다중임계치의 세분화방법에 의한 근접관계모델

Proximity relational model by refinement of multi-threshold

류경현, 정환묵

대구가톨릭대학교 컴퓨터정보통신공학부

요 약

일반적으로 의사결정의 대상이 되는 현실 시스템은 매우 가변적(variable)이며 때로는 많은 불확실성(uncertainty)이 포함된 상황에 놓일 수 있다. 이러한 문제의 처리를 위한 통계적 방법으로 유의수준이나 확신도, 민감도 분석 등이 사용된다.

본 논문에서는 먼저 근접관계 행렬에서 근접도를 구하는 방법으로 상대적 해밍거리와 max-min 방법을 이용한 다음, 다중임계치를 사용하여 최적구간분할을 하는 방법을 제안한다. 결과적으로 max-min방법을 이용하여 다중임계치을 적용한 근접관계의 분류가 상대적 해밍거리로 근접도를 구하여 다중임계치를 구하는 방법보다 계산과정이 더 간단하고 명확하며 분할과정을 줄일 수 있고 최적의 의사결정에 효율적이라는 것을 알 수 있다.

Key Words : Fuzzy decision-making, Fuzzy relation, Similarity relation , Proximity relation, multi-threshold, Proximity degree

1. 서 론

퍼지이론은 1965년 L. A. Zadeh에 의해 처음 제안되어 비선형적이고 복잡한 시스템의 특성을 해석하는데 적용함으로써 제어, 클러스터링, 영상처리, 분류문제와 같은 다양한 분야에서 좋은 결과를 가져왔다 [1]. 그리고 퍼지 알파컷(FAC) 분석은 퍼지 논리와 퍼지 집합 이론을 기반으로 불확실한 지식을 표현하는데 널리 사용되며 이러한 매개변수는 주어진 소속함수를 가지고 퍼지수를 다루는 퍼지 논리기반 알파컷 분석이다. 그러나 애플리케이션에서 불확실한 매개변수의 단조함수는 계산은 빠르지만 비단조성의 결과는 여전히 연구가 필요하다는 단점을 가지고 있어서 무한 상관관계 거리의 가정에서만 적용할 수 있다 [2]. Son [3]은 퍼지함의 연산방법에 퍼지 $\alpha-level$의 개념을 적용하여 특정 레벨 이상의 추론패턴들을 분류할 수 있는 방법을 제안하였다.

본 논문에서는 먼저 근접관계 행렬에서 근접도를 구하는 방법으로 상대적 해밍거리와 max-min방법을 이용한 다음, 다중임계치를 사용하여 최적구간분할을 분류하는 방법을 제안한다.

본 논문의 구성은 다음과 같다. 2장에서는 퍼지관계의 분류에서 퍼지동치관계와 근접관계에 대한 개념과 근접도를 구하는 방법을 알아본다. 3장에서는 다중임계치의 세분화에 의한 최적구간분할로 먼저 다중임계치를 구하고 최적구간분할을 하는 방법에 대하여 알아본다. 4장에서는 적용예를 통하여 해밍거리에 의한 근접도를 가진 다중임계치에 대한 분류와 max-min에 의한 근접도를 가진 다중임계치에 대한 분류를 모의실험한다. 마지막으로 5장에서는 제안된 방법의 결론 및 향후 연구과제에 대해서 간략히 언급한다.

2. 관련 연구

2.1 퍼지동치관계

정의 1: 퍼지동치관계

한 집합 A에 정의된 퍼지관계 $R \subseteq A \times A$가 임의의 $x, y, z \in A$에 대하여 다음 조건을 만족하면 이 관계를 동치관계 또는 유사관계라 한다.

(1) 반사관계

$$\forall x \in A \Rightarrow \mu_R(x, x) = 1$$

(2) 대칭관계

$$\forall (x, y) \in A \times A,\ \ \mu_R(x, y) = \mu \Rightarrow \mu_R(y, x) = \mu$$

