

# 비모수적 상관계수를 이용한 시맨틱 온톨로지 음성 정보 추출

이병욱\*

가천대학교 컴퓨터미디어융합학과

## Semantic Ontology Speech Information Extraction using Non-parametric Correlation Coefficient

Byungwook Lee\*

Dept. of Computer Media Convergence, Gachon University\*

**요약** 질의 키워드의 출현 빈도수가 높은 문서를 검색하면 키워드의 의미가 다양하여 정확한 정보를 인지하지 못하며, 기존 검색 시스템의 온톨로지 구성만으로는 검색된 문서들이 사용자의 요구에 부합되지 않는 문제점을 가진다. 본 연구에서는 시맨틱 웹 기술을 기반으로 인사관리에서 인선에 필요한 다양한 개념들과 지식으로 구성된 인선 온톨로지와 인선 규칙들을 구축하고 이들을 지원하는 인선 절차와 인선 결과의 적합성을 확인할 수 있는 시스템을 제안한다. 또한, 이를 기반으로 비모수적 상관 계수를 이용하여 음성 정보를 추출하는 방법을 사용하여 평균 SNR이 0.752dB 감소됨을 보임으로써 제안한 방법의 우수성을 확인하였다.

**주제어** : 키워드, 정보 검색, 온톨로지, 시맨틱 웹

**Abstract** On retrieving high frequency keywords in information retrieval system, mismatches to user's request are problems because of the various meanings of keywords in the existing ontology configuration. In this paper, it is to construct personnel selection ontology and rules in personnel management which are composed of various concepts and knowledges based on semantic web technology and suggest selection procedures to support these rules and knowledge retrieval system to verify suitability of selection results. This system utilizes a method of extraction of speech features by using non-parametric correlation coefficient. This proposed method has been validated by showing that the result average SNR of the experiment evaluation of the proposed techniques was shown to be decreased by .752dB.

**Key Words** : ontology, semantic web, non-parametric correlation coefficient

### 1. 서론

전통적인 인선 시스템들은 인사 데이터베이스를 대상

으로 인선 조건에 맞는 인물을 키워드 기반으로 검색하는 방식이다. 키워드 기반의 질의 방식으로는 정보 규모가 방대할 때 너무 많은 검색 결과가 제공되어 오히려 불

\* 이 논문은 2013년도 가천대학교의 교내연구비 지원에 의한 결과임.(GCU-2013-R188)

Received 8 July 2013, Revised 2 August 2013

Accepted 20 September 2013

Corresponding Author: Byungwook Lee(The Society of Digital Policy)

Email: leebw@gachon.ac.kr

ISSN: 1738-1916

© The Society of Digital Policy & Management. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

편하다. 그 이유는 질의 키워드의 출현 빈도수가 높은 문서를 검색하기 때문에 관련 문서를 많이 찾아주지만, 키워드의 의미가 다양하여 검색된 문서들이 사용자의 요구에 부합하기는 어렵다.

시맨틱 웹이란 컴퓨터 프로그램이 웹상의 정보를 이해하고 인간 대신 자동으로 정보를 검색하고 추론할 수 있는 차세대 웹을 의미한다. 시맨틱 웹 기술을 이용하면 컴퓨터가 인사 정보를 이해하여 직원의 경력과 전공과 업무 능력에 관한 다양한 지식들을 활용하여 지식 검색을 수행하고 결과로 얻은 인선 자료를 인사 담당자에게 제공할 수 있다. 시맨틱 웹 기술에는 XML, RDF, 온톨로지, OWL, SWRL 등이 있으며, 온톨로지는 XML과 RDF로 표현된 정보의 개념과 개념들 사이의 관계를 명확하게 설정하기 위한 수단이다.

