

## 고객지향 세분시장 획득을 위한 데이터 마이닝 기법 적용방안

김종호\*

### 요약

공급자 관점의 상품시장 정의는 공정, 원자재, 상품기능의 유사성 등으로 세분시장을 정의함으로써 특정상황이 배제되고 불연속성에 대한 고려가 결여되어 있어 기업의 시장 활동에 여러 가지 문제를 발생시킬 수 있다. 이러한 정의는 다분히 정적이고 일반적이어서 시장의 여러 상황과 시간의 추이에 따른 시장변화를 표현, 예측하는 것이 매우 어렵다. 반면 고객지향의 시장 세분화는 특정 혜택이 추구되는 상황에서 대체 가능한 상품과 관련 고객들의 집합을 정의함으로써 획득 가능하다. 이러한 상품시장 정의는 시장에서 부상하는 위협과 기회를 발견하고 효과적인 실적평가와 효율적인 자원배분을 가능케 한다. 그러나 고객관점의 시장정의를 실제로 구현되기 위해서는 많은 전제조건이 있다. 즉, 고객의 행위와 판단에 관련된 충분한 자료의 확보가 가능해야 하며 이들 자료의 분석을 통해 의미 있는 정보를 제공할 수 있는 정보처리기법이 필요하다. 본 연구의 목표는 고객관점 상품시장 정의가 귀납적 방법에 상당히 근거한다는 데 착안하여 데이터 마이닝 기술을 활용하여 실제적 적용에 어려움이 많은 고객관점의 시장 세분화를 지원할 수 있는 방안을 제시하는 데 있다.

## Application of data mining techniques for finding customer-oriented product market segments

Jongho Kim\*

### Abstract

The definition of the product market in a supplier's point of view can cause various problems in the market activities of companies because specific situations are excluded and the consideration for discontinuity is lacking by identifying segmented markets with processes, raw materials, the similarity of product functions and so forth. Furthermore, as this definition is static and general, it is difficult to express and predict the dynamic market changes. Meanwhile, customer-oriented market segment can be obtained by grouping substitutable products and related customers in the situation pursuing specific benefits. This definition of the product market enables us to find threats and opportunities emerging in markets and promotes effective performance assessments and resource allocation. The purpose of this paper is suggesting a framework to select data mining techniques proper for the customer data characteristics to identify customer oriented product market.

Keywords : Customer-Oriented Market Segmentation, Data Mining

## 1. 서론

### 1.1 연구의 배경

※ 제일저자(First Author): 김종호  
접수일:2012년 08월 21일, 수정일:2012년 09월 10일  
완료일:2012년 09월 12일

\* 경성대학교 경영정보학과  
jonghokim@ks.ac.kr

■ 본 연구는 경성대학교 2012학년도 신입교수정착 연구비에 의하여 연구되었음

공급자 관점의 상품시장 정의는 특정상황이 배제되고 불연속성에 대한 고려가 결여되어 있어 기업의 시장 활동에 여러 가지 문제를 발생시킬 수 있다. 즉, 시장의 위협과 기회를 포착해서 대응책을 만드는 데 민첩하지 못하며 실적평가나 자원분배에 어려움을 줄 수 있다. 따라서 상황변수를 고려하는 고객관점 상품시장 정의에 대한 연구의 필요성이 증대되고 있으며 이를 획득하기 위한 계량적 방법론들이 제시되고 있다 [1,2]. 그러나 고객관점 시장정의를 마케팅 관리자들에게 친숙하지 못하며 실행하기가 어렵고

마케팅조사에 많은 비용이 든다는 단점이 있다.

공급자 관점의 시장정의는 필연적으로 단기적인 자원의 투입을 통한 소비자 욕구의 충족이라는 시각과 세분시장의 부상과 쇠퇴, 기술과 가격, 촉진 등이 시장의 구조에 미치는 영향을 고려하는 장기적 시각의 조정을 통해 도출된다. 결과적으로 공정, 원자재, 기능의 유사성 등으로 세분시장이 정의되어 왔다. 그러나 이러한 정의는 다분히 정적이고 일반적이어서 시장의 여러 상황과 시간의 추이에 따른 시장변화를 표현, 예측, 통제하는 것이 매우 어렵다. 이러한 단점에도 불구하고 공급자 관점의 시장정의는 마케팅 조사가 수월하고 명확한 경계를 얻을 수 있다는 점에서 널리 이용되고 있다[3].

