http://dx.doi.org/10.7236/JIIBC.2013.13.5.183
JIIBC 2013-5-22

러프 엔트로피를 이용한 범주형 데이터의 클러스터링

lustering of Categorical Data using Rough Entropy

박인규*

Inkyoo Park

요 약 객체를 분류하기 위하여 유사한 특징을 기반으로 하는 다양한 클러스터해석은 데이터 마이닝에서 필수적이다. 그러나 많은 데이터베이스에 포함되어 있는 범주형 데이터의 경우에 기존의 분할접근방법은 객체간의 불확실성을 처리하는데 한계가 있다. 범주형 데이터의 분할과정에서 식별불가능에 의한 동치류의 불확실성에 대한 접근논리가 러프집합의 대수학적인 논리에만 국한되어서 알고리즘의 안정성과 효율성이 떨어지는 요인으로 작용하고 있다. 본 논문에서는 범주형 데이터에 존재하는 속성의 의존도를 고려하기 위하여 정보이론적인 척도를 기반으로 러프엔트로피를 정의하고 MMMR이라는 알고리즘을 제안하여 분할속성을 추출한다. 제안된 방법의 성능을 분석하고 비교하기 위하여 K-means, 퍼지에 의한 방법과 표준편차를 이용한 기존의 방법과 비교우위를 ZOO데이터에 국한하여 알아본다. ZOO데이터를 이용하여 기존의 범주형 알고리즘과의 비교우위를 살펴보고 제안된 알고리즘의 효율성을 검증한다.

Abstract A variety of cluster analysis techniques prerequisite to cluster objects having similar characteristics in data mining. But the clustering of those algorithms have lots of difficulties in dealing with categorical data within the databases. The imprecise handling of uncertainty within categorical data in the clustering process stems from the only algebraic logic of rough set, resulting in the degradation of stability and effectiveness. This paper proposes a information-theoretic rough entropy(RE) by taking into account the dependency of attributes and proposes a technique called min-mean-mean roughness(MMMR) for selecting clustering attribute. We analyze and compare the performance of the proposed technique with K-means, fuzzy techniques and other standard deviation roughness methods based on ZOO dataset. The results verify the better performance of the proposed approach.

Key Word : Cluster analysis, Clustering, Rough Set, Rough Entropy, Uncertainty

Ⅰ. 서 론

애매하고 불분명한 상황에서 여러 문제들을 판단하고 결정하는 과정에서 수학적으로 접근하려는 방법들이 많이 개발되어 왔다. 인공지능 분야에서 애매성(vagueness)과 식별불가능(indiscernibility)으로 대표되는 불확실한 정보의 핵심에는 임의의 대상을 구성하는 객체간의 명확한 경계에 대한 연구가 이루어져 왔다. 애매성은 자연어의 범주가 점진적인 개념으로 불분명한 경계를 갖는 집합이고, 식별불가능은 지식의 알갱이성(granularity)과 관련이 있다. 모호성(ambiguity)을 다룬 증거이론(evidence theory)은 확률이론의 확장으로 믿음(belief)척도, 근사(plausibility)척도, 확률(probability)척도, 가능(possibility)척도, 필요(necessity)척도와 같은 퍼지척도(fuzzy measure)에 의해 불확실성을 표현하였다. 궁극적으로 모든 알고리즘의 목적은 부정확한 데이터로

*정회원, 중부대학교, 컴퓨터학과
접수일자 : 2013년 9월 4일, 수정완료 : 2013년 10월 5일
게재확정일자 : 2013년 10월 11일

Received: 4 September, 2013 / Revised: 5 October, 2013 / Accepted: 11 October, 2013
*Corresponding Author: fip2441g@gmail.com
Dept. of Computer Science, Joongbu Univ, korea.

부터의 추론, 보다 정확하게 말하면 데이터간의 관계를 발견하는 것이다. 그러나 통계적인 방법을 이용하지 않고 불확실성을 접근하는 러프집합이 관심의 대상이 되어 왔다.

