

고객정보와 상품네트워크 유사도를 이용한 시장세분화 기법

정석봉 · 신용호 · 구서룡 · 윤협상*

A Market Segmentation Scheme Based on Customer Information and QAP Correlation between Product Networks

Seok-Bong Jeong · Yong Ho Shin · Seo Ryong Koo · Hyoup-Sang Yoon*

ABSTRACT

In recent, hybrid market segmentation techniques have been widely adopted, which conduct segmentation using both general variables and transaction based variables. However, the limitation of the techniques is to generate incorrect results for market segmentation even though its methodology and concept are easy to apply. In this paper, we propose a novel scheme to overcome this limitation of the hybrid techniques and to take an advantage of product information obtained by customer's transaction data. In this scheme, we first divide a whole market into several unit segments based on the general variables and then agglomerate the unit segments with higher QAP correlations. Each product network represents for purchasing patterns of its corresponding segment, thus, comparisons of QAP correlation between product networks of each segment can be a good measure to compare similarities between each segment. A case study has been conducted to validate the proposed scheme. The results show that our scheme effectively works for Internet shopping malls.

Key words : Market segmentation, Product network, Hierarchical agglomerative clustering, QAP correlation

요약

시장세분화를 위해 일반변수와 트랜잭션 기반 변수를 동시에 사용하는 하이브리드 방법이 널리 사용되고 있지만, 하이브리드 방법에는 일반변수의 기준에 따라 정확하게 세분화가 되지 않는 문제점이 존재한다. 본 연구에서는 이러한 문제점을 해결함과 동시에 상품 정보를 이용한 네트워크 분석을 활용하는 새로운 시장세분화 방법을 개발하는 것을 목표로 한다. QAP 상관관계 분석을 이용하여 상품네트워크의 유사도를 계산하는 새로운 시장세분화 방법은 일반 변수 기준으로 시장을 명확하게 세분화하고, 상품 정보를 기반으로 하여 세분화된 집단 간의 구매패턴을 효과적으로 비교할 수 있도록 하는 장점을 갖고 있다. 본 연구를 통해 개발된 상품구매정보를 활용한 네트워크 기반 시장세분화 방법의 활용 가능성과 성과를 입증하기 위해 실제 운영 중인 온라인 쇼핑몰의 고객정보와 상품구매정보를 수집하여 시장세분화 방법의 절차를 설명하고 결과를 제시한다. 본 연구에서 제안된 시장세분화방법은 기본적인 고객정보 및 상품구매정보를 이용하여 상품구매패턴이 유사한 고객 집단을 인구통계학적인 일반변수 기준으로 세분화할 수 있기 때문에 대다수의 온오프라인 유통업체에서 폭넓은 활용이 가능할 것으로 기대된다.

주요어 : 시장세분화, 상품네트워크, 계층적병합군집화기법, QAP 상관관계

1. 서론

* 이 논문은 2015년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2015S1A5A8012533).

Received: 23 November 2015, **Revised:** 9 December 2015, **Accepted:** 10 December 2015

***Corresponding Author:** Hyoup-Sang Yoon
E-mail: hsyoon@cu.ac.kr
Dept. of MIS, Catholic University of Daegu

시장세분화는 기업이 고객에 대한 이해도를 더욱 높임으로써 새로운 고객을 유인하거나 기존 고객의 만족도를 높이기 위한 마케팅 분야의 핵심적 개념이다. 시장을 세분화하기 위한 기준이 되는 세분화 변수를 선정하는 것은 성공적인 시장세분화를 위해 대단히 중요하다. 세분화 변

수는 크게 일반변수(general variable)와 트랜잭션 기반 변수(transaction based variable)로 구분할 수 있다^{1, 2)}.

일반변수를 사용한 시장세분화 연구는 고객의 인구통계적 특성, 생활습관, 태도, 심리와 같은 기초적인 고객의 속성을 이용하여 고객을 구분하고 이들의 대상으로 마케팅을 차별화하여 진행할 수 있도록 한다³⁻⁶⁾. 일반변수를 기반으로 하는 시장세분은 직관적이고 쉽게 적용 가능하다는 장점이 있는 반면에 유사한 일반변수의 특성을 갖는 고객이 유사한 구매 행태를 나타내지 않을 수 있으며, 데이터를 수집하기 어렵다는 단점이 존재한다^{1, 7)}.

반면 트랜잭션 기반 변수는 고객의 구매 행위, 고객의 잠재가치, life time value 등과 같은 고객 가치 지향적인 속성을 포함한다. 트랜잭션 기반 변수를 사용하는 시장세분화 방법은 기업의 이익을 높이기 위해 통상적으로 고객 가치에 초점을 맞추고 있다. 그러나, 고객 유지 관점에서 보면 이러한 방법은 기업이 고객의 취향이나 기호에 대해 파악하기 어렵고 상품에 관한 중요한 정보가 누락되어 있어 장기적으로 기업과 고객의 관계를 확고하게 하기 어렵다는 단점이 존재한다.

일반 변수와 트랜잭션 기반 변수를 사용하는 시장세분화 방법의 장점을 모두 취하기 위해 하이브리드 방법이 제안되었다⁸⁾. 하이브리드 방법은 트랜잭션 변수를 이용하여 고객을 군집화 하고 각 군집에 대해 일반변수를 이용하여 판별분석을 수행한다. 이러한 하이브리드 방법의 가장 큰 문제점은 판별분석의 결과로 도출되는 세분 시장의 분류 기준이 일반 변수를 기준으로 정확하게 나뉘지 않아 실질적 활용에 한계가 있다는 것이다.