본 연구에서는 시맨틱 웹에서 사용하기 위한 음성 인식 시스템의 인식을 향상을 위하여 입력된 음성으로부터 원하는 음성만 선택하여 추출하는 선택적 음성 추출을 위한 특징 추출 기법을 제안한다. 선택적 음성 추출을 위한 특징 추출은 시간의 연속성과 채널 간 유사성을 이용하여 모델링하였고 비모수적 상관 계수를 이용한 음성 특징을 추출하는 방법을 사용하였다. 제안 기법을 OHIO 대학 PNL의 Non speech Sounds와 ETRI 445 PBW 음성 데이터베이스와 혼합하여 선택적 음성 특징 추출 실험을 수행 평가한 결과 평균 SNR이 0.752dB 감소됨을 보임으로써 기존 방법보다 제안한 방법의 우수성을 확인하였다.

본 논문의 구성은 I장의 서론에 이어 II장에서는 기존의 온톨로지 추론기술과 음성 특징 추출 관련 연구들에 대해 분석하고 III장에서는 시맨틱 인선 규칙 및 음성 특징 추출 기법을 설명하고, IV장의 실험 결과 및 고찰, 마지막으로 V장에서는 결론을 제시한다.

## 2. 관련 연구

### 2.1 시맨틱 웹과 온톨로지

XML은 웹 문서 자료를 구조화하기 위하여 만든 표식 언어로서 기계가 정보의 내용을 이해할 수 있게 하였고, RDF는 웹상의 자원들을 3진구조로 표현함으로써 표현의 명확성을 향상하였다. OWL은 클래스와 속성을 이용

하여 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[1]. 온톨로지 기반의 검색 기법에 관한 연구는 TAP과 SEWISE 시스템 등이 진행되었다. [2]에 의하면 TAP에서는 RDF 기반의 RDQL 형태로 표현된 질의문을 처리해야 하는 문제점이 있다.

### 2.2 음성 특성 추출

음성 특정 구간별로 주파수 응답 특성을 표현하기 위한 모델링 방법으로 감마톤(Gamma Tone) 채널을 가진 ERB(Equivalent Rectangular Bandwidth) 필터뱅크(Filter Bank)를 사용하게 되는데 감마톤 필터는 시간 영역에서 식 (1)과 같이 임펄스 응답  $h(t)$ 로 정의된다[3].

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

$k$ 는 출력 이득,  $B$ 는 필터 대역폭,  $n$ 은 필터의 차수,  $f_c$ 는 중심 주파수,  $\phi$ 는 위상을 나타낸다. 필터의 주파수 응답  $|H(f)|$  와 최대 이득  $|H(f_{\max})|$  이 주어졌을 때, ERB는 다음 식 (2)과 같이 정의된다.

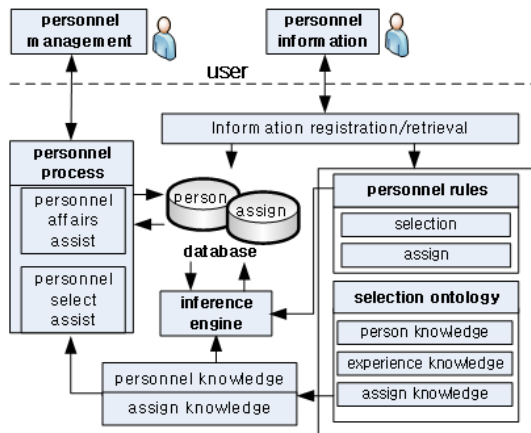
$$ERB = \frac{\int |H(f)|^2 df}{|H(f_{\max})|^2} \quad (2)$$

ERB는 일정한 이득  $|H(f_{\max})|$  과 에너지를 갖는 사각(rectangular) 필터의 대역폭, 감마톤 필터의 에너지가 같은 대역폭을 나타낸다. 외이, 중이, 내이의 청각 기관을 모델링하여 나온 T-F 단위 계수는 하나의 장면(scene)으로 표현하여 달팽이관 지도(cochleagram)[4]로 표현된다.