반면 고객지향의 시장 세분화는 특정 혜택이 추구되는 상황에서 대체 가능한 상품과 관련 고객들의 집합을 정의함으로써 획득 가능하다. 따라서 상황변수와 대체가능성의 두 기준 중 적어도 하나는 필수적인 세분변수로서 활용되어야 한다. 이러한 상품시장 정의는 시장에서 부상하는 위협과 기회를 발견하고 효과적인 실적평가와 효율적인 자원배분을 가능케 한다[4]. 그러나 고객관점의 시장정의가 실제로 구현되기 위해서는 많은 전제조건이 있다. 즉, 고객의 행위와 판단에 관련된 충분한 자료의 확보가 가능해야 하며 이들 자료의 분석을 통해 의미 있는 정보를 제공할 수 있는 통계적기법 또는 정보처리기법이 필요하다. 따라서 고객관점 시장정의는 대부분의 경우 마케팅 조사에 소요되는 비용이 혜택을 초과함으로써 실제적 활용이 이루어 지지 못하고 있다.

기업의 시장활동이 누적되고 정보기술의 발달로 고객의 반응을 수집, 저장하는 활동이 빈번해지면서 최근 기업들은 축적된 데이터를 전략적으로 활용하기 위한 움직임이 활발하다. 즉 기존 고객과의 거래 데이터를 활용하여 소비자 행동에 대한 통찰을 얻는 데 많은 관심을 가지고 있다[5]. 어떤 고객들이 해당기업의 브랜드에 머무르는지 또한 어떤 고객들이 타 기업의 브랜드로 전환하는 구매행동을 보이는지를 간파하려 하고 있다. 이러한 기업분석 행위의 궁극적인 목적은 소비자의 충성도를 예측하는 데 있다. 이러한 활동은 고객관점의 시장발견을 추구하는 것과 유사한 결과를 얻을 수 있다.

소비자 행동의 통찰과 고객관점의 시장발견을 위한 방법으로서 계량적, 통계적 기법과 최근 부상하고 있는 데이터 마이닝 기법 등 많은 도구들이 때로는 경쟁적관계를 가지면서 때로는 보완적 관계를 가지면서 마케팅 관리자들에 지원하고 있다. 데이터 마이닝 기술은 기존의 연역적 추론에 근거한 정보기술 들의 단점을 극복하고 귀납적 지식을 제공할 수 있다는 관점에서 적용 가능성이 높다. 즉 의미 없는 데이터의 집합으로부터 의미 있는 관계를 추출하고 이의 신뢰성을 획득할 수 있는 기술이다[6].

## 1.2 연구의 범위와 목적

이에, 본 연구는 고객관점 상품시장 정의가 귀납적 방법에 상당히 근거한다는 데 착안하여 데이터 마이닝 기술을 활용하여 고객관점 상품시장을 파악하는 것을 효과적으로 지원하는 데 그 목표를 두고 있다. 이와 같은 연구 목표를 달성하기 위해 본 연구는 고객 관점 세분시장 발견을 위한 계량화된 기법들을 기존 연구에서 찾아 체계적으로 정리한다. 그리고 데이터마이닝 기법을 기존 분류방법에서 탈피하여 창의적으로 재분류한 후 고객 관점 세분시장 발견을 위한 적용방안을 제시한다. 이로써 많은 장점을 지니고 있으면서도 마케팅 조사의 어려움으로 실제적 활용이 힘든 고객관점 상품시장의 발견이 용이해 질 수 있을 것으로 기대된다.

## 2. 고객관점 세분시장의 발견

### 2.1 고객관점 세분시장을 정의하기 위한 요건

시장 세분화는 마케팅과정의 핵심과정으로 단기적으로 기업 내 자원의 분배와 실적평가의 근본이 되며 장기적으로 기업경쟁력 향상의 기회를 제공한다. 그런데 대부분의 기업은 공급자관점에서 세분시장을 발견하고 이를 활용하여 다양한 마케팅 프로그램을 전개해 나가고 있다. 그러나 고객이 바라보는 시장과 공급자가 바라보는 시장과는 상당한 인식의 차이가 발생한다. 공급자 관점의 시장 세분화는 제조과정, 원자재, 기능의 유사성 등이 세분화의 기준이 되는 반면 고객 지향의 세분시장은 특정상황에서 대체 가

능한 상품과 고객들의 집합으로 정의되어진다. 또한 도출과정의 논리적 전개에서도 공급자 관점의 세분시장은 연역적 추론의 결과로, 고객지향의 세분시장은 귀납적 추론으로 획득 가능하다.

서론에서 언급한 바와 같이 고객관점의 시장 세분화가 많은 장점을 가지고 있음에도 불구하고 다음과 같은 전제조건을 충족시켜야만 한다는 점에서 실제적 활용에 많은 장애가 있다.

