

미디어 레퍼토리를 이용한 스마트폰 애플리케이션 이용 패턴 유형 분석

Mobile App Analytics using Media Repertoire Approach

권성은(Sung Eun Kwon)*, 장서인(Shu In Jang)**, 황보현우(Hyunwoo Hwangbo)***

초 록

현대인에게 가장 보편적이고 융합적인 미디어인 스마트 폰은 애플리케이션이라는 비히클을 갖는 뉴미디어이다. 이 연구는 미디어 사용자들은 어떻게 레퍼토리를 구성하여 미디어를 이용하고 있는지를 파악하고자 2019년 11월, 4주 동안의 개인별 모바일 이용행동 로그 데이터를 이용하여 모바일 애플리케이션 카테고리별 미디어 이용량을 중심으로 군집 분석을 실시하고, 최종적으로 8개의 모바일 미디어 레퍼토리 유형별 집단을 분류하였다. 8개의 각 미디어 레퍼토리 그룹은 애플리케이션 카테고리별 절대적 이용량과 타 그룹 대비 상대적 이용량에서 차이를 보였으며, 데모그래픽적 분포에서도 집단간 차이를 보였다. 이 연구는 모바일 미디어 레퍼토리를 규명해 냈다는 학문적 기여뿐만 아니라 기존의 k-means clustering에 의존적이었던 군집 분석을 SOM(Self-Organized Map)을 이용하여 프로토크터를 추출하고 이 프로토크터를 이용하여 k-means clustering을 실시하는 이단계 접근법(two-step approach)을 시도함으로써, 기존 k-means clustering이 갖고 있는 ‘이상치(outlier)’나 ‘결측치’에 민감했던 한계점을 극복하고 더 나은 성능의 분석 결과를 도출하고 있음을 보여준다는 점에서 방법론적으로도 의미를 갖는다. 또한 모바일 미디어 이용 행동의 유형 분류 연구는 전자거래 서비스를 이용하는 고객을 유형분류하고, 각 고객 유형에 맞는 고객 관리 서비스를 집행해야 하는 실무진이 고객 행동 로그 데이터를 기반으로 고객의 구조를 파악하고 각 고객 집단에 적합한 서비스 또는 마케팅 의사결정을 차별적으로 집행해야 하는 전자거래 커뮤니티에 실무적 가이드를 제공한다는 점에서도 의미를 갖고 있다.

ABSTRACT

Today smart phone is the most common media with a vehicle called ‘application’. In order to understand how media users select applications and build their repertoire, this study conducted two-step approach using big data from smart phone log for 4 weeks in November 2019, and finally classified 8 media repertoire groups. Each of the eight media repertoire groups showed differences in time spent of mobile application category compared to other groups, and also showed differences between groups in demographic distribution. In addition to the academic contribution of identifying the mobile application repertoire with large scale

본 연구는 산업통상자원부 산업기술혁신사업의 지원을 받아 진행되었음(20009185).

* First Author, CEO, Zero to One Partners(sekwon@0to1partners.com)

** Co-Author, Data Intelligence Lab, Zero to One Partners(sijang@0to1partners.com)

*** Corresponding Author, CDO(Chief Data Officer), Hana Financial Group(scott@hanafn.com)

Received: 2021-10-25, Review completed: 2021-11-24, Accepted: 2021-11-27

behavioral data, this study also has significance in proposing a two-step approach that overcomes ‘outlier issue’ in behavioral data by extracting prototype vectors using SOM (Self-Organized Map) and applying it to k-means clustering for optimization of the classification. The study is also meaningful in that it categorizes customers using e-commerce services, identifies customer structure based on behavioral data, and provides practical guides to e-commerce communities that execute appropriate services or marketing decisions for each customer group.

키워드 : 미디어 레퍼토리, 스마트폰 어플리케이션, k-평균 군집분석, 자기조직화지도, 이단계 접근법, 마케팅 의사결정
Media Repertoire, Smart phone application, k-means clustering, SOM, 2-step approach, Marketing planning

1. 서 론

모바일 인터넷의 보급으로 인해 가장 일상적이고 보편적인 미디어가 된 스마트폰은 모바일 기기가 갖는 이동성과 함께 통신 기능, 미디어 기능, 그리고 PC를 대체하는 인터넷 정보 검색 및 공유의 기능이 통합된 융합미디어로 자리매김하게 되었다[3]. 스마트폰은 우리의 모든 생활에 임베딩되어 있는 스마트폰은 현대인에게 가장 일상적이고 보편적인 미디어로 자리잡고 있음을 보여주고 있다[1].

제일기획이 2021년 발표한 ‘2021 광고연감’에 따르면 국내 모바일 광고 시장은 2020년 3.8조, 2021년 (추산) 4.2조로 전년 대비 약 10% 성장이 예고되고 있다. 2021년 추산 기준 모바일 광고 시장 규모 4.2조는 약 33.9%의 점유율에 해당하며, 이는 지상파 TV 9.2%, 케이블/중편 15.3%, 그리고 PC 광고 시장 15.5%의 점유율과 비교해 봐도 두 배 이상 시장 규모를 갖는 가장 핵심적인 시장이다[2]. 모바일 광고 시장의 규모와 성장세로 볼 때, 스마트폰이 광고 미디어로서 산업적으로 가장 중요하다는 것은 자명한 현실이다.

미디어로서의 스마트폰은 수많은 애플리케이션을 가지고 있다. 이 모바일 애플리케이션

들은 텔레비전이라는 미디어에 채널이나 프로그램이라는 비히클(Vehicle)이 있는 것처럼 스마트폰이라는 미디어에서 하나의 비히클로 기능한다. 비히클로서의 모바일 애플리케이션 이용 행동을 이해하는 방법 중 하나가 텔레비전, 채널 그리고 프로그램에 대한 미디어 연구 접근 방법인 “레퍼토리 접근법(Repertoire approach)”이다[12, 13, 18] 개인은 애플리케이션 마켓에서 애플리케이션을 선택하여 다운로드하고 자신만의 미디어 환경을 구축하여 일상에서 차별적으로 이용하면서 자신만의 고유한 애플리케이션 레퍼토리를 구성하는데, 이러한 애플리케이션 레퍼토리 구성에 대한 연구는 스마트폰이라는 모바일 미디어 이용 행태를 이해하고자 하는 미디어 연구에서 필수적이다 [3] 스마트폰 애플리케이션 레퍼토리의 이해는 모바일 애플리케이션을 비히클로 하여 모바일 광고를 집행하고자 하는 마케팅 커뮤니케이션 실무 담당자에게는 모바일 타겟팅의 정교화를 통해 마케팅 커뮤니케이션의 효율성과 효과성을 극대화할 수 있기 때문에 실무적 가이드를 제공할 수 있다. 또한 학문적으로도 개인 미디어

로서의 스마트 폰 애플리케이션 레퍼토리에 대한 이해는 개인간 커뮤니케이션 및 매스 커뮤니케이션 도구로서의 스마트 폰을 조명한다는 점에서 다양한 후속 연구를 이끌어 낼 수 있기 때문에 그 의의가 적지 않다.

개인 미디어이자 광고 미디어로서 스마트 폰은 학문적으로 그리고 산업적으로도 중요함에도 불구하고 Cable TV, IP TV, PC(인터넷 웹사이트)와 같은 기존 뉴미디어 이용 행동 연구와 비교해 볼 때 그 조명이 상대적으로 부족한 현실이다. 이는 빠르게 변화하는 모바일 앱 사용 행동을 설문조사와 같은 기억에 의한 측정을 하기 어려운 한계에서도 그 원인을 찾아볼 수 있다.

최근에는 모바일 로그 데이터 수집 기술의 발전으로 인해 모바일 소비자 행동 측정이 용이해지고 있어서 정확한 모바일 미디어로서의 스마트 폰 이용 행동 연구가 가능해지고 있는데, 이 연구는 모바일 로그 데이터 수집 기술의 발전에 힘입어 스마트 폰 이용자로 구성된 모바일 패널을 통하여 수집된 모바일 로그 데이터를 이용하여 스마트 폰 이용 패턴을 분석하고자 한다. 또한 미디어 레퍼토리 개념을 이용하여 모바일 미디어의 애플리케이션 이용 패턴을 분석을 시도한다. 이를 위해 모바일 애플리케이션을 소셜, 커뮤니케이션, 게임 등 34개의 유형으로 분류하여, 서로 다른 모바일 애플리케이션 레퍼토리를 갖고 있는 모바일 사용자의 특성을 이해하고자 한다. 연구 분석 방법적으로는 기존의 k-means clustering이 이상치와 잡음에 민감하기 때문에 대규모 빅데이터(large scale big data) 분석에는 한계를 갖고 있다. 따라서 이 연구에서는 자기조직화지도(Self-Organized Map)를 1 단계로 적용하여 프로토타입 벡터를 구하여 이상치와 잡음을 완화하고, 이를 입력값으로 적용

하여 k-means clustering을 통해 군집 분석을 완료하는 2단계 군집분석(Two-step)분석을 시도한다.

구체적인 이 연구의 목적은 모바일 미디어 애플리케이션 레퍼토리를 규명하고 각 모바일 미디어 레퍼토리 유형별로 사용자 특성을 이해하는 것이다. 이를 위해 다음 제2장에서는 이론적 틀에 대한 고찰을, 제3장에서는 데이터 및 분석 방법론을, 제4장에서는 분석 결과를 논의한 후, 마지막으로 연구 결과가 가지는 학문적 그리고 산업적 함의와 미래 연구를 위한 제언을 하고자 한다.

