

# 커뮤니티 사이트 특성과 navigation pattern 연관성의 세분시장별 이질성분석

- 믹스처모델의 구조방정식 적용을 중심으로 -

김소영\* · 곽영식\*\* · 남용식\*\*\*

## (요약)

기존의 인터넷소비자의 방문행동에 관한 연구들이 대부분 전체시장 수준에서 이루어졌고, 시장세분화를 하더라도 사전적 시장세분화로서 연구자의 주관이 반영되는 경우가 많다는 문제점이 있었다. 본 연구에서는 세분시장 수준에서의 인터넷소비자 방문행동 연구가 필요하다는 점과 함께 연구자의 임의성을 배제한 사후적 시장세분화의 필요성을 제기하고 이를 믹스처모델로 실증적으로 분석하였다. 또한 기존 연구가 웹사이트 평가요인과 방문행동 간의 단편적인 인과관계를 파악한 것과는 달리, 웹사이트 방문행동에 영향을 미치는 요인들 간의 간접효과에 따른 통합적인 인과관계를 파악하였다. 연구에서는 실제 커뮤니티 사이트를 방문한 1,765명의 웹상의 움직임과 설문조사를 통해 세분시장별로 이질적인 navigation pattern과 원인변수와의 관계가 존재함을 입증하였다. 마케터는 이런 세분시장별 이질적 관계를 이용하여 eCRM을 실행함에 있어 navigation pattern의 이질성에 영향을 미치는 변수를 세분시장별로 조절함으로써 기업이 원하는 방향으로 고객의 웹사이트 방문행위를 유도할 수 있는 기회를 얻게 되었다.

주제어 : eCRM, mixture model, 사후적 시장세분화, 웹사이트 평가요인, navigation pattern

## I. 서론

소수의 우량고객이 기업 전체 이익의 상당부분에 영향을 미친다는 20:80규칙 측면에서 보면, 기존고객의 가치 분석을 통해 우량고객이 누구인지를 파악하고 관리하여 지속적으로 높은 충성도를 유지시키고, 이탈가능성 있는 고객을 알아내어 적절한 마케팅 활동에 의해 그 가능성을 낮추는 활동은 매우 중요하다. 이를 위해

온라인과 오프라인에서 고객가치에 의한 세분화를 실시해 오고 있다. 일반적으로 고객가치는 기업입장에서 개별고객이 갖는 가치를 의미하며 고객의 기업 이익 기여도라고 할 수 있다.

상거래사이트의 경우에는 구매관련 자료(구매액, 구매량, 구매빈도 등)를 토대로 고객가치에 의한 세분화를 할 수 있지만, 포털사이트(portal site)나 커뮤니티 사이트의 경우에는 방문빈도와 사이트 밀착도(stickiness)가 기업의 수익성을 설명하는 중요한 요인이 된다(Bucklin and Sismeiro

\* 한국문화관광정책연구원 문화산업정책연구실 연구실장(sykim@kctpi.re.kr)

\*\* 진주산업대학교 벤처경영학과 전임강사(yskwak@jinju.ac.kr)

\*\*\* 삼성오프타이드 차이나 마케팅담당 이사(namys118@opentide.com.cn)

2000). 최근에는 사이트 밀착도나 방문빈도와 같은 navigation pattern의 이질성의 성격규명을 위한 시장세분화가 활발히 이루어지고 있기도 하다(Moe, 2001, Han and Park, 2003). 더 나아가 사이트 밀착도와 방문빈도에 영향을 미치는 변수들을 규명하는 연구도 이루어지고 있는데 (Bucklin and Sismeyro, 2000; Moe and Fader, 2000; Lee et al. 2000; Spiller and Lohse, 1998; Mandel and Johnson, 1999; 김소영·한상만·박세환, 2002), 이 중 김소영 외(2002)는 사이트 밀착도와 방문빈도에 영향을 미치는 변수로 해당 사이트의 구조적 요인, 정보적 요인, 오락적 요인과 플로우(flow)를 실증분석하여 확인한 바 있다.

그러나 이 같은 navigation pattern의 유형화와 이질성 연구 및 navigation pattern에 영향을 미치는 변수를 확인하는 연구들은 몇 가지 한계점을 갖고 있다. 첫째, eCRM이라는 이름으로 진행된 navigation pattern 유형화(시장세분화)에서는 연구자가 사전에 선정한 변수를 기준으로 세분화를 실시하거나 군집분석(cluster analysis)을 주로 사용하고 있다. 시장세분화 변수의 사전 결정의 경우에는 연구자의 주관적 판단이 개입된다는 단점이 있고, 군집분석은 척도와 관계없이 거리(distance)를 이용한 측정이 이루어지며, 동시에 최적 세분시장수를 제공하지 못한다는 문제점이 지적되었다. 이러한 문제점들을 극복하기 위한 대안적 시장세분화 방법으로 믹스처모델(mixture model) 방법론이 꾸준히 제기되어 왔으나 (Wedel and Kamakura, 2000), 이를 사용한 연구는 드문 실정이다. 이 같은 이유로

navigation pattern의 이질성을 충실히 측정하는 연구는 드문 실정이다. 둘째, navigation pattern의 이질성이 확인되었다 하더라도 이에 영향을 미치는 원인변수와 연관된 상태에서 이질성을 분석하는 연구도 거의 없는 상황이다. 즉, 이질성의 결과만을 보여주고 이질성 원인변수의 이질적 영향을 파악하는 연구는 거의 없다.

따라서 본 연구에서는 위에서 지적한 기존 연구들의 한계점을 극복하기 위해, navigation pattern의 이질성을 반영하기 위한 새로운 시장세분화 방법론인 믹스처 모델을 사용하여 네티즌의 이질적 사이트 방문행위를 시장세분화하고, 이 시장이질성의 원인변수라고 규명된 변수들의 이질적 영향을 파악하고자 한다.

이를 통해 각 세분시장별로 이질적인 navigation pattern과 원인변수와의 관계가 존재함을 실증하고자 한다. 이러한 일련의 과정은 마케터가 eCRM을 실행함에 있어 navigation pattern의 이질성에 영향을 미치는 변수를 세분시장별로 조절함으로써 회사가 원하는 방향으로 고객의 사이트 방문행위를 유도할 수 있을 것이며, 학술적으로는 navigation pattern의 이질성 분석에 사후적 시장세분화방법인 믹스처모델이라는 새로운 방법론을 적용시켰다는 공헌점을 갖게 된다.

## II. 이론적 배경

### 1. navigation pattern의 평가요인

웹에서 사이트의 효율성을 평가하는데 자주 제시되는 두 가지 기준은 웹사이트의 “접근성(reach)”과 “밀착도(stickiness)”이라 할 수 있다(Hanson 2000). “접근성”은 어떤 사이트가 유인하는 특정 방문자의 수를 의미하고 “밀착도”은 어떤 사이트가 의도한 목적을 이행할 만큼 오랫동안 그 방문자들을 보유하는 사이트 효율성을 의미한다(Bucklin and Sismeiro 2000). 판매가 이루어지려면 방문자들이 사이트에 오랫동안 머무르거나 자주 방문하여야 할 것이다. 또한, 배너광고를 판매하는 커뮤니티사이트나 검색포털 사이트 등 관련정보를 온라인상에서 제공하는 사업자들에게 있어서도 사이트 밀착도는 사이트 노출을 향상시키는데 매우 중요한 요소이다.

최근의 연구들은 고객분류 및 방문행동 특성에 대한 기존의 연구를 확장하여 고객집단별 방문행동 특성과 기업의 수익성의 관계에 관한 연구 등으로 확장을 시도하고 있다. Bucklin and Sismeiro(2000)는 특정 웹사이트의 체류시간(duration)과 페이지뷰(page viewed)가 웹사이트 밀착도를 구성하는 것이며, 웹사이트 밀착도에 영향을 미치는 요인들을 파악하기 위해 웹로그분석(web log)을 실시하였다. 또한 Moe and Fader(2000)는 인터넷쇼핑몰업체를 대상으로 고객의 방문행동을 재방문 소요시간(inter-visit time) 관점에서 접근하여 재방문 소요시간은 시간이 경과함에 따라 고객별로 서로 다른 유형이 존재한다고 주장하면서, 이러한 동적인 방문패턴의 변화가 제품구매와 밀접한 관계를 가지고 있다고 주장하였다. 또한 Lee et al. (2000)은 일정기간 동안 특정 웹사이트 방

문빈도가 높은 고객이 대상 사이트의 충성도가 높은 고객임을 제시하고 이에 영향을 미치는 요인들을 탐색하였다.