본 연구는 하이브리드 시장세분화 방법의 문제점을 해결함과 동시에 상품 네트워크 분석을 활용하는 새로운 시장세분화 방법을 개발하는 것을 목표로 한다. 즉, 일반 변수 기준으로 시장을 명확하게 세분화하는 동시에 상품 구매 정보를 활용하여 고객 집단 간의 구매패턴을 비교할 수 있는 효과적인 분석 방법을 제안한다.

본 논문의 구성은 다음과 같다. 먼저, 제 2장에서는 상품 구매의 유사성을 네트워크 분석 기법을 사용하여 표현하고 이를 이용하여 시장을 세분화하는 방법을 제안한다. 다음으로, 제 3장에서는 실제 운영 중인 온라인 쇼핑몰의 데이터를 분석하여 본 연구에서 제안하는 상품네트워크 분석 기반의 시장세분화 방법의 절차를 설명하고 결과를 분석한다. 마지막으로 결론과 향후연구 내용에 대해서는 제 4장에서 다루고자 한다.

2. 상품네트워크분석 기반의 시장세분화

시장 세분화에 대한 다양한 정의가 존재하지만, 효과적인 시장세분화의 기준으로 널리 받아들여지고 있는 기준은 다음과 같다⁹⁾.

- 기준1) 동일 세그먼트 내의 고객 간에는 동질성이 높아야 한다(homogeneity within segment).
- 기준2) 서로 다른 세그먼트에 속한 고객 간의 동질성은 낮아야 한다(heterogeneity between segments).
- 기준3) 하나의 세그먼트의 크기는 이익이 발생할 만큼 충분히 커야 한다(substantial).

본 연구에서는 위의 기준에서 ‘동질성’에 대한 개념을 ‘구매 패턴의 유사성’으로 간주하고 이러한 세 가지 기준에 부합하는 시장세분화 기법을 제안하고자 한다.

본 연구에서 제안하는 시장세분화 기법의 전체 프레임워크는 Fig. 1과 같이 크게 4 단계로 구성된다.

단계 1은 시장세분화를 위한 고객변수의 선택 및 선택된 변수가 갖는 속성 값의 상태를 정의하는 단계이다. 단계 2에서는 선택된 고객변수 및 각 변수의 상태 정의에 따라 고객 세그먼트를 구성하고, 각 세그먼트 별 상품 네트워크를 구성한다. 단계 3에서는 단계 2에서 구성된 고객 세그먼트 간의 구매패턴의 유사성을 상품네트워크를 이용하여 비교하고, 군집화 알고리즘을 이용하여 각 고객

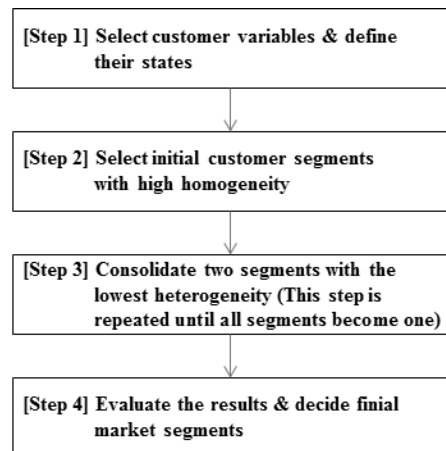


Fig. 1. Framework of the proposed scheme

세그먼트를 통합한다. 마지막으로 단계 4는 통합의 결과 평가를 통해 세분화 하고자 하는 시장의 개수를 정하고 최종적인 시장세분화 기준을 도출하는 단계이다.

한편, 제안된 기법은 다음과 같은 점에서 효과적 시장 세분화의 기준에 부합할 수 있다. 먼저 ‘세그먼트 내 높은 동질성’ 기준을 고려하기 위하여, 단계 1에서는 가능한 많은 고객변수를 도입하고 각 변수의 상태를 세밀히 정의 하여 고객 세그먼트를 최대한 동질적인 속성을 갖는 작은 그룹으로 세분화 한다. 둘째, ‘세그먼트 간 낮은 동질성’ 기준을 충족하기 위하여 단계 3에서는 군집화 알고리즘을 이용하여 구매패턴의 유사성이 높은 고객 세그먼트부터 순차적으로 통합을 진행한다. 마지막으로 ‘세그먼트의 크기’ 기준을 위하여 단계 4에서 각 세그먼트가 적절한 크기를 가질 수 있도록 세분화된 시장의 개수를 정하게 된다.

각 단계별 상세 설명은 다음 절부터 제시되어 있다.

2.1 단계 1: 고객변수 선정 및 각 변수의 상태 정의

제안된 시장세분화 기법의 첫 번째 단계는 고객변수 선택 및 각 변수의 상태를 정의하는 단계이다. 일반적으로 인터넷 쇼핑물에서 고객변수는 고객의 가입정보를 통해서 획득할 수 있는데, 대표적으로 성별, 직업, 나이, 결혼유무, 전화번호, 이메일 주소 등의 인구통계학 정보가 활용 가능하다. 단계 1에서는 이러한 고객정보 중에서 시장세분화에 활용할 고객변수들을 선정한다.

고객변수의 선정 후에는 각 변수가 갖게 되는 속성 값의 상태를 정의한다. 본 연구에서는 각 고객변수는 범주형 변수로 가정하며, 만약 연속 값을 갖는 변수일 경우 해당 변수를 범주형 변수로 변환한다. 예를 들어, ‘성별’과 ‘나이’, 두 개의 고객변수만이 존재하는 상황을 가정해보자. 이 경우 고객변수 ‘성별’은 ‘남자’와 ‘여자’의 두 개의 상태를 갖게 되며, 고객변수 ‘나이’는 ‘20세 미만’, ‘20~50세’, ‘50세 이상’ 등 몇 개의 범주로 그 상태를 정의할 수 있다.

