

신규시장 성장모형의 모수 추정을 위한 전문가 시스템*

이동원

한성대학교 경영학부
(dongwonlee@hansung.ac.kr)

정재권

국민대학교 경영학부
(jchung@kookmin.ac.kr)

정여진

국민대학교 경영학부
(ychung@kookmin.ac.kr)

박도형

국민대학교 경영정보학부
(dohyungpark@kookmin.ac.kr)

시장 수요 예측은 일정 기간 동안 소비자에게 판매되는 동종 제품 또는 서비스의 수량 혹은 매출액의 규모를 추정하는 활동으로서, 기업경영활동에 있어 효율적인 의사결정을 내릴 수 있는 근거로 활용된다는 점에서 중요하게 인식되고 있다. 신규 시장의 수요를 예측하기 위해 다양한 시장성장모형이 개발되어 왔다. 이런 모형들은 일반적으로 시장의 크기 변화의 동인을 신기술 확산으로 보고 소비자인 개인에게 기술이 확산되는 과정을 통해 시장 크기가 변하는 과정을 확산모형으로 구현하게 된다. 그러나, 시장이 형성된 직후에는 수요 관측치의 부족으로 인해 혁신계수, 모방계수와 같은 예측모형의 모수를 정확하게 추정하는 것이 쉽지 않다. 이런 경우, 전문가의 판단 하에 예측하고자 하는 시장과 유사한 시장을 결정하고 이를 참고하여 모수를 추정하게 되는데, 어떤 시장을 유사하다고 판단하느냐에 따라 성장모형은 크게 달라지게 되므로, 정확한 예측을 위해서는 유사 시장을 찾는 것은 매우 중요하다. 그러나, 이런 방식은 직관과 경험이라는 정성적 판단에 크게 의존함으로써 일관성이 떨어질 수밖에 없으며, 결국, 만족할 만한 수준의 결과를 얻기 힘들다는 단점을 지닌다. 이런 정성적 방법은 유사도가 더 높은 시장을 누락시키고 유사도가 낮은 시장을 선택하는 오류를 일으킬 수 있다. 이런 이유로, 본 연구는 신규 시장의 모수를 추정하기 위해 필요한 유사시장을 누락 없이 효과적으로 찾아낼 수 있는 사례기반 전문가 시스템을 설계하고자 수행되었다. 제안된 모형은 데이터 마이닝의 군집분석 기법과 추천 시스템의 내용 기반 필터링 방법론을 기반으로 전문가 시스템으로 구현되었다. 본 연구에서 개발된 시스템의 유용성을 확인하고자 정보통신분야 시장의 모수를 추정하는 실험을 실시하였다. 전문가를 대상으로 실시된 실험에서, 시스템을 사용한 모수의 추정치가 시스템을 사용하지 않았을 때와 비교하여 실제 모수와 더 가까움을 보임으로써 시스템의 유용성을 증명하였다.

주제어 : 시장수요예측, 시장성장모형, 군집 분석 데이터 마이닝 기법, 내용 기반 필터링, 추천 시스템

논문접수일 : 2015년 9월 27일 논문수정일 : 2015년 12월 11일 게재확정일 : 2015년 12월 12일
교신저자 : 이동원

1. 서론

시장 수요 예측은 일정 기간 동안 소비자에게 판매되는 동종 제품 또는 서비스의 수량 혹은 매출액의 규모를 추정하는 활동이다. 새로운 제품의 도입시기 결정, 제품 설계, 생산계획 수립, 마

케팅 전략 수립 등에 활용됨으로써 경영활동에 있어 효율적인 의사결정을 내릴 수 있는 근거로 활용되기 때문에 정확한 시장 예측은 중요성이 크다고 할 수 있다(Hwang, 2012). 이런 이유로, 정보통신(Ahm et al., 2007; Kim et al., 2011), 전력(Nam et al., 2008; Lim et al., 2013; Song and

* 본 연구는 한국과학기술정보연구원(KISTI)의 지원을 받아 연구되었음.

Lim, 2013), 관광(Kim, 2014), 모바일기기(Noh et al., 2011) 등 다양한 분야에서 시장수요를 예측하기 위한 연구가 이루어져 왔다.

시장 수요 예측에 대한 기존의 연구는 주로 확산모형에 기초하여 성장곡선을 얻어내는 방식으로 진행되어 왔다. 즉, 시장의 크기 변화의 동인을 신기술 확산으로 보고 소비자인 개인에게 기술이 확산되는 과정을 통해 시장 크기가 변하는 과정을 모형으로 구현한 것이다. 구체적으로는, 각 산업별로 고유한 특징을 가정하고 이를 모형에 반영함으로써 좀 더 정교한 예측이 이루어질 수 있도록 연구되어 왔다. 이런 성장모형은 성장과 관련된 시장규모의 초기 관측치를 기반으로 이후의 성장곡선을 예측하는 방식으로 추정한다.

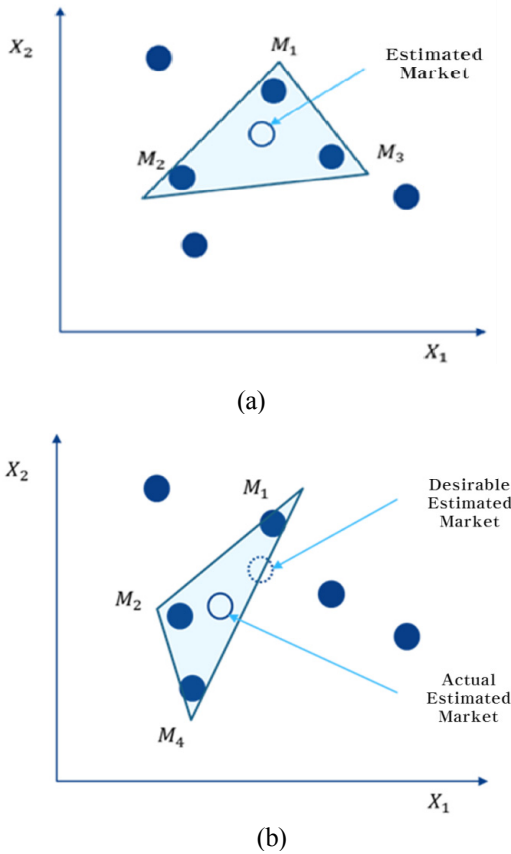
하지만, 이런 추정방식은 시장수요의 관측치가 충분한 경우 정교한 추정을 할 수 있는 반면, 그렇지 못한 상황에서는 미래 시장규모를 예측하는 데 있어 정확성이 낮은 단점을 갖는다 (Van den Bulte and Lilien, 1997; Srinivasan and Mason, 1986; Heeler and Hustad, 1980). 시장이 형성된 초기에는 누적된 관측치가 부족하여 시장수요의 예측이 쉽지 않기 때문에 기존의 모형을 활용하는 대신 정성적인 방식으로 접근하게 된다. 즉, 예측시장 또는 관련시장의 전문가 등의 판단에 의존하여 대략적인 수치를 추정하는 것이다. 이 때, 전문가는 예측시장과 유사한 시장을 선택하고 이 시장들의 수요변화를 참고하는 방식으로 예측시장의 미래 성장규모를 예측하게 된다(Hwang, 2012).

이런 방식은 예측자의 경험과 직관과 같은 주관적 기준에 의존함으로써 일관성이 떨어질 수 있다는 한계점을 지닌다. 즉, 서로 다른 시장에 대한 예측뿐만 아니라 동일한 시장에 대해 예측 시에도 서로 다른 결과를 나타낼 수 있다. 이는,

어떤 시장을 예측시장과 더 유사하다고 판단할 것인가를 결정하는 객관적 근거를 적용하지 못하기 때문이다. 따라서, 동일한 시장을 유사시장으로 선택하기도 하고, 때로는 그렇지 않다고 판단하게 되는데, 이는 시장 간 유사성에 대한 판단이 예측자의 주관적이고 정성적 기준에만 의존하여 내려지기 때문이다. 이로 인해, 예측자는 본인의 예측에 대해 확신을 갖지 못할 뿐 아니라, 예측의 일관성과, 나아가 정확성까지 낮추는 결과를 초래하게 된다.

