

유전 알고리즘 기반 제품구매예측 모형의 개발*

민재형** · 정철우***

A GA-based Classification Model for Predicting Consumer Choice

Jae H. Min** · Chulwoo Jeong***

■ Abstract ■

The purpose of this paper is to develop a new classification method for predicting consumer choice based on genetic algorithm, and to validate its prediction power over existing methods. To serve this purpose, we propose a hybrid model, and discuss its methodological characteristics in comparison with other existing classification methods. Also, we conduct a series of experiments employing survey data of consumer choices of MP3 players to assess the prediction power of the model. The results show that the suggested model in this paper is statistically superior to the existing methods such as logistic regression model, artificial neural network model and decision tree model in terms of prediction accuracy. The model is also shown to have an advantage of providing several strategic information of practical use for consumer choice.

Keyword : Consumer Choice Model, Binary Classification Method, Genetic Algorithm

1. 서론

제품구매예측은 소비자의 선택에 영향을 미치는 제품의 특성이 무엇이고, 그 영향력은 어느 정도인지

에 대한 정보를 객관적으로 제공해 줌으로써 의사결정자의 합리적인 판단에 도움을 준다는 점에서 의의를 가진다. 특히, 기업의 의사결정자는 제품구매예측 모형이 제공하는 정보를 제품의 생산 및 마케팅

논문접수일 : 2008년 11월 17일 논문게재확정일 : 2009년 06월 02일

논문수정일(1차 : 2009년 03월 16일, 2차 : 2009년 05월 09일)

* 이 연구는 2008년도 서강대학교 교내연구비 지원에 의한 연구임(200810051.01).

** 서강대학교 경영전문대학원

*** 서강대학교 대학원 경영학과

† 교신저자

팅 과정에서 효과적인 전략을 구사할 수 있는 수단으로 이용하게 된다.

기존의 제품구매 예측 모형은 크게 두 가지로 나누어볼 수 있는데, 하나는 통계 기반의 전통적 기법에 의해 구축된 모형이고, 다른 하나는 인공지능 기반의 모형이다. 통계 기반 모형의 예로는 MNL(multinomial logit model), MNP(multinomial probit model), 다변량판별분석(multidiscriminant analysis : MDA), SVM(support vector machine) 등을 이용한 모형을 들 수 있으며, 인공지능 기반의 모형으로는 인공신경망(artificial neural network : ANN), 의사결정나무(decision tree : DT), 사례기반추론(case-based reasoning : CBR), 자료포괄분석(data envelopment analysis : DEA) 등을 이용하여 구축된 모형들이 있다.

제품구매예측을 위한 이상의 모형들은 각각 고유한 장점과 함께 단점도 가지고 있다. 이는 곧 기존에 많은 제품구매예측 모형이 개발되었음에도 불구하고 기존의 모형들이 가지는 단점을 보완하고 새로운 장점을 가진 모형을 개발하고자 하는 노력의 근거가 된다.

본 연구에서는 이러한 노력의 일환으로 제품구매 예측을 위한 새로운 분류모형을 개발하고, 모형의 타당성을 확인하고자 한다. 이를 위해 본 연구에서는 우선, 새로운 분류모형의 구조를 수리적으로 제시하였다. 그리고 국내에서 유통되는 5개 사 12종류 MP3 플레이어 제품의 구매의도에 관한 설문자료를 이용하여 본 연구에서 개발한 모형의 예측성과를 도출하고, 이를 기존 모형의 성과와 비교하여, 본 연구모형의 우수성을 확인하였다. 또한 본 연구모형의 분석결과를 토대로 본 모형이 제공하는 정보의 실무적 유용성에 대하여 논의하였다.

2. 이론적 배경

2.1 이산선택모형

제품구매예측 및 소비자 행태분석을 위한 이산선택모형(discrete choice model)의 발달과정은 크게

두 가지로 나누어 볼 수 있다. 하나는 1980년대 이후로 소비자 패널자료의 이용이 가능해짐에 따라 많은 연구가 진행되어온 통계기반 모형이고, 다른 하나는 1990년대 이후 활발하게 연구가 이루어지고 있는 인공지능 기반 모형이다.

우선, 통계기반 모형의 발달과정을 살펴보면, 마케팅 분야를 비롯한 여러 분야의 연구에서 이산선택모형의 발달과정 초기부터 가장 흔하게 이용되어진 모형은 MNL(multinomial logit model)이다(Ben-Akiva and Boccara, 1995; Guadagni and Little, 1983). MNL은 비관련 대안들의 독립성(IIA; independence of irrelevant alternatives)과 같은 매우 엄격한 가정을 필요로 하는 한계점은 가지고 있지만, 모수 추정이 비교적 쉽다는 장점 때문에 많은 연구자들에게 선호되었다. 그럼에도 불구하고 엄격한 IIA의 가정을 완화시켜 보다 일반적인 상황에 적용될 수 있는 모형에 대한 연구가 계속되었는데, 그 일환으로 MNP(multinomial probit model)의 적용 연구들이 발표되었다(Bolduc, 1999; Bolduc et al., 1997; Paap and Franses, 2000).

또한 컴퓨팅 환경의 급속한 발달에 따라 선택확률이 닫힌 형태(closed form)가 아니라 할지라도 시뮬레이션을 이용해 쉽고도 빠르게 우도함수를 추정할 수 있는 모형에 대한 연구가 이루어질 수 있는 환경이 조성되었다(Borsch-Supan and Hajivassiliou, 1993; Keane, 1994; McFadden, 1989). 아울러 2000년대에 들어와서는 베이지안 추론을 이용하여 기존의 MLE(maximum likelihood estimation) 방법과는 달리 추정계수에 대해 고정적인 값이 아닌 확률분포를 가정함으로써 우도함수의 직접적인 계산 없이 복잡한 모형의 구축을 가능하게 하는 일련의 연구가 발표되었다(Allenby et al., 2004; Gilbride and Allenby, 2004; Lachaab et al., 2006; McCulloch et al., 2000; Rossi and Allenby, 2003). 그리고 이와 함께 기존에 나온 통계기반의 모형들을 결합한 혼합모형에 대한 연구도 이루어졌다(Brownstone et al., 2000; McFadden and Train, 2000). 이외에도 다변량판별분석 모형(Curram and Mingers, 1994)과

SVM(support vector machine)(안현철 외, 2005)이 구매예측을 위한 통계기반의 이산선택모형으로 연구되었다.

다음으로, 제품구매예측을 위해 적용된 인공지능 기반 모형으로는 ANN(송수섭, 이의훈, 2001; 한상만 외, 2000; Agrawal and Schorling, 1996; Ha et al., 2005; Kumar et al., 1995; Venugopal and Baets, 1994; West et al., 1997; van Wezel and Baets, 1995), DT(Curram and Mingers, 1994), CBR(Ahn et al., 2006; Ahn et al., 2007; Chiu, 2002), DEA(Athanasopoulos and Curram, 1996) 등을 들 수 있다.

