

패턴분류를 위한 온톨로지 기반 퍼지 분류기

Ontology-based Fuzzy Classifier for Pattern Classification

이인근 · 손창식 · 권순학[†]

In K. Lee, Chang S. Son and Soon H. Kwon[†]

영남대학교 전기공학과

Dept. of Electrical Engineering, Yeungnam University

요 약

최근, 패턴분류에 온톨로지를 이용하려는 연구가 다양한 분야에서 시도되고 있다. 그러나 대부분의 이러한 연구에서는 패턴분류 관련 지식을 표현한 온톨로지가 패턴분류 과정에서 단순히 참조되는 수준에 머물고 있다. 본 논문에서는 퍼지 규칙기반 분류기를 확장한 온톨로지 기반 퍼지 분류기를 제안한다. 이를 위해 퍼지규칙 기반 패턴분류 방법을 개념화하여 온톨로지를 구성하고, 패턴분류를 위한 온톨로지 추론 규칙을 생성한다. 그리고 IRIS 데이터집합의 패턴분류 실험을 통해 온톨로지 기반 퍼지 분류기의 타당성을 보인다.

키워드 : 온톨로지, 온톨로지 추론, 패턴분류, 퍼지 분류기

Abstract

Recently, researches on ontology-based pattern classification have been tried out in many fields. However, in most of the researches, the ontology which represents the knowledge about pattern classification is just referred during the processes of the pattern classification. In this paper, we propose ontology-based fuzzy classifier for pattern classification which is extended from the fuzzy rule-based classifier. In order to realize the proposed classifier, we construct an ontology by conceptualizing the method of fuzzy rule-based pattern classification and generate ontology inference rules for pattern classification. Lastly, we show the validity of the proposed classifier through the experiment of pattern classification on the Fisher's IRIS dataset.

Key Words : ontology, ontological inference, pattern classification, fuzzy classifier

1. 서 론

온톨로지는 지식의 표현과 확장이 자유롭고, 추론과 검색을 통해 쉽게 활용할 수 있는 장점이 있다. 이런 온톨로지의 장점으로 인해 최근 영상처리[1,2], 패턴분류[3,4], 로봇[5,6], 의료[7] 등 다양한 분야에서 온톨로지를 활용하기 위한 연구가 진행되고 있다. 이들 연구에서는 관련 분야의 지식을 계층화된 개념으로 표현함으로써 온톨로지를 구축하고, 온톨로지를 추론을 통해 지식을 확장한다. 또한 온톨로지 표현된 체계화된 지식을 이용하여 각 분야의 시스템의 성능을 향상시키려 노력하고 있다. 이에 발맞추어 퍼지이론, 신경회로망, 유전자 알고리즘 등을 주로 사용하던 패턴분류[8-10]에 대하여 최근 온톨로지를 적용하려는 연구가 시도되고 있다[1-4]. 그러나 현재 패턴분류 분야에서의 온톨로지 활용에 관한 연구에서는 개념의 계층을 나타내는 온톨로지는 구현하였으나, 추론을 통한 온톨로지의 활용 방법이나 온톨로지 추론을 위한 온톨로지 모델은 제시하지 못하였다. 온톨로지가 지식 표현의 자유성과 확장성이 용이하다는 장점에 반해 온톨로지 추론 과정은 복잡하다. 온톨로지 추론 규칙은 '조건부→결론부'의 형태로 구성되는데, 온톨로지

추론은 온톨로지 규칙의 조건부를 구성하는 서술문(1)들을 검색하고, 조건부가 참일 경우 결론부에 선언한 사건들을 새롭게 생성하거나 특정 명령을 수행하는 과정이다. 즉, 수치 기반의 함수를 논리 기반의 온톨로지 추론 방법으로 표현하는 것이 쉽지 않기 때문에 패턴분류와 같이 복잡한 수치적 계산이 필요한 분야에서는 온톨로지의 활용이 제한적일 수밖에 없다. 또한, 온톨로지 규칙이 복잡할수록 조건부의 서술문에 대한 온톨로지 검색 시간이 길어지는 문제가 있다. 따라서 온톨로지 구조의 복잡성, 추론 규칙 생성의 용이성, 응용 시스템의 성능 등을 고려하여 패턴분류를 위한 온톨로지가 설계되어야 한다.

본 논문에서는 참고문헌 [11]에서 제안한 퍼지 분류기를 온톨로지 확장하여, 패턴분류를 위한 온톨로지 기반 퍼지 분류기를 제안한다. 이를 위해 퍼지규칙 기반 패턴분류 방법을 개념화하여 온톨로지를 구성하고, 패턴분류의 과정을 표현하기 위한 온톨로지 추론 규칙을 생성한다. 그리고 Fisher[12]의 IRIS 데이터 집합의 실제 패턴분류 실험을 통해 제안한 퍼지 분류기의 타당성을 보인다.

1) 온톨로지서 subject(domain)-predicate(property)-range(object)로 표현되는 이진 관계(binary-relation) 또는 그 이상의 관계(n-ary relation)를 서술문(statement)이라 한다. 문, 문장, 선언, 사건이라고도 한다.

접수일자 : 2008년 3월 19일

완료일자 : 2008년 6월 5일

[†] Corresponding author

2. 온톨로지 기반 퍼지 분류기 설계

2.1 IRIS 데이터의 구조

온톨로지 기반 퍼지 분류기의 설계를 위해 Fisher의 IRIS 데이터 집합을 이용한다. IRIS 데이터 집합은 IRIS 꽃을 3가지 종류(setosa, versicolor virginica)의 품종별로 50개씩 선택하고, 각 IRIS 꽃의 꽃잎과 꽃받침의 크기를 'sepal length(SL)', 'sepal width(SW)', 'petal length(PW)', 'petal width(PW)'와 같이 4가지의 속성으로 나타낸 것이다. 표 1은 IRIS 데이터의 구조를 보인다.