예측시장의 모수를 추정하기 위해 선택되어야 할 유사시장이 누락되고, 유사성이 낮은 시장을 유사시장으로 대신 선택하는 경우, 모수의 예측 결과에 대한 정확성을 보장받기 어렵게 만든다. 고려해야 할 유사시장의 후보가 너무 많은 경우, 사용자는 이들을 모두 검토하는 것이 어려울 수 있기 때문에 이런 상황은 종종 발생할 수 있다. <Figure 1(a)>는 예측에 포함되어야 할 유사시장인 M_1, M_2, M_3 를 모두 포함한 경우의 예측시장의 위치를 보여주고 있다. 이에 반해 M_3 를 누락시키고 대신 M_4 가 선택된 경우에는 <Figure 1(b)>에서와 같이 예측결과가 기대한 값과는 차이를 보이게 됨을 알 수 있다.

따라서, 본 연구는 사례를 기반으로 시장의 수요를 예측할 때, 전문가가 유사시장을 선택하는 과정에서 오분류를 방지함으로써 일관성을 보장하고, 고려해야 할 유사시장 후보를 누락시키지 않도록 지원하며, 이로부터 예측시장의 모수를 추정함으로써, 예측의 정확성을 확보할 수 있는 시스템의 설계를 목표로 수행되었다. 오분류를 방지하기 위한 목적으로 적용되는 데이터 마이닝의 군집 분석 기법은 유사시장으로 선택된 시장들이 동질적인 특성을 지니는지를 평가함으로써 유사시장 선택의 일관성을 보장한다. 또한,



〈Figure 1〉 Adjacent Market Classification

추천 시스템의 내용 기반 필터링 기법은 고려할 대상이 되는 유사시장 후보의 범위를 줄여줌으로써 예측자가 누락 없이 유사시장에 대한 검토를 수행할 수 있도록 돕는다. 이에 대한 자세한 내용은 2장의 이론적 배경에서 다룬다.

2. 이론적 배경

2.1. Bass의 확장 로지스틱 모형

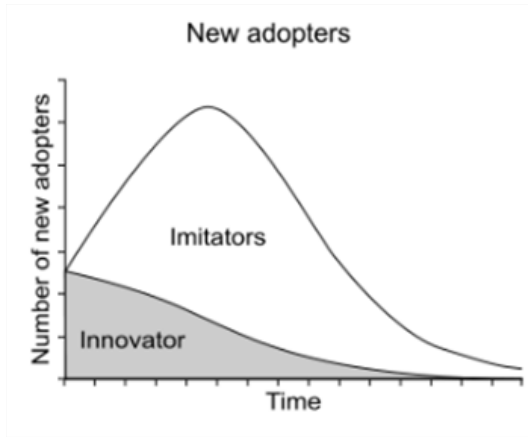
Bass, Logistic, Gompertz 모형과 같이 시장의

수요를 예측하기 위해 다양한 시장성장모형이 연구되어 왔다. 그 중에서 대표적인 모형인 Bass (1969)의 확장 로지스틱 곡선은 새로운 제품이나 서비스의 수요를 예측하는 데 가장 많이 쓰이는 모형이다. 이는 일반적으로 다른 성장곡선 모형보다 정확한 예측을 한다고 알려져 있기 때문이다(Noh et al., 2011). Bass가 소비재를 대상으로 처음으로 모형을 검증한 이후, 확장 로지스틱 곡선은 다양한 분야에 사용되고 있으며, 특히 전자 기기와 첨단 기술을 적용한 분야의 수요 성장을 예측하는 데 있어 널리 사용되고 있다.

Bass 모형은 신제품의 구매자를 혁신자만 고려한 수정지수곡선 모형(modified exponential curve; Woodstock et al., 1960)과 모방자만 고려한 로지스틱 성장곡선(simple logistic curve; Mansfield, 1961)을 결합한 형태이다. 혁신자(innovator) 그룹의 구매는 타구매자들의 결정과는 독립적으로 대중매체(외부영향)에 의해서만 영향을 받으며, 그 비중은 시간이 지남에 따라 감소한다. 모방자(imitator) 그룹은 다른 구매자들의 의사결정에 의해 영향을 받으며 전파과정은 구전(word-of-mouth)을 통해서 이루어진다. Bass 모형은 지수곡선과 로지스틱 곡선의 특성을 모두 지닌다.

Bass (1969)는 Fourt and Woodstock (1960)과 Mansfield (1961)의 모형을 종합하여, 혁신 구매자와 모방 구매자 모두 신제품의 확산에 기여한다고 제안하였다 (Figure 2참조). 즉, 신제품의 확산은 두 개의 집단에 의해 이루어지는데, 혁신 구매자의 구매의사결정은 외부 영향인 대중매체에 의존하는 반면, 모방구매자의 구매의사결정은 내부 영향인 구전효과를 통해 이루어진다고 주장하였다.

Bass의 확장 로지스틱 모형은 확률함수인 위



〈Figure 2〉 Extended Logistic Model

해함수(hazard function)로부터 유도된다. 여기서 위해율(hazard rate)이란, 신제품의 채택이 아직 일어나지 않았다고 주어진 경우 t시점에서 채택이 일어날 확률, 즉 t시점까지 비채택자로 남아 있던 개인이 다음 시점에서 채택자가 될 확률을 뜻한다. Bass 모형의 가정에 따라, 혁신 구매자의 채택은 이전의 채택과는 독립적인 사건인 반면, 모방 구매자의 채택은 이전의 채택에 비례하여 증가한다고 생각할 수 있다. 이를 정리하면 다음과 같다.

$$(f(t))/(1 - F(t)) = p + qF(t)$$

$$S(t) = p(m - N(t)) + q/m N(t)(m - N(t)) \quad (1)$$

여기서, N(t)는 t기까지의 누적 판매량 (t기까지의 누적 채택자의 수), S(t)는 t기의 판매량 (t기의 채택자의 수), F(t)는 t기까지의 누적 구매확률 (누적확률함수), f(t)는 t기의 구매확률 (확률밀도 함수), m은 제품의 잠재적 시장규모 (potential

market)를 의미한다.

이 식은 t시점에서의 구매확률 f(t)가 t시점까지 상품을 구매한 누적 구매확률에 비례하는 것을 나타낸다. 위 식에서 앞의 p(m-N(t))는 채택 시기에 기존 구매자에 의한 영향을 받지 않은 구매자, 즉 혁신 구매자에 의한 채택을 나타내며, q/m N(t)(m-N(t))는 구매의사결정이 기존 구매자에 의해 영향을 받은 구매자, 곧 모방 구매자에 의한 채택을 나타낸다. 위 식을 정리하면 다음과 같은 Bass의 확장 로지스틱 모형을 얻을 수 있다.

$$N(t) = m * (1 - exp(-(p + q)t)) / (1 + q/p exp(-(p + q)t)) \quad (2)$$

확장 로지스틱 곡선의 특징은 다음과 같다. (Lee and Lee, 2002) 첫째, 변곡점을 중심으로 좌우가 비대칭적이다. 즉, 확장 로지스틱 곡선은 좌우 대칭이라는 특성 때문에 시장상황을 반영하는 데 한계를 지닌 로지스틱 곡선의 단점을 보완할 수 있다. 둘째, 변곡점의 위치가 0과 잠재적 시장 규모의 절반인 m/2에 도달하는 시점 t 사이에 위치한다. 셋째, m의 규모를 알지 못해도 누적 수요인 N(t)만을 알고 있다면 모수 p, q를 직접 추정할 수 있다.

본 연구에서는 확장 로지스틱 모형을 대상으로, 모수인 p, q를 추정하기 위한 시스템의 개발을 목표로 이루어졌다. 시장규모를 나타내는 m 또한 중요한 모수임에는 분명하나, 기술의 전파와 확산 이외에도 소득과 같은 다양한 외부변수의 영향을 받기 때문에, 유사한 시장을 찾는 것만으로는 추정이 사실상 힘들다는 한계가 있다. 따라서, 본 연구에서 추정하는 대상에서는 제외

하기로 한다.

2.2. 데이터 마이닝의 군집 분석 기법

군집 분석은 분석 대상이 되는 레코드에 포함된 변수들에 기초하여 유사한 특성을 가진 레코드를 서로 같은 군집으로 구분하는 데이터 마이닝 기법이다(Jain and Dubes, 1988; Hartigan, 1975). 군집 분석은 다양한 분야에서 활용되고 있는데, 마케팅에서는 시장 세분화의 방법으로 활용되고 있다. 고객의 나이, 성별, 거주지역과 같은 인구통계학적 정보와 거래내역 특성을 기반으로 유사한 고객들로 군집을 형성함으로써, 각 군집에 대한 차별화된 마케팅 전략을 적용할 수 있도록 하기 위한 목적으로 활용되고 있다. 군집 분석은 군집을 형성하는 알고리즘에 따라 크게 2가지 유형인 계층적 방법(hierarchical methods)과 비계층적 방법(nonhierarchical methods)으로 나뉜다(Jain and Dubes, 1988; Hartigan, 1975).