한편, 기존의 연구에서는 서로 다른 고객변수의 선정은 서로 다른 시장 세분화 결과를 가져올 수 있기 때문에 고객변수의 선정은 그 자체로도 매우 중요한 문제였다^{11, 7)}. 그러나 제안된 기법에서는 무의미한 변수가 고객변수로 포함되거나 특정 고객변수의 상태가 초기에 의미 없이 너무 세분화되어 정의되더라도, 전체 프레임워크의 진행 과정 중 의미 없는 변수나 상태는 통합되므로 최종적인 시장세분화 기준에는 영향을 미치지 않게 된다. 본 연구에서는 ‘세그먼트 내 높은 동질성’을 위하여 이 단계에서 가능한 많은 고객변수를 사용하고, 각 변수의 상태를 상세

히 정의할 것을 권고한다.

2.2 단계 2: 고객 세그먼트 및 상품네트워크 구성

제안된 기법의 두 번째 단계는 고객변수 및 각 변수의 상태 정의에 따라 전체 고객을 다수의 고객 세그먼트로 구분하고, 각 세그먼트 별 상품네트워크를 구성하는 단계이다. 본 연구에서 고객 세그먼트는 각 고객변수 별로 모두 동일한 속성 값을 갖는 고객들의 집합을 의미한다. 앞에서 살펴본 예와 같이 ‘성별’과 ‘나이’ 등 두 개의 고객변수가 존재하고, 각 변수의 상태가 이전과 같이 각각 2개 및 3개로 정의된 상황에서는 총 6개의 고객 세그먼트가 구성될 수 있다(남자 & 20세 미만, 남자 & 20~50세, 남자 & 50세 이상, 여자 & 20세 미만, 여자 & 20~50세, 여자 & 50세 이상 등).

고객변수와 각 변수의 상태 정의에 따라 고객 세그먼트를 구분한 이후, 다음으로 각 세그먼트 별 고객들의 구매내역을 바탕으로 상품네트워크를 작성한다. 상품네트워크를 구성하는 방법은 몇 가지가 존재하지만 본 연구에서는 기존 문헌에서 가장 많이 활용되는 ‘연결구매’ 상품네트워크를 이용한다¹⁰⁻¹²⁾. 연결구매란 두 종류 이상의 상품이 동일 고객에 의해 함께 구매되는 것을 의미하며, 연결구매 상품네트워크란 연결구매가 발생한 상품 간에 연결(link)을 유지하고, 그 빈도를 가중치로 표현한 네트워크이다.

연결구매 상품네트워크는 링크에 가중치가 있는 가중네트워크로써, 이 가중치는 연결된 두 상품을 동시에 구매한 고객의 수를 의미한다. 즉, 가중치 w 로 연결된 두 상품은 w 명의 고객에 의해 함께 구매된 ‘연결상품’이 되며, 이 가중치가 클수록 두 상품의 연결의 강도가 높다고 할 수 있다.

상품네트워크에서 상품 간의 연결 형태(연결의 유무나 강도)는 해당 고객 세그먼트의 구매패턴을 반영할 수 있다^{10, 13)}. 본 연구에서는 상품네트워크의 연결 형태가 해당 고객 세그먼트의 구매패턴을 반영한다는 관찰에 의거하여, 단계 3에서는 각 세그먼트 간 상품네트워크의 연결 형태를 비교함으로써 유사 고객 세그먼트 간의 통합 작업을 수행한다.

2.3 단계 3: 네트워크 유사성 비교를 통한 고객 세그먼트 통합

제안된 기법의 세 번째 단계는 상품네트워크의 유사성 비교를 통하여 구매패턴이 유사한 고객 세그먼트들을 통합하는 단계이다. 이 단계는 두 과정으로 진행되는데, (1)

각 고객 세그먼트 간 구매패턴의 유사도를 계산하는 과정과, (2) 유사도에 따라 고객 세그먼트들을 통합하는 과정이다.

첫 번째 과정에서 우리는 각 고객 세그먼트의 상품네트워크 간 유사도를 QAP(Quadratic Assignment Procedure) 상관계수를 이용하여 계산한다. QAP 상관계수는 Pearson 상관계수를 네트워크를 대상으로 확대 적용한 것으로, 동일한 개체들로 구성된 두 네트워크에서 개체들 간의 연결이 두 네트워크워크에 걸쳐 서로 어느 정도 관련이 있는지 확인하는 방법이다¹⁴⁾. 이 과정을 통해 우리는 각 고객 세그먼트 간에 구매패턴의 유사성을 파악할 수 있다. 부록에는 본 연구의 사례분석으로 실시한 23개 고객 세그먼트 간의 QAP 상관계수가 나타나 있다.

단계 3의 두 번째 과정은 각 고객 세그먼트간의 유사도를 기반으로 유사도가 높은 세그먼트들을 하나로 통합하는 과정이다. 이 과정은 효과적 시장 세분화의 기준 중 하나인 세분 시장의 크기를 키우기 위해 필요하다. 본 연구에서는 계층적 병합 군집화(hierarchical agglomerative clustering) 기법을 사용하여 세그먼트 간의 통합을 수행한다. 계층적 병합 군집화 방법은 관측치(고객 세그먼트)들 간의 거리를 바탕으로 단일 군집으로부터 출발하여 연결방법에 따라 두 개의 군집을 하나로 병합하는 과정을 전체 관측치들이 하나의 군집으로 묶여 질 때까지 반복적으로 수행하는 방법이다¹⁶⁾. 고객 세그먼트 i 와 j 의 QAP 상관계수를 γ_{ij} 라고 하면, 두 세그먼트의 거리 d_{ij} 는 식 (1)과 같이 비유사성(dissimilarity)으로 정의할 수 있다. 이는 유사성이 낮을수록 두 세그먼트 간의 거리가 멀어짐을 의미한다.

$$d_{ij} = 1 - \gamma_{ij}^2 \quad (1)$$

인접한 두 군집을 병합하기 위해 두 군집 간의 거리를 계산할 때는 시장세분화 기법에서 일반적으로 사용하는 완전연결법을 이용하였다.

