

러프집합을 이용한 퍼지 규칙의 효율적인 감축

The Optimal Reduction of Fuzzy Rules using a Rough Set

노은영¹, 정환묵²

¹ 경북 경산시 대구가톨릭대학교 컴퓨터정보통신공학부
E-mail: eyroh@cu.ac.kr

² 경북 경산시 대구가톨릭대학교 컴퓨터정보통신공학부
E-mail: hmchung@cu.ac.kr

요 약

퍼지 추론은 애매한 지식을 효과적으로 처리할 수 있는 장점이 있다. 그러나 규칙의 연관속성은 규칙을 과다하게 생성하기 때문에 유용하고 중요한 규칙을 결정하는데 여러 가지 문제점이 있다.

본 논문에서는 퍼지 규칙에서 규칙간의 상관성을 고려하여 불필요한 속성을 제거하고, 퍼지 규칙의 상대농도를 이용하여 추론결과의 정확성을 유지하면서 규칙의 수를 최소화 하는 방법을 제안한다. 제안한 방법의 타당성을 검증하기 위하여 기존의 규칙 감축 방법에 따른 추론 결과와 비교 검증하였다.

Key Words : Rough set, attribution reduction, relative cardinality

1. 서 론

규칙의 연관 속성은 불필요한 규칙을 생성하기 때문에 유용하고 중요한 규칙을 결정하는데 여러 가지 문제점이 있다[2]. 이를 개선하기 위해 규칙의 수를 감축하고 추론의 정확성을 향상시키기 위한 연구가 지속되고 있다.

Ishibuchi는 규칙의 이행정도가 가장 높은 규칙을 추출하여 규칙을 최소화 하여 최소의 규칙과 분류 정확도를 향상 시켰지만, 입력패턴에 대응되는 모든 규칙을 고려하지 않았기 때문에 추론 결과의 객관성을 보장 할 수 없다[3,4]. 이를 해결하기 위한 방법으로 러프집합을 이용하여 규칙의 상관성을 고려하여 불완전한 정보의 의존성과 속성의 수를 감축함으로써 불완전성을 제거하고 규칙을 감축하는 방법들이 제안되었다.

본 논문에서는 퍼지 규칙에서 규칙간의 상관성을 고려하여 불필요한 속성을 제거하고, 퍼지 규칙의 상대농도를 이용하여 추론결과의 정확성을 유지하면서 규칙의 수를 최소화 하는 방법을 제안한다.

제안한 방법의 타당성을 검증하기 위하여 규칙을 감축 후 기존방법과의 추론결과를 비교하였다.

2. 러프집합의 기본개념

1980년대 초에 Pawlak에 의해 소개된 러프집합 이론은 어떤 집합에서 확실하게 분류되는 하한 근사 공간과 불확실하게 분류되는 상한 근사 공간을 집합이론을 통해서 나타낸다.

동치 관계에 의해 정보 객체 집단은 동치류로 나뉘어 질 수 있으며, 동치류 원소의 집합을 기본 집합이라 하고, 기본집합에 의해 정의되는 집합 공간을 근사공간이라고 한다. 근사 공간에 하나의 결정(decision)에 대한 정보 객체를 분류하는 경우, 동일한 기본집합에 있으면서도 서로 다른 결정을 내는 경우가 발생할 수 있다. 이러한 결정상의 불일치를 나타내고 처리하기 위해서 러프 집합 이론에서는 두 가지 근사를 정의한다. 결정에 의해 나타내어지는 개념 X 에 항상 포함되는 기본집합으로 정의되는 하한 근사와 개념 X 와 일치하는 부분이 하나라도 존재하는 모든 기본집합으로 정의되는 상한 근사이다. 그리고 상한근사에서 하한근사를 제외시키면 불확실한 개체들만 또 다른 부분집합으로 표현될 수 있으며 이것을 경계영역이라고 한다.