1.2 연구 문제

이 연구는 미디어 레퍼토리 접근법을 차용하여 한국 스마트 폰 사용자의 모바일 미디어 레퍼토리를 파악하고자 하는 탐색적 연구로 다음과 같은 두 가지 연구 질문에서 시작한다.

연구문제 1: 모든 모바일 미디어 애플리케이션에 대하여, 어떤 유형의 미디어 레퍼토리가 나타나는가?

연구문제 2: 각 모바일 미디어 레퍼토리 유형별 모바일 애플리케이션 사용 행동과 사용자 특성은 어떤 차이가 있는가?

2. 관련 연구

2.1 미디어 레퍼토리의 의미

레퍼토리 개념은 텔레비전 채널 선택 연구뿐

만 아니라 새로운 미디어가 등장하여 미디어 환경이 변화하는 상황에서 미디어 소비 행동을 분석하는 주요한 접근방법으로 다양한 연구에 적용되어 왔다. 미디어 레퍼토리(Media repertoire)는 케이블 텔레비전의 등장으로 다수의 시청 가능한 채널 환경에서 이용자의 채널 선택 과정에 대한 연구에서 시작되었다. 미디어 이용자들은 시청 가능한 수많은 채널 중에서 무작위로 시청 채널을 선택하는 것이 아니라 소수의 채널을 선택해서 규칙적으로 이용하는 경향을 보이는데, 이것이 “채널 레퍼토리”이며 “개인이나 세대가 규칙적으로 시청하는 채널 묶음(the set of channel)”으로 정의했다[9, 16].

레퍼토리 개념은 새로운 미디어의 도입에 따른 다매체 다채널 환경에서의 유용한 분석틀로 사용되면서 다양한 연구에 활발하게 적용되고 있으며, 최근에는 다중 미디어 이용 현상을 설명하기 위하여 다양한 뉴미디어 레퍼토리 연구로 그 영역을 확대하게 되었다.

2.2 미디어 레퍼토리에 대한 기존 연구

미디어 레퍼토리 개념을 처음 제안한 Heeter는 케이블 TV에서 ‘채널 레퍼토리’를 ‘시청자가 시청한 서로 다른 채널의 개수’로 정의하여 개인별 채널 레퍼토리의 다양성을 비교했다. Neuendorf, Atkin, & Jeffres 등은 전체 채널 중 한 번이라도 시청한 적이 있는 채널의 개수를 ‘기본 레퍼토리’로 정의하고 나아가 전체 채널의 표준화된 시청 빈도의 합은 ‘가중된 레퍼토리’로, 매일 시청하는 채널의 개수는 ‘1차 레퍼토리’, 매주 시청하는 채널의 개수는 ‘2차 레퍼토리’, 그리고 일주일 이하의 빈도로 시청하는 채널의 개수는 ‘3차 레퍼토리’로 정의하는

등, 레퍼토리 ‘크기’로 레퍼토리를 정의하여 개인별 미디어 이용패턴을 분석하였다. 그러나 미디어 이용자들의 이용 패턴을 보다 명확히 이해하기 위해서는 레퍼토리의 크기보다는 레퍼토리의 구성을 바탕으로 한 미디어 이용자 유형 분석의 필요성이 제기되었다. Kees van Rees and Koen van Eijck는 TV, 신문, 잡지, 라디오, 인터넷 등 19개의 미디어를 대상으로 요인분석을 실시하여 8개의 미디어 유형으로 차원을 축소한 후 군집분석을 실시하여 미디어 레퍼토리 집단을 분류해 내었다. 이들은 미디어 이용의 조합적 특성에 따라 성, 연령, 학업수준, 경제적 수준, 사회적 지위 등의 미디어 이용자 분포 특성이 달라지고 있음을 보여줌으로써 미디어 레퍼토리가 미디어 이용자의 미디어 선택행위를 구분 짓는 변수가 될 수 있음을 보여주었다[19].

국내에서도 TV, 신문, 잡지, 라디오, 인터넷 등의 전통적 매체를 중심으로 미디어 레퍼토리에 대한 연구가 이루어졌다. 강남준 등은 TV, 신문, 잡지, 라디오, 인터넷, 휴대전화, MP3 등 7개 미디어의 미디어별 사용량을 하루 사용량으로 나눈 값인 상대적인 미디어 사용량을 기준 변수로 K-means 군집분석을 실시하여 레퍼토리 유형을 분류하였다[12]. 저자들은 미디어 이용시간에 따른 군집 분석으로 도출된 미디어 레퍼토리 유형을 인구통계학적 및 경제적 특징 등과 같이 미디어 레퍼토리 영향 변인들과 상호 비교하여 미디어 이용량(이용 시간)이 미디어 이용자를 분류하는 주요한 분류 단서가 될 수 있음을 보여주었다. 한편 김은실 등은 지상파TV, 케이블 TV, 인터넷, 신문, 라디오, DMB 등 7개 미디어 이용량을 기준으로 K-means 군집 분석을 실시하여 8개 유형의 미

<Table 1> Previous Research on Media Repertoire

Author	Media	Analysis Variable	Data Collection	Analysis Method
Kees van Rees and Koen van Eijck (2003)	Newspapers, TV, Radio, Magazines, Books, Computer, and Video	Time spent on media types	Diary	Factor Analysis(CFA), K-means clustering
Kang, Lee, & Lee (2008)	TV, Newspapers, Radio, Magazines, Internet Mobile phone, MP3	Time spent on media types	Survey	K-means clustering
Kim, Ha, & Park (2009)	TV, Cable TV, Internet News paper, Radio, DMB	Time spent on media types	Survey	K-means clustering
Ferguson and Perse (2010)	World Wide Web sites	Site visit records among top 100 site lists on whether visited or not in the past week	Survey Diary	PCA/(EFA/CFA)
Yuan(2011)	TV, Newspapers, Internet, Mobile phone, Radio, and News Magazines	Amount of time spent on each media, Size of media repertoire	Survey	Factor Analysis(CFA), K-means clustering
Jung, Kim & Chan-Olmsted. (2014)	Smartphone application	Total time spent of smart phone	Mobile log	Comparison of concentration measures (ex. CR4, HHI, Gini coefficient)
Choi, Part & Lee (2015)	Smartphone application	Time spent of main application	Survey	Factor Analysis(CFA), K-means clustering
Sanphil Han (2018)	Smartphone application	Tme spent	Mobile log	Two-Dimensional MDA
Our study	Smartphone application	Time spent	Mobile log	SOM, K-mens clustering

디어 레퍼토리 집단을 분류하였으며, 이 8개 미디어 레퍼토리 유형 집단간 라이프 스타일 유형이 다르게 분포하고 있음을 발견하였다[13]. 김은실 등의 연구는 미디어 레퍼토리를 미디어 이용자의 활동, 흥미, 관심사와 연관되어 있음을 보여줌으로써 인구통계학적 변인 이외에 미디어 이용자 세분화의 근거로 라이프스타일과 상관관계가 높은 미디어 레퍼토리가 변수로 기

능할 수 있음을 보여주었다.

한편, World Wide Web의 등장은 미디어 연구에 새로운 영역 및 개념을 확장하며 큰 영향을 미쳤다. 기존의 채널 레퍼토리 개념이 수많은 웹페이지(Web page)를 소비하는 웹 오디언스 행동(Web audience behavior) 연구에 적용되어 그 개념이 확장되었다. 웹 오디언스(Web audience)들은 수많은 웹 사이트를 모두 활용

하는 것이 아니라 몇 개의 웹사이트에 사용성이 집중되는데[13], 특히 Webster & Lin은 Top 20%의 웹 사이트가 전체 웹 오디언스(Web audience)의 61%의 전체 방문자(Total Visitor)를 점유하고 있음을 확인하였다[21].

스마트 폰의 등장으로 레퍼토리 이론은 애플리케이션(Application)의 사용성 연구에도 적용되기 시작하였다. 특히, Jung & Chan-Olmsted는 Top 10 applications가 전체 사용시간의 54.1%를 차지하고, Top 20%의 applications가 전체 사용시간의 84.1%를 차지한다는 empirical study를 실시하였다[11]. 이러한 연구들을 통해 스마트 폰 사용자들이 모든 App을 공평하게 사용하지 않고, 소수의 Apps을 선택하여 사용한다는 Application 레퍼토리 존재를 확인하였다. 한 상필 등은 11,548 application에 대한 모바일 로그 데이터를 바탕으로 사용 시간을 Correlation 과 2-Dimensional MDA를 분석하여 상관관계가 높은 애플리케이션이 하나의 미디어 레퍼토리로 묶일 가능성이 있음을 시사한 바 있다[6]. 그러나 이러한 모바일 미디어에 대한 선행 연구들은 여전히 모바일 미디어 레퍼토리를 구체적으로 규명하지는 못했다.