1980년대에 Pawlak이 제안한 러프집합의 개념은 하나의 집합을 상한근사와 하한근사로 불리는 두 개의 집합으로 근사화하는 것으로 이루어진다[1]. 또한 이러한 근사화(approximation)를 통하여 집합의 원소에 관해 얻을 수 있는 데이터를 바탕으로 생성된 위상(topology)내의 내부(interior)와 닫힘(closure) 연산에 해당한다고 볼 수 있다. 러프집합에 의한 방법은 어떤 실체에 관한 주체가 가지고 있는 지식과 어떤 현상, 과정, 객체를 식별하는 능력에 기반을 두고 있고, 관찰 또는 측정으로부터 얻어진 데이터를 분류하는 능력이 있다. 그러나 러프집합에서 데이터의 패턴을 추출하기 위한 데이터를 분류함에 있어서 분할속성의 추출과 비결정성(non-deterministic)데이터의 감축에서 문제점이 발생한다.

본 연구는 이러한 러프집합의 문제점을 개선하고자 엔트로피의 정보이론에 기반하여[2], 범주형 데이터의 속성의 중요도의 불확실성에 대한 새로운 척도를 정의하고 이를 기반으로 하는 분할 알고리즘을 제안한다. 우선 조건 및 결정속성에 두 변량간의 러프엔트로피의 개념을 이용하여 가장 안정적인 속성을 추출하는 최적의 분할속성을 생성하는 알고리즘을 제안한다. 또한 가변정확도 러프집합(variable precision rough set)을 이용한 종래의 분할 알고리즘에 의한 속성간의 중요도에서 변별력이 저조하여 최적의 리덕트의 생성에 문제점이 있었다[3]. 따라서 러프 엔트로피에 기반으로 한 정보이론적인 엔트로피 척도를 적용하여 속성간의 변별력을 향상시키는 min-mean-mean-roughness (MMMR)알고리즘을 제시한다.

II. 관련연구

1. 클러스터링의 불확실성

지금까지 범주형 데이터에 관한 많은 클러스터링 알고리즘이 개발되어 왔다. Dempster[4]는 확률을 이용하여 EM(expectation maximation)알고리즘을 제안하였고, Han는] 객체간의 관계를 규칙으로 나타내는 그래프 개념을 이용한 알고리즘을 제안하였다. K-modes는 K-means를 확장하여 범주형 데이터에 대한 속성간의 비유사성의 척도를 새롭게 정의하였지만, 최적의 해가 국부성에 의존하기 때문에 안정성을 확보하기 위해서는 많은 데이터를 필요로 한다. Ralambondrainy[5]는 여러 개의 속성을 이진속성{0, 1}으로 변환하여 K-means알고리즘으로 전환하는 알고리즘을 제안하였다. Huang[6]는 수치데이터와 범주형 데이터를 같이 처리할 수 있는 CACTUS알고리즘을 제안하였고, Guha[7]는 두 객체간의 공통된 관계를 이용하여 객체간의 유사성의 척도로 이용하는 계층적인 알고리즘을 제안하였다.

이러한 모든 알고리즘은 하나의 객체는 반드시 하나의 클러스터에 할당되어진다. 따라서 임의의 클러스터에 속하는 모든 객체는 동일한 신뢰도를 가지고 있다. 그러나 실제의 경우에는 클러스터간의 명확한 경계를 가지기 어렵기 때문에 클러스터내의 객체들의 불확실성을 처리해야 할 필요성이 있다.

퍼지 K-means에서는 모든 객체가 각 클러스터와 연관이 있으며 Krishnapuram과 Keller[8]는 확률을 이용하여 임의의 클러스터에 대한 객체의 멤버쉽 값이 다른 클러스터에 영향을 미치지 못하면 멤버쉽을 보정하는 방법을 제안하였다. Kim외[9]는 퍼지무게중심법을 이용하여 퍼지 K-means를 확장하여 퍼지K-modes에서 범주형데이터를 사용할 수 있도록 하였다. 이러한 방법은 멤버쉽 값의 초기설정등과 같은 부수적인 부담을 가지고 있다. 따라서 클러스터링 과정에서 발생하는 불확실성을 처리하기 위한 러프집합을 이용한 방법들이 보다 안정적인 결과를 보이고 있다.