표 1. IRIS 데이터 집합의 구조
Table 1. Structure of IRIS dataset

번호	속성				분류
	SL	SW	PL	PW	
1	5.1	3.5	1.4	0.2	setosa
~50	...				
51	7.0	3.2	4.7	1.4	versicolor
~100	...				
101	5.9	3.0	5.1	1.8	virginica
~150	...				

IRIS 데이터 집합의 구조를 온톨로지로 표현하기 위해서는 IRIS 데이터의 속성과 분류를 개념화할 필요가 있다. 즉 그림 1과 같이 의미 네트워크(semantic network)를 이용하여 IRIS 데이터를 개념화하여 그 관계를 표현할 수 있다.

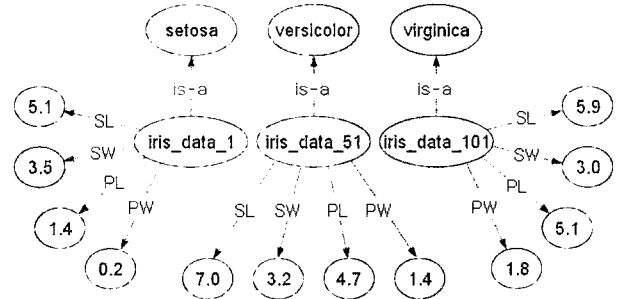


그림 1. IRIS 데이터 집합의 의미 네트워크
Fig. 1. Semantic network for IRIS dataset

IRIS 데이터 집합의 패턴분류 결과는 그림 1과 같이 각각의 IRIS 데이터가 특정 품종 클래스에 포함되는 형태가 되어야 한다는 것을 알 수 있다. 즉, 품종 정보가 없는 IRIS 데이터 집합의 경우에는 4가지 속성에 기반하여 각 데이터가 특정 품종 클래스에 속하도록 함으로써 패턴분류가 이루어진다. 따라서 온톨로지를 이용한 패턴분류에서도 분류 결과가 그림 1과 같은 형태가 되도록 온톨로지를 설계한다.

2.2 온톨로지 기반 퍼지 분류기

품종 정보가 없는 IRIS 데이터 집합의 패턴분류를 위해서 패턴분류 되어야 하는 IRIS 데이터를 그림 2와 같이 표현한다. 그림 2에서 'IrisData'는 IRIS 데이터 집합을 개념화한 것으로서 'sepal_length', 'sepal_width', 'petal_length', 'petal_width'를 속성으로 가지며 'xsd:double'과 같이 수치 형태의 속성 값을 가진다. 그리고 IRIS 데이터는

'iris_data_1', ..., 'iris_data_150'과 같이 개념 'IrisData'의 개체로 표현되며, 상위 개념인 'IrisData'의 속성과 속성 값의 형태에 따라 각 데이터의 속성을 표현한다. 그림 2의 우측 하단은 온톨로지를 표현하기 위한 요소로서 개념(concept), 개체(instance), 숫자나 문자와 같은 값(value), 그리고 개념이나 개체 사이의 관계 설정을 위한 속성(property)의 기호를 나타낸다.

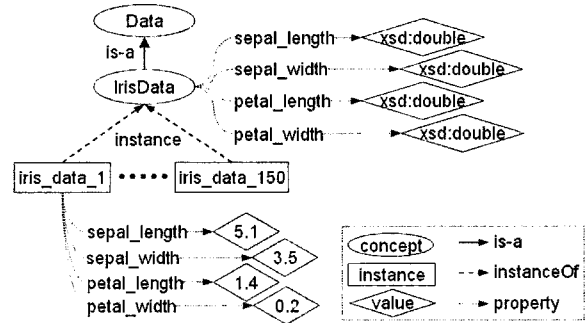


그림 2. IRIS 데이터 표현
Fig. 2. Representation of IRIS data

OWL(Web Ontology Language)[13]에서는 속성을 이용한 관계 표현을 'property(domain, range)'와 같이 이진 관계의 형태로 나타낸다. 여기서 'domain'은 개념이나 개체가 되고, 'range'는 개념이나 개체, 또는 'integer, double, boolean, string'과 같은 데이터 형태의 값이 될 수 있다. 따라서 온톨로지에서는 속성은 'range'의 형태에 따라 'object property'와 'datatype property'로 구분된다. 즉, 'object property'는 속성의 'domain'과 'range'가 모두 개념이나 개체가 되며, 'datatype property'는 속성의 'domain'은 개념이나 개체, 'range'는 데이터 형태가 되는 것이다. 그림 3과 표 2는 각각 온톨로지의 속성 계층과 정의를 보인다. 여기서 'active_linguistic_term'은 'object property'이고, 나머지 속성은 'xsd:double' 형태의 데이터 형태를 갖는 'datatype property'이다. 여기서 'xsd:'는 XML Schema Definition[14]으로 W3C에서 정의한 데이터 형태를 의미하는 어미(prefix)이다.

온톨로지 기반 퍼지 분류기를 구현하기 위해서는 퍼지 소속 함수를 개념화 하여야 한다. 또한 이 개념을 기존의 패턴 분류에 사용되는 함수와 같이 온톨로지 추론 과정에서 직접 활용할 수 있도록 설계되어야 한다. 따라서 그림 4와 같이 퍼지 소속 함수를 개념화하여 온톨로지로 구성한다.

