

프로세스 마이닝을 이용한 쇼핑몰 웹로그 데이터 분석

김채영, 용혜련, 황현석*
한림대학교 인터랙션디자이너

Analyzing the weblog data of a shopping mall using process mining

Chae-Young Kim, Hye-Ryeon Yong, Hyun-Seok Hwang*
Graduation school of Interaction Design, Hallym University, Republic of Korea

요약 인터넷의 발전과 모바일 기기 보급의 확산으로 온라인 시장이 급속하게 성장하였다. 특히 쇼핑몰 이용이 폭발적으로 증가함에 따라 데이터를 활용한 이용자 행태 분석, 개인화된 상품 추천 및 서비스 개발 등의 연구가 이루어지고 있다. 이에 본 논문은 프로세스 마이닝을 통해 온라인 쇼핑몰의 전반적인 프로세스를 분석하고, 사용자의 구매에 영향을 미치는 요소를 파악하고자 하였다. 분석에는 대형 온라인 쇼핑몰인 모 기업의 데이터를 사용하였으며 분석 도구로는 R을 활용하였다. 분석 결과 가격세일, 월경품행사와 같은 이벤트 요소를 가진 카테고리에서의 고객 활동이 가장 두드러졌다. 이에 반해 검색, 로그인, 캠페인 액티비티는 중요도에 비해 적절한 활동이 이루어지지 않은 것으로 나타났다. 해당 액티비티는 고객의 정보와 니즈를 파악할 수 있는 단서가 될 수 있어 매우 중요하다. 따라서 연관검색어 추천의 정교화, 로그인 시 제공되는 쿠폰 등의 액티비티 관리가 필요하다고 사료된다. 본 논문에서는 앞서 논의된 내용 이외에도 쇼핑몰의 경쟁력 제고 및 이윤 증대를 위한 다양한 비즈니스 전략을 제안한다.

Abstract With the development of the Internet and the spread of mobile devices, the online market is growing rapidly. As the number of customers using online shopping malls explodes, research is being conducted on the analysis of usage behavior from customer data, personalized product recommendations, and service development. Thus, this paper seeks to analyze the overall process of online shopping malls through process mining, and to identify the factors that influence users' purchases. The data used are from a large online shopping mall, and R was the analysis tool. The results show that customer activity was most prominent in categories with event elements, such as unconventional discounts and monthly giveaway events. On the other hand, searches, logins, and campaign activity were found to be less relevant than their importance. Those are very important, because they can provide clues to a customer's information and needs. Therefore, it is necessary to refine the recommendations from related search words, and to manage activity, such as coupons provided when customers log in. In addition to the previous discussion, this paper proposes various business strategies to enhance the competitiveness of online shopping malls and to increase profits.

Keywords : Weblog, Process mining, BupaR, Data mining, Shopping mall

본 논문은 2019년도 한림대학교 교비 학술연구비(HRF-201910-005)에 의하여 연구되었음.

*Corresponding Author : Hyun-Seok Hwang(Hallym Univ.)

email: hshwang@hallym.ac.kr

Received July 15, 2020

Accepted November 6, 2020

Revised August 13, 2020

Published November 30, 2020

1. 서론

네트워크 환경의 급속한 성장과 확장으로 인해 많은 서비스가 웹사이트를 통해 제공되고 있다. 그중 대표적인 서비스라고 할 수 있는 것이 온라인 쇼핑이다. 다양한 채널의 등장으로 온라인 쇼핑을 하는 이용자들의 선택 폭이 넓어졌는데[1], 특히 스마트폰, 태블릿 등 모바일 기기 보급의 활성화는 온라인 쇼핑의 경쟁력을 높이는 중요한 요인이 되었다[2].

한편, 온라인 쇼핑몰에서 판매하는 상품이 다양해짐에 따라 데이터베이스의 종류와 양도 폭발적으로 증가하였다[3,4]. 방대한 데이터의 적절한 분석은 쇼핑몰을 보다 효율적으로 운영하는데 있어 유용한 역할을 할 수 있다[5].

정보화된 데이터는 단순한 상품 판매를 넘어 고객을 집단별로 분류하는 것을 가능하게 한다. 또한, 분류된 고객의 특성과 선호도를 파악하여 맞춤 상품을 추천할 수 있다[6]. 이를 바탕으로 하는 개인화된 상품 제공은 고객이 지각하는 가치의 정도를 높일 수 있으며, 나아가 상품을 구매하는 행동으로 이어지게 한다[7]. 이병엽 외(2010)의 연구에서는 데이터 마이닝을 기반으로 자동차 산업에서의 고객 집단 분류 및 속성 파악을 통해 구매패턴을 파악하고자 하였으며, 이를 바탕으로 기업의 전략적 마케팅을 수립하고자 하였다[8]. 김지혜 외(2004)는 전자상거래의 고객 개인화를 위해 연관규칙 기법을 바탕으로 효율적인 상품추천시스템을 설계하고 구현하였다[9]. 이처럼 데이터를 활용한 비즈니스 전략 수립은 차별성, 관계지향성 등의 역량 및 이윤을 증진시킬 수 있다.

온라인 쇼핑몰이 경쟁력을 가지기 위해서는 앞서 언급한 것처럼 고객의 성향 및 특성 파악, 구매패턴 예측, 개인화된 추천 시스템 등의 이용 패턴 분석이 매우 중요하다. 그러나 기존의 연구들을 살펴보면 전체적인 이용 패턴의 파악보다는 쇼핑몰 만족도, 인터페이스 사용성 평가 등과 같은 특정 주제의 연구가 주를 이루는 실정이다.

이에 본 연구에서는 온라인 쇼핑몰 웹로그 데이터를 활용하여 웹사이트 사용자들의 전반적인 사용 프로세스를 분석하고자 하였으며, 특히 사용자의 구매패턴에 영향을 미치는 다양한 요소를 파악하고자 하였다.

본 논문은 다음과 같이 구성되어 있다. 2장에서는 프로세스 마이닝과 웹로그 데이터 분석에 대한 정의 및 기존 연구를 소개한다. 3장에서는 분석할 데이터를 파악하고 연구절차에 따른 전처리 과정을 설명한다. 4장에서는 도출된 연구결과에 대하여 논의한다. 마지막으로 5장에

서는 본 연구가 갖는 함의 및 실무적인 시사점에 대해 논의하고 추후 연구 방향을 제시한다.

