퍼지의사결정나무 개선방법을 이용한 CRM 적용 사례

Case Study of CRM Application Using Improvement Method of Fuzzy Decision Tree Analysis

양승정, 이종태
동국대학교 산업시스템공학과

Seung-Jeong Yang(yeod@dongguk.edu), Jong-Tae Rhee(jtrhee@dongguk.edu)

요약

의사결정나무는 대량의 데이터를 몇 개의 집단으로 분류하고, 미래상황을 예측하기 위해 자주 사용되는 분석기법 중의 하나이며, 각 노드에서 분할이 일어나면서 자라게 되고, 각 노드에 속하는 자료의 순수도가 효과적으로 증가하도록 진행된다. 또한 의사결정나무를 생성하는 과정에서 필요 이상의 가지(leaves)를 갖게 되면 노드의 분할을 정지하거나, 분류성능 향상에 큰 도움이 되지 못하는 가지를 잘라내게 된다. 이러한 가지치기의 결과로 의사결정나무의 형태가 변하게 되는데 이는 기존의 가지분할이 효율적이지 않았음을 의미하는 것이다.

본 연구에서는 가지치기의 교정뿐 아니라 새로운 분할과정을 혼합한 우수한 의사결정나무 추출 방법을 제안한다. 특히, 새로운 분할 노드의 선택에 있어 퍼지이론을 적용하여 분할의 효과성을 제고할 수 있는 방법을 제시하고자 한다.

■ 중심어 : | 고객관계관리 | 의사결정나무분석 | 퍼이지론 | 데이터마이닝 |

Abstract

Decision tree is one of the most useful analysis methods for various data mining functions, including prediction, classification, etc, from massive data. Decision tree grows by splitting nodes, during which the purity increases. It is needed to stop splitting nodes when the purity does not increase effectively or new leaves does not contain meaningful number of records. Pruning is done if a branch does not show certain level of performance. By pruning, the structure of decision tree is changed and it is implied that the previous splitting of the parent node was not effective. It is also implied that the splitting of the ancestor nodes were not effective and the choices of attributes and criteria in splitting them were not successful. It should be noticed that new attributes or criteria might be selected to split such nodes for better tries. In this paper, we suggest a procedure to modify decision tree by Fuzzy theory and splitting as an integrated approach.

■ keyword : | CRM | Decision Tree | Fuzzy Theory | Data Mining |

접수번호 : #070423-001
접수일자 : 2007년 04월 23일
심사완료일 : 2007년 06월 26일
교신저자 : 양승정, e-mail : yeod@dongguk.edu

I. 서 론

CRM이란 고객과 관련된 자료를 분석하여 고객특성에 기반을 둔 마케팅 전략을 계획하고, 실시 · 재평가하기 위한 일련의 프로세스를 말한다. 메타그룹에 의하면 CRM은 분석 CRM, 운영 CRM, 협업 CRM으로 나뉘어진다[1].

이 중 분석 CRM에서 가장 중요한 부분을 차지하는 데이터 마이닝은 데이터에서 중요한 정보를 추출하는 기법을 말한다. 그 중 가장 많이 사용되는 기법으로 의사결정나무 분석이 있다. 의사결정나무는 관심대상이 되는 집단을 몇 개의 소집단으로 분류하거나 또는 예측 기법으로 사용된다. 이러한 의사결정나무 기법의 가장 큰 장점은 의사결정규칙이 나무구조로 표현되기 때문에 다른 기법들에 비해서 쉽게 이해되고, 설명이 용이하다는 점이다. 반면 의사결정나무의 단점은 생성되는 규칙이 반드시 입력공간에서 수직적인 경계면을 형성한다는 것이다. 또한 연속형 변수의 값을 이산적 형태로 취급하기 때문에 경계면 부근에서의 분류오류가 발생할 위험성이 커지게 된다.

본 논문에서는 수직적인 경계면을 평활화(smoothing)할 수 있을 퍼지이론을 이용하여 상기한 문제를 보다 효율적으로 극복할 수 있는 퍼지 의사결정나무 생성기법을 제시하고자 한다. 특히 각 분할단계에서의 변수 및 분할기준치를 선정하는 과정에서 퍼지함수 분석에 의한 의사결정나무 효율극대화 방안을 적용하고자 한다.

