

네트워크 분석을 이용한 애플리케이션 서비스 하위 카테고리 분류: 헬스케어 어플리케이션 중심으로

Categorizing Sub-Categories of Mobile Application Services using Network Analysis: A Case of Healthcare Applications

하소희(Sohee Ha)*, 김영정(Youngjung Geum)**

초 록

모바일 애플리케이션 서비스 시장의 폭발적으로 성장함에 따라 애플리케이션 서비스를 고객과 개발자 관점에서 분류하는 것이 필요한 실정이다. 그러나 모바일 애플리케이션 서비스의 체계적 분류에 관한 연구는 제한적이다. 이에 본 연구에서는 네트워크 모듈성 분석을 통하여 모바일 애플리케이션 서비스의 분류하고 하위 카테고리를 제안하고자 한다. 구글플레이(GooglePlay)를 통해 총 1,607개의 헬스케어 관련 애플리케이션 서비스를 수집한 후 각 애플리케이션 서비스의 설명(description) 텍스트를 활용하여 유사도를 측정하고 이를 바탕으로 네트워크 분석을 수행하였다. 모듈성 분석을 수행하여 전체 네트워크의 커뮤니티를 탐지한 후, 각 클러스터를 기반으로 서비스 분류법을 도출한다. 이번 연구는 모바일 애플리케이션 서비스를 체계적으로 탐색하고자 하는 고객과 모바일 애플리케이션 서비스의 트렌드를 분석하고자 하는 개발자 모두에게 도움이 되는 서비스 분류에 대한 체계적인 접근방식이 될 것으로 기대된다.

ABSTRACT

Due to the explosive growth of mobile application services, categorizing mobile application services is in need in practice from both customers' and developers' perspectives. Despite the fact, however, there have been limited studies regarding systematic categorization of mobile application services. In response, this study proposed a method for categorizing mobile application services, and suggested a service taxonomy based on the network clustering results. Total of 1,607 mobile healthcare services are collected through the Google Play store. The network analysis is conducted based on the similarity of descriptions in each application service. Modularity detection analysis is conducted to detects communities in the network, and service taxonomy is derived based on each cluster. This study is expected to provide a systematic approach to the service categorization, which is helpful to both

본 연구는 서울과학기술대학교 교내연구의 지원으로 수행되었음.

* First Author, Master course of Datascience, Seoul National University of Science and Technology (newhap.in.here@ds.seoultech.ac.kr)

** Corresponding Author, Associate Professor, Dept. of Industrial & Systems Engineering, Seoul National University of Science and Technology (yjgeum@seoultech.ac.kr)

Received: 2020-06-01, Review completed: 2020-07-28, Accepted: 2020-08-05

customers who want to navigate mobile application service in a systematic manner and developers who desire to analyze the trend of mobile application services.

키워드 : 서비스 분류, 텍스트 마이닝, 모바일 애플리케이션 서비스, 네트워크 분석, 헬스케어 서비스
Service Taxonomy, Text Mining, Service Categorization, Mobile Application Service, Network Analysis, Healthcare Service

1. 서 론

최근 몇 년 동안 스마트폰과 태블릿과 같은 모바일 디바이스의 빠른 성장과 함께 모바일 애플리케이션 시장도 빠르게 성장하기 시작했다[27]. 모바일 디바이스만 소유하고 있으면 언제 어디서든 사용하기 쉽다는 장점으로 인해 오늘날 많은 사람들이 애플리케이션 서비스 플랫폼을 통해 애플리케이션 서비스를 다운 받아 사용하고 있다[11]. 2019년 기준으로 전 세계 애플리케이션 서비스 다운로드 수는 약 2040억 회이며, 2016년 이후 3년 동안 전 세계 애플리케이션 서비스 다운로드 성장률은 45%에 달하는 실정이다[1].

모바일 애플리케이션 시장이 성장함에 따라 모바일 애플리케이션 서비스와 관련된 연구들이 다양한 방면으로 수행되어 왔다. 모바일 애플리케이션 서비스 관련 연구 초창기에는 애플리케이션 서비스를 자유롭게 사고팔 수 있는 온라인 장터(Marketplace)를 의미하는 애플리케이션 서비스 플랫폼 자체에 대해 분석하는 연구들이 대부분이었다 [10, 14]. 이들 연구에서는 주로 비즈니스 확장, 차별적 서비스 제공, 마케팅, 기술 등 다양한 관점에서 애플리케이션 서비스 플랫폼을 이해하고 분석하고 이를 통해 서비스 프레임워크를 제안하거나 기업 성과 제고 방안을 모색하는데 초점이 맞춰져 있다.