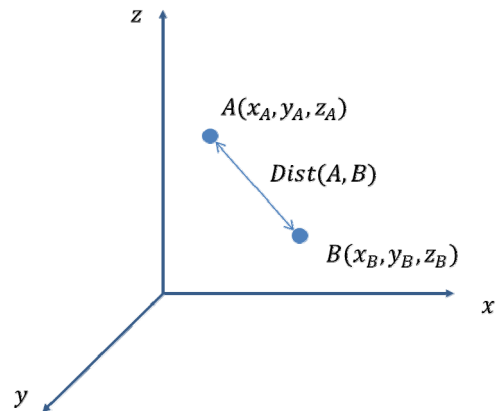
계층적 방법은 병합방식과 분할방식으로 나뉜다. 병합방식은 모든 레코드들을 각각 하나의 군집으로 보고 서로 가장 가까운 군집을 병합하는 과정을 반복하여 원하는 개수의 군집이 형성되면 중단하는 방법을 취한다. 반면, 분할방식은 모든 레코드를 하나의 통합된 군집으로 보고 시작하여 서로 특성이 가장 뚜렷이 구분되는 두 군집으로 나누는 과정을 반복한다.

비계층적 방법 중 k-평균(k-means) 기법은 k개의 군집수를 미리 정해두고, 각 레코드들을 정해진 군집에 할당하는 방식을 취한다. 동일한 군집에 속한 레코드들은 최대한 동질적 특성을 갖도록 하고 서로 다른 군집들 간에는 이질적 특성을 갖도록 하는 것이 이 방법의 최종 목표이다. 이

를 위하여 이미 하나의 군집에 속한 개별 레코드를 임의의 다른 군집으로 이동하는 과정을 반복함으로써 군집 내 이질성이 최소화되고, 군집 간 이질성이 최대화 되도록 한다.

군집 내에 속한 레코드들 간의 이질성을 계산하기 위해서는 이들 간의 유사성을 판단하는 척도가 필요한데, 두 레코드 사이의 거리가 일반적으로 많이 활용된다. 그 중에서도 가장 일반적으로 사용되는 거리척도는 유클리드 거리(Euclidean distance)로서 <Figure 3>에서와 같이 x, y, z라는 변수를 갖는 두 레코드 A, B 사이의 거리는 아래와 같은 수식으로 계산된다.

$$Dist(A, B) = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2 + (z_A - z_B)^2} \quad (3)$$



<Figure 3> Euclidean distance

일반적으로 군집 내의 이질성을 측정하는 분산은 이런 유클리드 거리의 제곱을 합한 값을 보편적으로 사용한다. 주어진 레코드들에 대한 군집 분석은 하나의 군집에 속한 각 레코드를 다른 군집으로 옮겨 군집 내 분산을 감소시키는 과정

을 반복하여 분산의 감소가 더 이상 없을 때까지 진행하게 된다. n 개의 관측치(O_1, O_2, \dots, O_n)를 포함하는 군집 C 의 중심이 O_c 인 경우, 분산은 아래와 같은 식으로 계산될 수 있다.

$$Var(C) = \sum_i^n Dist(O_i, O_c)^2 \quad (4)$$

군집화의 수준을 평가하기 위한 척도로 분산의 비율을 활용한다. 군집화가 이루어지기 전 모든 시장을 모두 하나의 군집에 속하는 것으로 보고 계산한 분산에 대해 군집화 이후에 계산된 각 군집의 분산의 합의 비율을 계산하는 것이다.

군집 분석 기법은 유사시장 분류의 일관성을 보장하기 위한 수단으로 활용될 수 있다. 즉, 유사시장으로 선택된 군집 내에서의 시장 간 동질성을 높이고, 선택되지 않은 시장과의 이질성을 높이는 방향으로 분류의 정확성을 높이는 방법이다. 우선, 모수를 변수로 갖는 서로 다른 두 시장 간의 유클리드 거리를 계산함으로써, 이들 간의 유사도를 판단할 수 있다. 즉, 거리가 가까울수록 서로 유사한 시장이라고 보는 것이다. 예측 시장과 유사한 시장을 선별하는 상황은, 주어진 모든 시장을 유사시장 군집과 비유사시장 군집 중 하나로 분류하는 문제로 간주할 수 있다. 즉, 예측 대상이 되는 시장과 가깝다고 선택된 유사 시장들을 하나의 군집으로, 그 외의 시장들을 또 다른 군집으로 분류하는 것이다. 유사시장과 기타시장의 군집에 대해 분류의 정확성을 더 높이기 위한 과정에서는, 각 군집의 분산을 더 낮추기 위한 군집 분석의 접근과 마찬가지로 방법을 적용할 수 있다. 즉, 유사시장 군집에 속한 시장을 기타 군집으로 옮기거나, 반대로 기타 군집의 시장을 유사시장 군집으로 옮기는 과정에서 군집

내의 분산이 낮아지도록 하는 것이다.

2.3. 내용 기반 필터링 추천 기법

추천 시스템이란 사용자가 자신이 원하는 상품이나 서비스를 찾을 수 있도록 돕기 위한 목적으로 개발된 시스템으로서, 사용자가 아직 사용하지 않은 아이템 중 사용자가 선호할 것으로 기대되는 것을 예상하여 제공하는 기능을 수행한다(Resnick and Varian, 1997). 1990년대 중반에 전자상거래의 급격한 발전과 함께 독립적인 연구분야로 각광을 받기 시작했다. 아마존(Amazon.com)에서는 하나의 상품 페이지를 보고 있으면 추가적인 아이템들이 현재 선택된 상품을 구입한 다른 고객들이 구매한 이력을 근거로 추천된다. 넷플릭스(Netflix)는 사용자가 이전에 관람한 영화에 대한 평점(rating)과 관람 습관을 근거로 보고 싶어할 만한 영화를 예측하는데, 영화의 장르와 같은 속성을 함께 이용한다(Montaner et al., 2003). 추천시스템은 다양한 분류가 있을 수 있으나 보편적으로 Balabanović et al. (1997)가 제시한 다음의 두 가지 종류로 분류하는 것이 일반적이다. 한 모형은 사용자의 구매 행동을 반영하는 협업 필터링(collaborative filtering) 기법을 활용하며, 또 다른 모형은 아이템의 속성(또는 특성)을 반영하는 내용 기반 필터링(content-based filtering) 기법을 적용한다.

협업 필터링 기법은 사용자들의 행동으로부터 사용자의 기호(preference)를 파악하는 방법을 사용하는데, 어떤 사용자와 유사한 기호를 지닌 사용자가 찾아지면 이 사용자가 호의적인 반응을 보인 아이템을 추천 대상으로 선정하는 방식이다. 즉, 특정 사용자에 대한 어떤 아이템의 추천 여부를 그 사용자와 유사한 다른 사람이 추천 후

보 아이템에 보인 반응을 근거로 예측하는 방법을 사용한다. 아이템에 대한 선호도는 (1) 사용자가 개별 아이템에 부여한 평점이나, 사용자가 작성한 선호 아이템 목록을 이용하는 방법을 사용하거나, (2) online store에서 사용자가 관심을 보인 상품과 그 상품 페이지에 방문한 횟수, 그 상품의 구매여부 등을 통해 간접적으로 선호 여부를 파악하는 방법을 활용하여 얻는다.

이와는 다르게 내용 기반 필터링 기법은 사용자가 선호하는 것으로 이미 확인된 아이템과 유사한 아이템을 찾아 이를 사용자에게 추천하는 방식이다. 아이템 간의 유사도는 각 아이템이 지닌 속성을 활용하는데, 사용자가 선호하는 것으로 파악된 아이템의 속성과 유사한 속성을 지닌 다른 아이템을 찾아 이를 추천하는 방식이다. 예를 들어, 영화를 어떤 사용자에게 추천하려고 하면, 그 사용자가 높게 평가한 영화들 간의 공통점을 찾으려고 한다. 이런 공통점은 그 영화의 속성 즉, 배우, 감독, 장르, 등등이 될 수 있다. 아직 사용자가 보지 않은 영화 중, 이런 속성에 있어서 높은 유사성을 보이는 것을 골라 추천하게 되는 것이다.

