

제품별 구매고객 예측을 위한 인공신경망, 귀납규칙 및 IRANN 모형*

정수미** · 이진호**

Artificial Neural Network, Induction Rules, and IRANN to
Forecast Purchasers for a Specific Product*

Su-Mi Jung** · Gun-Ho Lee**

■ Abstract ■

It is effective and desirable for a proper customer relationship management or marketing to focus on the specific customers rather than a number of non specific customers. This study forecasts the prospective purchasers with high probability to purchase a specific product. Artificial Neural Network(ANN) can classify the characteristics of the prospective purchasers but ANN has a limitation in comprehending of outputs. ANN is integrated into IRANN with IR of decision tree program C5.0 to comprehend and analyze the outputs of ANN. We compare and analyze the accuracy of ANN, IR, and IRANN each other.

Keyword : Forecasting Purchasers, Customized Marketing, Induction Rule, Neural Network

1. 서 론

최근 기업의 대량 판촉메일 혹은 가격할인 등 푸시 마케팅(Push Marketing) 전략은 법적, 제도적인 한계로 마케팅 관련자와 학자들은 개인화 마케팅에

보다 많은 관심을 가지고 그 중요성을 강조하고 있다. 수익 면에서 가장 큰 고객을 가진 시장은 푸시 마케팅을 선호하지 않는 경향이 있다[19, 21]. 또한 구매량, 구매빈도 및 인구통계학적 변수 중심의 연구를 통한 마케팅은 고객을 단편적으로 이해하거나

논문접수일 : 2005년 7월 5일 논문게재확정일 : 2005년 11월 23일

* 본 연구는 숭실대학교 교내연구비 지원으로 수행되었음.

** 숭실대학교 산업정보시스템공학과

잘못된 정보를 제공할 수 있다. 이에 비해 일대일 (1:1) 마케팅 방식은 구매 의도가 높은 고객을 예측한 후 소수의 고객에게 제품정보를 제공하는 직접 마케팅 전략으로 무작위로 추출된 고객의 반응에 비해 더 좋은 고객 반응을 얻을 수 있다[16].

기업 수익의 80%는 고객 20%에 의해서 창출된다는 법칙에 따라, 모든 고객을 대상으로 일괄적인 마케팅 활동보다는 기업의 진정한 고객을 대상으로 그 범위를 한정함으로써 보다 정교한 고객별 대응이 가능하게 된다[14].

선별된 고객에게 차별화 되고 집중화 된 판매촉진 활동으로 구매 가능성이 높은 우수 고객을 확보할 수 있고, 필요한 고객에게 필요한 마케팅과 제품정보를 제공할 수 있을 것이다. 나아가 구매 가능성이 낮은 고객에게 불필요한 마케팅 비용을 절감하여 전체 수익을 극대화시킬 수 있는 것이다.

마케팅에 관한 연구는 차별화된 마케팅을 위한 많은 연구가 진행되어왔고 고객 중심의 제품 추천, 고객 세분화, 시장 세분화의 연구는 활발히 진행되어 왔다. Zahavi and Levin[20]는 인공 신경망을 직접 마케팅의 목표시장 선정에 적용하였으나, 입력값에 대하여 1개의 출력 값만 존재하는 결정적 문제이므로 고객 특성을 출력할 수 없다. 김시환, 권영식[2]은 데이터마이닝을 이용한 인터넷 쇼핑물 고객 세분화에 관한 연구에서 고객 그룹의 세분화와 고객 구매 경향에 관한 예측을 하였으나 입력변수가 인구 통계학적 항목 변수와 구매거래 관련 변수항목으로 구성되어 고객의 특성을 단편적으로 이해하거나 잘못된 정보를 제공할 수 있다. 이정환, 최문기[5]는 고객 세분화를 통해 인터넷 쇼핑물 구매 경험자의 재 구매 의도에 영향을 미치는 요인을 분석하였다.

개인화 마케팅을 위하여 제품별 예상 구매고객의 특성을 알고 잠재구매 고객들을 분류하여 할 필요가 있으나 예상 구매고객의 특성을 예측하는 방법에 대한 연구는 이루어지지 않았다.

본 연구에서는 제품별 구매고객의 특성 예측방법을 제안하고자 한다. 구매고객의 특성을 예측하기

위하여, 고객의 DB(Data Base)에서 구매경험이 있는 고객에 대한 다양한 데이터를 수집하여 분석하는 것이 필요할 것이다.

인공신경망(Artificial Neural Network : ANN)을 이용할 경우 입력과 출력 변수가 연속형 뿐만 아니라 범주형의 경우 그리고 입력과 출력 변수의 관계를 나타내기가 어려운 복잡한 데이터에 대해서도 좋은 결과를 도출할 수 있다. 하지만, 예측결과에 대한 해석의 어려운 점이 있다. 본 연구에서는 이러한 ANN의 단점을 보완하기 위하여 출력변수 설명과 해석이 용이한 귀납규칙(Induction Rules : IR)을 통합한 예측모형(IRANN)을 제시하고 또한 ANN모형과 IR모형을 구축하여 각각 예측정확도를 비교 평가하고자 한다.

2. 구매 데이터와 변수별 유의성 검증

특정 제품에 대한 고객들의 반응은 기호나 특성에 따라 차이가 있고, 제품을 적절한 고객에게 추천을 할 경우 구매 가능성이 높다는 것을 가정 하여, 봄제품(SPA) 2개(공기청정기, 가습기)와 여름제품(SUA) 2개(에어컨, 선풍기)로 총 4개의 제품을 선택하였다. SPA 및 SUA의 각 제품은 동일한 계절 상품으로 가격차이가 있는 제품을 선택했고, SPA와 SUA를 동시에 선택한 것은 고객의 다양한 특성을 예측하고 IRANN모형이 서로 다른 계절제품에 대해서도 높은 예측 결과를 나타내는지를 보이고자 하였다.

2.1 데이터 수집

인구통계학적항목과 고객의 여러 특성을 예측하기 위한 8개의 고객의식항목을 설문조사를 통해 데이터를 수집했다. 설문 대상은 20세 이상의 남녀이며 공기청정기, 가습기, 에어컨, 선풍기를 구매한 경험이 있는 약 1,092명의 고객을 대상으로 온, 오프라인으로 조사하였다. 오프라인 조사는 12명의 교육된 조사자를 통해 개인면접을 실시하였고, 조사

의 시간적, 경제적 제한으로 대상지역을 수도권, 강원도 일부, 대구로 한정하여 진행하였다. 온라인 조사는 메시지를 통한 극히 일부분으로 연구결과에 큰 차이가 없다고 본다.

고객의식항목은 리커트 6점 척도를 사용하였고, 성의 없는 답변으로 인한 데이터를 제외한 총 1,000개의 설문 데이터를 표본으로 각 사례에 포함되는 카테고리는 입력력 요소로서 <표 1>과 같다.