이후 Apple 앱 스토어(App Store)에 이어 Google Play Store, BlackBerry Store 등 다양한 애플리케이션 서비스 플랫폼이 출시됨에 따라 모바일 애플리케이션 서비스의 개발 및 공급이 급속도로 확대되었다. 이에 따라 플랫폼 자체를 이해하고 분석하는 비즈니스 관점의 연구에서, 다양하고 새로운 애플리케이션 서비스의 개발 현황과 트렌드를 이해하는 것에 초점을 두는 서비스 관점의 연구들이 수행되기 시작했다[20]. 또한 특정 도메인 모바일 애플리케이션 서비스의 유형 및 특성을 분석하거나[15, 30], 더 나아가 유형에 따른 애플리케이션 서비스 속성의 중요 요인을 도출하고 이를 애플리케이션 설계 프로세스에 반영하여 추후 사용자의 만족도를 높이하고자 하는 연구[16], 애플리케이션 서비스의 시각적인 요소에 관한 연구 [18] 등이 수행되기도 하였다. 일부 연구에서는 애플리케이션 서비스 플랫폼 데이터를 바탕으로 새로운 서비스를 발굴, 기획하고자 하는 시도가 수행되기도 하였다[8, 19].

이렇듯 애플리케이션 서비스 플랫폼 내 서비스를 심층적으로 분석하고 활용하는 다양한 유형의 기존 연구에도 불구하고, 실무적인 측면에서 플랫폼 내 서비스의 구조 및 유형을 체계적으로 분석하기 위한 연구는 아직 부족한 실정이다. 이는 모바일 애플리케이션 서비스 시장이 급격히 성장함에 따라 소비자와 개발자

모두에게 심각한 문제로 인식되고 있다.

먼저 소비자 측면에서, 애플리케이션 서비스의 수가 폭발적으로 증가함에 따라 소비자들은 애플리케이션 플랫폼 내에서 원하는 애플리케이션 서비스를 탐색하고 발견하는 것에 많은 어려움을 호소하고 있다. 현재 구글 플레이 스토어(Google Play) 기준으로 제공되는 카테고리 종류는 총 37개에 불과하며, 이 중 2개에서만 하위 카테고리가 제공되고 있는 실정이다. 즉, 제공되는 애플리케이션 서비스의 숫자에 비해 이를 체계화하여 전달하는 체계가 매우 미흡한 것이다. 또한 카테고리 당 소비자에게 노출되는 애플리케이션 서비스의 개수는 약 200개 내외에 불과하기 때문에 현재 등록되어 있는 약 280만 개 애플리케이션 서비스 중 소비자들에게는 약 1만 2천(약 0.4%) 개만 보이고 있다. 애플리케이션 플랫폼 내에서 제공되는 다양한 서비스에 대한 소비자 접근성을 높이기 위해서는 각 서비스 내 하위 카테고리를 특성에 맞게 재정의하여 충분한 정보를 제공하는 것이 요구된다.

개발자 측면에서는 현재 제공되는 카테고리의 수가 실질적으로 제공되는 서비스의 수 및 서비스 유형에 비해 매우 적어 애플리케이션 서비스의 노출이 실질적으로 매우 어렵고, 이에 따라 개발된 서비스를 잠재적 소비자들에게 효과적으로 전달하기 위한 많은 노력이 요구된다. 일반적으로 애플리케이션 서비스가 한 카테고리의 상위 200개 안에 도달해야만 잠재적인 소비자들에게 노출될 수 있기 때문에 이러한 문제는 더욱 심각하다고 볼 수 있다[13]. 뿐만 아니라 서비스 개발자 측면에서, 현재 제공되고 있는 서비스의 주요 유형 및 특성을 쉽게 파악하고 이들에 대한 체계적 분석이 수반되어야 새로운 서비스의 개발이 원활하게 이루어진

다는 것을 감안하면 애플리케이션 서비스의 하위 카테고리 재정의는 서비스 혁신의 관점에서도 매우 중요한 의미를 가진다.