연속형 변수를 퍼지변수 값으로 변환하기 위한 퍼지 소속 함수값 도출을 위하여 설문을 실시했으며, 그 이유는 각 변수들의 가장 일반적이고, 대중적인 값을 얻기 위함이다. 또한 객관적인 소속함수 생성을 위하여 표준편차와 평균값, 최빈값, 최대값, 최소값 등을 이용하였다. 변수값의 퍼지 변환을 위해서 객체지향언어인 파이썬으로 퍼지값 자동 변환 프로그램을 구축하였다. 의사결정나무분석을 이용한 CRM 실행은 기존 CRM 툴인 Clementine 8.5를 활용하였다. 본 1장에서는 논문의 전체적인 소개를, 2장에서는 퍼지의사결정나무의 기존 연구를 살펴보며, 3장에서는 본 논문에서 제시하고자 하는 알고리듬과 사례에 대해 자세하게 설명한다. 마지막 4장에서는 적용결과와 추후 연구과제에 대해서 논한다.

II. 퍼지의사결정나무 분석

1. 기존 의사결정나무 구조와 분할알고리듬

의사결정나무는 1984년 Breiman, Friedman, Olshen, Stone이 의사결정 나무 역사상 가장 중요한 알고리듬인 CART(Classification and Regression Tree)를 구현하였으며, 이로써 의사결정 나무 분할규칙에 의한 성장, 가지치기 등이 정립되었다[2].

의사결정나무는 분류 · 예측의 목적으로 주로 사용되고 있다. 그러나 분석의 정확도보다는 분석과정의 설명이 필요한 경우에 더 유용하게 사용된다[3].

의사결정나무의 구조는 나무구조가 시작되는 뿌리노드(Root-node)와 하나 이상의 노드를 연결하는 부모노드(Parent-node), 여기에 연결된 자식 노드(child-node)들로 구성되어 있다. 각 노드는 각 변수들로 이루어져 있다. 더 이상 분할이 이루어지지 않는 노드를 종료 노드(또는 leaf)라고 부르며 이 노드의 값이 바로 클래스 값이 된다. 최종 노드의 수가 클래스의 수가 되며, 이것이 의사결정 규칙의 수가 된다[4].

본 논문에서 사용한 분할알고리듬인 CART알고리듬은 의사결정나무분석을 형성하는데 있어서 가장 보편적인 알고리듬이며 이진(binary)분할을 한다. 지니 지수(Gini Index) 또는 분산의 감소량을 이용하여 분할을 수행하는 알고리듬이다[5][6].

2. 퍼지의사결정나무 분석의 기존 연구

퍼지이론은 1965년에 캘리포니아 버클리대학의 Zadeh교수에 의해서 제창된 것으로서, 모호성을 다루는 시스템으로 넓게 사용되고 있다. 퍼지이론의 대상이 되는 프로세서는 종래의 방법에서는 모델링이 어려운 것으로써, 「비선형」, 「시계열」, 「확률적」 인 성질을 가지고 있다[7].

1990년대 후반, 퍼지이론이 한참 각광을 받고 있을 때 퍼지이론을 이용한 의사결정나무 분석 방법에 대한

연구 또한 활발하게 진행되었으며, 자세히 살펴보면 다음과 같다.

Hall & Lande[12]는 4개의 변수와 6개의 데이터로 이루어진 테스트 데이터를 이용해 학습을 한 뒤, 화학공장 점화 제어와 메탄과 산소가 연소할 때 가스용광로에서 발생하는 CO_2를 예측하기 위하여 퍼지 의사결정나무 분석방법을 이용하였다. 10개의 입력 변수를 이용해 175개의 퍼지집합과 7개의 출력 퍼지집합을 구성하여 최종적으로 76개의 규칙을 생성하였다.

또한 Yuan[13] 등은 각 변수 속성값에 대해 퍼지집합을 정의한 다음 소속정도 정보를 이용하여 퍼지 의사결정나무를 생성하였다. 이 경우에서는 변수의 속성값이 퍼지집합들에 대한 소속정도로 주어지는 것을 미리 가정하고 있으며, 속성값 자체가 퍼지값인 경우는 고려하지 않았다.