2. 이론적 배경

2.1 프로세스 마이닝

프로세스 마이닝(Process Mining)이란 이벤트 로그를 포함하고 있는 업무수행 데이터를 자동으로 분석하고, 워크플로우를 시각화함으로써 비즈니스 프로세스의 전반적인 모습을 직관적으로 빠르게 이해할 수 있도록 하는 분석 기법이다[10]. 프로세스 마이닝을 통한 비즈니스 분석은 기업의 생산비용 절감, 인력 및 자원 낭비 제거, 가치혁신 등과 같은 문제를 해결하는 데 도움을 줄 수 있다[11].

프로세스에서 발생한 이벤트를 분석하는 방법에는 통계, 딥러닝, 인공지능, 소셜 네트워크 등의 기법들이 있으며 분석을 위한 툴의 개발도 증가하고 있다. 이처럼 다양한 프로세스 마이닝 기법들을 통해 발견된 프로세스의 특징 및 모델은 관리자의 효율적인 업무를 위해 활용되기도 하며, 추후 프로세스 개선을 위한 지표가 되기도 한다[12].

프로세스 마이닝은 프로세스 모델링 및 분석과 데이터 마이닝 두 가지가 근간이 된다. 이러한 접근은 작업 결과의 평가에 유용할 수 있으며, 프로세스 모델의 품질을 측정 및 개선하는 데 도움이 된다. 즉, 작업의 전체적인 흐름을 발견(Process Discovery)하거나 프로세스 내의 특정 작업이 적합한지를 검증하는 적합성 검사(Conformance Checking)와 프로세스 모델의 확장(Extension)을 가능하게 한다[13]. 김현아(2019)는 프로세스 마이닝 기법을 이용하여 개방형 교육 플랫폼에서 제공하는 학습 데이터를 분석하여 개선된 학습 프로세스 모델을 발견하고자 하였다[14]. 또한, 박중구(2019)의 연구에서는 프로세스 마이닝을 활용하여 해양플랜트 배관재 제작 공정을 관리하는 방법을 연구하였다. 이를 통해 이상 프로세스의 패턴을 발견하고 분석하여 공정 모니터링 시스템에 적용 및 응용하였다[15]. Poggi와 Khalaf(2013)의 연구에서는 대규모 온라인 여행 및 예약 대행사의 웹로그 데이터를 활용하여 고객의 프로세스를 파악하고자 하였으며, 특정 이탈지점 및 주요 경로 등을 파악하고자 하였다[16].

온라인 쇼핑몰의 특성상 대부분의 프로세스는 매우 복잡하며 임시적이다[17]. 따라서 이 분야에서의 프로세스 마이닝을 통해 시스템을 개발하는 것은 한계를 가질 수

밖에 없다. 그럼에도 불구하고 많은 온라인 쇼핑몰들이 프로세스 마이닝 기법을 활용하여 성공적으로 구매 프로세스를 파악하고자 하며, 이를 통해 보다 나은 비즈니스 수익 모델로 개선하고자 하는 시도가 증가하고 있다[18].

2.2 웹로그 데이터 분석

웹사이트를 이용하면 이용자가 해당 사이트에서 활동한 결과가 웹로그(Weblog)라는 형태로 기록된다[19]. 웹로그란, 웹사이트 방문자들이 제품이나 서비스를 구매하는 과정을 통해 발생하는 데이터이다. 웹로그를 수집하는 방법에는 자바스크립트를 통한 페이지 태깅(Page Tagging), 웹 서버 로그 페이지에 원격 객체 삽입, 패킷 스니핑(Packet Sniffing) 등이 존재한다[16]. 이렇게 수집된 웹로그 데이터에는 방문 수, URL 액세스 빈도, 지리적 위치, 클라이언트 웹 등의 고유한 정보가 포함되어 있다.

한편, 웹 서버에 축적되는 대용량의 로그 데이터로부터 데이터 마이닝 기술을 도입하여 유용한 정보를 발견하고 새로운 패턴을 찾으려는 시도가 이루어지고 있는데, 이처럼 웹로그를 활용하여 의미 있는 정보를 찾아내는 일련의 활동을 웹 마이닝이라고 한다[20]. 장재영의 연구에서는 온라인 쇼핑몰 리뷰를 오피니언 마이닝을 활용하여 분석하고, 감성분석 알고리즘을 제시하였다[21]. 김성근과 김영철의 연구에서는 쇼핑몰 웹로그를 통해 고객의 가치를 판단하고, 다양한 형태의 마케팅 전략을 구사할 수 있는 마케팅로그 기술을 소개하였다[22]. 또 다른 연구에서는 온라인 쇼핑몰 데이터를 활용하여 마케팅 분석의 목적인 매출 증대에 관한 방안을 모색하고자 하였으며, 쇼핑몰을 위한 시스템 구성, 서비스, 마케팅 및 판매 등에 활용하고자 하였다[23].

웹로그 데이터를 활용한 고객 행동 분석은 적절한 마케팅 방향 및 개인화된 추천을 제안할 수 있으며, 이탈 및 전환 고객의 특정 지점을 파악할 수 있게 한다[24]. 또한, 사이트 사용성에 대한 평가도 가능하기 때문에 추후 인터페이스를 개선하는 데 도움이 될 수 있다[25].

본 연구에서는 위의 이론적 배경을 바탕으로 다음과 같은 연구가설을 도출하였다.

- RQ1. 온라인 쇼핑몰의 전반적인 프로세스의 흐름은 어떠한가?
- RQ2. 온라인 쇼핑몰을 이용하는 데 있어 성별 프로세스의 차이가 존재하는가?
- RQ3. 온라인 쇼핑몰을 이용하는 데 있어 연령별 프로세스의 차이가 존재하는가?

3. 연구절차

본 연구는 쇼핑몰 웹로그 데이터를 분석하고자 하였으며, 이를 위해 프로세스 마이닝을 통해 전반적인 쇼핑몰 이용행태를 파악하고자 하였다. 본 연구에서 분석한 사례의 연구절차는 Fig. 1과 같다.

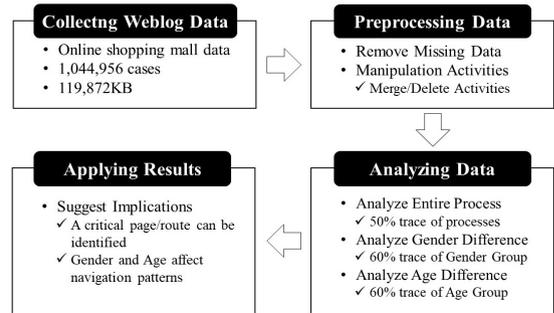


Fig. 1. Procedure of a Case Study

3.1 데이터 수집

본 연구는 Fig. 1과 같은 연구절차를 통하여 수행되었다. 데이터는 모 G기업의 온라인 쇼핑몰 웹로그 데이터를 활용하였으며, 2018년 8월부터 약 한 달간의 데이터를 사용하였다.

