

베이지안 망을 이용한 온톨로지의 구축에 관한 연구

장성원^a 이진창^b

^a 삼성경제연구소

137-072, 서울시 서초구 서초2동 1321-15번지 삼성생명 서초타워
Tel: 02-3780-8147, Fax: 02-3780-8006, E-mail: serijsw@naver.com

^b 성균관대학교 경영학부

110-745, 서울시 종로구 명륜동 3-53
Tel: 02-760-0505, Fax: 02-745-4566, E-mail:leekc@skku.ac.kr

Abstract

의미적 지식기반인 온톨로지(ontology)에 대한 관심이 높아지고 있다. 온톨로지란 어휘나 개념의 정의 또는 명세로서, 인간과 컴퓨터의 의사소통 또는 지식의 표현과 저장, 활용 및 재사용을 위해 이용된다. 그러나 온톨로지를 구축하는 대부분의 방법은 체계적이거나 자동적이지 못하다. 도메인 전문가에 의존하는 전통적인 온톨로지 구축 방법은 시간과 비용이 많이 소요된다. 온톨로지 구축 틀은 많이 있지만 아직 인간의 노력을 필요로 한다. 또한 변화하는 도메인 지식을 온톨로지에 신속하게 반영하는 것은 어려운 일이다. 본 연구는 이러한 한계를 해결하기 위해, 도메인 전문가의 지식이나 경험을 최소화하면서 자동적으로 도메인 지식을 얻을 수 있는 방법을 제시하였다. 이 방법은, 데이터 기반의 도메인 지식을 대상으로, 베이지안 망(Bayesian network)이 갖고 있는 데이터 분석에서의 장점과 온톨로지와의 관련성을 이용하여 온톨로지를 자동적으로 구축하는 것이다. 평판(flat panel) TV 경기예측 사례를 통하여 온톨로지를 구축하는 과정을 알아보았다. 구축과정의 타당성을 확보하기 위하여 디스플레이 산업 전문가들과의 인터뷰를 통하여 온톨로지를 완성하고, 해당 온톨로지의 타당성 검증을 위하여 멤버체크를 한 결과 매우 높은 타당성을 얻을 수 있었다. 본 연구에서 제안하는 온톨로지는, 실제로 산업경기 예측을 계획하고 구축하며 미래 의사결정지원시스템을 설계하기 위한 주요 구성요인으로 제공될 수 있을 것이다.

Keywords:

온톨로지, 베이지안 망, 디스플레이, 평판TV

1. 서론

지식경영에서 온톨로지는 정보를 제공하고 저장하는 역할을 한다. 그러나 온톨로지를 구축하는데 몇

가지 부족한 점이 있다[4]. 첫째, 기존 온톨로지를 재이용하거나 통합하는 표준이 부족하다. IEEE 워킹그룹이나 스탠포드 대학과 같이 온톨로지의 표준을 만드는 몇몇 기관이 있으나, 온톨로지 언어의 다양성 때문에 기존 온톨로지의 통합은 매우 어려운 일이고, 또한 이로 인해 온톨로지의 재이용도 어렵다. 둘째, 자동적으로 지식을 획득하는 방법이 부족하다. 온톨로지의 구축은 시간과 비용이 소요되는 절차이다. 셋째, 분류하는데 융통성이 부족하다. 예를 들면, 웹 페이지가 급속히 증가하고 없어지고 있기 때문에, 수동 분류 구조는 웹 페이지의 동태적 변화를 따라가기 어렵다. 비록 현재 온톨로지 구축 방법이 부분적으로 자동 분류 구조를 이룰 수 있다 하더라도, 거기에는 한계가 있다.

위와 같은 온톨로지의 한계 중에서 본 연구는 온톨로지의 지식획득 방법과 분류의 융통성에 대해 초점을 맞춘다. 온톨로지를 구축하는 대부분의 방법은 체계적이거나 자동적이지 못하다[18]. 전통적인 온톨로지 구축은 도메인 전문가에 의존하지만, 이것은 시간과 비용이 소요되고 논쟁의 대상이 된다. METHONTOLOGY[8], OntoTrack[11] 등과 같이 많은 온톨로지 구축 틀이 있지만 온톨로지 구축은 아직 인간의 노력을 필요로 한다. 따라서 온톨로지 구축과 응용에 대한 대부분의 연구는 메뉴얼 구축을 가정하고 있다. 또한 변화하는 도메인 지식을 온톨로지에 신속하게 반영하는 것은 어려운 일이다.