Zeidler[14] 등은 연속적인 수치속성을 갖는 변수 데이터에 대한 퍼지 결정트리 생성방법을 제안하였으나, 변수값이 퍼지값이 아닌 일반값으로 처리하는 한계가 있다.

또한 이우향 & 이건명[15]은 공간을 사선으로 분할하는 퍼지결정트리 유도에 대해 제안하였다. 이때 사용된 데이터는 약 150개이며, 입력변수는 총 4개였다. 결과의 유효성을 위하여 다층신경망 분석방법과 비교하여 더 나은 결과를 보였다. 그러나 생성되는 분류경계면만으로 결과를 도출하기 때문에 내부적으로 표현된 분류규칙을 이해하기가 어렵다. 또한 결정트리가 생성되는 시간이 다소 긴 단점을 가지고 있다. 이와 비슷한 연구로 Lee & Kim[16]의 연구결과가 있다.

연속적인 값의 수치속성을 갖는 변수만을 퍼지화한 연구로는 이건명[17]의 결과가 있다. 의사결정나무 분석의 알고리즘으로 엔트로피 지수를 사용하였으며, 퍼지소속함수로는 사다리꼴 함수를 이용하였으나 소속함수에 대한 정확한 언급이 없으며, 범주형 속성을 갖는 데이터의 퍼지변환은 고려하지 않은 한계가 있다.

또한 Min 등은[10] 효율적인 퍼지 규칙 생성을 위하여 퍼지 결정 트리를 이용하였으며, 히스토그램에 기반한 퍼지소속함수를 생성하였다. 총 150개의 데이터를 실험데이터로 사용하였으며, ID3 알고리즘을 이용하여 의사결정나무를 구축하였다.

최근 연구를 살펴보면, Mendonça & Vieira & Sousa [18]의 연구로 의사결정나무를 생성하는 접근방법으로 퍼지 모델링을 이용하였다. 이 연구에서는 의사결정나무를 생성하기 위한 방법론으로 'top-down algorithm'과 'bottom-up algorithm'을 퍼지모델링을 이용하여 생성, 비교하였으며, 간단한 4가지 사례를 들어 검증하였다.

그리고 Wang & Nauck & Spott [19]는 퍼지의사결정나무를 이용하여 SPIDA(a Soft computing Platform for automatic Intelligent Data Analysis)의 프레임워크를 구성하였다. SPIDA는 도메인 전문가이지만 데이터 마이닝에는 비전문가인 사람들을 위해 개발된 데이터 자동 분석 툴이며, 데이터 처리와 데이터 분석 방법 선택, 전문가적인 인터페이스를 제공한다. Wand 등은 이러한 SPIDA의 지식베이스 구축 시 연속적인 수치값을 갖는 변수를 퍼지의사결정방법을 이용하여 구축하였다. 또한 자동차 표면의 품질제어를 위한 사례를 통하여 검증하였다. 이때 사용된 데이터의 수는 모두 273개이며, 15개의 변수를 이용하였다.

기존의 연구 결과를 살펴보면 다음과 같은 몇가지 문제점을 발견할 수가 있다.

1) 입력변수가 10개 미만이다.
2) 주로 기계학습 또는 시스템제어 분야에 응용되어 CRM에 실용적이지 못하다.
3) 입력 데이터의 수가 제한적이다. 이 또한 대량의 데이터가 사용되는 CRM에서는 효과적이지 못하다.
4) 퍼지 소속 함수 도출 방법이 명확하지 않으며, 일반적으로 응용하기에 매우 어렵다.
5) 알고리즘 개선에 초점을 맞추다 보니 비실용적인 데이터를 샘플링하는 경우가 있다.

이에 본 논문에서는 CRM에서 사용하는 데이터가 대용량의 실제 고객관련 데이터임을 고려하여, 퍼지의사결정나무 개선방법을 이용한 CRM의 적용사례를 보이고자 한다.

III. 제안 알고리듬과 적용 예

1. 제안 알고리듬

본 논문에서는 위에서 언급된 퍼지의사결정나무 분석방법의 문제점을 해결하기 위해서 다음과 같은 방법을 제안하고자 한다.