첫째, 고객관점의 시장 세분화가 귀납적 추출의 결과라는 점에서 고객의 구매행위와 판단에 관련된 데이터가 충분히 확보 되어야 한다. 단기적 관점에서 세분 고객시장의 획득은 구매행동에 관련된 데이터를 이용하게 되는데 여기에는 수요의 교차탄력성, 행동의 유사성, 상표전환에 관련된 데이터들이 속하게 된다. 이들 데이터를 통해 발견된 세분시장은 기업의 전술설정에 활용되어진다. 반면, 고객의 판단과 관련된 데이터를 이용하여 발견된 세분시장은 기업의 장기적 전략계획에 활용 가능하다. 만약, 이들 데이터에 대한 확보가 불가능하다면 고객관점의 시장 세분화는 관리자의 직관에 의존할 수밖에 없다.

둘째, 세분화의 기준 변수인 상황변수와 대체가능성 관련변수의 설정과 이들의 타당성, 신뢰성에 대한 밀도 높은 검증이 요구된다. 상황변수의 도입은 고객이 특정혜택을 추구하는 상황을 세분화하고 각 상황에 대해서 대체가능성 관련 변수를 통해 고객의 미충족 욕구를 충족시킬 수 있는 도구를 발견하고 그들 간의 대체정도를 측정함으로써 고객지향의 세분시장 발견이 가능하다.

셋째, 세분결과에 대한 시간적, 경제적 고려가 필요하다. 대체적으로 고객지향의 세분시장은 공급자관점 세분시장보다 시간의 흐름에 더욱 민감한 경향이 있다. 따라서 마케팅 조사를 계획함에 있어서 시간적 고려와 이에 근거해서 손익 및 비용을 철저히 따져보는 경제적 고려가 요구된다.

**2.2 고객관점시장발견을 위한 방법론**

본 연구가 추구하고 있는 고객관점 시장 발견을 위한 데이터 마이닝 기술의 활용을 논하기 전에 기존의 계량적, 통계적 기법들에 대한 문헌 조사를 통해 상황변수와 대체가능성 관련 변수,

그리고 그룹화, 기법들을 요약하고자 한다.

세분시장 발견을 위한 방법론은 <표 1>과 같이 크게 고객의 활용하는 데이터 유형에 따라 행동데이터 의존형과 판단데이터 의존형으로 분류될 수 있다.

<표 1> 고객관점 세분시장 발견을 위한 기법

구분	분석방법
행동데이터 의존형	교차 가격탄력성 분석
	구매행동의 유사성 분석
	브랜드 전환 분석
판단데이터의존형	결정과정 분석
	인식맵핑
	기술대체 분석
	고객 대체가능성 판단 분석

**2.2.1 교차가격탄력성 분석**

행동데이터 의존형의 대표적인 방법은 수요의 교차 가격탄력성 분석 방법이다. 즉, 경쟁회사의 가격변화라는 상황변수를 개입시켜 해당 브랜드의 수요에 미치는 정도로서 세분시장을 획득하는 방법이다. 그러나 교차가격탄력성을 이용하는 방법에는 여러 가지 문제점이 있다. 우선 한 브랜드의 가격변화로 타 브랜드의 대응이 없는 경우는 찾아보기가 드물며, 교차탄력성 자체가 시간에 의존하지 않는 정적인 측정도구이기때문에 상품구성의 변화를 반영하기 힘들다. 또한 동시적인 가격변화를 반영하고 다른 요소들의 영향으로 인한 수요변화를 분리시키는 작업이 필요하다.

**2.2.2 구매행동 유사성 분석**

이 분석방법은 특정 사용상황에서 사용될 수 있는 상품들의 군을 미리 분류한 다음 각 상품의 혜택이 유사성을 갖는다면 그들을 세분시장으로 분류하는 방법이다.

**2.2.3 브랜드 전환분석**

브랜드 전환에 대한 측정도구로서 널리 사용되고 있는 것은 가장 최근에 브랜드 B를 구매했을 때 브랜드 A로 전환하는 조건부 확률이다. 이 확률이 높을수록 브랜드 간 대체가능성이 높다. 그러나 브랜드 전환 분석방법이 활용될 수 있는 상품범주는 구매빈도가 높아야 하고, 동시에 여러 브랜드를 구매하는 상황이라면 그 유용성에 한계가 있다.

2.2.4 결정과정 분석

이 분석방법은 구매행위 과정에서 고객의 마음속을 스쳐가는 상품 속성의 우선순위를 정하고, 이를 근거로 나무형태의 계층적 구조를 만들어 장단기적 관점에서 상품시장의 세분화를 시도하는 분석 방법이다.

2.2.5 인식맵핑

인식 맵핑은 고객의 선호와 브랜드의 위치를 여러 속성으로 조합하여 지도에 표현하고 고객의 선호와 브랜드간, 브랜드와 브랜드간의 차이를 여러 의미로 해석하는 방법이다. 대부분의 경우 이들간의 차이는 대체가능성보다는 경쟁구조로서 이해될 수 있으며 새로운 상품의 도입, 기존 브랜드의 이동, 기존 고객의 선호도 변화를 예측하는 등 여러 가지 용도로 사용될 수 있다.