웹 사이트 특성과 관련된 기존의 연구들은 홈페이지에 나타난 온라인 쇼핑물의 규모, 제공서비스, 디자인 요소와 정보내용이 소비자 의사결정에 영향을 미치며 (Spiller and Lohse 1998), 배경 그림과 배경색 등 배경자극이 소비자의 정보탐색과정과 선택에 영향을 미치고(Mandel and Johnson 1999), 적당한 콘텐츠, 콘텐츠의 양, 콘텐츠의 순서 등 콘텐츠의 타겟팅이 사이트 이용의 증가, 충성도 상승 및 높은 이익을 유도할 수 있다고 제시하였다 (Ansari and Mela 2000). Jarvenpaa and Todd(1997)는 인터넷 쇼핑에 영향을 미치는 요소를 연구하기 위해서 전통적인 소매점의 소비자 행동에 영향을 미치는 변수를 추출하여 연구 모형을 구성하고 이를 인터넷 쇼핑과 비교하기도 하였다. 또한 Alba et al.(1997)은 인터넷 쇼핑의 장단점을 제시하고 인터넷 쇼핑에 영향을 미치는 중요한 속성들로 제품정보의 신뢰성, 폭넓은 상품, 상품의 효율적인 스크리닝, 점포간/브랜드간 탐색가능성, 과거의 구매자료 보유여부 등을 제시하였다. Chen and Wells(1999)는 웹사이트에 대한 태도가 웹사이트 가치의 유용한 지표가 된다고 주장하면서, 웹사이트의 태도를 측정하는 척도를 개발했는데, 웹사이트의 오락적(entertainment) 측면, 정보적(informative) 측면, 구조적(organization) 측면의 세 가지 차원이 웹사이트에 대한 전반적인 태도를 설명하는 주요 척도라는 사실을 제시하였다.

한편, 웹상에서의 소비자행동을 설명하는 중요변수로 플로우(flow)라는 개념이 Hoffman and Novak(1996)에 소개된 이후, 이에 관한 연구도 지속적으로 진행되었다(Novak and Hoffman, 1997; Zinkhan, 1998). 플로우란 소비자가 인터넷을 사용하면서 주관적인 최적 경험을 하게 될 때 얻어질 수 있는 것으로 Novak and Hoffman(1997)은 전자/가전제품을 구매하는 소비자를 대상으로 플로우의 선행요인이 탐색행동과 구매행동에 영향을 준다고 한다. Novak, Hoffman and Yung(1998)은 웹에서의 플로우의 선행변인과 결과변인들에 대한 공변량 구조분석을 수행하여 포괄적인 모형을 제시하였다.

지금까지의 연구결과들을 종합해 보면 웹사이트의 오락적 요인, 정보적 요인, 구조적 요인 및 플로우가 웹사이트 소비자의 navigation pattern에 영향을 미치는 주요 요인이라고 정리할 수 있다. 김소영 외(2002)는 이 같은 사실을 토대로 이 네 가지 요인이 웹사이트 소비자의 체류시간, 방문빈도, 페이지뷰에 어떻게 영향을 미치는지를 실증적으로 분석하였다. 그러나 이 연구의 분석단위는 시장전체(표본전체)였기 때문에, 표본 내 세분시장의 이질성에 대한 반영이 이루어지지 않았다.

이와 관련하여 최근의 시장세분화에 대한 연구들은 인과관계에 있어서도 이질적인 시장세분화가 필요함을 지적하고 있다(McLachlan and Basford, 1988; McLachlan and Peel, 2000). 즉, 독립변수와 종속변수로 인과관계를 살펴볼 경우, 표본은 몇 개의 이질적인 세분시장의 인과관계의 합이라고 가정하는 것이다. 예를 들

어, 광고라는 독립변수와 매출이라는 종속변수간의 인과관계를 기존 회귀분석으로 분석하여 양의 상관관계로 0.2가 나왔다고 가정할 때, 이 데이터가 0.8의 양의 상관관계를 가진 60%와 -0.3의 음의 상관관계를 가진 40%로 구성되어 있을 가능성을 배제할 수 없다. 따라서 이 같이 숨어있는 이질적 인과관계가 navigation pattern과 평가요인 사이에 존재하고 있음을 밝힐 필요가 있는 것이다.

## 2. 시장세분화와 믹스처모델

시장세분화는 기존에 너무나 많은 연구들이 이루어졌기 때문에, 이를 일일이 나열하기도 어려울 정도다. 따라서 본 연구에서는 Wedel and Kamakura(2000)가 사용한 시장세분화 방법의 통계기법 성격분류를 이용하여 각 방법의 장단점을 살핀 후, 단점을 극복할 수 있는 새로운 시장세분화 방법으로써 믹스처모델에 대해서 살펴본다. <표 1>은 Wedel and Kamakura(2000)가 1) 연구자가 시장세분화변수를 집단 분류 이전인 사전에 결정했느냐, 아니면 집단 분류 후 집단성격을 사후에 밝혀냈느냐 하는 것과 2) 시장세분화 변수간의 인과관계가 있느냐 없느냐 하는 기준으로 시장세분화 방법을 분류한 것이다. 이 분류표의 각 방법론의 장단점을 비교함으로써 마케터가 시장세분화 작업 및 CRM(eCRM) 등에 사용할 수 있는 최적의 방법론을 찾을 수 있는 기초를 마련할 수 있다.

<표 1> 시장세분화를 방법론상으로 분류한 경우(Wedel and Kamakura, 2000)

구 분	마케터가 사전에 시장세분화기준을 정한 경우(a priori)	세분화후 시장세분화 기준을 확인한 경우(post hoc)
변수간의 인과관계가 없는 경우 (descriptive)	상황표(카이제곱분석)	군집분석(nonoverlapping, overlapping, fuzzy clustering) 믹스처모델
변수간의 인과관계가 있는 경우(predictive)	회귀분석, 로짓분석, 판별분석	Fuzzy clusterwise regression mixture regression model

제 1국면은 마케터가 사전에 시장세분화 기준을 정한 경우로 변수간의 인과관계가 없는 경우 (a priori and descriptive) 로써 카이제곱 분석에 의한 상황표의 작성이 대표적인 방법이다. 이 방법은 마케터가 시장을 세분화하기 위해 변수를 선정한 다음, 그 변수의 상하/고저 등을 시장세분화 기준으로 제시한 후, 시장세분화를 하는 경우를 말한다. 이 방법의 결정적인 단점은 마케터의 주관적 판단에 의해 시장이 사전에 분류된다는 점이다. 만약 마케터가 판단한 시장세분화 기준이 고객의 행동에 중요한 영향을 주는 변수가 아니거나, 세분화 기준의 상하/고저 적용 의미 없는 분류치가 사용되는 경우에는 시장세분화 효과를 기대하기 어렵다. 이 방법에 해당하는 연구로는 하나하나 열거할 수 없을 정도로 많은 연구결과가 있다 (Green, 1977; Shapiro and Bonoma, 1984; Bass, Tigert와 Lonsdale, 1968; Frank, 1972; Cuba, 1977; Rao and Winter, 1978; 이용기 외, 2000). 인터넷상의 고객관리를 위해 시장세분화를 실시한 연구 중 김소영 외(2002)의 STIVI분석이 여기에 해당한다.

제 2국면은 마케터가 사전에 시장세분화 기준을 정한 경우로 변수간의 인과관

계가 있는 경우(a priori and predictive)로써 판별분석, 로짓분석, 회귀분석 등이 대표적인 방법이다. 먼저 판별분석을 살펴보면, 판별분석은 사전에 분류된(classified) 표본들을 종속변수로 하고, 이 분류된 집단을 가장 잘 설명하는 변수를 찾아가는 과정이다. 이 방법은 시장세분화 기준으로 독립변수와 종속변수를 사전에 지정한다는 점에서 마케터의 주관적 판단에 의해 시장이 분류된다는 단점에서 벗어날 수 없다. 한편 로짓분석(logit analysis)에서도 종속변수가 사전에 결정된다는 점에서는 판별분석과 차이가 없다.

