
음성 단어를 이용한 구간검출에 의한 패턴인식

최재승*

*신라대학교

Pattern Recognition by Section Detection Using Speech Word

Jae-Seung Choi*

*Silla University

E-mail : jschoi@silla.ac.kr

요 약

본 논문에서는 화자 식별에서 음성신호의 애매한 점을 보완할 수 있는 신경회로망의 오차역전과 학습 알고리즘과 모음구간 검출에 기초하여 입력되는 음성의 화자 패턴을 구분하는 일본어 단어 패턴인식 알고리즘을 제안한다. 제안하는 알고리즘에서는 일본어 데이터베이스로부터의 단어를 사용하여 음성의 특징벡터를 추출하여 분석하고 이러한 음성의 특징벡터의 차이를 이용하여 일본어 화자에 대한 패턴인식 실험을 수행하였다.

키워드

화자 패턴, 단어 식별, 모음구간, 검출 방법, 식별 알고리즘

1. 서 론

근래 신경회로망을 패턴인식에 응용하려는 연구가 활발히 진행되어 왔는데, 오차역전과 학습 알고리즘(Error Back Propagation Learning Algorithm, EBPL)이 등장함으로써 여러 분야에서 신경회로망을 응용하였다[1, 2, 3].

최근에도 여러 연구자들에 의하여 음성인식기술의 연구개발이 활발하게 수행되고 있다. 이러한 음성인식 중에서 성별 화자인식에 대한 연구개발도 보고되고 있다[3]. 남성 및 여성 화자를 대상으로 성별 인식실험을 하는 성별 화자인식은 은닉 마르코프 모델(Hidden Markov Model, HMM), 신경회로망(Neural Network, NN)을 이용한 연구가 보고되고 있다[1, 5].

본 논문에서는 이러한 신경회로망의 오차역전과 학습 알고리즘과 모음구간 검출에 기초하여 입력되는 음성의 화자 패턴을 구분하는 일본어 단어

인식 알고리즘을 제안한다. 오차역전과 학습 알고리즘에 의한 다층 퍼셉트론을 사용하여 기초적인 인식에 관한 연구를 진행하였다. 오차역전과 학습 알고리즘은 다층 퍼셉트론에 있어서 보편화된 학습 알고리즘으로 음성에 관련한 패턴 인식 분야에 있어서 상당히 유용한 학습 알고리즘이라는 것이 여러 연구에 의하여 증명되고 있다.

II. 본 론

선형예측 분석법은 음성 분석 및 합성에 사용되는 등 음성신호 처리에서 자주 사용되는 방법이다. 선형 예측 분석을 다음과 같이 기술한다.

선형예측오차 e_n 의 2승 평균값을 최소로 하도록 선형예측계수를 결정한다. 이때, 시간신호 n 에

서의 음성신호의 표본 예측값 \hat{x}_n 은 식 (1)처럼 이전 p 개의 과거의 표본값 $x_{n-1}, x_{n-2}, \dots, x_{n-p}$ 로부터 예측한다.

$$\hat{x}_n = \sum_{k=1}^p a_k x_{n-k} + e_n \dots\dots\dots (1)$$

따라서 선형예측오차를 e_n 이라고 하면 식 (2)와 같이 구할 수 있다.

$$e_n = x_n - \sum_{k=1}^p a_k x_{n-k} \dots\dots (2)$$

여기에서, $a_k(k=1, \dots, p)$ 를 LPC 계수라고 하며, 본 논문에서는 이 LPC 계수를 특징 파라미터로써 사용한다.

본 논문의 실험에서 모음구간만을 검출한 프레임에 대해서, 위에서 기술한 LPC 캡스트럼 계수를 추출하여 신경회로망에 입력한다. 본 알고리즘은 입력된 LPC 계수를 네트워크에 의하여 학습된 후에 화자패턴을 식별하게 된다. 본 논문에서 제안하는 단어 성별식별 알고리즘에서 신경회로망의 입력으로는 LPC 캡스트럼 계수는 12차의 LPC 캡스트럼 데이터를 사용하였다.

본 논문에서는 LPC 캡스트럼 계수를 신경회로망에 입력하였을 경우에 각 일본어 화자에 대한 평균 식별률이 96% 이상인 것을 실험을 통하여 알 수 있었다.

III. 결 론

본 논문에서는 일본어 단어에 대한 화자패턴 식별을 위하여 모음구간 검출 방법을 사용한 일본어 단어 패턴식별 알고리즘을 제안하였다. 제안한 알고리즘은 일본어 화자 패턴식별을 위하여 음성신호의 애매성을 보완할 수 있는 신경회로망을 이용하여 정확한 모음구간 검출을 통하여 패턴인식을 양호하게 수행하였다.

본 논문에서 제안하는 알고리즘에 의한 인식결과가 효과적으로 개선되는 것을 실험으로 확인할 수 있었으며, 실험 결과 모든 원음에 대한 인식률은 96% 이상 나타났다.

참고문헌

[1] T. T. Le, J. S. Mason and T. Kitamura, "Characteristics of multi-layer perceptron models in enhancing degraded speech", Proc. ICSLP-94, pp. 1611-1614, 1994.
 [2] D. E. Rumelhart, G. E. Hinton, and R. J. Williams,

"Learning representations by back-propagation errors", *Nature*, Vol. 323, pp. 533-536, 1986.
 [3] J. S. Choi, "Japanese Word Gender Classification Using the Detection of the Vowel Sections", *Korean Institute of Information Technology*, Vol. 13, No. 8, pp. 47-53, 2015.
 [4] J. G. van Velden and G.F. Smoorenburg, "Vowel recognition in noise for male, female and child voices", *International Conference on Acoustics, Speech, and Signal Processing*, pp.989-992, Vol.2, April 1991.
 [5] A. A. M. Abushariah, T.S. Gunawan, O.O. Khalifa and M.A.M. Abushariah, "English digits speech recognition system based on Hidden Markov Models", *International Conference on Computer and Communication Engineering*, pp. 1-5. May 2010.