2.2.6 기술대체 분석

원자재, 생산과정, 신상품 등, 시장의 동적인 변화를 줄 수 있는 요소들이 개입되었을 때 시간의 흐름에 따라 변화과정을 이해하고자 하는 분석 방법이다. 이 방법은 기술대체의 자취와 시간적 과정은 보통 S자형을 따르는 것으로 가정하고 자취를 실제와 합치시키는 데 오차를 줄이는 것을 목표로 하고 있다.

2.2.7 고객대체가능성 판단 분석

이 방법은 여러 상황에서 대체가능정도를 설문을 통해 직접 소비자로부터 얻는 방법이다. 설문의 방식에 따라 여러 방법이 개발되어 있다.

<표 2> 고객관점 세분시장 발견을 위한 세분변수와 그룹화 기법의 분류

구분	단기적 (정적)	장기적 (동적)
세분 변수	수요의 교차가격 탄력성 특정사용상황에서 구매행동 유사성	구매과정에서의 상품속성 고객선호와 브랜드의 속성 조합
	브랜드 전환의 조건부 확률	원자재, 생산과정, 신상품 등 시장의 동적인 변화를 줄 수 있는 요소
그룹화 기법	상호작용 정도로서 그룹화	상품속성의 우선 순위에 따른 그룹화와 계층화
		시간의 추이에 따른 상품속성과 고객 선호의 변화를 그룹화에 반영

지금까지 공급자 관점의 시장 세분화의 대체 수단으로서 고객관점의 시장 세분화 방법들을

정리하였다. 이를 토대로 단기적(정적), 장기적(동적) 관점에서 이용되는 세분변수와 그룹화 기법을 <표 2>에서 정리하였다.

3. 데이터 마이닝 분석기법

데이터 마이닝은 데이터의 집합으로부터 패턴을 찾아 의사결정 과정을 지원해주는 프로세스로서 정의되고 있다. 따라서 데이터 마이닝은 단순한 정보기술이 아닌 의사결정의 한 부분으로서 사람, 기술, 절차가 복합된 형태로 이해하는 것이 옳바르다. 따라서 데이터 마이닝은 여러 각도에서 조명하는 것이 가능한데 본 연구에서는 고객 세분시장 획득을 위한 기법의 지원이라는 목적에 따라 절차적 관점, 사용자 관점 그리고 기술적 관점에서 데이터 마이닝을 창의적으로 재분류하고 설명한다.

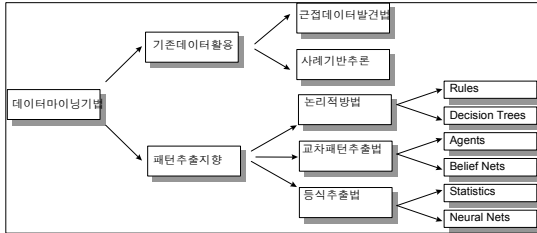
3.1 절차적 관점

절차적 관점에서 데이터 마이닝은 의사결정의 어떤 부분을 지원해 주는가에 따라 발견지향적, 예측지향적, 검증지향적 데이터 마이닝의 세가지로 분류 가능하다[7,8].

발견지향적 데이터 마이닝은 아무런 사전지식이나 사용자의 가설 없이 데이터의 집합에서 어떤 패턴을 발견하는 절차를 지칭한다. 다시 말해서 프로그램 자체적으로 관심도 높은 패턴을 자동적으로 인식하여 사용자에게 제공하게 된다. 이 기법에서 주요한 이슈는 발견되는 패턴의 수와 질이다.

예측지향적 데이터 마이닝은 발견된 패턴을 이용하여 미래를 예측하는 데 활용된다. 따라서 이 과정에서 사용자는 시스템에게 시장점유율이나 경상이익과 같은 관심 있는 변수와 모수를 입력하게 하고 시스템은 모수와 발견된 패턴에 근거해서 변수의 값을 예측하게 된다.

검증지향적 데이터 마이닝은 발견된 패턴을 이용하여 비정상적인 또는 유의적인 상태를 발견하는 데 목적이 있다. 이들 절차적 관점의 데이터 마이닝은 (그림 1)과 같이 여러 가지 용도에 따라 더욱 세분화될 수 있다.