실제 온라인상에서 사용되고 있는 고객의 데이터를 이용하여 고객의 행동 예측을 위한 모형을 구축하고자 하며, 제안하고자 하는 방법은 다음과 같다.

단계 1 : 분석에 사용되는 데이터 중 연속형 값을 갖는 데이터의 속성을 파악한다.

단계 2 : 변수값을 퍼지화 하기 위해 퍼지 소속함수를 정의한다. 제안 퍼지소속함수 도출 방법은 다음과 같다.

1) 각 변수의 일반적인 값 도출을 위하여 설문조사를 실시한다.
2) 각 변수별로 변수값에 대한 분포도를 도식한다.
3) 이상치 선별을 위하여 최대값과 최소값을 탈락시킨다.
4) 표준편차 $\sigma(x)$를 구한다.
5) 소속함수의 최대값($\mu(x)=1$)과 최소값 ($\mu(x)=0$)을 다음과 같이 구한다.

(이때, a는 최소값, b는 최대값, $\bar{x}$는 평균, x_M은 최빈값을 나타내며, [그림 1]은 일반적인 퍼지 소속함수를 나타내고 있다.)

① $\mu(x)=0$ 인 경우

if $\sigma(x)-\bar{x} < a$ then $x_1=a$

otherwise $x_1=\sigma(x)-\bar{x}$

if $\sigma(x)+\bar{x} > b$ then $x_4=b$

otherwise $x_4=\sigma(x)+\bar{x}$

② $\mu(x)=1$ 인 경우

if $\bar{x} = x_M$ then $\mu(\bar{x}=x_M) = 1$

(-> 삼각형함수)

if $\bar{x} \neq x_M$ then $\mu(\bar{x})=\mu(x_M)=1$

(-> 사다리형 함수)

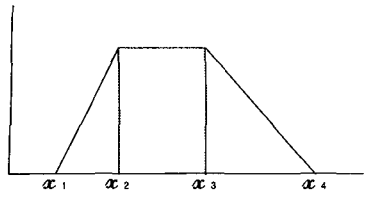


그림 1. 퍼지소속함수

단계 3 : 소속함수를 정의한 뒤, 변수값을 퍼지화 하여 해당 변수값의 퍼지 집합군을 형성한다. 이때 객체지향언어인 파이썬을 이용하여 퍼지값 자동 변환 프로그램을 완성하였으며, 연속형 값을 갖는 대량의 변수를 퍼지화 할 수 있게 된다.

단계 4 : CART에서 사용되는 지니 지수를 이용하여 분할기준을 정한다. 정해진 분할 기준을 이용하여 퍼지의사결정나무를 구성한다.

단계 5 : 의사결정나무를 모델링하기 위해서 클레멘타인 툴을 사용한다.

단계 6 : 최종적으로 구축된 모형의 검증을 위하여 정오분류표를 이용한다.

2. 사례를 이용한 고객행동 예측 모형 구축

인터넷 쇼핑몰에 접속하여 로그인 하는 고객들의 데이터를 분류하여 고객들의 흡연여부를 예측하고자 한다. 고객 흡연 여부 예측결과에 따라 대고객 캠페인과 이벤트 전략을 기획할 수 있다. 실험에 사용된 실험데이터는 2,000여개로, 고려된 변수는 연령대, 주소, 흡연여부, 학력, 결혼여부, 관심분야, 직업, 소득, 음주여부이다. 이 중 흡연여부를 목표변수로 사용하였다. [표 1]은 실험에 사용된 원 데이터의 일부를 나타내고 있다.