<표 1> 입력력 요소

변수	인구통계학적항목	
x _g	gender	성별을 이분형으로 표현
x _a	age	나이를 정수형으로 표현
x _m	marital	결혼여부를 이분형으로 표현
x _c	child	미취학 자녀여부를 이분형으로 표현
x _e	education	교육정도를 4개 범주형으로 표현
x _i	income	연간소득을 정수형으로 표현
x _j	job	직업유형을 9개 범주형으로 표현
x _h	house type	거주형태를 3개 범주형으로 표현
변수	고객의식항목	
x1	에너지 소비량에 민감	
x2	저가 제품 선호	
x3	마음에 들면 고가라도 구입함	
x4	고가의 제품 구매 시 기분 좋음	
x5	브랜드나 상표를 먼저 선택함	
x6	유행에 민감함	
x7	건강에 많은 신경을 기울임	
x8	미취학 자녀가 있어서 안전한 제품 선호	
변수	출력변수	
X _{SPA}	구매한 봄 가전제품 (공기청정기, 가습기)	
X _{SUA}	구매한 여름가전제품(에어컨, 선풍기)	

설문지의 타당성과 신뢰성 검증을 위해 요인분석과 신뢰도 검사를 하였고, <표 1>의 입력 요소들이 출력 요소에 대하여 유의한지 검토하기 위해서 Chi-square검정을 실시하였다. 검정을 위해 SPSS 10.0을 사용했고, 검정결과는 <표 3>과 같다.

본 연구의 모형은 많은 변수를 가지고 있는 대규모의 데이터를 대상으로 하며 다양한 방법론에 의

한 분석을 포함하고 있기 때문에 모형의 타당성을 평가하고 여러 모형을 비교 검토하는 작업이 필요하다. 이를 위해 전체 표본 1,000개 데이터를 훈련용, 테스트용, 검증용 3개의 데이터 Set을 학습(6) : 테스트(2) : 검증(2) 비율로 무작위로 구성한다. 훈련Set을 이용하여 모형을 구축하고 테스트Set과 검증Set으로 모형의 비교와 최종적인 평가를 수행하는 것이다.

테스트Set은 ANN이 훈련Set으로부터 모형을 일반화하는 것이 아니라 암기할 경우 추적으로 사용한다. 여러 번 반복과정을 통한 훈련Set 상에서 높은 결과를 나타내지만 테스트Set에서 아주 낮은 정확도를 보인다면 암기한 것으로 본다.

ANN모형에서는 네 가지 제품의 구매 수를 동일하게 하여 학습률과 모멘텀은 0.1로 하였다.

한편 C5.0에서는 테스트Set을 이용할 필요가 없는 관계로 훈련Set에 포함시켜 분석한다.

2.2 변수의 특성 및 선정

입력변수는 인구통계학적항목 변수와 고객 의식항목 변수를 사용하고, 출력 변수는 선정된 봄 제품인 공기청정기 구매, 가습기 구매와 여름제품인 에어컨 구매, 선풍기 구매를 각각 이분형으로 출력하였다. 인구통계학적항목의 각 특성분포는 <표 2>와 같고, 고객의식항목은 각 문항의 타당성을 검증하기 위해 요인분석을 하였다. 요인추출 방법으로 주성분분석을 이용하고 고유치가 1이상인 요인들을 추출하여 varimax 직각회전을 실시하였다. 0.5*0.6 이상의 적재치를 기준으로 추출하는 경우가 일반적이므로 본 연구에서는 각 요인별 적재치가 0.6이상인 변수들만 추출하였다.

요인 분석한 결과 SPA의 요인 1은 X3(0.74), X4(0.782), X5(0.722)로 요인 2는 X1(0.839), X2(0.853)로 요인 3은 X6(0.63), X7(0.807)로 묶인다.

SUA의 요인 1은 X1(0.821), X2(0.815)로 요인 2는 X3(0.816), X4(0.74), X5(0.725)로 요인 3은 X6(0.813), X7(0.812)로 묶인다.

〈표 2〉인구통계학적항목 응답의 특성분포

입력변수	제품	SPA 빈도(%)	SUA 빈도(%)
성 별	남성	142(28.4)	217(43.4)
	여성	358(71.6)	283(56.6)
연 령	20대	150(30)	99(19.8)
	30대	139(27.8)	122(24.4)
	40대	122(24.4)	164(32.8)
	50대	67(13.4)	97(19.4)
	60대	22(4.4)	18(3.6)
결혼여부	미혼	94(18.8)	114(22.8)
	기혼	406(81.2)	386(77.2)
미취학 자녀	있음	238(47.6)	177(35.4)
	없음	262(52.4)	323(64.6)
교 육	고졸이하	153(30.6)	23(4.6)
	전문대졸/재	104(20.8)	68(13.6)
	대졸/재	186(37.2)	162(32.4)
	대학원졸/재	57(11.4)	40(8)
연간소득	1000만원미만	28(5.6)	84(16.8)
	1000만원이상 ~2000만원미만	109(21.8)	129(25.8)
	2000만원이상 ~3000만원미만	179(35.8)	142(28.4)
	3000만원이상 ~4000만원미만	102(20.4)	72(14.4)
	4000만원이상 ~5000만원미만	52(10.4)	36(7.2)
	5000만원이상	30(6)	37(7.4)
직 업	학생	16(3.2)	52(10.4)
	서비스/판매	98(19.6)	155(31)
	제조/유통	76(15.2)	67(13.4)
	법률/사무	45(9)	20(4)
	가사/무직	100(20)	98(19.6)
	의료	56(11.2)	16(3.2)
	교수/교사	42(8.4)	23(4.6)
	공무원/군인	20(4)	29(5.8)
	보험/금융	47(9.4)	40(8)
주거형태	자가	254(50.8)	326(65.2)
	전세	218(43.6)	132(26.4)
	월세	28(5.6)	42(8.4)

요인분석을 통해 선택된 문항은 Cronbach's α 를 이용한 신뢰성 검증을 실시했다. 검증결과 SPA의 요인 1은 $\alpha=0.672$, 요인 2는 $\alpha=0.642$, 요인 3은 $\alpha=0.230$ 으로 나타나고, SUA의 요인 1은 $\alpha=0.568$, 요인 2는 $\alpha=0.650$, 요인 3은 $\alpha=0.540$ 으로 나타났다. α 계수가 0.6이상으로 신뢰성에 큰문제가 없는 문항이 선택되었고, 〈표 3〉과 같이 타당성과 신뢰성에 문제가 있는 SPA의 X6, X7, X8문항과 SUA의 X1, X2, X6, X7, X8문항이 제거되었다.

〈표 3〉 Chi-square검정 결과

변수	SPA			SUA		
	X ²	df	p-value	X ²	df	p-value
x _g	3.187	1	0.074	2.939	1	0.086
x _a	40.345	44	0.646	92.080	46	0.000*
x _m	0.838	1	0.360	24.043	1	0.000*
x _c	0.289	1	0.591	3.157	1	0.076
x _e	29.336	3	0.000*	11.244	3	0.010
x _i	94.251	75	0.010	175.051	95	0.000*
x _j	20.559	8	0.008*	76.971	8	0.000*
x _h	7.011	2	0.030	42.206	2	0.000*
x ₁	15.825	5	0.007*	-	-	-
x ₂	8.236	5	0.144	-	-	-
x ₃	33.689	5	0.000*	21.696	5	0.001*
x ₄	11.373	5	0.044	2.479	5	0.780
x ₅	15.627	5	0.008*	8.643	5	0.124
x ₆	-	-	-	-	-	-
x ₇	-	-	-	-	-	-
x ₈	-	-	-	-	-	-

주) *: p-value < 0.01 수준에서 유의함.