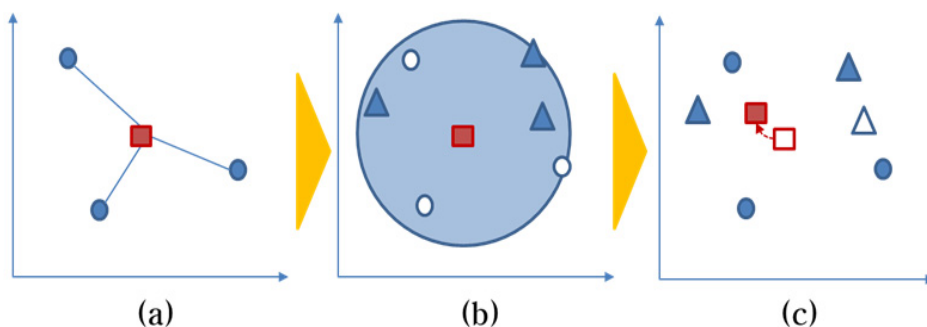
본 연구에서 예측시장과 유사한 시장을 선택하는 과정에서 내용 기반 필터링 기법을 활용할

수 있을 것이다. 시장을 추천 대상이 되는 상품으로, 모수를 이에 대한 속성으로 간주다면 속성 간의 유사성을 기반으로 유사시장을 선별하여 사용자에게 추천할 대상으로 결정하는 추천 시스템을 구현하는 상황으로 볼 수 있다.

상품의 추천 시스템에서 사용되는 내용 기반 필터링 기법은, 이미 소비자가 구매한 상품과 유사한 특성을 갖는 상품을 추천하게 된다. 먼저, 이미 구매한 상품들로부터 상품 프로파일을 작성하고 이 프로파일에 근거하여 이와 유사한 특성을 갖는 다른 상품을 찾아 이를 추천하는 방식이다.

<Figure 4(a)>에서와 같이 기존에 선택된 상품(세 개의 원)에 기반하여 상품 프로파일(사각형)을 작성하고 <Figure 4(b)>에서와 같이 이 프로파일로부터 속성의 특성이 일정한 거리 안에 인접한 다른 상품(세 개의 삼각형)을 추천 대상으로 선정한다. <Figure 4(c)>에서, 추천된 상품에 대해 사용자가 수용하여 구매가 이루어지면(속이 채워진 두 개의 삼각형), 다시 이를 반영하여 기존의 프로파일(속이 빈 사각형)을 새로운 프로파일(속이 채워진 사각형)로 수정하게 되며, 이로부터 추천 대상 상품을 선정하는 과정을 반복한다.

본 연구에서는, 시장을 상품으로, 시장의 모수



<Figure 4> Content-based filtering

를 상품의 속성으로 간주하면, 내용 기반 필터링 기법을 유사시장 후보의 선정에 활용할 수 있다. 즉, 사용자가 이미 선택한 유사시장들과 모수에 근거한 거리가 가깝지만 유사시장으로 분류되지 못한 시장들을 찾는다. 이렇게 발견된 시장을 유사시장 후보로 사용자에게 추천하면, 사용자는 검토해야 할 유사시장의 범위를 줄이고, 유사시장의 누락을 방지하는 효과를 기대할 수 있게 된다.

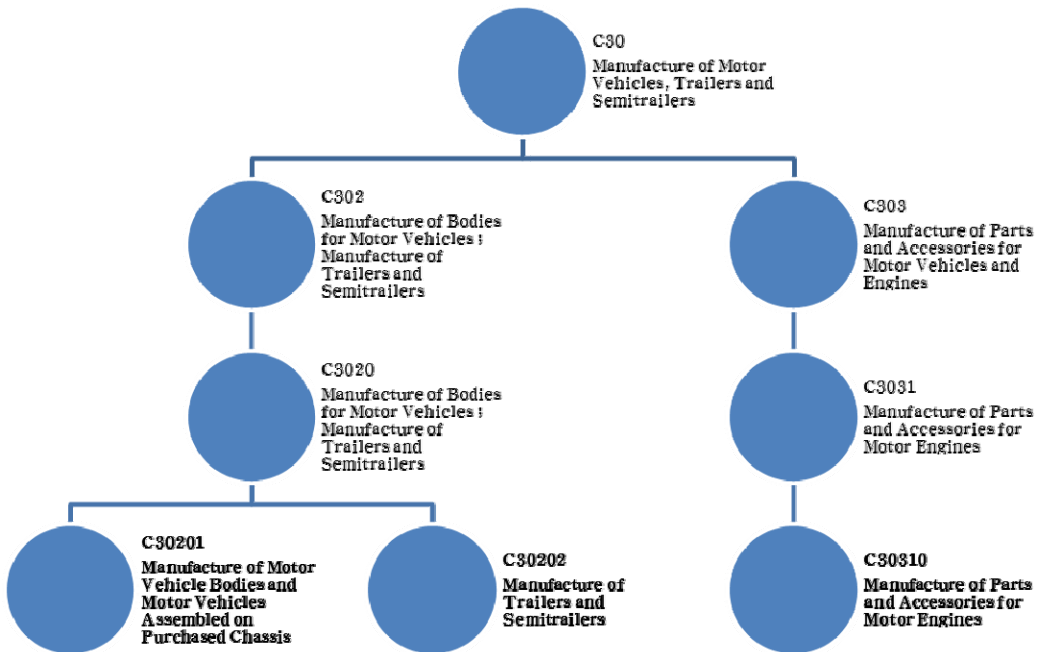
2.4. 산업분류코드에 기반한 거리 계산

시장의 속성 중에서 p, q 등의 모수는 비율적도(ratio scale)를 갖는 반면, 산업분류코드는 명목척도(nominal scale)를 갖는다. 모수를 이용하여 두 시장 간의 거리를 측정할 수 있으나, 산업분류코드를 유사도의 측정에 활용하기 위하여

두 시장 간 거리를 측정하는 데에 반영하기에는 한계가 있다.

산업분류코드는 계층구조를 갖는다는 점에 착안하여, 가계도에서 촌수를 계산하는 것과 같은 방식으로 두 시장 간의 거리를 측정하였다. 즉, 두 시장 간의 거리는 두 산업분류코드체계 내에서 두 시장이 속한 산업을 잇는 경로에 놓인 링크의 개수로 정의한다.

이렇게 계산된 산업분류코드 간 거리는 군집 분석 과정에서 여전히 한계점을 지닌다. 분석 과정 중 하나의 시장을 군집에 포함하고 제외시키는 과정에서 시장과 군집 간의 거리를 계산할 수 있어야 하는데, 일반적인 방법인 군집의 중심으로부터 시장까지의 거리를 계산할 수 없다는 것이다. 이는 군집 중심을 하나의 대표 시장의 산업코드로 설정할 수 없다는 점에서 기인한다. 다른 모수(p, q)의 경우, 평균과 같은 대푯값을 계

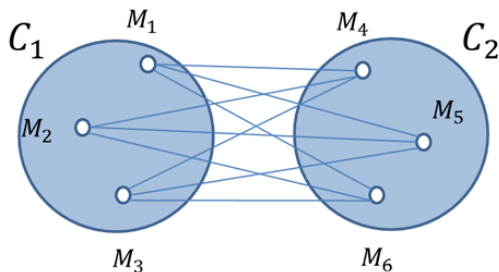


(Figure 5) Industry-Classification-Code-based Distance

산하여 중심시장의 속성으로 사용할 수 있다. 그러나, 명목변수인 코드값의 평균은 중심으로서의 의미를 지닐 수 없기 때문이다. 예를 들어, $A(p_A, q_A)$ 과 $B(p_B, q_B)$ 의 평균인 $C((p_A+p_B)/2, (q_A+q_B)/2)$ 는 두 시장 간 중심에 놓인 대표시장으로 볼 수 있으나, 두 개의 산업코드 '2100'과 '2300'의 평균인 '2200'은 중심 시장의 산업분류 코드로 볼 수 없다는 것이다.

이런 이유로 군집의 중심에 대한 코드값을 활용하지 않는 군집 간 거리의 계산방법이 필요하다. 군집 분석에서 두 군집 간 거리의 계산에 활용되는 척도인 평균거리를 활용하는 것으로 해결한다. 이는 하나의 군집 내에 속한 모든 관측치와 다른 군집에 속한 모든 관측치들 간에 존재하는 모든 가능한 연결 간의 거리를 구한 후 이들의 평균을 계산한 값이다. <Figure 6>에서 두 군집 C_1, C_2 에 속한 모든 관측치 간에 계산 가능한 모든 거리는 아래의 식과 같이 계산될 수 있다. 여기서, $n(C_1), n(C_2)$ 는 각각 C_1, C_2 군집에 속한 시장의 수를 의미한다.

$$Avg(Dist(C_1, C_2)) = \frac{\sum_i \sum_j Dist(M_{1,i}, M_{2,j})}{n(C_1) \times n(C_2)} \quad (5)$$



<Figure 6> Average Distance between two Clusters

두 군집에 대한 거리를 위와 같이 정의하면, 하나의 시장이 어떤 군집과 떨어진 거리도 계산할 수 있다. 이는 해당 시장을 하나의 시장만 포함된 군집으로 간주하고 계산함으로써 가능하다.

