

음소 기반의 유사율 알고리즘을 이용한

Home Network 환경에서의 음성 인식

이창섭 유재봉 박준석 양수호 김유섭 박찬영
한림대학교 컴퓨터공학과

e-mail: ds2shg, blueblood, joon, sooho79, yskim01, cypark@hallym.ac.kr

Voice Recognition using a Phoneme based Similarity Algorithm in Home Networks

Chang sub Lee, Jae bong Yu, Joon Seok Park,

Soo Ho Yang, Yu Seop Kim, Chan Young Park
Dept. of Computer Engineering, Hallym University

요 약

네트워크상에서 전달되는 음성데이터는 전달되는 과정에서 잡음 등의 외부 요인으로 인하여 데이터에 손실이 생기는 문제가 발생한다. 이렇게 전달된 음성데이터가 음성 인식을 통과하면 바로 음성 인식을 통과했을 때 보다 인식률이 낮아진다. 본 연구에서는 홈 네트워크를 제어하는데 있어서 음성 인식을 향상시키기 위해서 음성 데이터를 입력받아, 이를 음소단위 기반의 유사율 알고리즘을 적용시켜 이미 구축된 홈 네트워크 용어 관련 사전에 등록된 단어와의 유사성을 검토하여 추출된 결과로 홈 네트워크를 제어하는 방안을 제안한다. 음소단위 기반의 유사율 알고리즘과 다중발화를 이용했을 때 Threshold 값이 85% 일 경우 사전에 구축된 단어와 매칭된 인식률은 100%였으며, 사전에 없는 단어의 오인식률은 2%로 감소되었다.

1. 서론

홈 네트워크는 가정 내의 PC를 비롯한 정보가전 기기를 유선 또는 무선의 네트워크로 연결하여 주변 기기 공유 및 상호제어를 가능하게 하며, 인터넷이나 휴대용 정보 단말기를 이용한 외부 네트워크와의 연동으로 언제 어디서나 자유롭게 가정의 디지털 가전기기를 원격 제어할 수 있는 시스템을 말한다.[1]

지금까지 홈 네트워크 시스템에 제어 명령을 보낼 때는 휴대폰이나 인터넷에서 화면을 보고 여러 절차를 통해 버튼을 눌러 조작하도록 되어 있어서 사용이 불편하고 번거로웠다. 현재의 홈 네트워크 기술은 음성통화를 사용하므로 유무선 전화를 사용할 수 있는 곳이면 어디서나 이용할 수가 있고, 사용법이 간편하여 누구나 손쉽게 이용할 수 있다는 장점이 있다. 그러나 이렇게 음성통화를 이용해서 홈 네트워크 시스템을 제어하는 경우, 네트워크에서 VoIP 기술을 사용하여 음성데이터가 전달될 때 네트워크 상에서 발생하는 왜곡·손실로 인해 인식률이 현저

하게 감소된다는 단점이 있다.[2] 이러한 장벽은 음성인식 분야의 발전과 정보통신의 발달과 더불어 해결해야 할 당면 문제이다.

위의 문제를 해결하기 위하여 [3]에서는, 다중 발화된 음성데이터가 네트워크를 통해 음성 인식을 거쳐 얻어진 결과(text data)와 홈 네트워크 사전의 단어들의 유사성을 음절 단위 유사율 알고리즘을 이용하여 비교했을 때 음성 인식률이 향상됨을 볼 수 있었다. 반면에 단어의 수를 증가하고 사전에 없는 단어들을 실험한 결과, 음성 인식률이 급격하게 저하되는 현상을 볼 수 있었다. 이에 본 논문은 음성 인식을 거쳐 나온 결과와 이미 구축된 홈 네트워크 관련 용어 사전에 있는 단어와의 유사율을 비교하는데 있어서, 가전제품에 쓰일 단어를 음절단위 기반 유사율 알고리즘에서 음소단위 기반의 유사율 알고리즘으로 바꾸어 인식률을 좀 더 향상시키므로 이동 단말 상에서 사용자가 입력한 음성 정보를 정확하게 인식할 수 있도록 하여 홈 네트워크 시스템의 제어를 수월하게 하고자 한다.

2. 홈 네트워크 환경에서의 음성인식 시스템

기계에 의한 음성인식의 결과를 실생활의 여러 분야에 응용하는 사례가 나날이 늘어가고 있다. 예를 들어, 가전제품을 동작시킬 때 사용자의 음성명령에 따라 구동시키는 사례를 생각해 볼 수 있다. 하지만, 음성이 발화되는 환경은 보통 매우 열악하여 (자동차 소음, 군중의 소음, 또는 여러 음원이 섞이거나 반사되는 등), 기계가 정확히 내용을 인식하여 그 결과를 출력한다는 것은 대단히 어려운 일이다.