2. 러프집합의 개념

지식에 관한 이론에서의 기본적인 개념은 분류(classification)와 범주(category)이다. 범주는 지식으로 표현될 수 있는 객체(object)들로 구성된 집합으로 애매성을 포함하고 있다. 또한 전체집합 U의 부분집합이 X이고, R을 동치관계(equivalence relation)라고 할 경우에 지식기반 K=(U, R)가 가능하다. X가 어떤 R-기본 범주들의 합집합이면 X가 R-정의 가능하다(R-definable)라고 하고, 그 외는 R-정의 불가능하다(R-undefinable)라고 하고 이와 같은 집합을 R-러프집합(R-rough set)이라고 한다. R∈IND(K)에 대해서도 X가 R-러프하면 X는 K내에서 러프하다(rough in K)라고 한다.

표 1. 걷기 결정표
Table 1. Decision table of walking

객체(U)/속성(A)	나이	렌즈	걷기
x_1	16-30	50	예
x_2	16-30	0	아니오
x_3	31-45	1-25	아니오
x_4	31-45	1-25	예
x_5	46-60	26-49	아니오
x_6	16-30	26-49	예
x_8	46-60	26-49	아니오

불확실성을 다루는 방법으로 퍼지집합과 러프집합이 있다. 두 방법은 불확실한 부분에 대해 다루는 방법이나 그 의미하는 영역이 다르다. 퍼지집합은 불확실한 영역에 대한 부분은 멤버쉽 함수로 나타내며, 러프집합은 식별하기 어려운 부분에 대한 내역과 근사치에 대한 접근법으로 그 값을 나타내는 방법적인 차이가 있다.

간단한 예로, 표 1에서 집합 A=(U, A)을 가지고 있는 정보 시스템에서 B⊆A라는 관계가 형성되는 표현이 있다면 이것은 INDA(B) 형식으로 표현할 수 있다. INDA(B)는 INDA({Age})={{x_1,x_2,x_6}, {x_3,x_4}, {x_5,x_7}}, INDA({LEMS}) = {{x_1}, {x_2}, {x_3,x_4}, {x_5,x_6,x_7}}, INDA({Age,LEMS}) = {{x_1}, {x_2}, {x_3,x_4}, {x_5,x_7}, {x_6}} 가 된다.

한편, 지식기반 K=(U, R)에서 U의 부분집합 X가 주어졌을 때 이 집합 X를 정보시스템 내의 속성 요인 R의 동치관계 R∈IND(K)를 이용하여 X의 하한근사(lower approximation)인 $\underline{R}(X)$와 $\overline{R}(X)$에 해당하는 상한근사(upper approximation)로 정의된다.

$$\begin{aligned}&\underline{R}(X)=\{x\in U|[x]_R\subseteq X\}\\&\overline{R}(X)=\{x\in U|[x_i]_R\cap X\neq\varnothing\}\\&BN_R(X)=\overline{R}(X)-\underline{R}(X)\end{aligned}\quad(1)$$

임의의 범주집합이 부정확한 것은 경계영역에 있기 때문이고 집합의 경계영역이 커질수록 그 집합의 분류의 정확성은 낮아진다. 이러한 개념을 숫자로 표현되는 정확성 척도(accuracy measure)로 나타낼 수 있다.

$$\alpha_R(X)=\frac{|\underline{R}(X)|}{|\overline{R}(X)|}\quad(2)$$

|X|인 X의 절대값에 대한 표현은 X≠0이고 분명히 0 $\leq\alpha_R(X)\leq1$, X는 I에 대해 크리스프(crisp) 하다. 다른 표현으로, $\alpha_R(X)<1$ 이면 X는 R에 대해 러프(rough)하다.