본 연구는 온톨로지 구축 과정에서 지식을 획득하는데 시간과 비용이 많이 소요되는 한계를 보완하기 위해 다음과 같은 두 가지 연구목적을 제안한다. 첫째, 도메인 전문가의 지식이나 경험을 최소화하면서, 자동적으로 도메인 지식을 얻을 수 있는 방법을 제시하는 것이다. 두 번째 연구목적은, 변화하는 도메인 지식이 비교적 신속하게 온톨로지에 반영될 수 있는 방법을 제시하는 것이다.

본 연구는 이러한 두 가지 연구목적을 달성하기 위해 새로운 연구방법을 제시한다. 이것은 데이터

기반의 도메인 지식을 대상으로, 베이지안 망을 이용하여 온톨로지를 자동적으로 구축할 수 있는 방법이다. 베이지안 망은 다수의 변수들간의 확률적 관계를 표현하는 그래프 모델이고, 온톨로지는 개념들과 그들의 관계에 의해서 지식을 표현하는 도구로서, 본 연구에서는 베이지안 망이 갖고 있는 데이터 분석에서의 장점과 온톨로지와의 관련성을 이용하여 온톨로지를 구축한다. 즉, 베이지안 망을 이루는 노드와 그 확률적 인과관계를, 온톨로지의 개념(속성)과 그들의 관계로 대응시킴으로써, 도메인 지식을 자동적이고 신속하게 표현할 수 있는 것이다. 또한 도메인 지식이 변화하면 언제든지 베이지안 망을 이용하여 온톨로지를 갱신할 수 있다.

본 연구는 총 다섯 개 절로 구성되어 있다. 서론에 이어 2절에서는 온톨로지와 베이지안 망의 이전 연구로부터 중요한 개념을 알아본다. 3절은 베이지안 망과 온톨로지의 관련성 및 비교를 통해, 베이지안 망을 이용하여 온톨로지를 구축할 수 있는 연구 방법을 제시한다. 4절은 산업경기예측 중 디스플레이 사례에 대해 데이터 기반의 도메인 지식을 대상으로 베이지안 망을 이용한 온톨로지 구축 과정을 설명한다. 마지막으로 5절에서는 본 연구가 기여하는 부분과 향후 연구방향에 대해 알아본다.

2. 문헌 연구

2.1 온톨로지(ontology)

지식을 공유하고 재이용하는 수단으로서 개발된 방법 중의 하나가 바로 온톨로지이다. 온톨로지는 철학적 관점에서 본다면 세상의 어떤 관점을 설명하는 분류체계를 제공하는 것이라 할 수 있다[2].

Bunge[2]의 온톨로지를 Wand and Weber[21, 22]가 정보시스템 분야에 채택한 이후 많은 연구가 이루어져 왔다. 온톨로지에 관한 기존 연구는 다음과 같이 크게 세 가지 분야로 구분할 수 있다. 첫째, 온톨로지 자체에 관한 연구이다. 이는 온톨로지에 대한 기초 연구로서 온톨로지 연구의 초창기부터 오늘날까지 꾸준히 진행되고 있는 분야다. 온톨로지가 무엇이며[16, 20], 온톨로지 구축을 위한 문법을 정의하거나[14], 어떻게 구축할 것인가[12] 등을 통해 온톨로지 구축을 보다 용이하게 하기 위한 연구를 추진하고 있다.

둘째, 온톨로지 구축에 관련된 포괄적이고 통합적인 일련의 방법과 기술에 관한 연구가 다수 진행되었다. 지식프로세스[17], 시멘틱 B2B 통합[5]을 비롯해 온톨로지 구축 틀에 대한 연구가 다수 진행되었다[8, 11]. Chen et al.[4]은 온톨로지를 구축하는데 인공신경망(artificial neural network)과 베이지안 망(Bayesian network)을 기반으로 한 방안을 제안하였다.

마지막으로 온톨로지 응용에 대한 연구이다.

온톨로지로는 기존 지식을 표현하는 사례연구가 중심으로, 기상예측[1], 수질관리모델을 위한 지식경영시스템[3] 등 다양한 분야에서 온톨로지가 이용되고 있다.

온톨로지 연구에 관해서 표 1에 정리하였다.