표 1. 실험 데이터

번호	나이	주소	직업	결혼	학력	관심분야	소득	음주여부	담배여부
3405	34	제주	경영/관리직	기혼	전문대/대졸	여행	100~150만원	마신다	안피운다
3400	34	경기	경영/관리직	미혼	전문대/대졸	결혼	100~150만원	마신다	안피운다
3707	37	서울	경영/관리직	기혼	대학원 재학	여행	100~150만원	마신다	피운다
3099	30	서울	경영/관리직	미혼	대학원 재학	결혼	100~150만원	마신다	안피운다

3900	39	광주	경영/관리직	미혼	전문대/대졸	창업	100~150만원	마신다	피운다
3300	33	서울	경영/관리직	미혼	전문대/대졸	결혼	100~150만원	안마신다	안피운다
4101	41	인천	경영/관리직	기혼	전문대/대졸	재테크	100~150만원	마신다	안피운다
3407	34	경기	경영/관리직	기혼	전문대/대졸	여행	100~150만원	마신다	안피운다
3706	37	대구	경영/관리직	미혼	전문대/대졸	패션	100~150만원	안마신다	안피운다
3303	33	서울	경영/관리직	기혼	전문대/대졸	금융/투자	100~150만원	마신다	피운다
3503	35	경기	경영/관리직	기혼	전문대/대졸	컴퓨터	100~150만원	마신다	안피운다
3298	32	서울	경영/관리직	미혼	전문대/대졸	금융/투자	100~150만원	마신다	피운다
4304	43	서울	경영/관리직	기혼	대학원 졸이상	교육	100~150만원	마신다	피운다
3401	34	경남	경영/관리직	기혼	전문대/대졸	교육	100~150만원	마신다	안피운다
4504	45	경남	경영/관리직	기혼	전문대/대졸	컴퓨터	100~150만원	마신다	피운다
...	...	...	...	...	...	...	...	...	...

본 논문에서는 연속적인 값을 갖는 변수 즉, 입력데이터에서 나이에 해당하는 속성 값을 퍼지소속함수를 이용하여 퍼지값으로 재구성한 뒤 분할기준을 결정하였다.

퍼지소속함수 도출을 위하여 설문조사 방법을 이용한 것은 각 해당 변수의 가장 일반적인 값의 범위를 파악할 수 있기 때문이다. 또한 좀 더 객관적인 소속함수 도출을 위하여 평균값과의 분포정도와 최빈값, 최대, 최소값을 이용하였다. 연속형 변수를 퍼지값으로 변환하기 위하여 객체지향언어인 파이썬을 이용하여 자동퍼지값 변환 프로그램을 구축하였다.

위와 같은 방법으로 도출한 변수 '나이'의 퍼지소속함수는 [그림 2]와 같다.

또한 위의 퍼지소속함수를 이용하여 퍼지화한 '나이' 값의 일부를 [표 2]에서 나타내고 있다.

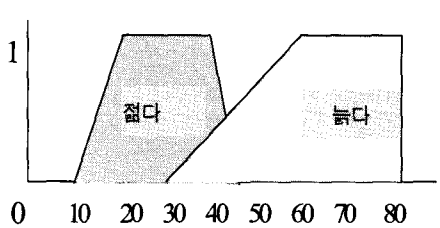


그림 2. 나이의 퍼지소속함수

표 2. 변수 '나이'의 퍼지화

번호	나이
3405	{(young:0.99),(old:0.1333)}
3400	{(young:0.99),(old:0.1333)}
3707	{(young:0.96),(old:0.2333)}
3099	{(young:17)}
3900	{(young:0.95),(old:0.3)}
3300	{(young:0.99),(old:0.1)}
4101	{(young:0.875),(old:0.366)}
3407	{(young:0.99),(old:0.1333)}
3706	{(young:0.96),(old:0.2333)}
3303	{(young:0.99),(old:0.1)}
3503	{(young:0.98),(old:0.1667)}
3298	{(young:0.625),(old:0.433)}
4304	{(young:0.99),(old:0.066)}
3401	{(young:0.375),(old:0.5)}
4504	{(young:0.99),(old:0.1)}
...	.....

다음 단계로 CART 알고리즘을 이용하여 퍼지의사결정나무를 설계한다. 목적변수는 흡연여부가 되며, 고객행동예측 모형의 결과로 새로운 고객의 흡연 여부를 예측하려고 한다. 예측된 결과를 대고객 마케팅 정보로 사용하고자 한다. 이때 사용한 데이터마이닝 툴은 SPSS의 클레멘타인 8.5이다.

3. 결과에 따른 효율성 검증

[그림 3]은 퍼지변수를 이용하여 형성된 의사결정나무를 나타내고 있다.

