

쇼핑 웹사이트 탐색 유형과 방문 패턴 분석

최경빈

한양대학교 일반대학원
비즈니스인포매틱스학과
(iris122413@gmail.com)

남기환

KAIST 경영대학원
경영공학부
(namkh@kaist.ac.kr)

온라인 소비자는 쇼핑 웹사이트에서 특정 제품군이나 브랜드에 속한 제품들을 둘러보고 구매를 진행할 수 있고, 혹은 단순히 넓은 범위의 탐색 환경을 보이며 여러 페이지들을 돌아보다 구매를 진행하지 않고 이탈할 수 있다. 이러한 온라인 소비자의 행동과 구매에 관련된 연구는 꾸준히 진행되어왔으며, 실무에서도 소비자들의 행동 데이터를 바탕으로 한 서비스 및 어플리케이션이 개발되고 있다. 최근에는 빅데이터 기술의 발달로 소비자 개인 단위의 맞춤형 전략 및 추천 시스템이 활용되고 있으며 사용자의 쇼핑 경험을 최적화하기 위한 시도가 진행되고 있다. 하지만 이와 같은 시도에도 온라인 소비자가 실제로 웹사이트를 방문해 제품 구매 단계까지 전환될 확률은 매우 낮은 실정이다. 이는 온라인 소비자들이 단지 제품 구매를 위해 웹사이트를 방문하는 것이 아니라 그들의 쇼핑 동기 및 목적에 따라 웹사이트를 다르게 활용하고 탐색하기 때문이다. 따라서 단지 구매가 진행되는 방문 외에도 다양한 방문 형태를 분석하는 것은 온라인 소비자들의 행동을 이해하는데 중요하다고 할 수 있다. 이러한 관점에서 본 연구에서는 온라인 소비자의 탐색 행동의 다양성과 복잡성을 설명하기 위해 실제 E-commerce 기업의 클릭스트림 데이터를 기반으로 세션 단위의 클러스터링 분석을 진행해 탐색 행동을 유형화하였다. 이를 통해 각 유형별로 상세 단위의 탐색 행동과 구매 여부가 차이가 있음을 확인하였다. 또한 소비자 개인이 여러 방문에 걸친 일련의 탐색 유형에 대한 패턴을 분석하기 위해 순차 패턴 마이닝 기법을 활용하였으며, 같은 기간 내에 제품 구매까지 완료한 소비자와 구매를 진행하지 않은 채 방문만 진행한 소비자들의 탐색 패턴에 대한 차이를 확인할 수 있었다. 본 연구의 시사점은 대규모의 클릭스트림 데이터를 활용해 온라인 소비자의 탐색 유형을 분석하고 이에 대한 패턴을 분석해 구매 과정 상의 행동을 데이터 기반으로 설명하였다는 점에 있다. 또한 온라인 소매 기업은 다양한 형태의 탐색 유형에 맞는 마케팅 전략 및 추천을 통해 구매 전환 개선을 시도할 수 있으며, 소비자의 탐색 패턴의 변화를 통해 전략의 효과를 평가할 수 있을 것이다.

주제어 : E-Commerce, 온라인 소비자 행동, 웹 사용 마이닝, 순차 패턴 마이닝, 클릭스트림 데이터

논문접수일 : 2018년 11월 16일 논문수정일 : 2019년 3월 6일 게재확정일 : 2019년 3월 11일
원고유형 : 일반논문 교신저자 : 남기환

1. 서론

1.1 연구 배경 및 목적

온라인 유통 산업의 규모는 지속적으로 증가하고 있는 현재의 추세와 함께, 스마트 디바이스의 활용과 다양한 부가 서비스의 발전으로 인해

더 많은 소비자들이 온라인 쇼핑 채널을 통해 제품을 구매하고 있다. 국내의 경우 2018년 8월 기준, 전년도 동월 대비 22.7% 증가한 9조 4,567억 원의 거래 규모를 기록했으며, 특히 모바일 기기를 통한 온라인 쇼핑 거래 규모는 전년도 대비 33.5%가 증가한 5조 9,201억 원을 기록하였다 (KOSIS, 2018). 이러한 흐름은 전세계의

E-commerce 산업 동향에도 적용되고 있다. Statista의 조사에 따르면, 전세계 E-commerce 산업의 매출은 2017년 기준 2조 3,040억 달러로 추산되며, 2021년에는 2배 이상 증가한 4조 8780억 달러에 다다를 것으로 예측되고 있다. 이와 같은 추세에 자연스럽게 소비자들이 남긴 온라인 상에서의 다양한 종류의 데이터들이 규모, 속도, 형태 면에서 “빅데이터”의 요소를 가지게 되었다(Bradlow et al., 2017). 이와 같은 데이터는 소비자의 쇼핑 경험을 설명하기 때문에 많은 기업들은 소비자의 구매를 유도하기 위해 관련 데이터를 활용한 어플리케이션 및 서비스를 응용하고 있으며 활발하게 연구가 진행되고 있다. 대표적으로 웹 사이트의 콘텐츠 개인화(Eirinaki and Vazirgiannis, 2003; Hong and Tam, 2006), 추천 시스템(Huang et al., 2006; Lu et al., 2015; Lee, 2018), 광고 최적화(Goh et al., 2015; Bae and Park., 2018) 등이 활용되고 있다.

온라인 쇼핑 산업의 가파른 성장에도 불구하고, 전세계적으로 온라인 소비자의 구매 전환율 자체는 2%~3%대의 매우 낮은 수치를 기록하고 있다(Statista, 2018). 또한 소비자는 이전에 탐색한 제품들을 온라인 쇼핑 카트에 넣어 구매 진행의 편의를 높일 수 있지만(Kukar-Kinney and Close, 2010), 실제로 카트에 담긴 채 구매까지 진행되지 않는 비율은 70%가 넘는다(Statista, 2018). 이는 곧, 데이터와 IT의 활용을 통해 웹사이트로 소비자들을 유입시키는 여러 노력에도 불구하고 실제로 구매로 이어지는 확률은 상대적으로 매우 낮다는 것을 의미한다. 실제로 온라인 소비자들은 제품 구매를 목적으로 하는 것이 아닌 여러 동기와 목적에 따라 다양한 형태로 웹사이트를 탐색할 수 있으며, 이러한 여러 형태의 행동을 이해하기 위한 선행 연구들이 진행되어

왔다(Moe, 2003, 2006; Montgomery et al., 2004; Bucklin et al., 2002;). 특히 구매를 진행하기까지 소비자는 정보를 수집하는 과정을 거치며 정보를 수집하는 행동의 특성은 구매를 예측하는데 유의미한 요인이 될 수 있다(Huang et al., 2009; Mallapragada et al., 2016).

본 연구에서는 온라인 쇼핑을 진행하는 소비자의 다양한 탐색 행동과 구매 행동을 이해하기 위한 마이닝 기법을 활용한다. 탐색을 한 페이지의 수, 머무른 시간, 페이지의 유형, 구매 여부 등의 정보를 포함하는 대규모의 세션 단위 데이터를 대상으로 다소 복잡하고 이질적인 행동의 특성을 이해하기 위해 웹 사이트 내에서의 행동 패턴을 유형화할 수 있는 클러스터링 분석을 진행한다. 또한 온라인 쇼핑 소비자는 개인의 쇼핑 경험과 목적, 동기에 따라 제품 구매를 확정하기까지 다양한 탐색 행동들에 대한 일련의 패턴 혹은 규칙을 가질 수 있다. 이에 대한 분석을 위해 클러스터링 분석을 통해 여러 탐색 유형을 도출한 후, 순차 패턴 마이닝 기법을 적용하여 온라인 소비자들의 탐색 형태에 대한 패턴을 분석하는 과정을 진행한다.

본 연구의 첫 번째 시사점 온라인 소비자들의 쇼핑 행동 데이터를 기반으로 제품 탐색과 구매에 대한 행동의 복잡성을 설명하기 위한 마이닝 방법을 적용한다는 점에 있다. 특히 기존의 관련 연구들과는 다르게 본 연구에서 활용될 8백만 건이 넘는 페이지 클릭 기록이 포함된 데이터셋은 500여가지의 유관 브랜드들이 포함되어 소비자의 제품 선택 및 구매 결정에 대한 다양한 상황을 반영할 수 있다. 두 번째 시사점은 복잡하고 다양한 형태의 온라인 소비자의 쇼핑 패턴을 설명하기 위해 클러스터링과 순차 패턴 마이닝 두 가지의 방법들을 제안했다는 점에 있

다. 여러 형태의 웹 페이지 단위의 탐색 행동 특징들을 클러스터링 분석을 통해 유형화할 수 있었으며, 기존의 관련 연구와 유사하면서도 차이점이 존재하는 탐색 유형들이 도출되었다. 그리고 대부분의 온라인 쇼핑 탐색 혹은 방문 유형에 대한 연구들은 해당 유형의 도출 및 분석만을 진행한 반면, 본 연구에서는 해당 유형들이 어떤 순서와 패턴을 가지는지를 패턴 마이닝 기법을 통해 도출하였다. 특히 같은 기간 내에 제품 구매를 완료한 소비자와 구매를 하지 않으면서 제품을 둘러보기만 하는 소비자들의 탐색 패턴 간의 차이점을 도출할 수 있었다. 그러므로 본 연구의 분석 과정은 온라인 유저의 행동 분석에 대한 연구를 확장했다는 점에서 의미가 있다고 할 수 있다.