(3) 전이관계

$$\forall\ (x, y), (y, z), (x, z)\ \in\ A \times A,$$

$$\mu_R(x, z)\ \geq\ \underset{y}{Max}\,[Min\,(\mu_R(x, y), \mu_R(y, z))]$$

유사관계를 이용하여 $\alpha-level$에 의한 분할을 나타내고 분할을 추론하며 프로세스에서 각 분할은 퍼지 소속 값과 관련있다 [7]. 분할은 유사관계가 α의 수준값에 대한 제한조건이 더 이상 만족될 수 없을 때 까지 스칼라 도메인의 원소를 병합하며 관찰하는 것이 중요하다. 다시 말하면 α_i가 클수록 분류경향은 각 데이터 포인트가 자신의 클래스에 할당하는 것이 자명하며 또한 퍼지 관계 경우, 분할에서 각 동치 클래스는 스칼라 도메인에서부터 서로에 대해 α만큼의 레벨에 대한 유사도를 산출하는 원소를 포함하는데 이것은 max-min 전이관계를 만족한 결과이다.

2..2 근접관계(proximity relation)

어떤 도메인에 속하는 여러 데이터 개체간의 근접도를 행렬로 표현한 것을 근접관계라 하고 퍼지 관계 모델은 스칼라 도메인에 대해 유사 관계를 근접 관계로 대체함으로써 확장된다 [7]. 근접관계는 다음과 같이 정의한다.

정의 2 : 근접 관계는 $x, y \in D_i$와 같이 $s_i : D_i \times D_i \rightarrow [0,1]$ 에 사상된다.

(1) 반사관계 : $s_i(x, x) = 1$
(2) 대칭관계 : $s_i(x, y) = s_i(y, x)$

2.3 근접도를 구하는 방법

관계에서 값을 결정하는 가장 일반적인 형태중의 하나는 데이터의 조작이다 데이터집합이 더 견고할수록 관계 항목들도 더 정확하게 두개 또는 그 이상의 데이터집합의 원소사이에서 확실한 관계가 있다 [6].

근접도(proximity degree)란 주어진 관점에 대한 데이터 개체간의 유사한 정도를 정량적으로 나타낸 것이다.

2.3.1 해밍거리(Hamming distance) 방법

해밍 거리는 다음 식 (1)과 같이 각 원소의 소속함수 값의 차이에 절대값을 취하여 합한 값과 같다.

$$d(A,B) = \sum_{\substack{i=1 \\ x_i \in X}}^{n} |\mu_A(x_i) - \mu_B(x_i)| \qquad (1)$$

전체집합 X의 원소가 n개 있을 때 즉 $|X| = n$ 일 때 상대 해밍거리는 다음 식 (2)와 같이 계산한다.

$$hd(A,B) = \frac{1}{n} d(A,B) \qquad (2)$$

이 해밍거리는 다음의 성질을 만족한다.
(1) $d(A,B) \geq 0$
(2) 교환법칙 : $d(A,B) = d(B,A)$
(3) 전이법칙 :
$$d(A,C) \leq d(A,B) + d(B,C)$$
(4) $d(A,A) = 0$

2.3.2. max-min 방법

관계의 각 요소 r_{ij} 는 두개의 데이터 샘플(x_i, x_j)의 쌍비교로부터 얻어지며 데이터 샘플 x_i와 x_j사이의 관계의 세기(strength)는 $r_{ij} = \mu_R(x_i, y_j)$로 표현하는 소속값으로 주어지고 관계행렬은 $n \times n$으로. 모든 근접도는 $0 \leq r_{ij} \leq 1$이다.

max-min 방법은 다음과 같은 식 (3)에서 r_{ij} 을 계산한다.

$$r_{ij} = \frac{\sum_{k=1}^{m} \min(x_{ik}, x_{jk})}{\sum_{k=1}^{m} \max(x_{ik}, x_{jk})}, \qquad (3)$$
$$단,\ i,j = 1,2,\ldots,n$$