(그림 1) 데이터 마이닝의 기술적 분류

### 3.2 사용자 관점

기업 내의 정보사용자는 직무와 직위에 따라 상이한 데이터를 요구하게 된다. 따라서 각 사용자에게 가치 있고 적합하게 만들어진 데이터를 공급하는 것은 데이터 마이닝에서도 예외일 수 없다. 데이터 마이닝은 운영계층, 관리계층과 임원계층 중 어느 계층을 지원하는가에 따라 일시적, 전략적, 지속적 데이터 마이닝으로 분류할 수 있다.

일시적 데이터 마이닝은 현 상태에 대한 스냅 샷을 얻는 방법이다. 즉 시간적 추이나 결과변수 보다는 통제 가능한 변수, 비통제 가능한 변수들에 대한 현재 값을 구하는 데 목적이 있다. 이 데이터 마이닝은 운영계층을 지원해 줄 수 있다.

전략적 데이터 마이닝은 통제변수나 비통제 변수보다는 시장점유율, 고객만족지수와 같은 요약된 형태의 정보를 얻는데 중점을 두고 있다.

지속적 데이터 마이닝은 일정시간 동안 비통제 변수나 통제변수 또는 결과변수의 변화를 파악하는 데 그 목적이 있다. 예를 들어 “1/4분기 동안 고객의 세일즈 추세는 어떻게 변화할 것인가?”라는 물음에 대한 해답을 구하는 것이다. 지속적 데이터 마이닝은 일시적 데이터 마이닝과 전략적 데이터 마이닝이 선행되는 것을 전제로 한다. 전략적 데이터 마이닝과 지속적 데이터 마이닝은 관리계층과 임원계층을 지원해 줄 수 있다.

### 3.3 기술적 관점

데이터 마이닝을 기술적으로 분류하는 데 첫 기준이 될 수 있는 것은 패턴추출 후에 기존의 데이터에 대한 유용성이다. 검증지향적 데이터 마이닝과 예측지향적 데이터 마이닝의 경우 기존 데이터를 활용하게 되지만 발견지향적 데이터 마이닝에서 기존의 데이터는 불필요하다. 협

의 데이터 마이닝이 발견지향적 데이터 마이닝에 국한되어 있고 검증, 예측 지향 데이터 마이닝이 핸드리 모델처럼 기존의 데이터를 활용할 경우 그 자체로서 타당하지 못하다는 약점을 지니고 있어 대부분의 경우 기존 데이터를 활용하는 경우는 드물다.

#### 3.3.1 기존데이터 활용기법

(그림 1)의 기존데이터 활용기법은 패턴 발견 후에도 패턴 매칭을 위하여 기존의 데이터를 계속 활용하는 방법으로 대표적인 방법이 예측을 위해 새로운 데이터 인스턴스가 추가되었을 때, 그것과 기존 데이터들과 비교를 통하여 가장 비슷한 데이터 인스턴스를 추출해 내는 최근접데이터발견법이 있다. 예를 들어, 은행고객 중 30대 중반이며 연봉이 3000만원에 가장 근접한 10명의 고객을 찾고자 할 때 이 방법을 사용하게 된다. 그러나 보통 기업에서는 지속적으로 누적되는 엄청난 데이터를 모두 보유하는 데는 어려움이 있기 때문에 “전형적인 사례”를 만들어 두고 활용하게 되는데 이것이 소위 사례기반추론 접근법이다. 그러나 이 방법에는 “전형적인 사례”를 어떻게 만들어야 하며 시간의 변화에 따른, 새로운 변수의 개입에 따른 사례의 갱신이 어렵기 때문에 정적인 방법이라는 한계가 있다. 더욱 비판적인 것은 비수치적 변수의 경우 사례간의 차이를 두는 것과 추론상의 어려움이 있다는 점이다. 보통은 근사적 방법과 기호적 추론으로 해결하고 있다.

#### 3.3.2 패턴추출지향방법론

기존 데이터를 활용하지 않는 경우는 패턴을 발견하는 데 중점을 두고 일단 패턴 발견 후에는 기존 데이터를 폐기 처분하게 된다. 이 방법에는 (그림 3)처럼 논리적 추출, 교차 패턴추출법, 등식추출법으로 더욱 세분화 할 수 있다. 이 분류는 수학적 체계에 그 기원을 두고 있다. 예를 들어 논리적 추출이 결과적 지식을 조건식으로 표현 (e.g. IF  $6 < X < 7$  THEN  $1 < Y < 2$ )하는 반면 등식추출법은 방정식 또는 등식의 형태 (e.g.  $(a * X) + b$ )로 표현한다. 따라서 실제적으로 추출되는 방식과 활용되는 데 상당한 차이가 있다. 등식추출법은 변수들간의 관계를 직선이나 곡선으로 표현하는데 용이하며 논리적 추출방법은

비수치적인 변수들을 처리하는데 많은 장점이 있다.