2.4 단계 4: 결과평가 및 시장세분화 기준 도출

제안된 기법의 마지막 단계는 단계 3의 군집화 결과를 바탕으로 세분화할 시장의 개수를 정하고 각 시장의 속성을 파악하여 일반적인 시장세분화 기준을 도출하는 단계이다.

단계 3의 계층적 병합 군집화 과정을 거치면서 군집의 수는 고객 세그먼트의 개수로부터 한 개까지 순차적으로 감소한다. 이러한 단계별 통합 과정은 계층형 트리 형태

의 덴드로그램(dendrogram)을 통해 확인할 수 있다.

시장세분화를 결정하는 의사결정자는 덴드로그램을 통해서 적절한 수준의 군집의 수를 최종적으로 선택해야 한다. 이때 선택하는 군집의 수가 최종적인 세분 시장의 개수가 된다. 적절한 군집의 개수 선택에 대한 일부 연구가 존재하지만, 해당 내용은 본 연구의 핵심을 벗어나므로 논의를 생략하도록 한다. 단, 너무 많은 수의 군집을 선택하면, 하나의 세분 시장에 포함되는 고객의 수가 적어지므로 효과적인 시장 세분화 기준의 ‘세그먼트 크기’에 대한 기준에 부합하지 않게 된다. 반면, 너무 작은 수의 군집을 선택할 경우 하나의 세분 시장 내에 다양한 속성을 지니는 고객들이 함께 포함되므로 ‘세그먼트 내 높은 동일성’ 기준을 만족시키지 못하게 된다.

적절한 시장의 개수를 선택하고 나면, 각 세분 시장에 포함된 고객 세그먼트를 확인할 수 있다. 이때, 하나의 세분 시장에 포함된 고객 세그먼트들의 고객변수 별 속성 값이 해당 세분시장을 정의하는 ‘시장 속성’이 되며, 이는 전체 시장을 세분화 하는 시장세분화의 기준이 된다.

한편, 고객변수 중 일부는 시장세분화 기준에서 사라질 수도 있다. 예를 들어, 앞의 예제에서 고객변수로 ‘성별’ 및 ‘나이’를 선택하여 총 6개의 고객 세그먼트를 구성했다라도, 군집과정을 거치면서 최종 세분 시장은 성별에 상관없이 나이만으로 구분되어 질 수도 있다.

이렇듯, 시장의 개수가 선정되면 각 시장에 포함된 고객 세그먼트들의 고객 변수 및 그 속성을 이용하여 해당 시장의 세분화 기준을 도출할 수 있다.

3. 사례연구

본 연구에서 제안하는 시장세분화 기법의 타당성 및 활용 가능성을 검증하기 위하여 실제 운영중인 D인터넷 쇼핑몰을 대상으로 사례 분석을 실시하였다. D인터넷 쇼핑몰은 연 매출 10억 원 정도에 3만 5천여 명의 회원을 보유하고 있는 국내 중형의 쇼핑몰이다. D인터넷 쇼핑몰의 취급 상품은 의류, 명품, 화장품, 생활·주방·가구, 스포츠, 식품, 잡화·제화 등으로, 일반적인 백화점에서 취급하는 거의 모든 제품을 다루는 종합쇼핑몰이다. 본 장에서는 D인터넷 쇼핑몰의 거래내역 데이터를 기반으로 제안된 기법에 따른 시장 세분화를 시도해 본다.

3.1 분석 데이터

본 분석의 구매 트랜잭션에 포함된 상품은 Fig. 2에 표현된 바와 같이 총 3계층의 상품 분류체계에 따라 분류된

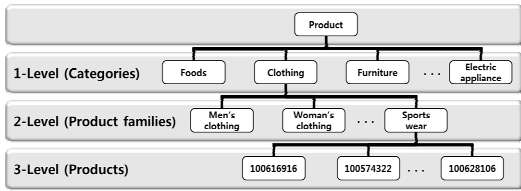


Fig. 2. Hierarchical product taxonomy

다. D인터넷 쇼핑물의 상품분류 체계에 따르면, 각 상품은 제 1계층에서 12개의 분류(Categories)로 구분되며, 제 2계층에서 49개의 상품군(Product families), 제 3계층에서는 53,381개의 개별상품(Products)으로 구분된다.

본 연구에서는 상품네트워크의 노드를 구성하는 단위로 3계층의 개별상품 대신에 제 2계층의 상품군을 사용한다. 개별 상품을 노드로 사용할 경우 네트워크 내 노드 수의 증가로 인해 소프트웨어의 분석 수행에 한계가 있으며, 또한 네트워크 내의 노드가 많아지면 노드 간의 연결이 상대적으로 적어져서 희박성(sparsity) 문제가 야기되기 때문이다¹⁷⁾. 이후 논문에서 사용되는 상품이라는 용어는 제 2계층의 상품군을 의미한다.

사례분석에 사용한 구매 데이터 18개월의 구매기간 동안 수집되었으며 구매 이력이 존재하는 고객은 총 6,942명이고, 구매 건수(트랜잭션 수)는 9,699건이며, 49개의 상품이 존재한다. 한편, 고객 1인당 구매 횟수의 평균이 1.40회(9699/6942)에 그쳐, 고객 충성도가 높지 않은 인터넷 쇼핑물의 특성을 확인할 수 있었다.