쇼핑몰 웹로그 데이터는 열 가지의 변인 정보를 바탕으로 총 1,044,956건의 데이터가 수집되었다. 수집된 변인 중 '상품 ID'와 '검색어키워드타입' 변인 내에는 'NA' 값이 대부분이었으며, 분석에 중요한 변인이 될 수 있는 '성별' 및 '연령' 변인에도 '모름'이라는 응답이 다수 존재하여 데이터 셋의 전처리가 필요하다고 사료되었다.

3.2 데이터 전처리

3.2.1 행과 열 제거

먼저 '성별' 및 '연령'은 쇼핑몰의 이용행태를 특정 그룹별로 구분 및 파악하는데 중요한 변인이 될 수 있다고 판단하여 '모름'이라고 응답한 행 데이터를 모두 제거하였다. 그 결과, '성별' 및 '연령' 변인 각각 86,131건과 205건이 제거되어 총 958,620건의 데이터 셋으로 구축되었다.

또한 '상품 ID'와 '검색어키워드타입' 변인 내에 존재하는 수많은 'NA' 값을 대체할 만한 다른 값을 입력하는 것은 불가능하다고 판단하였으며, 두 변인이 쇼핑몰의 전반적인 프로세스를 파악하는데 있어 크게 영향을 주지

않는다고 사료되어 해당 변인의 열 데이터를 모두 제거하였다.

3.2.2 전처리

본 연구에서 매우 중요한 변인 중 하나는 액티비티(Activity) 값이다. 전처리 전의 액티비티는 총 44개이며 그 수는 백만 건 이상이다. 그러나 전처리 전의 모든 Activity를 프로세스 맵에 나타내기에는 무리가 있다. 이에 직관적으로 프로세스 맵을 이해하기 위해 Table. 1과 같이 재 카테고리화 하였다.

Table 1. Activities Re-categorization

No	New Activity	Existing Activity	# Case
1	Main_일반	main_식품 main_자주구매 main_인테리어 main_생활용품 main_시즌상품	26,480
2	Main_특별	main_파격세일 main_월경품행사 main_오늘의행사	78,618
3	검색	searchMobile	71,539
4	로그인	login_initLogin snslogin login	8,019
5	장바구니담기	product_addTocart	111,441
6	장바구니삭제	product_cartRemoval order_deleteBasket	54,062
7	장바구니주문	order_basket	101,228
8	주문시작	order_initOrder	16,874
9	주문완료	order_결제완료	7,211
10	캠페인	main_쿠폰이벤트 event_createEventBasketCoupon specialEvent_roulette specialEvent_retrieveStampEvent	19,001
Total: 494,473			

재 카테고리화의 첫 번째는 분석 결과와 관련이 없는 변인을 제거하는 것이다. 기존에 가장 많은 케이스 수를 차지했던 'main_홈화면(178,488건)' 변인은 쇼핑물 데이터에 접속만 한다면 무조건 집계되는 활동이기 때문에 의미 있는 결과와는 관련이 없다고 생각되어 제거하였다.

두 번째는 결과와 관련 없는 Activity를 정리하는 것이다. 쇼핑물 프로세스 분석에서 중요하다고 생각되는 결과변인은 'order_결제완료'인데, 결제완료에 도달하기까지 중요한 활동이라고 보기 어려운 변인 또한 제거하였다.

세 번째는 비슷한 속성을 가진 여러 액티비티를 하나

의 액티비티로 나타내고자 하였다. 예를 들어 기존의 카테고리인 'main_식품', 'main_자주구매', 'main_인테리어' 등은 쇼핑물 사용자 또는 관계자들이 상품을 찾고 관리하기 편리하도록 구분해 놓은 일반적인 카테고리라고 볼 수 있다. 따라서 위의 액티비티는 'main_일반'이라고 재 카테고리화 하였다. 마찬가지로 'main_파격세일', 'main_월경품행사', 'main_오늘의행사' 카테고리는 이벤트 요소가 강한 액티비티라고 볼 수 있으므로 'main_특별'이라고 재 카테고리화 하였다.

마지막으로 데이터만으로 설명하기 어려운 액티비티를 제거하였다. 예를 들어, 기타 발생 로그를 나타내는 변인은 설명만으로 프로세스의 흐름을 파악하고 이해하기에는 어려움이 있다. 이렇듯 프로세스 과정에서 이해하기 어려운 액티비티는 제거하였다. 일련의 과정을 통해 44개의 액티비티는 10개의 액티비티로 축소되었다.

4. 데이터 분석

본 연구에서는 프로세스 마이닝을 이용하여 쇼핑물 웹 로그 데이터를 분석하고자 하였다. 분석에 사용된 주된 도구는 R 프로그램의 BupaR이다. BupaR은 비즈니스 프로세스 데이터 처리 및 분석을 위한 오픈소스 통합 R 패키지로, 주요 패키지를 포함하여 여덟 개의 패키지를 제공하며, 프로세스 마이닝의 워크플로우를 제공한다.

웹로그 데이터는 트레이스가 매우 복잡하며 일관적인 사용 패턴을 찾아내기 어려운 특성이 있다. 따라서 다수의 트레이스를 포함하여 결과를 도출하고, 이를 이해하는 것은 매우 어렵다. 또한, 전체적인 프로세스를 특정한 기준 없이 프로세스 맵 상에 모두 나타낼 경우, 직관적인 이해를 돕는 데 한계가 있다. 이에 본 연구에서는 전체 중 50%의 트레이스를 보고자 하였으며 이에 따른 프로세스 맵을 그려보았다. 해당 트레이스에 대한 결과는 Fig. 2와 같다.

50% 트레이스 분석으로 다음과 같은 몇 가지의 결과를 도출할 수 있다.

첫 번째, 메인 카테고리 일반과 특별 중 특히 이벤트적인 요소를 가진 'main_특별' 액티비티가 'main_일반' 카테고리보다 훨씬 굵은 트레이스를 가진다. 즉, 시작 지점에서부터 끝나는 지점까지 모든 단계에서 매우 중요한 역할을 하고 있다. 이는 일반적인 카테고리보다는 특별한 요소를 가진 카테고리가 쇼핑물 사용자들의 이목을 집중시킨다고 볼 수 있다.