(1) 논리적 추출

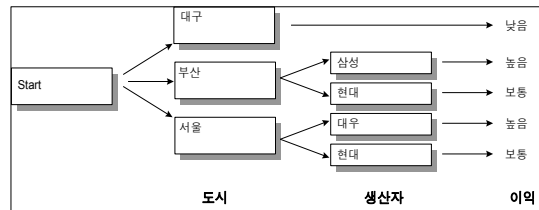
우선 논리적 추출에서 지식을 조건식으로 표현할 때 두 가지 표현법이 있는데 조건논리와 조합논리로 분류된다. 조건논리는 IF/THEN문장으로 표현되며 예로서 “만약 비가 온다면, 날씨가 흐리다 (IF it is raining THEN it is cloudy)”를 들 수 있다. 조건 논리는 두개의 조건으로 구성되며 각 조건은 변수와 값의 등식관계로서 표현되고 선후관계를 갖게 된다. 반면 조합논리는 WHEN/ALSO문장으로 표현되며 예로서 “페인트를 구매할 때 페인트 브러쉬도 구매한다 (WHEN paint is purchased ALSO paint brush is purchased)”를 들 수 있다. 조합논리는 두개의 사건으로 구성되며 이들은 동시적 발생관계를 맺게 된다. 또한 조건논리와 조합논리는 결합해서 사용될 수 있다. 즉 “IF day = Sunday WHEN Paint-brush is purchased ALSO paint is purchased”으로 결합되는 것이 가능하다.

논리적으로 표현된 지식을 데이터베이스로부터 자동적으로 얻어내는 방법에는 귀납적 추론과 유전 알고리즘, 결정나무분석 방법들이 있다. 귀납적 추론은 우선 특정변수의 분포상태를 확인하고 자동적으로 가설을 만든 후 가설과 함께 신뢰성(Confidence)을 사용자에게 보여주게 된다. 예를 들어 데이터베이스로부터 고객의 직업 분포를 확인하고 직업 중에 체육인을 선택한 후 그들의 연령분포를 확인해서 “IF 직업 = 체육인 THEN 연령 < 30 (confidence = 70%)”라는 규칙을 자동적으로 생성해 내게 된다. 유전 알고리즘은 일단 귀납적 추론을 통해 만들어진 여러 패턴들을 교배시켜 돌연변이적인 패턴을 찾고자 하는 방식이다. 귀납적 추론과 유전 알고리즘은 수행방식에서 많은 유사점이 있으며 입력물이 데이터이나 패턴이냐에만 차이가 있을 뿐이다. 결정나무 분석 방법은 우선 단계를 나누고 각 단계마다 변수들을 하나씩 도입해서 우선되는 변수의 값들로부터 후행 되는 변수의 값들을 링크 시켜 나무형태로 만들어 조건 논리를 획득하게 된다. 그러나 이 방법은 우선적으로 어떤 변수를 각 단계에서 도입해야 하는 문제와 변수의 수가 많을 때 실행에 어려움이 많다는

단점이 있다. 즉 <표 3>의 표본 데이터로부터 (그림 2)와 같은 결정 나무형태를 만들 때 “만약 상품색깔이 파랑이면 이익이 높다”라는 규칙은 얻기 힘들다.

<표 3> 표본 데이터

생산자	도시	상품색깔	이익
삼성	부산	파랑	높음
삼성	대구	초록	낮음
대우	서울	파랑	높음
대우	대구	빨강	낮음
현대	서울	초록	보통



(그림 2) 결정나무

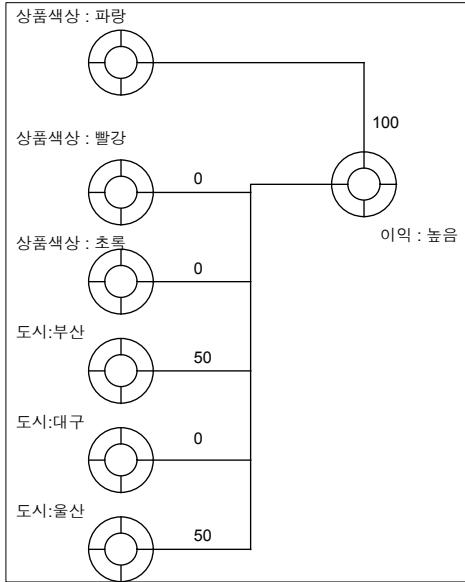
(2) 교차패턴추출법

<표 3>에서 생산자라는 변수는 제외시키고 각 변수의 값들을 교차 배열시킨 후 동시에 발생하는 경우들을 계산해서 <표 4>와 같이 표현할 수 있다.