표 1 온톨로지에 관한 연구 분류

범주	내용	연구자
온톨로지 개념	온톨로지의 본질	[16, 20]
	온톨로지 문법	[14]
	구축도구(Protégé 등)	[12]
온톨로지 구축	Semantic B2B Integration	[5]
	지식 프로세스	[17]
	METHONTOLOGY	[8]
	OntoTrack	[11]
	인공지능망, 베이지안망	[4]
온톨로지 응용	기상예측	[1]
	수질관리모델	[3]

2.2 베이지안 망(Bayesian network)

베이지안 망은 경영과학의 문제 해결, 특히 재무와 마케팅 분야에 유용하게 사용되어온 의사결정지원 방법이다[13, 15]. 베이지안 망은 주어진 의사결정문제의 영역지식을 확률적으로 표현한다. 즉, 주어진 문제를 구성하는 변수들간에 존재하는 확률적 의존관계를 방향을 나타내는 아크(arc)로 나타내고 각 변수들이 갖는 조건부확률을 계산함으로써 문제에 포함된 변수들간의 인과관계를 나타낸다[9]. 따라서 하나의 베이지안 망은 각 노드마다 하나의 조건부 확률표(conditional probability table)를 갖는 하나의 비순환 유향 그래프(direct acyclic graph)인 $G = \langle N, A \rangle$ 로서, 베이지안 망을 B 로 표기하면 $B = \langle N, A, \theta \rangle$ 으로 정의할 수 있다. 이 때 각 노드 $n \in N$ 는 하나의 영역변수를, 각 아크 $a \in A$ 는 두 변수간의 확률적 의존성을 나타내며, θ 는 조건부 확률들의 집합을 나타낸다. 일반적으로, 하나의 베이지안 망은 다른 노드들에 배정된 값들을 기초로 특정 노드가 가질 값에 대한 조건부 확률을 계산하는데 이용할 수 있다. 따라서 하나의 베이지안 망은 한 개체의 다른 속성들의 값이 주어졌을 때 분류 클래스 노드(또는 결과변수)의 사후확률분포(posterior probability distribution)를 구해줌으로써 주어진 의사결정문제의 클래스 노드에 대한 분류함수(classifier)로 이용될 수 있다[13]. 즉 하나의 데이터 집합으로부터 베이지안 망을 학습할 때 베이지안 망의 각 노드는 데이터 집합의 각 속성을, 각 아크는 속성들 간의 의존성을 표현하게 되며, 이렇게 학습된 베이지안 망을 기초로 분류 클래스를 확률적으로 예측할 수 있다. 또한 일반적으로 베이지안 망에서 나타내는 관계는 인과관계가 아니다. 즉, 인지지도(cognitive map)와는

달리 베이지안 망의 아크는 반드시 인과관계를 의미하지 않고, 아크(또는 아크가 없는 것)는 조건부 독립 가정을 나타낸다.

한편, 베이지안 망은 많은 장점을 갖고 있다[9]. 첫째, 모든 변수들간의 의존관계(dependency)를 표현하기 때문에 데이터 영역에 대한 지식을 기술할 수 있다. 둘째, 변수들간의 인과관계(causal relationship)를 학습하는데 사용될 수 있기 때문에 응용분야에 대한 이해를 도울 수 있다. 셋째, 모델 자체가 원인(causality)과 확률적 의미(probabilistic semantic)를 표현하고 있기 때문에 사전지식(prior knowledge)과 학습 데이터를 결합하는데 적합하다. 넷째, 주어진 결합확률분포로부터 필요한 조건부확률을 추론함으로써 새로운 예제에 대한 예측을 할 수 있다.

2.3 베이지안 망과 온톨로지

베이지안 망은 비순환 유향 그래프 구조로 이루어져 있는데, 이 그래프의 노드는 응용 도메인에서 변수를 나타낸다. 온톨로지는 개념들과 그들의 관계에 의해서 지식을 표현한다. 두 가지 정의를 정리하면, 온톨로지에서의 개념은, 베이지안 망 그래프에서 노드로 나타내면서 특정 값을 가질 수 있고 연관 확률분포를 가지는, 도메인 변수를 의미한다[6]. 따라서 온톨로지의 개념과 베이지안 망에서의 변수를 표현하는데 개념, 변수 그리고 노드를 상호 교환적인 용어로 이용할 수 있다. 또한, 베이지안 망에서 서로 아크들로 연결된 노드의 집합을 온톨로지에서도 하나의 클래스로 볼 수 있다. 이 때 집합을 이루는 노드들은 한 클래스를 구성하는 속성으로 간주할 수 있다. 따라서 베이지안 망을 이루는 노드와 그 확률적 인과관계를 온톨로지의 개념(속성)과 그들의 관계로 대응시킴으로써 도메인 지식을 자동적이고 신속하게 표현할 수 있다. 표 2는 베이지안 망과 온톨로지의 관계를 비교한 것이다.

