http://dx.doi.org/10.14400/JDC.2014.12.10.265

ERB 필터를 이용한 시맨틱 온톨로지 음성 인식 성능 향상

이종섭
세명대학교 교양과정부

Semantic Ontology Speech Recognition Performance Improvement using ERB Filter

Jong-Sub Lee
Dept. of General Education, Semyung University

요 약 기존의 음성 인식 알고리즘은 어휘들 간의 순서가 정해져 있지 않으며, 음성 인식 환경 변화에 따른 잡음으로 인한 음성 검출이 정확하지 못한 단점을 가지며, 검색 시스템은 키워드의 의미가 다양하여 정확한 정보를 인지하지 못한다. 본 연구에서는 사건 기반 시맨틱 온톨로지 추론 모델을 제안하였으며, 제안된 시스템에서 음성 인식 특징을 추출하기 위해 ERB 필터를 이용하여 특징 추출하는 모델을 구축하였다. 제안된 모델은 성능 평가를 위해 지하철역, 지하철 잡음을 사용하였고 잡음 환경의 SNR -10dB, -5dB 신호에서 잡음 제거를 수행하여 왜곡도를 측정한 결과 2.17dB, 1.31dB의 성능이 향상됨을 확인하였다.

주제어 : 정보 검색, 온톨로지, 시맨틱 웹, ERB

Abstract Existing speech recognition algorithm have a problem with not distinguish the order of vocabulary, and the voice detection is not the accurate of noise in accordance with recognized environmental changes. and retrieval system, mismatches to user's request are problems because of the various meanings of keywords. In this article, we proposed to event based semantic ontology inference model, and proposed system have a model to extract the speech recognition feature extract using ERB filter. The proposed model was used to evaluate the performance of the train station, train noise. Noise environment of the SNR-10dB, -5dB in the signal was performed to remove the noise. Distortion measure results confirmed the improved performance of 2.17dB, 1.31dB.

Key Words : Information Retrieval, Ontology, Semantic web, ERB

1. 서론

시맨틱 웹은 인터넷에서 웹상의 정보를 분석하여 사용자가 필요로 하는 정보를 자동으로 검색하여 추론할 수 있는 차세대 웹이다. 시맨틱 웹 기술을 사용하여 컴퓨터의 데이터베이스에 저장된 정보를 가지고 다양한 정보

* 이 논문은 2014학년도 세명대학교 교내학술연구비 지원에 의해 수행된 연구임.
Received 17 July 2014, Revised 17 August 2014
Accepted 20 October 2014
Corresponding Author: Jong-Sub Lee
(The Society of Digital Policy)
Email: 99jslee@semyung.ac.kr

ISSN: 1738-1916

를 지식 검색으로 사용자에게 제공할 수 있으며,시맨틱 웹 기술에는 XML, RDF, 온톨로지, OWL, SWRL 등이 있으며, 온톨로지는 XML과 RDF로 표현된 정보의 개념과 개념들 사이의 관계를 명확하게 설정하기 위한 수단으로 사용된다.

하드웨어 기술의 발전으로 일상에서 실용적인 음성 인식 시스템의 결과를 활용하기 위한 연구가 진행되고 있다. MFCC(Mel-Frequence Cepstral Coefficient)는 인간의 청취 영역에서 특정 잡음의 특성을 고려한 필터 설계 방법이며[1], 다양한 잡음 환경에서 사용자가 원하는 음성만을 선택적으로 추출하기 위한 시스템이 연구되고 있다[2]. 그러나 음성 인식 환경 변화에 따른 잡음으로 인한 음성 검출이 정확하지 못하므로 본 연구에서는 잡음에 강한 음성 특징을 추출하기 위해 ERB(Equivalent Rectangular Bandwidth) 필터를 이용하여 특징 추출하는 방법을 이용하였다. 일반적인 음성 특징 추출은 HMM(Hidden Markov Model) 특징 추출 방법을 이용하지만, 시맨틱 웹에서 사용하기 위한 음성 인식 시스템의 인식률 향상을 위해 ERB 필터를 사용한 방법을 제안한다. ERB 필터는 인간이 음성을 인지하는 과정을 컴퓨터로 수행하기 위해 시간-주파수 영역에서 음성 신호를 시간의 주기성과 채널간의 유사성을 파악 및 비교 연산하는 과정을 수행하고 세분화된 영역을 음원에 따라 그룹화 과정을 수행한다. 실험 및 분석을 위해 MFCC를 사용하여 음성에 대한 특징을 추출하고 잡음 분리율과 음성 인식률에 대하여 성능 평가를 수행하였다. 성능 평가를 위해 지하철역, 지하철 잡음을 사용하였고 잡음 환경의 SNR -10dB, -5dB 신호에서 잡음 제거를 수행하여 왜곡도를 측정한 결과 2.17dB, 1.31dB의 성능이 향상됨을 확인하였다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구에 대해 언급하고 3장에서는 음성 인식 성능 향상을 위한 모델 구성에 대해 설명하며, 4장에서는 시스템 평가를 수행하고 마지막으로 5장에서 결론을 맺는다.