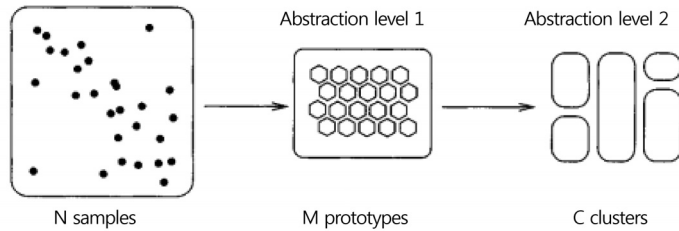
이 연구에서는 이 미디어 레퍼토리 개념을 이용하여 모바일 미디어의 애플리케이션 이용 패턴을 분석하고자 한다. 즉 스마트 폰 이용자의 모바일 애플리케이션을 ‘커뮤니케이션’, ‘엔터테인먼트(게임)’, ‘쇼핑’, ‘재무 관리’ 등과 같이 미디어 이용자의 이용 동기에 따라 유형 분류를 하여 각 유형별로 분류하여 어떤 유형의 미디어 이용자들이 어떤 유형의 모바일 미디어를 이용하는 지 모바일 미디어 레퍼토리 유형을 분류하고자 한다. 미디어 이용 동기는 특정 미디어를 선택하고 지속적으로 사용하게 하는

중요한 변수이자 개인의 라이프스타일에 영향을 받는 변수이다[13]. 따라서, 특정 이용 동기가 강하면 특정 애플리케이션의 이용량은 많게 나타날 것이며, 이는 이용자가 자주 이용하는 미디어 레퍼토리에도 영향을 미칠 것이다[3]. 이 아이디어를 바탕으로 어떤 유형의 미디어 이용자들이 어떤 유형의 모바일 미디어 유형을 이용하는 지를 군집분석 기법을 이용하여 분석하여 모바일 미디어 레퍼토리 유형을 분류하고자 한다.

2.3 SOM과 k-means clustering을 이용한 Two-step 군집분석

다수의 미디어 레퍼토리 연구에서 적용되어 왔던 k-means clustering은 동질적인 그룹을 분류해 내는데 유용한 분석 방법이다. 그러나 오랫동안 많은 연구에서 보편적으로 적용해 온 k-means clustering은 몇 가지 한계점들에 대해 비판받아 왔는데, 그 중의 하나가 극단치나 이상치(outlier), 그리고 잡음(noise)에 민감하다는 점이다[5]. k-means clustering의 단점을 해결하기 위해 Vesanto와 Alhoniemi는 초기 데이터 셋을 이용하여 학습을 하는 과정에서 얻게 되는 프로토타입 벡터를 이용하여 k-means clustering을 시도하는 2단계 접근법을 제안하였다[20]. 이들의 제안은 SOM(Self-Organized-Map)과 같이 프로토타입 벡터를 얻을 수 있는 분석을 통해서 1단계 그룹화된 데이터를 추출하고, 이를 다시 입력값으로 이용하여 2단계의 최종 클러스터를 추출하는 방식이다(<Figure 1> 참조).

SOM(Self-Organized-Map)은 비지도학습의 한 방법으로 고차원의 데이터를 2D(two-dimensional) 그리드 맵의 저차원 상으로 투영



〈Figure 1〉 2-step Cluster Analysis Approach

시켜 데이터의 특징을 추출해 내는 방법 중의 하나이다. SOM은 각 훈련 단계에서 하나의 샘플 벡터 x 를 랜덤하게 선택해서 이 x 와 다른 M 개의 프로타입 벡터들 사이의 거리를 기반으로 학습을 진행한다. 학습 과정에서 업데이트 되는 것은 프로타입 벡터로 수렴될 때까지 학습을 진행하여 최적의 프로타입 벡터를 구한다. 이때 프로타입 단위 벡터 i 를 위한 업데이트 규칙은 다음과 같이 나타낼 수 있다.

$$m_i(t+1) = m_i(t) + \alpha(t) h_{bi}(t) [x - m_i(t)]$$

여기서 t 는 시간, $\alpha(t)$ 는 학습률(learning rate)로 기능하는 일종의 상수(adaptation coefficient)이며, $h_{bi}(t)$ 는 최적화 벡터(winner unit i)에서의 근방 커널(neighborhood kernel centered on the winner unit)을 의미하는데, 구체적으로 산출식을 표현하면 다음과 같다.

$$h_{bi}(t) = \exp\left(-\frac{\|r_b - r_i\|^2}{2\sigma^2(t)}\right)$$

위 식에서 r_b 와 r_i 는 인공신경망 모델인 SOM의 b 와 i 로 표현되는 뉴런 속에서의 위치를 의미하며, $\alpha(t)$ 와 $\sigma(t)$ 는 시간이 지남에 따라 단조감소하게 된다[14].

SOM(Self-Organized-Map)의 학습을 통해 고차원의 관찰값들이 저차원 상의 노드에 최적화된 프로타입 벡터들로 투영된 지도(Map)를 얻게된다. 이와 같이 1단계에서 SOM을 통해 M 개의 프로타입 벡터들을 구하고 이를 SOM 그리드 상에 일종의 좌표 값으로 표현한 후, 이를 새로운 관찰값으로 취급하여 k-means clustering을 2단계로 진행한다. 이와 같은 Two-step 군집분석은 변수와 관찰값이 많거나 잡음과 이상치가 있을 때 특히 높은 성능을 보인다[14].

2.4 군집의 개수 선정

군집분석에서 최적의 클러스터 개수, k 를 결정하기 위해 이 연구는 실루엣 인덱스(Silhouette Index)를 이용하였다. 군집분석에서 최적의 군집 개수를 정하는 방법을 제안하는 다양한 군집 타당성 지수들은 공통적으로 군집간의 거리는 최대이면서 군집내의 거리는 최소가 될 수 있는 군집화를 판단할 수 있는 아이디어를 제시한다. 이 연구에서는 Kaufman과 Rousseeuw이 제시한 ‘데이터간 유클리디안 거리’에 바탕을 둔 실루엣 인덱스가 최대점을 기준으로 하기 때문에 명확한 판단근거를 제공하기에 이를 바탕으로 최적의 군집 개수를 도출하였다. 실루엣

인덱스의 구체적인 산출식은 군집 C1, C2, ..., Ck와 xi $x_i \in C_i$ 인 경우 아래와 같은데, $a_K(i)$ 는 한 클러스터 안에 있는 개별 데이터가 같은 클러스터 안의 다른 데이터와 갖는 거리 평균이고, $b_K(i)$ 는 그 개별 데이터가 다른 클러스터에 속하는 개별 데이터와 갖는 거리들 중 가장 작은 값, 즉 최소값이다. 분모인 $b_K(i) - a_K(i)$ 가 크면 클수록 군집타당성이 높은 것으로 판단할 수 있다[14].

$$Silhouette(K) = \frac{1}{N} \sum_{i=1}^N \frac{b_K(i) - a_K(i)}{\max(a_K(i), b_K(i))}$$

a_K : 같은 클러스터 안의 다른 데이터들과의 평균 거리

b_K : 다른 클러스터의 다른 데이터들과 거리들 중 최소값

2.5 군집 분석 방법별 성능 평가

군집분석 결과의 타당성을 평가하는 지수로 이 연구에서는 두 군집분석 결과간 일치도를 평가하는 지표인 랜드 인덱스(랜드 인덱스(Rand Index))를 이용하여 2-step 군집 분석의 성능을 평가하고자 한다. 랜드 인덱스(랜드 인덱스(Rand Index))는 군집분석 성능 평가 방법이 간결하고 해석이 명쾌하기 때문에 다수의 미디어 레퍼토리 연구에서 일반적으로 사용하고 있다. 허명희와 이용구 연구에 따르면, 랜드 인덱스(랜드 인덱스(Rand Index)) 산출은 행 I와 열 j로 분류되는 개체의 수를 n_{ij} 라고 하고($i, j = 1, \dots, K$), 분석 대상 표본을 두 개의 표본으로 분할할 때, 임의의 개체 a와 b의 군집 분류를 같은 군집으로 분류될 때와 다른 군집으로 분

류될 때로 나누어 군집 타당성을 평가하는 경우의 수를 생각해 볼 수 있다. 이 아이디어를 바탕으로 랜드 인덱스(랜드 인덱스(Rand Index)) 산출식은 각 표본에서 도출할 수 있는 총 쌍(pair) 중 군집분석의 분류 결과가 일치하는 쌍의 비율을 바탕으로 아래와 같이 나타낼 수 있다[10].

$$\text{랜드 인덱스(Rand Index)} = \frac{N}{\binom{n}{2}}$$

N: 같은 그룹으로 분류된 쌍(pair)의 수

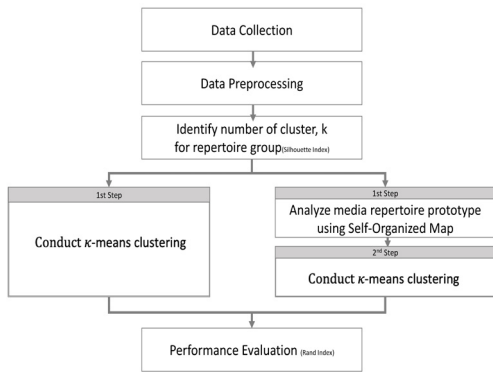
n: 표본에서 도출할 수 있는 전체 쌍(pair)의 수

최근에는 우연적인 일치 쌍을 제외하는 수정된 랜드 인덱스(랜드 인덱스(Rand Index))가 제안되기도 하였으나 군집 크기에 따라 일괄적으로 수정 지수를 낼 수 없는 한계점에 대한 합의된 해결 대안이 아직 뚜렷하지 않으므로 연구에서는 랜드 인덱스(랜드 인덱스(Rand Index))를 중심으로 군집 분석 성과를 평가하고자 한다.