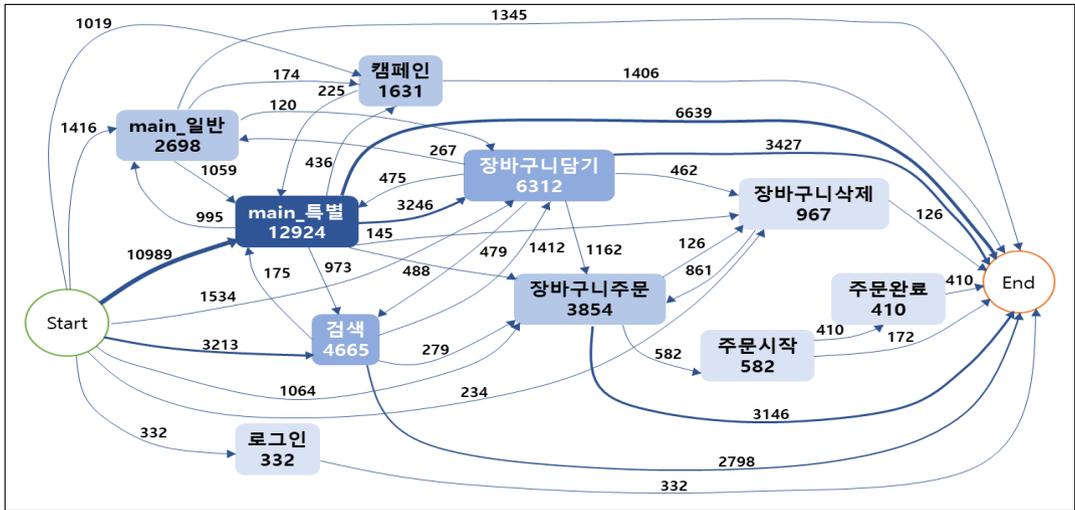


Fig. 2. 50% trace

두 번째, ‘main_특별’ 액티비티만큼 중요한 것은 ‘검색’이다. 특정한 상품이 필요한 사용자의 경우, 검색하여 상품을 구매하는 경우가 많아 ‘검색’ 액티비티도 구매로 이어지는 데 있어 매우 중요한 허브라고 할 수 있다.

세 번째, 회원과 비회원 간 차이가 없는 것으로 보인다. Fig. 2를 보면 ‘로그인’ 액티비티가 시작 지점과 끝 지점으로만 연결될 뿐, 다른 액티비티와의 관계는 찾아볼 수 없다. 즉, 회원 관리나 특별한 혜택이 없다고 볼 수 있으며, 있다고 하더라도 그 혜택은 구매에 별다른 영향을 미치지 못하는 것으로 보인다.

마지막으로, 캠페인의 적절성이다. ‘캠페인’ 액티비티는 장바구니 할인이나 스탬프 이벤트와 같이 물건 구매 시 사용할 수 있는 쿠폰이다. 다른 액티비티에서는 ‘캠페인’ 액티비티로 대부분 이어지나, ‘캠페인’ 액티비티에서 다른 액티비티로는 ‘main_특별’을 제외하고는 이어지지 않는 것을 볼 수 있다. 이는 현재 진행 중인 캠페인이 사용자들의 관심을 끌지 못한다고 볼 수 있다.

위의 분석을 통하여 전체 중 특히 시작 지점이 ‘main_특별’ 카테고리로 시작하는 트레이스가 많으므로, 시작 지점으로 지정하여 60%의 트레이스를 파악하였다. 그 결과, ‘장바구니담기’가 구매를 하는 데 있어 중요한 액티비티라는 것을 발견하였다. 따라서 추후 ‘장바구니담기’와 관련된 액티비티에서 구매로 이어질 수 있는 이벤트가 필요할 것으로 사료된다.

4.1 성별 프로세스 비교

다음은 성별에 따라 쇼핑몰 이용에 차이가 있는지 보

기 위하여 남녀별 분석을 진행하였다.

수집된 데이터는 남성 집단이 81,111건, 여성 집단 413,362건으로, 여성이 남성보다 월등히 높은 수치로 쇼핑몰을 이용한다는 것을 알 수 있다. 또한, 남녀 각각의 트레이스는 1,904건, 5,565건으로 여성이 남성보다 많은 트레이스를 가지는 것으로 나타났다.

수집된 데이터 내 성별의 차이가 큰 이유 외에도, 기존 문헌에 따르면 쇼핑몰 이용 및 상품 구매에 있어 여성이 남성보다 훨씬 높은 비율로 다양하게 온라인 쇼핑몰을 이용한다[26]. 따라서 본 연구에서도 성별에 따른 구매 패턴 프로세스를 파악하고자 하였다.

우선 특별한 시작 지점이나 끝 지점을 지정하지 않고 50%에 해당하는 트레이스를 남녀별로 분석하였다. 분석 결과는 다음과 같다.

첫 번째, 두 집단 모두 ‘main_특별’ 액티비티가 눈에 띄는 것으로 나타났다. 이는 앞서 결과와 마찬가지로 메인 카테고리 중 이벤트적인 요소가 있는 카테고리가 중요하다는 것을 뜻한다.

두 번째, ‘장바구니삭제’와 관련된 트레이스에서 차이를 보인다. 남성 집단의 경우 ‘장바구니담기→장바구니삭제→장바구니주문’의 트레이스를 보인다. 그러나 여성 집단의 경우 ‘장바구니담기→장바구니주문→장바구니삭제’의 트레이스를 보인다. 즉, 남성은 장바구니에 상품을 담은 후 주문하지 않을 상품을 삭제한 후, 담겨있는 상품을 주문한다. 그러나 여성은 장바구니에 상품을 담은 후 장바구니 주문을 실시한 후에도 기타 여러 가지 이유로 인해 장바구니 주문에서 빠져나와 상품을 삭제하는 양상을

보인다. 또한, 남성은 '장바구니삭제→장바구니주문'의 트레이스만 존재하지만, 여성의 경우 장바구니 삭제가 바로 끝 지점이 되는 경우도 발견되었다. 즉, 여성이 남성보다 더 다양한 루트로 쇼핑몰 사이트를 이용하는 것을 알 수 있다.

남성 집단과 여성 집단 모두 'main_특별'이 눈에 띄는 액티비티이므로, 시작 지점으로 설정하고 전체 중 60%를 설명할 수 있는 트레이스 값을 주어 상대빈도로 나타냈다. 아래의 Fig. 3은 남성 집단의 분석 결과이다.

이어서 여성 집단도 같은 조건으로 설정한 후 프로세스 맵을 나타내고자 하였다. 다음의 Fig. 4는 여성 집단

의 분석 결과이다.

'main_특별' 액티비티를 시작점으로 60%의 트레이스를 설정하였을 때, 남성의 경우에는 'main_특별→주문시작'의 루트가 존재한다. 그러나 여성의 경우에는 동일 루트가 존재하지 않는다. 즉, 메인 내 이벤트 요소는 여성보다는 남성에게 있어 구매로 이어지는데 더욱 큰 영향을 미친다고 할 수 있다.