이상에서 살펴본 기존의 모형들은 해당 모형이 어떠한 방법에 기반을 두었는가에 따라 고유한 장점과 함께 단점도 가지고 있다. 예를 들어, MNL은 보수추정이 비교적 쉽다는 장점은 있지만, IIA라는 엄격한 가정을 필요로 하므로, 일반적인 상황에서의 적용이 용이하지 않다는 한계가 있다. 베이지안 추론을 적용한 모형의 경우에는 우도함수의 직접적인 계산 없이 모형 구축이 가능하다는 장점은 있지만, 추정모수에 대한 적합한 분포 가정이 필요하다는 한계를 가진다. ANN의 경우는 상대적으로 다른 모형들에 비해 예측력이 높다고 평가받고 있지만, 이러한 높은 예측정확도에도 불구하고 예측결과에 대한 설명력이 떨어져 실무적으로 합의하는 바가 적다는 한계를 가지고 있다. 또한 DT의 경우, 분류 또는 예측을 위한 규칙을 도출한다는 장점은 갖고 있지만, 과적합 문제가 발생할 수 있는 위험이 다른 기법들에 비해 상대적으로 높다는 단점이 있다.

이러한 이유로 기존에 많은 제품구매예측 모형이 개발되었음에도 불구하고 기존의 모형들이 가지는 단점을 보완하고 새로운 장점을 가진 모형을 개발하고자 하는 노력은 계속되고 있다. 또한 같은 이유로 올바른 예측을 위해서는 어떠한 특정 모형에 집착하기보다는 문제 상황에 맞게 여러 가지 모형의 장점을 함께 활용하는 것이 권장된다고 하겠다.

본 연구에서 제안하는 모형은 통계 기반 기법인 k -means 군집분석과 인공지능기법인 CBR에서 이용하는 거리의 개념, 그리고 최적화 방법으로 유전

알고리즘을 함께 이용하여 개발한 혼합모형(hybrid model)이다. 본 연구모형은 k -means 군집분석과 CBR의 원리를 일부 이용하였지만, 그러한 기법들과 차별화되는 다음과 같은 특징을 가진다.

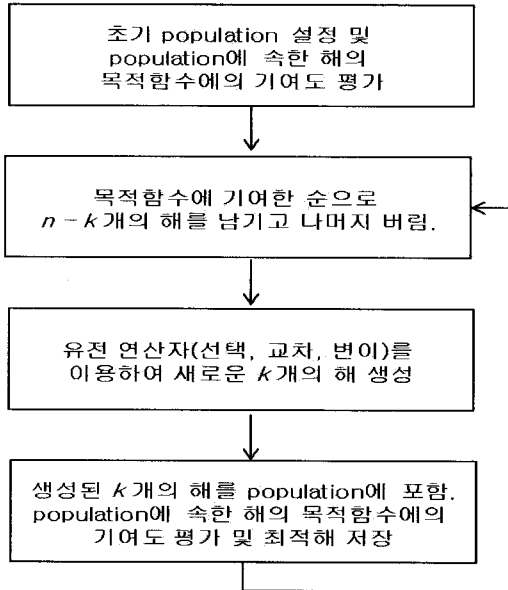
첫째, 본 연구모형은 특정 개수의 중심을 설정하고, 그 중심과 개체간 거리를 측정하여 중심으로부터 가까운 거리에 있는 개체끼리 군집화 한다는 점에서 k -means 군집분석과 유사하다. 그러나 본 연구모형은 독립변수와 함께 종속변수를 모형에 포함하여 모형을 구축한다는 점에서 독립변수만을 가지고 군집화를 하는 k -means 군집분석과 차별된다.

둘째, 본 연구모형은 예측이 필요한 개체와 비교 대상 개체들 사이의 거리를 측정하여 가장 거리가 가까운 비교 대상 개체의 특성에 따라 예측 대상 개체를 분류한다는 점에서 CBR과 유사하다. 그러나 CBR에서는 비교 대상 개체들이 실제로 데이터 중에 존재하는 개체인데 반해, 본 연구모형의 비교 대상 개체들은 실제로 존재하는 것이 아닌 가상 개체(virtual subjects)이다. 이러한 차이는 CBR이 국지적 학습방법(local learning approach)인데 반해, 본 연구모형은 전체적 학습방법(global learning approach)이라는 사실을 시사한다. 즉, CBR은 예측하고자 하는 개체와 비교 대상 개체 모두를 비교한 후, 가장 가까운 k 개의 개체를 참조하여 예측결과를 생성하는 국지적 학습방법을 이용한다. 반면, 본 연구모형은 유전 알고리즘을 이용하여 특정 개수의 가상 개체를 찾은 후, 그 가상 개체들과 예측하고자 하는 개체 간의 거리가 가장 가까운 가상 개체의 제품구매여부에 따라 예측 대상 개체를 분류하는 전체적 학습방법을 이용한다.

2.2 유전 알고리즘

유전 알고리즘은 생태계의 선택(selection), 교차(cross-over), 변이(mutation)와 같은 유전 메커니즘을 수리적으로 모형화한 것으로 최적해를 구하기 위해 방대하고 복잡한 공간을 확률적으로 탐색하는 특징을 갖고 있다. 선택, 교차, 변이 연산자를 모두 포

합한 유전 알고리즘의 완성된 구조는 Holland(1975)에서 처음으로 정리되었는데, 유전 알고리즘의 전형적인 구조를 간단히 도시하면 [그림 1]과 같다.



자료) Kecman, V., Learning and Soft Computing, MIT Press, Cambridge, Mass, 2001, p.499.

[그림 1] 유전 알고리즘의 구조

[그림 1]을 간단히 설명하면 우선, 유전 알고리즘은 분석자가 설정한 특정 개수(n)의 해로 이루어진 집단을 생성해내는데, 이를 population이라고 부른다. 다음으로, population에 속한 해를 목적함수에 기여하는 순으로 평가한다. 그리하여 목적함수에 가장 큰 기여를 한 $n-k$ ($k \leq n$)개의 해를 population에 남겨두고, 나머지 k 개의 해는 선택, 교차, 변이 등의 연산자에 의해 생성된 새로운 k 개의 해로 대체하여 크기 n 의 새로운 population을 만들어낸다. 이러한 과정을 미리 설정한 정지조건(stopping condition)을 만족할 때까지 반복 수행한 후 population에 남은 k 개의 해 중 목적함수에 가장 큰 기여를 한 해를 최적해로 도출하게 된다. 이와 같은 알고리즘의 구조적 특징으로 인해 유전 알고리즘은 다양한 제약조건을 포함하는 문제 상황에서 최적해를 추정하는데 널리 활용되어 왔다.