2. 관련연구

2.1 온톨로지 추론

시맨틱 웹 기술에는 XML(extensible markup language), RDF(Resource Description Framework), 온톨로지, OWL(Web Ontology Language), SWRL(Semantic Web Rule Language) 등이 있다. XML은 웹 문서 자료를 구조화하기 위하여 만든 표식 언어로서 기계가 정보의 내용을 조금 이해할 수 있게 하였고, RDF는 웹상의 자원들을 3진구조로 표현함으로써 표현의 명확성을 향상하였다. 온톨로지는 XML과 RDF로 표현된 정보의 개념과 개념들 사이의 관계를 명확하게 설정하기 위한 수단이다. OWL은 클래스와 속성을 이용하여 온톨로지를 정의할 수 있다. SWRL은 온톨로지 규칙을 이용하여 추론할 수 있는 언어이다. OWL에서 클래스는 동일한 특성을 가진 개체들의 그룹을 나타내며, 속성은 클래스 간의 관계를 표현하는 객체속성과 속성 값이 자료형의 종류임을 표현하는 자료형 속성으로 구분된다. SWRL을 사용하면 인선 의사결정을 도와주는 인선 지식을 규칙 형태로 정의할 수 있다. SWRL은 OWL을 확장시킨 언어로서 OWL의 클래스와 속성을 이용하여 규칙을 정의한다. 온톨로지 기반의 검색 기법은 온톨로지를 구성하는 개념과 개념간의 관계를 이용하여 검색하는 기법이다. 이 기법에서는 온톨로지의 계층 구조간의 관계를 정의하고 관계를 이용한 의미적인 검색을 지원한다. 온톨로지 기반의 검색 기법에 관한 연구는 TAP과 SEWISE 시스템 등이 진행되었다. TAP는 미국의 스탠포드 대학에서 SUO(Standard Upper Ontology) 온톨로지와 Cyeupper 온톨로지를 이용하여 컨텐츠에 대한 검색 영역을 확장하기 위해 제안된 시스템이다.

2.2 MFCC

음성 인식하여 특징을 추출하기 위한 방법은 사용자의 음성 인식 능력을 주관적으로 반영하므로 기존 주파수를 mel-scale로 변형한 필터를 비선형적으로 분포시켜 이용한다. 이러한 필터를 이용해서 구한 음성 벡터를 MFCC(Mel-Frequence Cepstral Coefficient)라 한다[7]. MFCC의 특징을 추출하는 과정은 음성 신호를 프레임별로 추출한 후 pre-emphasis를 수행하게 되며 다음 수식으로 나타낼 수 있다[4, 5].

$$\widehat{S}_n = S_n - \alpha S_{n-1} \quad (1)$$

pre-emphasis를 거친 음성신호를 가지고 해밍 윈도 필터를 가지고 다음 수식으로 표현할 수 있다.

$$W_H(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right) \quad (2)$$

또한, 이 수식은 FFT 분석을 가지고 파워 스펙트럼에 대한 내용을 다음 수식으로 나타낼 수 있다.

$$X(k) = \sum_{n=0}^{N-1} x(k)\, W h^{kn} \quad (3)$$

식 (3)을 mel-scale 필터 뱅크에 적용하면 다음 같이 수식으로 나타낼 수 있다.

$$mel = 2595 \cdot \log_{10}\left(1 + \frac{f}{700}\right) \quad (4)$$

이 결과를 주파수 스케일로 역변환을 하면 다음 같수식으로 표현할 수 있다.

$$f = 700 \cdot \left(\left(\frac{\exp_{10}(mel)}{2{,}595}\right) - 1\right) \quad (5)$$