3. 유사시장 선정 모형

3.1. 시스템 기능

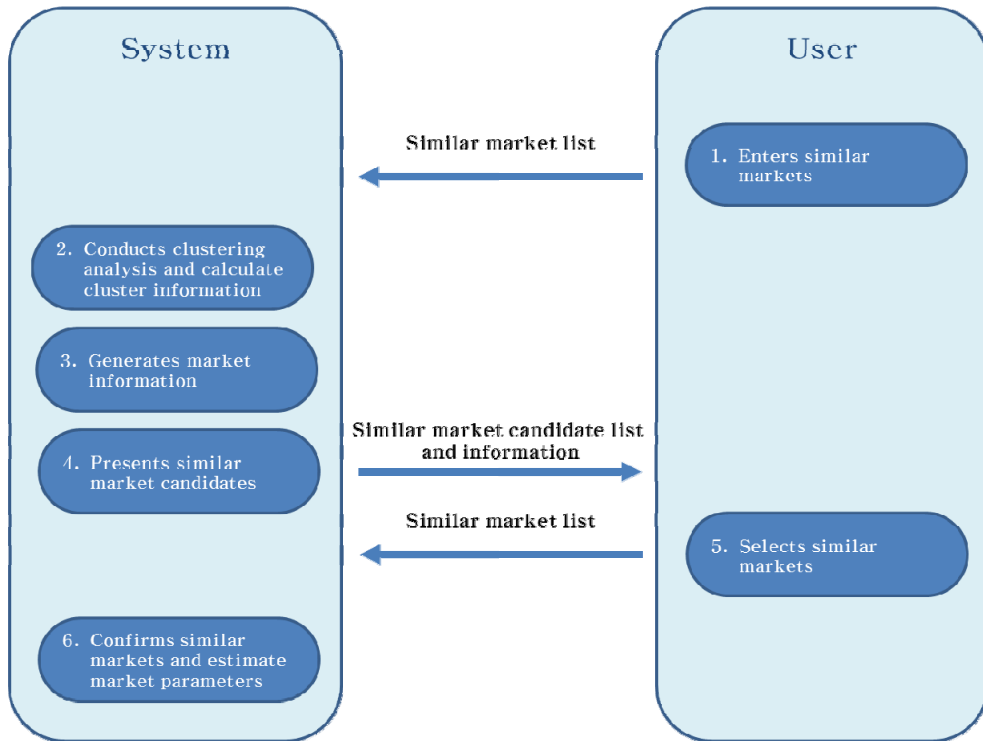
제안된 시스템의 기능은 크게 군집 분석과 추천으로 나뉜다.

첫째, 군집 분석은 유사시장 분류의 일관성을 보장하기 위한 기능이다. 모든 시장은 유사시장 또는 비유사시장으로 분류되는데, 각 군집에 속한 시장들 간의 동질성은 커지고, 서로 다른 군집에 속한 시장들 간에는 이질성이 커지는 방향으로 군집화가 이루어지도록 한다.

둘째, 추천 기능은 유사시장 군집에 속한 시장들의 속성들로부터 유사시장 군집의 프로파일 정보를 생성하고, 이로부터 가장 인접한 시장을 찾아, 사용자에게 제시한다. 사용자의 판단에 따라 유사시장으로 분류된 새로운 시장을 포함하고, 비유사시장으로 재분류된 기존 유사시장을 제외시킨 후 유사시장 군집의 프로파일을 다시 생성함으로써, 새로운 추천 후보를 선정하는 과정을 반복한다.

3.2. 프로세스

본 연구에서 제안하는 추정 프로세스는 <Figure 7>과 같다. 시스템이 제공하는 유사시장 후보에 대해 사용자의 선택여부에 따라 유사시장 군집을 재조정하고, 재조정된 군집에 기반하

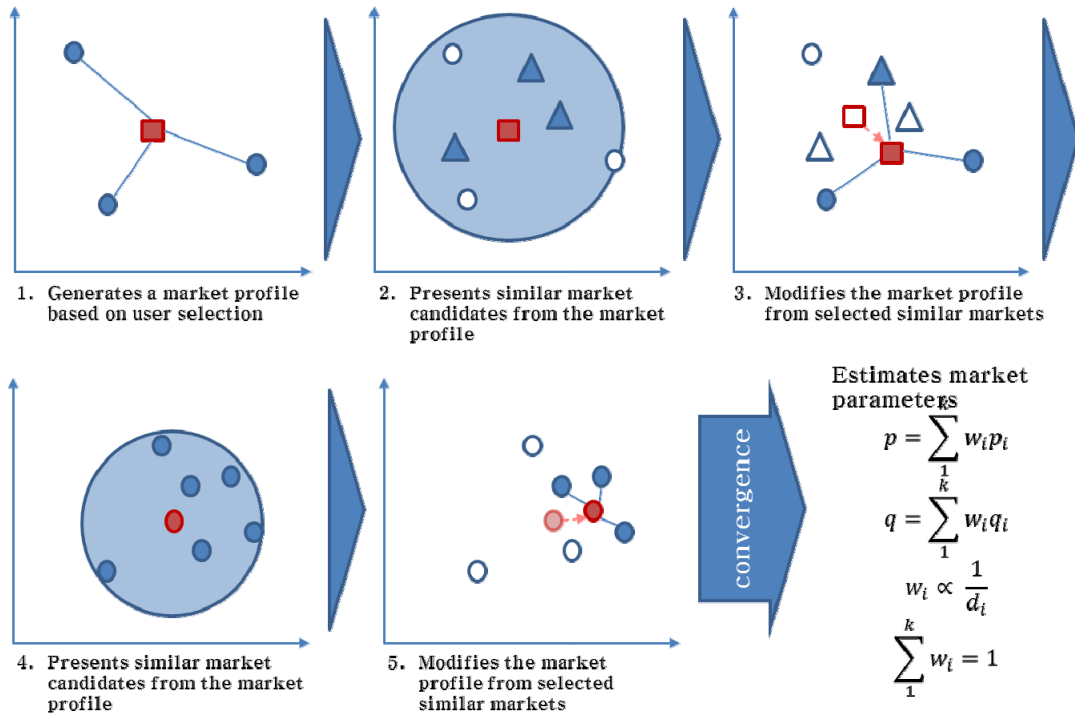


〈Figure 7〉 Market Parameter Estimation Process

여 새로운 유사시장 후보를 제시하는 과정이 반복적으로 이루어진다.

우선, 1단계에서 사용자는 예측시장과 가깝다고 판단되는 유사시장들을 시스템에 입력한다. 다음으로, 2단계에서, 시스템은 사용자로부터 입력된 유사시장들을 이용하여 두 개의 군집을 형성한다. 즉, 사용자에게 의해 입력된 시장을 유사시장 군집, 그 외의 시장을 비유사시장 군집으로 분류한다. 그리고, 이렇게 형성된 두 개의 군집에 대해 군집 관련된 정보로서, 개별 군집의 중심값과 군집 내 분산을 각각 계산한다. 3단계에서는, 이렇게 형성된 군집을 기반으로 이들 군집에 속한 시장의 군집화 정보를 생성한다. 즉, 각 시장으로부터 유사시장 군집과 비유사시장 군집까지의 유클리드 거리를 모수에 기반하여 계산

하고, 산업분류코드에 기반하여 유사시장 군집과의 거리를 계산한다. 네 번째로, 앞에서 계산된 거리를 근거로 유사시장의 후보를 선정하고 선정된 유사시장후보를 거리 및 군집 관련 정보와 함께 사용자에게 제시한다. 5단계에서, 사용자는 유사시장을 재조정하여, 새로이 유사시장 군집을 구성할 시장을 선택한다. 2단계에서와 마찬가지로, 선택되지 않은 시장들은 비유사시장 군집으로 분류된다. 이 때, 사용자는 시스템으로부터 제공 받은 시장-군집 간 거리뿐만 아니라, 하나의 시장을 유사시장 군집이나 비유사시장 군집 중의 하나로 포함시킬 때 각 군집 내의 분산과 군집 간의 분산이 어떻게 변화하는지를 확인함으로써, 하나의 시장이 유사시장과 비유사시장 중 어느 쪽에 속하는 것이 더 적합한지에



<Figure 8> Market Parameter Estimation

대한 객관적인 지표로 활용할 수 있다. 이 단계에서 사용자의 선택에 의해 유사시장/비유사시장 군집에 속한 시장이 변하게 되면, 다시 2단계로 돌아가 이후 과정을 수행하는 것이 반복되지만, 더 이상 수정을 하지 않는 경우에는 6단계로 넘어가 유사시장 군집을 확정하고 이 군집에 속한 시장들로부터 예측시장의 모수를 추정함으로써 모든 프로세스는 종료된다.