본 연구에서는 이진분류의 예측정확도를 최대화 하는 변수별 기준치와 구매집단과 비구매집단을 대표할 수 있는 중심의 변수값을 추정하기 위해 유전 알고리즘을 이용하였다.

3. 수리적 모형

본 연구에서 제안하는 이진분류모형을 수리적으로 표현하면 다음과 같다. 우선, 집단 중심과 개체¹⁾ 사이의 동질성 또는 이질성을 나타내는 거리 개념을 표현하면 식 (1)과 같다.

$$d_{ki} = \sum_j (w_j / \sum_j w_j) |X_{kj} - X_{ij}| \quad (1)$$

여기서, d_{ki} : 중심 k 와 개체 i 사이의 거리

($i=1, 2, \dots, n$)

X_{kj} : 중심 k 의 변수 j 관측값

($j=1, 2, \dots, m$)

X_{ij} : 개체 i 의 변수 j 관측값

w_j : 변수 j 의 가중치

식 (1)에서 개체 i 에 대하여 d_{ki} 는 중심의 개수²⁾만큼 계산되며, 이는 각 중심으로부터 개체 i 가 떨어져있는 거리를 의미한다. 다음으로, 식 (1)에서 구한 거리를 이용하여 개체 i 의 구매확률을 나타내면 식 (2)와 같다.

$$P(i) = \frac{\min(d_{q_i}'s)}{\min(d_{p_i}'s) + \min(d_{q_i}'s)} \quad (2)$$

여기서, $P(i)$: 개체 i 의 구매확률

d_{p_i} : 구매집단 중심 p 와 개체 i 사이의 거리

- 1) 여기서 '개체'는 각 제품을 가리키고, '집단 중심'은 연구대상 제품을 특정 개수의 집단으로 군집화 할 때 각 집단을 대표할 수 있는 중심을 가리킨다.
- 2) 중심의 개수는 구매집단과 비구매집단에 대해 각각 1이상의 정수이다. 이 값은 서로 같을 수도 있고 다를 수도 있는데, 이는 모형을 적용하는 데이터에 따라 달라진다. 따라서 최적의 중심 개수를 찾는 것이 본 연구모형의 설계에서 중요한 논점이 된다.

d_{qi} : 비구매집단 중심 q 와 개체 i 사이의 거리

식 (2)에서 $P(i)$ 는 다음과 같은 원리에 따라 계산된다. 먼저, 개체 i 와 구매집단의 중심들 사이의 거리를 구하여 이 값의 최소값을 A 라고 하고, 개체 i 와 비구매집단의 중심들 사이의 거리를 구하여 이 값의 최소값을 B 라고 하면, $P(i)$ 는 A 와 B 의 합에 대해 B 가 차지하는 비율과 같다. 여기서, A 와 B 의 합에 대해 B 가 차지하는 비율이 커질수록 개체 i 는 비구매집단 중심과는 멀어지고 구매집단 중심과는 가까워지며, 따라서 개체 i 의 구매확률 $P(i)$ 는 커지게 된다. 마찬가지로 A 와 B 의 합에 대해 B 가 차지하는 비율이 작을수록 개체 i 의 구매확률 $P(i)$ 는 작아지게 된다.

다음으로 식 (3)은 $P(i)$ 에 따라 개체 i 를 이진분류하는 과정을 표현한 것이다.

$$\begin{aligned} CP_i &= 1 \text{ if } P(i) \geq 0.5 \quad \forall i \in \{1, 2, \dots, n\}, \\ CP_i &= 0 \text{ otherwise} \end{aligned} \quad (3)$$

여기서, CP_i : 개체 i 에 대한 구매여부 예측

식 (3)에서 $P(i)$ 의 값이 0.5보다 크게 되면 개체 i 를 구매할 확률이 구매하지 않을 확률보다 높아진다. 따라서 0.5를 임계값(cut-off value)으로 하여, $P(i)$ 가 임계값보다 크면 개체 i 를 구매하는 것으로 예측하고, $P(i)$ 가 임계값보다 작으면 개체 i 를 비구매하는 것으로 예측한다.

한편, 식 (4)는 실제 구매여부에 대하여 옳게 예측한 비율을 목적함수로 하여 이 값을 최대화하도록 식 (1)에 있는 집단 중심의 변수값과 변수별 가중치를 탐색하도록 하는 모형이다.

$$\begin{aligned} \text{Max } HR &= \frac{1}{n} \sum_{i=1}^n H_i \\ \text{s.t. } H_i &= 1 \text{ if } CP_i = C_i \quad \forall i \in \{1, 2, \dots, n\}, \\ H_i &= 0 \text{ otherwise.} \end{aligned} \quad (4)$$

여기서, HR : 실제 구매여부에 대한 예측정확도

C_i : 개체 i 의 실제 구매여부($C_i = 0, 1$)

H_i : 개체 i 의 구매여부에 대한 실제와 예측의 일치여부

n : 개체 수

여기서 목적함수를 최대화하는 집단 중심의 변수값 (X_{kj})과 변수별 가중치(w_j)는 유전 알고리즘을 이용하여 탐색된다.

4. 실증분석

4.1 데이터 정리

본 연구의 실증분석에 사용된 데이터는 2006년 11월 10일부터 동년 12월 20일 사이에 서울 시내 4개 대학의 학생들을 대상으로 국내에서 유통되는 5개 사 12종류의 MP3 플레이어 제품에 대한 구매의도를 설문조사하여 수집한 것이다.³⁾ 12종류의 MP3 플레이어는 설문지 작성 기간에 야후 쇼핑몰(<http://kr.shopping.yahoo.com>)에서 판매되고 있는 1002종류의 MP3 플레이어 가운데 약 600종류가 포함된 가격범위를 설정한 후, 그 가격범위를 상·중·하 세 가지 가격층으로 나누고 각 가격층에서 4종류씩 뽑아서 선정되었다.

설문항목은 다음과 같이 구성되었다. 우선, 12종류의 MP3 플레이어의 가격, 브랜드, 기능 및 디자인 등에 대한 정보를 제공하고, 구매하기 원하는 제품을 선택하도록 하였다. 그리고 선택한 제품과 선택되지 않은 제품을 포함하여 각각의 제품에 대해 브랜드 점수를 0점부터 1점 사이의 비율척도로 평가하도록 하였고, 기능과 디자인 측면에서 호감이 가는 정도(선호도)를 1점부터 9점까지의 리커트 척도로 평가하도록 하였다. 설문지에 기록된 12개 MP3 플레이어의 가격은 설문지 작성 기간 중 야후 쇼핑몰에 게시된 판매가를 이용하였다.

