies and Phases,” Journal of Broadcast Engineering, Vol.20, No.3, pp. 408-413, May 2015.

[3] J. Gil and M. Kim, “Subimage Detection of Window Image Using AdaBoost,” Journal of Broadcast Engineering, Vol.19, No.5, pp. 578-589, September 2014.

[4] R. D. Kent, G. Weismer, J. F. Kent, H. K. Vorperian, and J. R. Duffy, “Acoustic Studies of Dysarthric Speech: Methods, Progress, and Potential,” Journal of Communication Disorders, vol. 32, no. 3, pp. 141-186, 1999.

[5] A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, “Accurate Telemonitoring of Parkinson’s Disease Progression by Noninvasive Speech Tests,” IEEE Trans. Biomedical Engineering, vol. 57, no.4, pp. 884 – 893, April 2010.

[6] http://scikit-learn.org/stable/modules/preprocessing.html.

[7] F. Rudzicz, A.K. Namasivayam, and T. Wolff, “The TORGO database of acoustic and articulatory speech from speakers with dysarthria,“ Language Resources and Evaluation, vol. 46, no. 4, pp. 523-541, 2012.