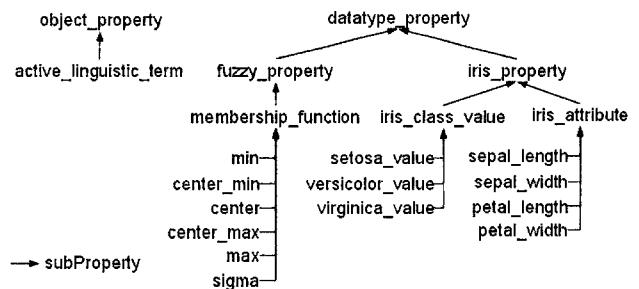


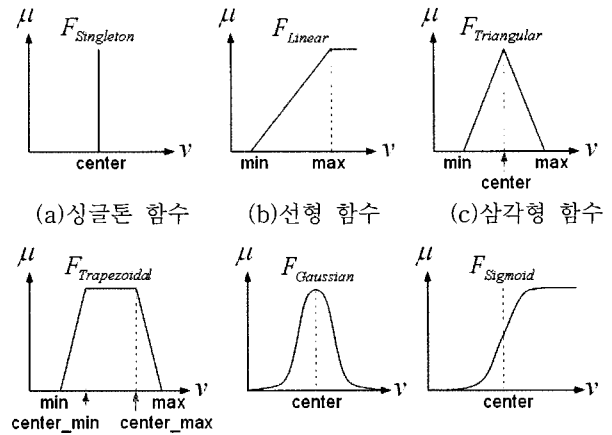
그림 3. 속성 계층
Fig. 3. Hierarchy of properties

표 2. 속성 정의

Table 2. Definition of properties

Property Name	Domain	Range/Datatype
active_linguistic_term	Data	FuzzyAttribute
min	FuzzyMembership Function	xsd:double
center_min	FuzzyMembership Function	xsd:double
center	FuzzyMembership Function	xsd:double
center_max	FuzzyMembership Function	xsd:double
max	FuzzyMembership Function	xsd:double
sigma	FuzzyMembership Function	xsd:double
setosa_value	Data	xsd:double
versicolor_value	Data	xsd:double
virginica_value	Data	xsd:double
sepal_length	Data	xsd:double
sepal_width	Data	xsd:double
petal_length	Data	xsd:double
petal_width	Data	xsd:double

함수의 형태를 복원할 수 있다. 또한, 그림 5(e)의 '가우시안 함수'나 그림 5(f)가 '시그모이드 함수'의 경우 각 함수의 기준 값을 'center', 'sigma'로 표현하여 식 (2), 식(3)과 같이 퍼지 소속 함수의 형태를 복원한다.



(a)싱글톤 함수 (b)선형 함수 (c)삼각형 함수
(d)사다리꼴 함수 (e)가우시안 함수 (f)시그모이드 함수

그림 5. 퍼지 소속 함수의 종류

Fig. 5. Some cases of fuzzy membership functions

$$f_{Triangular}(v) = \begin{cases} \max\left(0, \frac{(v - \min)}{|\text{center} - \min|}\right), & v \leq \text{center} \\ \max\left(0, \frac{-(v - \max)}{|\text{center} - \max|}\right), & v > \text{center} \end{cases} \quad (1)$$

$$f_{Gaussian}(v) = e^{-\frac{(v - \text{center})^2}{2(\text{sigma})^2}} \quad (2)$$

$$f_{Sigmoid}(v) = \frac{1}{1 + e^{(\text{sigma}) \cdot (v - \text{center})}} \quad (3)$$

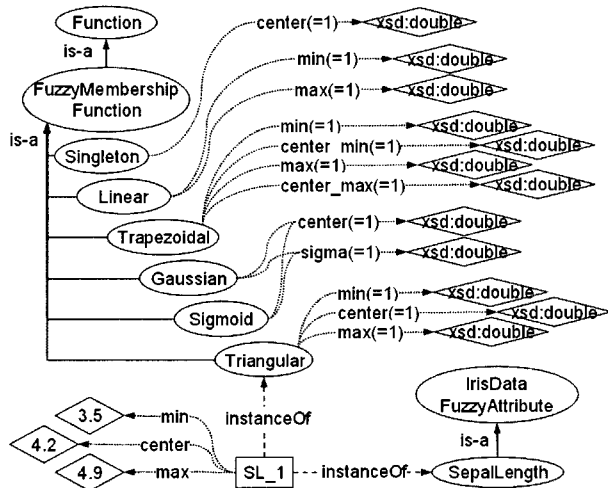


그림 4. 퍼지 소속 함수 정의

Fig. 4. Description of fuzzy membership functions

퍼지 소속 함수의 언어항을 그림 4의 'SL_1'과 같이 개체로 정의할 수는 있으나, 퍼지 소속 함수의 다양한 형태를 개념화 하기는 쉽지 않다. 따라서 퍼지 소속 함수를 정의하기 위해 일반적으로 많이 사용되는 '싱글톤(Singleton)', '선형(Linear)', '삼각형(Triangular)', '사다리꼴(Trapezoidal)', '가우시안(Gaussian)', '시그모이드(Sigmoid)'와 같은 함수의 형태를 개념화하고 이들을 'FuzzyMembershipFunction'의 하위 개념으로 표현한다. 또한, 퍼지 소속 함수를 표현하기 위한 최소의 기준 값들을 이용하여 이들의 제약조건을 형성한다. 그림 4에서 '(=1)'은 속성에 대한 제약조건으로서 각 함수 형태를 정의할 때 속성 값을 반드시 한번만 지정해야 한다는 것을 뜻한다. 예를 들면, 그림 5(c)와 같은 '삼각형 함수'를 표현하기 위해서는 퍼지 소속 함수의 속성 중 'min', 'center', 'max'가 반드시 설정되어야 하며, 설정된 퍼지 소속 함수의 속성 값과 식(1)을 이용하여 퍼지 소속