3) 본 연구의 실증분석을 위한 기초 데이터를 제공해 준 서강대학교 경영전문대학원의 김주영 교수에게 감사를 표한다.

이상의 설문조사를 통해 구매여부, 가격, 브랜드, 선호도 등 총 4개 변수에 대해 구매응답 418개, 비구매응답 2314개의 데이터를 얻을 수 있었다. 그리고 구매응답 418개 중 결측치를 갖는 응답 53개를 제외한 나머지 365개와 비구매응답 2314개 중 무작위로 추출한 365개를 합쳐 총 730개의 데이터를 분석을 위한 최종 데이터로 선정하였다.⁴⁾

다음으로 설문조사에서 구매로 응답한 데이터와 비구매로 응답한 데이터를 대상으로 층화표본추출 방식을 이용하여 각각 219개씩을 뽑아 총 438개로 이루어진 훈련용 데이터를 만들고, 이는 모형의 구축을 위해 이용하였다. 또한 동일한 방식으로 뽑은 구매 73개, 비구매 73개로 이루어진 146개 데이터는 최적의 모형을 찾기 위한 테스트용 데이터로 이용하였다. 그리고 동일한 방식으로 뽑은 나머지 146개 데이터는 구축된 모형을 검증하기 위한 검증용 데이터로 이용하였다. 이상의 방식으로 정리한 최종 데이터의 용도별 크기를 나타내면 <표 1>과 같다.

<표 1> 용도별 데이터 크기

	훈련용	테스트용	검증용
구매	219	73	73
비구매	219	73	73
합계	438	146	146

한편, 최종 데이터에 대한 변수별 요약통계량을 정리하면 <표 2>와 같다. 본 연구에서는 <표 2>의 변수별 평균과 표준편차를 이용하여 각 개체의 관측값을 표준화한 후, 이를 분석에 이용하였다.

<표 2> 변수별 기술통계

	표본 크기	최소 값	최대 값	평균	표준 편차
가격	730	9.8	18.2	14.04	2.25
브랜드	730	0	1	0.17	0.13
선호도	730	1	9	4.77	1.59

4) 본 연구에서 자료의 비대칭성을 조정하여 구매제품과 비구매제품의 자료 수를 동일하게 맞추는 이유는 본 연구에서 제시하는 모형의 실무적 목적이 구매제품 자료를 비구매제품 자료로부터 분류해내고, 그 특성을 파악하는 것에 초점이 맞추어져 있기 때문이다. 즉, 그 중요성에 있어서 구매제품에 속하는 제품의 특성을 파악하는 것이 비구매제품에 속하는 제품의 특성을 파악하는 것보다 우선시 된다고 하겠다. 만약 데이터의 비대칭성을 조정하지 않고 모형을 구축한다고 하면, 모형의 목적함수인 예측정확도를 최대화하기 위해서 자료의 대부분에 해당하는 비구매제품에 대한 예측정확도를 높이는 방향으로 변수별 가중치 및 각 중심의 변수값이 추정되어질 것이며, 이 경우 모형은 구매제품의 특성은 제대로 반영하지 못할 것이다. 따라서 자료의 비대칭성을 조정하는 것이 본 연구모형의 목적을 위해서 타당하다고 하겠다. 다만, 모형이 추구하는 목적이 무엇이냐에 따라 자료의 비대칭성을 그대로 유지하고 분석을 수행할 수도 있는데, 이 경우 모형의 목적함수는 이진분류의 예측정확도를 최대화하는 것이 아닌 다른 형태를 취해야 할 것이다. 예를 들어, 만약 구매제품을 비구매제품으로 잘못 예측하는 오류비용과 비구매제품을 구매제품으로 잘못 예측하는 오류비용을 정확히 추정할 수 있다면, 총오류비용을 최소화하도록 변수 가중치와 중심의 변수값을 추정하는 모형을 만들 수 있을 것이다.

4.2 모형의 설계

본 연구에서 제안하는 모형의 설계와 관련하여 중요한 논점 중 하나는 최적의 중심 개수를 찾는 것이다. 최적의 중심 개수를 찾는데 있어, 본 연구에서는 두 가지 기준을 고려하였다. 첫째, 훈련용 데이터에 대해 모형의 예측정확도가 높고, 동시에 훈련용 데이터와 테스트용 데이터에 대한 예측정확도의 차이가 크지 않는 모형이 좋은 모형이다. 둘째, 모형이 개체들을 각 중심의 집단으로 분류하였을 때, 각 집단에 할당된 개체의 수가 지나치게 적으면 좋은 모형이 아니다. 이상의 두 가지 기준은 중심의 개수가 지나치게 많을 때 발생할 수 있는 과적합 문제를 방지하기 위해 마련되었다.

이러한 기준 하에 본 연구에서는 집단중심의 수를 구매집단과 비구매집단에 대해 각각 1개부터 4개까지 늘려가면서 예측정확도의 추이를 살펴 집단중심의 최적 개수를 구하였다. 집단중심의 수가 구매집단과 비구매집단에 대해 각각 1개인 것을 [모형 1], 2개인 것을 [모형 2], 3개인 것을 [모형 3], 4

개인 것을 [모형 4]로 하여 실험을 진행하였다.

유전 알고리즘을 적용하는데 있어 본 연구모형이 가지는 염색체(chromosome) 구조를 설명하면 다음과 같다. 각 염색체는 가중치에 대한 유전자 3개와 집단중심의 변수값에 대한 (집단중심의 수) \times 3개의 유전자로 이루어져 있다. 가중치 유전자는 [0, 1] 범위의 실수이고, 집단중심별 변수 유전자는 [-3, 3] 범위 내에서 실수값을 취하도록 설정하였다. 초기 값은 가중치 유전자에 대해서는 [0, 1] 범위의 난수를 발생시켜 이용하였고, 집단중심 변수값의 유전자는 표준정규분포를 따르는 확률변수를 발생시켜 이용하였다.

최적해를 찾기 위해 population 크기는 100으로 하였고, mutation rate는 변수별 가중치(w_j)와 집단중심 변수값(X_{kj})에 대해서 모두 0.1로, cross-over rate는 w_j 와 X_{kj} 에 대해서 모두 0.5로 하였다. Stopping condition은 모든 실험에 대해 동일하게 iteration의 수가 10,000이 되면 자동으로 멈추도록 설정하였다. 분석도구로는 유전 알고리즘 소프트웨어인 Evolver 4.0을 사용하였다.

4.3 분석결과와 논의

<표 3>은 모형별 예측정확도를 정리한 것이다. <표 3>을 보면 훈련용 데이터와 테스트용 데이터에 대해 중심의 개수가 증가할수록 예측정확도 역시 증가하는 것을 볼 수 있다.