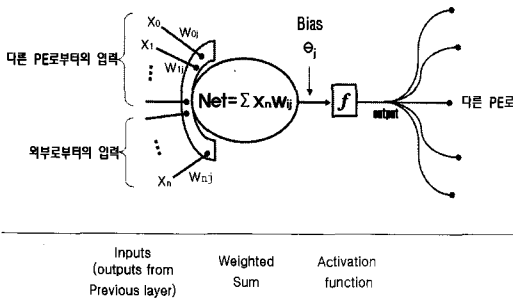
요인분석과 신뢰성 검증에서 제거된 변수 외에 남은 변수는 Monte Carlo 추정 값을 이용한 Chi-square검정을 실시했다. Chi-square검정에서 기대 빈도가 5이하가 되는 Cell들이 있을 경우에 출력변수와 입력변수간의 관련 있는 것으로 사실 왜곡 가능성이 있다. 즉, 데이터 변수군이 너무 커서 정확한 유의수준을 계산할 수 없을 경우 Monte Carlo 추정 값을 유용하게 사용된다.

검정 결과를 통해 출력 변수와 유의한 최종 입력 변수를 선정하여 학습에 적용하였다. 검정은 p-value가 0.01보다 큰 변수는 통계적 유의성이 없는 것이므로 입력 변수에서 제외되고, p-value가 0.01보다 작은 변수만 최종 입력 변수로 사용한다. <표 3>의 SPA 검정결과에서 인구통계학적항목의 X_e , X_j 와 고객의식항목의 X_1 , X_3 , X_5 가 p-value<0.01이므로 최종 입력 변수로 선택한다. SUA 검정결과에서 인구통계학적항목의 X_a , X_m , X_i , X_j , X_h 와 고객의식항목의 X_3 이 p-value<0.01이므로 최종 입력 변수로 선택되었다. 선정된 최종 입력 변수는 예측기법에 적용하게 된다.

3. 인공신경망(ANN)모형 구축

ANN은 많은 양의 간단한 Processing Element를 사용하여 인간의 신경망과 같이 사용하도록 하는 준 수학적 모델이다.

다양한 ANN알고리즘 중 역전파(Back propagation) 방식은 퍼셉트론의 한계점을 극복하기 위하여 입력층과 출력층에 한 개 이상의 은닉층(hidden layer)을 사용하여 다층퍼셉트론(Multi Layer Perceptron)으로 구성하고, 모든 입력패턴에 대하여 발생하는 오차함수를 최소화하는 방향으로 연결강도를 조정하는 학습 알고리즘이다. 다층퍼셉트론은 퍼셉트론과 유사한 구조를 가지고 있지만 은닉층과 비선형 입출력 특성을 이용함으로써 학습능력을 향상시키고 퍼셉트론의 단점을 극복할 수 있다[1].



[그림 1] Processing Element(PE)의 기본구조

ANN모형은 다양한 분야의 문제에 적용될 수 있고, 입력 변수와 출력 변수의 관계를 그리기가 어려운 복잡한 데이터에 대해서도 좋은 결과를 주는 것으로 알려져 있다. 또한 입력 변수와 출력 변수의 유형이 연속형이거나 이분형인 경우를 모두 다룰 수가 있는 반면 분석 과정이 불투명하며, 해석이 부족하다는 단점을 가지고 있다. 따라서 설명력(Comprehensibility)보다는 더욱 정확한 예측을 고려하는 문제 분석에 유용하게 이용될 수 있다.

ANN모형 구축 시 은닉층 수와 은닉층 노드 수의 결정은 경험적 방법에 의해 결정할 수 있다. 일반적으로 은닉층의 PE의 개수는 결과에 큰 영향을 주지 않는다고 한다[6, 18]. 즉, 은닉층 수와 PE의 개수 결정은 어떻게 응용하는가에 따라 달라지고, 주관적일 수 있으므로 실험으로 확인하는 것이 바람직 할 것이다[4]. 만약 입력 자료가 특성 추출이 용이하지 않은 자료로 구성되어 있으면 고수준의 특성을 추출하기 위해서는 여러 개의 은닉층이 요구된다. 반면 어느 정도 고수준의 특성치를 나타내고 있으면 하나 또는 두개의 은닉층만 있어도 거의 모든 형태의 문제 해결 공간을 구성할 수 있다[4, 6, 18]. 은닉층 노드 수는 입력층 노드 수보다 두 배 이상이 되어서는 안 된다는 것이 일반적이고, 시작하기 적합한 수는 은닉층과 입력층의 노드 수를 같은 크기로 만드는 것이다.

본 연구에서는 은닉층 수를 하나만 사용하고, 은닉층 노드 수는 <표 4>와 같이 총 6개 set으로 구성된다. SPA의 입력층 노드는 범주형 변수 $x_e(4)$, $x_j(9)$ 와 정수형 변수 x_1 , x_3 , x_5 로 총 16개이고, SUA는 이분형 변수 x_m 과 범주형 변수 $x_j(9)$, $x_h(3)$ 그리고 정수형 변수 x_a , x_i , x_3 로 총 16개의 입력층 노드를 사용한다.

경험적 방법에 의하여, SPA와 SUA에 적합한 은닉층 노드 수를 나타내는 식 (1), 식 (2)를 각각 얻었다. 여기서 n 은 입력층 노드 수를 나타낸다.

$$h = 2(n - 5) \tag{1}$$

$$h = 2(n - 1) \tag{2}$$

경험률에 의한 방법은 S_n 은 훈련 데이터 수, m 이 출력층 노드 수 일 때 은닉층 노드수를 학습 자료의 수에 비례하여 다음과 같이 사용하고 있다 [22].

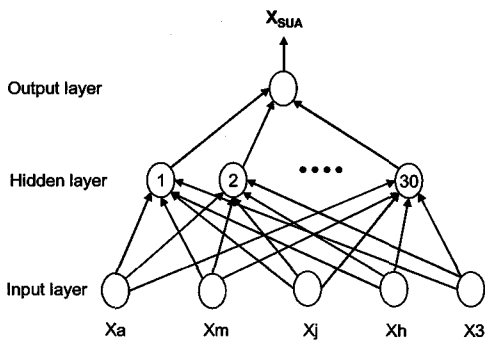
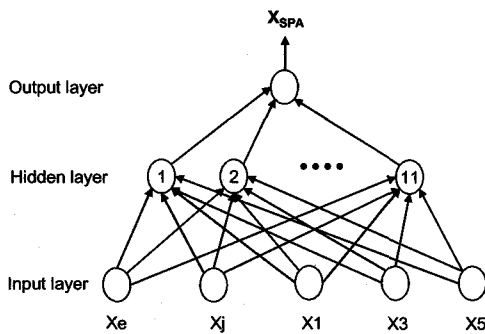
$$h = S_n / (5(m + n)) \quad (3)$$

$$h = \{2(m + n)\} / 3 \quad (4)$$

Set 1은 입력층 노드 수와 같다는 제안에 따라 n 개로 구성하고, Set 2는 $(n \times 1/2)$ 개, Set 3은 $(n \times 2)$ 개, Set 4는 식 (1)과 식 (2)에 의한 구성, Set 5는 식 (3)에 의한 구성, Set 6은 식 (4)에 의해 구성된다.