$$\text{하한근사 } \underline{R}X = U \{Y \in U/R : Y \subseteq X\} \quad (1)$$

$$\text{상한근사 } \overline{R}X = U \{Y \in U/R : Y \cap X \neq \emptyset\} \quad (2)$$

$$\text{경계영역 } BN_R(X) = \bar{R}X - \underline{R}X \quad (3)$$

러프집합에서는 경계영역이 적고 하한근사에 해당되는 범위가 큰 집합일수록 그 집합이 정확도가 높다고 한다. 전체집합의 정확성 정도는 식 (4)의 정확성 척도 α 를 이용하여 계산하며 근사화의 정확도라고 한다.

$$\alpha_R(X) = \frac{\text{card}R}{\text{card}\bar{R}} \quad (X \neq \emptyset) \quad (4)$$

3. 러프집합과 퍼지 상대농도를 이용한 규칙감축 방법

러프집합을 이용하여 규칙 내의 포함된 불완전한 속성을 제거함으로써 간략화된 규칙으로도 분류할 수 있는 방법들이 제안되고 있다. 따라서 본 논문에서는 제안된 방법에 따라 속성을 제거함으로써 규칙을 간략화 시키고, 각 규칙의 퍼지 상대농도가 최대인 규칙을 선택하여 규칙의 개수를 최소화 시키는 방법을 제안한다.

입력 데이터로부터 추론을 위한 퍼지 규칙은 다음과 같다.

$$R_i : \text{if } x_1 \text{ is } A \text{ and } \dots \text{ and } x_m \text{ is } M \\ \text{then } y \text{ is } E_i \quad (5)$$

여기서 x_1, \dots, x_m 은 입력부 변수이고, y 는 출력변수이다. $A, \dots, M, E_i (i = 1, 2, \dots, n)$ 는 소속함수이다.

3.1 러프집합을 사용한 규칙의 속성제거

규칙의 개수를 감축하기 위해서 퍼지 if-then 규칙으로부터 식 (6)을 이용하여 식별가능 행렬을 구한다.

$$c_{ij} = \{a \in A : a(x_i) \neq a(x_j)\} \quad (6)$$

여기서 $c_{ij} (i, j = 1, 2, \dots, m)$ 는 i 번째 입력 속성 x_i 와 j 번째 입력속성 x_j 를 구별하는 모든 속성의 집합이다.

행렬의 모든 단위 원소 엔트리의 집합으로부터 식 (7)을 이용하여 코어는 계산한다.

$$\text{CORE}(A) = \{a \in A : c_{ij} = (a)\}, \exists i, j \quad (7)$$

여기서 a 는 각각 i 와 j 번째 속성을 유일하게 구별할 수 있는 속성이다.

식별가능 행렬로부터 식별가능 함수 $f(a)$ 는 식 (8)와 같다.

$$f(A) = \prod_{(x,y) \in U^2} \left(\sum \delta(x,y) : (x,y) \in U_2 \right) \quad (8) \\ \text{and } \delta(x,y) \neq 0$$

코어와 리덕트를 계산하기 위한 식별가능 함수는 식 (9)와 같다.

$$f^x(A) = \prod_{y \in U} \left(\sum \delta(x,y) : y \in U \right) \quad (9) \\ \text{and } \delta(x,y) \neq 0$$

3.2 퍼지 상대농도를 이용한 규칙감축

본 절에서는 이전 절에서 획득한 감축된 규칙들 간에 redundant attribute들을 제거하기 위해서 퍼지 상대농도를 이용하여 퍼지추론을 위한 최종 규칙을 획득하였다.

다음은 퍼지추론을 위해 주어진 각 수치적인 입력 패턴들이 대응하는 규칙들의 퍼지 상대농도를 계산하기 위한 식을 나타낸다.

$$X = X / U \quad (10)$$

여기서 U 는 수치적인 입력패턴들이 대응된 퍼지 소속함수들의 수이고, X 는 그 입력패턴들이 대응된 소속함수들의 소속값들의 합을 나타낸다. $\|X\|$ 는 각 퍼지 if-then 규칙의 퍼지 상대농도를 의미한다.