<표 3> 모형별 예측정확도

모형	훈련용	테스트용
1	76.5%	74.7%
2	78.5%	74.7%
3	78.5%	75.3%
4	78.8%	76.7%

한편, <표 4>는 모형별로 설정한 중심에 할당된 훈련용 데이터와 테스트용 데이터의 개수를 보여준다.

<표 4> 모형별 각 집단에 할당된 관측값 도수

모형	집단	중심	훈련용	테스트용
1	구 매	1	228	78
	비구매	2	210	68
2	구 매	1	193	71
		2	20	5
	비구매	3	153	49
		4	72	21
3	구 매	1	1	1
		2	60	21
		3	167	58
	비구매	4	83	20
		5	121	43
		6	6	3
4	구 매	1	157	59
		2	5	2
		3	0	0
		4	56	14
	비구매	5	59	17
		6	160	54
		7	1	0
		8	0	0

<표 4>에서 [모형 4]는 훈련용 데이터와 테스트용 데이터 모두에 있어서 구매집단의 [중심 3]과 비구매집단의 [중심 8]에 할당된 개체는 전혀 없음을 알 수 있다. 또한 [모형 3]은 훈련용 데이터와 테스트용 데이터 모두에 대하여 구매집단의 [중심 1]에 할당된 개체가 각각 1개에 불과하였다. 이러한 결과는 [모형 4]는 물론이고, [모형 3] 역시 중심의 개수가 필요 이상으로 많이 설정되었음을 의미한다. 따라서 [모형 3]의 경우, 구매집단에 대해서는 중심의 개수를 2개로 줄이는 것이 바람직한 것으로 판단된다. 그렇게 함으로써 각 중심의 변수값과 변수별 가중치를 찾는 시간을 절약할 수 있고, 동시에 필요 없는 중심이 들어감으로써 발생할 수 있는 추정값의 편의(bias)를 제거할 수 있는 효과를 거둘 수 있다. 그런데 [모형 3]에서 [중심 6]의 경우도 [중심 1]과 마찬가지로 할당된 개체의 수가 지나치게 적지

얇은가라는 의문이 제기될 수 있다. 그러나 만약 [모형 3]에서 [중심 6]까지 제거하면 중심의 개수는 구매집단에 대해 2개, 비구매집단에 대해서도 2개가 되어 결국 [모형 3]은 [모형 2]와 같아지게 된다. 하지만 <표 3>에서 [모형 2]보다는 [모형 3]이 예측 정확도 측면에서 더 우수한 모형으로 나타난 점을 고려하면, [모형 3]에서 중심의 개수를 2개 제거하는 것보다는 1개만 제거하는 것이 바람직할 것이다.

따라서 본 연구에서는 [모형 3]에서 구매집단에 대해 중심의 개수를 1개 줄여 2개로 설정하고, 비구매집단에 대해서는 [모형 3]과 같이 중심의 개수를 3개로 설정한 모형을 최적 모형으로 선택하였다. 이와 같이 중심의 개수를 설정한 최적 모형에 대해 유전 알고리즘을 이용하여 추정된 각 집단 중심의 변수값과 변수별 가중치를 정리하면 <표 5> 및 <표 6>과 같다. <표 5>와 <표 6>의 추정된 값은 훈련용 데이터에 대한 실제 구매여부와 구매여부의 예측이 일치하는 비율을 최대화하는 값이다.

<표 5> 변수별 가중치 추정값

	가격	브랜드	선호도
w_j	0.16	0.18	0.58
$w_j / \sum_j w_j$	0.17	0.20	0.64

<표 6> 집단 중심별 변수값

		가격	브랜드	선호도
구 매	중심 1	-1.22	0.46	0.92
	중심 2	0.52	0.64	0.94
비구매	중심 3	1.42	0.26	-0.80
	중심 4	-1.72	1.10	-0.34
	중심 5	-0.55	-1.22	-1.22

<표 5>의 변수별 가중치의 추정값은 구매확률에 가장 큰 영향을 미치는 변수가 무엇이고, 이때의 변수별 가중치는 얼마인지를 말해준다. 이 값은 앞서 식 (4)에서 실제 구매여부에 대한 예측정확도를 최대화하도록 식 (2)에 의해 각 개체의 구매확률이 결정되도록 추정된 값이다. <표 5>를 보면, MP3 플

레이어의 구매확률 결정에 가장 큰 영향을 미친 변수는 선호도로서, 64%의 상대적 가중치로 구매확률에 영향을 미친 것을 알 수 있다. 그리고 가격과 브랜드 변수는 상대적으로 낮은 17%와 20%의 가중치로 구매확률에 영향을 미친 것으로 나타났다. 다음으로 <표 6>은 각 중심의 추정된 변수값을 보여주고 있는데, 여기서 구매집단에 해당하는 [중심 1]과 [중심 2]의 선호도 변수의 추정값은 비구매집단에 해당하는 중심들의 선호도 변수 추정값에 비해 월등히 높은 것을 알 수 있다. 이상의 추정값을 이용하여 이진분류 하였을 경우의 훈련용, 테스트용, 검증용 데이터에 대한 예측정확도는 각각 79.2%, 75.3%, 72.6%로 나타났다.

이 경우, 각 중심의 집단에 할당된 개체의 도수를 정리하면 <표 7>과 같다. <표 7>에서 보듯이 각 중심에 할당된 개체의 수가 지나치게 적은 중심은 없는 것으로 나타났다.

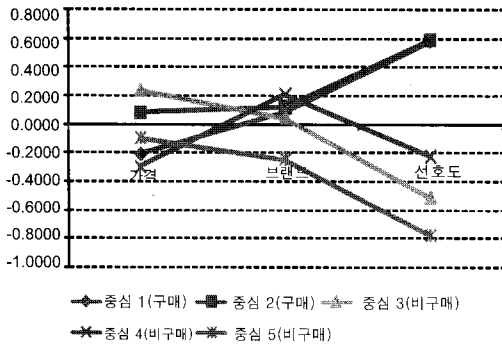
<표 7> 최적 모형의 각 중심 집단에 할당된 개체 도수

집단	중심	훈련용	테스트용	검증용
구 매	1	57	14	20
	2	163	61	39
비구매	3	58	18	29
	4	75	26	36
	5	85	27	22

이제 본 연구모형이 제공하는 결과물이 실무적 차원에서 어떠한 의미를 갖는지 살펴보자. 이를 위해 각 중심의 변수값과 변수별 가중치를 곱한 가중 변수값을 구하여 <표 8>과 같이 정리하였으며, 이를 그래프로 나타내면 [그림 2]와 같다.