4.2 연령별 프로세스 비교

쇼핑몰 이용 프로세스를 파악하는데 있어 연령대 또한 중요한 기준이 될 수 있다고 판단하여 연령대를 20대부

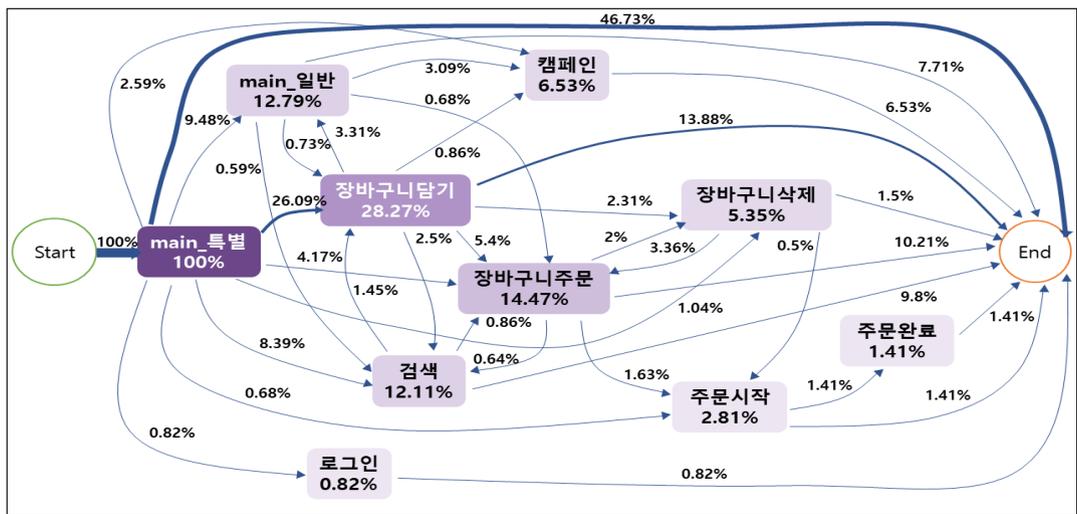


Fig. 3. Designating a starting point and 60% trace of Male

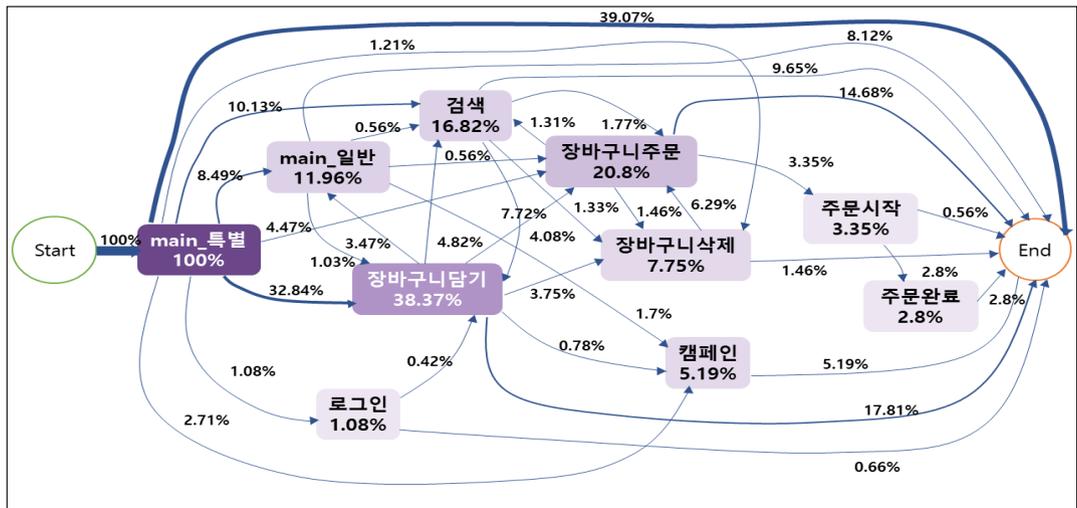


Fig. 4. Designating a starting point and 60% trace of Female

터 60대까지 세분화하였다. 10대의 데이터도 함께 수집되었지만 케이스 수가 매우 적기 때문에 연령별 분석에서는 제외하였다.

전 연령대 모두 특정한 조건을 주지 않는 경우 프로세스 맵을 이해하기 어렵다. 따라서 성별 프로세스를 분석한 것과 마찬가지로 일정 수준의 트레이스 값을 부여하여 프로세스를 파악하고자 하였다. 그 결과, 전 연령대에서 공통적으로 가장 중요한 시작점은 'main_특별' 액티비티이다. 또한, 모든 연령대에서 '캠페인'과 '로그인' 액티비티는 다른 액티비티와 관련이 없는 것을 확인하였다. 즉, 앞서 실행한 성별 분석 결과와 마찬가지로 로그인

나 캠페인의 혜택이 크지 않아 각 연령대의 관심을 끌지 못하는 것으로 나타났다.

특이한 점은 60대 미만 연령대에서 '캠페인'은 'main_특별' 액티비티와 끝점으로 이어졌는데 60대에서는 오직 끝점으로만 이어진다. 또한, 50% 트레이스에 '주문완료' 액티비티가 존재하지 않는다는 것이다. 이는 주문을 할 때 인터페이스가 노년층이 사용하기에도 적절한 수준인가를 생각해 볼 수 있는 근거가 될 수 있다고 생각된다.

전 연령대에서 중요한 시작점인 'main_특별' 액티비티를 활용하여 좀 더 자세히 비교해보고자 하였다. 그러나 웹로그 데이터 특성상 일정한 프로세스를 파악하기는