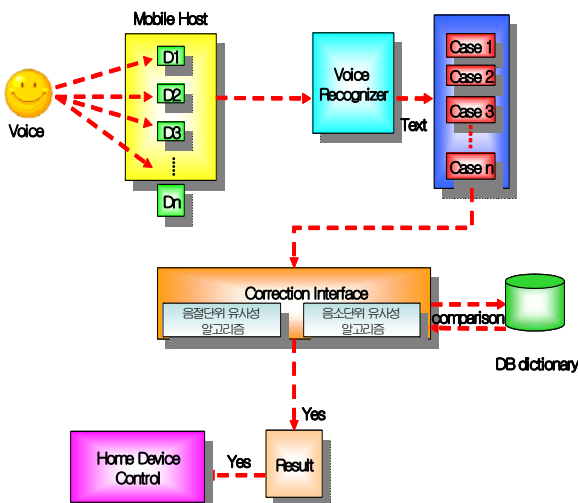
따라서, 지금까지 구현된 음성인식기의 출력은 어느 정도 오류를 포함하고 있으며, 이를 입력으로 하여 위에 열거한 응용에 사용하기 위해서는 오류를 검출하고 원문에 맞게 이를 수정하는 작업이 필요불가결하다.[4]

현재 음성인식 시스템분야에서는 기존의 음성통화에서 사용되던 off-line 전화통신이 on-line 상으로 빠르게 대체되어 가고 있다. 그러므로 인터넷망을 통하여 음성 데이터들을 효율적으로 전송하는 기술들이 연구되어 왔으며 VoIP(Voice over Internet Protocol)기술은 이러한 시도들에 대한 구체적인 방법론이라고 할 수 있다.[5]

네트워크상에서의 음성인식을 이용한 홈 네트워크 시스템에서는 유선의 네트워크 환경을 이용하므로 인터넷의 과부하로 인해 왜곡이나 손실로 인한 문제가 생길 수 있다.

본 연구에서 제안하는 시스템은 사람들이 많이 사용하는 인터넷상에서 음성데이터를 입력하여 음성인식기를 거쳐서 얻은 결과(text data)를 홈 네트워크 용어 사전에 있는 단어와 비교하여 가장 적절한 명령어로 홈 네트워크 시스템에 연결된 각각의 장치를 제어하고자 한다.

본 연구는 네트워크상에서 음성데이터를 PC to



[그림 1] 유사율 알고리즘을 적용한 음성인식시스템 PC 상태에서 전송하여 음성 인식기를 통해 나온 결과들로 음소단위/음절단위 유사성 알고리즘을 적용

시켜 음성데이터의 인식률을 향상시켜 홈 네트워크 시스템을 제어하는 방법을 제안하고자 한다.

본 연구에서 제안하고자 하는 시스템의 과정은 다음과 같다.[3]

- (1) 네트워크상에서 음성데이터를 다중 발화한다.
- (2) 단말기에서 입력받은 음성데이터는 home network system의 음성 인식기를 거쳐 텍스트 데이터로 추출 된다.
- (3) 추출한 텍스트 데이터를 홈 네트워크 용어 사전과 비교하여 해당하는 단어를 찾는다. 이때 음소단위별 유사율 알고리즘과 음절단위별 유사율 알고리즘을 적용한다.
- (4) 해당하는 단어를 찾지 못하는 경우에는 비교하지 않은 다른 case의 데이터를 가지고 다시 비교한다.
- (5) home device를 제어한다.

3. 음절단위 기반의 유사율 알고리즘

[그림 1]의 Correction Interface 부분에 음소단위 기반의 유사성 알고리즘과 음절단위 기반의 알고리즘을 적용시켰다. [3]에 쓰인 음절단위 기반의 유사율 알고리즘에서는 유사율의 threshold 값을 50%로 설정하였다. 이와 같이 Threshold 값을 50%로 설정한 이유는 발화된 단어와 홈 네트워크 사전에 구축된 단어는 두 음절일 경우 한 음절이상 매칭되어야 하며, 세 음절일 경우 두 음절이상 매칭되어야 발화된 단어는 홈 네트워크 사전에 있는 단어로 인식되어 처리되기 때문이다.