<표 8> 각 중심의 가중 변수값

	가격	브랜드	선호도
중심 1(구매)	-0.2064	0.0907	0.5858
중심 2(구매)	0.0887	0.1260	0.5948
중심 3(비구매)	0.2402	0.0505	-0.5094
중심 4(비구매)	-0.2914	0.2153	-0.2187
중심 5(비구매)	-0.0927	-0.2387	-0.7732



[그림 2] 각 중심의 가중 변수값

<표 8>과 [그림 2]를 보면 소비자들이 선택한 MP3 플레이어와 선택하지 않은 MP3 플레이어의 전형 (prototype)을 파악할 수 있다. 예를 들어, 구매집단에 해당하는 [중심 1]과 [중심 2]에 해당하는 MP3 플레이어는 선호도(기능과 디자인 측면의 선호도)에 있어 다른 중심들에 비해 높은 점수를 받았고, 브랜드 측면에서는 비슷한 점수를 받았지만, 가격에 있어서 [중심 1]에 해당하는 MP3 플레이어는 가격이 낮은 편에 속하고, [중심 2]에 해당하는 MP3 플레이어는 가격이 높은 편에 속하였다.

이러한 결과를 통해 MP3 플레이어 선택·구매하는 소비자는 크게 두 부류로 나뉜다는 사실을 유추할 수 있다. 한 부류는 가격이 싸면서도 기능과 디자인 측면에서 만족할 만한 MP3 플레이어를 원하는 소비자 집단이고, 다른 한 부류는 가격이 비싸더라도 기능과 디자인 측면에서 뛰어난 MP3 플레이어를 원하는 소비자 집단이다. 여기서 주목할 점은 두 부류 모두 기능과 디자인 측면에서 우수한 MP3 플레이어를 원하고 있다는 사실이다. 다시 말해, 가격이나 브랜드 측면에서 아무리 유리하다고 할지라도 기능과 디자인 측면에서 만족을 주지 못한다면 소비자의 선택을 받을 수가 없다는 것을 알 수 있다. 예를 들어, [중심 4]에 해당하는 MP3 플레이어는 가격이 싸고, 브랜드 측면에서 높은 점수를 받았음에도 불구하고 기능과 디자인 측면에서의 선호도가 낮음으로 인하여 소비자가 외면하는 경우라고 하겠다.

다음으로, 본 연구모형이 가지는 또 다른 실무적 유용성은 이미 출시되었거나 출시 예정인 MP3 플레이어에 대한 소비자들의 구매확률 정보를 식 (2)를 이용하여 파악할 수 있다는 것이다. 예를 들어, 출시 예정인 제품 A의 브랜드와 선호도에 대한 소비자들의 평가를 조사한 후, <표 2>의 평균과 표준편차를 이용하여 평가점수를 표준화한 값을 계산한 결과 각각 0.28과 0.77이 나왔고, 제품 A의 출시 예정 가격을 표준화한 값이 -0.91이라고 하자. 이 경우, 제품 A와 5개의 중심 사이의 거리를 <표 5>와 <표 6>의 값을 이용하여 측정하면 각각 0.1853, 0.4186, 1.3968, 1.0084, 1.6168이 된다. 여기서 제품 A와 구매집단 중심 사이의 거리 중 최소값은 0.1858이고, 제품 A와 비구매집단 중심 사이의 거리 중 최소값은 1.0084이다. 따라서 제품 A에 대한 소비자의 구매확률은 $1.0084 / (0.1852 + 1.0084) = 0.5444$ 로 계산된다. 그리고 또 다른 출시 예정 제품인 제품 B의 가격, 브랜드, 선호도에 대한 표준화된 값이 각각 1.76, 0.35, -0.49라고 할 때, 위와 동일한 방법으로 제품 B에 대한 소비자 구매확률을 계산하면 0.1901이 나온다.

이렇게 계산된 제품의 구매확률은 단순히 구매·비구매라는 이진분류결과에 비해 실무적 차원에서 보다 유용한 정보를 제공할 수 있다. 즉, 위의 예에서 제품 A의 경우, 이 제품에 대한 소비자들의 구매확률이 높게 나왔으므로, 해당 제조업체는 그 제품에 대한 집중화 전략이 유리한 반면, 제품 B의 경우에는 소비자들의 구매확률이 매우 낮게 나왔으므로 출시된다고 할지라도 소비자들이 외면할 가능성이 높다. 따라서 해당 제조업체는 제품 B의 기능이나 디자인 측면에서 소비자들의 선호도를 증진시킬 수 있는 방법을 강구하여 해당 제품의 구매확률을 높이거나, 이러한 노력이 가능하지 않을 경우, 해당 제품을 시장에서 과감히 철수시키는 전략을 취할 수도 있을 것이다. 이렇듯 본 연구모형이 제공하는 제품의 구매확률 정보는 제품에 대한 시장전략수립을 지원하는 역할도 수행할 수 있다.

4.4 기존 기법과의 성과비교

본 연구에서 개발한 분류모형의 예측성적을 평가하기 위해 기존의 분류모형과 그 성과를 통계적으로 비교하였다. 이를 위해 우선, 수집한 자료를 6:2:2의 비율로 무작위하게 할당하는 방식으로 5개의 데이터셋을 만들었다. 각 데이터셋에서 6:2:2의 비율로 할당된 데이터는 훈련용, 테스트용, 검증용 자료로 각각 이용된다. 그리고 5개 데이터셋 각각에 대하여 본 연구모형, 로지스틱 회귀분석(Logistic Regression : LR), ANN, DT를 이용하여 분석을 수행하였다. 기존 분석기법의 설계 특징을 간단히 기술하면 다음과 같다.

우선, LR의 경우, 가격, 브랜드, 선호도 변수가 독립변수로서 모두 입력되도록 설정하였다. ANN 모형은 입력계층과 출력계층, 그리고 하나의 은닉계층을 가지는 3층 퍼셉트론(three layer perceptron)

을 사용하였으며, 은닉층의 노드(node) 수는 입력변수의 개수와 동일한 3으로 설정하였다. 모형의 출력층은 1개의 출력노드로 구성되었고, 그 값은 [0, 1]의 범위에 존재하게 된다. DT 분석에서 분리기준(splitting criterion)은 유의수준 0.02에서 Chi-square 검정으로 하였고, 최대 뿌리 깊이는 6으로, 한 잎에서 최소 관측값의 수는 3으로 설정하였다.

이상의 방법으로 구축한 3가지 기존 분류모형의 각 데이터셋에 대한 예측정확도와 본 연구에서 제안한 모형의 예측정확도를 비교하면 <표 9>와 같다.