4. 시뮬레이션

4.1 시뮬레이션 환경

본 논문에서는 제안된 방법의 타당성을 검증하기 위하여 rice taste data를 기반으로 다음과 같은 2가지 실험을 하였다: (1) 단순히 러프집합으로 감축 후 생성된 퍼지 if-then 규칙으로 퍼지추론을 한 경우 (2) 제안된 방법인 퍼지 상대농도를 기반으로 감축된 규칙의 redundant attribute들을 제거한 규칙으로 퍼지추론을 한 경우

규칙은 rice taste data를 분류하기 위한 IF~THEN 규칙과 각 입력변수 대응하는 변수와 소속함수를 구성하였다[1].

규칙은 5개의 입력변수와 1개의 출력변수로 나타내었다. 입력변수 5개는 x_1, x_2, x_3, x_4, x_5 로 나타내었고, Flavor, Appearance, Taste, Stickiness, Toughness를 의미한다. 출력변수 1개는 y 로 나타내었고, Evaluation을 의미한다. 입력변수 Flavor, Appearance, Taste의 레벨은 각 2가지로 good/ bad, Stickiness는 sticky/ not sticky, Toughness는 tough/tender로 나누었다. 출력변수 Evaluation은 5가지 레벨, high/ medium high/ medium/ medium low/ low로 나누었다.

실험 데이터를 IF-THEN 형태로 표현하면 표 1과 같이 32개의 규칙으로 표현된다.

표 1. 속성 감축 전 If-then 규칙

Rule	Input					Output
	x1	x2	x3	x4	x5	y
1	G	G	G	S	To	H
2	G	G	G	S	Te	H
3	G	B	G	S	To	MH
4	B	G	G	S	Te	MH
5	B	B	G	S	Te	MH
6	G	B	G	S	Te	MH
7	G	G	G	NS	Te	MH
8	G	B	B	S	Te	MH
9	G	G	B	S	To	M
10	G	G	B	NS	Te	M
⋮						
28	B	G	B	NS	Te	ML
29	B	B	B	NS	Te	ML
30	B	B	G	NS	Te	ML
31	G	B	B	NS	To	ML
32	B	B	B	NS	To	L

	x1	x2	x3	x4	x5	y
1	G	G	G	S	-	H
2	G	B	G	S	TO	MH
3	B	-	G	S	TE	MH
4	-	B	G	S	TE	MH
5	G	G	B	-	-	M
6	G	B	G	NS	TE	M
⋮						
11	-	B	G	NS	TO	ML
12	B	-	B	NS	TE	ML
13	B	B	B	NS	TO	L

4.2 리프집합을 이용한 속성감축과 중복된 규칙 제거

rice taste data로 표 1과 같이 생성된 퍼지 규칙에서 속성을 감축하고, 중복된 규칙을 제거하면 표 2와 같이 18개의 규칙으로 감축된다. 표 2에서 '-'은 감축된 속성을 의미한다.

표 2. 속성감축과 중복된 규칙 제거

Rule	Input					Output
	x1	x2	x3	x4	x5	y
1	G	G	G	S	-	H
2	G	B	G	S	TO	MH
3	B	-	G	S	TE	MH
4	-	B	G	S	TE	MH
5	G	B	G	S	TE	MH
6	G	G	B	-	-	M
⋮						
14	B	G	B	-	TO	ML
15	B	B	B	S	TO	ML
16	B	-	B	NS	TE	ML
17	B	B	G	NS	-	ML
18	B	B	B	NS	TO	L

18개로 감축된 규칙과 감축전의 32개 규칙의 입력부에 rice taste의 수치 데이터 105 set를 입력하고 각 규칙의 출력부 소속함수의 소속정도를 그림 1-5로 나타냈다. 소속정도가 정확하게 일치하였다.

4.3 제안하는 방법을 이용한 규칙감축

동일한 데이터를 이용하여 감축된 18개의 규칙에서 각 규칙의 입력속성에 대한 퍼지 상대농도가 가장 큰 규칙을 추출하여 규칙을 감축하였다. 규칙은 13개로 감축되어 개수가 상대적으로 최소화 되었다.