3. 연구 방법

3.1 연구 분석 절차

이 연구는 아래 <Figure 2>와 같이 데이터 수집 및 전처리, 미디어 레퍼토리 유형 수 k의 결정, 그리고 k-means clustering만 적용한 1단계 군집 분석과 Self-Organized map과 k-means clustering을 순차적으로 적용하는 2단계 군집분석의 실시, 그리고 마지막으로 두 분석 방법의 모형 평가 및 해석의 절차로 진행되었다.



<Figure 2> Research Procedure

3.2 모바일 로그 데이터

스마트 폰 사용 행동이 로그(log)로 기록되면서 쌓이는 로그 데이터를 활용할 수 있게 되자 미디어 연구진은 새로운 연구 분야를 확장시켜 나갔다. 스마트 폰 로그 데이터는 크게 센서리 데이터(Sensory Data)와 사용 행동 추적 데이터(Usage pattern tracking data)로 나눌 수 있다. 스마트의 센서로 모을 수 있는 데이터이다. 일례로, GPS를 통한 위치 데이터(location data), 가속도계(accelerometer) 및 자이로스코프(gyroscope) 등을 통한 수직 운동/수평 운동 데이터(vertical/lateral movement data), 또는 블루투스(Bluetooth)를 통해 다른 기기와의 상대적인 위치 데이터 등이 있다. 스마트 폰의 사용 행동 추적 데이터(Usage pattern tracking data)는 보통 스마트 폰 사용자들이 스마트 폰 애플리케이션 이용 행동을 기록(logging)하는 데이터로 스마트 폰 화면이나 부속 버튼을 클릭할 때마다 행동 기록을 저장한 데이터를 의미한다[17].

<Figure 3>의 예시에서 보듯이 “글로벌 정보(global)” 부분에는 스마트 폰 기기 정보, 운

영 체계 정보 등이 기록되며, “데이터 정보(dgp)” 부분에는 모바일 애플리케이션을 화면 클릭할 때 마다 1) 시작 시작 2) 종료시간, 3) 애플리케이션 이름(즉, 패키지 네임) 4) 애플리케이션 종료 방식(기기 종료, 애플리케이션 종료, 백그라운드 전환으로 인한 종료 등)에 대한 정보가 차례대로 기록된다. 이 로그 데이터를 활용하면 스마트 폰 사용자 각 개인별로 어떤 모바일 애플리케이션을 사용했는지의 여부, 사용 순서, 사용 시작 시간과 종료 시간을(start time, end time) 알 수 있다. 특히 사용 시간은 YYMMDD뿐만 아니라 초(Second) 단위로 기록되기 때문에 애플리케이션별 이용 시간(time spent)과, 이용 월, 요일, 시간대 등과 같이 다양한 파생적 이용 정보 데이터를 생성하여 스마

```

{
  "global": {
    "settingsRevision": 3,
    "osVersion": "2.3.4",
    "appVersion": "1.2.0",
    "deviceManufacturer": "samsung",
    "deviceModel": "Nexus S",
    "telNum": "",
    "osName": "Android",
    "deviceId": "354494040409457" },
  "dgp": {
    "startTime": "2011\07\03 10:21:29",
    "endTime": "2011\07\03 12:12:32",
    "app": {
      "installedThisDgp": [],
      "uninstalledThisDgp": [],
      "session": [
        {
          "endResult": "appNonForeground",
          "endTime": "2011\07\03 10:30:43",
          "startTime": "2011\07\03 10:30:23",
          "name": "com.android.phone" },
        {
          "endResult": "deviceIdle",
          "endTime": "2011\07\03 10:30:53",
          "startTime": "2011\07\03 10:30:43",
          "name": "com.android.launcher"},
        ...
      ]
    }
  }
}
  
```

<Figure 3> Example of Mobile Log Data

트 폰의 이용 패턴에 대한 입체적 분석이 가능하게 한다. 나아가 각 애플리케이션은 구글 드라이브나 앱 스토아 등을 활용하면 애플리케이션별로 쇼핑 카테고리의 애플리케이션인지, 게임 카테고리의 애플리케이션인지와 같은 메타 정보를 얻을 수 있기 때문에 애플리케이션 유형별 사용 시간 및 사용 순서 등에 대한 분석이 가능해진다.

모바일 미디어 소비자의 행동 패턴 파악에 대한 연구에서 로그 데이터가 각광을 받기 시작한 것은 스마트 폰이 사람들의 행동에 깊숙이 자리하고 있다는 이유 외에도 데이터 자체의 정확도도 공헌하고 있다[15].

3.3 모바일 로그 데이터 수집 및 전 처리

이 연구를 위해 모바일 미디어 패널 회사인 마켓링크(Market Link)로부터 대규모 패널 데이터(large scale panel data)를 제공받았다. 마켓링크는 2019년 약 25,000명의 스마트 폰 로그 데이터 수집 패널을 구축하여 데이터를 수집하고 있다. 스마트 폰 이용 행동 로그(log) 데이터는 개인별로 저장되었다가 하루에 1번 일정한 시간에 배치(batch) 파일로 데이터 수집 서버로 전송된다. 수집된 25,000여 명의 패널 데이터는 매일 데이터 정제 가공 과정을 거쳐 데이터베이스화 되는데, 이때 최종적으로 분석에 사용되는 로그 데이터는 1개월 기준으로 볼 때, 1일~30일/31일 중 28일 이상 안정적으로 로그 데이터를 송신한 패널의 데이터만 포함한다. 따라서 이 연구에서는 마켓링크로부터 제공받은 25,000여 명의 로그 데이터 중 안드로이드 스마트 폰 사용자 9,189명의 데이터를 안정적 데이터로 추출하였다. 안드로이드(Android) 운

영체계(OS)는 대한민국에서 약 93%에 이르는 대중적인 플랫폼이며, 아이오에스(iOS)에 비해 상대적으로 안드로이드 스마트 폰의 로그 데이터 수집기가 보다 잘 발달되어 있으며, 로그 데이터 구조가 단순하기 때문에 데이터 정제 및 분류가 용이하고 보다 정확한 모바일 애플리케이션 데이터를 수집할 수 있다는 강점이 있다.

또한 안정적인 데이터를 송신한 9,189 패널들의 데이터 중에서 1,200명의 데이터 패널을 선정하여 분석에 포함했다. 이는 대한민국 인구 분포를 반영할 수 있도록 성별, 연령대별 분포를 반영하여 특정 인구통계적 집단 특성이 미치는 영향력을 최소화하고 표본의 대표성을 확보할 수 있도록 하기 위함이다. 이 1,200명의 표본 중 이상치(outlier)를 제거한 1072 표본을 최종적으로 분석에 포함하였다.

한편, 분석 대상 기간은 2019년 1년 중에서 2019년 11월 2일부터 11월 29일까지 4주간 사용한 스마트 폰 로그 기록 데이터를 분석에 활용하였다. 이는 11월은 기념일이나 법정 공휴일이 없어서 계절성(Seasonality)의 영향이 없는 유일한 월(month)이기 때문이다. 요일 속성에서는 월요일~일요일까지의 요일 영향력을 최소화하기 위해 11월 2일(토)부터 29일(금)까지의 28일의 데이터를 사용하여 월요일부터 일요일까지의 요일들이 각각 4번씩 동일한 횟수로 포함될 수 있도록 하였다.

1,072명의 4주간 스마트 폰 이용 데이터는 13,238개의 모바일 애플리케이션의 사용 기록 정보를 담고 있다. 모바일 애플리케이션은 구글 앱스토어 기준으로 32개의 카테고리로 분류하고 있으며 각 카테고리 별 대표 애플리케이션은 다음 <Table 2>와 같다.

<Table 2> Mobile Application Typology

No	Category	Repreentative Application	Count
1	ART_AND_DESIGN	‘이비스 페인트’, ‘스틱맨:만화그리기’ 등	21
2	AUTO_AND_VEHICLES	‘보배드림’, ‘오일나우:차계부’, 등	74
3	BEAUTY	‘화해’, ‘페이스보드’, ‘뷰티포인트’, 등	40
4	BOOKS & REFERENCE	‘네이버’, ‘에스24’, ‘밀리의서재’, 등	178
5	BUSINESS	‘사람인’, ‘워크넷’, ‘알바몬’, 등	420
6	COMICS	‘네이버웹툰’, ‘카카오웹툰’, 등	46
7	COMMUNICATION	‘카카오톡’, ‘다음’, ‘텔레그램’, 등	395
8	DATING	‘빠른톡’, ‘바닐라브릿지’, ‘글랩’, 등	74
9	EDUCATION	“네이버 사전’, ‘캐치잇 잉글리시’, 등	827
10	ENTERTAINMENT	‘Netflix(넷플릭스)’, ‘wavve(웨이브)’, 등	509
11	EVENTS	‘돼지저금폰’, ‘개이득’, ‘숫자 뽑기’, 등	17
12	FINANCE	‘토스’, ‘우리카드’, ‘페이북’, ‘PASS’, 등	708
13	FOOD_AND_DRINK	‘스타벅스’, ‘배달요기요’, ‘마켓컬리’, 등	87
14	GAMES	‘리니지M’, 등	2169
15	HEALTH_AND_FITNESS	‘캐시슬라이드스텝업’, ‘캐시닥’, ‘만보기’ 등	318
16	HOUSE_AND_HOME	‘호갱노노’, ‘직방’, ‘네이버 부동산’	80
17	LIBRARIES_AND_DEMO	‘토니스트리트(TONYSTREET)’	40
18	LIFESTYLE	‘OK캐쉬백’, ‘해피포인트’, ‘HPoint’ 등	1073
19	MAPS & NAVIGATION	‘T map’, ‘카카오지하철’, ‘아틀란’	223
20	MEDICAL	‘닥터나우’, ‘모아베베’, ‘365mc멤버십’	69
21	MUSIC_AND_AUDIO	‘멜론’, ‘FLO’, ‘Google Play Music’, 등	269
22	NEWS & MAGAZINES	‘한국경제’, ‘YTN’, 등	120
23	PARENTING	‘BabyTime:수유’, ‘놀이의 발견’, 등	71
24	PHOTOGRAPHY	‘QuickPic포토 갤러리’, ‘삼성 갤러리’, 등	244
25	PRODUCTIVITY	‘Microsoft Word’, 등	423
26	SHOPPING	‘쿠팡’, ‘11번가’, ‘SSG.COM’, 등	1092
27	SOCIAL	‘Instagram’, ‘Facebook’, 등	318
28	SPORTS	‘LIVE스코어’, ‘골프락’, ‘라이브스코어’	120
29	TOOLS	‘자명종’, ‘계산기’, ‘국세청 홈택스’	838
30	TRAVEL_AND_LOCAL	‘네이버 지도’, ‘카카오맵’, ‘지도’	275
31	VIDEO_PLAYERS	‘U+모바일tv’, ‘MX 플레이어’, ‘틱톡’	237
32	WEATHER	‘웁기날씨’, ‘Windy.com’	66