〈표 4〉 구성된 Set의 은닉층 노드 수

구분	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6
SPA	16	8	32	22	4	11
SUA	16	8	32	30	4	11



〈그림 2〉 인공신경망 구조

4. 귀납규칙(IR) 구축

귀납적 학습방법인 C5.0은 생성된 의사결정트리 가 지나치게 많은 단계와 잎 노드를 가질 경우에 학습된 의사결정트리의 일반화 능력을 제고하기 위하여 잎 노드를 제거하는 가지치기(Pruning)를 하는데 유용하다. 이 방법은 예측 정확도를 향상시킬 수 있으나 오류율도 함께 증가시키므로 오류기반 가지치기로 오류율의 증가를 통제한다. 또한 자동적으로 If-Then Rule을 생성해주는 규칙들의 집합 형태인 명제형 지식(Propositional Knowledge)으로 해석 및 분석이 용이하기 때문에 일반 사용자들도 쉽게 이해할 수 있다[3, 17].

ANN모형과 같은 표본 데이터를 사용하므로 동일한 변수를 사용하고, 테스트Set을 훈련Set에 포함시켜 분석하였다. 가지치기는 지나치게 많은 마디를 가지는 의사결정나무는 새로운 자료에 적용할 때 예측 오차가 매우 클 가능성이 있으므로 SPSS사의 Clementine 6.5에서 제안한 방법으로 75%를 사용하였다.

Boosting은 잘못 분류된 개체들에 집중하여 분류 규칙을 새로 만드는 단계를 반복하고 각 단계의 분류 규칙을 결합하여 최종 모형을 산출한다.

다음은 Boosting 알고리즘을 나타낸 것이다[8].

[Step 1] 개체 i 에 초기 가중치 $w_i = 1/n (i = 1, \dots, n)$ 을 부여한다.

[Step 2] 단계 $m (= 1, \dots, m)$ 에서의 반복과정은 다음과 같다.

- 분류규칙 $G_m(x)$ 를 만들어 개체 x 를 분류한다.
- 오분류율 err_m 을 산출한다.
- $a_m = \log\{(1 - err_m)/err_m\}$
- 오분류 개체 i 에 대하여 가중치 w_i 를 다음과 같이 바꾼다.
 $w_i \leftarrow w_i \cdot \exp\{a_m\}$
- 다음 단계에서 반복한다.

[Step 3] 최종 분류규칙을 산출한다.

$$G(x) \leftarrow a_1 \cdot G_1(x) + \dots + a_m \cdot G_m(x)$$

Boosting 옵션의 사용은 시도 수만큼의 선택할 수 있는 규칙모형을 생성하고 C5.0의 예측정확도를 보다 월등하게 제시할 수 있다. <표 5>와 <표 6>은 Boosting 옵션을 이용하여 시도 수를 1부터 10까지 증가시키는 과정을 보여준다. 결과를 보면 시도 수를 1로 하고 실행할 경우 분류자가 너무 적어서 향상이 취소되고, 마지막 분류자가 부정확할 경우 시도 수만큼 증가하지 않는다. 본 연구에서는 시도 수를 10까지 증가시켜도 시도 수 2이상은 예측정확도가 나아지지 않았기 때문에 Boosting의 시도 수를 3으로 설정하였다.

<표 5> SPA의 Boosting 시도에 따른 정확도

시도	예측정확도(%)									
1	분류자가 너무 적어서 향상이 취소									
2	73.7	-	-	-	-	-	-	-	-	-
3	73.7	66.7	59.0	-	-	-	-	-	-	-
4	73.7	66.7	59.0	59.5	-	-	-	-	-	-
5	73.7	66.7	59.0	59.5	60.2	-	-	-	-	-
6	73.7	66.7	59.0	59.5	60.2	66.2	-	-	-	-
7	73.7	66.7	59.0	59.5	60.2	66.2	62.0	-	-	-
8	73.7	66.7	59.0	59.5	60.2	66.2	62.0	62.0	-	-
9	마지막 분류자가 부정확하여 증가 못함									
10										

<표 6> SUA의 Boosting 시도에 예측정확도

시도	예측정확도(%)									
1	분류자가 너무 적어서 향상이 취소									
2	71.5	-	-	-	-	-	-	-	-	-
3	71.5	67.0	67.7	-	-	-	-	-	-	-
4	71.5	67.0	67.7	64.7	-	-	-	-	-	-
5	71.5	67.0	67.7	64.7	65.5	-	-	-	-	-
6	71.5	67.0	67.7	64.7	65.5	67.0	-	-	-	-
7	71.5	67.0	67.7	64.7	65.5	67.0	49.5	-	-	-
8	71.5	67.0	67.7	64.7	65.5	67.0	49.5	66.0	-	-
9	71.5	67.0	67.7	64.7	65.5	67.0	49.5	66.0	49.5	-
10	71.5	67.0	67.7	64.7	65.5	67.0	49.5	66.0	49.5	67.0

5. IRANN모형 구축

5.1 모형의 절차

기업의 고객 DB에서 과거에 특정 제품에 대해 구매 경험이 있는 고객관련 데이터를 수집한 고객 리스트는 IRANN모형의 제품에 대한 고객특성 예측을 위한 기본 자료로 사용된다.

IRANN모형 도출절차는 [그림 3]과 같다. 사용된 데이터는 두 가지의 제품군을 구매한 고객 리스트로 구분하고, 예측기법에 따라서는 IR기법 C5.0모형은 테스트Set을 적용할 필요가 없으므로 훈련Set과 검증Set으로 분류한다.

다음은 입력 변수와 출력 변수를 선정한다. 인구통계학적항목과 고객의식항목으로 구성된 입력 변수는 출력 변수에 대하여 유의한지를 검토하기 위해서 Chi-square검정을 실시한다. 이를 통해 최종 입력 변수를 선정하고 적용하여 ANN의 역전파 알고리즘으로 예측정확도를 측정하고, IR기법은 적절한 가지치기(Pruning)와 Boosting을 사용하여 측정한다.

ANN모형의 측정결과가 우수한 모형을 선택한 후 이 모형을 검증Set에 적용하여 학습시키고 IR기법을 이용하여 예측결과의 원인을 분석한다. IRANN모형과 ANN모형의 우수성을 <표 7>의 예측정확도로 비교 평가한다. 예측된 특성에 따라 선별된 고객이 광고 등의 마케팅 대상이 되는 것이다.

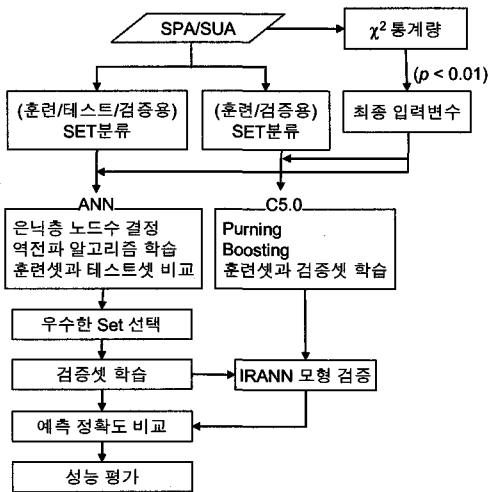
<표 7> 제품별 예측정확도(%)

	XSPA	XSUA
ANN	69.56	69.77
IRANN	96.00	90.00

ANN모형은 내부의 노드들의 가중치가 결정되어지는 내부 학습 과정의 특성상 신경망 모형 결과의 타당성에 대한 추론을 하기는 어려우므로 예측결과의 과정을 구체적으로 알기는 불가능 하다.