<표 9>의 순위는 각 데이터셋의 검증용 데이터에 대한 분석기법의 예측정확도를 순위화한 것이다. 본 연구모형의 예측정확도는 4개 데이터셋에서 1위를, 나머지 1개 데이터셋에 대해서는 2위를 기록하였다. 다음으로, <표 9>에 나와 있는 검증용 데이터에 대한 분석기법별 예측정확도가 통계적으로 유의한 차이를 보이는지 검정하기 위해 5개 데이터셋을 대상으로 쌍대 t-검정을 수행하였다. 그 결과를 정리하면 <표 10>과 같다.

<표 10>을 보면, 검증용 자료에 대해 유의수준 0.05에서 본 연구모형의 예측정확도는 로지스틱 회귀분석 모형, ANN 모형, DT 모형 등 모든 비교 대상 모형의 예측정확도와 비교하여 통계적으로 유의한 차이를 보이는 것으로 나타났으며,⁵⁾ 신뢰구간 결과로 유추해 볼 때 기존 기법보다 통계적으로 우월한 예측성적을 보이는 것으로 나타났다. 다시 말해, 본 연구모형은 기존의 예측모형에 비해 예측정확도 측면에서도 우수한 성과를 보이고 있음을 알 수 있다.

이러한 결과는 본 연구모형이 가지는 방법론적 장점 및 실무적 유용성과 함께 본 연구모형을 제품 구매예측을 위한 유망한 대안 기법으로 자리 잡게

<표 9> 분석기법별 예측정확도 비교

데이터셋	분석기법	훈련용	검증용	순위
1	본 연구 모형	79.2%	72.6%	1
	LR	74.4%	69.9%	4
	ANN	74.4%	70.5%	2
	DT	74.9%	70.5%	2
2	본 연구 모형	76.7%	74.0%	1
	LR	73.5%	71.9%	3
	ANN	75.6%	71.9%	4
	DT	73.5%	73.3%	2
3	본 연구 모형	78.3%	76.0%	1
	LR	73.3%	75.3%	3
	ANN	74.4%	74.0%	4
	DT	74.2%	76.0%	2
4	본 연구 모형	77.6%	76.7%	1
	LR	73.7%	76.0%	2
	ANN	76.5%	75.3%	3
	DT	74.0%	74.7%	4
5	본 연구 모형	78.8%	75.3%	2
	LR	74.7%	71.9%	4
	ANN	78.3%	76.7%	1
	DT	74.9%	72.6%	3

5) <표 10>에서 본 연구모형과 ANN 모형의 예측정확도 차이에 대한 검정통계량의 p-값은 0.0501로 나타났는데, 이는 유의수준 0.05와 거의 같은 수치라고 판단하였다. 또한 두 모형의 예측정확도 차이에 대한 95% 신뢰구간을 고려할 때 본 연구모형이 ANN 모형보다 우월한 성과를 보인다고 판단할 수 있다.

〈표 10〉 본 연구모형과 타 기법의 예측정확도 차이에 대한 쌍대 t-검정 결과

	Paired Differences							
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
				Lower	Upper			
본 연구 모형-LR	0.0219	0.0132	0.0059	0.0056	0.0383	3.7199	4	0.0205
본 연구 모형-ANN	0.0153	0.0123	0.0055	0.0000	0.0305	2.7752	4	0.0501
본 연구 모형-DT	0.0179	0.0141	0.0063	0.0003	0.0354	2.8307	4	0.0473

할 수 있다. 본 연구모형이 가지는 실무적 유용성을 정리하면 다음과 같다.

첫째, 각 제품의 구매여부에 대한 예측과 함께 구매확률을 제공해준다. 둘째, 본 연구모형은 변수별 가중치를 제공함으로써 소비자의 제품구매여부에 영향을 미치는 제품의 특성은 무엇이고, 그 영향력은 얼마나 되는지에 대한 정보를 제공해준다. 셋째, 본 연구모형은 제품들을 몇 개의 전형적인 특징을 가지는 집단으로 나누고, 각 집단의 중심이 가지는 변수값의 추정치를 제공한다. 이 추정치를 이용하여 구매예측이 필요한 제품과 각 중심과의 거리를 측정하면 그 제품이 어느 중심의 집단에 속하는지 판단할 수 있다. 또한 그 제품이 속한 집단의 중심이 가지는 전형적인 특징을 기반으로 하여 그 제품의 어떠한 특징이 소비자에게 매력적이거나 매력적이지 않은지를 쉽게 추측할 수 있다.

5. 결 론

본 연구에서는 제품구매예측을 위한 새로운 분류모형을 개발하고, 모형의 타당성을 실증자료를 이용하여 검증하였다. 모형의 타당성 검증을 위하여 국내 5개사 12종류의 MP3 플레이어 제품에 대한 설문자료를 이용하여 본 연구에서 제안한 모형과 기존 분류모형들의 예측성적을 비교하였다. 분석 결과, 본 연구모형은 예측정확도 측면에서 비교 대상 모형인 LR, ANN, DT 모형보다 모두 통계적으로 우수한 성과를 보임을 확인할 수 있었다.

본 연구에서 제안한 모형은 군집분석과 CBR에서

이용하는 거리의 개념, 그리고 유전 알고리즘을 함께 이용한 혼합모형으로, 제품군의 중심을 설정하고, 그 중심과 각 제품 사이의 유사성을 측정하여 그에 따라 제품을 분류하는 접근방식을 취하였다. 그리고 이 과정에서 유전 알고리즘은 실제 구매여부와 모형을 통한 구매여부 예측이 일치하는 비율을 최대화하는 변수별 가중치와 중심의 변수값을 추정하는 역할을 하였다.

본 연구의 모형은 예측성과의 우수성 뿐만 아니라 그 결과물로서 실무적으로 유용한 전략적 정보를 제공해 주는 장점도 가진다. 구체적으로, 본 연구모형은 각 제품의 구매여부 예측 뿐만 아니라 구매확률을 제공해주고, 소비자가 제품을 구매하는데 영향을 미치는 제품 특성과 그 영향의 정도에 대한 정보도 제공해주며, 각 제품의 분류와 함께 그 제품이 소비자에게 소구하는 특징 또한 쉽게 알 수 있게 한다. 이러한 결과는 본 연구의 모형이 제품구매예측을 위한 유망한 기법으로 활용될 수 있는 가능성을 기대하게 한다.