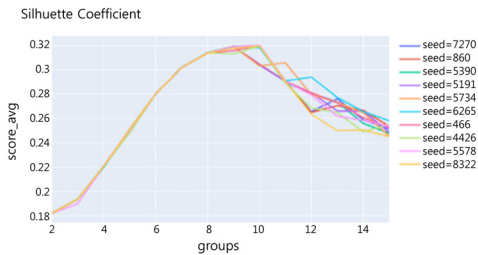
3.4 클러스터 개수 k의 결정

실루엣 인덱스를 이용하여 클러스터링 결과가 재현성(reproducibility; 타당성, validity)을 갖는가를 확인하기 위해서 1,200명의 전체 데이

터 셋을 반분하여 두 개의 데이터 셋으로 나누 후, 데이터 셋에 2부터 15까지의 그룹 개수 즉 k값을 부여하면서 k-means 모델을 각각 실행 하여 두 개의 클러스터링 결과를 얻어냈다. 이 때 k-means 클러스터링이 초기값에 영향받는

것을 최소화하기 위해서 10개의 초기값을 무작위로 부여하여 이 중 가장 안정적으로 고득점을 획득하는 k 값을 선정하는 방식을 택하였다.

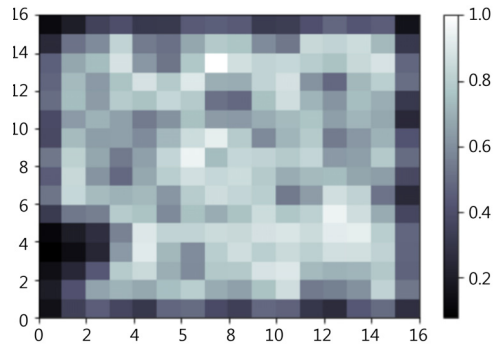
최적의 k 값을 결정하기 위해 두 훈련 자료에서 데이터들이 얼마나 비슷하게 분류가 되었는지 확인하기 위해 실루엣 지수(Silhouette Index)는 아래 <Figure 4>와 같이 도출되었다. <Figure 4>에서와 같이 실루엣 지수(Silhouette Index)는 $k=8$ 일 때 가장 높은 성과 수준(Performance)을 보이고 있으므로, 이 연구에서는 최종적으로 미디어 레퍼토리 유형의 개수, k 값을 8개로 확정하였다.



<Figure 4> Silhouette Index According to the Number of Clusters k and Seed

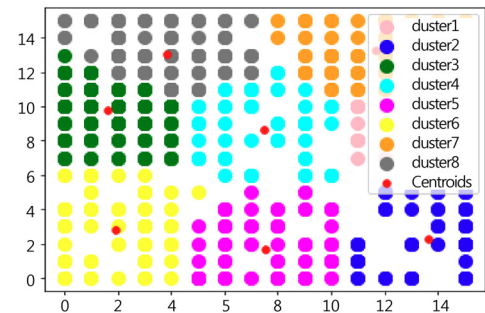
3.5 2단계 군집분석

모바일 미디어 레퍼토리를 규명하기 위해 32개 모바일 미디어 애플리케이션 사용 시간을 기준으로 자기조직화지도(Self-Organized Map)와 k -means 클러스터링을 실시하였다. 먼저 32개 변수로 나타나는 고차원의 관찰값들에 대하여 자기조직화분석을 실시한 결과 수렴된 프로토타입 벡터를 얻었다. 이를 저차원, 즉 이차원(2-dimension)의 SOM 그리드(grid) 상에 매핑해 본 결과 아래 <Figure 5>와 같이 프로토타입 벡터가 분포하고 있음을 알 수 있었다.



<Figure 5> Prototype Vector According to Self-Organizing Map

위 <Figure 5>의 프로토타입 벡터들의 2차원 그리드 상의 위치 값을 입력 값으로 사용하고, 앞에서 구한 k 값 8을 적용하여 k -means 클러스터링을 진행하였다. 그 결과 각 군집들의 분포를 2차원 평면 상에 매핑해 본 결과 아래 <Figure 6>과 같이 8개의 군집들이 배타적으로 분포되어 있음을 확인할 수 있었다.



<Figure 6> Result of K-means Clustering using Prototype Vector

3.6 모형 평가

이 연구에서는 2단계 군집분석과 함께 k -means clustering만 적용한 1단계 군집 분석을 실시하여, 두 가지 분석 방법의 성능을 평가했다.

성능평가를 위해 먼저 1,072명의 표본을 40:40:20으로 무선추출기법을 이용하여 훈련 데이터, 테스트 데이터, 그리고 검증 데이터로 구분하여 데이터 셋1, 데이터 셋2, 데이터 셋3를 얻었다. 이 각각의 데이터 셋들에 1단계 k-means clustering만 적용한 군집 분석 결과1과 자기조직화지도와 k-means clustering을 단계적으로 실시한 군집 분석결과 2를 통하여 도출된 군집결과 값(Cluster membership)이 저장되었으며, 이 군집결과 값을 비교하여 랜드 인덱스(Rand Index)를 도출했다.

서 보는 바와 같이 k-means clustering만 실시할 경우의 랜드 인덱스(Rand Index)는 50%로 나타난 반면 Two-step 군집 분석을 실시한 경우에는 73%의 개체들이 두 개의 데이터 셋에서 같은 군집으로 묶이는 성능을 보였다.

이와 같은 분석 방법의 성능 평가를 통해 이 연구에서는 Two-step 군집 분석 결과를 채택하여 데이터 결과 해석을 진행하였다.

4. 연구 결과

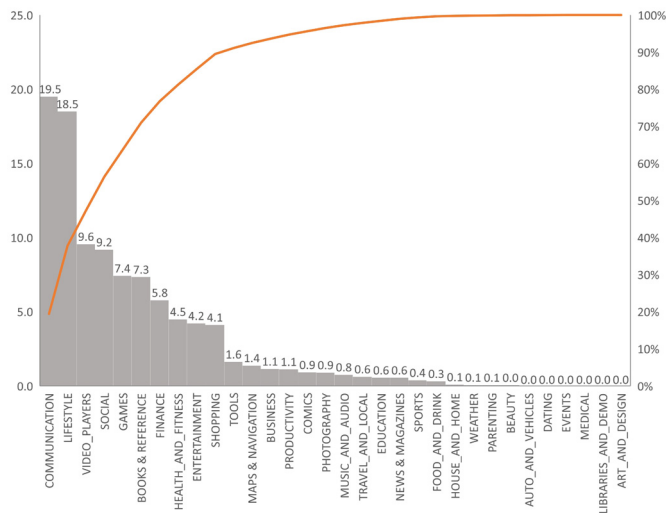
4.1 스마트 폰 애플리케이션 사용 행동 기초 통계 분석

스마트 폰 사용 시간은 월 평균 1006 시간으로 나타났다. 우리나라 사람들이 가장 많이 사용하는 애플리케이션은 32개 카테고리 중에서 ‘커뮤니케이션’ 카테고리로 월 평균 19.5 시간

<Table 3> Performance Comparison

No	Cluster Anlysis Method	Rand Index
1	k-means clustering	50%
2	SOM + k-means clustering	73%

위의 <Table 3>은 성능평가 지수로 활용한 랜드 인덱스(Rand Index)의 비교표이다. 위에



<Figure 7> Average Monthly Usage Time by Mobile Application Category (Unit: hours/Month/person, n = 1200)

사용되고 있으며, 그 다음이 ‘라이프스타일’ 애플리케이션에 월 평균 18.5시간을 사용하고 있으며, 세 번째는 ‘동영상 플레이어’ 월 평균 9.6시간을, 근소한 차이로 네 번째는 ‘소셜미디어’ 애플리케이션으로 월 평균 9.2시간, 다섯 번째는 ‘게임’ 애플리케이션으로 월 평균 7.4시간 사용하는 것으로 나타났다. 그 외, 도서 및 참고자료, 금융, 건강/운동, 엔터테인먼트, 쇼핑 애플리케이션도 많이 사용되는 애플리케이션이며, 각각 7.3시간, 5.8시간, 4.5시간, 4.2시간으로 상위 10위에 포함되었다(<Figure 7> 참조).