쉽게 설명되지 않는 내부적인 작업을 수행하고,

이를 통해 얻어진 결과물을 제공할 뿐 어떤 변수가 얼마나 중요한지, 어떻게 상호작용이 이뤄져서 그러한 결과물을 주게 되는지에 대한 해석이 불가능하다[7]. 따라서 IRANN모형에서는 이와 같은 ANN의 문제점을 보완하기 위해 IR 추론에 근거하여 If-Than 규칙을 생성함으로써 명쾌하고 이해하기 쉬운 결과물을 나타내고 원인 해석이 용이하도록 하였다.



[그림 3] IRANN모형의 절차

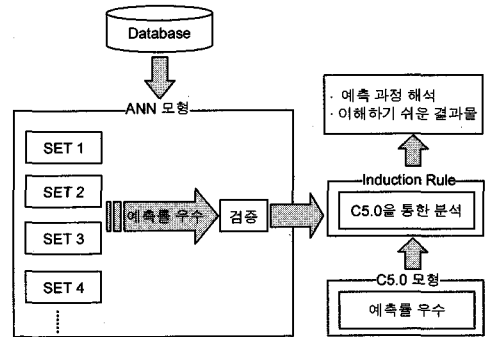
5.2 IRANN모형 구조

예측에 대한 원인을 좀 더 구체적으로 이해하기 위하여 해석이 용이하도록 한 IRANN모형 처리과정은 다음과 같다.

- [Step 1] 고객의 구매정보 DB로부터 set들을 구축한다.
- [Step 2] 예측정확도가 우수한 set을 선택한다.
- [Step 3] ANN모형을 구축하여 Step 2에서 선택한 set으로 ANN모형을 검증한다.
- [Step 4] Induction Rule을 이용하여 해석한다.

따라서 Step 3과 Step 4의 ANN 및 IR의 구축은 앞에서 구축한 모형과 크게 다르지 않아 그대로 사용한다. 고객 DB에서 ANN모형을 적용시켜 훈련을

시키고 그 결과 예측정확도가 우수한 Set을 선택한다. 그리고 이 우수한 Set의 모형을 그대로 적용하여 검증Set에서 검증한다. 검증된 Set을 해석하기 위하여 IR기법을 통합하여 분석한다. 그럼으로써 우수한 예측정확도뿐만 아니라 ANN모형에서 설명할 수 없었던 예측과정을 Decision Tree로 해석할 수 있으며, 결과물 또한 이해하기 쉽다.



[그림 4] IRANN모형 구조

6. 결과 및 분석

본 장에서는 모형별 분석과 분석 결과의 원인을 제시하고, 각 모형들을 비교 분석하여 최적의 모형을 통해 제품별 고객의 특성을 추출하는 것이다. 각 모형 테스트는 제품에 따라 크게 두 부분으로 나눠 실행하였다. 봄제품(SPA)에 대한 고객 특성 예측과 여름제품(SUA)에 대한 고객 특성 예측이 그것이며, 선정된 제품은 다른 제품과의 결과 차이를 보이는지 알아보기 위한 것이다.

6.1 ANN모형 분석 결과

훈련Set과 테스트Set의 예측정확도 결과는 <표 8>과 같고, ANN모형실험에서 출력 변수의 예측정확도 차이는 표본 데이터 크기와 내용에 의한 것이라고 할 수 있다. 두 개의 출력 변수의 예측정확도 차이보다는 출력 변수의 각 Set의 비교 분석을 위한 것이다.

예측정확도는 10번의 실험의 평균값이다. 각 Set

은 비교적 낮은 예측정확도를 보이며 아주 큰 차이가 없다. 변수 X_{SPA}에서 11개의 은닉층 노드수를 가지는 Set 6(64.06%)이 가장 우수한 예측정확도를 나타냈고, 변수 X_{SUA}에서는 30개의 은닉층 노드수를 가지는 Set 4(67.87%)이 가장 우수한 예측정확도를 나타냈다.

훈련Set에서 가장 우수한 예측정확도를 나타냈더라도 테스트Set을 통해 나타난 결과가 반드시 높다고 할 수 없으므로 훈련Set을 통해 산출된 모형을 바탕으로 테스트Set에 적용하여 가장 예측정확도의 변화가 없는 모형을 선택하는 것이 올바른 것이다. 본 연구에서는 테스트Set과의 결과 차이가 크지 않으므로 훈련Set에서 우수한 결과를 적용한다.

<표 8> ANN모형 간의 예측정확도(%)

Set	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	
X _{SPA}	HLN	16	8	32	22	4	11
	훈련	61.85	62.36	61.80	62.47	60.29	64.06
	테스트	56.28	58.91	60.92	61.86	60.40	63.73
X _{SUA}	HLN	16	8	32	30	4	11
	훈련	65.29	67.59	66.50	67.87	65.30	66.61
	테스트	63.71	59.13	63.80	63.92	63.64	61.92

6.2 C5.0모형 분석결과

C5.0모형 테스트는 보다 우수한 예측 결과를 위해서 Boosting 옵션을 사용하였고, Boosting 옵션에서 시도 수만큼 증가시킬 수 없는 경우는 마지막 분류자가 부정확할 때 발생한다.

시도 수는 <표 5>와 <표 6>의 결과와 같은 이유로 3으로 했으며, 검증Set의 실험에서도 동일하게 사용하였다.

<표 9> C5.0의 예측정확도(%)

Boosting	훈련Set		검증Set	
	SPA	SUA	SPA	SUA
rule 1	73.70	71.50	78.00	69.00
rule 2	66.70	67.00	73.00	65.00
rule 3	59.00	67.70	74.00	52.00

6.3 IRANN모형 결과 분석

IRANN모형은 ANN모형을 학습 시킨 결과가 가장 우수한 훈련Set 모형을 그대로 검증Set에 적용하여 검증한 모형에 C5.0모형을 통합한 모형이므로 검증Set을 적용한 예측정확도를 나타내고 C5.0모형의 형태로 해석된다. C5.0모형실험 결과 Boosting 옵션에 의해 다소간 정확도의 차이가 있었고, IRANN 모형에도 동일한 실험을 위해서 Boosting 옵션을 사용하였다.

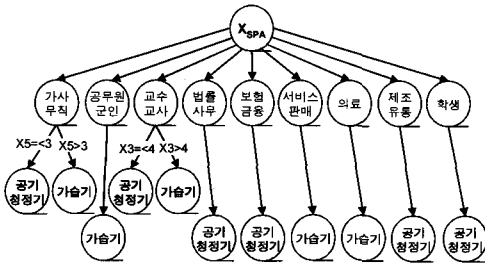
<표 10> IRANN모형의 예측정확도(%)

Boosting	SPA	SUA
rule 1	96.00	95.00
rule 2	93.00	94.00
rule 3	88.00	82.00

IRANN모형의 예측정확도에서 가장 우수한 결과를 Rule set으로 나타내고, 이 Rule set은 if, then 규칙 형식으로 변수X_{SPA}의 0.0(공기청정기)과 1.0(가습기) 그리고 변수X_{SUA}의 0.0(에어컨)과 1.0(선풍기)에 대한 고객 특성이 각각 어떤 규칙에 의해 분류됐는지 해석된 결과이다.