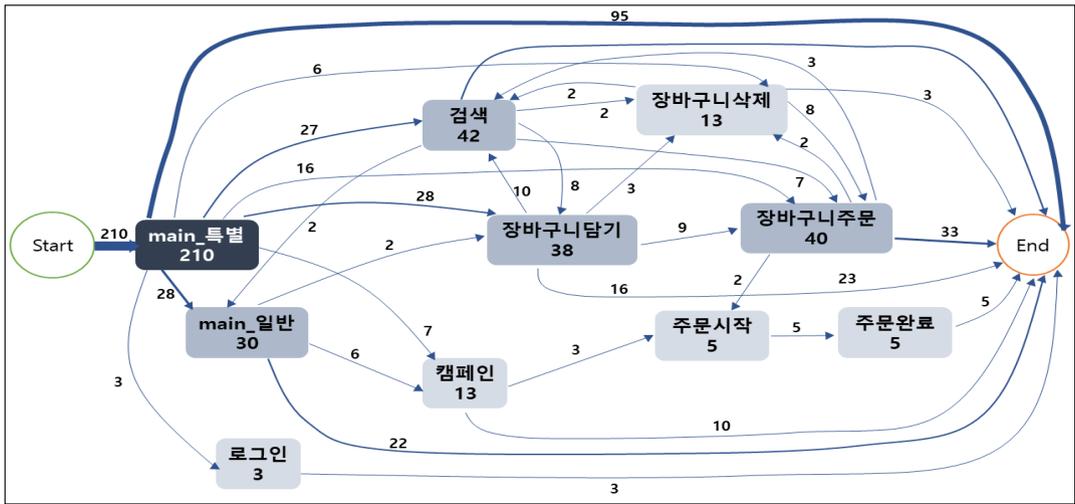


Fig. 5. Designating a starting point and 60% trace of 20's

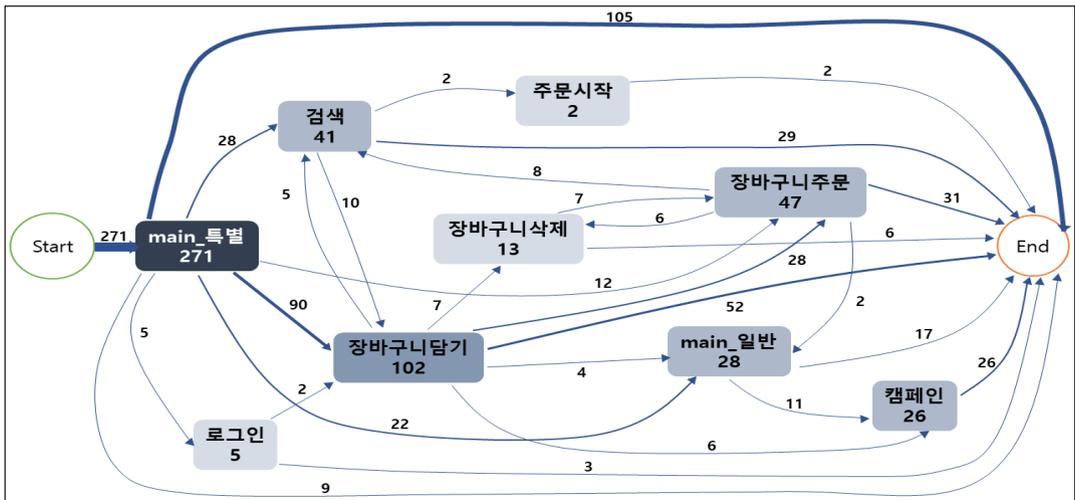


Fig. 6. Designating a starting point and 60% trace of 60's

매우 어렵고, 연령별로 두드러지는 차이는 없는 것으로 나타났다. 따라서 극단 범주인 20대와 60대만 비교하였다.

Fig. 5는 20대의 60% 트레이스를 빈도로 나타낸 것이다. 아래의 Fig. 6은 20대와 마찬가지로 60대의 트레이스를 빈도로 나타낸 것이다.

앞서 서술한 것처럼, 'main_특별'을 시작점으로 하여 두 극단 연령의 프로세스를 파악하고자 하였다. 두 연령대의 차이는 Fig. 6 이후에서 서술한다.

20대와 60대의 'main_특별'을 시작점으로 하는 60% 트레이스 차이는 다음과 같은 특징을 보인다.

우선, 20대는 60대보다 데이터가 많음에도 불구하고 'main_특별' 액티비티로 가는 빈도가 더욱 낮은 것을 볼 수 있다. 이는 20대가 60대보다 다른 액티비티 활동 또한 활발하게 하기 때문이라고 보인다. 프로세스 맵을 직관적으로 보면 20대의 프로세스는 60대의 프로세스보다 좀 더 복잡한 것을 알 수 있다.

두 번째로 60대에서는 '주문완료' 액티비티가 존재하지 않는다. 이는 앞서서도 언급되었듯이 주문을 완료하기까지의 인터페이스가 적절하지 못하거나, 결제 구조가 어려운 등의 사용성 문제와 연관 지어 생각해 볼 수 있다고 판단된다.

5. 결론

본 연구에서는 프로세스 마이닝을 활용하여 온라인 쇼핑몰 웹로그 데이터를 분석하고자 하였으며 세 가지의 연구문제를 바탕으로 사례 분석을 실시하였다. 이를 위해 R 프로그램의 오픈 소스 패키지인 BupaR이 사용되었다. 본 연구는 과거 연구와는 다음의 몇 가지 측면에서 차별점을 가진다.

- 웹 로그를 이용하여 온라인 쇼핑몰 방문고객의 활동을 프로세스를 파악한 점
- 프로세스 간 전이(transition) 확률을 계산하여 모델링한 점
- 프로세스 내 주요 액티비티를 파악하고 성별/연령별 차이점을 파악하고자 한 점

또한 분석된 결과에 따른 실무적 제언을 다음과 제시하고자 한다.

첫 번째로 메인페이지 중 특히 이벤트 요소를 지닌 카테고리 매우 중요하다. 앞의 전반적인 프로세스 및 성별, 연령별 프로세스 마이닝에서도 보았듯이 시작점에 가

장 많은 비중을 차지하는 것이 'main_특별' 액티비티이다. 식품, 생활용품과 같은 일반적인 카테고리인 'main_일반' 액티비티는 'main_특별' 액티비티에 비해 중요도가 많이 떨어진다. 즉, 다수의 쇼핑몰 사용자가 이벤트 요소가 있는 특정 카테고리를 이용하고자 하며 이러한 분석 결과는 효과적인 쇼핑몰 기획전 구축의 필요성을 시사한다. 또한, 이러한 시사점은 Fig. 3의 남성 집단 분석 결과와도 관련이 있다. Fig. 3의 결과에 따르면 여성 집단보다는 남성 집단이 'main_특별' 액티비티가 '주문시작'으로 바로 갈 수 있음을 보여주었다. 따라서 남성을 타겟으로 하는 기획전도 필요하다고 판단된다.

두 번째, '검색' 액티비티가 중요하다. 'main_특별' 액티비티와 같이 두드러지는 액티비티는 아니지만, 다른 모든 액티비티와 대체적으로 연관이 있으며, 전반적인 트레이스에 있어서 허브 역할을 한다. 따라서 연관 검색어 추천 및 검색 시 나오는 관련된 상품들을 잘 선별하여 사이트를 구축하는 것이 중요할 것으로 사료된다.