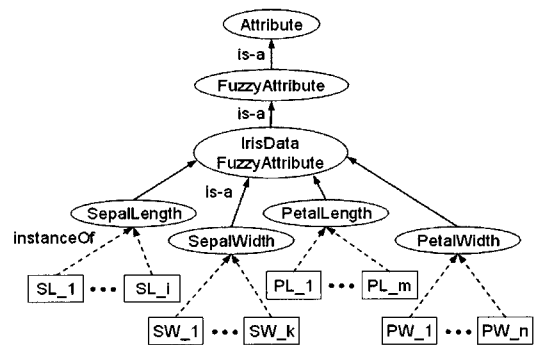


그림 6. 속성별 언어항

Fig. 6. Linguistic terms at each attribute

그림 6은 IRIS 데이터 집합의 각 속성별로 정의된 퍼지 소속 함수의 언어항을 나타낸다. IRIS 데이터 집합의 속성 중 'sepal_length'의 퍼지 소속 함수를 표현하기 위해 개념 'SepalLength'를 정의하였고, 퍼지 소속 함수의 언어항으로 개체 'SL_1'을 'SepalLength'의 개체로 설정하였다. 예를 들어, 그림 4와 그림 6에서 개체 'SL_1'은 IRIS 데이터 집합의 속성 중에서 'SepalLength'의 소속 함수들 중 하나의 언어항을 나타내며, 함수의 형태는 'Triangular'이고, (v, μ) 로 표현되는 2차원 공간에서 $(3.5, 0)$, $(4.2, 1)$, $(4.9, 0)$ 을 꼭지

점으로 하는 삼각형 모양의 함수를 의미한다. 그리고 언어 변수 'v'에 대한 소속 함수 값(μ)은 식 (1)로부터 구한다.

3. 패턴분류를 위한 온톨로지 추론

2장에서 설계한 온톨로지에서 개념 'IrisData'의 개체로 등록되어 있는 분류되지 않은 IRIS 데이터를 온톨로지 추론을 통해 품종을 결정한다. 3.1절에서는 온톨로지 추론 과정을 단순화하기 위해 IRIS 데이터의 속성별로 활성화된 언어항을 선택하고, 3.2절에서는 참고문헌 [11]에서 사용한 퍼지규칙을 온톨로지 추론 규칙으로 변환하여 IRIS 데이터의 분류를 위한 적합도를 결정하며, 3.3절에서는 IRIS 데이터의 분류 적합도에 따라 IRIS 데이터 객체의 IRIS 품종을 결정하는 온톨로지 추론 과정을 보인다. 여기서 분류 적합도는 퍼지 추론에서 규칙의 이행 정도를 의미한다.

3.1 속성에 대한 언어항 선택

참고문헌 [11]과 같이 퍼지 분류기를 설계할 때 퍼지 추론의 연산량을 줄이기 위해 규칙의 이행 정도(degree of fulfillment)가 '0'인 경우는 고려하지 않는다. 따라서 온톨로지 추론에서도 추론 과정을 단순화하기 위해 'active_linguistic_term'의 속성을 이용하여 IRIS 데이터의 속성별 퍼지 소속 함수 값이 '0' 이상으로 활성화된 경우의 언어항과 IRIS 데이터 객체와의 관계를 설정하고, 분류 적합도를 결정하는데 활용한다. 즉, [Rule-1]의 온톨로지 추론 규칙을 통해 'sepal_length(iris_data_1, 5.1)'의 속성 관계로부터 'active_linguistic_term(iris_data_1, SL_2)'와 같이 IRIS 데이터 객체와 활성화된 언어항과의 관계를 새롭게 생성한다.

[Rule-1]에서 '^'는 논리곱을 나타낸다. 그리고 'IrisData(?data)'는 'IrisData'의 하위 개념이나 개체를 변수 '?data'로 지정하는 것을 뜻하고, 'sepal_length(?data, ?sl)'은 변수 '?data'에 지정된 개념이나 개체와 'sepal_length'의 속성으로 관계가 설정된 개념, 개체, 특정 형태의 데이터를 "?sl" 변수에 지정함을 뜻한다. 여기서 'sepal_length'는 표 2에서 정의한 바와 같이 'datatype property'이므로 "?sl"에 지정되는 데이터는 'xsd:double' 형태이다. 따라서 [Rule-1]의 조건부를 만족하는 개념과 개체들 중에서 "?data"와 "?term"의 변수에 지정된 개념이나 개체를 'active_linguistic_term'이라는 속성으로 새로운 관계를 설정한다.

$$\begin{aligned}
 & \text{[Rule-1]} \\
 & \text{IrisData}(\text{?data}) \wedge \text{SepalLength}(\text{?term}) \\
 & \wedge \text{sepal_length}(\text{?data}, \text{?sl}) \\
 & \wedge MF_{\text{Triangular}}(\text{?term}, \text{?sl}, \text{?mv}) \wedge F_{\text{greaterThan}}(\text{?mv}, 0) \\
 & \rightarrow \text{active_linguistic_term}(\text{?data}, \text{?term})
 \end{aligned}$$

여기서, ' $MF_{\text{Triangular}}(\text{?term}, \text{?sl}, \text{?mv})$ '는 변수 '?term'에 지정된 언어항의 삼각형 모양의 소속 함수에서 '?sl'에서의 소속 함수 값을 변수 '?mv'에 지정하는 함수이다. 그리고 ' $F_{\text{greaterThan}}(\text{?mv}, 0)$ '는 변수 '?mv'의 값이 '0'보다 큰 경우 'true'를 반환하는 함수이다. 즉, 그림 7과 같이 개체 'iris_data_1'의 'sepal_length'값으로 퍼지 소속 함수 'SL_2'에서 구한 소속 함수 값이 '0'보다 큰 경우, 'iris_data_1'과 언어항 'SL_2'의 관계를 'active_linguistic_term'으로 설정한다는 의미이다.