그러나 이러한 문제에도 불구하고, 플랫폼 내의 애플리케이션 서비스를 체계적으로 분석하여 재분류하기 위한 실질적 연구는 매우 부족한 실정이다. 일부 연구에서 플랫폼 내에 존재하는 애플리케이션을 분류하여 서비스 유형을 세분화한 연구가 시도된 바 있지만[28], 서비스 플랫폼 내에 존재하는 애플리케이션 서비스를 수집하여 LDA(Latent Dirichlet Allocation) 기반 토픽 모델링을 통해 서비스를 1차적으로 분류하여 기존 카테고리와의 비교하였을 뿐, 개발자 및 소비자의 접근성 향상을 위해 서비스의 하위 카테고리를 새롭게 정의하고 그 특성을 분석하기 위한 연구는 매우 부족한 실정이다. 특히 카테고리의 정의뿐만 아니라 이를 계층적 관점에서 체계적으로 구성하는 것이 요구된다. 계층적 카테고리의 구축은 소비자가 자발적으로 원하는 서비스를 찾을 수 있도록 지원하는 동시에 개발자로 하여금 현재 서비스 트렌드의 전체적 구성을 이해하고 각 서비스 카테고리의 특성 및 현황을 명확히 파악함으로써 서비스 개발에 도움을 줄 수 있다.

이에 본 연구는 소비자 및 개발자 접근성을 높이기 위해 애플리케이션 서비스의 하위 카테고리를 제안하는 방법론을 제안하고자 한다. 본 연구에서는 각 애플리케이션 서비스 설명을 포함하는 텍스트 데이터를 추출하여, 애플리케이션 서비스의 특징을 구체화하는 데 활용한다. 이후 전처리된 텍스트 데이터를 기반으로 TF-IDF(Term Frequency-Inverse Document Frequency) 매트릭스를 구축하여, 이에 기반한 애플리케이션 서비스 유사도 매트릭스를 구축

하고, 각 서비스 간 유사도를 계산하여 네트워크 분석을 수행한다. 이후 네트워크 내 커뮤니티를 탐지하여 이를 클러스터로 활용하기 위해 네트워크 모듈성(modularity) 분석을 수행하고, 수행된 군집 결과에 따라 각 서비스의 특성을 해석하여 하위 카테고리를 재정의한다.

본 연구는 애플리케이션 플랫폼 내에서 사용자들이 원하는 애플리케이션 서비스를 보다 빠르고 편리하게 접근할 수 있는 접근성을 개선하기 위해 활용될 수 있다. 뿐만 아니라 하위 카테고리를 제공함으로써 소비자들에게 더 많은 애플리케이션 서비스를 노출시켜 개발자들의 치열한 경쟁을 완화시킬 수 있으며 더 나아가 모바일 애플리케이션 서비스 플랫폼 사업자들에게도 효율적인 서비스 제공 방안에 대해 도움이 될 것이다.

본 연구의 구성은 다음과 같다. 제2장에서는 본 연구와 관련된 선행 연구들을 살펴본다. 제3장에서는 연구의 전체적인 프로세스에 대해 서술하며, 제4장에서는 연구의 분석 결과에 대해 서술한다. 마지막으로 제5장에서는 연구의 결론과 의의, 한계점 및 추후 연구 방안에 대해 서술한다.

2. 이론적 배경

2.1 모바일 애플리케이션에 대한 기존 연구

모바일 애플리케이션 서비스와 관련된 연구들이 다양하게 수행되었으며 다음과 같이 크게 3가지로 나누어진다. 애플리케이션 서비스 플랫폼에 대한 연구, 애플리케이션 서비스 초점에 맞춰진 연구 그리고 서비스 플랫폼 데이터를 활용한 연구이다.

먼저 애플리케이션 서비스 플랫폼에 대한 연구로는 플랫폼에 대한 전반적인 이해와 비즈니스 관점에서 이루어진 연구들이 수행되었다 [14, 10]. Kim[14]의 연구에서는 애플리케이션 서비스 플랫폼에 대해 각각 소비자 관점, 개발자 관점 그리고 플랫폼 사업자 관점에서 시장성을 분석하였다. Huh et al.[10]의 연구에서는 주요 애플리케이션 서비스 플랫폼의 현황을 분석하고 비교하였고 개방형 애플리케이션 서비스 플랫폼을 위한 서비스 프레임워크를 제안하였다. 위와 같은 연구들에서는 하나의 관점이 아닌 다양한 관점에서 플랫폼의 시장성을 분석하여 애플리케이션 서비스 플랫폼의 선순환을 제시하였다.