### 3. 시맨틱 웹에서 선택적 음성특징 추출

#### 3.1 시맨틱 웹

시맨틱 웹 기술 지원을 위한 인선 시스템은 [그림 1]과 같다. 인사 규칙들은 SWRL로 인선 지식들은 OWL을 이용하여 온톨로지로 데이터베이스에 저장한다. 인선/보직 지식은 데이터베이스 안에 있는 RDF 형식의 인선 지식과 보직 지식에 대한 검색 기능을 제공한다.



[Fig. 1] personnel select system structure

<Table 1> Example of personnel selection rules for mechanical engineer

rules	expression of rules
machinery_subsection chief	employee(?a) ∩ major(?a, mechanical engineering) ∩ work for years(?a,?c) ∩ swrla:greaterThan(?c,4) ∩ performance rating(?d) ∩ swrlb:greaterThan(?d,2) → machinery_subsection chief(?a)
machinery_section chief	employee(?a) ∩ major(?a, mechanical engineering) ∩ work for years(?a,?c) ∩ swrla:greaterThan(?c,7) ∩ performance rating(?d) ∩ swrlb:greaterThan(?d,3) → machinery_subsection chief(?a)
machinery_manager	employee(?a) ∩ major(?a, mechanical engineering) ∩ work for years(?a,?c) ∩ swrla:greaterThan(?c,12) ∩ performance rating(?d) ∩ swrlb:greaterThan(?d,4) → machinery_manager(?a)

인선 규칙은 <표 1>에서와 같이 인선에 도움을 주는 인선 자료 지식을 SWRL로 표현한 것을 나타낸다. 본 논문에서는 인선 지식을 인선 규칙과 보직 규칙으로 구분

하여 정의하였다.

시맨틱 웹 기술로 운용되는 인선 시스템에서는 인선 온톨로지를 기반으로 표현된 RDF 형태의 직원과 보직 정보가 필요하다. RDF의 객체에서 시간에 따라 변하는 정보는 타원으로 변하지 않는 정보는 사각형으로 표현하며, 등록 시스템을 통해 입력된 정보는 정보 등록 모듈에 보내지고 다시 미리 정의된 변환 규칙에 따라 RDF 형태로 변환된 뒤에 데이터베이스의 인물(직원) 테이블과 보직 테이블에 저장된다.

#### 3.2 음성 특징 추출

특징 추출 과정은 청각 기관을 모델링하여 얻어진 감마톤의 필터뱅크 주파수 응답 계수의 시간 축 주기성 특징과 주파수 축의 채널 간 주파수 유사성의 특징을 얻는 과정이다. 주기성 특징을 찾는 전통적인 방법은 자기 상관 계수(Autocorrelation)를 이용하며 식 (3)과 같이 나타낸다[5].

$$R(k) = \sum_{t=1}^{n-k} (X_t - \mu)(X_{t-k} - \mu) \quad (3)$$

시간  $t$ 에서 입력되는  $n - k$ 까지의 계수를 계산하며  $\mu$ 는 평균을 나타내고  $\sigma^2$ 는 분산을 나타낸다[6].

Kendall's Tau는 두 변수의 일치성을 이용하여 계산한다. 모든 가능한 두 관측치의 조합에 대하여 한 변수의 값이 증가 또는 감소할 때 다른 변수의 값의 증가 또는 감소의 일치성을 확인하기 위해 사용하였다[7].

$n$ 은 변수의 관측치를 나타내고  $t_i$ 는 주어진 순위에서 동일한  $X$ 의 관측치를 나타내며 식 (4)와 같이 나타낸다.

$$T_1 = \left( \frac{n(n-1) - \sum t_i(t_i - 1)}{2} \right) \quad (4)$$

$u_i$ 는 주어진 순위에서 동일한  $Y$ 의 관측치를 나타내며 식 (5)와 같이 나타낸다.

$$T_2 = \left( \frac{n(n-1) - \sum u_i(u_i - 1)}{2} \right) \quad (5)$$

$sgn$ 은  $sgn(z) = 1$ 이면  $z > 0$ 를 나타내고  $sgn(z) = 0$ 이면  $z = 0$ 를 나타내며  $sgn(z) = -1$ 이면  $z < 0$ 를 나타낸다. 식 (4)와 식 (5)에서 관측된 타우를 식 (6)과 같이 나타낸다.