제 3국면은 마케터가 시장세분화 이후에 세분시장별 성격을 가장 잘 설명하는 변수를 찾을 수 있는 경우로, 군집분석과 믹스처모델이 대표적인 방법이다. 이 국면은 사전에 마케터의 주관적인 의견이 반영되지 않는다는 점에서 시장세분화방법으로 우월한 방법이다. 마케팅 반응특성에 따른 이 사후적 기준에 따른 시장세분화는 소비자의 마케팅반응을 시장세분화 과정에서 직접 반영하고 있다는 점에서 마케팅전략 수립에 사전기준에 의한 시장세분화방법 보다 더 유용한 시장세분화 방법이라고 지적되어 왔다(안광호과 임병훈, 1995). 인터넷상의 고객 navigation

pattern 연구 중 Moe (2001)이나 Han and Park (2003)의 연구가 여기에 해당한다.

그런데 군집분석의 경우, 우선 어떤 군집분석 알고리즘을 사용하느냐에 따라 시장세분화 결과가 달라진다는 단점이 있다. 동일한 표본에 의한 데이터일지라도 계층화된 군집분석(단일기준결합방식, 완전기준결합방식, 평균기준결합방식, Ward의 오차제곱합방식), 비계층적 군집분석(k-mean 분석)의 채택여부에 따라 군집분석의 결과는 모두 달라진다. 이렇게 제각기 다른 시장세분화 결과가 진실로 동질적인 집단을 이질적으로 구분했다고 말하기 어렵다.

두 번째 문제점은 군집분석이 채택하고 있는 분류방법인 non-overlapping 군집방법이 인간행동의 다양성을 반영하지 못하고 있다는 점이다. non-overlapping 방법으로 군집이 이루어진 경우, 특정 관찰치(고객)는 오직 한 집단에만 속하는 것으로 나타나고, 그 관찰치는 그 집단의 성격만을 반영하는 것으로 판단된다. 반면에 over-lapping 방법은 한 관찰치가 두 개 이상의 세분시장에서도 발견될 수 있는 방법이지만 어느 집단에 더 나타날지는 알 수 없다. 반면에 fuzzy cluster 방법은 한 관찰치가 두 개 이상의 세분시장에서 발견될 수 있을 뿐만 아니라, 특정 세분시장에 속한 정도를 파악할 수 있어서 가장 선호되는 방법이다. 결국 어떤 한 개인의 행동의 다면성을 non-overlapping 방법보다는 overlapping 방법이나 fuzzy clustering 방법이 더 잘 반영하고 있다.

세 번째 문제점은 군집분석이 ward법 알고리즘의 일부를 제외하고는 통계적으

로 적절한 세분집단의 수를 제시하지 못한다는 결정적인 단점을 갖고 있다는 데 있다(Lehmann, Gupta and Steckel, 1998).

넷째 문제는 군집분석에 사용할 수 있는 척도는 등간척도 이상으로 명목척도 변수를 포함하는 시장세분화가 제한된다는 점이다. 군집분석에서는 유사성(similarity)과 거리(distance)를 이용하여 비슷한 성격을 가진 관찰치를 묶어 준다. 군집방법에서는 사용하는 변수가 명목척도인 경우에는 binary data만이 사용될 수 있으며, 3개 이상의 명목척도(categorical data)는 사용할 수 없다. 왜냐하면 비계층적이건 계층적이건 군집분석에 사용하는 기본 조건은 거리(Euclidian distance 등)를 측정하는 것인데, 명목척도가 3개 이상이면 거리를 측정할 수 없기 때문이다(Lehmann et al., 1998). 가령 봄/여름/가을/겨울이 군집분석에 중요한 정보일 경우, 거리를 기초로 군집분석이 이루어지므로 봄과 가을간의 거리가 여름과 가을간의 거리와 차이가 있다고 말할 수 없다.

지금까지 논의한 것을 정리하면, 바람직한 시장세분화 방법이 가져야 할 기준을 다음과 같이 정리할 수 있다. 첫째, 인간행동 다양성을 반영하기 위해 non-overlapping 방법보다 fuzzy clustering 방법이 사용되는 알고리즘으로 시장세분화가 이루어지는 것이 바람직하다. 둘째, 마케터가 사전에 시장세분화 변수를 지정하는 것보다, 고객성격에 따라 분류가 완성된 다음, 세분시장별로 성격이 변수별로 사후적으로 정의되는 것이 바람직하다. 셋째, 적정 수의 세분시장 수를 통계적으로 유의하게 확인해 줄 수 있는 것이

바람직하다. 넷째, 시장세분화에 사용되는 변수의 척도에 구애받지 않아야 한다. 이 같은 기준들을 만족시킬 수 있는 시장세분화 분석방법이 바로 믹스처모델이라 할 수 있다.

### 3. 믹스처모델

이 모델은 육안으로 보이는 한 변수의 분포는 우리가 인지하지 못하는 몇 개의 하위분포로 구성되어 있다는 점에서 출발한다(McLachlan과 Basford, 1988). 한 표본이  $n$ 개의 data로 구성되어 있고, 각 데이터는  $k$ 개의 변수로 되어 있다고 하면(식 1),

$$y_n = (y_{nk}) \quad (1)$$

$y_n$ 값은  $s$ 개의 세분시장에서 시장크기가 모두 다르고  $\pi$ 로 표시된  $\pi_1, \dots, \pi_s$ 의 비율로 이루어진 모집단에서 측정된 관찰치의 값이다. 이 시장크기의 합은 1이다(식 2).

$$\sum_{s=1}^s \pi_s = 1 \quad (2)$$

그러면  $y_{nk}$ 는 벡터  $y_n$ 의 조건부분포함수로써 세분시장  $s$ 에서 계산된 값이 된다.  $y_n$ 은 우리가 모르는 모수(parameter)를 가진 특정 밀도함수의 벡터로 표시된  $\theta_s$ 로써  $f_s(y_n/\theta_s)$ 의 일반적인 형태를 가진다. 이 확률밀도함수는 정규분포, 포아송, 이항분포, binomial, negative binomial 등의 각종 형태를 모두 지닐 수 있다. 따라서 사용하는 변수의 척도와 관계없이 시장세분화를 할 수 있다는 특징이 있다.

이 특징은 기존 군집분석이 사용할 수 있는 척도가 likert 척도 이상이어야 된다는 한계점을 극복한 것이다. 이런 측면에서 믹스처모델은 척도와 관계없이 세분시장별로 그 분포에 상관없이 평균과 분산을 가지며, 한 관찰치  $i$ 가 특정 세분시장  $s$ 에서 발견된 확률은 특정세분시장  $s$ 의 변수들의 조건( $\theta$ ) 아래에서 관찰치  $y_n$ 이 관찰될 가능성으로 표시된다(식 3).

$$p_{.i} = \frac{\pi_s \cdot f_s(y_i, \theta_s)}{\sum_s \pi_s \cdot f_s(y_i, \theta_s)} \quad (3)$$

즉, 기존 시장세분화방법이 한 변수의 분포를 동질적(homogeneous)으로 가정하고 이 변수의 상하/고저로 시장세분화를 실시한다는 시각을 가졌다면, 믹스처모델은 별도의 평균과 분산을 가진 독특한 하위분포들의 합(heterogeneous)으로 이루어져 있다는 시장세분화 시각을 가졌다. 또한 식(3)의  $\pi$ 가 일정할 때 특정세분시장에서의 조건인 식(2)의  $\theta_s$ 는 바로 특정시장 크기의 하위시장의  $k$ 개 변수의 성격을 반영하므로, 집단을 사전에 연구자의 주관적 판단으로 묶은 것보다는 집단이 묶여진 후에 집단의 성격을 확인해야 한다는 분석 사후적 세분시장별 평가의 성격을 가지고 있다. 위의 식(3)을 보면 한 관찰치가 반드시 한 세분시장에 속한다는 기존 군집분석의 단순한 가정에서 벗어나(non-overlapping clustering), 한 관찰치가 여러 개의 세분시장에 들어갈 확률로 표시되므로(fuzzy clustering) 인간의 다양한 측면을 반영하게 된다는 장점이 있다.