음절단위 유사율은 다음과 같이 구할 수 있다.[6]

$$\text{유사율}(\%) = \frac{\text{일치음절수}}{\text{추측된단어의전체음절수}} \times 100$$

예를 들어, 발화된 단어 “보일러”와 홈 네트워크 사전에 등록되어있는 단어 “보일러”와의 유사율은 다음과 같이 계산된다.

$$\text{유사율}(\%) = \frac{2}{3} \times 100 = 66.66$$

음소단위 기반의 유사율 알고리즘에서는 음절을 음소로 분해하여 비교하기 때문에 유사율의 threshold 값을 좀 더 유연하게 변경할 수 있다. 그러므로 적당한 threshold 값을 설정하여 유사율이 threshold값 이상이면 홈 네트워크 사전에 있는 단어로 인식하여 처리하는 알고리즘을 적용한다. 유사율은 다음과 같이 구할 수 있다.

$$\text{유사율}(\%) = \text{weight} \times \frac{\text{일치음소수}}{\text{발화된단어의음소수}} \times 100$$

이 때, weight는 발화된 단어와 홈 네트워크 사전에 등록된 단어의 음소의 길이를 비율로 나타낸 것이

다. wieght는 다음과 같이 구할 수 있다.

$$\begin{aligned} &\text{if } (a \leq b) \\ &\quad \text{weight} = \frac{-a}{b} \\ &\text{else} \\ &\quad \text{weight} = \frac{b}{a} \end{aligned}$$

단, a는 발화된 단어의 음소수를 b는 홈 네트워크 사전에 등록된 단어의 음소수를 나타낸다. 예를 들어, 발화된 단어 “보일러”와 홈 네트워크 사전에 등록되어있는 단어 “보일러”를 음소로 분해하면 다음과 같다.

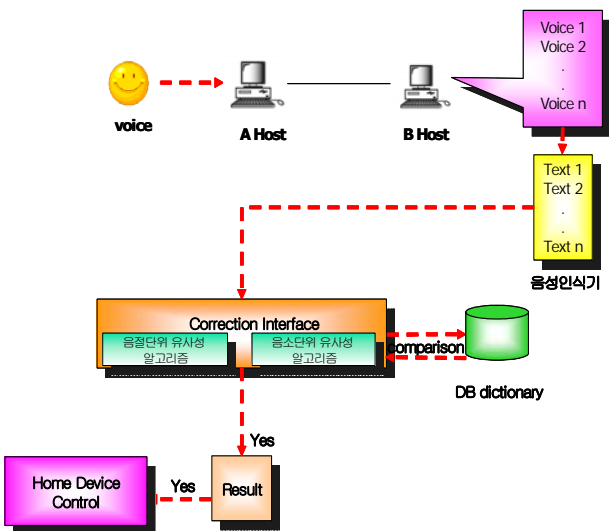
$$\begin{aligned} \text{보일러} &= \text{ㅂ} + \text{ㅇ} + \text{ㅇ} + \text{ㅇ} + \text{ㅇ} + \text{ㅇ} + \text{ㅇ} + \text{ㅇ} + \text{ㅇ} + \text{ㅇ} \\ \text{보일러} &= \text{ㅂ} + \text{ㅇ} + \text{ㅇ} + \text{ㅇ} + \text{ㅇ} + \text{ㅇ} + \text{ㅇ} + \text{ㅇ} + \text{ㅇ} + \text{ㅇ} \end{aligned}$$

모든 음절은 초성, 중성 그리고 종성으로 이루어져 있다. 이 때, ㉠은 중성이 없다는 것을 의미하며, 한 음소로 취급한다. 두 단어의 유사율은 다음과 같이 계산된다.

$$\text{유사율}(\%) = \frac{9}{9} \times \frac{8}{9} \times 100 = 88.88$$

4. 실 험

본 연구에서는 인식률을 향상시키기 위한 하나의 방법으로 이동 단말에서 사전에 있는 단어와 사전에 없는 단어를 다중발화로 입력된 음성데이터들을 음성 인식기로 입력하여 나온 결과(text data)에 각각 음절단위 기반 유사율 알고리즘을 사용하였을때와 음소단위 기반 유사율 알고리즘을 사용하였을때의 인식률을 비교 실험 하였다. 오인식된 단어를 올바른 단어로 수정하여 홈 네트워크 시스템에서 각각의 장치를 제어할 수 있도록 하였다.



[그림 2] 네트워크상에서 음성데이터 전송 실험

4.1 유사율 알고리즘 적용하지 않은 경우의 인식률

본 연구에서는 네트워크상에서 음성데이터를 전송하여 홈 네트워크를 제어할 때 음성데이터가 네트워크를 거쳐 음성인식기로 전달될 때 유사성 알고리즘을 고려하지 않았을 경우의 인식률을 평가하였다.