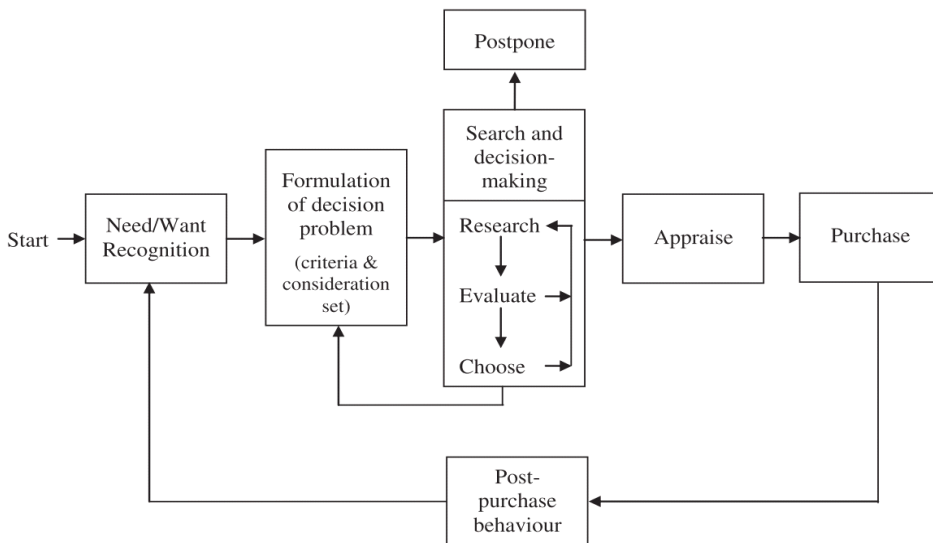
본 연구의 구성은 다음과 같다. 2장에서는 온라인 소비자 행동 연구를 포함한 관련 연구들을 검토하였다. 3장에서는 본 연구에서 제안하는 마

이닝 방법론들과 분석 절차에 대해 기술하였으며 4장에서는 제안된 분석 절차에 따른 분석 결과를 설명하였다. 마지막으로 본 연구의 결론 및 한계점과 향후 추가 연구 방향을 제시하였다.

2. 관련 연구

2.1 온라인 쇼핑 행동과 동기

소비자 행동에 관한 연구들에 따르면, 소비자들은 제품을 구매하기까지 여러 단계에 걸쳐 의사결정 및 다양한 형태의 행동을 취하게 된다 (Howard and Sheth, 1969; Bettman et al., 1998). 특히 정보 처리 이론은 소비자의 행동 모델 소비자 행동에 관련된 모델을 구성하는데 가장 중요한 요인을 포함하고 있다. <Figure 1>에서 알 수 있듯이, 소비자의 의사 결정 과정은 정보 처리 이론의 관점에서 문제/니즈의 인지, 정보 탐색,



<Figure 1> Online Consumer Decision Process (Karimi et al., 2015)

대안 평가 및 선택, 구매, 구매 후 행동의 다섯 단계로 구분이 가능하다. 물리적 매장에서 쇼핑을 하는 소비자의 행동을 설명하지만, 온라인 소비자의 행동을 설명하기 위한 연구에서도 기존의 소비자 구매 의사결정 과정이 적용되었다 (Kukar-Kinney and Close, 2010; Jansen and Schuster, 2011; Karimi et al., 2015). Karimi et al., 2015의 연구에서는 기존의 소비자 행동과 의사결정 연구들을 바탕으로 온라인 소비자의 구매 의사결정 과정 단계를 <Figure 1>과 같이 구성하였다. 온라인 소비자의 구매 의사결정은 기존의 의사결정 모델과 다르게 선형적 방향성을 가지지 않고, 단계 별 행동에서 전 단계나 혹은 다른 단계로 되돌아가는 동적인 형태를 보이는 점이 가장 큰 특징이다.

온라인 쇼핑 환경에서 소비자 행동은 시간적, 물리적 환경의 분리와 개인의 정보 처리의 방식 면에서 기존의 물리적인 쇼핑 환경과 큰 차이를 보인다(Ba et al., 2002). 소비자의 행동 관점에서, 소비자가 웹 사이트로 유입되고 쇼핑을 마치고 웹 사이트의 창을 닫을 때까지의 온라인 쇼핑 과정에는 정보 획득과 제품 구매의 두 가지 주요한 단계가 있다(Pavlou and Fygenson., 2006). Mallapragada et al., (2016)는 온라인 소비자의 행동을 탐색(Browsing) 단계와 구매(Purchasing) 단계로 모델링하였으며, 탐색 과정에서의 시간 및 탐색 페이지의 수와 제품의 특성, 웹사이트의 특성이 제품 구매 결정과 구매를 결정했을 때의 구매 금액에 유의한 영향을 미침을 확인하였다.

온라인 쇼핑 환경의 주요 특징들 중 소비자와 판매자와의 시간적, 물리적 분리는 기존의 물리적인 쇼핑 환경의 특징과 가장 다른 특징이라 할 수 있다. 이러한 특성은 온라인 쇼핑 소비자가 제품의 품질을 정확하게 평가하기 어렵기 때문

에 불확실성에 의한 위험을 인지하게 만든다(Ba et al., 2002). 특히 제품을 내재적 특성에 따라 탐색재와 경험재로 구분할 때(Nelson, 1970; 1974), 경험재에 속하는 제품들은 온라인 쇼핑 환경에서 소비자의 행동과 구매 의사결정에 유의한 영향을 미칠 수 있다(Girard and Dion, 2010; Kim and Krishnan, 2015). 온라인 소비자들은 쇼핑 과정에서 인지하는 위험의 정도가 높아질수록 정보를 더 많이 수집하게 된다(Peck and Childers, 2003). Huang et al., 2009는 온라인 쇼핑 환경에서 소비자의 탐색 행동을 탐색의 깊이 (Depth of Search)와 탐색의 폭 (Breadth of Search)을 정의하였으며, 각각을 웹 페이지 내에서 머무른 시간과 페이지 뷰의 수로 측정하였다. 그들은 온라인 쇼핑 환경에서 경험재에 속하는 제품들은 탐색재에 속하는 제품들보다 탐색의 깊이가 크다는 것과 반대로 탐색재에 속하는 제품들은 경험재에 속하는 제품들보다 탐색의 폭이 넓다는 것을 실증적 분석을 통해 확인하였다.

소비자의 탐색 행동이 제품 특성에 영향을 받기도 하지만, 그들의 쇼핑 목적과 동기가 무엇인지에 따라서도 행동이 달라진다. 소비자의 쇼핑 동기는 여러 기준에서 설명이 가능한데, 그 중 실용적 동기(Utilitarian)와 쾌락적 동기(Hedonic)는 주요한 기준이 될 수 있다. 먼저 실용적 동기란 임무 중심적, 효과적인 의사결정, 목적 지향적 동기로 정의할 수 있다(Holbrook and Hirschman, 1982; Engel and Roger, 1993; Babin et al., 1994;). 실용적 동기는 쇼핑 활동을 일종의 업무 혹은 임무로 간주하며, 쇼핑 과정 동안 효율적으로 이를 완료할 수 있는가 혹은 그렇지 못하는가에 따라 이익이 결정된다고 믿게 만든다(Babin et al., 1994). 반대로 쾌락적 동기는 경험적, 감정적인 이득을 추구하는 동기이다. 쾌락적

동기는 행복, 환상, 감각, 기쁨을 위한 소비 행동을 추구하게 한다(Holbrook and Hirschman, 1982). 이 두 가지의 주요 쇼핑 동기는 온라인 환경에서 쇼핑 동기와 행동 간의 관계에 대한 다양한 연구들에 적용되어왔다(Bridges and Florsheim, 2008; Chiu et al., 2012; Scarpi, 2012). 실용적 동기와 쾌락적 동기와 유사한 맥락에서 소비자의 쇼핑 행동은 목표 지향적 행동과 탐색 지향적 행동으로 구분할 수 있다(Janiszewski, 1998). 목표 지향적 탐색 행동은 말 그대로 계획된 행동을 통해 구체적인 목표를 달성하기 위한 행동을 의미한다. 이러한 탐색 행동을 가지는 소비자들은 특정 제품을 구매하고자 하는 목표를 가진 채 관련 제품에 대한 정보를 탐색하기 때문에 매우 뚜렷한

방향을 보이는 특징을 가진다. 반대로 탐색 지향적 행동은 뚜렷한 구매 계획을 가지지 않은 상태에서 오로지 둘러보는 행위를 지향한다. Moe, 2003는 온라인 소비자의 탐색 행동을 목표 지향적인 형태와 탐색 지향적 형태 그리고 구매를 탐색 시점에 결정하는지 혹은 미래 시점으로 미루는지에 따라 <Table 1>의 형태로 4가지 유형의 탐색 행동을 분류하였다.