표 2 베이지안 망과 온톨로지의 비교

	베이지안 망	온톨로지
목적	불확실하거나 불완전한 특정 도메인의 전문가 지식을 표현	특정 도메인 지식을 공유하기 위해 공통의 개념으로 표현
노드	변수	개념(속성)
아크 (방향)	조건부 독립 (확률적 의존관계)	개념들의 관계 (상하 위계관계, 유사 성질)

온톨로지를 구축하는데 베이지안 망을 이용하는 방법에 대한 이전 연구로는 Chen et al.[4] 연구가 있다. 그들은 온톨로지를 구축하는데 인공지능 망과 베이지안 망을 기반으로 하고 있다. 이외 베이지안

망을 이용한 온톨로지 구축에 대한 연구는 거의 없는 실정이고, 관련 연구로서 베이지안 망을 이용한 온톨로지를 사상(mapping)하는 연구가 다수 진행되었[19].

3. 연구 방법

전통적인 온톨로지 구축 방법은 경험과 지식을 바탕으로 한 도메인 전문가에 의존한다. 하지만 이것은, 서론에서 밝혔듯이, 많은 시간과 비용이 소요된다. 본 연구에서 제시한 방법은 전문가 의존에서 벗어나 데이터 기반의 도메인 지식을 대상으로, 베이지안 망을 이용하여 온톨로지를 자동적으로 구축할 수 있는 방법이다(그림 1 참조). 즉, 베이지안 망이 갖고 있는 데이터 분석에서의 장점과 온톨로지와의 관련성을 이용하여 온톨로지를 구축한다. 또한 도메인 지식이 변화하면 언제든지 베이지안 망을 이용하여 온톨로지를 갱신할 수 있다.

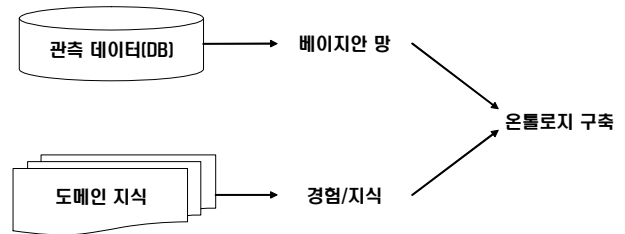


그림 1 온톨로지 구축 방법

본 연구의 분석방법은 네 가지 단계로 구성된다(그림 2 참조). 첫 번째 단계는 데이터 기반의 도메인 지식을 대상으로 하는 것이다. 온톨로지 구축 과정에서 도메인 전문가의 지식이나 경험을 최소화하여 시간과 비용을 줄이기 위해 정량적 데이터를 대상으로 하는데, 여기에는 객관적으로 이산화(discretization) 할 수 있는 데이터도 포함된다. 두 번째 단계는 도메인 지식을 베이지안 망을 통해 분석하는 것이다. 도메인의 데이터로부터 베이지안 망의 그래프, 즉 모델의 구조를 학습하는 것으로 도메인을 구성하는 변수들간에 존재하는 확률적 의존관계를 나타낸다. 세 번째 단계는 베이지안 망의 분석 결과를 온톨로지로 전환하는 것이다. 즉, 베이지안 망 분석을 통해 나타난 변수들의 구조와 관계를 기초로 온톨로지의 개념과 관계를 설정하는 것이다. 베이지안 망에서 서로 아크들로 연결된 노드의 집합을 온톨로지에서도 하나의 개념으로 볼 수 있다. 이 때 집합을 이루는 노드들은 그 개념을 구성하는 속성으로 볼 수 있다. 마지막 단계는 설정된 개념과 관계를 바탕으로 온톨로지를 구축하는 것이다. 클래스와 이를 이루는 속성, 그리고 이들간의 관계를 바탕으로 하나의 온톨로지를 완성한다.

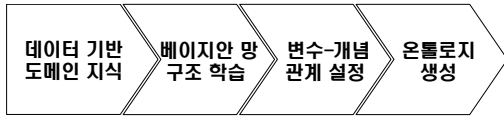


그림 2 분석방법

본 연구에서는 베이지안 망의 모델링을 위해 Hugin 툴[23]을 이용한다. 데이터로부터 베이지안 망의 그래프, 즉 구조를 학습하고 확률적 의존관계를 학습한다. 베이지안 망에 대한 기초와 Hugin 분석방법에 대한 자세한 설명은 Kjærulff and Madsen[10]를 참조하였다.