[그림 2]에서 보는 것과 같이 PC to PC 환경에서, 즉 A Host에서 음성을 전달하여 B Host에 전달된 음성데이터가 음성 인식기를 통해 Text가 나오는 결과를 확인하였다.

[3]에서는 10개의 단어를 추출하여 실험을 하였으나, 본 실험에 쓰인 단어는 홈 네트워크를 제어하는데 필요로 할 수 있는 단어를 100개를 추출하여 음성 인식기에서 해당되는 단어를 학습을 시킨 후 전송된 데이터의 인식률이 얼마인지 평가하였다.

[표 1] 실험에 쓰인 100개 단어 중 샘플항목

실험에 쓰인 단어	
제품단어	제품기능별 제어 단어
보일러	온수, 예약, 정지
냉장고	탈취, 온도조절
선풍기	약풍, 강풍
전자레인지	해동
텔레비전	채널변경
에어컨	강, 약, 운전, 정지
비디오	녹화, 예약
세탁기	탈수, 행굼, 빨래

본 실험에서는 남녀 각각 3명씩 각 단어 당 100회를 반복하여 실험한 결과, 단어의 유사성 알고리즘을 고려하지 않고, 단지 음성데이터를 네트워크에서 전송을 했을 때의 인식률은 평균 80.4%가 인식이 되는 것으로 나타났다. 네트워크 환경에서 음성 정보가 전달될 경우 데이터의 손상이나 왜곡 및 음성 정보가 음성 인식기에 입력될 때 외부의 잡음과 같은 변인들이 인식률을 떨어뜨리는 요인이 된다.

4.2 음절단위 유사율 알고리즘 사용 시 다중 발화 인식률

[3]에서 제안하는 다중발화시스템에서는 10개의 단어들로 음성데이터를 음성 인식기를 통해 추출하여 그 결과로 인식률을 평가하였으나, 본 실험에서는 홈 네트워크에서 사용될 수 있는 단어를 100개를 추출하고 또한 홈 네트워크 사전에 없는 단어 100개를 추가하여 실험을 하였다.

본 실험에서 사전에 없는 단어를 추가 실험한 이유는 화자가 사전에 없는 단어를 발화하였을 경우 홈 네트워크 제어 시스템에서는 잘못된 입력이라고 판단해야 하기 때문이다. 따라서 사전에 없는 단어의 오인식률이 감소할수록 홈 네트워크 제어 시스템의 신뢰성을 증가시킬 수 있다. 본 논문에서는 사전에 없는 단어를 발화했을 경우 사전에 있는 단어로 잘못 인식된 경우는 오인식률이라 한다.

아래 [표 1]은 [3]의 연구에서의 음절단위 유사성 알고리즘을 이용하여 사전에 등록되어 있는 단어와

사전에 등록되지 않은 단어로 다중 발화될 때 홈 네트워크 사전과 비교한 인식률을 나타내고 있다.

[표 1] 음절단위 유사성 알고리즘 적용 시 다중 발화 인식률

단위 : %

Threshold	사전에 있는 단어의 인식률	사전에 없는 단어의 오인식률
50	100	24

실험결과 사전에 있는 단어를 입력하였을 때는 100% 인식되는 결과를 볼 수 있었으나, 사전에 없는 단어의 경우 24% 가 사전에 있는 단어와 매칭이 되는 결과를 얻을 수 있었다. 3절에서 언급한 것과 같이 두 음절의 단어인 경우 어느 한 음절만 매칭이 되도 인식이 되고, 세 음절의 경우에는 두 음절만 매칭이 되더라도 인식이 된다. 이러한 현상이 일어나는 것은 음절단위 유사율 알고리즘을 적용하는데 있어서 음성 인식기에서 인식한 결과와 홈 네트워크 사전의 단어의 유사율인 Threshold를 50%로 고정하여 사용했기 때문이다.

4.3 음소단위 유사율 알고리즘 사용 시 단일/다중 발화 인식률

본 논문에서는 이러한 Threshold를 조절하기 위해서 음절단위의 유사율 알고리즘을 대신하여 음소단위 유사율 알고리즘을 사용하여 단일 발화와 다중 발화일 경우에 인식률을 비교 실험하여 아래 [그림 3]과 같은 결과를 얻을 수 있었다.