믹스처모델은 또한 최적의 세분시장 수

를 통계적으로 밝히는 데 유리하다. 원래 분포 아래 그 하위분포는 몇 개인지는 모르지만 세분시장 수를 늘릴 때마다 설명력은 높아질 것이라 가정하고, 세분시장 수를 늘렸지만 설명력이 유의적으로 증가하지 않거나, 세분시장 수를 늘리는데 사용한 모수에 비해 설명력 증가가 적어서 모델의 parsimony를 위해 더 이상 세분화 작업을 할 필요가 없을 때 최적 하위시장 수를 판단할 수 있게 된다(한상만, 곽영식, 1997). 이를 측정하기 위한 모델 적합도(goodness-of-fit)는 adjusted-p2과 AIC(Akaike Information Criterion)로 측정한다. Adjusted-p2의 적합도 측정치는 회귀분석의 R2와 비슷한 해석이 가능하다(Ben-Akiva and Lerman, 1985). 이 모델의 계수는 log-likelihood값을 기초로 EM(expectation-maximization)알고리즘에 의해 추정된다.

$$\text{adjusted-}p^2 = (\text{특정세분시장의 log-likelihood값} - \text{모수(parameter)수}(p)) / \text{귀무모형의 LL}$$

$$AIC = -2(LL - p(\text{모수 수})) / \text{관찰치}(N)$$

$$L(\phi; y) = \prod_{i=1}^N \left\{ \sum_{j=1}^K \pi_j f(y_i | \theta_j) \right\} \quad (4)$$

여기에서 LL은 (식 4)의 L(f;y)로 표시된 값으로 특정세분시장에서의 로그우도(log-likelihood)값이다. 이 값은 각 모델이 특정세분시장 수에서 얼마만큼의 설명력을 가지는 지로 환원할 수 있는 값이다. 특정 세분시장별, 세분시장크기별 LL값을 계산한 후 adjusted-p2과 AIC(Akaike Information Criterion)를 계산하게 된다. 이 값의 변화를 추적하여 최적의 세분시장 수를 확인할 수 있는 통계적 유의성을 찾을 수 있게 된다.

<표 2>에서 정리된 바와 같이 믹스처 모델은 군집분석의 한계점을 극복한 새로운 시장세분화 방법론으로 부각되고 있다. 이를 navigation pattern에 사용할 경우, 사후적 시장세분화를 통한 이질성의 영역 확대에 좋은 방법론이 될 것으로 판단된다. 특히 navigation pattern에 사용된 변수들간의 인과관계가 존재할 경우에는 믹스처모델을 회귀분석에 적용하여, 여러 개의 숨어있는 인과관계를 가진 선형관계를 찾는 꼴로 모델을 변경할 수 있게 된다.

하지만 믹스처모델의 계수 추정시 문제점이 있는데 이것이 바로 추정된 모수가 global maximum이 아니라 local maximum일 가능성을 배제할 수 없다는 점이다. 만약에 하위분포의 밀도함수가 제대로 분리되지 않을 경우, 개별관찰치의 사후확률을 구하기 위해 계수추정을 시작하

<표 2> 군집분석과 믹스처모델의 시장 세분화 적용 비교

구분	군집분석	믹스처모델
군집 알고리즘	거리, non-overlapping clustering	조건부확률, fuzzy clustering
최적시장세분 수	통계적 유의성 설명못함	통계유의성 설명가능
척도(scale)	2단계명목,서열,등간, 비율	명목,서열,등간,비율
세분시장별 성격규명	서열이상의 척도로 군집분석 후 명목척도의 빈도분석 실시 2단계	척도에 구애 받지 않고 군집분석을 실시하는 1단계



는 값이 얼마냐에 따라 사후적으로 도달하는 계수추정치가 불안정하게 된다. 따라서 만약 global maximum 앞에 조그마한 local maximum이 존재하면, 믹스처모델의 EM알고리즘은 그 local maximum에서 iteration을 멈추고 계수값을 산출하게 된다. 따라서 만약 local maximum에서 산출된 최적 시장세분 수는 최적이지 아닐 가능성이 있으므로 주의가 요망된다.

#### 4. 문헌연구를 통한 본 연구의 방향

우리는 문헌고찰을 통해 navigation pattern 연구의 이질성 확보를 위한 새로운 대안인 믹스처 모델이 아직까지 인터넷상의 이질적인 고객행동연구에는 사용되지 않았음을 알 수 있었다. 따라서 본 연구에서는 인터넷 고객의 navigation pattern의 이질성을 반영하기 위한 새로운 시장세분화 방법인 믹스처 모델을 사용하여 인터넷 고객의 이질적인 사이트 방문행위를 시장세분화하고, 이 때 시장 이질성의 원인변수로 규명된 변수들을 포함시켜 시장세분화를 실시하고자 한다. 이를 통해 각 세분시장별로 이질적인 navigation pattern과 원인변수와의 관계가 존재함을 보여줄 수 있을 것으로 예측된다.

### III. 연구방법

#### 1. Data

본 연구는 국내의 기업 홍보용 커뮤니티 포털사이트를 대상으로 분석하고 있으며, 해당 사이트는 회원수가 약 20만 명에 상당한다. 실증분석에 필요한 변수들(예, 방문빈도, 페이지뷰, 체류시간)을 추출하기 위하여 우선, 해당 사이트의 서버에 기록된 두 달 동안의 웹 로그 데이터를 DB로 만들고 난 후 데이터 정제작업을 실시하였다. 고객 개개의 확인은 이용자 쿠키(cookie) 정보를 이용하였다. 본 연구에서 사용한 데이터는 크게 세 부분으로 구성되어 있다.

첫째 데이터는 고객 클릭스트림 데이터(Customer Click-stream Data)이다. 본 연구에서는 특정 웹사이트의 고객들이 두 달 동안 해당 사이트를 돌아다니면서 남긴 흔적을 나타내는 마우스 클릭 스트림(mouse click stream) 데이터를 추출하여 정제과정을 거쳐 분석에 사용하였다. 로그 파일 데이터는 고객의 흔적을 사이트 서버에 그대로 남겨주기에 웹사이트 방문특성과 관련하여 객관적 소비자행동 분석이 가능하다는 이점이 있다. 여기서는 로그파

<표 3> 사이트 방문행동 변수(종속변수)의 요약 및 기초 통계

변수	변수의 측정	평균
방문빈도	고객별 2달 동안 N사이트 총 방문빈도	124 번
체류시간	고객별 2달 동안 N 사이트에서 머문 총 시간	32.4 시간
평균체류시간	한번 방문 시 머문 평균 시간	0.2 시간(12분)
총페이지뷰	고객별 2달 동안 N사이트에서 요청한 총 페이지수	1460.7 페이지
평균페이지뷰	한번 방문 시 요청한 평균 페이지수	10.7 페이지

일 분석을 통해 고객이 해당사이트에서 얼마나 많은 페이지를 요청했는지를 나타내는 페이지뷰(page viewed), 얼마나 오랫동안 머물렀는지를 나타내는 체류시간(duration time) 및 얼마나 사이트를 자주 방문했는지의 방문빈도(visit frequency) 데이터를 추출하였다<표 3 참조>. 고객들은 평균 124번을 즉, 하루에 약 2번 정도 방문을 했고, 한번 방문해서 평균 12분 정도를 머무르며 10.7페이지를 보고 나가는 것으로 밝혀졌다.

두 번째 데이터는 고객 서베이 데이터(Customer Survey Data)이다. 고객의 웹사이트 특성에 대한 평가요인으로 문헌고찰 결과 정리된 정보적 차원, 오락적 차원, 구조적 차원 및 웹사이트 사용 시 플로우 경험 그리고 일반적인 인터넷 이용과 관련한 항목들을 측정하였다.