<Table 2>을 통해 알 수 있듯이, 기존의 관련 연구에서 도출된 여러 방문 유형들에 따라 각각 어떤 유형의 웹페이지를 집중해서 방문하는지 다양한 제품 유형들을 탐색하는지 등의 웹상에서의 구체적 행동들이 다름을 알 수 있다. 결론적으로, 온라인 쇼핑을 진행하는 소비자들은 개

<Table 1> Online shopping types (Moe, 2003)

Purchasing horizon	Search Behavior	
	Directed	Exploratory
Immediate	Directed buying	Hedonic browsing
Future	Search/Deliberation	Knowledge building

<Table 2> Characteristics of visit types (Moe, 2003; Liu et al., 2012; Pallant et al., 2017)

Visit Type	Main characteristics
Directed Buying	Repeatedly searching for specific products, the most likely to buy
Hedonic Browsing	Lowest purchase probability, product navigation across a wide range of brands and categories
Search / Deliberation	Explore and compare different products within a specific product category
Knowledge Building / Learners & Novices	Navigate only to information-oriented pages such as company and brand information, shipping exchange refunds, and more
Shallow / Touching Base	Describe the habit of entering websites as a type that deviates from a few pages within a short period of time
Cart-Only	Type of revisiting and purchasing only the products in the shopping cart
Promotion Finder	Product referrals, advertising, and a look at discount offer pages. Inflow of product page is low

인의 구매 의사결정의 단계, 제품의 특성에 따른 불확실성, 쇼핑을 진행하게 만드는 목적 및 동기에 따라 웹 사이트 내에서의 탐색 행동이 다르게 나타나며, 이는 다시 여러가지의 유형들로 설명될 수 있다. 본 연구에서는 소비자들의 다양한 쇼핑 상황 및 구매 결정 단계를 고려할 수 있는 대규모의 클릭스트림 데이터를 분석함으로써 기존 연구에서 밝혀진 탐색 유형들과의 차이를 분석하였다.

2.2 클릭스트림 데이터 연구

클릭스트림 데이터는 인터넷 사용자가 남긴 전자적 활동 기록들로 정의할 수 있다(Bucklin and Sismeiro, 2009). 이러한 기록은 사용자가 하나의 웹사이트 내에 여러 페이지들에 대한 일련의 선택 과정이나 여러 웹사이트들을 거쳐 방문한 기록을 포함한다(Bucklin et al., 2002). 클릭스트림 데이터는 상세한 수준으로 온라인 환경에서의 사용자 행동을 포함하기 때문에 온라인 소비자들의 정보 탐색 및 활용, 웹 사이트 선택, 구매 전환 등 다양한 주제의 연구에 활용될 수 있다(Moe, 2003; Bucklin et al., 2004; Park and Cho., 2017). Moe (2003)의 연구에서는 웹 페이지들을 특정 기준으로 분류하고(제품 상세, 카테고리, 브랜드, 구매 등), 사용자의 온라인 행동을 설명하기 위해 페이지 뷰와 머무른 시간 등의 변수를 통해 클러스터링 분석을 진행하였다. 이를 통해 온라인 소비자들 개개인의 쇼핑 계획 및 전략의 유형 4가지(Directed buying, Search/deliberation, Hedonic browsers, Knowledge-building)를 발견하였고 각각의 유형들의 탐색 패턴이 다름을 확인하였다. Montgomery et al., 2004의 연구에서도 비슷한 방법으로 웹 사이트 내의 페이지들을 분

류하여 제안된 동적 모델링을 통해 이전의 모델보다 페이지의 전환을 고려하면서 제품 구매를 예측하는데 개선된 성능을 보일 수 있음을 확인하였다. Moe (2003)의 연구는 페이지 단위의 이동을 포함하는 클릭스트림 데이터를 처음으로 활용해 소비자의 탐색적 혹은 목표 지향적 쇼핑 동기에 따른 행동 유형화 방법을 처음으로 제안하였으며, 이를 기반으로 온라인 소비자 행동의 유형화에 대한 후속 연구들이 진행되었다(Phang et al., 2010; Liu et al., 2012; Pallant et al., 2017).

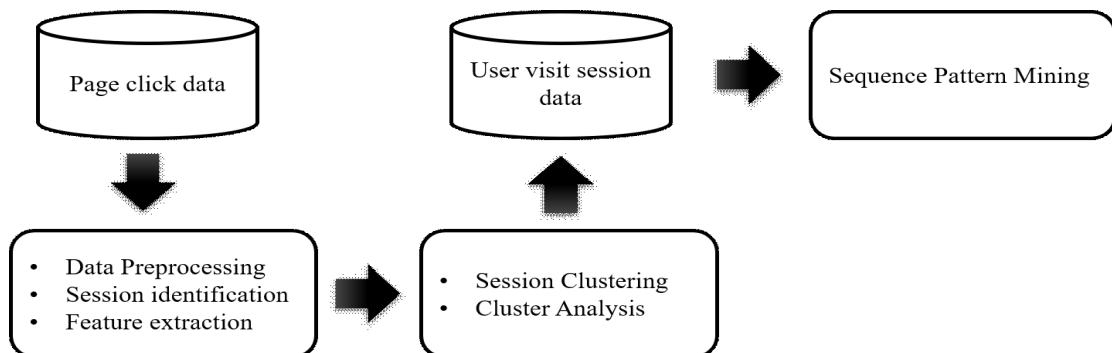
E-commerce 웹 사이트 클릭스트림 데이터를 활용한 대부분의 연구들에서는 소비자가 웹 사이트를 방문했을 때, 구매가 이루어지는지를 모델링했으며, 온라인 소비자의 탐색 행동은 구매 행동을 예측하는 모델에 주요 요인이 될 수 있음을 확인했다(Moe et al., 2004; Sismeiro and Bucklin, 2004; Moe, 2006; Mallapragada et al., 2016). Moe et al., 2004는 소비자의 웹사이트 방문 기록과 구매 이력을 기반으로 방문 시점 내 제품 구매 확률을 예측하는 모델을 제안하였다. 그들은 제안된 모델을 통해 E-commerce 웹사이트의 클릭스트림 데이터를 활용해 온라인 소비자의 방문 효과와 구매 임계점 효과로 구매 전환을 설명할 수 있음을 확인하였다. Moe, 2006는 온라인 소비자가 제품을 구매하기까지 과정을 대안이 될 수 있는 전체 제품들 중 탐색과 검색의 대상이 되는 제품을 선택하는 단계와 탐색 과정을 거친 제품을 구매하는 단계로 구분한 구매 전환 모델을 제안하였다. Sismeiro and Bucklin, 2004는 온라인에서 제품 구매가 완료되기 전까지의 과정을 제품의 확인, 개인 정보 입력, 주문 완료의 3가지 업무들로 구분하였고, 각 단계에서 탐색 행동, 정보 수집, 재방문 등의 소비자 행동이 어떤 영향을 미치는지를 분석하였다.

클릭스트림 데이터 연구들 중 E-commerce 웹 사이트 방문자들의 온라인 구매 행동을 이해하고 모델링에 대한 연구 주제가 가장 활발하게 진행되는 분야에 속한다(Bucklin and Sismeiro, 2009). 본 연구에서도 클릭스트림 데이터 연구들의 내용을 기반으로 온라인 소비자의 쇼핑 행동과 관련된 주요 요인들을 고려해 탐색 행동 및 쇼핑 패턴에 대한 분석을 진행하였다.

3. 활용 데이터 및 분석 방법

온라인 소비자의 탐색 및 구매 행동에 대한 분석을 위해 본 연구의 분석 단계에서는 크게 2가지의 방법을 적용한다. 기존의 다양한 온라인 소비자 모델 관련 연구에서 웹 사이트 내 행동을 설명하기 위해 페이지 뷰의 수와 페이지에 머무른 시간이 주요 요인으로 활용되었다(Mallapragada et al., 2016; Huang et al., 2009; Montgomery et al., 2004). Moe, 2003의 연구에서는 여러 페이지의 클릭을 담고 있는 세션 단위로 어떤 유형의 페이지를 탐색했는지에 대한 비율을 변수로 활용해 웹 사이트 방문 형태를 유형화

하였다. 본 연구의 분석 과정에서는 이전의 관련 연구에서 활용된 요인들을 포함해 온라인 소비자의 웹 사이트 내에서의 행동을 설명할 수 있는 변수들을 활용해 세션 단위의 클러스터링을 진행한다. 이를 통해 웹사이트를 방문하는 하는 여러 목적과 소비자의 행동을 유형화해 설명한다. 관련 연구에서 언급되었듯이, 온라인 소비자의 의사결정 과정은 정보를 수집하고 구매를 결정하기까지 여러 단계를 순환하는 특징을 가지며 개인의 쇼핑 경험과 목적, 동기에 따라 제품 구매를 확정하기까지 다양한 탐색 행동들에 대한 일련의 패턴 혹은 규칙을 가질 수 있다. Bronnenberg et al. 2016는 온라인 소비자들의 클릭스트림 데이터 분석을 통해, 온라인 소비자들은 이전 방문에 선택한 대안을 다시 탐색하기 위해 재방문을 하며, 탐색을 진행해 나갈수록 제품 속성의 범위를 좁혀 나감을 실증적으로 밝혔다. 이는 곧 구매를 하기 전까지 대안을 설정하기 위한 정보 수집, 대안의 평가 등의 행동이 여러 번의 방문을 통해 진행될 수 있다는 점을 의미한다. 클러스터링 분석을 통해 도출된 유형은 소비자의 방문 목적과 행동을 설명할 수 있기 때문에 이를 활용하여, 구매를 결정한 소비자 집단의 방



〈Figure 2〉 Analysis Process in this study

문 패턴과 단순히 탐색만 진행하고 이탈한 소비자 집단의 방문 패턴 간의 차이를 분석하여 구매 의사결정 과정 상에 다른 단계에 있는 소비자들의 이질적인 행동을 파악하고자 한다. 이를 위해 각 유형들의 비율만을 분석하기 보다는 여러 방문 유형의 순서 조합을 도출할 수 있는 순차 패턴 마이닝 기법을 활용한다. <Figure 2>는 전체 분석 단계의 요약 과정을 담고 있으며, 각 단계 별로 포함된 상세 내용을 다음 절에서 기술하였다.