Ⅲ. MMMR 분할 알고리즘

Mazlack는 러프집합을 이용하여 속성에 대한 분류를 위한 클러스터링을 제안하였으며 전체적인 러프정도(total roughness)를 이용하여 각 분할의 크리스프정도(crispness)를 결정하였다.[10] 그러나 일반적으로 속성에 대한 분할은 속성의 값이 두 개인 경우에서 출발하여 그 이상의 값들로 진행한다. 따라서 모든 속성들에 대하여 분할을 수행함으로써 이러한 단점을 극복하고자 한다. MMMR은 엔트로피의 개념을 기반으로 범주형 데이터 간의 유사성을 측정하는 새로운 척도를 정의한다. MMMR은 Beaubouef가 제안한 러프집합의 불확실성에 대한 정보이론적인 척도와 비교하여 러프 엔트로피(rough entropy)라는 척도를 이용하였다. $a_i\in A$, $V(a_i)$는 속성 a_i의 값의 집합이고, X는 속성의 특정한 값이 α인 객체들의 집합 $X(a_i[\alpha])$일 경우에 $RE_{Y|a_j}(X|a_i[\alpha])$는 러프 엔트로피에 의해서 $\{a_j\}$에 대한 X의 러프 엔트로피에 대한 정의는 식(3)과 같다.

$$RE_{Y|a_j}(X|a_i[\alpha])=-Q_j\log_2\frac{|X\cap Y_{a_j}|}{|Y_{a_j}|},\quad a_i,a_j\in A,a_i\neq a_j\quad(3)$$

Q_j는 전체집합 U에 대한 동치류 j의 확률이고, $|X\cap Y_{a_i}|/|Y_{a_i}|$는 동치류 j의 객체에 대한 객체 $\{a_i\}$의 확률이다. 또한 임의의 속성간의 유사성을 구하기 위하여 동치류간의 관계에서 가장 유사성이 많은 동치류는 러프 엔트로피 값이 작다. 따라서 임의의 동치류간의 유사성인 min-roughness(MR)은 식(4)에 의하여 구할 수 있다.

$$MR_{a_j}(a_i[\alpha_1])=min(RE_{Y|a_j}(X|a_i[\alpha_1])\quad(4)$$

임의의 속성 $\{a_j\}$에 대하여 임의의 속성 $\{a_i\}$와의 유

사성인 mean-roughness(MeR)는 식(5)에 의하여 구할 수 있다.

$$MeR_{a_j}(a_i) = \frac{MR_{a_j}(a_i[\alpha_1]) + .. + MR_{a_j}(a_i[\alpha_{|V(a_i)|}])}{|V(\alpha_i)|} \quad (5)$$

임의의 속성 $\{a_j\}$에 대하여 모든 속성 $\{a_i\}$와의 유사성인 mean-mean-roughness(MMR)은 식(6)에 의하여 구할 수 있다.

$$MMR_{a_j}(a_i) = \frac{MeR_{a_j}(a_i[\alpha_1]) + . + MeR_{a_j}(a_i[\alpha_{|V(a_i)|}])}{|V(\alpha_i)|} \quad (6)$$

최종적으로 n개의 속성이 주어질 경우에 가장 중요한 속성인 MMMR은 식(7)에 의하여 구할 수 있다.

$$MMMR(a_i) = \min_{a_i \in A, i \in [1, card(A)]}(MMR(a_1), ..., MMR(a_i)) \quad (7)$$

러프 엔트로피가 작을수록 클러스터링의 러프정도가 높다. MR은 각각의 속성이 가질 수 있는 가장 우수한 러프니스를 나타낸다. 따라서 MMMR은 다른 속성을 대표하는 가장 중요한 분할속성(clustering attribute)을 나타낸다. 이러한 원리를 반복적으로 적용하여 범주형의 데이터를 원하는 수로 분할할 수 있다.

MMMR을 이용하여 분할속성을 결정하는 예를 표 2에서 고려한다. 속성의 범주는 나이, 렌즈, 근력, 비만, 걷기이다.