웹사이트 특성에 대한 평가 항목으로는 Chen and Wells(1999)의 연구에서 사용된 문항들을 중심으로 해서 측정항목을 만들었다. 플로우는 Novak, Hoffman and Yang(1998)이 정의한 개념을 토대로 측정하였다. 총 16개 문항으로 구성되어 있고

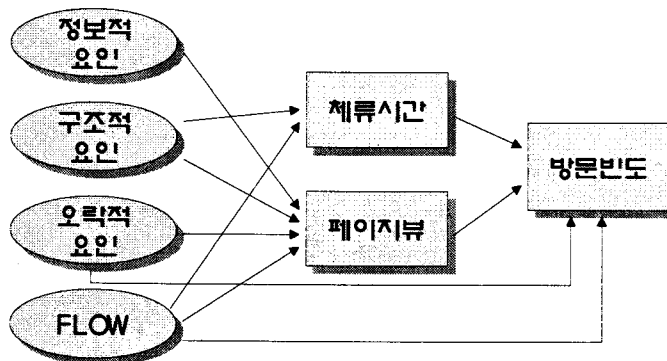
7점 척도로 되어있다. 연구자의 의도대로 4가지 차원으로 구성되는 개념인지가 이미 신뢰성, 타당성 분석을 통해 검토되었다(Cronbach- $\alpha$ 값이 0.852에서 0.884로 나타남). 온라인 서베이는 2000년 12월 4일부터 12월 8일까지 연구대상 사이트인 N 사이트에 HTML 문서로 제작된 설문지를 띄워 획득했으며, 총 1,765명의 응답자료를 얻었다.

세 번째는 고객 프로파일 데이터(Customer Profile data)이다. 고객 프로파일 데이터에는 고객의 인구통계적 특성(성별, 연령,직업, 교육수준, 등) 자료들이 포함되어 있다. 고객의 클릭스트림 데이터와 서베이 데이터와 이 인구통계학적 변수를 통합하였다.

## 2. 연구모델

본 연구에서는 웹사이트 평가요인과 navigation pattern의 관계 분석을 위해 아래 <그림 1>을 기본 연구모델로 활용하고 있다.

<그림 1> 웹사이트 평가요인과 navigation pattern의 관계(김소영 외, 2001)



이 모델은 김소영 외(2001)의 연구결과에서 요인과 변수 간에 유의한 관계가 도출된 것만을 종합한 것이다. 화살표로 연결되지 않은 것은 유의하지 않은 관계가 나타난 바 있다. 즉, 고객의 웹사이트에 대한 평가요인과 페이지뷰, 체류시간 및 방문빈도와 같은 웹사이트 방문행동 간의 인과관계를 전체고객집단 수준에서 분석된 결과이다. 김소영 외(2001)는 웹사이트 평가요인과 페이지뷰의 관계는 포아송 회귀모델(Poisson Regression Model), 웹사이트 평가요인과 체류시간의 관계는 지수회귀모델(Exponential Regression Model), 웹사이트 평가요인과 방문빈도는 음이항회귀모델(Negative Binomial Regression)을 토대로 하여 개별적인 모델을 사용하여 웹사이트 평가요인과 방문행동 간의 개별적인 인과관계를 분석한 바 있다. 하지만, 이들 관계를 통합하여 분석하지 않았다.

따라서 본 연구에서는 김소영 외(2001)에서 유의하게 도출된 웹사이트 평가요인과 navigation pattern의 인과관계를 통합하여, 기존 연구의 개별인과관계를 벗어나고자 한다. 또한 개별온라인 고객을 믹스처모델에 의해 시장세분화를 실시한다. 이를 통합모델에 적용하여 세분시장별 웹사이트 평가요인과 방문행동간의 인과관계를 전체 변수 간 간접효과를 종합적으로

로 고려하는 실증분석을 실시해보고자 한다. 즉, 김소영 외(2001)의 연구에서는 전체 고객집단(표본집단)을 대상으로 인과관계를 제시하였으나, 본 연구에서는 이 같은 관계가 세분시장별로 이질적일 것이라는 가정 하에, 믹스처모델에 의한 시장세분화와 구조방정식모형에 기초한 간접효과를 포함한 인과관계를 파악함으로써, 세분시장별 웹사이트 평가요인과 navigation pattern간의 인과관계를 실증적으로 분석하고 있는 것이다.

#### IV. 연구결과

##### 1. 믹스처모델에 의한 적정 세분시장 수의 확인

본 연구는 세분시장별 요인구조분석 및 요인간의 상대적 중요성의 차이를 확인하는 것이 목적이므로, 이 표본을 대상으로 믹스처모델에 의한 시장세분화를 시도하였다. 믹스처모델은 시장을 세분화할 때, 세분시장의 수를 늘려가면서 그때마다 AIC값을 계산할 수 있는데, 이 값의 감소폭이 현저히 적어지거나, 오히려 증가하면 더 이상 세분시장 수를 늘리더라도 설명력이 거의 증가하지 않거나 오히려 설명

<표 4> 세분시장 수를 증가에 따른 모델 설명력 증가

세분시장 수	Loglikelihood값	AIC값	추정해야 할 모수 수
1	-37251.243	74599.6	13
2	-35730.260	71617.5	21
3	-35031.216	70279.2	29
4	-34596.887	69470.3	37
5	-34533.170	69520.7	45

력이 떨어짐을 뜻한다(한상만과 곽영식, 1997; Bucklin and Gupta, 1992, Bucklin, Gupta and Han, 1995).

본 연구에서 사용된 모든 변수를 대상으로 믹스처모델을 적용한 결과, 위에서 보듯이 4개의 세분시장에서 5개의 세분시장으로 증가할 때, AIC값을 보면 설명력이 오히려 감소하였다(AIC값은 0으로 갈수록 설명력이 높다고 해석하는데, 이 연구에서는 4개 세분시장에서 5개로 세분시장이 커지자 AIC값이 0에서 멀어졌다). 그러므로 연구 대상 웹사이트의 전체고객에 대한 적정 세분시장의 수는 4개로 확인되었다. 각 세분시장별 크기를 보면 세분시장 1에는 전체표본의 80.9%가, 세분시장 2에는 7.9%, 세분시장 3에는 7.1%, 세분시장 4에는 전체표본의 3.9%가 포함되어 있다.

2. 세분시장별 연구모형의 적합도 및 계수

시장을 세분화하기 전에 전체시장(전체 고객)에 대한 연구모형의 적합도를 분석한 결과, adjusted GFI 값이 0.819로 나타났다. 이순목(1990)에 의하면, adjusted GFI 값이 0.9이상 이어야 모델의 적합도에 문제가 없다고 했기에, 시장을 세분화하지

않은 상태에서 전체시장을 대상으로 분석했을 때는 모델의 이론적 관계에 표본이 잘 부합한다고 할 수 없었다.

믹스처모델은 각 개별관찰치가 어느 집단에 속하는지를 사후에 확률적으로 산출할 수 있으므로(fuzzy clustering), 각 개별치를 세분시장별로 할당하였다. 이 4개의 세분시장에 해당하는 관찰치를 세분시장별로 <그림 1>의 모델에 적용하였다. 이 때 전체시장과 세분시장의 모델 적합도는 amos 패키지를 이용하였다.

믹스처모델에 의해 제시된 4개의 세분시장별 적합도를 분석한 결과, 전체시장을 대상으로 했을 때의 적합도보다 높을 뿐만 아니라, 모델의 이론적 구조를 확신할 수 있는 적합도 수치인 adjusted GFI 값이 4개 세분시장 모두에서 0.9이상을 보였다. 따라서 통합모형에 대해 자료가 잘 부합한다고 할 수 있었다.