3.2 분석결과

먼저, 제안된 기법의 단계 1의 수행을 위해 D인터넷 쇼핑물의 고객 회원가입정보로부터 추출 가능한 ‘age’, ‘gender’, ‘region’ 등 총 3개의 변수를 고객변수로 선정하였다. 또한 각 고객 변수 별로 age는 8개의 상태(0: 20대 미만, 1: 20-24세, 2: 25-29세, 3: 30-34세, 4: 35-39세, 5: 40-44세, 6: 45-49세, 7: 50세 이상), gender는 2개의 상태(0:남자, 1:여자), region은 2개의 상태(0: 수도권, 1: 비수도권)로 정의하였다.

단계 2에서는 단계 1에서 선정된 고객변수 및 상태 정의에 따라 총 32개의 고객 세그먼트가 구성되었으며, 각 세그먼트 별 고객의 분포는 Table 1에 표기되어 있다.

한편, 32개의 고객 세그먼트 중 고객 수가 너무 적어 (100명 이하) 특정 세그먼트를 대표하기 어렵다고 판단되는 집단은 분석 대상에서 제외하여 총 23개의 고객 세그먼트를 대상으로 상품네트워크를 구성하였고 다음 단계

Table 1. Customer distribution

Seg. No.	Customer variables			No. of customers	Avail.
	age	gen.	reg.		
1	0	0	0	20	NA
2	0	0	1	52	NA
3	0	1	0	49	NA
4	0	1	1	121	
5	1	0	0	73	NA
6	1	0	1	197	
7	1	1	0	160	
8	1	1	1	431	
9	2	0	0	142	
10	2	0	1	288	
11	2	1	0	260	
12	2	1	1	544	
13	3	0	0	170	
14	3	0	1	288	
15	3	1	0	306	
16	3	1	1	710	
17	4	0	0	107	
18	4	0	1	221	
19	4	1	0	220	
20	4	1	1	631	
21	5	0	0	83	NA
22	5	0	1	146	
23	5	1	0	246	
24	5	1	1	518	
25	6	0	0	43	NA
26	6	0	1	99	NA
27	6	1	0	110	
28	6	1	1	304	
29	7	0	0	35	NA
30	7	0	1	114	
31	7	1	0	87	NA
32	7	1	1	167	

의 분석을 진행하였다.

단계 3에서는 23개의 고객 세그먼트의 상품네트워크 간 QAP 상관계수를 계산하였고(부록 참고), 이를 식 (1)과 같이 각 세그먼트 간의 거리로 변환하여 완전연결법의 방법을 이용한 계층적 병합 군집화를 수행하였다. 군집화

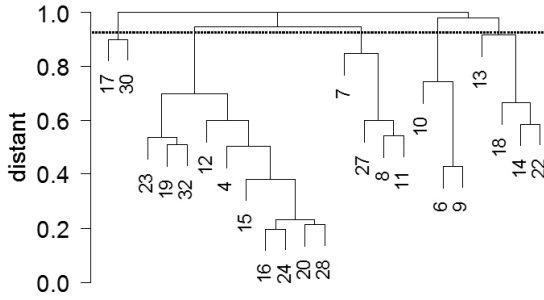


Fig. 3. Dendrogram

Table 2. Results of market segmentations

	Segment	Market attributes			# of Cust.
		A	G	R	
M1	17, 30	4	0	0	221
		7	0	1	
M2	4, 12, 15, 16, 19, 20, 23, 24, 28, 32	0,2	1	1	3767
		3,4,5	1	all	
		6,7	1	1	
M3	7, 8, 11, 27	1	1	0,1	961
		2	1	0	
M4	6, 9, 10	1	0	1	627
		2	0	all	
M5	13, 14, 18, 22	3	0	all	825
		4,5	0	1	

과정을 보여주는 덴드로그램은 Fig. 5에 나타나있다.

Fig. 3 및 부록 A에 따르면, 상품네트워크 간의 유사도가 0.90으로 가장 높은 16번과 24번 고객 세그먼트가 가장 먼저 통합되며, 이어서 유사도가 0.89인 20번과 28번 세그먼트가 통합된다. 이러한 통합 과정이 순차적으로 진행되면서 최종적으로 모든 고객 세그먼트는 하나의 군집으로 통합된다.

마지막으로 단계 4에서 의사결정자는 덴드로그램을 통해 세분화할 시장의 개수를 정해야 한다. 적절한 군집의 개수 선택에 대한 몇몇 이론이 존재하지만, 해당 내용은 본 연구의 핵심을 벗어나므로 논의를 생략하도록 하며, 본 연구에서는 Fig. 3의 덴드로그램에서 거리 약 0.95를 기준으로 고객을 총 5개의 군집(세분 시장)으로 분류하였다. 이에 따른 최종 시장세분화 결과는 Table 2에 제시되어 있다.

Table 2에 따르면 전체 시장은 남성이 3개의 세분 시장(M1, M4, M5), 여성이 2개의 세분 시장(M2, M3)으로 분류되었다. 각 시장에 포함된 고객 세그먼트의 속성이 해당 세분 시장의 속성이 되는데, 첫 번째 시장(M1)의 속성은 수도권에 거주하는 30대 초반 남성 및 비수도권에 거주하는 50대 이상 남성으로 정의된다.

두 번째 시장(M2)은 비수도권에 거주하는 20대 미만 및 거주지에 상관없이 30대 전체와 40대 초반 여성, 수도권에 거주하는 20대 후반 및 45세 이상 여성으로 구성된다. 세 번째 시장(M3)은 거주지에 상관없이 20대 초반 여성, 수도권에 거주하는 20대 후반 및 40세 후반 여성이 이에 해당한다. 네 번째 시장(M4)은 20대 남성으로만 구성된다. 마지막으로 다섯 번째 시장(M5)의 속성은 거주지에 상관없이 30대 초반의 남성, 비수도권에 거주하는 30대 후반 및 40대 후반의 남성으로 정의된다.