2.3 HMM(Hidden Markov Model)

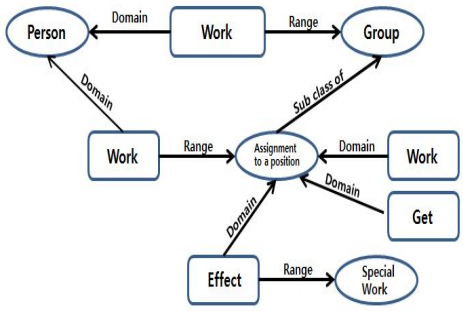
HMM(Hidden Markov Model)은 음성 신호의 스펙트럼 변화와 시간에 따른 변화를 동시에 모델링할 수 있으며 유한개의 상태와 상태전이들을 사용한다. 인식단계에서는 주어진 데이터를 사용하여 파라미터를 추정하고 새로 입력된 음성에 대하여 가장 적합한 모델을 찾는다.

HMM은 일련의 연속된 상태들로부터 이산 신호를 생성하는 확률 과정 모델로 표현된다. 모델은 전이 확률에 따라 상태를 바꾸며 특정 상태는 그 상태의 출력 확률에 따라 하나의 관측을 발생시킨다. 모델의 파라미터를 추정하기 위하여 카테고리 정보가 있는 음성 데이터베이스를 사용하며 각 모델을 위한 충분한 데이터가 있을 경우 실제 음성에 존재하는 다양성을 잘 표현할 수 있는 강인한 모델링이 가능하다. 음성 인식을 HMM의 접근 방법으로 수행하기 위해서는 관측 열 특징 벡터 x가 입력되었을 때 확률이 어떻게 계산되느냐의 문제로 전향 알고리듬과 후향 알고리듬을 이용하여 계산한다[6].

3. 시스템 모델

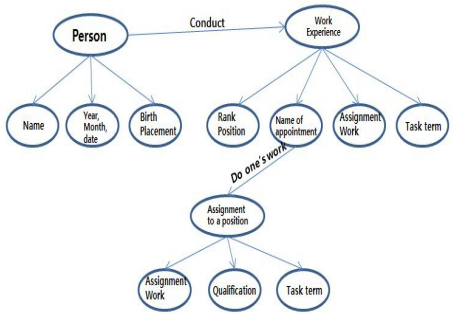
3.1 시맨틱 온톨로지 추론

본 연구에서는 시맨틱 온톨로지 시스템 적용을 위해 인적 정보가 보직에 연결되는 사건이 성립하는 추론을 설정하고, 가변적인 속성들을 이용하여 보직에 적합한 인물을 검색하는 시스템을 설계한다[10]. RDF 스키마의 자원 특성을 정의하는 기능과 다른 자원과의 관계를 정의하는 기능을 이용하여 개념을 검색하도록 RDF를 설계한다. RDF 스키마를 이용하면 RDF의 부족한 어휘 정의 능력을 보완할 수 있어서 새로운 용어 정의가 용이하다. 인물과 보직 개념의 명확한 정의를 위하여 주체(subject)를 정의역으로 제한하고, 객체(object)를 치역으로 제한하여 각 클래스들의 형태(type)를 유지하도록 설정한다.



[Fig. 1] Event ontology inference model

[Fig. 1]은 인선을 위한 온톨로지를 설계하기 위하여 사건형 RDF를 설계한 것이다. 인선 온톨로지는 [Fig. 2]에서와 같이 인선 의사결정에 사용되는 다양한 개념들과 그들의 관계를 정의한 온톨로지이다. 인선 온톨로지는 인물 지식, 경력 지식, 보직 지식 등의 지식으로 구성된다. 본 연구에서 추론 엔진은 정의된 인사 규칙, 인선 온톨로지와 지식들을 대상으로 데이터베이스에서 정보를 선발하는 모듈이며, OWL과 SWRL로 작성된 지식 정보를 이용하여 실시간으로 입력된 인선 정보에 대한 추론을 실행하고, 새로운 정보가 추론될 경우 추론된 결과를 메모리에 저장한다.



[Fig. 2] Ontology for personnel selection

3.2 ERB 필터 이용한 음성 특징 추출

음성 특징 추출 과정을 위한 일반적인 과정은 자기 상관 계수(Autocorrelation)를 사용하며 식 (6)과 같이 나타낸다[5].

$$R(k) = \sum_{t=1}^{n-k} (X_t - \mu)(X_{t-k} - \mu) \qquad (6)$$

시간 t에서 입력되는 $n-k$까지의 계수를 계산하며 μ는 평균을 나타내고 σ^2는 분산을 나타낸다[9].