4.2 모바일 애플리케이션 미디어 레퍼토리

모바일 애플리케이션 미디어 이용자는 어떤 미디어 레퍼토리 군집으로 분류되는가라는 연구문제에 답하기 위해 이단계(Two-step) 군집 분석이 실시되었다. 실루엣 인덱스(Silhouette Index) 비교를 통해 k=8일 때 가장 재현성이 높아 최적의 클러스터 개수, 즉 미디어 레퍼토리의 수는 8개의 유형인 것으로 분석되었다. 8개의 모바일 애플리케이션 미디어 레퍼토리는 <Table 4>와 같다.

<Table 4> Mobile Application Media Repertoires

Cluster	Number of users (n)	Size (%)
1	361	33.7
2	100	9.3
3	93	8.7
4	134	12.5
5	57	5.3
6	140	13.1
7	78	7.3
8	109	10.2
Total	1,072	100.00%

<Table 4>에서 보듯이, 클러스터 1이 33.7%로 가장 큰 군집으로 나타났으며, 클러스터 6(13.1%)>클러스터 4(12.5%)>클러스터 8(10.2%)>클러스터 2(9.3%)>클러스터 3(8.7%)>클러스터 7(7.3%)>클러스터 5(5.3%) 순으로 크기를 갖는 8개의 클러스터로 분류되었다.

4.3 모바일 애플리케이션 미디어 레퍼토리 특성

4.3.1 모바일 애플리케이션 레퍼토리 집단별 카테고리 이용 행동

클러스터링 결과로 형성된 8개의 클러스터들은 <Table 5>에서와 같이 모바일 애플리케이션 카테고리 별 사용 시간에서 차이를 가지는 집단으로 분류된 것을 알 수 있다. 특히, ‘도서 및 참고자료(Books and reference)’, ‘만화(Comics)’, ‘커뮤니케이션(Communication)’, ‘엔터테인먼트(Entertainment)’, ‘금융(Finance)’, ‘음식(Food and drink)’, ‘게임(Game)’, ‘건강/운동(Health and fitness)’, ‘라이프스타일(Lifestyle)’, ‘지도와 내비게이션(Maps and navigation)’, ‘음악/오디오(Music and audio)’, ‘사진(Photography)’, ‘쇼핑(Shopping)’, ‘소셜미디어(Social)’, ‘도구(Tools)’, ‘여행 및 지역정보(Travel and local)’, ‘동영상 플레이어(Video Player)’ 등의 18개 카테고리 사용성은 ANOVA 검증 결과 미디어 레퍼토리 클러스터별로 유의미한 차이를 보였다(<Table 5>). 18개 카테고리 중 특히 유의수준 $p < .001$ 수준에서 집단 간 평균 사용 시간의 차이가 있는 것으로 평가된 모바일 애플리케이션 카테고리는 ‘도서 및 참고자료(Books and reference)’, ‘커뮤니케이션(Communication)’,

‘엔터테인먼트(Entertainment)’, ‘금융(Finance)’, ‘게임(Game)’, ‘건강/운동(Health and fitness)’, ‘라이프스타일(Lifestyle)’, ‘쇼핑(Shopping)’, ‘소셜 미디어(Social)’, ‘동영상 플레이어(Video Player)’ 등 10개 카테고리이었다. 따라서 클러스터링 결과 분류된 8개 미디어 레퍼토리 그룹은 이 10개

<Table 5> Mobile App Category Usage Time According to Each Type of Media Repertoire (Unit: hours/month)

Application categories	Total mean	Cluster								Sig*
		1	2	3	4	5	6	7	8	
ART_AND_DESIGN	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
AUTO_AND_VEHICLES	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	*
BEAUTY	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0	
BOOKS_AND_REFERENCE	7.3	12.3	4.5	5.1	9.0	4.9	11.2	5.2	5.2	***
BUSINESS	1.1	1.1	1.3	0.7	0.6	0.9	0.9	3.1	0.9	
COMICS	0.9	0.6	0.3	3.3	0.5	0.4	0.7	0.5	1.1	
COMMUNICATION	19.4	21.8	13.7	18.1	14.2	34.2	23.4	18.3	20.2	***
DATING	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	
EDUCATION	0.6	0.3	0.1	0.3	0.2	4.6	0.2	0.2	0.2	
ENTERTAINMENT	4.2	2.6	11.8	1.7	1.8	2.3	2.3	7.4	3.3	***
EVENTS	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	
FINANCE	5.7	4.9	4.6	3.3	4.1	5.9	4.1	17.0	4.2	***
FOOD_AND_DRINK	0.3	0.2	0.2	0.3	0.2	0.3	0.7	0.2	0.3	*
GAMES	7.4	4.7	3.6	29.6	3.2	3.1	6.6	3.2	5.6	***
HEALTH_AND_FITNESS	4.5	3.4	9.6	1.1	3.1	2.3	2.7	3.2	8.5	***
HOUSE_AND_HOME	0.1	0.2	0.0	0.0	0.1	0.1	0.2	0.0	0.0	
LIBRARIES_AND_DEMO	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
LIFESTYLE	18.4	18.9	24.4	10.6	30.3	17.6	14.0	18.3	9.1	***
MAPS_AND_NAVIGATION	1.4	1.0	4.8	1.0	1.1	0.7	0.9	0.3	0.6	**
MEDICAL	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
MUSIC_AND_AUDIO	0.8	0.5	0.4	0.5	0.3	0.3	0.6	0.4	2.7	*
NEWS_AND_MAGAZINES	0.6	0.2	0.1	0.1	1.7	0.7	0.2	0.1	0.7	
PARENTING	0.1	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0	*
PHOTOGRAPHY	0.9	0.9	0.4	0.6	0.6	0.6	2.2	0.6	1.0	**
PRODUCTIVITY	1.1	1.0	0.4	0.7	0.7	1.4	1.0	3.7	0.7	
SHOPPING	4.1	3.1	2.8	2.6	3.7	4.3	9.7	3.0	2.5	***
SOCIAL	9.1	9.4	10.7	5.0	5.4	7.2	7.1	5.9	20.8	***
SPORTS	0.4	0.1	0.0	2.5	0.1	0.0	0.1	0.2	0.2	
TOOLS	1.6	1.5	1.0	1.5	2.6	1.3	1.6	1.1	1.7	*
TRAVEL_AND_LOCAL	0.6	0.5	0.4	0.4	0.3	0.4	1.9	0.3	0.5	*
VIDEO_PLAYERS	9.5	10.6	4.5	11.1	16.1	5.6	7.4	7.5	9.7	***
WEATHER	0.1	0.0	0.0	0.0	0.0	0.5	0.0	0.0	0.0	
Total Usage Time(hours/month)	100.6	26.5	146.3	137.4	144.4	134.0	114.1	156.8	144.1	***

* $\rho < .05$, ** $\rho < .01$, *** $\rho < .001$.

애플리케이션 카테고리에서 특히 사용성 차이가 있는 집단으로 볼 수 있다.

각 모바일 미디어 레퍼토리 그룹의 모바일 애플리케이션 카테고리 사용 특성이 요약된 표 5에서 보듯이 클러스터의 크기가 가장 큰 클러스터 1은 한 달 평균 스마트 폰 사용 시간은 26.5시간으로(하루 평균 약 1시간) 가장 적은 사용 시간을 보여주고 있으며, 주로 커뮤니케이션과 라이프스타일 애플리케이션을 중심으로 스마트 폰을 사용하는 모습을 보여주고 있다. 두 번째로 크기가 큰 클러스터 6은 한 달 평균 스마트 폰 사용 시간은 114.1시간인데(하루 평균 약 4.8시간), 이 클러스터 6은 커뮤니케이션, 쇼핑, 도서 및 참고자료 카테고리에 전체 사용 시간의 23.4%, 9.7%, 11.2%를 사용하고 있는 것으로 나타난다. 이 클러스터의 사용성은 단일 카테고리에 집중되기 보다는 3개 이상의 카테고리에 분산된 사용성을 보인다. 클러스터 4는 동영상 플레이어와 라이프스타일 카테고리 애플리케이션 사용시간이 해당 그룹 스마트 폰 한 달 평균 총 사용시간 144.4시간의 16.1%를 차지하며, 라이프스타일 카테고리는 30.3%를 차지하는 것으로 나타났다. 클러스터 8은 한

달 평균 사용 시간은 144.1시간을 사용하는 그룹으로, 이 중 20.8%를 소셜미디어 카테고리에 사용하고 있으며, 클러스터 2는 한 달 평균 146.3시간의 스마트 폰 사용성을 보이는데, 이 중 엔터테인먼트가 11.8%, 건강/운동이 9.6%, 그리고 라이프스타일이 24.4%의 사용 시간을 차지하고 있다. 클러스터 3은 게임 중심 사용성을 보이는 그룹으로 스마트 폰 총 사용 시간 137.4 시간의 29.6%를 게임에 사용하고 있으며, 클러스터 7은 총 사용 시간이 156.8 시간인 고사용 그룹인데, 이중 금융, 비즈니스, 생산성 등의 애플리케이션에 각각 17.0%, 3.1%, 3.7%의 시간을 할애하는 사용성을 보이고 있다. 마지막으로 클러스터 5는 가장 그룹 크기가 작은 클러스터로 총 사용 시간 134.0 시간 중 34.2%를 커뮤니케이션 애플리케이션에 할애하는 커뮤니케이션 중심 사용성을 보이는 그룹이다.