4. 사례연구

본 절에서는 베이지안 망을 이용한 온톨로지 구축 방법을 평판TV (flat panel) TV 경기예측 사례를 통해 알아본다. 기존 TV에 비해 두께가 얇고 선명하며 화면 사이즈의 대형화에 유리한 평판TV(LCD TV와 PDP TV)는 2005년을 전후하여 시장이 급속히 성장하면서 대중적인 제품으로 부상하고 있다. 세계 평판TV 수요는 2004년 1억 2천만대에서 2005년 2억 7천만대, 그리고 2006년 5억 5천만대로 연간 100% 이상 성장하였다[7]. 이에 따라 CRT TV와 프로젝션 TV를 포함한 전체 TV 시장에서 평판TV 수요가 차지하는 비중은 2004년 6%에서 2006년 29%로 급격히 증가하고 있고, 2008년경 전체 TV 시장의 반 이상이 평판TV로 대체될 것으로 예상되고 있다. LCD와 PDP는 고성장 분야인 평판TV 시장에서 경쟁 관계에 있는 제품으로서 경쟁 정도가 심화되고 있다. 평판TV 수요 중 가장 큰 비중을 차지하고 있는 30인치급 시장에서 LCD가 우위를 보이고 있지만, 시장이 급속히 확대되고 있는 40인치 이상 대형 TV 시장에서 치열한 시장 경쟁이 벌어지고 있다(그림 3 참조).¹

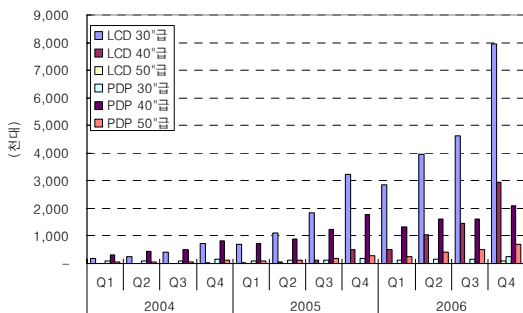


그림 3 평판TV 시장[7]

¹ 조사기관인 Display Search[7]에 의하면 2006년 평판 TV 수요 중 비중이 가장 큰 사이즈는 30인치급으로 36%를 차지하였다. 40인치 이상 대형TV도 2005년 22%, 2006년 26%로 수요비중이 크게 증가하고 있다. 사이즈대별 주요 제품은 30인치급은 32·37인치, 40인치급은 40·42·46·47인치, 50인치급은 50·52인치이다.

평판TV 시장이 본격적인 성장궤도에 진입하고 있는 가장 큰 이유는, 가격이 예상보다 빨리 하락하여 기존 TV를 대체하는 속도가 빨라졌기 때문이다. 최근 평판TV 시장에서 가장 큰 관심을 모으고 있는 40인치급의 경우, PDP TV는 2004년 초에 3,800달러 수준이었으나, 2006년 말에는 1/3 수준인 1,300달러까지 떨어졌다. 경쟁 제품인 LCD TV의 가격 하락폭은 더 커서, 같은 기간 9,300달러에서 1,800달러 수준까지 가격이 급락하였다. 40인치급 제품 뿐만 아니라 30인치급은 물론 50인치급 제품까지 평판TV의 가격이 하락하여 수요를 확대시키고 있다. 결국 평판TV의 기종별(LCD, PDP), 사이즈별(30인치급, 40인치급, 50인치급) 수요 및 가격이 서로 영향을 주고 받으며 경기를 좌우할 것이다.²

본 사례에서는 데이터 기반 변수들을 대상으로 평판TV 시장에 대해 경쟁의 관점에서 온톨로지를 구축하고자 한다. 이를 위해 평판TV 시장을 분석할 때 가장 관심을 갖게 되는 기종별, 사이즈별 수요 및 가격 등 12개 변수를 대상으로 한다³. 데이터는 평판TV 시장의 확대기인 2003년 1/4분기부터 2007년 1/4분기까지의 분기별 자료이다. 표 3에 사례연구에 사용한 관련 변수들과 기간별 값을 나타냈다.