세 번째로 '로그인' 액티비티는 쇼핑몰 사용자에게 필요한 과정인 동시에, 관리자에게도 원활한 회원 관리 및 마케팅을 위해 반드시 필요한 액티비티이다. 그러나 본 데이터에서는 로그인과 관련된 트레이스가 매우 낮은 것으로 나타났다. 이처럼 해당 액티비티가 중요함에도 불구하고 트레이스가 매우 낮은 이유는 회원에 대한 혜택이 적절하지 못하거나, 번거로운 로그인 인터페이스 등의 이유가 있을 것으로 생각된다. 따라서 추후에는 로그인 시스템 점검을 통해 회원 권한으로 보다 쉽게 이용할 수 있도록 사이트를 구축해야 하며, 나이가 회원 관리 및 차별화 전략 등의 회원만을 위한 혜택을 제공해야 할 것으로 사료된다. 또한, 웹 쿠키 활용 등의 적절한 방법을 통해 수많은 익명 회원에 대한 대처 및 분석이 이루어져야 할 것으로 보인다.

네 번째, 프로모션이 적절하지 못한 것으로 생각된다. '캠페인' 액티비티는 구매를 촉진하는 아주 중요한 액티비티가 될 수 있는데[27], 본 사례에서는 쇼핑몰 이용 프로세스에 있어 '캠페인'이 중요하지 않은 액티비티로 나타났다. 따라서 추후에는 사용자들을 위한 보다 적절한 프로모션을 제공해야 할 것으로 생각된다. 특히 적절한 프로모션을 제공하는데 있어 로그인이 중요한 요인이 될 수 있다. 회원 정보를 바탕으로 사용자를 분석한 후, 개인에게 맞춰진 프로모션을 제시할 수 있을 것으로 생각된다.

마지막으로, 장바구니와 관련된 액티비티에서 추가적인 단계가 필요할 것으로 사료된다. 현재 본 데이터에는 '장바구니담기', '장바구니주문', '장바구니삭제'와 같은 주

문 직전의 장바구니와 관련된 액티비티가 존재하는데, 장바구니 관련 액티비티에서 실제 주문 완료로 이어지는 트레이스는 매우 낮다. 대개 '장바구니담기→장바구니삭제→장바구니주문→주문시작→주문완료'의 루트로 이루어지는데 주문을 시작하기 전 단계인 '장바구니주문'이나 장바구니 관련 액티비티의 첫 시작점이라고 할 수 있는 '장바구니담기' 단계에서 적절한 프로모션을 하면 주문 완료로 이어질 가능성이 높을 것으로 생각된다. 예컨대 특정 장바구니 단계에서 장바구니에 담아두었던 제품의 행사정보를 팝업으로 띄워주거나, 담아둔 제품보다 더욱 저렴하게 판매하고 있는 최저가를 찾아서 배너로 알려주는 방법 등이 있을 수 있다. 주문 완료로 이어지는 것이 쇼핑물의 궁극적인 목적이므로 전 단계인 장바구니와 관련된 액티비티에 구매를 촉진할 수 있는 요인을 넣을 필요가 있다고 생각한다.

본 연구의 한계점은 다음과 같다.

프로세스 마이닝에 필요한 변인들이 대체적으로 부족하다. 각 액티비티별 경로과약 및 소요시간 등을 분석할 때 인스턴스(Instance)라는 요인이 매우 중요한데, 본 데이터에서는 인스턴스에 대한 정보를 찾아볼 수 없다. 방대한 양의 데이터에 비해 빈도로 표현된 프로세스 맵을 도출하는 것 이외의 다른 분석을 수행하기 어렵다는 점에서 한계를 가진다.

그뿐 아니라, 초기 데이터에서 지역 정보나 키워드 타입 칸이 상당 부분이 'NA' 값이라 제거되었다. 만약 이러한 변인들의 데이터가 풍부했다면 분석을 통해 '지역별 상품 배치' 등과 같은 특화된 프로모션도 고안해 볼 수 있었을 것이다.

이러한 한계점에도 불구하고 오픈소스 패키지를 활용하여 일관성을 파악하기 어려운 온라인 쇼핑물 사이트의 프로세스 마이닝을 시도하였다는 점에서 연구의 의의가 있다. 추후에는 인스턴스의 보완 및 체계적인 액티비티 전처리를 통해 보다 나은 프로세스 마이닝을 수행할 수 있을 것으로 사료된다.

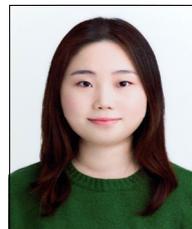
References

- [1] H. Cho, K. H. Kim, S. H. Byun, "Analysis on overseas direct purchasing utilizing global online shopping mall (B2C)", *The e-business studies*, Vol.15, No.4, pp.343-361, Aug. 2014.
DOI: <http://doi.org/10.15719/geba.15.4.201408.343>
- [2] A. A. Ozok, J. Wei, "An empirical comparison of consumer usability preferences in online shopping using stationary and mobile devices: results from a college student population", *Electronic Commerce Research*, Vol.10, No.2, pp.111-137, Apr. 2010.
DOI: <https://doi.org/10.1007/s10660-010-9048-y>
- [3] H. C. Hwang, S. H. Song, "A Study on the Factors Affecting the Acceptance of Logistics Robot in the Fulfillment Center Using the Technology Acceptance Model", *Journal of the Korea Academia-Industrial cooperation Society*, Vol.20, No.12, pp.287-297, Dec. 2019.
DOI: <http://doi.org/10.5762/KAIS.2019.20.12.287>
- [4] J. M. Yeo, U. J. Shin, B. W. Zheng, "The Data Model based Design and Implementation of a Shopping Mall System", *Korea Information Processing Society*, Vol.16, No.4, pp.273-288, Aug. 2009.
DOI: <http://doi.org/10.3745/KIPSTA.2009.16-A.4.273>
- [5] C. B. Yang, Y. S. Kim, C. K. Lee, "A Study on Time Series Analysis Using Purchasing Data of Online Shopping Mall", *Proceeding of Korea Intelligent Information Systems Society*, pp.80-81, Nov. 2017.
- [6] S. B. Jeong, "A Novel Approach to Customer Classification and Preference Goods Extraction based on Social Network Analysis for Internet Shopping Malls", *Entrue Journal of Information Technology*, Vol.13, No.1, pp.57-68, Apr. 2014.
- [7] Y. J. Seo, J. S. Han, "The Effect of Use Characteristics on Experiential Value and Behavioral Intention of Big Data based Online Travel Agency", *Journal of Tourism and Leisure Research*, Vol.31, No.7, pp.115-134, Jul. 2019.
DOI: <https://doi.org/10.31336/JTLR.2019.7.31.7.115>
- [8] B. Y. Lee, Y. H. Park, J. S. Yoo, "Buying Customer Classification in Automotive Corporation with Decision Tree", *Journal of the Korea Contents Association*, Vol.10, No.2, pp.372-380, Feb. 2010.
DOI: <http://dx.doi.org/10.5392/JKCA.2010.10.2.372>
- [9] J. H. Kim, K. H. Lee, H. J. Lee, D. S. Park, "Development of Recommendation System based on Data Mining Technique for E-Commerce", *The Journal of Korean Institute of Information Technology*, Vol.2, No.1, pp.47-54, March. 2004.
- [10] W. Van Der Aalst, *Data science in action*, In: Process mining, p. 488, Springer, Berlin, Heidelberg, 2016, pp.3-23.
DOI: https://doi.org/10.1007/978-3-662-49851-4_1
- [11] Y. S. Kang, B. K. Lee, *Process mining for business*, p.323, Hannarae, 2016, pp.34-47
- [12] Y. S. Kang, H. W. Lee, B. S. Kim, "Exploring the Prediction of Timely Stocking in Purchasing Process Using Process Mining and Deep Learning", *Information Systems Review*, Vol.20, No.4, pp.25-41, Dec. 2018.
DOI: <http://doi.org/10.14329/isr.2018.20.4.025>
- [13] W. Van Der Aalst, *Process mining: discovery*,