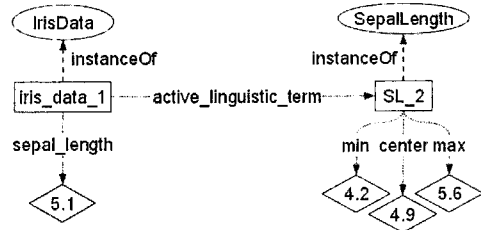


그림 7. IRIS 데이터와 활성화된 언어항과의 관계
Fig. 7. Relations between IRIS data and activated linguistic terms

3.2 규칙에 의한 분류 적합도 결정

IRIS 데이터의 패턴분류를 위한 분류 적합도를 결정하기 위해 참고문헌 [11]에서 정의한 퍼지 추론 규칙 [FuzzyRule]과 입력부 적합도의 판별을 위한 식 (4)를 변형하여 온톨로지 추론 규칙 [Rule-2]를 생성한다.

$$\begin{aligned}
 & \text{[FuzzyRule]} \\
 & \text{IF SepalLength}(SL_2) \text{ AND Sepal Width}(SW_6) \\
 & \text{THEN Class}(\text{setosa}) \\
 & \mu_{\text{setosa}} = \min(\mu_{SL_2}(v_{sl}), \mu_{SW_6}(v_{sw})) \quad (4)
 \end{aligned}$$

여기서, 언어 변수 ' v_{sl} '과 ' v_{sw} '는 각각 IRIS 데이터의 'sepal_length'와 'sepal_width'의 값이다.

$$\begin{aligned}
 & \text{[Rule-2]} \\
 & \text{IrisData}(\text{?data}) \\
 & \wedge \text{active_linguistic_term}(\text{?data}, SL_2) \\
 & \wedge \text{sepal_length}(\text{?data}, \text{?sl}) \\
 & \wedge MF_{\text{Triangular}}(SL_2, \text{?sl}, \text{?slmv}) \\
 & \wedge \text{active_linguistic_term}(\text{?data}, SW_6) \\
 & \wedge \text{sepal_width}(\text{?data}, \text{?sw}) \\
 & \wedge MF_{\text{Triangular}}(SW_6, \text{?sw}, \text{?swmv}) \\
 & \wedge F_{\text{min}}(\text{?slmv}, \text{?swmv}, \text{?mv}) \\
 & \rightarrow \text{setosa_value}(\text{?data}, \text{?mv})
 \end{aligned}$$

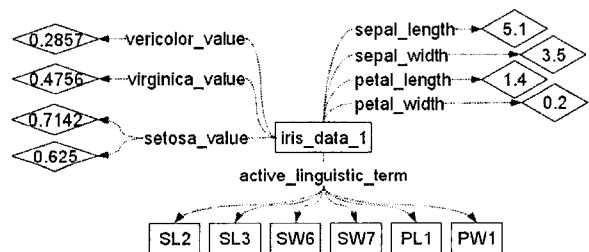


그림 8. 온톨로지 추론을 통한 분류 적합도
Fig. 8. 'iris_class_value' through ontological inference

그림 8은 개체 'iris_data_1'의 [Rule-2]에 대한 추론 결과를 보인다. [Rule-2]에서 개체 'iris_data_1'의 분류 적합도를 의미하는 'iris_class_value'의 하위 속성인 'setosa_value'의 속성 값을 변수 '?mv'에 지정된 '0.625'의 값으로 설정한다. 여기서, ' $\mu_{SL_2}(v)$ '나 ' $\mu_{SW_6}(v)$ '중 하나가 '0'인 경우에는 ' μ_{setosa} '도 '0'이 된다. 따라서 3.1절에서 설정한 'active_linguistic_term'을 이용하여 활성화된 언어항에 관

런된 추론 결과만을 생성함으로써 온톨로지에서 불필요한 연산과 관계 설정을 방지한다.

3.3 개체의 분류

온톨로지 추론을 통해 각 IRIS 데이터의 분류 적합도인 'setosa_value', 'versicolor_value', 'virginica_value' 속성 값을 결정하고, 식 (5)의 퍼지 연산에서 규칙의 이행정도가 최대인 클래스를 선택하는 것과 같이 [Rule-3]을 통해 IRIS 데이터의 분류 적합도가 최대인 값을 선택한다.

$$\mu_{dof} = \arg \max_{i = All Rules} (\mu_{class}^i) \quad (5)$$

여기서, 'i'는 퍼지 추론 규칙의 번호를 의미하며, ' μ_{class}^i '는 'i'번째 퍼지 추론 규칙의 이행정도, 그리고 ' μ_{dof} '는 IRIS 데이터의 규칙의 최대 이행정도를 나타낸다.