또한 퍼지변수를 이용하여 구축된 고객행동(흡연여부) 예측 모형의 정확도를 비교하기 위해 [그림 4]와 같이 퍼지 변수에 의한 흡연여부 예측 모형의 정오분류표를 정리하였다.

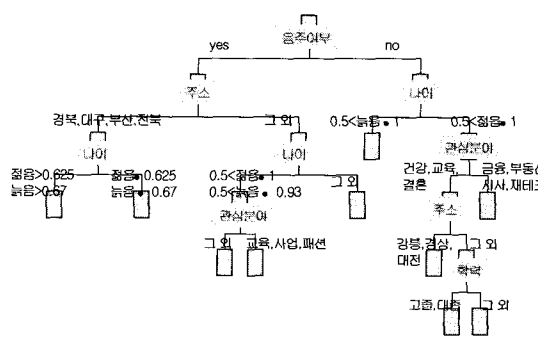
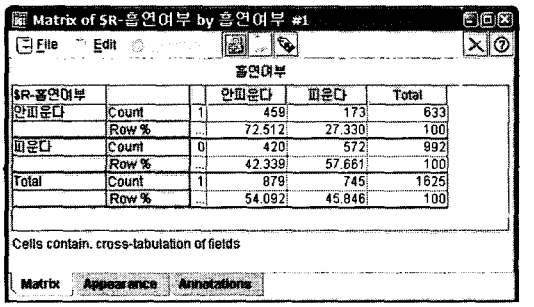


그림 3. 퍼지변수를 이용한 의사결정나무

정오분류율(hit ratio)은 입력변수의 값을 갖고 각 사례의 실제 결과를 예측하는 분류적 예측 모형의 성능 평가에 가장 일반적으로 사용되어온 방법이다[20]. 정오분류표란 실제 분류값과 예측값을 이용하여 예측오류와 예측의 정확도를 비교하는 표를 말한다.


Matrix of $R-흡연여부 by 흡연여부 #1

File Edit

흡연여부

$R-흡연여부			안피운다	피운다	Total
안피운다	Count	1	459	173	633
	Row %	...	72.512	27.330	100
피운다	Count	0	420	572	992
	Row %	...	42.339	57.661	100
Total	Count	1	879	745	1625
	Row %	...	54.092	45.846	100

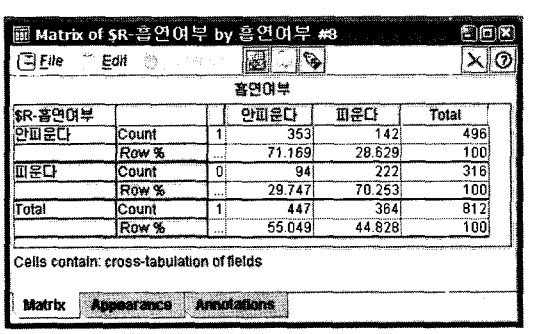
Cells contain: cross-tabulation of fields

Matrix Appearance Annotations

그림 4. 퍼지 변수에 의한 흡연여부 예측 모형의 정오분류표

이 정오분류표를 이용하여 퍼지변수를 이용한 예측 모형의 실제 흡연/비흡연 예측률의 정확도를 비교하였다. 퍼지 변수를 이용한 경우, 흡연예측률은 57.6%이고, 예측율 범위안에 포함되는 경우는 전체 1625명 중 총 339명이며, 이 중 실제 흡연을 하는 경우는 262명으로 예측의 정확도는 77%가 된다.

또한 비흡연의 예측률은 72.5% 이고, 이 범위안에 포함되는 경우는 총 1625명 중 311명이 된다. 이 중 실제 담배를 피지 않는 경우는 291명으로 예측 모형의 정확도는 93%가 된다. 정확도의 좀 더 구체적인 평가를 위하여 전체 데이터 중 50%를 샘플링 하여 흡연여부 예측 모형을 새롭게 구축하였다. [그림 5]는 그 결과로 나온 정오분류표를 나타내고 있다.