3.1 활용 데이터 및 전처리 과정

본 연구의 분석에서는 국내 패션 산업을 주도하고 있는 기업의 공식 쇼핑몰 웹사이트를 방문한 유저들의 클릭스트림 데이터와 유저들의 구매 기록 데이터를 활용한다. 2018년 10월 1달 동안 해당 웹사이트에 접속한 유저들의 페이지 단위 클릭 데이터를 수집하였으며 페이지에 머문 시간이 1초 이하인 비이상적 클릭 형태를 제외한 총 8,417,303개의 데이터 포인트가 포함된 데이터셋을 최종적으로 활용하였다. 해당 웹사이트로 유입된 유저는 총 500여개가 넘는 개별 브랜드와 15개 이상의 통합 제품 카테고리를 선택할 수 있다. 데이터 셋의 형태는 해당 웹사이트에서 탐색을 진행하는 유저를 식별할 멤버십 코드, 클릭해서 요청한 웹페이지의 URL, 페이지 요청 시간, 페이지의 유형을 포함하고 있다.

다양한 온라인 소비자 관련 연구들과 유사한 맥락으로, 본 연구에서도 하나의 세션 단위를 유저의 웹사이트 방문 단위로 정의한다. 하나의 세션에 포함된 여러 개의 페이지 단위 탐색 행동들에 따라 유저의 탐색 패턴이 설명될 수 있다 (Liaukonyte et al., 2015; Raphaeli et al., 2017). 유

저의 세션을 확인하는 과정에는 여러 방법들이 적용될 수 있다(Pabarskaite and Raudys, 2007; Liu and Kešelj, 2007). 관련 연구에서 자주 쓰이는 방법은 시간 기반의 휴리스틱 규칙을 적용하는 방법이다. 가장 일반적으로 그리고 상용화된 어플리케이션들에 적용되는 방법은 페이지의 요청 시간을 기준으로 세션을 구분하는 방법이다 (Pabarskaite and Raudys, 2007;). 이는 특정 세션에 포함되는 웹 페이지의 요청 시간 t 와 다음 웹 페이지의 요청시간 t' 의 차이인 $t'-t$ 가 미리 지정한 값보다 큰 경우 다음 웹페이지부터 새로운 세션으로 지정하는 방식이다. Guerbas et al., 2013의 연구에서는 온라인 상에서의 패턴을 예측하는 과정에서 세션을 정의할 때, 가장 많이 쓰이는 임계값인 30분보다 10분으로 설정할 때가 더 사용자 세션을 잘 구분할 수 있음을 확인하였다. 본 연구에서도 페이지 요청 시간의 차이를 기준으로 세션을 정의하였으며, 세션을 구분하는 시간 차이 기준 값은 10분으로 설정하였다.

유저의 세션을 도출하는 과정을 거치면 세션의 유형을 분석하기 위해 세션 단위의 변수들을 추출하는 과정을 진행한다. 이 변수들은 클러스터링 분석 과정에 활용된다. 기존의 온라인 소비자의 탐색 및 구매 행동 관련 연구에서 활용된 다양한 변수들을 토대로 세션 단위의 여러 특성들을 <Table 3>와 같이 정의하여 추출하였다 (Moe, 2003; Huang et al., 2009; Phang et al., 2010; Mallapragada et al., 2016). 세션 단위의 변수들의 통계량들은 <Table 4>의 내용과 같다. 변수 도출 후, 페이지 뷰가 1개 이거나 페이지에 머무른 평균 시간이 이상치에 해당하는 세션을 제외하고 총 549,872개의 유저 세션을 도출하여 분석을 진행하였다.

〈Table 3〉 Definition session-level variables

Variables	Definition	Classification
PAGEVIEW	Total number of Page views in session	Search of breadth
AVGTIME	Average time spent on page in session	Search of depth
CATPG	Percentage of product category pages in session	Focus page type
BRNDPG	Percentage of brand pages in the session	
PRODPG	Percentage of product detail pages in the session	
SEARCHPG	Percentage of search results pages in the session	
CARTPG	Percentage of shopping cart pages in the session	
PROMPG	Percentage of promotion pages in the session	
DIFFCAT	Percentage of product category that were unique in the session	
DIFFBRND	Percentage of brand that were unique in the session	
DIFFPROD	Percentage of product that were unique in the session	
MAXREP	Maximum number of times any one product page was viewed	Repeat product viewing

〈Table 4〉 Session-level variable statistics

Variables \ Statistics	Mean	Std Deviation	Min	Max
PAGEVIEW	8.69	7.51	2	36
AVGTIME	177.98	398.59	1	3083.75
CATPG	10%	20%	0%	100%
BRNDPG	36%	32%	0%	100%
PRODPG	32%	29%	0%	100%
SEARCHPG	4%	14%	0%	100%
CARTPG	2%	10%	0%	100%
PROMPG	14%	25%	0%	100%
DIFFCAT	36%	39%	0%	100%
DIFFBRND	45%	33%	0%	100%
DIFFPROD	46%	24%	3%	100%
MAXREP	5.81	5.03	1	36

3.2 클러스터링 분석

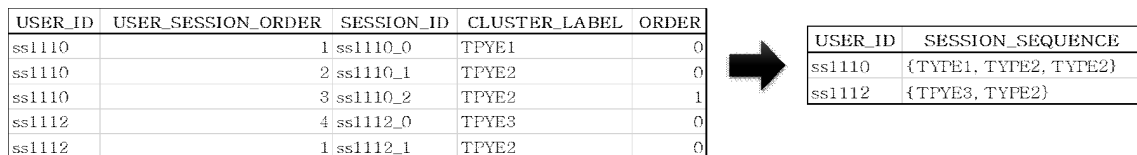
온라인 소비자의 방문과 방문 내에서의 이질적인 탐색 행동들을 분류하기 위해 클러스터링 분석을 진행한다. K-means 알고리즘은 학습의

효율과 성능이 비교적 우수해 다양한 클러스터링 문제에 적용되어왔으며, 온라인 쇼핑 방문 유형을 분석한 연구에서도 가장 많이 활용되었다 (Moe, 2003; Phang et al., 2010; Liu et al., 2012).

하지만 학습 데이터 셋의 크기가 증가하면 K-means 알고리즘의 효율성은 그 크기에 비례해 떨어지게 된다(Elkan, 2003; Jain, 2010). 분석에 활용될 데이터 역시 대규모의 클릭스트림 데이터로부터 추출된 사용자들의 방문 단위 데이터이기 때문에 K-means 알고리즘으로 학습을 진행할 경우 시간 및 연산 효율이 매우 저하될 수 있다. 이를 위해 기존의 K-means 알고리즘을 개선한 Mini-Batch K-means 알고리즘을 분석에 활용하였다. Mini-Batch K-means는 임의로 데이터 샘플들을 고정된 크기의 배치에 지정하고 배치 단위의 반복 학습을 통해 샘플들의 클러스터를 지정하는 과정을 거친다. 이러한 배치 단위의 학습은 수렴에 이르기까지의 계산 비용을 기존의 K-means 알고리즘보다 대폭 낮출 수 있다는 장점이 있다(Sculley, 2010). 클러스터의 수를 결정하는 과정은 초기의 2개 클러스터로부터 기존 클러스터와 동일한 클러스터가 나오거나 추가된 클러스터가 의미 있는 샘플 수를 충족하지 못할 때까지 클러스터의 수를 증가시켜 분석을 진행하였다(Moe, 2003; Phang et al., 2010). 클러스터링 분석에 활용되는 세션 단위의 변수들 중 비율을 제외한 실수 및 정수에 해당하는 변수는 표준화 처리하여 학습을 진행하였다. 또한 Mini-batch K-means는 배치 사이즈를 결정해야 하는데 이는 반복적으로 클러스터링을 진행하면서 학습 시간 및 클러스터링 결과를 고려해 50개로 결정하였다.

3.3 순차 패턴 마이닝

소비자의 구매 행동, 웹 접속 패턴, 사이버 보안 등 다양한 도메인을 통틀어 시간 상의 사건 혹은 품목들의 순서는 문제를 해결하는데 중요한 요인이 된다. 순차 패턴 마이닝은 시퀀스(순차) 데이터 내에 포함되어 있는 빈번하게 나타나는 하위 시퀀스를 찾아내는 기법이다(Agrawal and Srikant, 1995). 다시 말해, 시간 순서 상의 사건 혹은 품목들이 하나의 시퀀스 단위에 속하게 되고, 이러한 시퀀스가 데이터베이스 형태로 구축이 되어 있을 때, 순차 패턴 마이닝을 적용하면 최소 지지도 즉, 전체 시퀀스 중에서 일정 빈도수 이상을 가지는 순서를 도출해낼 수 있다. 본 연구의 분석 과정에서도 클러스터링 분석을 통해 도출된 여러 형태의 온라인 쇼핑 유형들에 대한 빈발 패턴을 도출하고 구매까지 완료한 유저와 구매를 하지 않은 유저의 빈발 패턴 간의 차이를 분석할 것이다. Figure 3에 표현한 것처럼, 세션 단위의 클러스터링 레이블을 활용하여 유저 단위의 방문 순서를 표현한 시퀀스 데이터를 구축하였다. 그리고 다양한 순차 패턴 마이닝 알고리즘 중에서도 후보 패턴을 데이터베이스 내에서 찾는 다른 알고리즘과 다르게 시퀀스 데이터베이스 내에서 깊이 우선 검색(Depth First Search)을 활용하는 PrefixSpan 알고리즘을 활용하였다(Pei et al., 2004).