모바일 애플리케이션 서비스에 초점이 맞춰진 연구로는 대표적으로 특정 애플리케이션 서비스 도메인의 동향을 분석하는 연구[20]와 애플리케이션 서비스의 유형 및 특성을 분석하는 연구[15, 30]가 있다. Liu et al.[20]의 연구에서는 애플리케이션 서비스 플랫폼의 상위 200개 애플리케이션에 대해 동향을 파악하고, 관련 기술, 아키텍처, 사용자 인터페이스 설계 문제를 분석하여 이를 개발자의 관점에서 검토하였다. 이를 통해 모바일 헬스 애플리케이션이 공공 의료 및 보건 교육을 개선점과 의료 센서와 애플리케이션의 상호운영성을 제안하였다. Yoon [30]의 연구에서는 교육(Education) 카테고리의 유아 교육용 유·무료 인기 애플리케이션 서비스의 유형을 분류하고 분석하여 추후 개발이 필요한 부분을 제안하였다. 이를 통해 애플리케이션 서비스의 사용자뿐만 아니라 교육 종사자, 개발자에게 현실적인 자료를 제공하였다는 의의가 있다. Kim et al.[15]의 연구에서는 모바일 애플리케이션 서비스를 탐색하여 텍스트 마이닝 기반 네트워크 분석을 통해 모바일 애플리케이션 서

비스 간의 관계를 분석하였다. Apple 앱 스토어 (App Store)에 존재하는 모든 카테고리를 사용하여 각 카테고리의 연관성을 네트워크로 시각화하여 각 부문의 모바일 애플리케이션 서비스의 영향력과 연관성이 다른 분야의 애플리케이션 서비스와는 다르다는 것을 보여주었다. 이 연구는 모바일 애플리케이션 서비스의 패턴과 구조를 보여주었으며 서비스 제공자, 개발자 등 모바일 생태계 참여자들에게 시사점을 제공하였다. 하지만 이와 같이 애플리케이션 서비스의 개발 현황, 트렌드, 영향력을 보여주는 단순한 서비스 특성의 분석을 넘어 새로운 서비스를 발굴하고 기획하는 것이 요구된다. 이러한 요구로부터 서비스의 주요 측면을 분석하여 새로운 서비스를 발굴하거나 더 나은 서비스를 제공할 수 있기 위한 요소들을 분석하는 연구들이 수행되었다[8, 19]. Geum et al. [8]의 연구에서는 애플리케이션 서비스 플랫폼의 데이터베

이스를 활용하여 제품 계층과 서비스 계층 사이의 종속적 관계를 식별하기 위해 연관 규칙 분석 (Association Rule Mining) 기반의 기술 로드맵을 구축하였다.

마지막으로 애플리케이션 서비스 플랫폼 내에서 얻을 수 있는 데이터를 기반으로 사용자 경험을 분석하는 연구가 있다[7]. 애플리케이션 서비스를 구매 및 설치한 사용자는 리뷰와 사용자 평점의 형태로 피드백을 제공할 수 있다. 사용자 리뷰는 해당 애플리케이션 서비스를 이용할 잠재적 사용자뿐만 아니라, 고객 피드백에 관심이 있는 개발자와 소프트웨어 회사들에게도 유용한 정보를 제공한다. 이를 바탕으로 사용자의 경험과 이의 의견, 기능 요청 및 버그 보고에 대한 정보를 분석하여 해결되지 못한 소프트웨어의 문제 및 더 나은 애플리케이션 서비스에 대한 방향을 제시한다. 이들 연구를 간략히 정리하면 <Table 1>과 같다.