표 3 변수들의 값[7]

(단위 : 수요는 천 대, 가격은 달러)

	'03.1Q	'03.2Q	...	'07.1Q
PDP 30"급 수요	54	70	...	187
PDP 40"급 수요	105	120	...	1,505
PDP 50"급 수요	35	44	...	627
LCD 30"급 수요	52	54	...	6,518
LCD 40"급 수요	-	-	...	2,687
LCD 50"급 수요	-	-	...	116
PDP 30"급 가격	4,617	4,528	...	1,173
PDP 40"급 가격	4,494	4,178	...	1,287
PDP 50"급 가격	10,621	10,003	...	2,340
LCD 30"급 가격	5,007	4,930	...	926
LCD 40"급 가격	9,738	9,738	...	1,708
LCD 50"급 가격	-	-	...	3,848

표 3의 데이터를 대상으로 베이지안 망을 이용해 변수들의 확률적 관계를 분석하였다. 그 결과는 그림 4의 위 그림과 같은데, 12개 변수 중, 10개 변수는 다른 변수들과의 관계가 나타났고, 나머지 2개 변수는 독립적으로 나타났다. 다른 변수와 관계를 갖는 7개 변수들도 각각 2~3개씩의 집합으로 나뉘어졌는데, 각 집합은 온톨로지의 개념이고

² 본 연구는 수요측면의 경기를 보는 것으로 공급측면(TV 제조업체의 시장 공급)은 제외하였다.

³ 분석 대상인 12개 변수는 디스플레이 산업 전문가와의 인터뷰, 관련 문헌 분석을 통해 선정하였다.

집합의 변수들은 해당 개념을 구성하는 속성들이 된다. 따라서 베이지안 망 분석의 결과를 온톨로지로 표현하면 그림 4의 아래 그림을 얻게 된다.

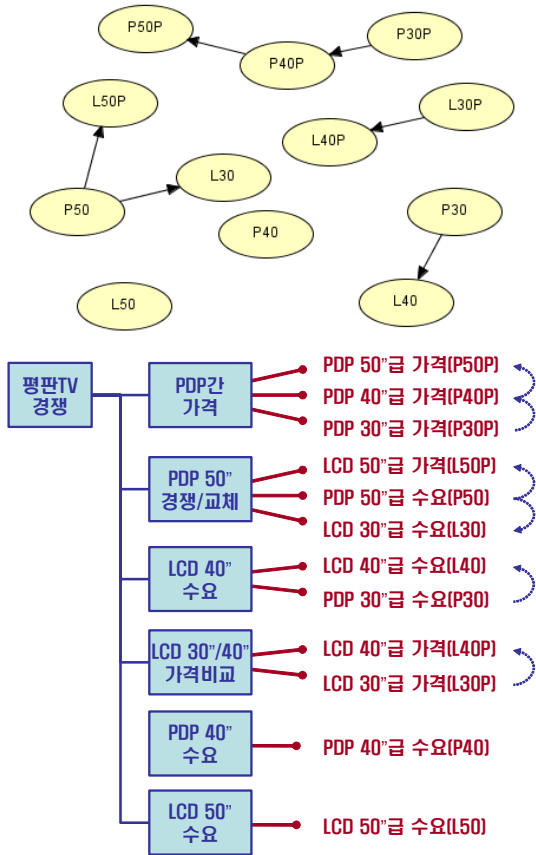


그림 4 베이지안 망 결과와 관련 온톨로지

평판TV 시장 온톨로지의 클래스는 최상위 클래스인 ‘평판TV 경쟁’을 제외하면 6개이다. 베이지안 망 분석에서 2~3개 변수로 이루어진 4개 집합과 독립적으로 나타난 2개 변수가 온톨로지의 클래스를 이룬다. 그런데 집합으로 이루어진 4개 클래스의 속성들은 새로운 정의가 필요하다. 이에 따라 새롭게 정의된 클래스 개념을 각각 ‘PDP간 가격’, ‘PDP 50인치 경쟁/교체’, ‘LCD 40인치 수요’, ‘LCD 30인치/40인치 가격비교’로 하였다(그림 4 아래 그림 참조). 새롭게 구성된 클래스에 대해 설명하면 다음과 같다.

(1) ‘PDP간 가격’ : PDP TV의 가격은 사이즈별로 상대적인 가격 수준에 따라 결정된다고 할 수 있다. 경쟁제품인 LCD TV의 가격에도 영향을 받지만, 우선적으로 PDP 동종 기종 내에서 타 사이즈대 가격 수준에 따라 해당 사이즈대의 가격이 영향을 받는다. 만약 PDP 50인치급 가격을 LCD 가격에 대응하려면 한 없이 내려야 하지만, 한 단계 작은 사이즈 제품인 PDP 40인치급 가격을 고려하여 결정해야 한다.

(2) ‘PDP 50인치 경쟁/교체’ : PDP 50인치급 수요에 가장 큰 영향을 미치는 것은 경쟁제품인 LCD 50인치급 제품의 수요를 결정하는 가격이다. 또한 LCD 30인치급 제품(37인치)을 사용했던 구매자에 의해 교체 수요로서 고려될 수 있는 제품이라고 할 수 있다. 이는 LCD 37인치의 교체수요로서 LCD 및 PDP 40인치급 제품은 사이즈상 크기에 큰 차이가 없어, 두 단계 큰 사이즈대인 PDP 50인치를 선호할 것으로 보이기 때문이다.