〈Figure 3〉 Constructing Sequence data

4. 분석 결과

4.1 온라인 쇼핑 탐색 유형

클러스터링 분석을 통해 방문 유형은 총 4가지 유형들로 도출되었으며 결과의 요약은 Table 6를 통해 확인할 수 있다.

먼저 클러스터 1의 가장 두드러진 특징은 세션 내에서 프로모션 관련 페이지에 가장 많이 집중을 한다는 점이다. Moe, 2003의 연구에서 활용한 페이지 분류와 다르게 본 연구에서는 추가로 프로모션 페이지 유형을 추가해 클러스터링 분석을 진행하였다. 전체 방문 세션 중에 12%가 이 유형에 속하며, 프로모션 페이지에서 가장 많이 머무름과 동시에 4가지 도출된 유형중 페이지 뷰가 평균적으로 가장 적은 점이 특징이다. 동시

에 상대적으로 한 페이지에 머무른 시간은 긴 편에 속하기 때문에 프로모션 혹은 이벤트에 해당하는 제품들 중 원하는 제품을 리스트에서 둘러보는 형태를 보였다고 유추할 수 있다. 또한 탐색한 제품의 카테고리의 다양도가 가장 낮기 때문에, 좁은 범위의 특정 제품 카테고리만 탐색했다고 볼 수 있다. 클러스터 1은 Liu et al., 2012의 연구에서 도출한 ‘Promotion finder’와 매우 유사한 특징을 가짐을 확인하였다. 클러스터 2의 가장 큰 특징으로는 전체 방문의 평균보다 높은 브랜드와 제품 카테고리의 다양도를 들 수 있다. 특히 넓은 범위에 걸친 제품 카테고리를 탐색하면서 쇼핑을 진행하는 것이 가장 두드러진 특징이다. 또한 검색 기능을 활용하는 비율이 가장 낮으므로, 특정 목적을 위한 쇼핑을 진행하는 것

<Table 6> Cluster Solution

Cluster Feature	Total Samples	Cluster1	Cluster2	Cluster3	Cluster4
N(samples)	549,872	65,174	154,904	121,802	207,992
AVGTIME	177.9799	213.8202	190.5014	243.7382	118.9153
PAGEVIEW	8.688131	4.516586	6.295105	5.003612	13.93519
DIFFBRAND	45%	38.91%	52.74%	58.48%	33.35%
DIFFCAT	36%	0.50%	88.46%	6.11%	25.79%
DIFFPROD	46%	37.93%	41.92%	75.57%	34.68%
MAXREP	5.812864	3.996302	4.738302	2.230555	9.280203
BRANDPG	36%	17.47%	54.02%	9.63%	42.78%
CARTPG	2%	3.44%	0.70%	1.46%	3.87%
CATPG	10%	0.26%	15.06%	4.69%	11.97%
PRODPG	32%	6.72%	21.13%	69.07%	27.30%
PROMPG	14%	67.09%	6.90%	10.22%	5.51%
SEARCHPG	4%	3.39%	1.76%	3.89%	6.26%
ORDER	6%	5.68%	2.49%	3.13%	9.60%

이 아닌 유희적인 쇼핑 동기를 가지고 탐색을 진행하는 Hedonic Browsing 유형에 속한다고 볼 수 있다(Moe, 2003). 클러스터 3은 클러스터 2와는 다르게 여러가지의 브랜드에 걸쳐 특정 제품 카테고리에 제한된 제품 탐색을 하는 것이 주된 특징이라고 할 수 있다. 세션 내에서 제품 상세 페이지의 비중이 70% 정도로 나타나며, 특정 제품 카테고리에 속하는 다양한 제품들을 탐색해 제품에 대한 다양도가 가장 높은 것으로 나타났다. 그렇기 때문에 자연스럽게 특정 제품을 반복해서 탐색하는 횟수 역시 가장 낮은 것으로 나타났다. 클러스터 3은 Moe, 2003의 연구에서 규명한 Search/Deliberation 유형과 가장 유사하지만 브랜드에 대한 다양도가 높은 점이 차이점이라 할 수 있다. 즉, 브랜드와 제품 카테고리라는 두 가지의 제품 선택 기준 및 고려 사항 중 제품 카테고리만 범위가 정해진 상태에서 쇼핑을 진행하는 상태로 해석할 수 있다. 마지막으로 클러스터 4는 구매율이 가장 높으면서 반복적으로 동일 제품을 탐색하는 점이 가장 큰 특징이다. 또한 검색 기능을 가장 많이 활용하는 유형이면서, 브랜드, 제품 카테고리, 제품에 대한 다양도가 모두 낮게 나타났다. 페이지에 머무른 평균 시간은 다른 유형들에 비해 가장 짧으면서 세션 내에 페이지 뷰는 가장 많은 것으로 나타났다. 일반적으로 온라인 쇼핑 환경에서 경험재에 속한 제품을 탐색할 때는 해당 제품의 속성에 대한 정보를 처리하는 시간은 탐색재에 비해 오래 걸리기 때문에 페이지에 머무르는 시간 또한 길어지게 된다(Huang et al., 2009). 클러스터 4의 경우는 가장 많은 페이지 뷰를 기록하면서 각 페이지에서는 머무르는 평균 시간이 가장 짧았으며 다양도 특성들과 반복 탐색의 특성을 종합해 보았을 때 이전 방문에서 구매를 위한 탐색을 충분히 한 상태

에서 구매를 고려하고 있는 특정 소수의 제품들을 반복적으로 탐색하기 위한 방문 유형으로 정리할 수 있다. 이는 이전 관련 연구에서 Direct Buying 유형의 특징과 가장 유사하다고 할 수 있다(Moe, 2003; Pallant et al., 2017).

4.2 쇼핑 방문 패턴 분석

온라인 쇼핑 유저의 탐색 유형들의 순서를 담고 있는 시퀀스 데이터에 순차 패턴 마이닝을 적용해 클러스터링을 통해 도출한 4가지의 유형들의 순서 조합 중 빈발 패턴을 도출하였다. 본 분석의 활용 데이터의 기간인 한 달 동안 제품 구매를 진행한 유저 집단과 구매를 진행하지 않고 제품 탐색만 진행한 유저 집단의 패턴을 따로 도출하여 이들의 차이를 분석하였다. 빈발 패턴을 도출하는 최소 지지도는 10%로 설정하였으며, 2개 이상의 탐색 유형 시퀀스에 대해서는 유형간 순서가 나타날 비율을 설명할 수 있는 신뢰도를 계산하였다.

먼저 제품 구매를 진행한 유저들의 탐색 유형 시퀀스에서 도출된 빈발 패턴은 Table 7에서 확인할 수 있다. 가장 주목할 만한 특징은 특정 제품을 반복적으로 탐색하며 구매를 진행할 확률이 높은 클러스터 4 유형이 가장 높은 지지도를 가지는 빈발 패턴으로 도출된 점이다. 클러스터 4 유형이 84%의 지지도를 보인다는 것은 제품 구매를 진행한 전체 유저들의 탐색 시퀀스 중 84%에 해당하는 시퀀스에 모두 해당 유형이 포함되어 있다는 것을 말한다. 이를 통해 다시 한번 클러스터 4가 제품 구매와 밀접한 관련이 있는 탐색 유형임을 확인할 수 있다. 그리고 다른 탐색 유형 후에 순서상 뒤따르는 유형이 될 가능성이 가장 높은 유형 역시 클러

스터 4임을 확인할 수 있다. 예를 들어 다양한 브랜드와 넓은 범위의 제품 카테고리 내에서 제품을 탐색하는 클러스터 2 유형을 포함하는 모든 하위 시퀀스 중 다음 순서로 클러스터 4로 이어질 확률은 56.95%라 할 수 있다. 그리고 {클러스터 1 -> 클러스터 4}의 패턴 이후에 클러스터 4로 이어질 확률은 42.41%로 이 패턴 역시 전체 시퀀스 중 10% 이상을 차지하는 빈발 패턴으로 도출되었다.