<Table 1> Summary of Research Papers

Category	Author	Description
Application Service Platform	Kim[14]	Analyze the marketability of the platform from the perspective of customer, developer and platform operator
	Huh et al.[10]	Analyze and compare the status of service platforms and propose a service framework for open application service platforms
Application Services	Liu et al.[20]	Identify trends, analyze technology, architecture, user interface design issues and customer review from the developer's perspective
	Yoon[30]	Categorize and analyze types of paid, free and popular application services for early childhood education
	Kim et al.[15]	Analyze relationships between mobile application services through text mining-based network analysis
	Geum et al.[8]	Propose a technology roadmap based on associative rule analysis to identify the dependencies between product layers and service layers.
	Lee et al.[19]	Morphological analysis on Mobile application services for service idea development
Application User Experience Analysis	Genc-Nayebi and Abran[7]	Identify problems with the software through customer feedback data from the application service platform and provide directions for future services

2.2 데이터 기반 모바일 애플리케이션 유형화 연구

그러나 시간이 지날수록 모바일 애플리케이션 시장은 더 확장되었고, 다양하고 새로운 애플리케이션들도 많아지면서 애플리케이션 서비스를 유형화하고 세분화하는 필요성이 적극적으로 제기됨에 따라 이와 관련된 다양한 연구들이 수행되었다[29, 28, 16]. Qiang et al.[29]의 연구에서는 다양한 머신러닝 분류 기법을 이용하여 헬스케어 관련 iOS 애플리케이션 서비스와 안드로이드(Android) 애플리케이션 서비스의 유형을 세분화하였다. 위 논문에서는 헬스케어 관련 애플리케이션 서비스의 자동 분류 체계를 제시한다는 의의를 갖지만 헬스케어 애플리케이션 서비스의 범위를 이미 분류되어 있는 기존 카테고리인 건강 및 운동(health&fitness) 카테고리 and 의료(medical) 카테고리로 한정했다는 한계가 존재한다. Vakulenko et al.[28]의 연구에서는 모든 카테고리 내에 존재하는 영어 애플리케이션 서비스를 수집 대상으로 하여 LDA 기반 토픽 모델링을 통해 애플리케이션 서비스의 유형을 분류하였고, 도출된 토픽들을 활용하여 기존 카테고리를 보완점으로 제시하였다.

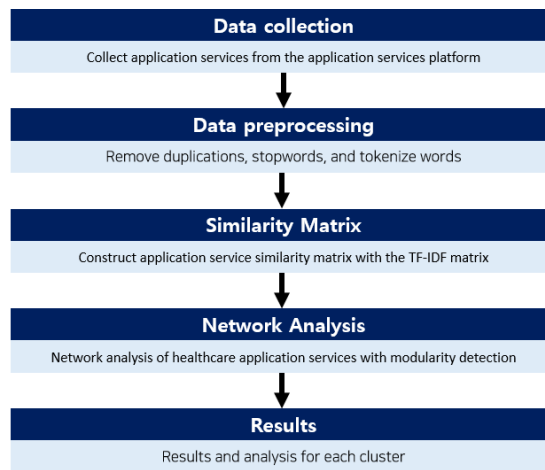
또한 LDA 알고리즘은 단편적으로 토픽모델을 생성하는 반면 본래 애플리케이션 서비스 카테고리는 계층적으로 구성되기 때문에 생성된 토픽모델을 바탕으로 전문가들이 계층적인 카테고리를 수동적으로 구성해야 한다는 한계가 존재한다. Kim and Park[16] 연구에서는 Apple 사의 iTunes App Store에 속해 있는 7,000개의 애플리케이션 서비스를 수집하여 Vakulenko et al.[28] 연구와 마찬가지로 LDA 기반 토픽모델링을 통해 애플리케이션 서비스

의 유형을 분류하였다.

이후 Random Forest 알고리즘을 통해 각 서비스 유형별로 사용자 평점에 주된 영향을 미치는 앱의 중요 요인을 파악하였다. 이는 앱 개발자의 주관적인 의사결정에 따른 기존 분류체계와 비교하였을 때보다 객관적이며 애플리케이션 서비스의 다기능, 다목적성화 추세에 알맞은 통합적인 분류체계를 제시하였다. 더 나아가 서비스 유형별로 사용자 평점에 영향을 미치는 중요 요인을 파악하여 이를 애플리케이션 설계 구성요소로 제시하였다는 의의가 있다[16]. 하지만 이러한 시도에도 불구하고, 토픽 모델 등을 통해 카테고리를 재정의하기 위한 대부분의 연구들이 단편적인 토픽모델을 생성한다는 한계가 존재한다. 애플리케이션 서비스들이 점점 더 다양해짐에 따라 이런 단편적 접근을 통해서만 체계적 분류를 시도하기 어려우며, 따라서 이를 체계적으로 분류하기 위한 계층적 분류 시스템이 필요한 것으로 보인다.