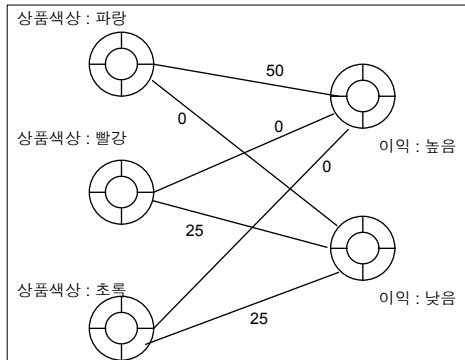
<표 4> 교차테이블

	부산	대구	서울	파랑	초록	빨강
높은이익	1	0	1	2	0	0
보통이익	1	0	1	0	1	1
낮은이익	0	2	0	0	1	1

<표 4>에서 건수를 가지는 조건인 “IF 상품 색깔 = 파랑 THEN 이익 = 높음”과 “IF 도시 = 대구 THEN 이익 = 낮음”이 신뢰성 높은 조건임을 쉽게 발견할 수 있다. 교차이블의 또다른 표현법으로서 변수와 값의 쌍으로 정의되는 에이전트(Agent)를 이용하는 방법이 있다. 즉 변수의 값이 고정된 객체를 에이전트(Agent)라고 한다. (그림 3)은 에이전트간의 링크를 발생횟수의 빈도로서 표현한 방법이고 (그림 4)는 링크를 조건부 확률로서 표현했다.



(그림 3) 에이전트 링크의 발생횟수 빈도 부여



(그림 4) 에이전트 링크의 조건부 확률 부여

(3) 등식추출법

지금까지 조건적 지식이나 동시적 상황의 발생수를 획득하는 방법론을 다루었다. 지금부터는 등식적 지식을 획득하기 위한 대표적 방법인 신경망 분석을 소개한다. 계량분석에서 등식적 지식의 획득의 대표적인 귀납적 방법은 회귀 분석이다. 단순 직선이나, 이차곡선의 경우 쉽게 얻어질 수 있지만 매우 복잡한 시스템의 경우 이러한 지식을 회귀분석을 이용해 추출하기는 매우 힘들다. 그러나 신경망 분석을 사용할 경우 학습을 통해 용이하게 얻을 수 있다. 신경망 분석은 데이터베이스로부터 데이터를 입력 받아 지속적으로 링크의 비중(weight)에 따른 계산결

과와 실제적 결과를 비교해서 연속적으로 링크 비중의 조정을 통해 학습시켜 오차가 최소화 되었을 때 학습을 멈춘다. 그리고 새로운 인스턴스를 입력노드에 투입하고 그 결과를 예측에 이용하게 된다.

4. 고객관점 세분시장 발견을 위한 데이터 마이닝 기술적용방안

지금까지 본 연구는 고객지향 시장 세분화 방법과 데이터 마이닝 기술들을 분류하고 정리하였다. 이 절에서는 고객 관점 시장 세분화를 위한 여러 방법론과 데이터 마이닝 기술과의 정합성을 고려해서 여러 각도에서 효과적인 데이터 마이닝의 활용을 살펴본다.

우선 단기적 관점에서 수요의 교차가격 탄력성, 특정 사용상황에서 구매행동의 유사성, 브랜드 전환의 조건부 확률 등으로 세분시장을 파악하는 정적인 방법은 절차적 관점에서 발견지향적 데이터 마이닝과 사용자 관점에서 운영계층을 지원하는 일시적 데이터 마이닝의 범주에 속한다. 따라서 기술적 관점에서 기존 데이터 활용보다는 패턴의 추출에 중점이 있다.

먼저 수요의 교차가격 탄력성을 이용한 세분시장의 발견은 등식적 지식보다는 조건적 지식을 요구한다. 즉, 정확한 탄력성의 정도보다는 일정이상의 수요변화를 야기하는 브랜드의 집합을 구하는 것이 목적이므로 이 경우, 귀납적 추론 또는 교차패턴 추출법을 활용하는 것이 추천된다. 특정 사용상황에서 구매행동의 유사성을 이용하는 세분시장 발견은 기존의 데이터를 활용하는 것이 바람직하다. 즉, 특정 사용상황을 새로운 인스턴스 또는 사례로서 표현하고 이와 유사한 데이터의 집합을 추출함으로써 세분시장의 발견이 가능하다. 브랜드 전환의 조건부 확률으로써 세분시장을 발견하는 방법은 교차패턴 추출법에서 각 브랜드간의 교차테이블이나 에이전트 링크를 활용하는 것이 추천되지만 구매행위의 우선순위를 표현하는 데이터의 확보도 동시에 이루어져야 할 것이다.