[Rule-3]

IrisData(?data)
 $\wedge iris_class_value(?p1) \wedge iris_class_value(?p2)$
 $\wedge ?p1(?data, ?cv1) \wedge ?p2(?data, ?cv2)$
 $\wedge F_{greater\ Than} (?cv1, ?cv2)$
 $\rightarrow C_{remove} (?p2(?data, ?cv2))$

[Rule-4]

IrisData(?data) $\wedge setosa_value(?data, ?setosa)$
 $\rightarrow Setosa(?data) \wedge C_{remove} (IrisData(?data))$

여기서, '*iris_class_value*(?p1)'은 '*iris_class_value*'의 하위 속성인 '*setosa_value*', '*versicolor_value*', '*virginica_value*'중의 한 속성을 변수 '?p1'에 지정하는 것이다. 또한, '*F_{greater Than}*(?cv1, ?cv2)'은 '?data'에 지정된 IRIS 데이터에서 변수 '?p1', '?p2'의 속성 값이 (?cv1 > ?cv2)인 경우 'true'를 반환한다. 그리고 '*C_{remove}*()'는 '삭제(remove)' 명령으로써 '?p2(?data, ?cv2)'의 관계를 삭제함을 뜻한다. 즉, 퍼지 추론에서 규칙의 이행정도가 가장 큰 값을 선택하는 것과 같이, 분류 적합도를 의미하는 '*iris_class_value*'가 가장 큰 값을 남기고 나머지는 모두 삭제함으로써 [Rule-4]를 통해 그림 9에서처럼 '*iris_data_1*'을 'Setosa'로 분류하게 된다. [Rule-4]는 개념 '*IrisData*'에 속해있는 IRIS 데이터 개체가 '*setosa_value*'의 속성 값이 존재하는 경우, IRIS 데이터 개체를 개념 '*IrisData*'로부터 개념 '*Setosa*'의 하위 개체로 이동하는 것을 뜻한다. 그림 9에서 개념 '*Setosa*', '*Virginica*', '*Versicolor*'는 서로 'disjoint' 관계로 설정되어 있어 특정 IRIS 데이터 개체가 반드시 이들 개념 중 단 하나의 개체로만 존재해야 한다.

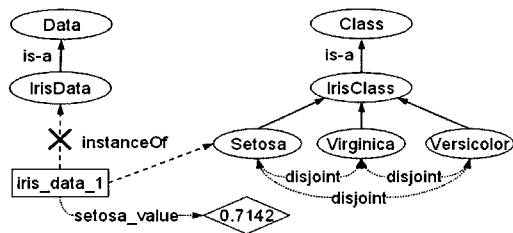


그림 9. IRIS 데이터의 분류
 Fig. 9. Classification of IRIS data

4. 패턴분류 실험

제안한 패턴 분류기의 타당성을 검증하기 위해 참고문헌 [11]에서 설정한 입력부의 퍼지 소속 함수와 퍼지 추론 규칙을 이용하여 온톨로지 기반 퍼지 분류기를 구축하고, IRIS 데이터 집합의 분류 성능을 비교한다. 온톨로지 기반 퍼지 분류기는 JAVA 기반의 Jena2[15]를 이용하여 구현하였다.

그림 10은 온톨로지 기반 퍼지 분류기를 통한 IRIS 데이터의 분류 과정을 보인다. 그림 10과 같이 분류되지 않은 IRIS 데이터를 '분류 전처리' 과정을 통해 개념 'IrisData'의 개체로 등록하고, 개체로 등록된 IRIS 데이터는 '언어항 선택', '분류 적합도 결정', '개체 분류' 과정의 온톨로지 추론을 통해 최종 분류된다.

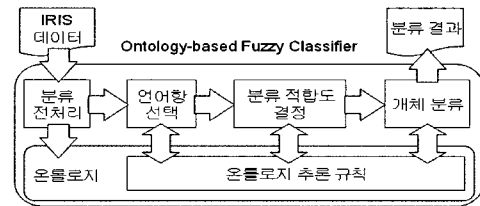


그림 10. 온톨로지 기반 퍼지 분류기
 Fig. 10. Ontology-based fuzzy classifier

그림 11과 표 3은 각각 참고문헌 [11]에서 설정한 입력부의 퍼지 소속 함수와 퍼지 추론 규칙이다. 참고문헌 [11]에서는 IRIS 데이터 집합의 속성을 그림 11과 같이 'sepal_length' 6구간, 'sepal_width' 8구간, 'petal_length' 5구간, 'petal_width' 5구간으로 나누었고, 표 3과 같이 총 47개의 퍼지 추론 규칙을 사용하였다. 그리고 150개의 IRIS 데이터에 대한 패턴분류 실험 결과로서 표 8과 같이 149개의 데이터가 정확히 분류되어 총 99.3%의 정확률을 보인다.

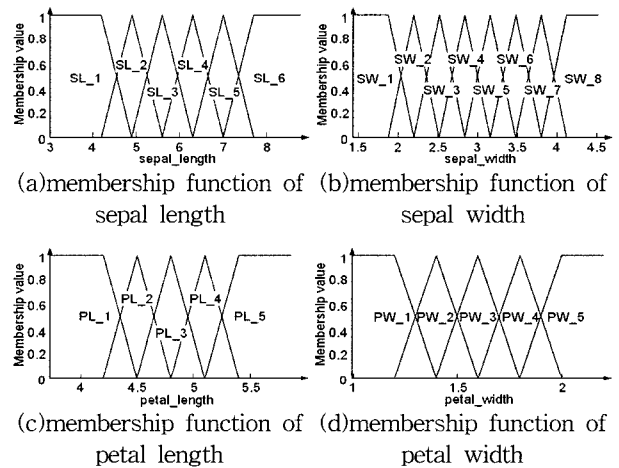


그림 11. 입력부 퍼지 소속 함수

Fig. 11. Fuzzy membership functions for antecedent parts

3장에서 언급한 추론 방법으로 IRIS 데이터가 분류되는 과정을 확인하기 위해 온톨로지 추론 과정에서 생성된 IRIS 데이터 개체와 설정된 관계를 온톨로지 검색을 통해 확인하였다. 여기에서는 IRIS 데이터 객체 '*iris_data_1*', '*iris_data_73*', '*iris_data_120*'에 대한 검색 결과만을 보인다.