<Table 5>와 같이 월 평균 스마트 폰 총 사용 시간과 모바일 애플리케이션 이용량의 절대 시간 및 상대적 사용 비중의 차이를 보여주는 결과를 종합적으로 분석하여 각 군집의 미디어 이용 행태 특성을 각각 <Table 6>과 같이 '커뮤니케이션 중심 저사용 그룹(Communication oriented

<Table 6> Clusters for Media Repertoires

(Unit: hours/month)

Cluster	Mobile Application Repertoires	Cluster Size(%)	Time Spent (hours/month)
1	Communication oriented light users	33.7	26.5
2	Entertainment/Health oriented heavy users	9.3	146.3
3	Game oriented users	8.7	137.4
4	Video player/Lifestyle users	12.5	144.4
5	Communication oriented users	5.3	134.0
6	Reference/Shopping oriented users	13.1	114.1
7	Finance/Business/Productivity oriented heavy users	7.3	156.8
8	Social media oriented users	10.2	144.1

light users)', '엔터테인먼트/건강 활용 다사용 그룹(Entertainment/Health oriented heavy users)', '게임 중심 그룹(Game oriented users)', '동영상 및 라이프스타일 중심 그룹(Video player & Lifestyle users)', '커뮤니케이션 중심 그룹(Communication oriented users)', '참고자료 및 쇼핑 중심 그룹(Reference & Shopping oriented users)', '금융/비즈니스/생산성 활용 다사용 그룹(Finance/Business/Productivity oriented heavy users)' 그리고 '소셜미디어 중심 그룹(Social media oriented users)'으로 명칭을 부여함으로써 최종적으로 모바일 애플리케이션의 미디어 레퍼토리 유형을 도출하였다.

4.3.2 모바일 애플리케이션 레퍼토리 집단별 데모그래픽 특성

두 번째 연구 문제인 모바일 미디어 레퍼토

리 유형별 사용자 특성은 무엇인가에 대해 답하기 위해 미디어 레퍼토리 특성에 따른 8개의 클러스터 별 인구통계학적 특성 차이를 살펴보았다. <Table 7>에서와 같이 성 및 연령 구성비에 대한 Chi-square test 결과 유의수준 $p < .05$ 수준에서 유의미한 차이를 보였다.

Chi-square test 결과를 바탕으로 군집 분석 결과 집단간 인구통계학적 특성 차이를 살펴보면, '커뮤니케이션 중심 저사용' 그룹은 남성(49.3%)과 여성(50.7%)의 비중이 유사하나 40대 이상의 고연령층(68.2%)이 많이 분포되어 있음을 알 수 있었다. '엔터테인먼트/건강 활용 다사용' 그룹은 여성(54.0%)의 비중이 다소 높은 가운데, 40대가 타 군집 대비 많이 분포되어 있으며(29.0%), '게임 중심' 그룹은 남성(61.3%)이 여성(38.7%) 보다 많고 20대(23.7%)와 30대(29.0%)의 젊은 층을 중심으로 분포하고 있었으며, '동영

<Table 7> Demographic Characteristics of Media Repertoire Groups Identified (Unit: hours/month)

	Communication oriented light users	Entertainment/Health oriented heavy users	Game oriented users	Video player/Lifestyle users	Communication oriented users	Reference/Shopping oriented users	Finance/Business/Productivity oriented heavy users	Social media oriented users	Sig.
Gender									**
Female	49.3%	46.0%	38.7%	49.3%	61.4%	61.4%	37.2%	46.8%	
Male	50.7%	54.0%	61.3%	50.7%	38.6%	38.6%	62.8%	53.2%	
Age, years									**
20~29	14.7%	15.0%	23.7%	17.9%	10.5%	20.7%	14.1%	32.1%	
30~39	17.2%	19.0%	29.0%	19.4%	14.0%	13.6%	14.1%	19.3%	
40~49	21.9%	29.0%	20.4%	21.6%	24.6%	20.7%	29.5%	19.3%	
50~59	25.5%	22.0%	12.9%	20.1%	24.6%	27.1%	26.9%	20.2%	
60~69	20.8%	15.0%	14.0%	20.9%	26.3%	17.9%	15.4%	9.2%	

* $p < .05$, ** $p < .01$, *** $p < .001$.

상 및 라이프스타일 중심' 그룹은 남녀의 비중은 비슷하지만(남성 49.3%, 여성 50.7%) 40대 이상의 고연령층(62.6%)의 분포가 높게 나타나고 있어 '커뮤니케이션 중심 저사용' 그룹과 유사한 분포를 보였다. 한편 '커뮤니케이션 중심' 그룹은 남성(38.6%)보다 여성(61.4%)이 많고 40대 이상의 고연령층(75.5%)이 가장 많으며, '참고자료 및 쇼핑 중심' 그룹은 여성(61.4%)이 큰 비중을 차지하는 가운데, 20대(20.7%), 40~50대(40대 20.7%, 50대 27.1%)가 상대적으로 많이 분포되어 있다. '금융/비즈니스/생산성 활용 다사용' 그룹은 여성(37.2%)보다 남성(62.8%)이 큰 비중을 차지하고 있으며 40대(29.5%)와 50대(26.9%)가 크게 분포되어 있으며, 마지막으로 '소셜미디어 중심' 그룹은 20대(32.1%)가 많은 비중을 차지하는 특징을 보였다. 이와 같이 미디어 레퍼토리 8개의 클러스터는 각각 인구통계학적 특성에 있어 유의미한 차이를 가지고 차별적으로 형성되어 있음을 확인하였다.

5. 결 론

이 연구는 현대인에게 가장 일상적이고 보편적인 미디어인 스마트 폰과 모바일 애플리케이션이라는 비히클을 통해 커뮤니케이션, 엔터테인먼트, 비즈니스 등 통합된 융합 미디어로서 기능하고 있는 점에 주목하여 미디어 사용자들은 어떻게 레퍼토리를 구성하여 미디어를 이용하고 있는지를 파악하고자 하는 문제 제기에서 시작되었다.

이를 위해 2019년 11월, 4주 동안의 개인별 모바일 이용행동 로그 데이터를 모바일 애플리케이션 카테고리별 미디어 이용량을 중심으로

군집 분석을 실시하고, 최종적으로 8개의 모바일 미디어 레퍼토리 유형별 집단을 분류하였다. 8개의 각 집단은 모바일 애플리케이션 카테고리별 절대적 이용량과 타 집단과의 비교를 통한 상대적 이용량을 비교하여, '커뮤니케이션 중심 저사용 그룹', '엔터테인먼트/건강 활용 다사용 그룹', '게임 중심 사용 그룹', '동영상 및 라이프스타일 중심 그룹', '커뮤니케이션 중심 그룹', '참고자료 및 쇼핑 중심 그룹', '금융/비즈니스/생산성 활용 다사용 그룹', '소셜미디어 중심 그룹'으로 명명했다. 이 8개 모바일 미디어 레퍼토리 그룹은 성별 및 연령별의 데모그래픽 분포에서도 차이를 보였다. '커뮤니케이션 중심 저사용 그룹'과 '동영상 및 라이프스타일 중심 그룹'은 남성과 여성은 유사하게 분포되었으나 연령별로는 40대 이상의 고연령층이 많았던 반면, '게임 중심 그룹'과 '금융/비즈니스/생산성 활용 고사용 그룹'은 여성보다 남성이 많이 분포되는 등 집단간 차이를 보였다.

이와 같은 모바일 미디어 레퍼토리 유형 분류 결과는 스마트 폰에서 애플리케이션이라는 비히클의 선택이 보편적이지 않고 개인별 이용 동기에 따라 선택적으로 차이를 보이고 있음을 명확하게 보여준다. 이는 모바일 플랫폼을 중심으로 마케팅 커뮤니케이션을 집행하거나 마케팅 예산을 분배해야 하는 실무자들에게 매체 효율성 기반의 타겟 마케팅에 대한 실무적 가이드를 제공해 줄 수 있다는 점에서 의의가 있다. 미디어 레퍼토리를 파악하는 것은 성, 연령과 같은 인구통계학적 변수를 바탕으로 모바일 마케팅 타겟 오디언스를 추출하여 마케팅 커뮤니케이션을 집행하는 기존의 방식을 정교화하는 새로운 타겟팅 변수를 생성하는 것을 의미한다. 예를 들어 젊은 여성을 타겟으로 새로운

‘패션 e커머스’ 애플리케이션을 출시해야 하는 경우라면, ‘20대 여성’이라는 인구통계학적 특성 이외에 ‘소셜미디어 중심 그룹’이라는 모바일 미디어 사용성적 특성을 추가한다면, 신규 서비스의 확산 가능성을 보다 제고시킬 수 있는 타겟 오디언스를 추출할 수 있을 것이다. 이때 이 연구에서 도출된 ‘소셜미디어 중심’ 그룹의 평균 스마트폰 총 사용 시간과 소셜미디어 애플리케이션 카테고리들을 포함한 주 사용 애플리케이션 카테고리들의 평균 사용 시간은 타겟 오디언스 추출에 대한 가이드를 제공할 수 있을 것이다. 이와 같은 방식으로 미디어 사용 특성에 따른 미디어 사용자 집단 분류는 타겟 마케팅을 효율화하고자 하는 실무 담당자들에게 시사점을 제공할 수 있다.