<표 5> 세분시장별 적합도

구분	GFI	Adjusted GFI
전체시장	0.968	0.819
세분시장 1	0.987	0.927
세분시장 2	0.988	0.930
세분시장 3	0.987	0.928
세분시장 4	0.987	0.929

<표 6> 전체시장 및 세분시장별 웹사이트 평가요인이 navigation pattern에 미치는 영향

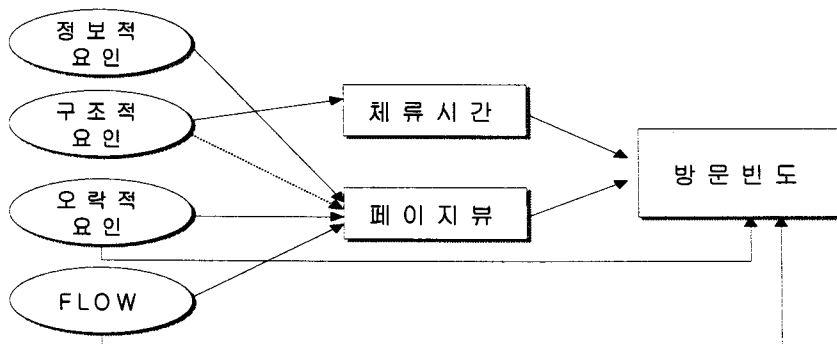
구분	전체시장	세분시장 1	세분시장 2	세분시장 3	세분시장 4
오락적-page	0.09***	0.07***	-0.07	0.06	0.12
구조적-page	0.03	0.04*	0.18**	-0.12	0.03
Flow-page	0.16***	0.10***	0.26***	0.06	0.33***
정보적-page	0.03*	0.08***	-0.06	0.09	0.16
구조적-duration	0.03*	0.07***	0.02	-0.05	-0.04
Flow-duration	0.07***	0.03	-0.04	0.145*	0.01
Duration-visit	0.81***	0.56***	-0.17**	0.775***	-0.04
Page-visit	-0.06***	-0.13***	-0.07	-0.16***	-0.06
Flow-visit	0.04***	0.03	0.12	0.03	0.12
오락적-visit	0.05***	0.03	0.11	0.00	0.13

\* p<0.1    \*\* p<0.05    \*\*\*p<0.01

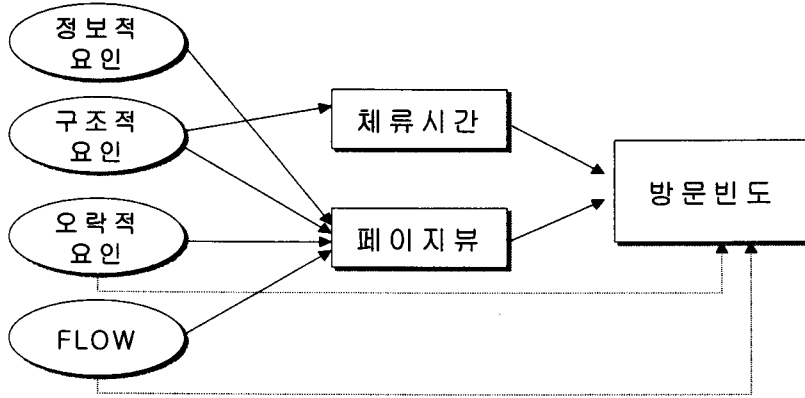
이 같은 결과를 토대로 각 세분시장별 경로 계수에 대한 세부적인 측정치를 분석함으로써 세분시장별 웹사이트 평가요인이 navigation pattern에 미치는 영향에 대한 파악과 함께 인과관계를 확인하였다<표 6 참조>. <표 6>에 제시된 구조방정식 분석결과를 전체시장 및 세분시장 수준에서 변수들 간의 인과

관계를 그림으로 나타내면 다음과 같다(<그림 2>, <그림 3>, <그림 4>, <그림 5>, <그림 6>참조).

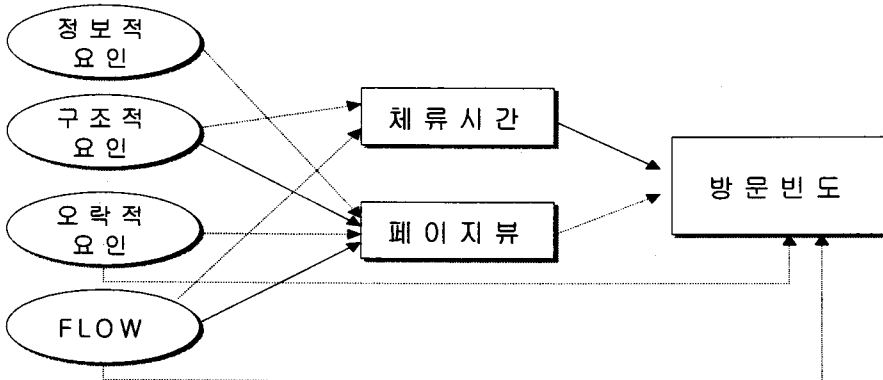
<그림2> 전체시장 수준에서의 웹사이트 평가요인과 navigation pattern 관계



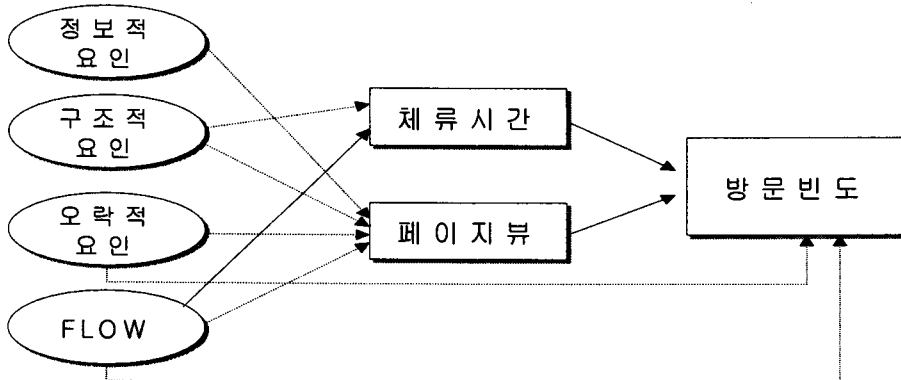
<그림3> 세분시장 1의 웹사이트 평가요인과 navigation pattern 관계



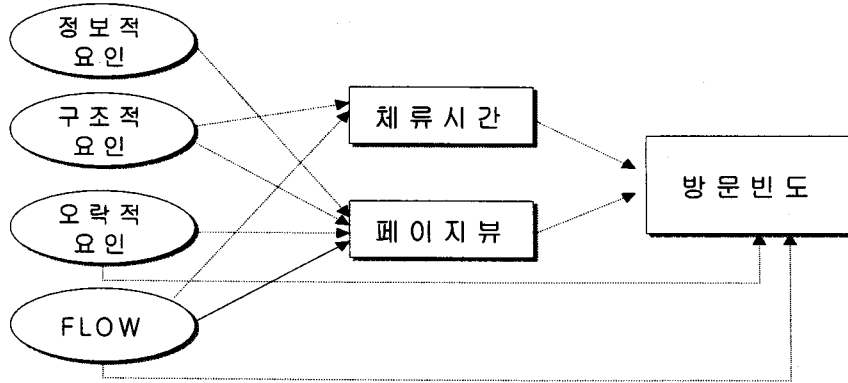
<그림4> 세분시장 2의 웹사이트 평가요인과 navigation pattern 관계



<그림5> 세분시장 3의 웹사이트 평가요인과 navigation pattern 관계



<그림6> 세분시장 4의 웹사이트 평가요인과 navigation pattern 관계



위의 결과들을 살펴보면, 세분시장별로 유의한 경로계수가 모두 다르게 나타나고 있는데, 이는 세분시장별로 웹사이트 방문행동이 이질적이라는 사실을 나타내는 것이다. 따라서 세분시장별로 마케터가 원하는 사이트 밀착도나 방문빈도를 달성하기 위해서는 세분시장별로 웹사이트의 제반요소 중 유익하게 제시된 웹사이트의 평가요인을 집중적으로 관리하고 개선하는 활동과 함께 세분시장별 상이한 마케팅커뮤니케이션 활동이 필요하다는 사실을 시사한다.

위의 결과를 변수 수준에서 해석하면 다음과 같다.

첫째, 웹사이트 평가요인이 페이지뷰에 미치는 영향을 살펴보면, 전체시장의 경우에는 웹의 구조적 요인이 페이지뷰에 미치는 영향이 없는 것으로 나타났다. 이는 김소영(2001)의 연구에서 웹사이트 평가요인과 방문행동 간에 개별적인 분석을 통해 제시된 결과와는 다른 것으로, 변수들 간 전체적인 상호관계를 살펴보기 위해서는 구조방정식 같은 통합적 모델분석을 통해 간접효과까지 측정할 필요성이 있음을 시사하는 결과라

볼 수 있다.