표 2. 걷기의 의사결정표
Table 2. Decision table of walking

U/A	나이	렌즈	근력	비만	걷기
1	16-30	26-49	아니오	양호	못함
2	16-30	1-25	예	불량	잘함
3	31-45	0	아니오	불량	못함
4	31-45	1-25	아니오	양호	못함
5	46-60	26-49	예	불량	잘함
6	46-60	0	아니오	불량	잘함
7	46-60	0	예	양호	못함
8	31-45	0	예	양호	잘함

첫째로, 가장 신뢰도가 높은 속성을 추출하기 위하여 지식의 속성에 대한 동치류를 조사한다. 둘째로, 식(3)을 이용하여 각각의 속성간의 의존성의 관계를 계산한다. 가장 신뢰도가 높은 속성을 군집화를 수행하는 기준이 되는 속성으로 간주하게 된다. 예를 들어 '나이' 속성에 대하여 '근력' 속성의 러프 엔트로피는 식(3)에 의해 구할 수 있다. 셋째로, 식(3)에 의하여 계산된 엔트로피는 속성간의 불확실성을 나타내기 때문에 이러한 값에서 가장 작은 값을 취함으로써 가장 신뢰도가 높은 속성간의 러프정도는 식(4)에 의하여 구할 수 있다. MRE('근력'='예') = min(0.173,0.415,0.152) = 0.152, MRE('근력'='아니오')| = min(0.173,0.152,0.415) = 0.152

넷째로, 다섯 가지의 속성에 대하여 하나의 속성의 평균 불확실성()의 평균을 식(5)에 의하여 구할 수 있다. 즉, '나이' 속성에 대하여 '근력' 속성의 평균 러프정도 다음과 같이 구해지고, V(ai)는 '예'와 '아니오'로 2가 된다. 표 3에서 가로의 '근력'과 세로의 '나이'의 중복성은 0.152로 결정 된다. MeR('근력') = ('나이'(X|'인성' = '아니오') + '나이'(X|'근력' = '예')) / V'나이' = (0.152 + 0.152) / 2 = 0.152

다섯째로, 결국 '나이' 에 대한 '근력' 속성의 전체적인 불확실성은 0.152로 나타났다. 이러한 방법으로 모든 속성간의 관계가 표 3에 나타나 있다. 또한 식(6)를 이용하여 다섯 가지의 속성에서 발생하는 신뢰도의 평균 러프정도 MMR을 구할 수 있고, 결국 주어진 모든 속성에서 분할속성이 식 (7)에 의해 MMMR이 정의되고 모든 속성의 분할 속성이 결정된다.

표 3. 걷기에 대한 군집화 결과
Table 3. Clustering of walking

속성	Rough Entropy Roughness					평균
	나이	렌즈	근력	비만	걷기	
나이		0.1733	0.4621	0.4621	0.4621	0.3899
렌즈	0.1662		0.5776	0.5776	0.5776	0.4748
근력	0.1520	0.1733		0.3466	0.1438	0.2039
비만	0.1520	0.1733	0.3466		0.1438	0.2039
걷기	0.1520	0.1733	0.1438	0.1438		0.1533

표 3에서 알 수 있듯이 속성에 대한 정확도가 가장 높은 속성은 '걷기' 이다. 따라서 이 속성은 지식 데이터를 분류하기 위한 분할속성으로 간주하게 된다. 추출된 '걷

기' 속성을 기준으로 객체를 분할하게 된다. 결국 '걷기' 속성에 의한 지식의 동치류는 U/'걷기' = {{1,3,4,7}, {2,5,6,8}}이기 때문에 U의 지식을 {1,3,4,7}와 {2,5,6,8}으로 분할 할 수 있게 된다. 러프엔트로피를 이용한 방법의 경우 평균값이 가장 작은 값이 신뢰도가 가장 양호가기 때문에 분할속성으로 검출되었다.