이런 과정을 일률적으로 자동화하지 않고 사용자와의 상호작용을 해야 하는 이유는 다음과 같다. 첫째, 최초 사용자가 입력한 유사시장에 크게 의존하게 되기 때문이다. 사용자가 어느 한 쪽으로 편향된 선택을 한 경우 이런 편향이 최종 결정에까지 크게 영향을 미치게 된다. 하지만, 상호작용을 하게 되면, 최초의 선택이 한쪽으로

편향되었다고 하더라도, 이후 수정 과정에서 이를 보완할 수 있다. 둘째, 모수나 분류코드에 근거한 거리만으로는 유사시장을 확정할 수 없기 때문이다. 우연히 서로 유사한 성장곡선을 갖는 경우도 있을 것이며, 서로 다른 분류코드 상에서도 서로 유사한 시장이 존재할 수도 있다는 점 때문이다. 따라서, 단계별로 제공되는 시장과 군집 관련 정보를 보고 사용자가 결정하는 과정이 필요하다고 판단하였다. 유사시장의 선정 과정은 <Figure 8>과 같다.

4. 실험

4.1. 실험 목적

실험은 본 연구에서 제안한 기법의 유효성을 검증하기 위한 목적으로 실시되었다. 즉, 시스템을 사용함으로써, 예측시장과 유사한 시장들을 더 잘 선별할 수 있는지, 그리고 이렇게 선별된 유사시장들로부터 추정된 모수는 실제 모수에 더 근접한지를 평가하기 위한 것이다.

본 연구에서는 유사시장 선택의 일관성을 높이기 위해 군집 분석 기법과 내용 기반 필터링 기법을 적용하였는데, 이로부터 제공되는 군집 및 시장 관련 정보가 사용자에게 실질적인 도움이 되었는지를 살펴보기 위한 실험을 실시했다.

4.2. 실험 방법

정보통신분야의 전문가 5명을 대상으로 주어진 16개의 예측시장을 제시한 후, 본 연구를 통해 설계된 시스템을 사용하여 유사시장을 선별하도록 했다.

본 실험은 3장에서 소개된 프로세스에 따라 진행되었으며, 시스템은 사용자에게 군집과 시장 관련 정보를 제공하고, 사용자는 이런 정보들을 활용하여 유사시장을 선별하는 과정을 반복한 후, 최종적인 유사시장 군집을 결정하고 예측시장의 모수를 추정하였다.

본 실험은 동일한 사용자에게 대해 시스템을 사용하기 전후의 개선된 정도를 비교함으로써 시스템의 유효성을 평가하였다. 즉, 사용자가 초기에 입력한 유사시장들로부터 추정된 모수에 비해, 시스템 사용 후 최종적으로 추정된 모수가 얼마나 예측시장의 실제 모수와 더 가까운 값을 갖는지를 비교하였다.

구체적인 방법은 다음과 같다. 우선, 정보통신 분야에서 성장곡선의 모수가 이미 알려진 16개의 시장을 예측시장으로 선정한다. 다음으로, 5

명의 해당 분야 전문가에게 모수값을 알려주지 않은 채 예측시장과 가장 유사한 3개의 시장을 선택하게 한 후, 시스템을 활용하여 3.2절의 프로세스에 따라 유사시장을 선택하도록 한다. 이때, 최초에 선택된 유사시장들로부터 추정된 예측시장의 모수와 시스템을 사용하여 선택된 유사시장으로부터 예측시장의 모수를 각각 추정한다. 각 모수와 실제 예측시장의 모수를 비교함으로써 유사도를 측정한다. 측정된 유사도를 비교함으로써 유사도의 개선도를 계산한다.

4.3. 실험 결과

<Table 1>은 5명의 실험으로부터 수집된 모수 거리의 개선도를 각 예측시장에 대해 정리한 것이다. 실험에 참여한 전문가에 따라 다소 차이가 있긴 하지만, 모두 시스템 사용 전보다 사용 후에 예측시장의 모수에 가까운 모수를 추정했음을 보이고 있으며, 향상도는 평균 12.6%에 이른다. 동일한 시장에 대해 실험자마다 서로 큰 편차를 보이는데, 이는 개별 시장에 대한 지식의 차이에 크게 기인한 것으로 판단된다.

<Table 2>는 시스템 사용 전후의 분산지수 개선도를 보여주고 있다. 시스템 사용 전에 사용자가 선택한 유사시장을 기준으로 형성된 두 개의 군집에 대한 분산지수와 시스템 사용 후에 형성된 군집의 분산지수의 변화율을 계산한 것이다. 대부분의 실험에서 주어진 예측시장에 대해 분산지표의 개선도는 양(+)의 값을 보이는데, 이는 분산지표의 제공이 유사시장 선정에 실질적인 도움을 준 것으로 이해될 수 있다.

<Table 1> Improvement of Market Parameter Distances

Prediction Market	Improvement Rate of Market Parameter Distances (%)				
	Exp. 1	Exp. 2	Exp. 3	Exp. 4	Exp. 5
Key telecommunications service switch	9.4	14.4	12.8	11.3	7.4
Private switch	62.1	69.7	20.1	2.5	21.8
Photo transmission systems	-6.7	5.9	3.7	20.6	2.7
Hand-held device	-12.5	-11.7	11.9	1.3	19.7
Wireless telecommunications switch	-9.5	15.5	5.1	5.1	15.5
Terrestrial broadcasting transmitter-receiver	4.0	4.0	4.8	4.1	2.5
Wired broadcasting transmitter-receiver	0.0	1.6	1.2	1.6	0.1
Digital camera	-9.6	-9.6	4.3	1.3	3.9
DVD player	0.0	52.7	50.1	39.3	41.4
CCTV camera	-19.4	2.1	0.6	0.6	17.0
Notebook PC	7.3	5.6	13.6	7.3	5.8
D RAM	0.0	2.6	6.1	3.7	1.2
Flash memory	0.0	11.5	28.3	20.4	8.8
Kimchi refrigerator	0.0	3.0	2.3	0.9	5.2
Air conditioner	68.2	0.0	17.5	87.4	34.5
Vacuum cleaner	29.3	45.5	45.5	28.6	17.5
Average	7.7	13.3	14.2	14.8	12.8

<Table 2> Improvement of Variance Indices

Prediction Market	Improvement Rate of Market Parameter Distances (%)				
	Exp. 1	Exp. 2	Exp. 3	Exp. 4	Exp. 5
Key telecommunications service switch	1.8	3.4	2.7	3.3	1.0
Private switch	1.1	1.3	0.3	0.4	1.2
Photo transmission systems	1.1	0.6	0.2	-0.8	0.3
Hand-held device	6.9	10.4	2.0	8.0	0.1
Wireless telecommunications switch	0.5	0.0	0.9	0.9	0.2
Terrestrial broadcasting transmitter-receiver	0.2	0.2	1.2	0.6	0.9
Wired broadcasting transmitter-receiver	0.0	0.2	1.1	0.2	0.8
Digital camera	0.1	-0.2	0.0	-0.3	0.1
DVD player	-0.3	0.1	0.2	-0.1	0.2
CCTV camera	0.1	0.3	0.6	0.6	0.3
Notebook PC	0.4	1.6	2.0	0.9	0.5
D RAM	0.0	0.4	0.4	0.0	0.2
Flash memory	0.2	0.1	0.5	-0.1	0.4
Kimchi refrigerator	0.0	0.1	0.1	0.2	0.2
Air conditioner	0.3	0.0	0.1	0.3	0.1
Vacuum cleaner	0.2	0.5	0.5	0.2	0.1
Average	0.8	1.2	0.8	0.9	0.4