3. 연구 방법

본 연구의 프로세스는 <Figure 1>과 같다. 먼저 애플리케이션 서비스 플랫폼에서 분석 대상이 되는 애플리케이션 서비스를 수집한다. 이후 수집한 애플리케이션 서비스의 설명(description) 텍스트에 대해 불용어를 제거하고 단어로 토큰화하는 전처리를 진행한다. 데이터 전처리 이후, 전처리된 텍스트를 바탕으로 TF-IDF 매트릭스를 구축하고 이를 활용하여 문서 유사도 매트릭스를 생성한다. 이후 문서 유사도 매트릭스를 활용하여 네트워크 분석을 수행하고, 네트워크 내 커뮤니티를 탐지하여 클러스터로 활용하기



〈Figure 1〉 Overall Process of the Proposed Approach

위해 모듈성(modularity) 분석을 진행한다. 마지막으로 네트워크 모듈성 분석 결과를 바탕으로 애플리케이션 서비스 유형의 세분화에 대해 해석한다.

3.1 데이터 수집

먼저 애플리케이션 서비스 플랫폼에서 분석 대상이 되는 애플리케이션 서비스 도메인과 관련된 키워드를 검색 키워드로 설정한다. 각 키워드를 검색하여 노출된 애플리케이션 서비스의 이름(title), 카테고리(category), 설명(description), 리뷰 수(review), 사용자 평점(ratings), 설치 수(installation)을 포함하여 크롤링 프로그램을 통해 데이터를 수집한다. 이때 애플리케이션 서비스는 영어 애플리케이션 서비스로 제한한다.

3.2 데이터 전처리

수집한 텍스트 데이터를 정형화된 데이터로

변환하기 위해 전처리 과정을 진행한다. 키워드를 검색하여 나온 결과에 해당하는 애플리케이션 서비스를 바탕으로 수집하였기에 검색된 애플리케이션 서비스가 중복으로 존재할 가능성을 고려하여 먼저 중복제거를 수행한다. 이후, 수집한 애플리케이션 서비스의 데이터 중 설명(description)에 해당하는 텍스트 데이터에 한해 불용어(stopwords)를 제거하고, 각 문장을 단어 단위로 구분하는 토큰화(tokenization) 작업을 거친 후 마지막으로 BoW(Bag of Words) 인코딩 벡터화로 변환하는 작업을 수행한다.

3.3 문서 유사도 매트릭스 생성

문서 유사도란 문서들 간에 동일한 단어 또는 비슷한 단어가 얼마나 공통으로 많이 사용되었는지 나타내는 것을 의미한다. 문서 유사도를 측정하기 위해 비정형 데이터인 텍스트를 벡터화하여 정형화된 데이터로 변환해야 한다. 텍스트 데이터를 수치화하는 방법으로 BoW, BoW에

기반한 단어 표현 방법인 DTM(Document Term Matrix), TF-IDF, Word2Vec 등 여러 가지 방법이 존재하며 본 연구에서는 TF-IDF 매트릭스를 이용하여 문서 유사도 매트릭스를 구축한다. 먼저 TF-IDF란 TF(단어빈도, Term Frequency)와 IDF(역 문서 빈도, Inverse Document Frequency)의 곱으로 나타내며 어떤 단어가 특정 문서 내에서 중요한 정도를 가중치로 주는 방법을 말한다. TF는 어떤 단어가 특정 문서에 얼마나 많이 쓰였는지 빈도를 나타내는 값이며 IDF는 특정 단어가 문서에 등장한 횟수를 나타내는 값인 DF(문서 빈도, Document Frequency)의 역수를 의미한다. 즉 TF-IDF는 단순 빈도에 가중치를 부여하여 특정 단어가 문서 내에 얼마나 많은 비중을 차지하는지 나타낸다 [24]. TF와 IDF를 계산하는 방법은 다양하게 존재하며 대표적으로 TF-IDF는 식 (1)과 같이 표현되며 식(1)에 포함된 요소들은 각각 식 (2)와 식 (3)으로 나타난다.