표 4에 러프엔트로피를 기반으로 분할정복을 이용한 MMMR 분할 알고리즘이 나타나 있다. 알고리즘은 두 개의 함수로 이루어져 있다. 첫 번째 함수에서 보면 7번 줄에서 각각의 속성에 대하여 동치류 관계를 이용하여 동치류 클래스를 계산하고, 9번 줄에서는 모든 속성 a_j에 대하여 $i \neq j$경우의 a_j의 러프 엔트로피를 계산한다. 10번 줄에서 모든 a_j에 대하여 $i \neq j$경우의 a_i의 평균 정확도를 계산한다. 그리고 12번 줄에서 속성에 대하여 최소 평균 정확도를 이용하여 최종적인 분할속성을 결정한다. 13번 줄에서는 이 분할속성을 이용하여 자료를 이분 분할(binary split)을 수행한다.

표 4. MMMR 분할 알고리즘
Table 4. MMMR clustering algorithm

알고리즘 : Min-Mean-Mean-Roughness(MMMR)

```
1:  Input:
2:  CNC=1, U:Parentnode,
3:  Loop1:
4:  If CNC<K And CNC≠1
5    Then  Parentnode = ParentNode(CNC)
6     End If
7:  For All Ai∈A(i=1,,,N)  Do
8:      Determinate [Xi]IND(Ai)
9:      For All Ai∈A Do
10:         Calculate RE(Aj(Ai))
11          End For
12:     MeR(Ai)=Mean(RE(Aj(Ai))
13      End For
14: MMeR=Min(MeR(Ai) // i=1,,,N
15: Determine Splitting Attribute Ai
16: Do Binary Split by Ai
17: CNC = No of Leaf Nodes
18: Go To Loop 1
19: ParentNode(CNC)
20    i=1
21: Do Until i<CNC
22    Size(i) = Count(Elements of Cluster  i)
23       i=i+1
24    Loop
25: Determine Max(Size(i))
26: Return (Set Of Elements In Cluster i)
```

다음에 현재노드의 수인 CNC는 단말노드의 개수가 되고 Loop1으로 분기하여 이분 분할을 계속하여 원하는 수의 분할까지 수행한다. 두 번째의 함수에서는 19번 줄에서 이분 분할을 통하여 얻은 두 개의 단말 노드에 대하여 크기가 큰 노드를 부모노드로 결정해준다. 반복적인 이분 분할과정에서 분할클래스들은 가장 작은 러프 엔트로피를 가지고 있다. 이 값은 러프 엔트로피가 안정적이라는 의미이고 즉, 속성의 신뢰도가 높다고 할 수 있다.

Ⅳ. 실험 및 결과

제안된 알고리즘의 성능을 알아보기 위하여 Matlab Ver. 2010a를 이용하여 시스템을 구축하여 UCI 기계학습 저장소(Machine Learning Reopsotory)에서 추출한 ZOO 데이터를 가지고 실험을 수행하였다. ZOO데이터의 7개의 결정속성의 클러스터들에 대하여 분할된 클러스터들의 결과를 비교하기 위하여 i분할의 분할비(partition ratio)는 i분할의 최대 데이터수를 i분할의 전체 데이터 수로 나눈 것이다. 따라서 전체 분할비는 각 분할의 최대 데이터수의 합을 전체 데이터 수로 나눈 값이다. 표 5에 분할결과를 나타내었다. 각각의 분할에서 음영부분이 가장 큰 분할을 나타낸다. 전체 분할비가 높을수록 분할이 양호하다는 것을 나타낸다.