세분시장 수준에서 보면, 세분시장 1과 2는 웹평가의 구조적 요인과 페이지뷰 사이에 인과관계가 존재하는 것으로 나타났으며, 세분시장 3과 4는 인과관계가 없는 것으로 나타났다. 따라서 전체시장에서 웹사이트 평가요인과 페이지뷰 간의 인과관계가 나타나지 않은 것은 세분시장 3, 4의 결과에 의한 영향으로 해석할 수 있을 것이다.

또한 웹사이트 평가요인 중 오락적 요인과 정보적 요인이 페이지뷰에 유의적인 영향을 미친다는 전체시장의 결과는 세분시장 1에서만 적용 가능함을 알 수 있었다. 더 나아가 플로우가 페이지뷰에 미치는 영향은 세분시장 3에는 해당되지 않는다는 사실을 알 수 있었다.

둘째, 웹사이트 평가요인이 지속시간(duration)에 미치는 영향을 살펴보면, 전체시장의 경우에는 구조적 요인과 플로우가 지속시간에 인과관계가 있는 것으로 나타났는데, 이런 인과관계가 있는 세분시장은 세분시장 1에 국한된 결과로 나타났다. 다른 세분시장에 속한 고객들의 경우에는 지속시

간에 영향을 미치는 독립변수가 나타난 것은 세분시장 3의 플로우 뿐 이었다.

셋째, 지속시간이 방문빈도(visit)에 미치는 영향을 살펴보면, 전체시장을 포함한 세분시장 1, 2, 3 모두에서 유의성이 검증된 반면, 세분시장 4의 경우에는 지속시간이 방문빈도에 영향을 미치지 않는 것으로 나타났다. 또한 페이지뷰가 방문빈도에 영향을 미치는 것을 살펴본 결과, 전체시장 수준에서의 유의한 인과관계가 세분시장 1과 3에서는 나타났지만 세분시장 2와 4에서는 나타나지 않았다.

넷째, 웹사이트 평가요인이 방문빈도에 미치는 영향은 전체시장 수준에서만 나타났으며, 세분시장에서는 전혀 영향이 없는 것으로 나타났다.

이상의 결과를 통해 웹사이트 마케터는 세분시장별 고객들의 독특한 웹사이트 방문 행위에 근거하여 세분시장별로 차별적인 고객관리를 실시해야 할 필요성이 있다는 사실을 확인할 수 있었다. 가령, 세분시장 1은 홈페이지의 오락적/구조적/정보적 요인이 웹사이트 방문행동(페이지뷰와 지속시간)에 영향을 미치므로, 이 세 요인에 대해 고객을 만족시키기 위한 전략을 강구해야 하는 것으로 나타났다. 반면에 세분시장 2는 구조적 요인의 개선과 플로우의 증가를 위한 방안들을 수립해야 하는 것으로 나타났다. 세분시장 3은 웹사이트 평가요인을 개선 또는 변경하기 보다는 고객들이 머문 콘텐츠 등을 확인하여 촉진활동(promotion)을 실시하는 것이 적절할 수 있을 것이며, 세분시장 4는 어떤 행동을 마케터가 해도 웹사이트 방문행위에 영향을 미칠 수 없으므로 오히려 de-marketing을 실시하는 것이 바람직할 수

도 있을 것이다. 특히, 믹스처 모델은 Fuzzy clustering 알고리즘을 채용하고 있어서, 어느 관찰치가 어느 세분시장에 속하는지를 알 수 있다. 이를 토대로 각 세분시장별로 별도의 사이트화면을 노출시키는 것도 가능하다. default로 세분시장과는 상관없이 전체 고객을 대상으로 한 contents를 화면에 노출시키고, 각 세분시장에 해당하는 고객들이 로그인하면 각 세분시장별 특성을 고려한 화면을 default 화면과 함께 보여줄 수 있을 것이다. 가령, 세분시장 1에 해당하는 고객이 방문을 하면 홈페이지에 오락적/구조적/정보적 요인이 강화된 화면을 이 고객에게 노출시킬 수 있게 되는 맞춤형 홈페이지가 가능하다.

## V. 요약 및 토의

기존의 인터넷소비자의 방문행동에 관한 연구들이 대부분 전체시장 수준에서 이루어졌고, 시장세분화를 하더라도 사전적 시장세분화로서 연구자의 주관이 반영되는 경우가 많다는 문제점을 확인하였다. 이에 본 연구에서는 세분시장 수준에서의 인터넷소비자 방문행동 연구가 필요하다는 점과 함께 연구자의 임의성을 배제한 사후적 시장세분화의 필요성을 제기하고 이를 믹스처모델로 실증적으로 분석하였다. 또한 기존 연구가 웹사이트 평가요인과 방문행동 간의 단편적인 인과관계를 파악한 것과는 달리, 웹사이트 방문행동에 영향을 미치는 요인들 간의 간접효과에 따른 통합적인 인과관계를 파악하였다.

본 연구는 navigation pattern의 이질성을



반영하기 위해 실제 커뮤니티 사이트를 방문한 1,765명의 웹상의 움직임과 설문조사를 실시하고 나서, 새로운 시장세분화 방법론인 믹스처모델을 이용하여 세분시장별 이질적인 웹사이트 방문행동을 확인하였으며, 이질적인 웹사이트 방문행동에 영향을 미치는 선행요인으로 웹사이트 평가요인들과 방문행동 간의 인과관계를 세분시장별로 확인하였다. 이를 통해 각 세분시장별로 navigation pattern과 원인변수 간 상이한 인과관계가 존재한다는 사실을 입증하였다.

마케터는 이런 세분시장별 이질적 관계를 이용하여 eCRM을 실행함에 있어 navigation pattern의 이질성에 영향을 미치는 변수를 세분시장별로 조절함으로써 기업이 원하는 방향으로 고객의 웹사이트 방문행위를 유도할 수 있는 기회를 얻게 되었다. 이것이 본 연구의 실무적 공헌점이다. 앞으로 마케터는 자신의 고객들의 인터넷상의 navigation pattern을 분석할 때 세분시장별 분석이 반드시 필요함을 인식하게 되었다. 또한 학술적으로는 navigation pattern의 이질성 분석에 기존에 사용되지 않았던 사후적 시장세분화를 위한 새로운 대안으로 믹스처모델을 적용하는 성과를 얻었다.

하지만 이 연구는 특성 커뮤니티 사이트에 국한된 연구라는 한계가 있다. 또한 세분시장별로 인구통계학적 특성의 차이는 발견되지 않았다. 이런 결과를 보면 태도나 가치와 같은 심리학적변수를 추가적으로 분석할 필요성이 대두되었다.

## 참고문헌

- 김소영, 한상만, 박세환(2001), 커뮤니티사이트의 eCRM 전략에 관한 연구, 한국마케팅학회 춘계학술대회 발표논문집, 2001. 3. 10
- 김소영, 한상만, 박세환(2002), 웹사이트 평가요인과 고객 방문행동의 관계에 관한 연구 - STIVI 매트릭스를 이용한 eCRM 전략을 중심으로, 경영학연구, Alba, Joseph, John Lynch, Barton Weitz, Chris Janiszewski, Richard Lutz, Alan Sawyer & Stacy Wood (1997), "Interactive Home Shopping: Consumer, Retailer, and Manufacturer Incentives to Participate in Electronic Marketplaces", *Journal of Marketing*, 61, 38-51.
- Ansari Asim & Carl Mela(2000), "Targeting Electronic Content in Interactive Media", *Working Paper*.
- Berry M.J.A. & Gordon Linoff(2000), *Mastering Data Mining*, John Wiley & Sons, Inc.
- Ben-Akiva, M. and Lerman, S. R.(1993) *Discrete Choice Analysis: Theory and Application to Travel Demand*. London: The MIT Press.
- Bucklin, R.E. and Gupta, S.(1993), "Brand Choice, Purchase Incidence, and Segmentation: An Integrated Modeling Approach", *Journal of Marketing Research*, 29(May), 210-215.
- Bucklin, R. E., Gupta, S., and Han, S.(1995), "A Brand's Eye View of