〈Table 7〉 Frequent pattern of user who finished product purchase

Frequent Pattern	Support (%)	Confidence (%)
{Cluster1}	32.26%	-
{Cluster1}, {Cluster1}	11.07%	34.31%
{Cluster1}, {Cluster4}	18.83%	58.36%
{Cluster2}	44.07%	-
{Cluster2}, {Cluster2}	16.11%	36.56%
{Cluster2}, {Cluster3}	10.69%	24.25%
{Cluster2}, {Cluster4}	25.09%	56.95%
{Cluster2}, {Cluster4}, {Cluster4}	10.64%	42.41%
{Cluster3}	42.10%	-
{Cluster3}, {Cluster3}	14.44%	34.31%
{Cluster3}, {Cluster4}	21.38%	50.80%
{Cluster4}	84.81%	-
{Cluster4}, {Cluster1}	12.28%	14.48%
{Cluster4}, {Cluster2}	24.67%	29.09%
{Cluster4}, {Cluster3}	23.05%	27.18%
{Cluster4}, {Cluster4}	50.62%	59.69%
{Cluster4}, {Cluster4}, {Cluster4}	22.70%	44.85%

같은 기간 내에 제품 구매를 진행하지 않은 유저들의 탐색 시퀀스 상에서 빈발 패턴을 도출한 결과는 Table 8의 내용과 같다. 우선 제품 구매까지 완료한 유저들의 빈발 패턴과의 큰 차이점은 동일한 지지도를 기준으로 패턴을 도출하였을 때 유형들의 조합으로 구성된 패턴의 다양함이 적다는 점을 들 수 있다. 또한 반복적으로 같은 제품을 탐색하면서 구매 확률이 높은 클러스터 4 유형의 패턴에 대한 지지도가 상대적으로 낮은 편이다. 또한 동일한 패턴에 대한 신뢰도가 구매를 완료한 유저들의 패턴 신뢰도에 비해 낮은 것을 확인할 수 있다. 예를 들어, 특정 제품 카테고리를 다양한 브랜드에 걸쳐 탐색을 진행하는 클러스터 3 유형 다음에 특정 제품을 반복적으로 탐색하면서 제품 구매를 할 확률이 높은 클러스터 4 유형으로 이어지는 확률은 23%로 제품 구매를 완료한 유저들의 동일한 패턴 신뢰도

〈Table 8〉 Frequent patterns of non-purchasing users

Frequent Pattern	Support (%)	Confidence (%)
{Cluster1}	22.75%	-
{Cluster2}	49.47%	-
{Cluster2}, {Cluster2}	19.26%	38.93%
{Cluster2}, {Cluster4}	18.50%	37.39%
{Cluster3}	42.65%	-
{Cluster3}, {Cluster3}	12.65%	29.65%
{Cluster3}, {Cluster4}	10.08%	23.63%
{Cluster4}	56.60%	-
{Cluster4}, {Cluster2}	19.31%	34.12%
{Cluster4}, {Cluster3}	12.09%	21.35%
{Cluster4}, {Cluster4}	25.01%	44.19%

의 절반 수준인 것을 볼 수 있다. 이런 특징들을 종합해 볼 때, 비교적 단순한 형태의 빈발 패턴이 도출되었다는 점에서 구매를 진행하지 않은 유저의 탐색 시퀀스는 매우 이질적인 형태를 가진다는 점과 온라인 소비자의 구매 프로세스 (Howard and Sheth, 1969; Karimi et al., 2014) 상 니즈 인식 및 형성 혹은 정보 수집 단계에 더 집중되어 있다고 할 수 있다.

공통 도출 패턴에 대해서는 두 집단에서 각각 1000개의 세션을 무작위로 추출하여 공통 빈발 패턴의 지지도와 신뢰도 값의 샘플 100개를 구성하였으며, 해당 패턴의 지지도(support)와 신뢰

<Table 9> Mean support(confidence) of Common patterns derived from the group of users

Common pattern \ Group	purchase group	non-purchase group
{Cluster1}	31.08%	24.89%
{Cluster2}	42.12%	50.23%
{Cluster3}	43.06%	45.45%
{Cluster4}	85.21%	54.12%
{Cluster2}, {Cluster2}	16.31% (36.56%)	19.26% (36.83%)
{Cluster2}, {Cluster4}	26.13% (56.95%)	17.27% (34.29%)
{Cluster3}, {Cluster3}	15.44% (32.31%)	12.65% (30.15%)
{Cluster3}, {Cluster4}	20.12% (52.90%)	10.02% (21.32%)
{Cluster4}, {Cluster2}	22.17% (28.13%)	20.31% (34.12%)
{Cluster4}, {Cluster3}	23.21% (24.18%)	12.09% (21.35%)
{Cluster4}, {Cluster4}	53.42% (60.89%)	25.01% (44.19%)

도(confidence)의 평균에 대한 차이를 T-test를 통해 유의한 차이가 있음을 확인하였다. 두 집단에서 도출된 패턴의 주요한 차이는 구매 확률이 가장 높은 클러스터4의 지지도이다. 클러스터4와 관련된 패턴과 다음 방문에 클러스터4 유형으로 이어지는 패턴의 지지도와 신뢰도 모두 구매를 확정한 집단에서 더 높게 도출되었다. 반대로 매우 다양하고 넓은 범위의 제품 카테고리나 브랜드를 탐색하는 클러스터2와 연속해서 클러스터2의 형태로 쇼핑을 진행하는 패턴은 구매를 진행하지 않은 집단에서 더 높게 도출되었다. 특정 제품 카테고리가 정해진 상태에서 여러 브랜드에 걸쳐 다양한 제품을 비교하는 형태인 클러스터3의 비율은 두 집단에서 비슷하지만 연속해서 같은 유형의 방문을 진행하거나 클러스터4의 형태로 쇼핑을 하는 비율은 구매 집단에서 더 높게 도출되었다. 특히 클러스터3에서 클러스터4로 이어지는 경우는 제품 탐색의 옵션이 더 좁혀지는 방문 형태라고 볼 수 있으며, 두 집단 간 해당 패턴의 신뢰도의 차이가 크기 때문에 구매를 결정하고자 하는 소비자일수록 이와 같은 쇼핑 행태를 보일 확률이 높다 할 수 있다.

5. 결론

5.1 결론 및 시사점

본 연구에서는 실제 온라인 쇼핑 유저들의 대규모 클릭스트림 데이터를 기반으로 그들의 다양한 탐색 행동을 유형화하고 해당 유형들의 순차적 패턴을 도출함으로써 복잡하고 다양한 온라인 소비자의 행동을 이해하기 위한 분석을 진행하였다.

본 연구의 학술적 시사점은 다음과 같다. 먼저 탐색 유형을 도출하는 과정은 기반 연구가 되는 Moe, 2003의 연구와 유사했지만, 본 연구에서는 제품을 구매하기까지 과정에 더 연관성이 있는 페이지 유형 분류를 통해 분석을 진행하여 다른 탐색 유형을 도출하였다. 또한 클릭스트림 데이터를 활용한 유형화 연구(Moe, 2003; Phang et al., 2010; Pallant et al., 2017)에서 활용된 데이터보다 더 다양한 쇼핑 상황을 포함할 수 있는 빅데이터에 가까운 형태의 데이터를 활용했다는 점에도 의미가 있다. 대부분의 온라인 소비자의 행동을 유형화하는 연구에서는 유형에 대한 특징과 해당 유형을 구별하는데 어떤 요인이 주요한지에 집중해 분석해왔다(Moe, 2003; Rohm and Swaminathan, 2004; Ganesh et al., 2010). 본 연구에서도 이와 같은 분석 과정을 진행하였으며 더 나아가 유형들이 서로 어떤 순서로 구성되어 일련의 탐색 패턴이 될 수 있는지를 분석하였다. 특히 제품 구매를 완료한 유저들의 탐색 과정 패턴에서는 넓은 범위의 탐색을 진행한 후 특정 제품으로 선택사항을 좁혀 나가는 행동을 보이는 반면 구매를 진행하지 않은 유저들의 탐색 과정에서는 탐색 범위가 좁혀지는 형태가 빈발 패턴이 도출되지 않았다. 소비자가 제품을 구매하기까지의 과정을 설명하는데 많이 쓰인 의사결정 모델(Howard and Sheth, 1969)은 선형적인 특징을 가지고 있으며 이는 전통적인 오프라인 쇼핑 환경의 구매 과정을 설명하기에는 적합했으나 온라인 쇼핑 채널에서의 실제 구매 과정을 설명하기에는 한계점을 가진다. 온라인 소비자의 구매 과정이 기존의 전통적 모델의 단계를 포함한다 해도, 단순히 과정을 지나쳐오는 것 아닌 여러 단계의 순환과 반복적 특성을 보이면서 구매에 다다르게 된다(Karimi et al., 2015; Karimi et

al., 2018). 본 연구의 분석 결과에서 도출한 각 탐색 유형들의 순차적인 패턴은 온라인 소비자의 구매 의사결정 과정이 기존의 연구에서 제안한 단계의 반복과 순환을 가질 수 있음을 데이터 기반으로 설명할 수 있음을 보였다라는 점에서 학문적 시사점을 가질 수 있다.

본 연구의 실무적 시사점은 다음과 같다. 온라인 마케터는 본 연구의 분석 과정 및 결과를 활용해 구매 전환율을 제고하기 위한 마케팅 전략을 실행할 수 있을 것이다. 특히 제품 구매를 완료하기까지의 탐색 패턴은 넓은 범위의 탐색에서 특정 제품의 반복 탐색으로 범위를 좁히는 형태를 보였다. 소비자들은 온라인 상에서 제품 자체에 대한 인지된 위험의 수준이 높고, 경험적인 속성이 높은 제품에 대한 정보 처리와 제품 품질 평가가 어렵기 때문에 구매 전에 일련의 탐색 과정을 거치게 되는데(Ha and Stephen, 2002; Jacoby et al. 1994), 여러 형태의 탐색 과정을 통해 특정 제품에 대한 평가로 이어지는 과정을 효율적으로 만드는 것이 제품 구매율을 높일 수 있다는 점을 본 연구의 분석 결과를 통해 알 수 있다. 즉, 특정 소수의 제품들을 반복적으로 탐색하기 위한 방문 유형이 연속된다면 제품 구매 결정 단계에 가까워지고 있음을 유추할 수 있을 것이다. 이 때 온라인 마케터는 제품 추천 시스템과 프로모션 전략을 활용해 특정 제품 페이지로의 유입과 제품의 구매 전환율을 개선할 수 있을 것이다. 또한 구매까지의 방문 유형의 과정이 비슷한 집단을 추출해 대조군과 실험군을 나누어 개별 마케팅 전략들의 효과가 유저들의 방문 패턴에 변화를 줄 수 있는지, 구매 결정을 빠르게 만들 수 있는지를 확인하기 위해 본 연구에서 제시한 방법들을 활용해 평가할 수 있을 것이다.