<표 11> X_{SPA}의 Rule set

<p>rules for 0.0(공기청정기) : rule #1 for 0.0 : if X_j=법률/사무 then → 0.0 rule #2 for 0.0 : if X_j=제조/유통 then → 0.0 rule #3 for 0.0 : if X_j=가사/무직 and X₅ ≤ 3 then → 0.0 rule #4 for 0.0 : if X_j=교수/교사 and X₃ ≤ 4 then → 0.0 rule #5 for 0.0 : if X_j=학생 then → 0.0 rule #6 for 0.0 : if X_j=보험 then → 0.0</p>	<p>rules for 1.0(가습기) : rule #1 for 1.0 : if X_j=의료 then → 1.0 rule #2 for 1.0 : if X_j=서비스/판매 then → 1.0 rule #3 for 1.0 : if X_j=공무원/군인 then → 1.0 rule #4 for 1.0 : if X_j=가사/무직 and X₅ > 3 then → 1.0 rule #5 for 1.0 : if X_j=교수/교사 and X₃ > 4 then → 1.0</p>
--	--



[그림 5] X_{SPA}의 Decision Tree

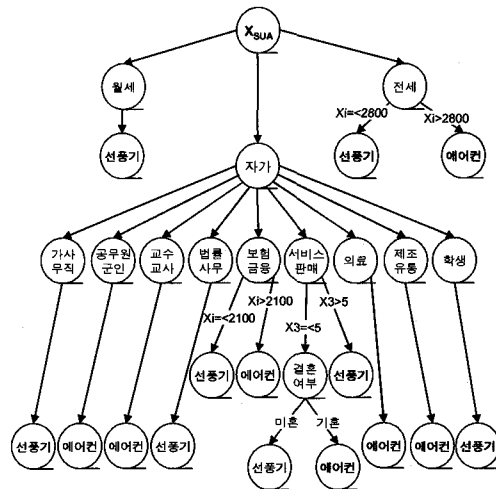
X_{SPA}의 공기청정기(0.0)는 5개의 규칙결과를 갖는다. 즉, rule #1과 rule #2는 법률 및 사무직에서 근무하는 응답자와 제조 및 유통직종을 가진 응답자가 공기청정기를 구매한다는 뜻이다. rule #3은 가사 및 무직이고, X5에서 척도3점 이하로 반응한 응답자로 브랜드나 상표를 중요시하는 특성을 가진다. rule #4는 교직에 근무하고, X3에서 척도점수의 4이하 응답자로 마음에 들면 고가라도 구입한다는 특성을 가진다. 공기청정기의 구매자는 브랜드를 중요시하고 고가의 제품을 구매할 가능성이 크다고 해석된다.

[표 12] X_{SUA}의 Rule set

<p>rules for 0.0(에어컨) : rule #1 for 0.0 : if X_m=기혼 and X_j= 서비스/판매 and X_h= 자택 and X₃ <= 5 then → 0.0 rule #2 for 0.0 : if X_j=제조/유통 and X_h=자택 then → 0.0 rule #3 for 0.0 : if X_j=교수/교사 and X_h=자택 then → 0.0 rule #4 for 0.0 : if X_i > 2800 then → 0.0</p>	<p>rules for 1.0(선풍기) : rule #1 for 1.0 : if X_i <=2800 and X_h=전세 then → 1.0 rule #2 for 1.0 : if X_j=가사/무직 then → 1.0 rule #3 for 1.0 : if X_m=미혼 then → 1.0 rule #4 for 1.0 : if X_h=월세 then → 1.0 rule #5 for 1.0 : if X_i <= 2100 then → 1.0 rule #6 for 1.0 : if X_j=법률/사무 then → 1.0 rule #7 for 1.0 : if X_j=보험/금융 then → 1.0</p>
--	--

가습기(1.0)에서 rule #1과 rule #2는 의료와 서비스, 판매 종사자들이 주로 구매한다는 의미이다. rule #4는 가사 및 무직이고 X5에서 척도점수의 3 초과로 응답한 자이다. 브랜드나 상표를 그다지 중요시 여기지 않는 특성을 가진다. rule #5는 교직에 근무하고 X3에서 척도점수가 4초과로 응답한 자이다. 마음에 들더라도 고가이면 구입하지 않는다는 특성을 가진다. rule #5에서도 볼 수 있듯이 가습기의 구매자는 제품의 질을 선호하는 실용위주의 고객이 많다고 해석할 수 있다.

따라서 각 규칙의 조건에 해당하는 고객들을 대상으로 각 제품에 대한 마케팅 활동이 효과적일 것으로 판단된다.



[그림 6] X_{SUA}의 Decision Tree

6.4 모형별 비교평가

[표 13]은 검증Set의 결과와 IRANN모형 결과 뿐만 아니라 ANN모형의 훈련Set과 테스트Set에서 가장 우수한 결과를 보인 모형의 예측정확도와 C5.0모형의 훈련Set의 예측정확도를 보여준다. 이것은 각 모형의 훈련 중에 어떤 우수한 Set을 적용 시켜서 나타난 검증 결과인가를 보여주기 위한 것이며, 우수성을 입증하기 위하여 IRANN모형의 예측정확도와 ANN모형의 예측정확도를 비교하는 것

이므로 ANN모형과 C5.0모형의 비교는 언급하지 않는다.

<표 13> 모형별 예측정확도(%)

기법 변수	ANN 훈련	ANN 테스트	C5.0 훈련	ANN 검증	C5.0 검증	IRANN
xSPA	64.06	63.73	73.70	70.83	78.00	96.00
xSUA	67.87	63.92	71.50	72.08	69.00	95.00

<표 14> 통계학적기법과의 특징 비교

모형 특징	IRANN	ANN	통계학적 방법론
작업유형	분류 예측	분류 예측 군집화	분류, 예측
독립변수와 종속변수의 관계	비모수 비선형	비모수 비선형	선형, 비선형(특정 함수식을 사용하는 가정 하에 비선형 문제도 추정가능)
변수유형	무관	연속형 이분형	연속형(선형 회귀분석), 범주형(로지스틱회귀분석)
원인 및 결과해석	가능	불가	가능하지만 부족
오차항 분포가정	불필요	불필요	엄격한 가정
가정 위배 시	무관	무관	신뢰성 및 기법의 적정성 문제시 됨 추정치에 대한 추론도 정당성 없음

실험결과를 분석해보면 몇 가지 특징이 있다.

첫 번째 X_{SPA}에서 ANN모형이 C5.0모형보다 나은 결과를 나타내고 있다. ANN기법은 대량 데이터나 많은 종류의 노이즈를 포함하고 있는 고객의 식 항목변수를 늘릴 경우 더 효과적이고, 이 실험에서 두 모형은 큰 차이를 보이지 않으므로 C5.0에 비해 성능이 떨어진다고 볼 수 없다.

두 번째 IRANN모형을 제외한 모형들이 대체로 낮은 예측정확도를 보이고 있다. 그러나 ANN기법을 통한 학습 성과는 사용된 데이터의 양과 질에 의존하므로 개선된 데이터만 확보된다면 보다 높은 예측정확도를 기대할 수 있다.

세 번째 봄제품(SPA)과 여름제품(SUA)의 ANN

모형과 IRANN모형의 예측결과를 <표 13>에서 보면 IRANN모형은 C5.0모형에 의한 해석 능력을 가질 뿐만 아니라 예측정확도에 있어서도 우수한 결과를 나타낸다.