장기적 관점에서 시간적 고려와 대체성의 계층구조를 확보하는 방법인 결정과정 분석, 인식 맵핑, 기술대체 분석 등은 절차적 관점에서 예측

지향적이며 사용자관점에서 전략적, 지속적인 데이터 마이닝의 범주에 포함된다. 또한 변수들 간의 정교한 관계를 요구하기 때문에 등식추출법에 중점을 두게 된다.

결정과정 분석을 통한 세분시장의 획득은 변수들간의 계층구조를 얻는 것이 핵심이다. 따라서 결정나무 분석이 적합한 데이터 마이닝 기법이 될 수 있다. 인식 맵핑은 상품속성의 조합에 대한 응집도를 파악함으로써 가능하다. 이를 위한 데이터 마이닝 기법은 교차패턴 추출법 또는 기존데이터 활용법이 추천된다. 기술 대체분석의 경우 시간의 흐름에 따른 여러 변수들의 변화과정을 추적하기 때문에 신경망 분석을 활용하는 것이 바람직하다.

## 5. 결론

본 연구는 많은 장점을 지니고 있으면서도 마케팅조사의 어려움으로 실제적 활용이 힘든 고객관점 세분시장의 발견을 위해 데이터 마이닝의 활용가능성을 검토하였다. 세분시장 발견을 위해 적합한 분석방법은 시장의 특성에 따라 선택되고, 연쇄적으로 분석방법에 따라 알맞은 데이터 마이닝 기법이 선택된다. 따라서 데이터 마이닝 기법은 시장의 특성에 많은 영향을 받게 된다. 다시 말해서 데이터 마이닝을 활용하고자 하는 기업이 진입한 산업의 특성과 데이터 마이닝 정보기술은 밀접한 연관관계를 맺게 된다. 뿐만 아니라 조직내의 사용자가 고객관점의 세분시장을 단기적관점에서 검토하고자 하는지, 장기적 관점에서 검토하고자 하는지에 따라 활용되는 기법이 달라지기 때문에 사용자의 직위와 직무와도 관련이 있다. 근본적으로는 획득코자 하는 지식의 형태에 따라 데이터 마이닝 기법의 선택이 이루어져야 한다[9]. 그리고 마케팅 관리자들은 고객관점 세분시장의 획득을 위한 필수적인 전제조건으로서 고객관련 데이터가 충분히 확보되고 정제되어야 한다는 점도 인식해야 한다[10].

또한 향후 과제로서 실험 등을 통하여 본 연구에서 제시된 기법들과 기존연구와의 비교평가를 통해 본 연구의 결과들이 기존방법보다 우수함을 실증적으로 제시할 예정이다.

## 참 고 문 헌

- [1] G. S. Day, et al., "Customer-Oriented Approach to Identifying Product-Markets", *Journal of Marketing*, Vol. 43, pp. 8-19, 1979.
- [2] R. K. Srivastava, et al., "A Customer oriented Approach for Determining Market Structures", *Journal of Marketing*, Vol. 48, pp.32-45, 1984.
- [3] P. R. Dickson and J. L. Ginter, "Market Segmentation, Product Differentiation, and Marketing Strategy", *Journal of Marketing*, Vol. 51, pp. 1-10, 1987.
- [4] P. Kotler and K. L. Keller, *Marketing Management*, Pearson Education, New York, 2012.
- [5] D. Tanier, L. I. Rusu, *Strategic Management in utilizing data mining and warehousing technologies*, Information Science Reference, Hershey PA, 2010.
- [6] G. Linoff and M. Berry, *Data Mining Techniques: For Marketing, Sales and Customer Relationship Management*, Wiley, New York, 2011.
- [7] K. Parsaye and M. H. Chignell, *Intelligent Database Tools and Applications*, John Wiley and Sons, New York, 1993.
- [8] U. Fayyad et al., "From Data Mining to Knowledge Discovery in Databases", *AI Magazine*, Vol. 17, No. 3, pp. 37-54, 1996.
- [9] 송임영, 이태석, 신기정, 김경창, "고객관계관리를 위한 통합 데이터 마이닝 모형 연구", *한국지능정보시스템학회논문지*, 제13권, 제3호, pp. 83-99, 2007.
- [10] 김진성, "데이터 마이닝 사례기반추론 기법에 기반한 인터넷 구매지원 시스템 구축에 관한 연구", *한국경영과학회지*, 제28권, 제3호, pp.135-148, 2003.



### 김 종 호

1994년 : KAIST 경영정책학과  
(공학사)

1996년 : KAIST 경영정보공학과  
(공학석사)

2003년 : KAIST 경영공학과  
(공학박사)

2003년~2006년: 삼성SDS 책임

2006년~2008년: 가톨릭대학교 연구조교수

2008년~2011년: 현대경제연구원 연구위원

2011년~현 재: 경성대학교 경영정보학과 조교수

관심분야 : 비즈니스인텔리전스, IT정책 등