3. 다중임계치의 세분화에 의한 근접관계의 최적구간분할

3.1 다중임계치를 구하는 방법

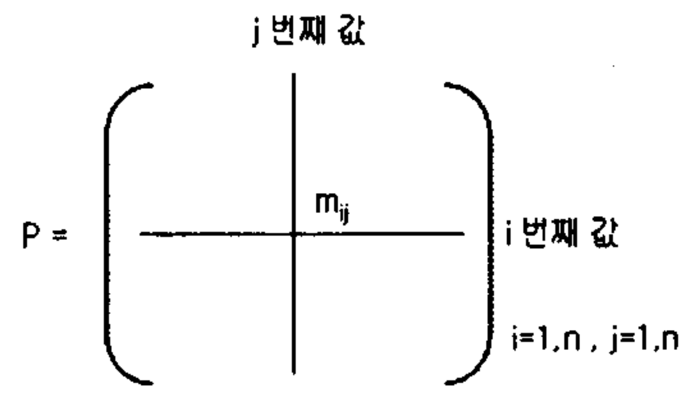
퍼지관계를 사용하여 전체집합에서 데이터 포인트를 분류하기 위해 연관된 근접관계를 이용한다. 도메인의 분할은 매우 유익하며 특히 데이터베이스에서 중복성의 개념과 관계 표현의 일치성, 명확한 관계 대수를 유지하는데 도움을 주기 때문에 중요한 일이다 [7].

크리스프 데이타 $X \times Y$에 있어서 의미를 가지도록 어떤 특정 값을 부여하여 퍼지화한 퍼지관계 R에 대하여, 만일 X와 Y가 $X = \{x_1, x_2, \cdots, x_m\}$, $Y = \{y_1, y_2, \cdots, y_n\}$인 유한집합이라면 퍼지관계 R은 다음과 같은 $m \times n$행렬 M_R에 의해 나타낼 수 있다 [8].

$$M_R = \begin{bmatrix} r_{11} & \cdots & r_{1n} \\ r_{21} & \cdots & r_{2n} \\ & \cdots & \\ r_{m1} & \cdots & r_{mn} \end{bmatrix}$$
$$단,\ r_{ij},\ i = 1, m,\ j = 1, n$$

퍼지관계 R로부터 근접도를 구하는 방법을 사용하여 근접관계 D를 생성하고 $n \times n$ 행렬로 배열된다.

$$D = \{d_1, d_2, \ldots, d_n\}$$



p는 대응되는 요소간의 근접도로서 m_{ij} 는 d_1과 d_2 사이에서 가지는 근접도를 단위구간에서 정량화한 것이며 m_{ij}가 클수록 두 요소간의 의견이 더 강하게 일치한다.

단계1) [0,1] 구간의 p 분할

D를 유한 도메인으로 하고 p는 D상에서 근접관계를 가진다고 하자. 다음과 같은 크기의 오름차순에서 관계의 값 p(x,y)을 배열한다.

$$m_1 < m_2 < \cdots < m_{k-1} < m_k = 1$$

단계2) p-분할 부분구간을 구한다.

$[0, m_1]$을 포함하는 부분구간($(m_r, m_{r+1}]$)은 다음과 같이 표현된다.

$$I_1 = \begin{cases} [0, m_1], & \text{if } m_1 \neq 0 \\ 0, & \text{if } m_1 = 0 \end{cases} \qquad I_2 = (m_1, m_2]$$
$$I_3 = (m_2, m_3] \cdots I_k = (m_{k-1}, m_k]$$

단계3) 다중임계치 a_i 의 각 값에 대해 k개의 클러스터 분포(D-분할)로 분할하여 임의의 원소 $x \in A$에 관

련되어 있는 원소와 유사한 집합으로 묶고 α_i 의 값이 클수록 더욱 효율적임을 알 수 있다.

3.2 최적구간분할

[0,1] 구간에서 최적구간을 찾기 위한 데이터포인트로 $\alpha(\neq 0)$ 가 임계치라면 다음과 같이 α^-를 정의한다.

$$\alpha^- = \begin{cases} (\alpha_p, 0) & \text{if } \alpha_p \neq 0 \\ \alpha^- = [0, \alpha) & \text{if } \alpha_p = 0 \end{cases}$$
$$\text{단, } \alpha_p \text{는 } \alpha \text{의 전노드}$$

만약 $\alpha(\neq 1)$ 가 임계치라면 다음과 같은 α^+를 정의한다.

$$\alpha^+ = \begin{cases} (\alpha, \alpha_n] & \text{if } \alpha \neq 0 \\ [0, \alpha_n) & \text{if } \alpha = 0 \end{cases}$$
$$\text{단, } \alpha_n \text{는 } \alpha \text{의 다음노드}$$

위의 정의로부터 α의 최소허용범위는 L(α)로 표현하고 다음과 같이 정의한다.

$$L(\alpha) = \frac{\alpha^- \text{의 분할개수}}{[0,1] \text{의 분할개수}}$$

α의 최대허용범위는 U(α)로 표현하고 다음과 같이 정의된다.

$$U(\alpha) = \frac{\alpha^+ \text{의 분할개수}}{[0,1] \text{의 분할개수}}$$

분류된 클래스에서 원하는 에이전트를 산출하여 조합을 찾는 것이 가능하며 다중임계치 α는 의사결정을 하는데 중요한 역할을 한다는 것을 알 수 있으며, 만약 모든 에이전트가 성취하는 대안이 있으면 이것이 퍼지 의사결정이다 [4].