- Response Segmentation in Consumer Choice Behavior. *Journal of Marketing Research*, 32(Feb.) 66-74.
- Bucklin, R. E., Gupta, S., & Siddarth S. (1991), "Segmenting Purchase Quantity Behavior: A Poisson Regression Model", *Working Paper*, Graduate School of Management, University of California, LA.
- Bucklin, Randolph E. and Catarina Sismeiro (2000), "How sticky is Your Web Site? Modeling Site Navigation Choices Using Clickstream Data", *Working Paper*, Anderson School, UCLA.
- Charnes, A., Cooper, W.W., and Rhodes, E.(1978), "Measuring the Efficiency of Decision Making Units", *European Journal of Operational Research*, 2, 429-444.
- Chen and Wells(1999), "Attitude toward the Site", *Journal of Advertising Research*, Sep · Oct. 27-37.
- DeSarbo, W. S., Manrai, A.K. and Manrai, L.A.(1994), "Latent Class Multidimensional Scaling: A review of Recent Developments in the Marketing and Psychometric Literature", in *Advanced Methods of Marketing Research*, Richard P. Bagozzi(ed.), Cambridge: Blackwell, 190-222.
- Dillon, W. R. and Kumar, A.(1994), "Latent Structure and other Mixture Models in Marketing: An Integrative Survey and Overview", in *Advanced Methods of Marketing Research*, Richard P. Bagozzi(ed.), Cambridge: Blackwell, 295-351.
- Dreze, Xavier and Fred Zufryden (1997), "Testing Web site Design and Promotional Content," *Journal of Advertising Research*, 37(2).
- Green, P. E.(1997), "A New Approach to Market Segmentation", *Business Horizons*, 20, 61-73.
- Han, S. & Park, S. (2003), "The Evolution of Navigation Pattern in Consumer Choice in e-Commerce Site", *2003 Marketing Science Conference Proceeding*, University of Maryland,
- Hanson, Ward (2000), *Principles of Internet Marketing*, South-Western College Publishing.
- Hoffman, D. & Novak, T.P.(1996), "Marketing in Hypermedia Computer-Mediated Environments: Conceptual Foundations", *Journal of Marketing*, 60, 50-68.
- Jarvenpaa Sirrka L. & Peter A. Todd(1997), "Consumer Reaction to Electronic Shopping on the World Wide Web", *International Journal of Electronic Commerce*, 1(2), 59-88.
- Kamakura, W.A. and Russell, G.J.(1989), "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure", *Journal of Marketing Research*, 15(Nov.), 379-390.
- Kamba, T., Bharat, K. and Albers, M. C. (1995), "The Krakatoa Chronicle-An Interactive, Personalized Newspaper

- on the Web," In *Proceedings of the 4th International WWW Conference*, Boston, Massachusetts, 159-170.
- Lehmann, D.R, Gupta, S., and Steckel(1998), *Marketing research*, New York: Addison Wesley publishing Inc. NY.
- Lohse, Gerald L. & Eric J. Johnson(1999), "Consumer Buying Behavior on the Internet ; Finding from panel data", *Working Paper*, Wharton School.
- McLachlan G. and Basford, K.E.(1998), *Mixture Model: Inference and Applications to Clustering*, New York: marcel Deckker.
- McLachlan, G. and Peel, D.(2000), *Finite mixture model*, New York: John Wiley & Sons Inc. 2000.
- Mandel, Naomi and Eric. J. Johnson(1999), "Construction Preferences Online : Can Web Pages Change What You Want?," *Working Paper*, The Wharton School.
- Morrison, Donald G. and David C. Schmittlein(1981), "Generalizing the NBD Model for Customer Purchases : What Are the Implications and Is It Worth the Effort?," *Journal of Business & Economics Statistics*, 6(2), 145-166.
- Moe, W. W. and Fader, P. S. (2000), "Capturing Evolving Visit Behavior in Clickstream Data," *Working Paper*, The Wharton School, University of Pennsylvania.
- Novak,T. & Hoffman, D.L.(1997), "Measuring the flow Experience among Web Users", *Working Paper*, Vanderbilt University.
- Novak,T. & Hoffman, D.L. & Yung, Y.F.(1998), "Measuring the flow Construct in On-line Environments: A Structural Modeling Approach", *Working Paper*, Vanderbilt University.
- Oral, M., Ossama, K., and Lang, P, A.(1991) "Methodology for Collective Evaluation and Selection of Industrial R&D Projects", *Management Science*, 37(7), 871-885.
- Parkan, C., Lam, K., and Hang, G.(1997), "Operational Competitiveness Analysis on Software Development", *Journal of the Operational Research Society*, 48, 892-905.
- Rosbergen, E.A. Pieters, F.G.M., and Wedel, M.(1997), "Visual Attention to Advertising: A Segment-level Analysis", *Journal of Consumer Research*, 24, 305-314.
- Shapiro, B.P. and Bonoma, T.V.(1984), "How to Segment Industrial Markets", *Harvard Business Review*, 62, 104-110.
- Siddarth, S., Bucklin, R.E., & Morrison, D.G. (1995). "Making the Cut: Modeling and Analyzing Choice set Restriction in Scanner Data", *Journal of Marketing Research*, 32(Aug.), 225-266.
- Spiller, P. & Lohse, G.L.(1998), "A Classification of Internet Retail Stores", *International Journal of Electronic Commerce*, 2.(2)

- Sukekyu Lee, Xavier Dreze, and Fred Zufryden (2000), "Modeling Consumer Visit Frequency on the Internet," , *Working Paper*, University of Southern California.
- Vriens, M., Wedel, M., and Williams, T.(1996), "Metric Conjoint Segmentation Methods: A Monte Carlo Comparison," *Journal of Marketing Research*, 32, 73-85.
- Wedel , M. and Desarbo, W.S.(1995), "A Mixture Likelihood Approach for Generalized Linear Models," *Journal of Classification*, 12, 1-35.
- Wedel, M, and Kamakura, W.A.(2000), *Market Segmentation: Conceptual and Methodological Foundations*, Kluwer Academic Publisher, Boston.
- Wedel, M., and Steenkamp, J. B.(1989), "Fuzzy Clusterwise Regression," *International Journal of Research in Marketing*, 6, 45-49.
- Wedel, M. and Steenkamp, J.B.(1991), "A Clusterwise Regression Method for Simultaneous Fuzzy Market Structuring and Benefit Segmentation," *Journal of Marketing Research*, 28: 385-396.

(Abstract)

## Exploring Navigation Pattern and Site Evaluation Variation in a Community Website by Mixture Model at Segment Level

Kim, So Young<sup>\*</sup> · Young Sik Kwak<sup>\*\*</sup> · Yong Sik Nam<sup>\*\*\*</sup>

Although the site evaluation factors that affect the navigation pattern are well documented, the attempt to explore the difference in the relationship between navigation pattern and site evaluation factors by post hoc segmentation approach has been relatively rare.

For this purpose, this study constructs the structure equation model using web-evaluation data and log file of a community site with 300,000 members. And then it applies the structure equation model to each segment. Each segment is identified by mixture model. Mixture model is to unmix the sample, to identify the segments, and to estimate the parameters of the density function underlying the observed data within each segment.

The study examines the opportunity to increase GFI, using mixture model which supposes heterogeneous groups in the users, not through specification search by modification index from structure equation model. This study finds out that AGFI increases from 0.819 at total sample to 0.927, 0.930, 0.928, 0.929 for each 4 segments in the case of the community site. The results confirm that segment level approach is more effective than model modification when model is robust in terms of theoretical background. Furthermore, we can identify a heterogeneous navigation pattern and site evaluation variation in the community website at segment level.

Key Words : Navigation Pattern, Structure Equation Model, Specification Search by Modification Index, Mixture Model, Log File Analysis, Web Evaluation Factors

---

\* Korea Culture and Tourism Policy Institute.

\*\* Jinju National University

\*\*\* Samsung Opentide China