각 세분 시장의 상품네트워크는 Figs. 4-8에 각각 나타나 있다. Figs. 4-8의 상품네트워크는 가독성을 높이기 위하여 3회 이상 구매된 상품만을 표기하였으며 연결이 없는 고립 노드는 표시하지 않았다. 남성으로 구성된 M1, M4, M5 시장의 상품네트워크는 Figs. 4, 7, 8로 각각 표현되며, 여성으로만 구성된 M2, M3 시장의 상품네트워크

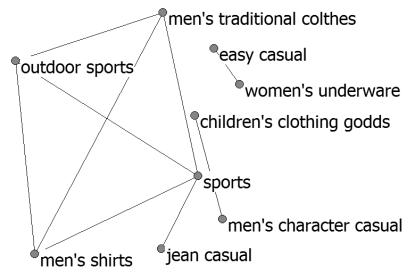


Fig. 4. Product Network for market M1



Fig. 5. Product Network for market M2

4. 결론 및 향후 연구과제

본 연구는 사회네트워크 분석을 활용한 새로운 시장세분화 기법을 제안하였다. 제안된 기법은 크게 4단계로 구성된다. 첫 번째 단계는 시장세분화를 위한 고객변수를 선정하고 각 변수의 속성 값의 상태를 정의하는 단계이며, 두 번째는 고객 세그먼트 및 각 세그먼트 별 상품네트워크를 구성하는 단계이다. 세 번째 단계는 상품네트워크 간 유사성을 측정하고 이를 이용한 고객 세그먼트의 군집화를 수행하는 단계이며, 마지막은 군집 결과를 바탕으로 세분 시장의 개수를 정하고 각 시장의 속성을 파악하여 시장 세분화 기준을 도출하는 단계이다. 각 단계는 효과적인 시장 세분화 기준에 부합하도록 설계되었다.

본 연구에서는 제안된 방법론의 타당성 및 적용 가능성을 검증하기 위하여 국내의 한 중형 인터넷 쇼핑몰의 구매내역 데이터를 확보하여 사례분석을 실시하였다. 사례분석 결과 제안된 기법은 합리적으로 시장 세분화 결과를 도출하는 것으로 확인되었으며, 더불어 각 시장별 상품네트워크 정보는 타겟 마케팅에 효과적으로 활용될 수 있을 것으로 예상된다.

한편, 본 연구의 시사점은 다음과 같다. 먼저 시장세분화와 관련된 기존 연구에서는 크게 인구통계학정보 등을 활용한 고객의 속성기반(attribute based)의 세분화 기법과 구매 트랜잭션 기반의 세분화 기법이 주를 이루었다. 그러나 본 연구에서는 이 두 가지 정보를 모두 활용하여 효과적인 시장세분화 기준에 부합하는 기법을 제시하였다는 것이 기존 연구들과 차별화 되는 점이라고 할 수 있다.

두 번째, 본 연구에서는 고객들의 구매패턴을 파악하기 위하여 기존 연구에서 시도된 적이 없었던 상품네트워크를 활용하였다. 상품네트워크에는 해당 고객들의 연결구매 패턴이 반영되어 있어 특정 집단의 구매패턴을 확인하는데 효과적이며, 더불어 네트워크 간의 유사성 비교 분석이 가능하여 집단 간의 구매 패턴의 차이를 밝히는데 유용하다. 제안된 기법은 이러한 사실에 착안하여 유사한 구매 패턴을 보이는 고객 세그먼트들을 통합해 감으로써 ‘세그먼트 내 높은 동질성’ 및 ‘세그먼트 간 낮은 동질성’의 기준을 반영할 수 있다.

본 연구에서 활용한 상품네트워크 간의 QAP 상관계수 분석은 네트워크를 구성하는 노드(상품) 간의 연결 측면에서 네트워크 간의 유사도를 측정한다. 즉, 고객의 구매 패턴을 고려할 때 고객이 구매한 상품의 목록만을 고려하며, 각 상품의 구매 수량은 감안하지 못한다. 또한 단일 품목만 구매한 고객은 상품네트워크 상의 상품 간 연결에