또한, 음성 인식 성능 향상을 위해 잡음을 분리하기 위해 ERB 필터를 사용하여 음성의 특징을 추출하며, 한 ERB 필터는 시간 영역에서 임펄스 응답 $h(t)$로 정의하며 다음과 같이 수식적으로 나타낼 수 있다.

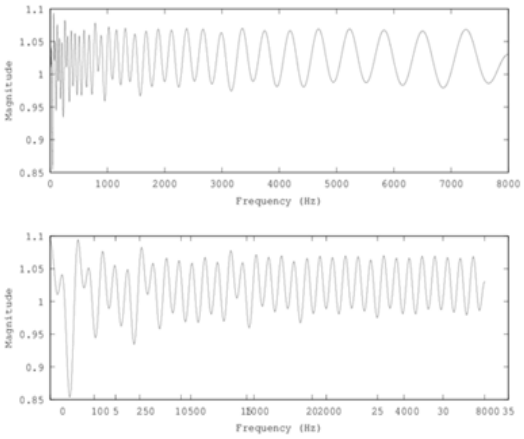
$$h(t) = kt^{n-1}\exp(-2\pi Bt)\cos(2\pi f_c t + \phi) \qquad (7)$$

k는 출력 이득, B는 필터 대역폭, n은 필터의 차수, f_c는 중심주파수, ϕ는 위상을 나타낸다.

음성 신호를 블록 단위로 구분하는 최소 단위는 벡터 단위이며, 피치 주기가 프레임과 프레임 사이에 존재할 경우 음성 신호의 정보를 손실할 우려가 있어 프레임 사이의 신호에 대해 오버래핑(Overlapping) 작업을 수행한다.

4. 실험 결과

본 연구에서 ERB 필터를 이용하여 시맨틱 온톨로지 모델에 적용하고, 음성 특징을 추출하여 음성 인식 성능을 평가하며, 실험 분석을 위해 Ohio주립대학의 PNL에서 채집한 100 Non-speech Sounds 생활 환경 잡음에서 10개의 잡음을 5명의 화자의 임의의 음성으로 실험을 수행하였다. 제안된 모델의 성능 평가를 위해 지하철역 , 지하철 잡음을 사용하였고 잡음 환경의 SNR -10dB, -5dB 신호에서 잡음 제거를 수행하여 왜곡도를 측정한다. 다음 [Fig. 3]은 ERB 필터를 사용한 음성 특징 추출과정에서 잡음이 제거되기 전과 제거된 후의 신호를 나타낸다.



[Fig. 3] Noise Speech Signa and Noise Removed Speech Signal.

식 (8)를 통하여 SNR의 향상 정도를 평가하였다[8].

$$SNR(dB) = 10\log_{10}\frac{\sum_{n=1}^{N} x^2(n)}{\sum_{n=1}^{N}[x(n)-\hat{x}(n)]^2} \qquad (8)$$

$x(n)$은 노이즈가 혼합된 음성을 의미하고 $\hat{x}(n)$은 분리된 음성을 의미하며 n은 시간 인덱스를 나타낸다.

제안된 모델의 성능 평가를 위해 지하철역 , 지하철 잡음을 사용하였고 잡음 환경의 SNR -10dB, -5dB 신호에서 잡음 제거를 수행하여 왜곡도를 측정한 결과 2.17dB, 1.31dB의 성능이 향상됨을 확인하였다.

〈Table 1〉 Distortion Evaluation of Noise Removal

Noise	-10dB		-5dB	
	MFCC	Proposed Method	MFCC	Proposed Method
Subway Station	4.66	2.40	3.15	2.27
	4.01	2.21	3.71	2.31
	4.37	2.31	3.23	2.21
Subway Noise	5.36	3.31	4.27	3.37
	6.53	3.41	4.31	3.27
	5.56	3.31	4.37	3.31
Improvement (Average).	2.17dB		1.31dB	