IRANN모형은 ANN모형과 달리 독립변수와 종속변수의 관계에 있어 비모수, 비선형 문제에도 적용이 가능하다. 또한 통계학적 모형에서 요구되는 엄격한 가정에 제한을 받지 않고 가정을 위배했을 경우에도 적용이 가능하지만 기존의 통계학적 방법론들은 엄격한 오차항 분포 가정을 요구하기 때문에 그 사용 범위에 제한을 받거나 가정을 위배한 경우에는 그 결과에 대한 신뢰성이나 사용 기법의 적정성들이 문제시 된다[12, 19].

7. 결 론

기존의 통계학적 방법론들은 엄격한 가정을 요구하기 때문에 그 사용 범위에 제한을 받거나 가정을 위배한 경우에는 그 결과에 대한 신뢰성이나 사용 기법의 적정성이 문제가 되었다. 제품별 구매고객을 예측과 특성을 분류하기 위하여, 정확한 예측 뿐만 아니라 해석까지 고려해야하는 문제를 ANN 혹은 IR기법 단독으로는 해결이 어려웠다.

본 연구는 마케팅 관련 연구에서 입력과 출력변수가 연속형 뿐만 아니라 범주형의 경우 그리고 입력과 출력 변수의 관계를 나타내기가 어려운 복잡한 데이터에 대해서도 좋은 결과를 도출할 수 있는 점에서 ANN을 적용하였다. 하지만 ANN은 중간과정과 출력결과에 대한 설명과 해석이 어려워 IR기법 C5.0으로 보완한 IRANN모형을 통해 특정 제품의 구매 의도를 가진 고객특성을 보다 정확히 예측할 수 있었다.

고객예측을 위한 문제에서 ANN와 IR의 두 기법의 이점을 취한 모형을 이용함으로써 ANN 혹은 IR을 단독으로 사용했을 때 보다 우수한 예측정확도를 나타내는 결과를 얻을 수 있었다. 예측정확도 높이기 위해서는 제품을 필요로 하는 구매경험이 있는 고객의 리스트를 충분히 확보하여 고객의 정

보를 정확히 수집하여야 할 것이다.

참고 문헌

- [1] 김대수, 「신경망 이론과 응용(1)」, 하이테크 정보사, (1992), pp.91-142.
- [2] 김시환, 권영식, “데이터마이닝을 이용한 인터넷 쇼핑물 고객 세분화”, 「2000 SPSS 사용자 사례 논문」, (2000), pp.41-69.
- [3] 이진창, 이철원, “사례기반추론과 인공지능망을 결합한 지능적인 경영전략수립 모형에 관한 연구”, 「경영정확연구」, 제8권, 제1호(1998).
- [4] 이진창, 정남호, 신경식, “신용카드 시장에서 데이터마이닝을 이용한 이탈고객 분석”, 「한국지능정보시스템학회논문지」, 제8권, 제2호(2002), pp.15-35.
- [5] 이정환, 최문기, “고객세분화를 통한 인터넷 쇼핑물 구매 경험자 재구매 의도 영향 요인”, 「*Journal of Information technology applications & management*」, 제10권, 제1호(2003).
- [6] 전웅섭, “인공 신경망을 이용한 소프트웨어 개발 공수 예측모델에 관한 연구”, 「한국정보처리학회 논문지」, 제3권, 제1호(1996).
- [7] 조용준, 이용규, “의사결정 나무분석에서 C5.0과 분류알고리즘의 비교연구”, 「수학 통계 논문집」, 제6권(1999), pp.21-47.
- [8] 허명희, “잘 알려지지 않은 Clementine 활용기법-복합적 모델링”, 「9th SPSS user conference」, 2004.
- [9] 백길호, “매스마케팅 한계 극복할 DB 마케팅”, 「제일기획마케팅연구소」, 1995.
- [10] Herz, J., A. Krogh, and R.G. Palmer, *Introduction the Theory of Neural Computation*, Addison Wesley, 1991.
- [11] Jiawei Han, Micheline Kamber, *Data Mining : Concepts and Techniques*, Morgan Kaufmann Publishers, 2001.
- [12] Michael J.A., Berry and Gordon Linoff, “Data Mining Techniques-For Sales and Customer Support,” *Wiley*, 1997.
- [13] Quinlan, J.R. and J. Quinlan, *C4.5 : Programs for Machine Learning*, Morgan Kaufman Publishers, 1997.
- [14] Richard Koch, *The 80/20 Principle : The Secret to Success by Achieving more with Less*, Bantam Doubleday Dell Pub, 1998.
- [15] Blattberg, R.C., “Research Opportunities in Direct Marketing,” *Journal of Direct marketing*, Vol.1, No.1(1987), pp.7-14.
- [16] Jain, B.A. and N.B. Nag, “Performance Evaluation of Neural Network Decision Models,” *Journal of Consumer Research*, Vol.9(1982), pp.179-180.
- [17] Quinlan, R., “Induction of Decision Tree,” *Machine Learning*, Vol.1(1996), pp.81-98.
- [18] Lipmann, R.P., “An Introduction to Computing With Neural Nets,” *IEEE ASSP Magazine*, Vol.3, No.4(1988), pp.4-22.
- [19] West, P.M., P.L. Brockett, and L.L. Golden, “A Comparative Analysis of Neural Networks and Statistical Methods for Predicting Consumer Choice,” *Marketing Science*, Vol.16, No.4(1997), pp.370-391.
- [20] Zahavi, J. and N. Levin, “Applying Neural Computing to Target Marketing,” *Journal of Direct marketing*, Vol.11, No.1(1997), pp.5-22.
- [21] Mckinsey and Company, “Internet Pricing-A Creator of Value Not a Destroyer,” *McKinsey Marketing Solution*, 2000.
- [22] Neural ware 社의 NeuralWorks S/W.

별첨 1 <표 1> SPA의 은닉층 노드수에 따른 훈련결과(%)