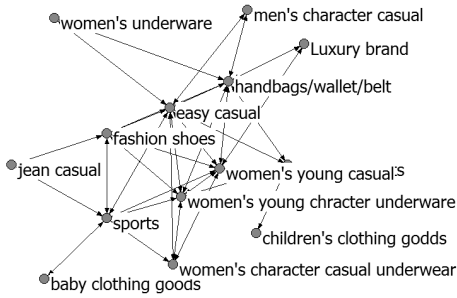


Fig. 6. Product Network for market M3

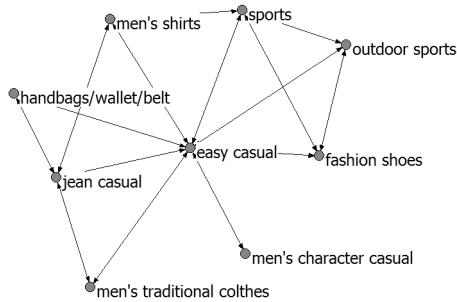


Fig. 7. Product Network for market M4

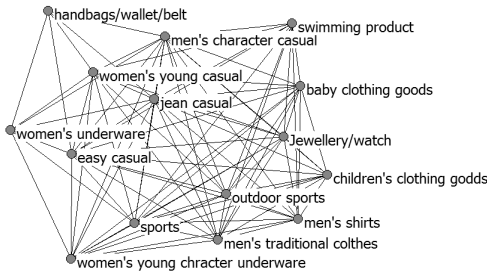


Fig. 8. Product Network for market M5

는 Figs. 5, 6으로 각각 표현된다. M1, M4, M5 시장은 남성 용품 및 아웃도어나 스포츠 관련 상품의 연결이 주를 이루며, M2, M3 시장은 여성이나 패션 용품 간에 연결이 나타나고 있다. 또한 동일한 남성으로 구성된 시장이라도 비교적 젊은 20대 남성으로 구성된 M4 시장의 경우 캐주얼 의류 및 패션, 아웃도어 및 스포츠 관련 상품간의 연결이 눈에 띄는 반면, 상대적으로 나이가 많은 남성 집단인 M1 시장의 경우 자녀 및 전통의상 관련 상품이 나타나고 있다. 이렇듯 각 시장의 상품네트워크는 각 시장을 구성하는 고객의 구매 패턴에 대한 이해 증대 및 타겟 마케팅에 유용한 도구로 사용될 수 있을 것으로 예상된다.

기여하지 못하기 때문에, 이들의 구매 패턴 역시 상품네트워크에 반영되지 못한다. 따라서 고객의 대다수가 단일 품목만을 구매하는 구매패턴을 보이거나, 구매의 수량에서도 뚜렷한 패턴이 존재할 경우 제안된 기법의 적용에는 한계가 있을 수 있다.

그럼에도 불구하고 본 연구에서 제안된 시장세분화방법은 기본적인 고객정보 및 상품구매정보를 이용하여 상품구매패턴이 유사한 고객 집단을 인구통계학적인 일반 변수 기준으로 세분화할 수 있기 때문에 대다수의 온오프라인 유통업체에서 폭넓은 활용이 가능할 것으로 기대된다.

References

1. S.P. Borgatti, M.G. Everett, and J.C. Johnson (2013), *Analyzing Social Networks*. London: SAGE Publications.
2. M.L. Doerfel, and G.A. Barnett (1999), "A semantic network analysis of the International Communication Association", *Human Communication Research*, Vol. 25, No. 4, pp. 389-603.
3. F.M. Hsu, L.P. Lu, and C.M. Lin (2012), "Segmenting customers by transaction data with concept hierarchy", *Expert Systems with Applications*, Vol. 39, No. 6, pp. 6221-6228.
4. J. Han and M. Kamber (2001), *Data Mining - Concepts and Techniques*. Morgan Kaufmann.
5. H.K. Kim, J.K. Kim, and Q.Y. Chen (2012), "A Product Network Analysis for Extending the Market Basket Analysis", *Expert Systems with Applications*, Vol. 39, pp. 7403-7410.
6. P. Kotler, L. Brown, A. Stewart, and G. Armstrong (2003), *Marketing*. Pearson Education Australia, 6th edition.
7. T. Radedder and N.V. Chawla (2009), "Modeling a store's product space as a social network", *Proceedings of the 2009 international conference on advances in social network analysis and mining*, pp. 164-169.
8. B. Sarwar, G. Karypis, J.A. Konstan, and J. Riedl (2000), "Analysis of recommendation algorithms for e-commerce", *Proceedings of ACM E-commerce 2000 conference*, pp. 158-167.
9. C.Y. Tsai and C.C. Chiu (2004), "A Purchase-based Market Segmentation Methodology", *Expert Systems with Applications*, Vol. 27, No. 2, pp. 265-276.
10. D.P. William and Jr. E.J. McCarthy (1992), *Basic Marketing: A Global-Managerial Approach*, IRWIN, 12th edition.
11. S.B. Jeong (2015), "The Relationship between Structure Attributes of Product Networks and Sales Performance of Integrated Internet Shopping Malls", *Business Education research*, Vol. 30, No. 3, pp. 433-453.
12. I.H. Choi, B.K. Kim, and S.B. Jeong (2014), "A Novel Approach to Product Classification based on Product Network Analysis", *Entrue Journal of Information Technology*, Vol. 13, No. 2, pp. 93-104.
13. M. Natter (1999), "Conditional market segmentation by neural networks: a Monte-Carlo study", *Journal of Retailing and Consumer Services*, Vol. 6, No. 4, pp. 237-248.
14. T.C. Lu, and K.Y. Wu (2009). "A transaction pattern analysis system based on neural network", *Expert Systems with Applications*, Vol. 36, pp. 6091-6099.



정 석 봉 (sbjung@kiu.ac.kr)

1999 한국과학기술원(KAIST) 산업경영학과 학사
 2001 한국과학기술원(KAIST) 산업공학과 석사
 2005 한국과학기술원(KAIST) 산업공학과 박사
 2011~현재 경일대학교 경영학부 부교수

관심분야 : 사회네트워크분석, 데이터분석, 제약이론(TOC)



신 용 호 (yhshin@ynu.ac.kr)

1993 서울대학교 산업공학과 학사
 1995 한국과학기술원(KAIST) 산업공학과 석사
 2003 한국과학기술원(KAIST) 산업공학과 박사
 2001~2003 한국IBM 경영컨설턴트
 2003~2005 삼성전기(주) 정보경영팀 과장
 2006~2008 한국전자통신연구원 선임연구원
 2009~현재 영남대학교 경영학과 부교수

관심분야 : 운영전략, 서비스 사이언스, 이산사건모형, Data Science



구 서 룡 (seoryong.koo@doosan.com)

1998 한국과학기술원(KAIST) 원자력공학과 학사
 2000 한국과학기술원(KAIST) 원자력 및 양자공학과 석사
 2005 한국과학기술원(KAIST) 원자력 및 양자공학과 박사
 2006~현재 두산중공업(주) 원자력I&C BU 재직 중

관심분야 : SW엔지니어링, SW 확인 및 검증(V&V), 애플리케이션 생명주기 관리(ALM)