4. 적용 예

7가지의 의견 $\{p1, p2, p3, p4, p5, p6, p7\}$에 대하여 6명의 에이전트 $\{Agt_1, Agt_2, Agt_3, Agt_4, Agt_5, Agt_6\}$가 각각 주관적인 선호도를 나타낸 것이다. <표 1>에서 1.0은 선호도가 가장 높은 것을 나타내며 0은 가장 낮은 선호도를 나타낸다.

<표 1> 선호도를 나타내는 퍼지관계 행렬

	p1	p2	p3	p4	p5	p6	p7
Agt_1	0.1	1.0	0.9	0.9	0.0	0.7	0.6
Agt_2	0.7	1.0	0.0	0.1	0.2	0.9	1.0
Agt_3	0.6	0.5	0.7	1.0	1.0	0.5	0.2
Agt_4	0.1	0.0	0.2	0.2	0.3	0.3	0.6
Agt_5	0.4	0.4	0.8	0.5	0.2	0.0	0.0
Agt_6	1.0	0.4	0.3	0.7	1.0	1.0	0.6

이러한 선호도 조사의 결과를 통하여 다중임계치 α-수준을 이용하여 각각의 근접도 지표에 의한 분류를 할 수 있다.

4.1 해밍거리에 의한 근접도를 가진 다중임계치에 대한 분류

에이전트 Agt_1과 Agt_2간의 상대 해밍거리는 다음과 같이 계산하고 <표 2>에 나타낸다.

$$hd(P_1, P_2) = \frac{1}{7} \cdot \left\{ \begin{array}{r} |0.1-0.7|+|1.0-1.0|+|0.9-0.0|+|0.9-0.1|+ \\ |0.0-0.2|+|0.7-0.9|+|0.6-1.0| \end{array} \right\} = 0.4$$

<표 2> 해밍거리의 근접도

	Agt_1	Agt_2	Agt_3	Agt_4	Agt_5	Agt_6
Agt_1	1.0	0.6	0.6	0.6	0.6	0.5
Agt_2	0.6	1.0	0.4	0.6	0.4	0.6
Agt_3	0.6	0.4	1.0	0.5	0.7	0.7
Agt_4	0.6	0.6	0.5	1.0	0.6	0.5
Agt_5	0.6	0.4	0.7	0.6	1.0	0.5
Agt_6	0.5	0.6	0.7	0.5	0.5	1.0

<표 2>에서 구한 해밍거리에 의한 근접도로 다중임계치 $\alpha-level$에 의한 분류는 다음과 같다.

단계1) [[0,1] 의 p분할은 다음과 같이 주어진다.
$0.1 < 0.2 < 0.3 < 0.4 < 0.5 < 0.6 < 0.7 < 0.8 < 0.9 < 1.0$

단계2) 해밍거리에 의한 다중임계치를 나타내면 다음과 같다.

$I_1 = [0, 0.4]$ $\quad I_2 = (0.4, 0.5]$
$I_3 = (0.5, 0.6]$, $\quad I_4 = (0.6, 0.7]$

다중임계치(=p분할 부분구간)로 구간별 에이전트를 분류하면 <표 3>과 같다.

<표 3> 해밍거리의 구간별 에이전트 분류

에이전트	근접도구간값
Agt1	[0.5 - 0.6]
Agt2	[0.4 - 0.6]
Agt3	[0.4 - 0.7]
Agt4	[0.5 - 0.6]
Agt5	[0.4 - 0.7]
Agt6	[0.5 - 0.7]

단계3) 다중임계치 α_i 의 각 값에 대한 분류의 결과는 그림 1에서 나타낸다.