5. 결론

하드웨어 기술의 발전으로 일상에서 실용적인 음성 인식 시스템의 결과를 활용하기 위한 연구가 진행되고 있다. MFCC(Mel-Frequence Cepstral Coefficient)는 인간의 청취 영역에서 특정 잡음의 특성을 고려한 필터 설계 방법이며, 본 연구에서는 잡음에 강한 음성 특징을 추출하기 위해 ERB 필터를 이용하여 특징 추출하는 방법을 이용하였다. ERB 필터는 인간이 음성을 인지하는 과정을 컴퓨터로 수행하기 위해 시간-주파수 영역에서 음성 신호를 시간의 주기성과 채널간의 유사성을 파악 및 비교 연산하는 과정을 수행하고 세분화된 영역을 음원에 따라 그룹화 과정을 수행한다. 실험 및 분석을 위해 MFCC를 사용하여 음성에 대한 특징을 추출하고 잡음 분리율과 음성 인식률에 대하여 성능 평가를 수행하였다. 성능 평가를 위해 지하철역, 지하철 잡음을 사용하였고 잡음 환경의 SNR -10dB, -5dB 신호에서 잡음 제거를 수행하여 왜곡도를 측정한 결과 2.17dB, 1.31dB의 성능이 향상됨을 확인하였다. 본 연구에서의 음성 특징 추출에서 추출된 음성 특징들의 연관성과 이들의 특성 정보를 목록으로 제공하는 연구를 필요로 한다.

ACKNOWLEDGMENTS

This paper was supported by the Semyung University Research Grant of 2014

REFERENCES

[1] Yun-Kyung Lee, Oh-Wook Kwon, Application of Shape Analysis Techniques for Improved CASA-Based Speech Separation. The Korean Society of Phonetic Sciences and Speech Technology:MALSORI. No. 65, pp. 153-168, 2008.

[2] Tae-woong Choi, Soon-Hyub Kim. Target Speech Segregation Using Non-parametric Correlation Feature Extraction in CASA System. The Journal of the Acoustical Society of Korea. Vol. 32, No. 1, pp. 79-85. 2013.

[3] T. T. pham, J. Y. Kim, S. Y. Na, S. T. Hwang, "Robust Eye Localization for Lip Reading in Mobile Environment," Proceddings of SCIS&ISIS in Japan, pp.385-388, 2008.

[4] P. Li, Y. Guan, B. Xu, W. Liu.. Monaural speech separation based on computational auditory analysis and objective quality assessment of speech. IEEE Trans, audio, speech, and language processing. Vol. 14, No. 6, pp. 2014-2022, 2006.

[5] A. P. Klapuri. Multipitch analysis of polyphonic music and speech signals using an auditory model. IEEE trans, Audio, Speech and Language Process. Vol. 16, No. 2, pp. 255-266, 2008.

[6] Y. Shao, S. Srinivasan, Z. Jin, D. Wang. A Computational Auditory Scene Analysis System for Robust Speech Recognition. Computer Speech & Language. Vol. 24, No. 1, pp. 77-93, 2010.

[7] G. Hu, D. Wang. A Tandem Algorithm for Pitch Estimation and Voiced Speech Segregation. IEEE Trans, Audio, Speech and Language Processing. Vol. 18, No. 8, pp. 2067-2079, 2010.

[8] Chanshik Ahn, Beomseung Kim, Taewoong Choi, Soonhyub Kim. Colored Noise Cancellation Algorithm using Average Estimator. Proceedings of the Acoustical Society of Korea Conference. Vol. 29, No. 1, pp. 71-74, 2012.

[9] T. T. Pham, M. G. Song, J. Y. KIm, S. Y. Na, S. T. Hwang, "A Robust Lip Center Detection in Cell Phone Environment," Proceedings of IEEE

Symposium on Signal Processing and Information Technology, pp.390-395, Sarajevo, December, 2008.
[10] Byungwook Lee, Improvement of the Semantic Information Retrieval using Ontology and Spearman Corelation Coefficients, The Journal of Digital Policy and Management. Vol. 11, No. 11, pp. 351-357, 2013.

이 종 섭(Lee, Jong Sub)

- 1993년 8월 : 광운대학교 대학원 전자계산학과 (이학석사)
- 1997년 2월 : 광운대학교 대학원 전자계산학과 (박사수료)
- 2012년 3월 ~ 현재 : 세명대학교 교양과정부 교수
- 관심분야 : 분산객체처리시스템, 음성/음향 신호 처리, 차량 통신
- E-Mail : 99jslee@semyung.ac.kr