표 3. 최종 선택된 규칙

Rule	Input					Output

퍼지 상대농도를 구할 때 규칙의 이행정도는 매우 낮았지만 상대 농도의 값이 가장 높은 규칙은 13개의 규칙 가운데 2개였고 나머지 규칙은 이행정도와 상대농도 값의 순위가 다소 근사하거나 일치하였다.

규칙 감축 전의 32개의 규칙과 중복된 속성을 제거한 18개의 규칙 그리고 본 논문에서 제안한 방법에 의한 13개 규칙의 입력부에 4.2절과 동일한 데이터를 입력하고, 각 규칙의 출력부 소속함수의 소속정도를 그림 1-5으로 나타냈다.

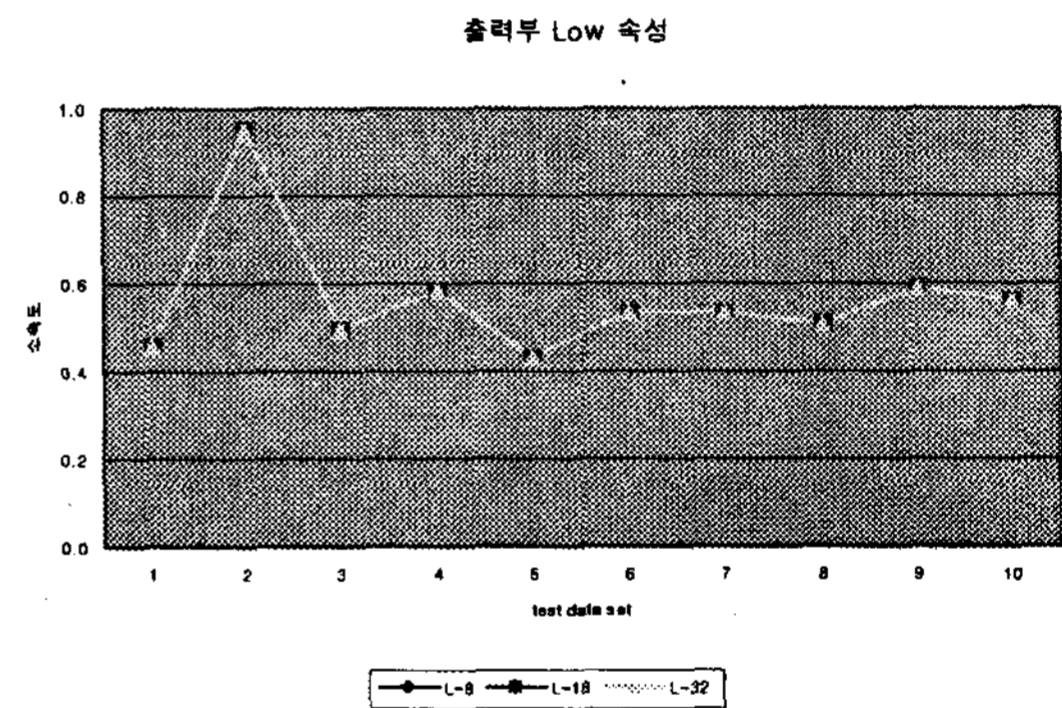


그림 1. 규칙의 출력부 소속함수 Low의 적합도

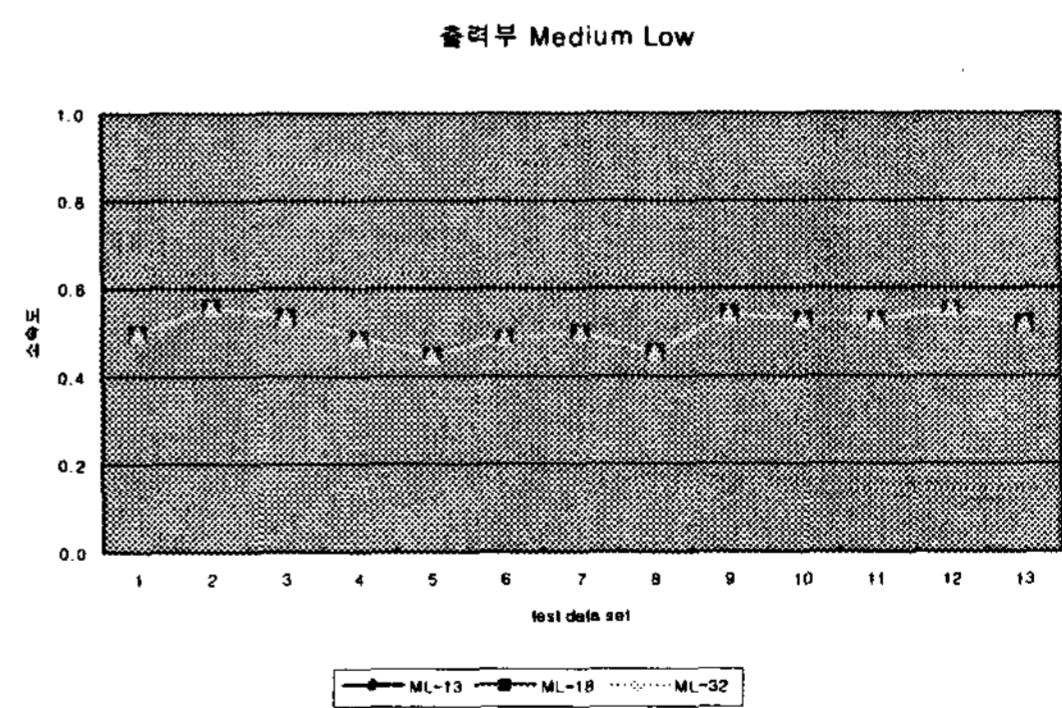


그림 2. 규칙의 출력부 소속함수 Medium Low의 적합도

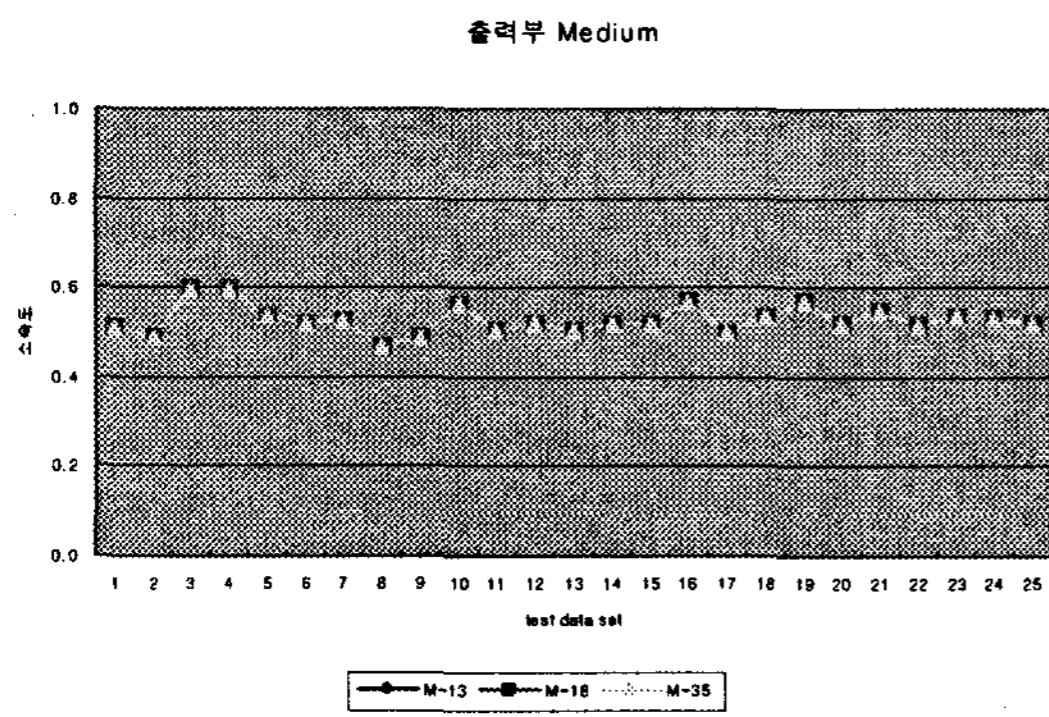


그림 3. 규칙의 출력부 소속함수 Medium의 적합도

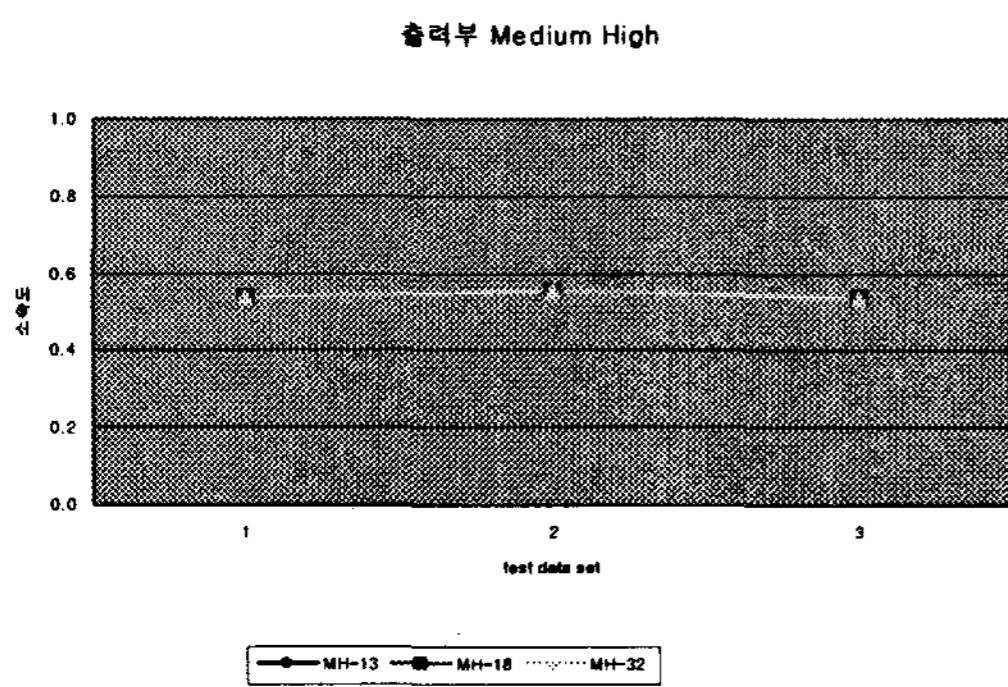


그림 4. 규칙의 출력부 소속함수 Medium High의 적합도