5.2 한계점과 향후 연구

본 연구의 한계점과 향후 추가 연구 방향은 다음과 같다. 첫 번째로, 분석에 활용된 데이터는 하나의 온라인 유통기업의 웹사이트 내 클릭스트림 데이터로 이는 패널 형태의 유저 중심 클릭스트림 데이터에 비해 유저의 모든 탐색 행동을 담지 못한다는 한계를 가진다(Bucklin and Sismeiro, 2009). 본 연구에서 활용한 데이터는 여러 브랜드의 선택 상황도 반영할 수 있지만, 같은 제품 카테고리에 대해 다른 기업 내의 브랜드에 속하는 제품 탐색은 분석에 반영할 수 없었다. 만약 유저 중심의 클릭스트림 데이터를 활용할 수 있다면, 탐색 유형에 대한 타당성을 더 확보하는 동시에, 탐색 패턴에 대한 더 깊이 있는 분석이 가능할 것이다. 두 번째 한계점으로는 유저의 접속 디바이스에 대한 영향을 고려하지 않았다는 점을 들 수 있다. 모바일 기기에서는 PC에서 웹사이트에 유입되었을 때와는 다른 행동 패턴과 정보 처리 과정을 거치게 된다(Ghose et al, 2012; Raphaeli et al., 2017). 따라서 디바이스에 따라 어떤 탐색 유형이 빈번하게 나타나며, 같은 사용자여도 각 디바이스를 어느 목적에 따라 사용하는지를 패턴 분석을 통해 탐색해볼 수 있을 것이다. 마지막으로 유저의 인구통계학적 특성과 RFM(Recency, Frequency, Monetary)과 같은 다양한 기준으로 유저 세분화 그룹 별로 탐색 유형과 그 패턴에 대한 분석을 진행된다면, 더 넓은 관점에서 온라인 소비자의 제품 탐색 및 구매 행동에 대한 이해가 가능할 것이다.

참고문헌(References)

- Agrawal, Rakesh, and Ramakrishnan Srikant. "Mining sequential patterns.", *Data Engineering, 1995. Proceedings of the Eleventh International Conference on.* IEEE, 1995.
- Ba, Sulin, Andrew B. Whinston, and Han Zhang. "Building trust in online auction markets through an economic incentive mechanism.", *Decision Support Systems*, Vol.35, No.3 (2003), 273~286.
- Babin, Barry J., William R. Darden, and Mitch Griffin. "Work and/or fun: measuring hedonic and utilitarian shopping value.", *Journal of consumer research*, Vol.20, No.4(1994), 644~656.
- Bae, S., and Park, D., "The Effect of Mobile Advertising Platform through Big Data Analytics: Focusing on Advertising, and Media Characteristics" *Journal of Intelligence and Information Systems*, Vol24, No.2(2018), 37~57.
- Bettman, James R., Mary Frances Luce, and John W. Payne. "Constructive consumer choice processes.", *Journal of consumer research*, Vol.25, No3 (1998), 187~217.
- Bradlow, Eric T., et al. "The role of big data and predictive analytics in retailing.", *Journal of Retailing*, Vol.93, No.1(2017), 79~95.
- Bridges, Eileen, and Renée Florsheim. "Hedonic and utilitarian shopping goals: The online experience.", *Journal of Business research*, Vol.61, No.4 (2008), 309~314.
- Bronnenberg, B. J., Kim, J. B., & Mela, C. F. "Zooming in on choice: How do consumers search for cameras online?", *Marketing Science*, Vol.35, No.5 (2016), 693~712.

- Bucklin, Randolph E., and Catarina Sismeiro. "Click here for Internet insight: Advances in clickstream data analysis in marketing.", *Journal of Interactive marketing*, Vol.23, No.1 (2009), 35~48.
- Bucklin, Randolph E., et al. "Choice and the Internet: From clickstream to research stream.", *Marketing Letters*, Vol.13, No.3 (2002), 245~258.
- Chiu, Chao-Misn, et al. "Re-examining the influence of trust on online repeat purchase intention: The moderating role of habit and its antecedents.", *Decision Support Systems*, Vol.53, No.4(2012), 835~845.
- Eirinaki, Magdalini, and Michalis Vazirgiannis. "Web mining for web personalization.", *ACM Transactions on Internet Technology (TOIT)*, Vol3, No.1(2003), 1~27.
- Elkan, Charles. "Using the triangle inequality to accelerate k-means.", *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, (2003) 147~153.
- Engel, James F., Roger D. Blackwell, and Paul W. Miniard, *Consumer Behavior*, Chicago: Dryden, 1993.
- Ganesh, Jaishankar, et al. "Online shopper motivations, and e-store attributes: an examination of online patronage behavior and shopper typologies.", *Journal of retailing*, Vol.86, No.1(2010), 106~115.
- Ghose, A., Goldfarb, A., & Han, S. P. "How is the mobile Internet different? Search costs and local activities.", *Information Systems Research*, Vol.24, No.3(2012), 613~631.
- Girard, Tulay, and Paul Dion. "Validating the search, experience, and credence product classification framework.", *Journal of Business Research*, Vol.63, No.9-10 (2010), 1079~1087.
- Goh. K, Y., Chu. J., and Wu. J., "Mobile advertising: an empirical study of temporal and spatial differences in search behavior and advertising response.", *Journal of Interactive Marketing*, Vol.30(2015), 34~45.
- Guerbas, Abdelghani, et al. "Effective web log mining and online navigational pattern prediction.", *knowledge-based systems*, Vol.49 (2013), 50~62.
- Ha, Y., and Stephen J., H., "Ambiguity, Processing Strategy, and Advertising-Evidence Interactions," *Journal of Consumer Research*, Vol.16, No.3(1989), 354~60.
- Holbrook, Morris B., and Elizabeth C. Hirschman. "The experiential aspects of consumption: Consumer fantasies, feelings, and fun.", *Journal of consumer research*, Vol.9, No.2 (1982), 132~140.
- Hong, S., and Kar Yan Tam. "Understanding the adoption of multipurpose information appliances: The case of mobile data services.", *Information systems research*, Vol.17, No.2 (2006), 162~179.
- Huang, Peng, Nicholas H. Lurie, and Sabyasachi Mitra. "Searching for experience on the web: an empirical examination of consumer behavior for search and experience goods.", *Journal of marketing*, Vol.73, No.2 (2009), 55~69.
- Howard, J., A., and Sheth. J., N., *The theory of buyer behavior*. New York, Wiley. 1969.
- Huang, Peng, Nicholas H. Lurie, and Sabyasachi Mitra. "Searching for experience on the web: an empirical examination of consumer behavior for search and experience goods.", *Journal of marketing*, Vol.73, No.2 (2009), 55~69.

- Huang, Yueh-Min, et al. "NP-miner: A real-time recommendation algorithm by using web usage mining.", *Knowledge-Based Systems*, Vol.19, No.4(2006), 272~286.
- Jacoby, Jacob, et al. "Tracing the impact of item-by-item information accessing on uncertainty reduction.", *Journal of Consumer Research*, Vol.21, No.2(1994), 291~303.
- Jain, A., K., "Data clustering: 50 years beyond K-means.", *Pattern recognition letters*, Vol.31, No.8 (2010), 651~666.
- Janiszewski, Chris. "The influence of display characteristics on visual exploratory search behavior.", *Journal of Consumer Research*, Vol.25, No.3(1998), 290~301.
- Jansen, Bernard J., and Simone Schuster. "Bidding on the buying funnel for sponsored search and keyword advertising.", *Journal of Electronic Commerce Research*. Vol.12, No.1 (2011), 1.
- Karimi, S., Holland, C. P., and Papamichail, K. N. "The impact of consumer archetypes on online purchase decision-making processes and outcomes: A behavioural process perspective.", *Journal of Business Research*, Vol.91(2018), 71~82.
- Karimi, Sahar, K. Nadia Papamichail, and Christopher P. Holland. "The effect of prior knowledge and decision-making style on the online purchase decision-making process: A typology of consumer shopping behaviour.", *Decision Support Systems*, Vol.77(2015), 137~147.
- Kim, Y.S, and Krishnan R., "On product-level uncertainty and online purchase behavior: An empirical analysis.", *Management Science*, Vol.61, No.10(2015), 2449~2467.
- KOSIS, "2018 online shopping trends", 2018, http://kostat.go.kr/portal/korea/kor_nw/2/1/index.board?bmode=read&aSeq=370428
- Kukar-Kinney, Monika, and Angeline G. Close. "The determinants of consumers' online shopping cart abandonment.", *Journal of the Academy of Marketing Science*, Vol.38, No.2(2010), 240~250.
- Lee, D., "A Regression-Model-based Method for Combining Interestingness Measures of Association Rule Mining", *Journal of Intelligence and Information Systems*, Vol.23, No.1(2017), 127~141.
- Liaukonyte, Jura, Thales Teixeira, and Kenneth C. Wilbur. "Television advertising and online shopping.", *Marketing Science*, Vol.34, No.3(2015), 311~330.
- Liu, Fang, et al. "A Typology Of Online Window Shopping Consumers." *PACIS*. 2012, 128.
- Liu, Haibin, and Vlado Kešelj. "Combined mining of Web server logs and web contents for classifying user navigation patterns and predicting users' future requests.", *Data & Knowledge Engineering*, Vol.61, No.2(2007), 304~330.
- Lu, Jie, et al. "Recommender system application developments: a survey.", *Decision Support Systems*, Vol.74(2015), 12~32.
- Mallapragada, Girish, Sandeep R. Chandukala, and Qing Liu. "Exploring the effects of "What"(product) and "Where"(website) characteristics on online shopping behavior.", *Journal of Marketing*, Vol.80, No.2 (2016), 21~38.
- Moe, W. W., & Fader, P. S., "Dynamic conversion behavior at e-commerce sites.", *Management Science*, Vol.50, No.3(2004) 326-335.