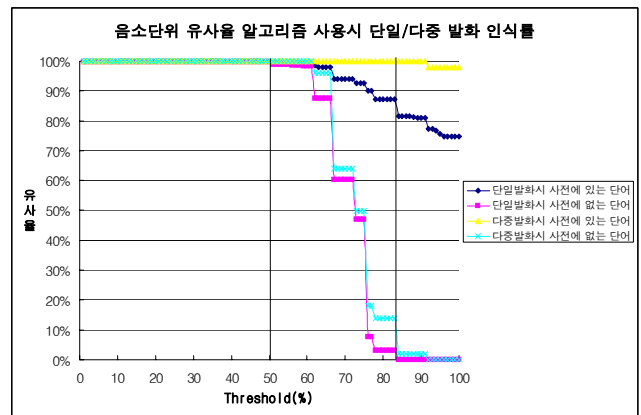
음소 단위 유사율 알고리즘과 단일발화를 사용한 경우, 실험결과 Threshold의 값이 50% 이상일 경우에는 사전에 있는 단어일 경우 유사율이 100%이고, 사전에 없는 단어일 경우에는 오인식률이 99.6%로 나타났다. 음절단위 유사율 알고리즘을 사용 시 다중발화에 비해서 사전에 없는 경우의 오인식률이 24%에서 99.6%로 급격히 증가하는 단점을 볼 수 있었다. 그러나 음소 단위 유사율 알고리즘의 Threshold값을 높임에 따라서 사전에 있는 단어는 인식률이 감소하나, 사전에 없는 단어는 급격히 감소하다가 Threshold 값이 85% 이상일 경우부터는 사전에 없는 단어의 오인식률이 4%로 현저히 감소되는 결과를 볼 수 있었다.

음소단위의 유사율 알고리즘과 단일발화를 사용할 때 단일발화를 하였을 때 전혀 인식이 안 된다는 지, 또는 네트워크상에 발생하는 음성데이터의 왜곡/손실에 대한 문제점을 보완하기 위해서 음소단위의 유사율 알고리즘과 다중발화를 적용하여 실험을 한 경우에는 Threshold 값이 85% 일 때 사전에 있는 단어의 인식률은 100%로 변함이 없으며, 사전에 없는 단어의 오인식률이 2%로 현저하게 감소되었다.

5. 결 론

네트워크 환경에서 음성 데이터를 입력할 때 데이터의 손실이나 왜곡의 정도가 크다. 그리고 음성

인식기에서 사용하는 알고리즘에서 나타나는 제약을 고려하여 본 논문에서 제안하는 알고리즘을 적용하지 않았을 경우 네트워크 상태에서 전달된 음성데이터가 음성인식기를 거쳐 나온 단어의 인식률이 80.4%였다.



[그림 3] 음소단위 유사성 알고리즘 적용 시 단일/다중 발화 인식률

이러한 네트워크 상에서의 음성데이터의 왜곡/손실된 데이터의 인식률을 보완하기 위해서 본 논문에서 제안한 음소단위 알고리즘과 [3]에서 제안한 음절단위 알고리즘에 다중발화를 이용하여 인식률을 사전에 있는 단어와 사전에 없는 경우에 대해서 비교분석한 결과 음절단위 유사성 알고리즘의 경우 24%였던 오인식률이 음소단위 유사성 알고리즘과 다중발화를 이용한 결과 오인식률이 2%로 현저히 감소되었다.

2%의 오인식이 되는 이유는 앞에서 언급하였듯이, 한음소만 틀린 경우에 나타난 문제이다. 향후에는 이런 문제를 해결하기 위해서는 본 논문이 제안한 방법이외에 가전제품별로 시간별·계절별 가중치를 두어 인식률 향상시키고자 하는 방법을 제안하고자 한다.

참고문헌

- [1] “최근의 홈 네트워크 기술동향 및 시장 전망”, ETRI IT 정보센터, 2003.6.
- [2] 이규환, 정민화, “G.723기반의 음성인식을 위한 변별적인 음성 특징 벡터 선정”
- [3] 황지수, 이창섭, 박준석, 김유섭, 박찬영 “이동단말에서 다중발화를 이용한 Home network 환경에서의 QoS 보장 연구” 한국정보처리학회 2004 춘계학술대회
- [4] 윤용욱, 정한민, 이근배, “어휘의미패턴을 이용한 음성인식 오류 검출 및 수정”, 제14회 한글 및 한국어 정보처리 학술대회 논문집, pp.62~68, 2002.
- [5] 윤현남, 김양우, 이필우 “VoIP 기술을 이용한 P2P 멀티미디어 채팅용 통신 프로그램의 구현
- [6] 서화정, 김주한, “음성인식기반의 전자의무기록 구현을 위한 오인식 의학용어 수정 시스템 개발”
- [7] 신중호, 박혁로, 이기호, “단어의 유사성 척도와 클러스터링 알고리즘”, 한국인지과학회 논문지, Vol.09, No.02, pp.37~44, 1998.