5. 결론

본 연구는 초기단계에 있는 시장의 미래 수요를 예측하기 위한 모형을 개발하기 위해 수행되었다. 기업의 입장에서 시장의 초기단계에는 향후 투자의 규모를 결정하는 등 장기적인 관점에서의 의사결정에 영향을 미친다는 점에서, 시장의 미래 수요에 대한 정확한 예측은 중요하다. 정부의 입장에서든 해당 산업을 육성하기 위한 정책을 결정하는 등에 중요한 지표로 활용될 수 있다. 시장의 미래수요 예측이 중요함에도 불구하고, 시장의 초기에는 수요예측이 쉽지 않다. 수요를 예측하기 위해 주로 이용되는 시장수요 곡선모형은 수요 데이터가 일정 수준 이상으로 확보되었을 때 정확성을 보장 받을 수 있기 때문이다. 따라서, 시장의 초기에 미래 수요를 예측할 수 있는 대안이 필요하다. 실무에서는 이런 대안으로서, 이미 성장곡선이 예측된 유사 시장을 참고하게 된다. 예측하고자 하는 시장과 유사한 기술을 사용하는 기존 시장의 수요변화를 통해 신규 시장의 수요를 예측해보는 것이다. 하지만, 이런 과정은 관련 시장 전문가의 경험과 직관과 같은 주관적 판단에 의존함으로써, 예측의 일관성이 떨어지는 단점을 갖고 있다. 이는 결국 예측 정확성의 저하로 이어질 뿐 아니라, 반복되는 예측 과정이 지식으로 축적되지 못하고 계속 직관에 의존하는 상황이 이어지게 된다.

이는 예측 시장에 대한 유사시장을 선택할 때에 포함되어야 할 유사시장을 누락시키거나, 유사성에 대한 객관적 지표가 부족하다는 점 등이 그 원인으로 파악되고 있다. 결국 두 시장의 유사도는 두 시장의 성장곡선의 유사도에 의해 결정될 것이며, 이는 성장곡선의 모양을 결정하는 두 모수인 혁신계수와 모방계수에 의해 결정될

것이다. 따라서, 유사시장의 선택은 예측시장과 모수의 값이 가까운 시장을 찾는 것을 목표로 수행하게 된다. 하지만, 예측시장의 모수가 미지수인 상황에서 이와 가까운 값을 직접적으로 찾을 수는 없다.

이런 이유로, 대안적인 방법을 적용하였다. 즉, 유사시장을 선택하는 사용자의 일관성을 높이는 것이다. 사용자가 예측시장에 대한 지식을 갖춘 경우, 유사시장의 후보를 선정하는 것이 가능할 것이다. 이렇게 선정된 시장 중에는 유사도가 높은 시장과 그렇지 않은 시장이 포함되어 있을 것이다. 사용자의 지식이 충분하다면 유사시장의 비율은 더 높을 것이고 이런 시장들과 그렇지 않은 시장들 간에는 유사도가 낮을 것이다. 군집 분석 기법은 이렇게 사용자가 선택한 유사시장 군집에 포함된 시장들 중 서로 유사도가 높은 시장들과 그렇지 않은 시장들을 구분함으로써 유사성이 높은 시장들끼리 동일한 군집에 남도록 하는 것을 목표로 한다. 또, 한 가지 기능으로써 추천 기법을 적용하게 되는데, 이는 사용자가 누락시킨 유사시장의 후보를 찾아나감으로써 유사 시장 선택의 폭을 넓히는 것을 목표로 한다. 본 연구는, 두 가지 기능이 조화를 이루게 되어 예측시장과 가장 유사한 시장을 찾고 이로부터 성장곡선의 모수를 추정해내려는 방향으로 수행되었다.

실무적인 관점에서, 본 연구는 전문가의 정성적 판단과 정량적 기준을 결합한 수요예측이 이루어질 수 있도록 했다는 점에서 기여했다고 할 수 있다. 즉, 사용자의 도메인 지식 중 정량화할 수 있는 영역을 계량화함으로써 수요예측의 정확성을 높이는 데에 기여했다.

학문적인 관점에서는, 명목형 변수인 산업분류코드를 이용한 시장 간 거리를 계산할 수 있는

방안을 제시했다. 이는 유사하게 계층적 구조를 갖는 관측치들 간에 적용할 수 있을 것이다. 추천 시스템의 관점에서 사용자에게 유사시장 후보를 제안한다는 점과 군집 분석 기법을 적용하여 유사시장 선정의 유효성을 검증하려는 새로운 시도를 했다는 점에서 높게 평가할만하다.

다만, 이렇게 제시된 지표는 사용자의 선택이 일관성이 있다는 것을 보여주는 지표이며, 선택의 결과가 올바른 모수를 추정할 수 있다는 것을 보장하지는 않는다. 시장성장곡선의 모수를 올바르게 추정하기 위해서는 사용자의 도메인 지식이 적절히 활용되어야 한다.

본 연구에서는, 사용자의 정성적 판단기준을 정량화함으로써 일관성을 확보하고, 유사시장 누락의 한계를 극복하고자 더 개선된 예측모형을 제시하고자 하였으나, 다음과 같은 한계점을 지니고 있다. 첫째, 시장수요곡선의 모수 중 시장규모(m)를 추정하지 않는다. 이는 이미 밝힌 바와 같이 해당 모수가 유사한 다른 시장으로부터 도출되기에는 더 고려해야 할 외부변수들이 많기 때문이다. 둘째, 사용자의 초기 선택이 최종 결과에 큰 영향을 미칠 수 있다. 이는 초기에 유사시장 중에 누락된 시장을 찾기 위한 추천 기법을 통해 찾는 방법으로 보완되기는 하였으나, 이 또한 사용자의 입력에 근거하고 있기 때문에 한계가 존재한다. 셋째, 관련분야의 전문가를 충분히 확보하기 어려운 한계점으로 인해, 실험에 참여한 전문가의 수가 상대적으로 부족했다는 점을 지적할 수 있다. 마지막으로, 군집 분석 기법을 활용함으로써 유사시장 간의 동질성에 근거하여 분류결과를 평가하려 하였으나, 이런 동질성이 분류의 정확성을 항상 보장하는 것은 아니라는 점이다. 예를 들어, 휴대전화, 디지털 카메라, PDA 등 여러 시장과 유사성을 가진 스마

트폰의 경우, 이들 시장 간에는 동질성이 낮아 하나의 군집으로 묶이기 쉽지 않을 것이다.

향후에는 위에 언급한 한계점을 극복하기 위한 연구가 추가적으로 진행되어야 할 것으로 판단된다. 본 연구는, 유사시장의 수요가 보이는 성장 패턴과 관련한 지표 및 산업 고유의 특성에 기반하여 새로운 시장의 수요 성장을 예측하는 방향으로 진행되어, 시장의 소비자의 성별, 연령, 직업 등에 관한 정보는 활용되지 않았으나, 향후 연구에서는 이런 추가적인 정보를 적용한다면, 스마트폰 시장과 같이 이질적인 특성을 보이는 여러 시장과 유사성을 보이는 신규시장의 유사 시장 발견에 더 좋은 성과를 기대해볼 수도 있을 것이다.

참고문헌(References)

- Ahn, C. and S. Kang, "An Analysis of the Cross Relation between The New Telecommunications Services and Demand Forecast based on Use Pattern of Consumers," *Proceedings of the Korean Institute of Industrial Engineers*, (2007), 585~592.
- Balabanović, M. and Y. Shoham, "Fab: content-based, collaborative recommendation," *Communications of the ACM*, Vol.40, No.3 (1997), 66~72.
- Bass, F. M., "A New Product Growth Model for Consumer Durables," *Management Science*, Vol.15, No.5(1969), 215-227.
- Fourt, L. A. and J. W. Woodstock, "Early Prediction of Market Success of New Grocery Products," *Journal of Marketing*, Vol.25, No.2(1960), No.2, 31~38.