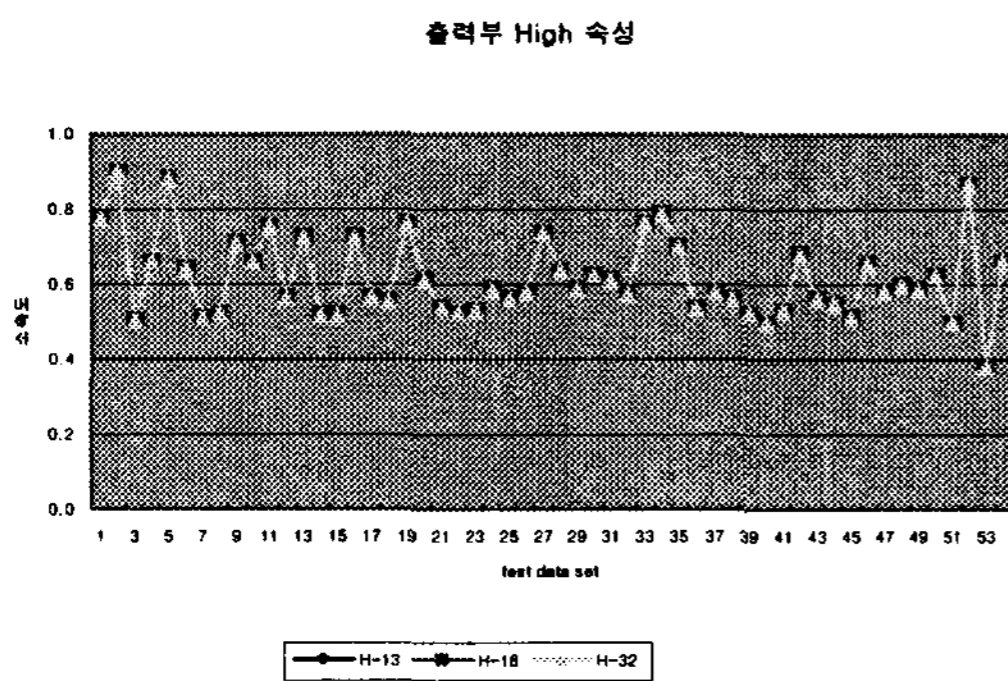


그림 5. 규칙의 출력부 소속함수 High의 적합도

각 출력부 소속함수의 비퍼지화 된 최소-최대 실수 구간을 분석하면 표 3과 같다.

표 3. 비퍼지화 된 실수 구간

	출력부	실수구간
rule 32개	Low	0 - 0.9535
	Medium Low	0 - 0.5654
	Medium	0.0341 - 0.5963
	Medium High	0 - 0.6178
	High	0 - 0.8433
rule 18개	Low	0 - 0.9539
	Medium Low	0 - 0.5540
	Medium	0 - 0.5963
	Medium High	0 - 0.5555

	High	0 - 0.9003
rule 13개	Low	0 - 0.9535
	Medium Low	0 - 0.5540
	Medium	0 - 0.5963
	Medium High	0 - 0.5555
	High	0 - 0.9003

감축전 32개의 규칙과 불필요한 속성제거에 의한 18개의 규칙, 그리고 상대농도에 의해 감축한 13개의 규칙은 출력부 소속함수의 적합도 패턴이 정확히 일치 하였다.