이러한 본 연구의 기여에도 불구하고 본 연구는 다음과 같은 한계점을 갖는다. 첫째, 본 연구모형의 설계에 있어 가격, 브랜드, 선호도와 같은 제품 특징만을 변수로 사용하였다. 향후 연구에서는 제품 특징 변수와 함께 소비자 특성 변수를 모형에 반영함으로써 소비자층에 따른 보다 세분화된 결과물 도출해낼 수 있으리라 기대한다. 둘째, 본 연구에서는 수리적 모형의 구축 시 집단 중심의 최적 개수를 알아내기 위한 알고리즘은 반영하지 않았으며, 이에 따라 집단 중심의 최적 개수를 찾기 위한 추

가적인 노력을 필요로 하였다. 향후 연구에서는 최적 집단의 개수를 찾는 알고리즘을 반영한 모형을 개발할 필요가 있을 것이다. 셋째, 본 연구에서는 변수별 가중치와 각 중심의 변수값을 추정하는데 있어 유전 알고리즘을 이용하였는데, 유전 알고리즘은 그 특성상 최적해의 추정에 많은 시간을 필요로 하는 한계점을 가지고 있다. 따라서 향후 연구에서는 본 연구모형의 최적해 추정을 위한 대안적 방법의 고안이 필요할 것으로 보인다.

참 고 문 헌

- [1] 송수섭, 이용훈, “인공신경망을 이용한 소비자 선택 예측에 관한 연구”, 『한국경영과학회지』, 제26권, 제4호(2001), pp.55-70.
- [2] 안현철, 김경재, 한인구, “Support Vector Machine을 이용한 고객구매예측모형”, 『한국지능정보시스템학회논문지』, 제11권, 제3호(2005), pp. 69-81.
- [3] 한상만, 박승배, 정남호, “인공신경망과 로짓모형을 이용한 내구재의 구매의도 예측에 관한 비교연구-대형 TV 시장을 중심으로”, 『마케팅 연구』, 제15권, 제3호(2000), pp.71-92.
- [4] Agrawal, D., C. Schorling, “Market share forecasting : An empirical comparison of artificial neural networks and multinomial logit model,” *Journal of Retailing*, Vol.72, No.4(1996), pp.383-407.
- [5] Ahn, H., K. Kim, I. Han, “Hybrid genetic algorithms and case-based reasoning systems for customer classification,” *Expert Systems*, Vol.23, No.3(2006), pp.127-144.
- [6] Ahn, H., K. Kim, I. Han, “A case-based reasoning system with the two-dimensional reduction technique for customer classification,” *Expert Systems with Applications*, Vol.32, No.4(2007), pp.1011-1019.
- [7] Allenby, G.M., T.S. Shively, S. Yang, M.J. Garratt, “A choice model for packaged goods : Dealing with discrete quantities and quantity discounts source,” *Marketing Science*, Vol.23, No.1(2004), pp.95-108.
- [8] Athanassopoulos, A.D., S.P. Curram, “A Comparison of data envelopment analysis and artificial neural networks as tools for assessing the efficiency of decision making units,” *Journal of the Operational Research Society*, Vol.47, No.8(1996), pp.1000-1016.
- [9] Ben-Akiva, M., B. Boccara, “Discrete choice models with latent choice sets,” *International Journal of Research in Marketing*, Vol.12, No.1(1995), pp.9-24.
- [10] Bolduc, D., “A practical technique to estimate multinomial probit models in transportation,” *Transportation Research B*, Vol.33, No.1(1999), pp.63-79.
- [11] Bolduc, D., B. Fortin, S. Gordon, “Multinomial probit estimation of spatially interdependent choices : An empirical comparison of two new techniques,” *International Regional Science Review*, Vol.20, No.1-2(1997), pp.77-101.
- [12] Borsch-Supan, A., V. Hajivassiliou, “Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models,” *Journal of Econometrics*, Vol.58, No.3(1993), pp.347-368
- [13] Brownstone, D., D.S. Bunch, K. Train, “Joint mixed logit models of stated and revealed preferences for alternative fuel vehicles,” *Transportation Research B*, Vol.34, No.5(2000), pp.315-338.
- [14] Chiu, C., “A case-based customer classification approach for direct marketing,” *Expert Systems with Applications*, Vol.22, No.2 (2002), pp.163-168.
- [15] Curram, S., J. Mingers, “Neural networks, de-

- cision tree induction and discriminant analysis : an empirical comparison," *Journal of the Operational Research Society*, Vol.45, No.4 (1994), pp.440-450.
- [16] Gilbride, T.J., G.M. Allenby, "A choice model with conjunctive, disjunctive, and compensatory screening rules," *Marketing Science*, Vol. 23, No.3(2004), pp.391-406.
- [17] Guadagni, P.M., J.D.C. Little, "A logit model of brand choice calibrated on scanner data," *Marketing Science*, Vol.2, No.3(1983), pp.203-238.
- [18] Ha, K., S. Cho, D. MacLachlan, "Response models based on bagging neural networks," *Journal of Interactive Marketing*, Vol.19, No.1 (2005), pp.17-30.
- [19] Holland, J.H., *Adaptation in natural and artificial systems*, The University of Michigan Press, Ann Arbor, MI, 1975.
- [20] Keane, M., "A computationally practical simulation estimator for panel data," *Econometrica*, Vol.62, No.1(1994), pp.95-116.
- [21] Kecman, V., *Learning and Soft Computing*, MIT Press, Cambridge, Mass, 2001.
- [22] Kumar, A., V.R. Rao, H. Soni, "An empirical comparison of neural network and logistic regression models," *Marketing Letters*, Vol.6, No.4(1995), pp.251-263.
- [23] Lachaab, M., A. Ansari, K. Jedidi, A. Trabelsi, "Modeling preference evolution in discrete choice models : A Bayesian state-space approach," *Quantitative Marketing and Economics*, Vol.4, No.1(2006), pp.57-81.
- [24] McCulloch, R., N. Polson, P.E. Rossi, "A Bayesian analysis of the multinomial probit model with fully identified parameters," *Journal of Econometrics*, Vol.99, No.1(2000), pp.173-193.
- [25] McFadden, D., "A method of simulated moments for the estimation of discrete response models without numerical integration," *Econometrica*, Vol.57, No.5(1989), pp.995-1026.
- [26] McFadden, D., K. Train, "Mixed MNL models for discrete response," *Journal of Applied Econometrics*, Vol.15, No.5(2000), pp.447-470.
- [27] Paap, R., P.H. Franses, "A dynamic multinomial probit model for brand choice with different long-run and short-run effects of marketing-mix variables," *Journal of Applied Econometrics*, Vol.15, No.6, Special Issue : Inference and Decision Making(2000), pp. 717-744.
- [28] Rossi, P.E., G.M. Allenby, "Bayesian statistics and marketing," *Marketing Science*, Vol.22, No.3(2003), pp.304-328.
- [29] Venugopal, V., W. Baets, "Neural networks and statistical techniques in marketing research : a conceptual comparison," *Marketing Intelligence and Planning*, Vol.12, No.7(1994), pp.30-38.
- [30] West, P.M., P.L. Brockett, 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.
- [31] van Wezel, M.C., W.R.J. Baets, "Predicting market responses with a neural network : the case of fast moving consumer goods," *Marketing Intelligence and Planning*, Vol.13, No.7(1995), pp.23-30.