표 3. 퍼지 추론 규칙

Table 3. Fuzzy inference rules

번호	입력부				출력부
	SL	SW	PL	PW	Class
1	SL_1	SW_8	-	-	setosa
2	SL_1		-	-	
3	SL_4	SW_8	-	-	
.					
11	SL_5	SW_5	PL_1	-	versicolor
12	-	SW_4	PL_1	-	
13		SW_3	PL_1	-	
.					
35		SW_2	PL_4	-	virginica
36	SL_6		-	PW_4	
47	SL_2	SW_3	-	-	

[Rule-1]의 추론을 통해 설정된 각 IRIS 데이터 객체와 활성화된 언어항과의 관계를 확인하기 위해 SPARQL[16] 문법에 기반한 [Query-1]의 온톨로지 검색 질의를 사용하였다. 그리고 온톨로지 검색 결과 표 4의 결과를 얻었다. 여기서, 'rdf:type'은 개념 'IrisData'에 속한 개체를, '?domain', '?range'는 변수를 의미한다.

```
[Query-1]
prefix isac:<http://isac.yu.ac.kr/classification/>
prefix rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
prefix rdfs:<http://www.w3.org/2000/01/rdf-schema#>
select ?domain, ?range
where {
    ?domain rdf:type isac:IrisData.
    ?domain isac:active_linguistic_term ?range.
}
```

표 4. IRIS 데이터 객체와 활성화된 언어항과의 관계
Table 4. Relations between instance of IRIS data and activated linguistic terms

domain	property	range
iris_data_1	active_linguistic_term	SL_2, SL_3, SW_4, SW_7, PL_1
iris_data_73	active_linguistic_term	SW_2, SW_3, PL_3, PL_4, PW_2
iris_data_120	active_linguistic_term	SL_5, SW_1, SW_2, PL_3, PL_4, PW_2, PW_3

활성화된 언어항과의 관계 설정 후, 표 3의 퍼지 추론 규칙을 변형한 [Rule-2]와 같은 형태의 온톨로지 추론 규칙으로부터 온톨로지 추론을 통해 IRIS 데이터 객체의 분류 적합도를 구하였다. 추론을 통해 설정된 각 IRIS 데이터 객체의 분류 적합도는 [Query-2]의 질의로부터 표 5의 결과를 얻었다. 그림 8을 통해 언급한 바와 같이 각 IRIS 데이터 객체의 분류 적합도는 표 5와 같이 다수 존재하였다. 그리고 이들 분류 적합도 중 [Rule-3]의 추론을 통해 표 6과 같은 최대 분류 적합도를 얻었다. 'iris_data_73'의 경우 'virginica_value', 'versicolor_value' 두 종류의 분류 적합도가 설정되었으나 [Rule-3]의 추론을 통해 'versicolor_value'

의 분류 적합도만이 선택되고, 나머지 관계는 삭제되었다.

```
[Query-2]
//prefix:<namespace> 선언 생략
select ?domain, ?property, ?range
where {
    ?domain rdf:type isac:IrisData.
    ?property rdfs:subPropertyOf isac:iris_class_value.
    ?domain ?property ?range.
}
```

표 5. 각 IRIS 데이터 객체의 분류 적합도
Table 5. 'iris_class_value' of each instance of IRIS data

domain	property	range
iris_data_1	setosa_value	0.333333333333333526 0.4545454545454546 0.5454545454545454
	versicolor_value	0.30434782608695654
iris_data_73	versicolor_value	0.5 0.6666666666666667
	virginica_value	0.0909090909090909 0.3333333333333352
iris_data_120	versicolor_value	0.3333333333333323 0.4545454545454535
	virginica_value	0.3333333333333327 0.5 0.5454545454545465

표 6. 최대 분류 적합도
Table 6. Maximum 'iris_class_value'

domain	property	range
iris_data_1	setosa_value	0.5454545454545454
iris_data_73	versicolor_value	0.6666666666666667
iris_data_120	virginica_value	0.5454545454545465

표 7. IRIS 데이터 객체의 분류
Table 7. Classification of each instance of IRIS data

domain	property	range
iris_data_1	rdf:type	Setosa
iris_data_73	rdf:type	Versicolor
iris_data_120	rdf:type	Virginica

```
[Query-3]
//prefix:<namespace> 선언 생략
select ?domain, ?property, ?range
where {
    ?range rdfs:subClassOf isac:IrisClass.
    ?domain rdf:type ?range.
}
```

마지막으로 각 IRIS 데이터 객체에 설정된 최대 분류 적합도를 나타내기 위해 사용된 속성 'iris_class_value'의 하위 속성에 따라 [Rule-4]의 추론을 통해 각 IRIS 데이터 객체를 개념 'Setosa', 'Virginica', 'Versicolor'의 하위 개체로 분류하였다. 표 7은 분류된 IRIS 데이터 객체를 [Query 3]의 질의를 통해 검색한 결과이다. 온톨로지를 이용한 패턴분류 실험 결과에서도 총 150 개

의 IRIS 데이터에 대해 149 개의 데이터가 정확히 분류되어 99.3 %의 패턴분류 정확률을 보였다. 즉, 표 8에서와 같이 참고문헌 [11]의 패턴분류 방법에 기반한 IRIS 데이터 집합의 분류 정확도와, 온톨로지를 이용한 패턴분류 방법의 분류 정확도가 일치함을 확인하였다.