모바일 미디어 이용 행동의 유형 분류 연구는 전자거래 관련 커뮤니티에게도 실무적 가이드를 제시할 수 있다. 모바일 이용 행동을 분류하는 것은 전자거래 서비스를 이용하는 고객을 유형분류 하고, 각 고객 유형에 맞는 고객 관리 서비스를 집행해야 하는 실무진에게 반드시 필요한 사항이다. 기존에는 전자 거래 서비스를 제공하는 실무진은 주로 고객의 유형을 거래 기반으로 분류를 하는 것이 전형적이었다. 그러나 고객의 전자거래 사이트에서의 행동 로그 데이터를 기반으로 고객을 분류하게 될 경우 새로운 시각으로 고객의 구조를 파악할 수 있으며, 나아가 각 고객 집단에 적합한 서비스 또는 마케팅 의사결정을 차별적으로 집행할 수 있다는 점에서 효율성이 높아지는 효과를 기대할 수 있다.

한편, 학문적으로도 역시 인쇄 및 전파매체, 또는 PC와 같은 전통적 미디어를 중심으로 한 미디어 레퍼토리 연구를 ‘스마트 폰’이라는 뉴미

디어와 ‘모바일 애플리케이션’이라는 비히클로 확장하여 심층적으로 미디어 이용 패턴을 규명했다는 점에서 의미를 가지며 나아가 다양한 후속 연구를 이끌어 낼 수 있다는 점에서 연구의 의의가 있다. 미디어 레퍼토리는 성, 연령과 같이 미디어 전체 사용자를 고유 특성이 존재하는 하위 그룹으로 세분화하는 변수의 성격을 갖는다. 따라서 미디어 레퍼토리 그룹에 따른 ‘라이프스타일이나 가치관’의 차이, ‘쇼핑 제품, 쇼핑 빈도, 지출 규모 등’과 같은 소비 특성의 차이 또는 ‘정치 성향 및 투표 성향’의 차이 규명과 같은 다양한 후속 연구를 시도해 볼 수 있다. 이는 기존 전통적 미디어를 중심으로 하는 미디어 레퍼토리에 따라 미디어 사용자의 정치, 경제, 사회, 문화적 인식 및 행동 차이를 유발하는 차이와 관계를 파악하고자 하는 기존 연구들이 ‘스마트 폰이라는 뉴미디어’에 초점을 두어 후속 연구를 시도해 볼 수 있음을 의미한다.

또한, 이 연구는 기존의 k-means clustering에 의존적이었던 군집 분석을 SOM(Self-Organized Map)을 이용하여 프로토크터를 추출하고 이 프로토크터를 이용하여 k-means clustering을 실시하는 이단계 접근법(Two-step approach)를 시도했다. 이러한 접근법은 기존 k-means clustering이 갖고 있는 ‘이상치(outlier)’나 ‘결측치’에 민감했던 한계점을 극복하고 더 나은 성능의 분석 결과를 도출하고 있음을 보여준다는 점에서 방법론적으로 기여를 갖는다. 특히 설문지 기반의 미디어 이용 행동 측정 데이터가 아니라 기계적으로 측정되는(Machine to Machine) 빅데이터를 분석해야 되는 뉴미디어 연구자들에게 이단계 접근법(Two-step approach)은 함의가 크다고 볼 수 있다. 이는 기계적 측정(Machine to Machine) 빅데이터는 이상치가 다수 나타나

는 매우 편포된 분포를 갖는데, 이러한 상황에서 문제를 해결할 수 있는 방법론을 시도하고 검증했다는 점에서 이 연구는 학문적으로도 그 의미를 갖고 있다.

이 연구는 모바일 미디어 애플리케이션 이용 로그 빅데이터를 분석하여 애플리케이션 카테고리 이용 시간에 따라 이용 유형을 분류한 연구로서 기존의 설문 데이터 기반 모바일 미디어 연구보다 정교하며 방법론적으로도 정확성을 높였다는 점에서 의미를 찾을 수 있었다. 이러한 연구 결과들이 미디어로서의 스마트 폰과 비히클로서의 애플리케이션에 대한 이해를 높이고 다양한 후속 연구를 통해 뉴미디어와 미디어 빅데이터 분석 방법론에 대한 다양한 제안이 이끌어 지기를 기대한다.

References

- [1] Boase, J., and Ling, R., "Measuring mobile phone use: Self-report versus log data," *Journal of Computer-Mediated Communication*, Vol. 18, No. 4, pp. 508-519, 2013.
- [2] Cheil Worldwide Inc, "Advertising Yearbook 2021," 2021.
- [3] Choi, Y., Park, J., and Lee, H., "Factors affecting smart phone application's functional composition and repertoires," *Social Science Review*, Vol. 46, No. 1, pp. 163-198, 2015.
- [4] Ferguson, D. A., and Perse, E. M., "The world wide web as a functional alternative to television," *Journal of Broadcasting & Electronic Media*, Vol. 44, No. 22, pp. 155-174, 2000.
- [5] Gan, G., and Ng, M. K. P., "K-means clustering with outlier removal," *Pattern Recognition Letters*, Vol. 90, pp. 8-14, 2017.
- [6] Han, S. P., Park, S., and Oh, W., "Mobile app analytics: A multiple discrete-continuous choice framework," *MIS Quarterly*, Vol. 40, No. 4, pp. 989-1008, 2018.
- [7] Ha, S., and Geum, Y., "Categorizing sub-categories of mobile application services using network analysis: A case of health-care applications," *The Journal of Society for e-Business Studies*, Vol. 25, No. 3, pp. 15-40, 2020.
- [8] Ha, S., Oh, J., and Lee, B., "The analysis of advertisement effect in smart phone environment: The comparison of users with providers of commercial," *The Journal of Society for e-Business Studies*, Vol. 16, No. 4, pp. 221-239, 2011.
- [9] Heeter, C., "Program selection with abundance of choice: A process model," *Human Communication Research*, Vol. 12, pp. 126-152, 1985.
- [10] Hu, M., and Lee, Y., "Reproducibility assessment of k-means clustering and applications," *The Korean Journal of Applied Statistics*, Vol. 17, No. 1, pp. 135-144, 2004.
- [11] Jung, J., Kim, Y., and Chan-Olmsted, S., "Measuring usage concentration of smartphone applications: Selective repertoire in a marketplace of choices," *Mobile Media & Communication*, Vol. 2, No. 3,

- pp. 352-368, 2014.
- [12] Kang, N., Lee, J., and Lee, H., "Classifying media repertoires groups using k-mean cluster analysis method," *Korean Journal of Broadcasting and Telecommunication Studies*, Vol. 22 No. 2, pp. 7-46, 2008.
- [13] Kim, E., Ha, Y., and Park, W., "A study on media repertoire and lifestyle," *The Korean Journal of Advertising and Public Relations*, Vol. 11, No. 1, pp. 61-95, 2009.
- [14] Kim, S., and Kim, J., "A new cluster validity index based on connectivity in self-organizing map," *The Korean Journal of Applied Statistics*, Vol. 33, No. 5, pp. 591-601, 2020.
- [15] Kim, Y., Choi, Y., Kim, S., Park, K., and Park, J., "A study on method for user gender prediction using multi-modal smart device log data," *The Journal of Society for e-Business Studies*, Vol. 21, No. 1, pp. 147-163, 2016.
- [16] Neuendorf, K. A., Atkin, D. J., and Jeffres, L. W., "Reconceptualizing channel repertoire in the urban cable environment," *Journal of Broadcasting & Electronic Media*, Vol. 45, No. 3, pp. 464-482, 2010.
- [17] Ømen, J., and Thorhauge, A. M., "Smartphone log data in a qualitative perspective," *Mobile Media & Communication*, Vol. 3, No. 3, pp. 335-350, 2015.
- [18] Shim, M., "A study on media repertoire," *Korean Journal of Broadcasting and Telecommunication Studies*, Vol. 21, No. 2, pp. 351-390, 2007.
- [19] Van Rees, K., and Van Eijck, K., "Media repertoires of selective audiences: The impact of status, gender, and age on media use," *Poetics*, Vol. 31, No. 5-6, pp. 465-490, 2003.
- [20] Vesanto, J., and Alhoniemi, E., "Clustering of the self-organizing map," *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 11, pp. 586-600, 2000.
- [21] Webster, J. G., and Lin, S. F., "The internet audience: Web use as mass behavior," *Journal of Broadcasting & Electronic Media*, Vol. 46, No. 1, pp. 1-12, 2002.

저 자 소 개



권성은 (E-mail: sekwon@oto1partners.com)
1990년 연세대학교 지질학과 (학사)
1994년 연세대학교 본 대학원 신문방송학과 (석사)
2008년 연세대학교 정보대학원 정보시스템학 (박사)
1994년~2016년 제일기획, LG전자, KANTAR, nielsen 등
2016년~현재 (주)제로투원파트너스 대표
관심분야 인공지능, 빅데이터, Business Analytics, 마케팅 의사결정, 정보시스템



장서인 (E-mail: sijang@oto1partners.com)
2019년 숙명여자대학교 통계학과 (학사)
2021년 숙명여자대학교 통계학과 (석사)
2021년~현재 (주)제로투원파트너스 데이터인텔리전스연구소 연구원



황보현우 (E-mail: scott@hanafn.com)
1998년 연세대학교 행정학과 (학사)
2009년 연세대학교 행정대학원 (석사)
2017년 연세대학교 정보대학원 정보시스템학 (박사)
2009~2018년 코오롱베니트(주) 빅데이터분석팀장
2018~2020년 ㈜하나벤처스 경영전략본부장
2020~2021년 한남대학교 글로벌IT경영학과 조교수
2021년~현재 하나금융그룹 CDO(Chief Data Officer)