따라서 이행정도가 가장 높은 값을 가진 규칙을 승자규칙으로 추출하는 것 보다 규칙간의 상관성을 고려한 불필요한 속성을 제거하고, 상대농도 값에 따라 규칙을 추출할 때 규칙의 개수가 감축되어도 추론 결과가 정확하고 객관적인 것을 알 수 있다.

5. 결론

본 논문에서는 퍼지규칙을 최적화 하기 위한 방법으로 러프집합의 특성과 퍼지의 상대농도를 이용한 방법을 제안하였다. 러프집합으로 규칙간의 상관성을 고려하고 불명확한 속성을 제거함으로써 규칙을 감축하여 간소화하고, 감축된 각 규칙의 상대농도가 가장 높은 것을 추출하여 규칙을 최소화 하였다.

rice taste 의 수치 데이터 105 set의 입력 패턴에 대해 출력부 소속함수를 비교하였으며 규칙의 최대 적합도 속성이 모두 동일함을 알 수 있었다.

참 고 문 헌

- [1] K.Nozaki, H.Ishibuchi and Tanaka, "A Simple But Powerful Heuristic Method For Generating Fuzzy Rules From Numerical Data", Fuzzy Sets and Systems, Vol. 86, pp. 251-270, 1997.
- [2] Jiye Li, Nick Cercone, "A Rough Set Based Model to Rank the Importance of Association Rules", The Tenth International Conference on Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing (RSFDGrC 2005), 109-118, August 31-September 3rd, 2005.
- [3] H. Ishibuchi, T. Murata and I.B. Turksen, Single-Objective and Two-Objective Genetic Algorithms for Selecting Linguistic Rules for Pattern Classification Problems, Fuzzy Sets and Systems, Vol.89, pp. 135-150, 1997.
- [4] H. Ishibuchi, T. Nakashima and T. Morisawa, Voting in Fuzzy Rule-Based Systems for Pattern Classification Problems, Fuzzy Sets and Systems, Vol.103, pp. 223-238, 1999.