(3) ‘LCD 40인치 수요’ : LCD 40인치급 수요는 평판TV 초기 도입 제품인 PDP 30인치급 제품의 교체수요로서 고려될 수 있다.

(4) ‘LCD 30인치/40인치 가격비교’ : LCD 30인치급과 40인치급은 현재 수요 주력제품으로서 상대적인 가격 설정에 따라 경쟁제품인 PDP TV와의 경쟁에 우위를 둘 수 있고, LCD 30·40인치 간의 시장 잠식을 피할 수 있다.

이상과 같이 새롭게 구축된 온톨로지는 평판TV 시장을 기존 방법과 같이 정적인 측면에서 본 것이 아니고, 기종 및 사이즈대간의 동적인 경쟁상황을 나타낸다고 할 수 있다. 즉, 6개 클래스와 구성 속성들이 현재의 평판TV 경쟁 상황을 축약적으로 나타낼 수 있다고 할 수 있다.

4명의 디스플레이 산업 전문가들로 하여금 위에서 구축된 온톨로지에 대한 구조적 타당성 검증에 의뢰하였다. 의뢰방법은 1차적으로는 온톨로지를 이메일로 송부하여 구조적 타당성을 5점 척도로 확인하였다. 그 결과 평균 4 이상의 만족도를 얻을 수가 있었다.

5. 결론 및 향후 연구방향

온톨로지를 구축하는 대부분의 방법은 체계적이거나 자동적이지 못하다. 전통적인 온톨로지 구축은 도메인 전문가에 의존하지만, 이것은 시간과 비용이 소요되고 논쟁의 대상이 된다. 많은 온톨로지 구축틀이 있지만 온톨로지 구축은 아직 인간의 노력이 필요하고, 또한 변화하는 도메인 지식을 온톨로지에 신속하게 반영하는 것은 어려운 일이다. 본 연구의 목적은 도메인 전문가의 지식이나 경험을 최소화하면서 자동적으로 도메인 지식을 얻을 수 있는 방법을 제시하는 것이다. 이를 위해 전문가 의존에서 벗어나 베이지안 망을 이용하여 온톨로지를 자동적으로 구축할 수 있는 방법을 제시하였다. 또한 도메인 지식이 변화하면 언제든지 베이지안 망을 이용하여 온톨로지를 갱신할 수 있다.

평판TV 경기예측 사례에 대해 온톨로지를 구축한 결과, 도메인 전문가의 지식이나 경험을 최소화하면서 자동적으로 도메인 지식을 얻을 수 있었다. 특히, 구축된 온톨로지는 기존 지식 체계에서 탈피하여 새로운 시각에서 도메인을 이해할 수 있는 기회를

제공한다.

본 연구의 공헌점은 다음과 같다. 첫째, 온톨로지 구축 과정에서 지식을 획득하는데 시간과 비용이 많이 소요되는 한계를 보완할 수 있는 방법을 제시하였다. 둘째, 변화하는 도메인 지식을 비교적 신속하게 온톨로지에 반영할 수 있다. 반면, 본 연구의 한계로는 첫째, 성과 및 비용 측면에서 새로운 연구방법과 기존 전문가 또는 직관적 방법과의 통계적 비교 평가가 미진하였다. 둘째, 도메인 변수가 데이터를 구하기 어렵거나 정성적 정보일 경우 이를 베이지안 망에 직접 적용하기 위해서는 전문가 지식 등에 의존해서 정량적 정보로 변환해야 한다. 셋째, 베이지안 망의 구조 학습과정에서 변수들의 이산화(discretization) 방법, 학습 알고리즘, 유의수준 등에 따라 결과가 다르게 나올 수 있다.

향후 연구방향으로는 첫째, 기존의 직관적 및 전문가 지식에 의한 온톨로지 구축과 데이터 기반 온톨로지 구축 방법을 결합한 방법에 대한 연구가 필요하다. 이는 비용, 노력 및 주관적 개입과 정성적 지식의 해석 등 양 방법의 단점을 서로 해결할 수 있기 때문에 의미 있는 연구가 될 것이다. 둘째, 베이지안 망에 의해 구축된 온톨로지의 클래스, 속성 등의 상호 연관성을 이용하여 문제 해결을 위한 접근방법에 대한 연구도 가능할 것이다.