- conformance and enhancement of business processes (Vol.2). p.352, Heidelberg: Springer, 2011, pp.59-91
- [14] H. A. Kim, "Learning process mining techniques based on open education platforms", *The Journal of the Convergence on Culture Technology (JCCT)*, Vol.5, No.2, pp.375-380, May. 2019.
DOI: <http://dx.doi.org/10.17703/JCCT.2019.5.2.375>
- [15] J. G. Park, M. G. Kim, J. H. Woo, "A Study on Process Management Method of Offshore Plant Piping Material using Process Mining Technique", *Journal of the Society of Naval Architects of Korea*, Vol.56, No.2, pp.143-151, Apr. 2019.
DOI: <http://dx.doi.org/10.3744/SNAK.2019.56.2.143>
- [16] N. Poggi, V. Muthusamy, D. Carrera, R. Khalaf, Business process mining from e-commerce web logs. In: Business process management, pp. 65-80, Springer, Berlin, Heidelberg, 2013.
DOI: https://doi.org/10.1007/978-3-642-40176-3_7
- [17] J. D. Jung, S. H. Kim, H. J. Yeo, "Clinical Pathway Verification through Process Mining", *Journal of the Korea Society of Computer and Information*, Vol.23, No.4, pp.115-120, Apr. 2018.
DOI: <https://doi.org/10.9708/jksci.2018.23.04.115>
- [18] J. S. Park, J. Y. Jung, "Analysis of Purchase Process Using Process Mining", *The Korea Journal of BigData*, Vol.3, No.1, pp.47-54, Aug. 2018.
DOI: <http://doi.org/10.36498/kbigdt.2018.3.1.47>
- [19] K. C. Lee, J. S. Kim, N. H. Chung, S. J. Kwon, "Fuzzy cognitive map approach to web-mining inference amplification", *Expert Systems with Applications*, Vol.22, No.3, pp.197-211, 2002.
DOI: [https://doi.org/10.1016/S0957-4174\(01\)00054-9](https://doi.org/10.1016/S0957-4174(01)00054-9)
- [20] J. Srivastava, R. Cooley, M. Deshpande, P. N. Tan, "Web usage mining: Discovery and applications of usage patterns from web data", *Acm Sigkdd Explorations Newsletter*, Vol.1, No.2, pp.12-23, Jan. 2000.
DOI: <https://doi.org/10.1145/846183.846188>
- [21] J. Y. Chang, "A sentiment analysis algorithm for automatic product reviews classification in on-line shopping mall", *The Journal of Society for e-Business Studies*, Vol.14, No.4, pp.19-33, Nov. 2009.
- [22] S. K. Kim, Y. C. Kim, "The Study of Online Log Data based Marketing log", *Electronic Commerce Research Association*, Vol.13, No.1, pp.23-39, Mar. 2012.
- [23] J. S. Choi, D. Y. Kim, I. S. Han, Y. C. Kim, "A Study on Data Analysis in on/off-line shopping mall", *Electronic Commerce Research Association*, Vol.12, No.2, pp.71-88, Jun. 2011.
- [24] K. H. Lee, D. S. Park, "A Design and Implementation of Goods Recommendation System using Web Mining", *Proceeding of Korea Academia-Industrial cooperation Society*, KAIS, South Korea, pp.222-225, Jun. 2003.
- [25] R. Bhushan, R. Nath, "Automatic recommendation of web pages for online users using web usage mining", In *2012 International Conference on Computing Sciences*, IEEE, India, pp.371-374, Sep. 2012.
DOI: <https://doi.org/10.1109/ICCS.2012.17>
- [26] B. S. Hong, E. J. Lee, H. Y. Ma, "The effect of merchandise characteristics of internet fashion shopping malls on the shopping values, purchase satisfaction, and repurchase intention of female consumers", *Journal of the Korean Society of Clothing and Textiles*, Vol.33, No.11, pp.1828-1838, Nov. 2009.
DOI: <https://doi.org/10.5850/JKST.2009.33.11.1828>
- [27] D. M. Gu, H. S. Jang, The influences of personal values on online shopping benefits, online store attributes and re-purchase intention to an online store. *Journal of the Korea Association of Business Education*, Vol.40, pp.399-433, Dec. 2005.

김 채 영(Chae-Young Kim)

[정회원]



- 2019년 2월 : 한림대학교 경영학과 학사
- 2019년 3월 ~ 현재 : 한림대학교 인터랙션디자인 석사과정

<관심분야>

프로세스 마이닝, 인터랙션 디자인, 디지털 콘텐츠

용 혜 련(Hye-Ryeon Yong)

[정회원]



- 2015년 2월 : 한림대학교 경영학과 학사
- 2017년 2월 : 한림대학교 대학원 인터랙션디자인 석사과정
- 2015년 9월 ~ 2016년 12월 : 한림ICT정책연구센터 석사연구원

<관심분야>

프로세스 마이닝, 인터랙션 디자인, 디지털 콘텐츠

황 현 석(Hyun-Seok Hwang)

[정회원]



- 2000년 2월 : POSTECH, 산업경영공학과 (Management Information Systems 석사)
- 2004년 2월 : POSTECH, 산업경영공학과 (Management Information Systems Ph. D.)
- 2004년 3월 ~ 현재 : 한림대학교 경영학과 교수

- 2004년 3월 ~ 현재 : 한림 경영연구소 연구위원

<관심분야>

빅데이터, 데이터마이닝, 프로세스 마이닝