HLN \ 실험	1	2	3	4	5	6	7	8	9	10	평균
3	59.39	56.03	50.31	61.63	61.03	59.85	58.70	59.12	61.97	57.85	58.59
4	62.07	58.12	60.83	59.45	61.36	60.68	58.55	62.09	58.49	61.22	60.29
5	59.44	12.00	60.99	59.02	59.00	61.94	52.77	60.41	61.11	64.82	55.15
6	57.72	60.45	60.12	57.35	62.82	66.66	57.14	61.74	61.78	61.58	60.74
7	52.50	59.73	60.00	59.85	58.82	61.93	61.18	59.42	61.53	64.58	59.95
8	63.41	60.14	61.90	55.55	63.63	66.07	61.74	62.26	61.70	67.24	62.36
9	59.18	58.82	58.70	62.06	62.50	60.13	60.14	59.35	59.72	57.69	59.83
10	61.25	59.42	60.15	55.42	58.86	58.44	61.94	57.85	65.83	57.24	59.64
11	57.04	60.69	71.18	71.15	58.22	66.66	61.15	67.44	65.38	61.72	64.06
12	60.13	61.07	62.89	62.00	65.80	59.21	57.53	60.64	60.71	65.41	61.54
13	59.35	59.29	63.39	58.15	64.47	64.13	64.02	62.50	54.34	64.58	61.42
14	62.26	60.90	59.39	56.11	55.92	60.58	64.41	62.74	60.83	58.52	60.17
15	58.49	59.72	59.37	60.92	61.97	62.16	58.55	59.11	52.22	61.43	59.39
16	66.00	60.00	61.18	57.14	63.26	60.43	61.07	60.26	63.01	66.18	61.85
17	61.59	63.88	58.86	57.63	57.14	63.91	61.76	62.26	63.26	59.86	61.02
18	59.44	58.44	59.28	59.74	57.53	52.22	59.23	67.11	64.33	61.07	59.84
19	58.44	62.17	58.62	58.33	58.89	57.03	60.13	61.15	63.41	57.14	59.53
20	57.94	60.62	60.56	64.66	58.66	63.01	61.37	57.48	57.69	58.33	60.03
21	60.00	56.79	55.47	59.21	60.13	59.71	62.96	58.78	59.86	60.99	59.39
22	59.51	64.53	61.03	65.27	57.32	65.10	63.12	65.33	63.51	60.00	62.47
23	58.78	65.10	62.89	61.43	61.68	61.70	58.33	63.70	56.09	61.29	61.10
24	61.11	60.00	60.93	61.53	57.71	60.25	63.26	56.93	61.11	59.87	60.27
25	58.22	58.97	58.59	51.90	65.78	60.00	61.07	59.73	61.87	55.63	59.18
26	58.27	53.94	62.06	51.97	59.11	61.72	59.87	53.95	60.00	58.00	57.89
27	55.80	61.07	59.73	57.14	58.33	58.44	60.68	57.63	64.10	64.23	59.72
28	60.26	61.87	66.90	55.03	62.74	58.82	60.52	57.82	61.11	56.11	60.12
29	60.53	58.53	59.73	58.15	59.85	56.71	63.03	60.28	55.17	62.41	59.44
30	64.20	51.72	55.23	58.22	65.51	63.57	62.14	59.42	53.42	61.53	59.50
31	61.88	66.43	58.08	58.21	62.00	62.67	58.33	59.42	64.05	54.60	60.57
32	63.63	62.00	57.53	57.53	62.55	64.70	69.23	62.26	63.82	54.79	61.80

<표 2> SPA의 은닉층 노드수에 따른 테스트결과(%)

HLN \ 실험	1	2	3	4	5	6	7	8	9	10	평균
16	59.26	66.66	42.30	61.53	44.00	71.69	54.71	45.65	65.11	51.85	56.28
8	59.18	53.70	61.90	58.82	51.02	67.92	66.00	52.38	57.69	60.46	58.91
32	60.86	57.79	63.46	63.82	58.62	56.25	63.82	63.04	62.00	59.52	60.92
22	64.44	60.71	63.63	65.95	53.33	42.85	70.21	59.25	68.62	69.64	61.86
4	65.22	58.00	59.57	52.83	62.72	65.38	65.30	55.55	57.39	62.00	60.40
11	72.72	72.22	63.26	60.41	61.90	64.58	59.25	62.74	56.60	63.63	63.73

주) HLN : Hidden Layer Node, n=16, m=1

별첨 2 <표 3> SUA의 은닉층 노드수에 따른 훈련결과(%)

HLN \ 실험	1	2	3	4	5	6	7	8	9	10	평균
3	66.02	61.48	66.44	69.38	64.33	66.66	66.03	64.93	63.01	64.81	65.31
4	63.44	66.47	68.91	66.44	62.58	70.13	66.23	62.83	64.75	61.17	65.30
5	62.17	67.08	70.19	68.00	66.04	64.51	71.62	65.69	63.64	66.25	66.52
6	58.99	66.22	62.80	63.33	67.37	63.87	66.44	64.70	66.90	67.55	64.82
7	63.09	65.13	57.71	63.33	69.33	67.16	64.90	70.00	66.21	63.97	65.08
8	71.24	73.14	58.94	70.46	69.53	69.01	67.92	65.75	63.69	66.25	67.59
9	60.00	64.28	66.66	71.05	65.62	63.69	63.69	72.46	66.66	61.25	65.54
10	60.83	62.87	61.53	68.62	65.75	64.86	66.27	67.04	66.43	64.28	64.85
11	64.37	62.83	69.48	64.55	67.13	70.00	67.76	63.41	67.78	68.78	66.61
12	66.21	63.22	63.63	62.65	64.17	64.08	63.29	67.58	60.89	73.50	64.92
13	68.38	61.93	58.55	66.25	68.21	65.35	63.39	59.17	65.83	63.90	64.10
14	66.42	71.62	63.94	65.51	62.42	71.33	64.53	68.75	69.03	65.54	66.91
15	67.36	60.92	60.78	67.07	66.89	64.74	65.27	66.66	60.13	63.44	64.33
16	64.49	66.66	67.12	72.50	67.97	63.25	61.00	64.28	65.49	60.13	65.29
17	62.16	58.00	63.69	67.36	61.33	58.51	68.42	63.31	58.91	66.06	62.78
18	65.89	64.33	68.65	68.75	59.73	69.93	71.15	63.44	61.15	67.34	66.04
19	61.84	65.06	62.58	68.83	57.51	67.88	69.12	65.06	64.82	62.25	64.50
20	68.05	65.27	68.54	72.25	69.72	61.53	67.34	65.30	59.42	70.13	66.76
21	62.66	65.62	71.61	66.66	64.18	62.74	69.18	62.34	62.16	63.69	65.08
22	64.96	67.87	66.90	63.38	60.26	68.62	64.74	70.06	62.17	67.80	65.68
23	65.51	67.10	60.16	68.02	62.87	64.10	64.33	67.33	63.12	66.66	64.92
24	67.28	63.57	70.62	62.50	61.74	62.42	68.11	60.13	71.12	65.06	65.26
25	60.91	64.42	70.88	66.24	64.47	64.23	65.38	70.94	69.86	66.66	66.40
26	69.67	64.05	66.23	64.74	65.38	60.28	68.96	66.19	60.28	68.02	65.38
27	67.36	64.58	63.08	61.37	69.67	64.74	61.48	68.38	70.80	71.61	66.31
28	67.30	67.30	66.44	71.14	64.63	65.58	59.71	65.33	65.43	67.36	66.02
29	68.51	69.17	64.10	71.24	57.93	67.36	70.58	70.74	64.66	67.53	67.18
30	66.66	67.54	70.50	69.07	65.54	68.65	68.18	64.53	71.32	66.66	67.87
31	68.75	63.50	65.64	64.66	65.78	67.80	64.51	68.49	66.22	70.74	66.61
32	64.41	60.12	67.51	69.28	64.82	67.14	70.86	67.85	68.08	64.96	66.50

<표 4> SUA의 은닉층 노드수에 따른 테스트결과(%)

HLN \ 실험	1	2	3	4	5	6	7	8	9	10	평균
16	68.75	58.82	66.66	63.41	63.41	63.81	68.00	54.38	63.82	66.07	63.71
8	50.00	50.00	62.96	55.55	56.86	64.00	54.34	61.70	66.10	69.81	59.13
32	61.70	58.00	61.90	62.26	56.25	71.73	59.18	69.76	65.51	71.69	63.80
30	64.06	65.30	64.28	60.78	69.04	56.00	62.74	58.49	65.95	72.54	63.92
4	63.04	60.00	56.89	65.38	72.34	67.39	59.57	62.50	70.58	58.69	63.64
11	68.96	64.58	54.00	65.21	59.64	63.79	64.15	62.50	55.55	60.78	61.92

주) HLN : Hidden Layer Node, n=16, m=1