표 8. 패턴분류 결과

Table 8. Results of pattern classification

패턴분류방법	퍼지기반	온톨로지 기반
입력	150 개	150 개
분류 개수	149 개	149 개
정확도	99.3 %	99.3 %

5. 결 론

본 논문에서는 퍼지규칙 기반 패턴분류 방법을 온톨로지 개념화 하였고, 함수 기반의 퍼지 추론 규칙을 논리 기반의 온톨로지 추론 규칙으로 변환함으로써 온톨로지를 이용한 퍼지 분류기를 제안하였다. 그리고 패턴분류 실험을 통해 기존의 퍼지 기반 패턴분류 방법과 동일한 분류 정확도 보임으로써 제안한 분류기의 타당성을 확인하였다.

온톨로지를 이용한 패턴분류보다 기존의 함수 기반의 패턴분류 방법이 더 빠르고, 추론 규칙 또한 더 단순하게 표현될 수 있다. 그러나 개념적인 패턴분류 방법 및 추론 규칙을 구현함에 있어 온톨로지를 이용함으로써 패턴분류 방법의 가독성을 높이고, 추론 규칙의 확장성을 높일 수 있는 장점이 있다. 또한, 현재 지식을 개념화한 온톨로지를 참조하는 연구는 많이 수행되고 있으나 추론을 통한 온톨로지의 활용 방법에 관한 연구는 부족한 상황이다. 따라서 본 논문은 패턴분류의 분야에 적합한 온톨로지의 설계와 추론 방법을 제시하였다는 것에 그 의미가 있고, 또한 이를 응용하여 온톨로지가 좀 더 다양한 분야에서 활용될 수 있을 것으로 판단된다.

참 고 문 헌

[1] N.E. Maillot and M. Thonnat, "Ontology based complex object recognition," *Image and Vision Computing*, Vol. 26, pp. 102-113, 2008.

[2] V. Mezaris, I. Kompatsiaris, and M.G. Strintzis, "Region-based Image Retrieval Using an Object Ontology and Relevance Feedback," *Eurasip Journal on Applied Signal Processing*, Vol. 6, pp. 886 - 901, 2004.

[3] M.H. Wang, C.S. Lee, H.C. Li, and W.M. Ko, "Ontology-based fuzzy inference agent for diabetes classification," *North American Fuzzy Information Processing Society (NAFIPS)*, pp. 79-83, 2007.

[4] K. Wolstencroft, P. Lord, and L. Taberero, "Protein classification using ontology classification," *Bioinformatics*, Vol. 22, No. 14, pp. e530-e538, 2006.

[5] R.E. Wray, S.A. Lisse, and J.T. Beard, "Ontology infrastructure for execution-oriented

autonomous agents," *Robotics and Autonomous Systems*, Vol. 49, pp. 113-122, 2004.

[6] R. Provine, C. Schlenoff, S. Balakirsky, S. Smith, M. Uschold, "Ontology-based methods for enhancing autonomous vehicle path planning," *Robotics and Autonomous Systems*, Vol. 49, pp. 123-133, 2004.

[7] C.S. Lee, and M.H. Wang, "Ontology-based intelligent healthcare agent and its application to respiratory waveform recognition," *Expert Systems with Applications*, Vol. 33, No. 3, pp. 606-619, 2007.

[8] J. Casillas, O. Cordón, F. Herrera, and M.J. Del Jesus, "Genetic feature selection in a fuzzy rule-based classification system learning process for high-dimensional problems," *Information sciences*, Vol. 136, pp. 135-157, 2001.

[9] V. Uebele, S. Abe, and M.S. Lan, "A neural network based fuzzy classifier," *IEEE Transactions on SMC*, Vol. 25, pp. 353-361, 1995.

[10] Y.C. Tsai, C.H. Cheng, and J.R. Chang, "Entropy-based fuzzy rough classification approach for extracting classification rules," *Experts systems with applications*, Vol. 31, No. 2, pp. 436-443, 2006.

[11] 손창식, 정환목, 권순학, "퍼지 규칙기반 분류시스템에서 퍼지 분할의 선택방법," *지능시스템학회 논문지*, 제18권, 3호, pp. 360-366, 2008.

[12] R.A. Fisher, "The Use of Multiple Measurements in Taxonomic Problems," *In Annual Eugenics*, Vol. 7, No. 2, pp. 179-188, 1936.

[13] OWL Web Ontology Language Guide, <http://www.w3.org/TR/owl-guide/>

[14] XML Schema, <http://www.w3.org/XML/Schema>.

[15] Jena-A Semantic Web Framework for Java, <http://jena.sourceforge.net/>

[16] SPARQL Query Language for RDF, <http://www.w3.org/TR/rdf-sparql-query/>

저 자 소 개

이인근(In K. Lee)
제18권 5호 (2008년 10월호) 참조
E-mail: iklee@ynu.ac.kr

손창식(Chang S. Son)
제18권 5호 (2008년 10월호) 참조
E-mail : fuzzyrisk@paran.com

권순학(Soon H. Kwon)
제18권 5호 (2008년 10월호) 참조
E-mail : shkwon@yu.ac.kr