- Moe, W. W., "An empirical two-stage choice model with varying decision rules applied to internet clickstream data.", *Journal of Marketing Research*, Vol.43, No.4 (2006), 680~692.
- Moe, W. W., "Buying, searching, or browsing: Differentiating between online shoppers using in-store navigational clickstream.", *Journal of consumer psychology*, Vol.13, No.1(2003), 29~39.
- Montgomery, A. L., et al. "Modeling online browsing and path analysis using clickstream data.", *Marketing science*, Vol.23, No.4(2004), 579~595.
- Nelson, Phillip. "Advertising as information.", *Journal of political economy*, Vol.82, No.4 (1974), 729~754.
- Nelson, Phillip. "Information and consumer behavior.", *Journal of political economy*, Vol.78, No.2(1970), 311~329.
- Pabarskaite, Zidrina, and Aistis Raudys. "A process of knowledge discovery from web log data: Systematization and critical review.", *Journal of Intelligent Information Systems*, Vol.28, No.1(2007), 79~104.
- Pallant, J. I., et al. "An empirical analysis of factors that influence retail website visit types.", *Journal of Retailing and Consumer Services*, Vol.39(2017), 62~70.
- Park J., and Cho Y., "Clickstream Big Data Mining for Demographics based Digital Marketing", *Journal of Intelligence and Information Systems*, Vol22, No.3(2016), 143~163.
- Pavlou, P. A., and Fygenon, M., "Understanding and predicting electronic commerce adoption: An extension of the theory of planned behavior.", *MIS quarterly*, Vol.30, No.1(2006), 115~143.
- Peck, J., and Childers, T. L., "To have and to hold: The influence of haptic information on product judgments.", *Journal of Marketing*, Vol.67, No.2(2003), 35~48.
- Pei, Jian, et al. "Mining sequential patterns by pattern-growth: The prefixspan approach.", *IEEE Transactions on Knowledge & Data Engineering*, Vol.11(2004), 1424~1440.
- Phang, C. W., Kankanhalli, A., Ramakrishnan, K., & Raman, K. S. "Customers' preference of online store visit strategies: an investigation of demographic variables." *European Journal of Information Systems*, Vol.19, No.3(2010), 344~358.
- Raphaeli, O., Goldstein, A., and Fink, L., "Analyzing online consumer behavior in mobile and PC devices: A novel web usage mining approach.", *Electronic Commerce Research and Applications*, Vol.26(2017), 1~12.
- Rohm, A. J., and Swaminathan, V., "A typology of online shoppers based on shopping motivations.", *Journal of business research*, Vol.57, No.7(2004), 748~757.
- Scarpi, D., "Work and fun on the internet: the effects of utilitarianism and hedonism online.", *Journal of interactive marketing*, Vol.26, NO.1(2012), 53~67.
- Sculley, David, "Web-scale k-means clustering.", *Proceedings of the 19th international conference on World wide web*. ACM, 2010.
- Sismeiro, C., and Bucklin, R. E., "Modeling purchase behavior at an e-commerce web site: A task-completion approach.", *Journal of marketing research*, Vol.41, No.3(2004), 306~323.

Statista, “*Conversion rate of online shoppers in the United States as of 2nd quarter 2018*”, 2018, <https://www.statista.com/statistics/439558/us-online-shopper-conversion-rate/>

Statista, “*Retail e-commerce sales worldwide from 2014 to 2021 (in billion U.S. dollars)*” 2018, <https://www.statista.com/statistics/379046/worldwide-retail-e-commerce-sales/>

Abstract

Analysis of shopping website visit types and shopping pattern

Kyungbin Choi* · Kihwan Nam

Online consumers browse products belonging to a particular product line or brand for purchase, or simply leave a wide range of navigation without making purchase. The research on the behavior and purchase of online consumers has been steadily progressed, and related services and applications based on behavior data of consumers have been developed in practice. In recent years, customization strategies and recommendation systems of consumers have been utilized due to the development of big data technology, and attempts are being made to optimize users' shopping experience. However, even in such an attempt, it is very unlikely that online consumers will actually be able to visit the website and switch to the purchase stage. This is because online consumers do not just visit the website to purchase products but use and browse the websites differently according to their shopping motives and purposes. Therefore, it is important to analyze various types of visits as well as visits to purchase, which is important for understanding the behaviors of online consumers.

In this study, we explored the clustering analysis of session based on click stream data of e-commerce company in order to explain diversity and complexity of search behavior of online consumers and typified search behavior. For the analysis, we converted data points of more than 8 million pages units into visit units' sessions, resulting in a total of over 500,000 website visit sessions. For each visit session, 12 characteristics such as page view, duration, search diversity, and page type concentration were extracted for clustering analysis. Considering the size of the data set, we performed the analysis using the Mini-Batch K-means algorithm, which has advantages in terms of learning speed and efficiency while maintaining the clustering performance similar to that of the clustering algorithm K-means. The most optimized number of clusters was derived from four, and the differences in session unit characteristics and purchasing rates were

* Department of Business Informatics, Hanyang University

** Corresponding Author: Kihwan Nam

Information Systems, Management Engineering,

College of Business, Korea Advanced Institute of Science and Technology (KAIST)

85 Hoegi-Ro, Dongdaemooon-Gu, Seoul, 130-722, Korea

E-mail: namkh@kaist.ac.kr

identified for each cluster.

The online consumer visits the website several times and learns about the product and decides the purchase. In order to analyze the purchasing process over several visits of the online consumer, we constructed the visiting sequence data of the consumer based on the navigation patterns in the web site derived clustering analysis. The visit sequence data includes a series of visiting sequences until one purchase is made, and the items constituting one sequence become cluster labels derived from the foregoing. We have separately established a sequence data for consumers who have made purchases and data on visits for consumers who have only explored products without making purchases during the same period of time. And then sequential pattern mining was applied to extract frequent patterns from each sequence data. The minimum support is set to 10%, and frequent patterns consist of a sequence of cluster labels. While there are common derived patterns in both sequence data, there are also frequent patterns derived only from one side of sequence data. We found that the consumers who made purchases through the comparative analysis of the extracted frequent patterns showed the visiting pattern to decide to purchase the product repeatedly while searching for the specific product.

The implication of this study is that we analyze the search type of online consumers by using large - scale click stream data and analyze the patterns of them to explain the behavior of purchasing process with data-driven point. Most studies that typology of online consumers have focused on the characteristics of the type and what factors are key in distinguishing that type. In this study, we carried out an analysis to type the behavior of online consumers, and further analyzed what order the types could be organized into one another and become a series of search patterns. In addition, online retailers will be able to try to improve their purchasing conversion through marketing strategies and recommendations for various types of visit and will be able to evaluate the effect of the strategy through changes in consumers' visit patterns.

Key Words : E-Commerce, Online consumer behavior, Web usage mining, Sequential pattern mining, Clickstream data

Received : November 16, 2018 Revised : March 6, 2019 Accepted : March 11, 2019

Publication Type : Regular Paper Corresponding Author : Kihwan Nam

저 자 소개



최경빈

한양대학교 경영대학에서 비즈니스인포매틱스 석사 학위를 취득하였다. 데이터 기반의 온라인 소비자 행동 분석과 빅데이터 분석, 데이터 마이닝, 머신러닝과 딥러닝 활용에 대한 연구에 관심을 두고 있다.



남기환

KAIST 경영대학원 경영공학부에서 MIS 박사학위를 취득하였다. 현재 KAIST 경영대학원 경영공학부, UNIST 경영학부 겸임교수로 재직 중이다. 주요 관심분야는 Business Analytics & Business Intelligence, Big Data Analytics, Data Mining, Statistical Analysis, Recommender Systems, Econometrics Models, Machine Learning, Deep Learning 등이다. 관련 연구들은 Decision Support Systems, Data Mining and Knowledge Discovery 등에 논문이 게재되었다. 학문적인 연구뿐만 아니라 이론을 바탕으로 실제 기업에서 다양한 프로젝트를 성공적으로 진행함으로써 학계와 산업계 모두에 실증적인 기여를 하고 있다.