Matrix of $R-흡연여부 by 흡연여부 #8

File Edit

흡연여부

$R-흡연여부			안피운다	피운다	Total
안피운다	Count	1	353	142	496
	Row %	...	71.169	28.629	100
피운다	Count	0	94	222	316
	Row %	...	29.747	70.253	100
Total	Count	1	447	364	812
	Row %	...	55.049	44.828	100

Cells contain: cross-tabulation of fields

Matrix Appearance Annotations

그림 5. 퍼지 변수를 이용한 예측 모형 테스트 결과

이 결과 흡연 예측율은 70%이며, 예측 범위 안에 포함되는 경우는 샘플링 숫자인 812명 중 총 76명이다. 이 중 실제 흡연인 경우는 57명으로 정확도는 75%가 된다. 또한 비흡연예측율은 71.16%이며, 이 범위 안에 포함되는 경우는 50% 샘플링한 숫자 812명 중 총 169명이고, 이 중 실제로 담배를 피우지 않는 경우는 159명으로 정확도는 94%이다. 이를 표로 정리한 것이 [표 3]이다.

표 3. 고객행동 예측 모형의 정확도

	퍼지 변수	
	전체 데이터	샘플링(50%)
흡연 예측율	57.6%	70.25%
예측 인원 중 실제 흡연율	77%	75%
비흡연 예측율	72.5%	71.16%
예측 인원 중 실제 비흡연율	93%	94%

Ⅳ. 결 론

의사결정나무는 이해가 쉽고, 규칙생성과정을 쉽게 설명할 수 있다는 장점이 있어 데이터 마이닝에서 가장 많이 사용되고 있다. 또한 분할 기준선에서의 오류를 줄이기 위하여 퍼지를 이용한 퍼지의사결정나무 분석 방법이 활용되고 있다. 그러나 기존 퍼지의사결정나무 분석의 경우, 기계학습 등의 사례에 주로 응용이 이루어졌으며, 10개 미만의 입력 변수와 연속적인 수치 데이터를 다루었다. 그러한 이유로 CRM이나 마케팅 시장에서 다루게 되는 대량의 데이터를 다루게 될 경우 한계를 보이게 된다.

이에 본 논문에서는 CRM 시스템에서 다루게 되는 대량의 연속형 변수를 퍼지화하기 위해서 새로운 퍼지 소속함수 도출 방법을 제시하였다. 제안 퍼지소속함수 도출 방법을 이용하여 개선된 퍼지의사결정나무를 구축하였다. 그리고 제안 방법의 검증을 위하여 CRM에서의 고객행동 예측 모형을 구축하였다. 이를 위하여 온라인상에서 실제 사용하는 고객데이터를 이용하였다. 또한 새로운 퍼지소속함수 도출 방법을 이용하여

퍼지의사결정나무를 설계하였다. 예측 모형의 정확도를 분석하여 제안된 퍼지소속함수 도출 방법과 개선된 퍼지의사결정나무를 검증하였다.

추후연구과제는 다음과 같다.

첫 번째, 최적의 퍼지함수 값 도출을 위하여 퍼지값에 대한 민감도 분석 등을 실시 한 뒤 가장 적합한 변수의 값을 선택할 수 있도록 퍼지함수 값 도출에 대한 심도 있는 연구가 진행되어야 할 것이다.

두 번째, 웹 또는 유비쿼터스 환경에서 실시간, 대용량으로 모아지는 데이터를 처리하는 방법에 대해서 활발한 연구가 이뤄져야 한다.

세 번째, 의사결정나무 개선방법에 관해 다양한 관점에서의 연구가 이루어져야 한다.

네 번째, 연속형 값을 갖는 변수 뿐만아니라 범주형, 언어형 값을 갖는 변수에 대한 퍼지의사결정나무 분석에 대한 방법론이 연구되어져야 한다.

참 고 문 헌

[1] J. S. Oh, *The problems, the present conditions and the prospects for CRM market,* KISDI IT FOCUS, 2001.

[2] W. H. Jung, J. Jones, and J. Chen, *Optimization of the Decision Tree*, Proc.of the IEEE Int.Conf.on Tools for AI, San Jose, CA, 1991.

[3] H. C. Kang, S. T. Han, and J. H. Choi, "Interpretation of Data Mining Prediction Model Using Decision Tree," The Korean Communications in Statistics, Vol.7, No.3, pp.937-943, 2000.