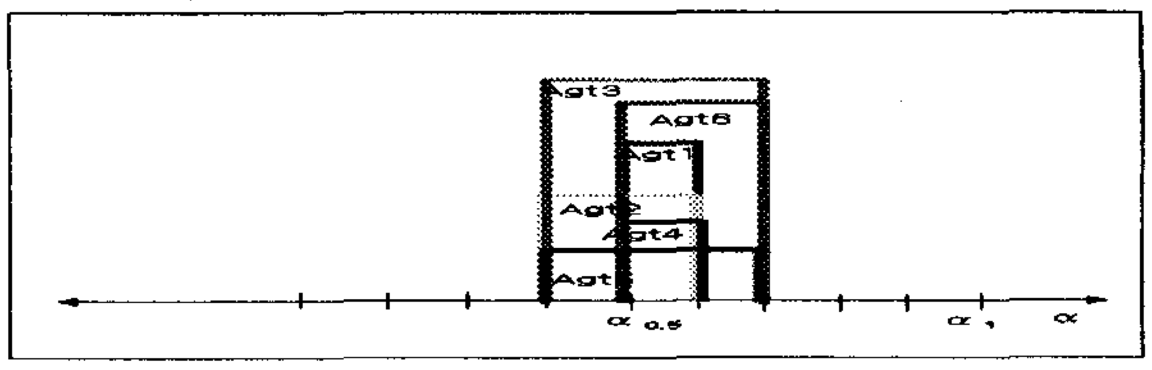


그림 1 해밍거리의 구간별 에이전트

그림 1에서 L(α)는 40%, U(α)는 30%임을 알 수 있다.

4.4 max-min 방법에 의한 근접도를 가진 다중임계치에 대한 분류

<표 1>에서 주어진 데이터를 가지고 식 (3)을 이용하여 i =1 j=2에 대해 계산하고 <표 4>에 나타낸다.

$$r_{12} = \frac{\sum_{j=1}^{7}(\min(0.1,0.7), \min(1.0,1.0), \min(0.9,0.0), \min(0.9,0.1), \min(0.0,0.2), \min(0.7,0.9), \min(0.6,1.0))}{\sum_{j=1}^{7}(\max(0.1,0.7), \max(1.0,1.0), \max(0.9,0.0), \max(0.9,0.1), \max(0.0,0.2), \max(0.7,0.9), \max(0.6,1.0))}$$
$$= \frac{0.1+1.0+0.0+0.1+0.0+0.7+0.6}{0.7+1.0+0.9+0.9+0.2+0.9+1.0}$$
$$= 0.5$$

<표 4> max-min의 근접도

	Agt_1	Agt_2	Agt_3	Agt_4	Agt_5	Agt_6
Agt_1	1.0	0.5	0.5	0.4	0.4	0.5
Agt_2	0.5	1.0	0.4	0.3	0.3	0.5
Agt_3	0.5	0.4	1.0	0.3	0.5	0.7
Agt_4	0.4	0.3	0.3	1.0	0.3	0.4
Agt_5	0.4	0.3	0.5	0.3	1.0	0.4
Agt_6	0.5	0.5	0.7	0.4	0.4	1.0

<표 4>에서 구한 max-min방법에 의한 근접도로 다음 <표 5>와 같은 구간별 에이전트 분류를 구하고 다중임계치 α_i 의 각 값에 대한 분류의 결과는 그림 2에서 나타낸다.

<표 5> max-min의 구간별 에이전트 분류

에이전트	근접도구간값
Agt1	[0.4 - 0.5]
Agt2	[0.3 - 0.5]
Agt3	[0.3 - 0.7]
Agt4	[0.3 - 0.4]
Agt5	[0.3 - 0.5]
Agt6	[0.4 - 0.7]