- Hartigan, J. A., *Clustering algorithms*, John Wiley & Sons, Inc., 1975.
- Heeler, R. M. and T. P. Hustad, "Problems in predicting new product growth for consumer durables," *Management Science*, Vol.26, No.10 (1980), 1007~1020.
- Hwang, Y., "A Hybrid Forecasting Framework based on Case-based Reasoning and Artificial Neural Network," *Journal of Intelligence and Information Systems*, Vol.18, No.4(2012), 43~57.
- Nam, B.-W., K.-B. Song, K.-H. Kim, and J.-M. Cha, "The Spatial Electric Load Forecasting Algorithm using the Multiple Regression Analysis Method," *Journal of the Korean Institute of Illuminating and Electrical Installation Engineers*, Vol.22, No.2(2008), 63~70.
- Jain, A. K., and R. C. Dubes, *Algorithms for clustering data*, Prentice-Hall, Inc., 1988.
- Kim, J.-H., K.-H. Hong, and J.-Y. Min, "A Real-Time Stock Market Prediction Using Knowledge Accumulation," *Journal of Intelligence and Information Systems*, Vol.17, No.4(2011), 109~130.
- Kim, S.-T., "The Study on Tourism Demand Forecasting: Focused on Panel Data Analysis Method," *Journal of Tourism and Leisure Research*, Vol.26, No.1(2014), 115~129.
- Lee, G. and Lee, C., "Comparison Study of Demand Forecasting Techniques using Growth Curve Models," *Sogang Journal of Business*, Vol.13, No.2(2002), 195~228.
- Lim, J.-H., S.-Y. Kim, J.-D. Park, and K.-B. Song, "Representative Temperature Assessment for Improvement of Short-Term Load Forecasting Accuracy," *Journal of the Korean Institute of Illuminating and Electrical Installation Engineers*, Vol.27, No.6(2013), 39~43.
- Mansfield, E., "Technical Change and the Rate of Imitation," *Econometrica: Journal of the Econometric Society*, (1961), 741~766.
- Montaner, M., B. López, and J. L. De La Rosa, "A taxonomy of recommender agents on the internet," *Artificial intelligence review*, Vol.19, No.4(2003), 285~330.
- Noh, K., S. Sim, S. Hong, and B. Jeong, "A study on hybrid demand forecasting for mobile phone," *Proceedings of the 2011 Spring KIIE/KORS Joint Conference*, (2011), 1222~1228.
- Resnick, P. and H. R. Varian, "Recommender systems," *Communications of the ACM*, Vol.40, No.3 (1997), 56~58.
- Song, K.-B. and J.-H. Lim, "Short-Term Load Forecasting for the Consecutive Holidays Considering Businesses' Operation Rates of Industries," *The Transaction of the Korean Institute of Electrical Engineers*, Vol.62, No.12(2013), 1657~1660.
- Srinivasan, V. and C. H. Mason, "Technical note-nonlinear least squares estimation of new product diffusion models," *Marketing science*, Vol.5, No.2(1986), 169~178.
- Van den Bulte, C. and G. L. Lilien, "Bias and systematic change in the parameter estimates of macro-level diffusion models," *Marketing Science*, Vol.16, No.4(1997), 338~353.
- Woodstock, L. W. and F. Skoog, "Relationships between growth rates and nucleic acid contents in the roots of inbred lines of corn," *American Journal of Botany*, Vol.47, No.9(1960), 713~716.

Abstract

An Expert System for the Estimation of the Growth Curve Parameters of New Markets

Dongwon Lee* · Yeojin Jung** · Jaekwon Jung** · Dohyung Park***

Demand forecasting is the activity of estimating the quantity of a product or service that consumers will purchase for a certain period of time. Developing precise forecasting models are considered important since corporates can make strategic decisions on new markets based on future demand estimated by the models. Many studies have developed market growth curve models, such as Bass, Logistic, Gompertz models, which estimate future demand when a market is in its early stage. Among the models, Bass model, which explains the demand from two types of adopters, innovators and imitators, has been widely used in forecasting.

Such models require sufficient demand observations to ensure qualified results. In the beginning of a new market, however, observations are not sufficient for the models to precisely estimate the market's future demand. For this reason, as an alternative, demands guessed from those of most adjacent markets are often used as references in such cases. Reference markets can be those whose products are developed with the same categorical technologies. A market's demand may be expected to have the similar pattern with that of a reference market in case the adoption pattern of a product in the market is determined mainly by the technology related to the product.

However, such processes may not always ensure pleasing results because the similarity between markets depends on intuition and/or experience. There are two major drawbacks that human experts cannot effectively handle in this approach. One is the abundance of candidate reference markets to consider, and the other is the difficulty in calculating the similarity between markets. First, there can be too many markets to consider in selecting reference markets. Mostly, markets in the same category in an industrial hierarchy can be reference markets because they are usually based on the similar technologies. However, markets can be classified into different categories even if they are based on the same generic technologies.

* Corresponding author: Dongwon Lee

School of Business Administration, Hansung University

116 Samseongyoro-16gil, Seongbuk-gu, Seoul 136-792, Korea

Tel: +82-2-760-4250, Fax: +82-2-760-4482, E-mail: dongwonlee@hansung.ac.kr

** School of Business Administration, Kookmin University

*** School of MIS, Kookmin University

Therefore, markets in other categories also need to be considered as potential candidates. Next, even domain experts cannot consistently calculate the similarity between markets with their own qualitative standards. The inconsistency implies missing adjacent reference markets, which may lead to the imprecise estimation of future demand. Even though there are no missing reference markets, the new market's parameters can be hardly estimated from the reference markets without quantitative standards.

For this reason, this study proposes a case-based expert system that helps experts overcome the drawbacks in discovering referential markets. First, this study proposes the use of Euclidean distance measure to calculate the similarity between markets. Based on their similarities, markets are grouped into clusters. Then, missing markets with the characteristics of the cluster are searched for. Potential candidate reference markets are extracted and recommended to users. After the iteration of these steps, definite reference markets are determined according to the user's selection among those candidates. Then, finally, the new market's parameters are estimated from the reference markets. For this procedure, two techniques are used in the model. One is clustering data mining technique, and the other content-based filtering of recommender systems. The proposed system implemented with those techniques can determine the most adjacent markets based on whether a user accepts candidate markets.

Experiments were conducted to validate the usefulness of the system with five ICT experts involved. In the experiments, the experts were given the list of 16 ICT markets whose parameters to be estimated. For each of the markets, the experts estimated its parameters of growth curve models with intuition at first, and then with the system. The comparison of the experiments results show that the estimated parameters are closer when they use the system in comparison with the results when they guessed them without the system.

Key Words : market demand forecast, market growth curve, clustering data mining technique, content-based filtering, recommender system

Received : September 27, 2015 Revised : December 11, 2015 Accepted : December 12, 2015

Corresponding Author : Dongwon Lee

저자 소개



이동원

LG CNS에서 시스템 엔지니어로 근무하였으며, KAIST 경영대학원에서 MIS 전공으로 석사/박사 학위를 취득하였다. 현재 한성대학교 경영학부 조교수로 재직 중이다. 현재 빅데이터에 기반한 연구를 주로 수행하고 있으며, 주요 관심분야는 고객관계관리, 추천 시스템, 데이터 마이닝 기법의 정교화, 디지털 콘텐츠 마케팅 등이다.



정여진

연세대학교에서 경제학 및 응용통계학 복수전공으로 학사를 취득하였으며, 동 대학원에서 응용통계학 석사학위를 취득한 후 The Pennsylvania State University에서 통계학 박사 학위를 취득하였다. 현재 국민대학교 경영대학 경영학부 조교수로 재직중이다. 주요 연구분야는 비모수분포추정, 모형기반 군집분석, hierarchical linear model, 일반화선형모형이다.



정재권

University of Liverpool 에서 경영학 석사/ 박사학위를 취득하였고, 현재 국민대학교 경영대학 경영학부 조교수(마케팅전공)로 재직 중이다. (주)세렌디피컴퍼니의 대표(제작자)로 재직하였으며(2012년 3월 ~ 10월), 현재 주요 관심분야는 신선식품 가격전략, 동태적 가격전략, 조절초점이론, SNS나 온라인 구전 등의 사용자 행동 이론, 문화예술마케팅 등이다.



박도형

KAIST 경영대학원에서 MIS 전공으로 석사/ 박사학위를 취득하였다. 현재 국민대학교 경영대학 경영정보학부 조교수로 재직 중이며, 한국과학기술정보연구원(KISTI)에서 유망아이템 발굴, 기술가치 평가 및 로드맵 수립, 빅데이터 분석 등을 수행하였고, LG전자에서 통계, 시선/뇌파 분석, 데이터 마이닝 등 활용한 연구 및 소비자 평가 모형 개발을 담당하였고, 스마트폰, 스마트TV, 스마트Car 등에 대한 Technology, Business, Market Insight 기반 컨셉 도출 프로젝트를 다수 수행하였다. 현재 주요 관심분야는 SNS나 온라인 구전 등의 사용자 행동 이론(User Behavior), 사용자 경험 디자인 프로세스 및 혁신 제품 발굴(User eXperience), 빅데이터 기반 사용자 분석(User Analytics) 등이다.