[4] R. Fleischer, "Decision Trees: Old and New Results," Information and Computation 152, pp.44-61, 1999.

[5] C. Apte and S. Weiss, "Data mining with decision trees and decision rules," Future Generation Computer Systems 13, pp.197-210, 1997.

[6] H. Almuallim, "An efficient algorithm for optimal pruning of decision trees," Artificial Intelligence 83, pp.347-362, 1996.

[7] M. Friedman, T. B. Noy, M. Blau, and A. Kandel, "Certain Computational Aspects of Fuzzy Decision Trees," Fuzzy Sets and Systems 28, 1998.

[8] X. Wang, B. Chen, G. Qian, and F. Ye, "On the Optimization of Fuzzy Decision Trees," Fuzzy Sets and Systems 112, pp.117-125, 2000.

[9] B. Apolloni, G. Zamponi, and A. Zanaboni, "Learning Fuzzy Decision Trees," Neural Networks 11, pp.885-895, 1998.

[10] C. W. Min, M. Y. Kim, and S. K. Kim, "Efficient Fuzzy Rule Generation Using Fuzzy Decision Tree," Journal of the Korean Institute of Electronics Engineers, Vol.35-C, No.10, pp.59-68, 1998.

[11] K. M. Lee and W. H. Lee, "Fuzzy Decision Tree for Fuzzy Data with Numeric Attributes and Non -Numeric Attributes," Journal of the Research Institute for Computer and Information Communication Vol.7, No.1, 1999(5).

[12] L. O. Hall and P. Lande, "Generating fuzzy rules from data," Proceedings of IEEE 5th International Fuzzy Systems, Vol.3, pp.1757-1762, 1996.

[13] Y. Yuan and M. J. Shaw, "Induction of fuzzy decision trees," Fuzzy Sets and Systems, Vol.69, pp125-139, 1995.

[14] J. Zeidler and M. Schlosser, "Continuous -valued attributes in fuzzy decision trees," Proc. of the 6th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, pp.395-400, 1996.

[15] 이후향, 이건명, "특징공간을 사선 분할하는 퍼지 결정트리 유도", 정보과학회논문지:소프트웨

어 및 응용, 제29권, 제3호, pp.156-166, 2002.

[16] K. M. Lee and H. J. Kim, "Fuzzy Classification Rule Learning by Decision Tree Induction," International Journal of Fuzzy Logic and Intelligent Systems, Vol.3, No.1, pp.44-51, 2003(6).

[17] 이건명, "Classification Rule Mining from Fuzzy Data based on Fuzzy Decision Tree", 정보과학회논문지:소프트웨어 및 응용, 제28권, 제1호, pp.64-72, 2001.

[18] L. F. Mendonça, S. M. Vieira, and J. M. C. Sousa, "Decision tree search methods in fuzzy modeling and classification," International Journal of Approximate Reasoning 22, pp.106-123, 2007.

[19] X. Wang, D. Nauck, and M. Spott, "Intelligent data analysis with fuzzy decision trees," Soft Comput 11, pp.439-457, 2007.

[20] 김충영, 장남식, 김준우, "이동통신서비스 해지 고객 예측 모형의 비교 분석에 관한 연구", 경영정보학연구, 제12권, 제1호, pp.139-158, 2002.

저 자 소 개

양 승 정(Seung-Jeong Yang) 정회원

- 1994년 2월 : 서울산업대학교 산업공학과 (공학사)
- 1997년 2월 : 서울산업대학교 대학원 산업공학과 (공학석사)
- 2007년 8월 : 동국대학교 대학원 산업공학과 (공학박사)
- 2005년 9월 ~ 현재 : 동국대학교 산업기술연구원 전임연구원

<관심분야> : 고객관계관리(CRM), 퍼지이론, u-CRM

이 종 태(Jong-Tae Rhee) 정회원

- 1981년 2월 : 서울대학교 산업공학과 (공학사)
- 1983년 2월 : KAIST 산업공학과 (공학석사)
- 1990년 2월 : U.C Berkeley 산업공학과 (공학박사)
- 1992년 3월 ~ 현재 : 동국대학교 산업시스템공학과 교수

<관심분야> : SCM, CRM, 유비쿼터스 및 RFID System, Neural Network