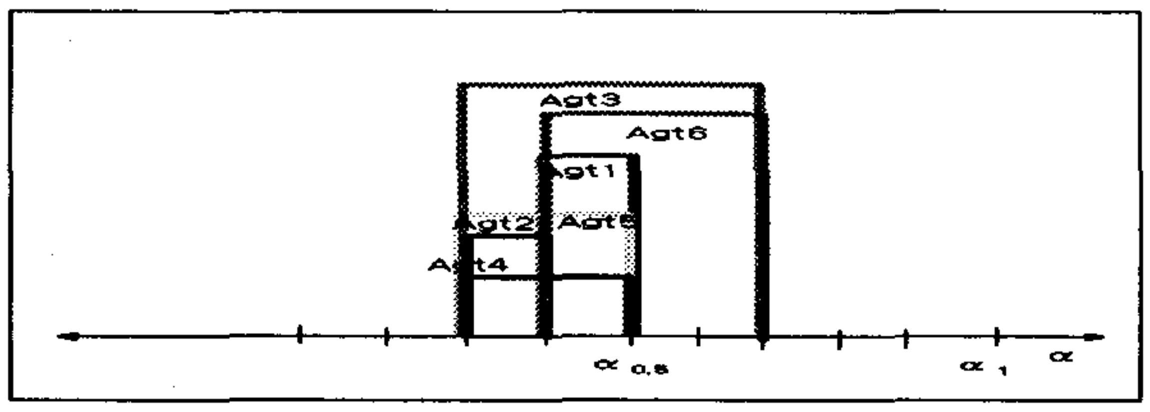


그림 2 max-min의 구간별 에이전트

그림 2에서 L(α)는 30%, U(α)는 30%임을 알 수 있다. 위의 두 가지 방법에서 Agt_3과 Agt_6의 의견이 가장 근접도가 높다는 것을 알 수 있으며 max-min방법에 의한 방법이 해밍거리에 의한 근접도로 다중임계치를 분류한 것보다 계산이 더 간단하고 빠르게 분류할 수 있다는 것을 알 수 있었다.

5. 결과 및 향후 연구과제

본 논문은 근접관계 행렬에서 근접도를 구하는 방법으로 상대적 해밍거리와 max-min방법을 이용한 다음, 다중임계치를 사용하여 최적의 구간분할을 하는 방법을 제안한다.

근접관계에서 max-min 전이관계 제한조건의 삭제는 퍼지 데이터베이스의 사용자에게 스칼라 도메인에 대한 값 구조를 표현하기 위해 더 자유롭게 제공된다는 장점을 가지고 있다 따라서 유사관계에서 $\alpha-level$에 의한 도메인 분할은 변화하지 않지만 근접관계를 가진 도메인 분할은 데이터베이스 갱신(update)할 수 있고 근접관계에 의해 결정된 클러스터의 경우는 유사관계에 의해 결정된 클러스터보다 데이터베이스의 일시적인 성질에 대해 훨씬 더 섬세하다. 왜냐하면 퍼지관계에서 튜플의 추가와 삭제는 형성된 클러스터에 대하여 완전한 효과를 가질 수 있기 때문이다. 또한 근접도는 데이터 집합으로부터 퍼지 관계내에서 관계의 세기나 거리를 전개하는데 사용하며 다중임계치를 사용한 방법은 의사결정 지원 시스템에 대한 결과물을 유도하는데 유용하며 더욱 중요한 것은 부정확추론에서 결정 로직을 추출할 수 있다.

참 고 문 헌

[1] 손창식, "구간값 퍼지집합과 규칙감축에 기반한 패턴분류", 대구가톨릭대학교 대학원 박사학위논문, 2006.

[2] A. J. Abebe, V. Guinot, D. P. Solomatine, "Fuzzy alpha-cut vs. Monte Carlo techniques in assessing uncertainty in model parameters", Proc. 4-th International Conference on Hydroinformatices, Iowa City, USA, July 2000.

[3] C. S. Son, H. M. Chung, "An Emotion Classification Based on Fuzzy Inference and Color Psychology", International Jpurnal of Fuzzy Logic and Intelligent Systems, Vol. 4, No. 1, pp. 18-22, 2004.

[4] R. E. Beliman and L .A. Zadeh, "Decision-making in a Fuzzy Environment", Management Science Vol. 17, No. 4, December, 1970.

[5] Supriya Kumar De, Ranjit Biswas, Akhil Ranjan Roy, "On extended fuzzy relational database model with proximate relation", Fuzzy Sets and Syatem 117, pp.195-201, 2001.

[6] J. Ross, Fuzzy logic with engineering applications, Timothy , McGraw-Hill, Inc, 1995.

[7] Sujeet Shenoi and Austin Melton, "Proximity relations in the fuzzy relational database model", Fuzzy Sets and Systems 100 Supplement, pp.51-62, 1999.

[8] 정환묵 편저, 소프트 컴퓨팅, 내하출판사, 2007.