$$\tau = \frac{\sum_{i < j} sgn(x_i - x_j)sgn(y_i - y_j)}{\sqrt{T_1 T_2}} \quad (6)$$

계산된 계수는 정규화를 통해 0~1사이의 값을 갖는다. 채널의 주기 패턴이 일치성이 높으면 1에 가까운 값을 보이고 반대의 경우에는 0에 값에 가까운 값을 보이게 된다. 비모수 상관계수 예측을 통해 음성 신호 정보를 만드는 과정에서는 현재 입력된 음성 신호 정보에서 특징 추출 확률을 구하고 이를 기반으로 기준 특징을 갱신하는 과정이 수행된다.

#### 4. 실험 결과 및 분석

본 논문의 시맨틱 웹기반 선택적 음성 특징 추출 기법의 성능 비교를 위하여 다양한 환경에서 채집한 OHIO 대학 PNL(Perception and Neurodynamics Laboratory)의 100 Non-speech Sounds[6]를 사용하였으며 ETRI 445 PBW 음성 데이터베이스와 혼합하여 선택적 음성 특징 추출 실험을 수행하였다.

특징 추출의 과정으로 채널 간의 유사성을 판단하기 위해 교차 상관 계수 방법과 제안한 방법을 비교하였다. <표 2>는 임의의 동일 음성 첫 번째 프레임에 해당하는 각 채널 별 상관 계수의 연산 과정을 수행한 결과 데이터를 각각 비교하였다.

<Table 2> compare cross with non-parametric correlation coefficient

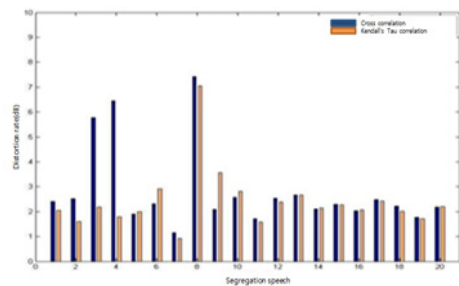
channel	cross	non-parametric
1	(0~1)	(0~1)
2	0.8407	0.7683
3	0.7864	0.8725
4	0.8823	0.8801
5	0.7136	0.6839

유사성이 약한 채널은 0에 가까운 값을 나타내고 유사성이 강한 채널은 1에 가까운 값을 나타내었다. 교차 상관 계수는 상관도에 의해 음성 특징을 추출하고 제안한 non-parametric은 일치성에 의해 음성 특성을 추출하여 유사성이 더 높은 결과를 얻을 수 있었다[8].

추출된 음성의 평가를 위하여 암묵 신호 제거 알고리즘에서 사용하는 공통 척도인 왜곡도 SNR (Signal to Noise Ratio)[9]를 사용하여 평가하였으며 식 (7)과 같이 나타낸다[10,11].

$$SNR = \frac{P_{signal}}{P_{noise}} \quad (7)$$

$P_{signal}$ 는 잡음과 혼합되기 전 깨끗한 음성을 의미하고,  $P_{noise}$ 는 잡음 신호를 입력으로 시스템에 통과하여 얻어진 선택적 추출 신호를 의미한다. SNR은 값이 작을수록 두 음성의 유사도가 더 높다고 판단한다. [그림 2]에서 일부 데이터에서는 교차 상관을 이용한 선택적 특징 추출 성능이 우수함을 보였지만 전체 평균에서는 제안한 비모수적 상관을 이용한 선택적 특징 추출이 더 우수한 성능을 보였다.



[Fig. 2] Segregation speech SNR evaluation

#### 5. 결론

시맨틱 웹 기술을 이용하면 컴퓨터가 인사 정보를 이해하여 직원의 경력과 전공과 업무 능력에 관한 다양한 지식들을 활용하여 지식 검색을 수행할 수 있으며, 시맨틱 웹 기술에는 XML, RDF, 온톨로지, OWL, SWRL 등이 있으며, 온톨로지는 XML과 RDF로 표현된 정보의 개념과 개념들 사이의 관계를 명확하게 설정하기 위한 수

단이다.