표 5. ZOO 데이터의 분할순도
Table 5. Purity of clusters in ZOO data

분할수	분할1	분할2	분할3	분할4	분할5	분할6	분할7	분할비
1(14)	0	0	1	13	0	0	0	0.928
2(9)	2	0	3	0	4	0	0	0.444
3(15)	0	0	1	0	0	4	10	0.666
4(20)	0	20	0	0	0	0	0	1
5(10)	6	0	0	0	0	4	0	0.6
6(17)	17	0	0	0	0	0	0	1
7(16)	16	0	0	0	0	0	0	1
전체 분할비					86/101=0.8514			

표 6. MMMR 알고리즘의 분할순도 비교
Table 6. Purity comparison of MMMR

번호	방법	분할순도
1	K-means	0.6
2	Fuzzy K-means	0.64
3	Fuzzy Centroids	0.75
4	SDR	0.7821
5	SSDR	0.7821
6	제안된 방법	0.8514

따라서 분할순도(partition purity)가 1이라는 것은 완벽한 분할을 의미한다. 범주형 데이터에 존재하는 불확

실성을 처리하는 퍼지논리를 기반으로 구축된 알고리즘은 K-modes, 퍼지 K모드(fuzzy K-modes), 퍼지 무게중심법(fuzzy centroids), SDR와 SSDR가 있다. 표 6에 기존 방법과 대비하여 분할순도를 나타내었다.

V. 결 론

일반적으로 범주형 데이터에서는 속성의 개수가 많고 그 값이 다양하기 때문에 하나의 객체가 여러 개의 집단으로 분류되는 불확실성으로 인하여 기존의 알고리즘에서는 애매함을 완전하게 처리하지 못하고 있는 실정이다. 따라서 범주형 데이터를 구성하는 속성간의 변별력을 향상시키기 위하여 새로운 러프 엔트로피척도를 정의하였다. 또한 정의된 러프 엔트로피 척도를 기반으로 데이터를 분할하는 알고리즘을 구현하여, ZOO 데이터의 분할결과에 대한 기존방법과의 비교우위에서 만족할 만한 결과를 보였다.

References

[1] Pawlak, Z. “Rough sets”, International Journal of Information and Computer Sciences, Vol.11,No. 5, pp. 341-356, 1982

[2] Beaubouef, T., Petry, F. E. and Arora, G.,, “Information-theoretic measurtes of uncertainty for rough sets and rough relational databases”, Information Science, Vol. 109, No. 1-4, pp. 185-195, 1998.

[3] Wojciech Ziarko, “Variable Precision Rough Set Model”, June 1, 1990 August 1, 1991

[4] A. Dempster, N. Laird, D. Rubin, “Maximum likelihood form incomplete data via the EM algorithm”, Journal of the Royal Statistical Society Vol. 39(1), pp. 1-38, 1997

[5] H. Ralambondrainy, “A Conceptual Version of the K-means Algorithm, Pattern Recognition Letters, Vol. 16, No. 11, pp. 1147-1157, 1995

[6] H.T. Lee et al., “AED System using Fuzzy Rules“, The Institute of Internet, Broading and Communication, Vol 13, No. 4, Aug. 2013
Z. Huang, “Extensions to the k-means algorithm for clustering large data sets with categorical values”, Data Mining and Knowledge Discovery, Vol. 2, pp. 283-304, 1998

[7] S. Guha, R. Rastogi, K. Shim, Information Systems, Vol. 25, pp. 345-366, 2000

[8] R. Krishnapuram, J. Keller, IEEE Transactions on Fuzzy Systems, Vol. 1, pp. 98-110, 1993

[9] J. Y. Kim, S S. Jo , K.K. Kim , S. H. Choi, Development of Localization and Three-dimensional hull map creation S/W for Underwater robot, Journal of Korean Institute of Information Technology, Vol.8 No.6 ,35-40, June 2010

[10] J. E. Chung, J. K. Ahn, A Study of Robust Design of FCM Gasket Using Taguchi Method, Journal of the Korea Academia-Industrial cooperation Society, v.14, no.7, 3177-3183, July 2013

저자 소개

박 인 규(정회원)

- 1985년 : 원광대학교 전기과(학사)
- 1987년 : 연세대학교 전기과(공학석사)
- 1997년 : 원광대학교 전자과(공학박사)
- 1997년 ~ 현재 : 중부대학교 컴퓨터학과 교수

<관심분야: 소프트 컴퓨팅, 러프집합>