참고문헌

- [1] Bally, J. Boneh, T., Nicholson, A.E. and Korb, K. (2004). "Developing An Ontology for the Meteorological Forecasting Process," *Decision Support in an Uncertainty and Complex World: The IFIP TC8/WG8.3 International Conference 2004*, pp.70-81.
- [2] Bunge, M. (1977). *Treatise on Basic Philosophy: Volume 3: Ontology I: The Furniture of the World*, Reidel, Dordrecht.
- [3] Chau, K.W. (2007). "An ontology-based knowledge management system for flow and water quality modeling," *Advances in Engineering Software*, Vol.38, pp.172-181.
- [4] Chen, R.C., Chuang, C.H., and Tseng C.C. (2006). "Constructing an Ontology Automatically by Projective ART Neural Network," *International Computer Symposium ICS 2006*, Taipei, Taiwan.
- [5] Cui, Z., Jones, D. and O'Brien, P. (2002), "Semantic B2B Integration: Issues in Ontology-based Approach," *SIGMOD Record*, Vol.31, No.1, pp.43-48.
- [6] Devitt, A, Danev, B. and Matusikova, K. (2006), "Constructing Bayesian Networks Automatically using Ontologies," *Proceedings of Second Workshop on Formal Ontologies Meets Industry*, Trento, Italy.
- [7] Display Search (2007). *Quarterly Global TV Shipment and Forecast Report*.
- [8] Fernandez, M., Gomez-Perez, A. and Juristo, N. (1997). "METHONTOLOGY: From Ontological Art Towards Ontological Engineering," presented to AAAI97, *Workshop on Ontological Eng.*, Stanford, pp.33-40.
- [9] Heckerman, D. (1995). "A Tutorial on Learning with Bayesian Networks," *Technical Report MSR-TR-95-06*, Microsoft Research, Redmond, WA.
- [10] Kjørulff, U.B. and Madsen, A.L. (2006). *Probabilistic Networks for Practitioners - A Guide to Construction and Analysis of Bayesian Networks and Influence Diagrams*, 272pages, Unpublished.
- [11] Liebig, T. and Noppens, O. (2005). "ONTOTRACK: A semantic approach for ontology authoring," *Web Semantics, Science, Services and Agents on the World Wide Web*, Vol.3, pp.116-131.
- [12] Noy, N.F. and McGuinness, D.L. (2001). "Ontology Development 101: A Guide to Creating Your First Ontology," Retrieved from http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html.
- [13] Pearl, J. (1988), *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, San Francisco.
- [14] Ram, S. and Park, J. (2004). "Semantic Conflict Resolution Ontology(SCROL): An Ontology for Detecting and Resolving Data and Schema-Level Semantic Conflicts", *IEEE Transactions on Knowledge and Data Engineering*, Vol.16, No.2, pp.189-202.
- [15] Sarkar, S. and Sriram, R.S. (2001). "Bayesian models for early warning of bank failures," *Management Science*, Vol.47, No.11, 1457-1475.
- [16] Shanks, G., Tansley, E. and Weber, R. (2004). "Representing Composites in Conceptual Modeling," *Communications of the ACM*, Vol.47, No.7, pp.77-80.
- [17] Staab, S., Schnurr, H.P., Studer, R. and Sure, Y. (2001). "Knowledge processes and ontologies," *IEEE Intelligent Systems*, Vol.16, No.1, pp.26-34.
- [18] Sugumaran, V. and Storey, V.C. (2006). "The role of domain ontologies in database design: An ontology management and conceptual modeling environment", *Source ACM Transactions on Database Systems*, Vol.31, No.3, pp.1064 - 1094.
- [19] Svab, O. and Svatek, V. (2006). "Combining Ontology Mapping Methods Using Bayesian Networks," *International Workshop on Ontology Matching collocated with the 5th International Semantic Web Conference (ISWC-2006)*.
- [20] Wand, Y., Storey, V. C. and Weber, R. (1999). "An Ontological Analysis of the Relationship Construct in Conceptual Modeling," *ACM Transaction on Database System*, Vol.24, No.4, pp.494-528.
- [21] Wand, Y. and Weber, R. (1990). *MARIO BUNGE's ontology as a formal foundation for information systems concepts*, in Weingartner, P. and Dorn, G.J.W.(Eds.), *Studies on MARIO BUNGE's Treatise* (pp.123-149), Amsterdam, Rodopi.
- [22] Wand, Y. and Weber, R. (1995). "On the deep structure of information systems," *Information Systems Journal*, Vol.5, No.3, pp.203-223.
- [23] <http://www.hugin.com>