본 연구에서는 시맨틱 웹에서 사용하기 위한 음성 인식 시스템의 인식을 향상을 위하여 입력된 음성으로부터 원하는 음성만 선택하여 추출하는 선택적 음성 추출을 위한 특징 추출 기법을 제안하였으며, 비모수적 상관 계수를 이용한 음성 특징을 추출하는 방법을 사용하였다. 제안 기법을 OHIO대학 PNL의 Nonspeech Sounds와 ETRI 445 PBW 음성 데이터베이스와 혼합하여 선택적 음성 특징 추출 실험을 수행 평가한 결과 평균 SNR이 0.75dB 감소됨을 보임으로써 기존 방법보다 제안한 방법의 우수성을 확인하였다.

## ACKNOWLEDGMENTS

This work was supported by the Gachon University research fund of 2013.”(GCU-2013-R188)

## REFERENCES

- [1] Horrocks, I., Patel-Schneider, P. F., Boley, H., Tabet, S., Grosz, B., Dean, M., "SWRL: A Semantic Web Rule Language Combining OWL and RuleML", W3C Member Submission 21 May 2004, <http://www-w3.org/submission/SWRL>.
- [2] R. Gaa, R. McCool and E. Miller, "Semantic Search." Proc. Int'l Conf. on WWW, pp. 700-709, 2003
- [3] Jean-Luc Schwartz, Frederic Berthommier and Christophe Savariaux, "Seeing to Hear Better: Evidence for Early Audio-Visual Interactions in Speech Identification," ERIC Journal Articles : Reports-Research, Cognition, vol.93, no.2, pp. 69-pp.78, Sep, 2004.
- [4] C. C. Chibelushi, F. Deravi, and J. S. Moson, "A review of speech-based bimodal recognition," IEEE Trans. Multimedia, vol.4, no.1, pp23-37, Mar. 2002.
- [5] T. T. pham, J. Y. Kim, S. Y. Na, S. T. Hwang, "Robust Eye Localization for Lip Reading in

Mobile Environment," Proceedings of SCIS&ISIS in Japan, pp.385-388, 2008.

- [6] 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.
- [9] Hu, G. and Wang, D.L., "Monaural speech segregation based on pitch tracking and amplitude modulation". IEEE Transactions on Neural Networks, Vol. 15, 1135-1150, 2004.
- [7] Raj, B., Seltzer, M.L., Stern, R.M., "Reconstruction of Missing Features for Robust Speech Recognition", Speech Communication, Vol. 43, Issue 4, pp.275-296, September 2004.
- [8] Shao Y. and Wang D.L. "Model-based sequential organization in cochannel speech". IEEE Transactions on Audio, Speech, and Language Processing (formerly IEEE Transactions on Speech and Audio Processing), vol. 14, 289-298, 2006.
- [10] Cooke, M. P. "A glimpsing model of speech perception in noise", Journal of the Acoustical Society of America, in press, 2006.
- [11] Cooke, M. P., Barker, J., Cunningham, S. P. and Shao, X., "An audio-visual corpus for speech perception and automatic speech recognition", submitted to J. Acoust. Soc. Amer. [status: submitted 29 Nov 2005].

## 이 병 욱(Lee, Byung wook)



- 1973년 2월 : 연세대학교 공학사
- 1984년 2월 : George Washington Univ. Dept. of Computer Science (공학석사)
- 1994년 2월: 중앙대학교 전자계산학과(공학박사)
- 1985년 3월 ~ 현재 : 가천대학교 컴퓨터미디어융합학과 교수

- 관심분야 : 데이터베이스, 분산 시스템, 디지털 융합
- E-Mail : [leebw@gachon.ac.kr](mailto:leebw@gachon.ac.kr)