윤 협 상 (hsyoon@cu.ac.kr)

1997 한국과학기술원(KAIST) 산업공학과 학사
 2000 한국과학기술원(KAIST) 산업공학과 석사
 2007 한국과학기술원(KAIST) 산업공학과 박사
 2010~현재 대구가톨릭대학교 경영정보학과 조교수

관심분야 : 실험계획법, 네트워크 시뮬레이션, 비디오 품질

부록 고객 세그먼트 간 QAP 상관계수

	4	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	22	23	24	27	28	30	32
4	1	0.06	0.61	0.55	0	0.02	0.47	0.63	0.14	0.17	0.73	0.72	-0.01	0.13	0.61	0.71	0.08	0.68	0.7	0.43	0.74	-0.01	0.68
6	0.06	1	0	0.11	0.76	0.57	0	0.15	0.27	0.37	0	0.07	0	0.33	0.18	0.06	0.14	0.06	0.15	0	0.17	0	0.09
7	0.61	0	1	0.62	0	0.03	0.59	0.52	0.17	0.19	0.76	0.63	0	0.12	0.37	0.6	0.14	0.23	0.61	0.39	0.5	0.05	0.45
8	0.55	0.11	0.62	1	0	0.08	0.68	0.62	0.22	0.27	0.7	0.74	0.22	0.22	0.6	0.67	0.24	0.42	0.71	0.63	0.68	0.02	0.56
9	0	0.76	0	0	1	0.51	0	0.12	0.36	0.39	0	0.05	0	0.41	0.1	0.05	0.2	0.08	0.14	0	0.2	0	0.12
10	0.02	0.57	0.03	0.08	0.51	1	0.16	0.26	0.26	0.46	-0.01	0.17	0.05	0.55	0.21	0.1	0.36	0.09	0.22	-0.01	0.16	0.15	0.05
11	0.47	0	0.59	0.68	0	0.16	1	0.56	0.11	0.07	0.66	0.7	0	0.1	0.53	0.63	0.1	0.39	0.65	0.64	0.63	-0.01	0.5
12	0.63	0.15	0.52	0.62	0.12	0.26	0.56	1	0.23	0.35	0.65	0.81	0.05	0.41	0.68	0.76	0.31	0.61	0.77	0.47	0.74	0.15	0.6
13	0.14	0.27	0.17	0.22	0.36	0.26	0.11	0.23	1	0.4	0.15	0.2	0.12	0.44	0.18	0.13	0.28	0.05	0.17	0.11	0.16	-0.01	0.07
14	0.17	0.37	0.19	0.27	0.39	0.46	0.07	0.35	0.4	1	0.15	0.25	0.2	0.62	0.18	0.24	0.64	0.14	0.38	0.09	0.25	0.29	0.12
15	0.73	0	0.76	0.7	0	-0.01	0.66	0.65	0.15	0.15	1	0.85	0.07	0.14	0.6	0.84	0.18	0.55	0.81	0.59	0.79	-0.01	0.64
16	0.72	0.07	0.63	0.74	0.05	0.17	0.7	0.81	0.2	0.25	0.85	1	0.04	0.24	0.73	0.88	0.23	0.7	0.9	0.6	0.88	0.09	0.71
17	-0.01	0	0	0.22	0	0.05	0	0.05	0.12	0.2	0.07	0.04	1	0.21	-0.01	0	0.38	-0.01	0.07	0.23	0.04	0.32	-0.01
18	0.13	0.33	0.12	0.22	0.41	0.55	0.1	0.41	0.44	0.62	0.14	0.24	0.21	1	0.28	0.21	0.58	0.15	0.29	0.1	0.22	0.34	0.14
19	0.61	0.18	0.37	0.6	0.1	0.21	0.53	0.68	0.18	0.18	0.6	0.73	-0.01	0.28	1	0.75	0.18	0.68	0.72	0.53	0.77	-0.01	0.7
20	0.71	0.06	0.6	0.67	0.05	0.1	0.63	0.76	0.13	0.24	0.84	0.88	0	0.21	0.75	1	0.2	0.74	0.89	0.52	0.89	0.02	0.73
22	0.08	0.14	0.14	0.24	0.2	0.36	0.1	0.31	0.28	0.64	0.18	0.23	0.38	0.58	0.18	0.2	1	0.06	0.32	0.12	0.16	0.44	0.04
23	0.68	0.06	0.23	0.42	0.08	0.09	0.39	0.61	0.05	0.14	0.55	0.7	-0.01	0.15	0.68	0.74	0.06	1	0.71	0.4	0.82	0.02	0.7
24	0.7	0.15	0.61	0.71	0.14	0.22	0.65	0.77	0.17	0.38	0.81	0.9	0.07	0.29	0.72	0.89	0.32	0.71	1	0.61	0.88	0.13	0.71
27	0.43	0	0.39	0.63	0	-0.01	0.64	0.47	0.11	0.09	0.59	0.6	0.23	0.1	0.53	0.52	0.12	0.4	0.61	1	0.58	-0.01	0.52
28	0.74	0.17	0.5	0.68	0.2	0.16	0.63	0.74	0.16	0.25	0.79	0.88	0.04	0.22	0.77	0.89	0.16	0.82	0.88	0.58	1	0.03	0.8
30	-0.01	0	0.05	0.02	0	0.15	-0.01	0.15	-0.01	0.29	-0.01	0.09	0.32	0.34	-0.01	0.02	0.44	0.02	0.13	-0.01	0.03	1	-0.01
32	0.68	0.09	0.45	0.56	0.12	0.05	0.5	0.6	0.07	0.12	0.64	0.71	-0.01	0.14	0.7	0.73	0.04	0.7	0.71	0.52	0.8	-0.01	1