$$tfidf_{i,j} = tf_{i,j} \times idf_i \quad (1)$$

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (2)$$

$$n_{i,j} = \text{단어 } t_i \text{가 문서 } d_j \text{에 출현한 횟수}$$

$$\sum_k n_{k,j} = \text{문서 } d_j \text{에서 모든 단어가 출현한 횟수}$$

$$idf_i = \log \left| \frac{D}{\{d_j | t_j \in d_j\}} \right| \quad (3)$$

D = 총 문서의 수

$\{d_j | t_j \in d_j\}$ = 단어 t_j 가 등장하는 문서의 수

이후 TF-IDF 매트릭스를 활용하여 문서 유사

도를 측정한다. 문서 유사도 측정 지표로는 Euclidean Distance, Cosine Similarity, Jaccard Coefficient, Pearson Correlation Coefficient, Averaged Kullback-Leibler Divergence 등 다양하게 존재하며[9] 본 연구에서는 코사인 유사도(cosine similarity) 측정 방법을 이용하여 문서 유사도를 계산한다. 코사인 유사도는 두 벡터 간의 코사인 각도를 이용하여 구할 수 있는 두 벡터의 유사도를 의미한다. 두 벡터의 방향이 완전히 동일한 경우에는 1의 값을 가지며, 90°의 각을 이루면 0, 180°로 반대의 방향을 가지면 -1의 값을 갖게 된다. 즉, 결국 코사인 유사도는 -1 이상 1 이하의 값을 가지며 값이 1에 가까울수록 유사도가 높다고 판단할 수 있다. 두 벡터 A, B에 대한 코사인 유사도는 식 (4)와 같다.

$$\cos(\theta) = \frac{A \cdot B}{|A||B|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}} \quad (4)$$

3.4 네트워크 분석

네트워크 분석은 개인과 집단들 간의 관계를 노드(node)와 엣지(edge)로 모형화하여, 그것의 구조나, 확산 및 진화 과정을 계량적으로 분석하는 방법이다. 본 연구에서는 문서 유사도 매트릭스를 활용하여 네트워크 분석을 수행하였다. 각 문서를 노드(node)로 두고 문서 간 유사도를 엣지(edge)로 하는 네트워크를 구축하였으며 이후 네트워크의 커뮤니티 구조를 탐지하기 위해 모듈성(modularity) 분석을 실시하였다 [21]. 본 연구에서는 Louvain method를 기반으로 하는 모듈성(modularity) 분석을 수행하였다 [3]. Louvain 알고리즘은 다음과 같은 두 단계로

구성된다. 먼저 첫 단계는 커뮤니티의 국소적 변화(local change)만을 허용하여 모듈성을 최적화하는 방식이며, 다음 단계는 도출된 커뮤니티를 다시 새로운 커뮤니티 네트워크로 만들기 위해 병합하는 방식이다. 이러한 과정은 더 이상의 모듈성 증가가 없을 때까지 반복해서 수행된다. 이를 도식화하면 식 (5)와 같다.

$$\Delta Q = \left[\frac{\sum_{in} + 2k_{i,in}}{2m} - \left(\frac{\sum_{tot} + k_i}{2m} \right)^2 \right] - \left[\frac{\sum_{in}}{2m} - \left(\frac{\sum_{tot}}{2m} \right)^2 - \left(\frac{k_i}{2m} \right)^2 \right] \quad (5)$$

ΔQ = a change in amount of modularity
 $\sum_{in}$ = the sum of the weights of the links inside community C
 $\sum_{tot}$ = the sum of the weights of the links incident to nodes in C
 k_i = the sum of the weights of the links incident to node i
 $k_{i,in}$ = the sum of the weights of the links from $BIGCIRC$ to nodes in C
 m = the sum of the weights of all the links in the network

4. 사례 연구

본 연구에서 제안한 프레임워크를 실제 적용하고 그 유용성을 밝히기 위해 사례연구를 수행하였다. 본 연구에서는 사례연구의 대상으로 헬스케어 관련 앱 서비스의 하위 카테고리를 재구성하고자 한다.

최근 모바일 헬스 시장의 성장으로 인해 헬스케어 관련 애플리케이션 서비스의 수가 급격히 증가하였다. 실제로 2018년 건강 및 피트니스 애플리케이션 서비스의 전 세계 소비자 지출은 2016년 대비 3배 성장하였고, 의료 애플리케이션

서비스의 2018년 전 세계 다운로드 수는 2016년 대비 15% 증가한 4억 건을 기록하였고 2020년 2월 기준으로 구글 플레이 스토어(Google Play)에 등록된 건강 및 운동(health&fitness) 애플리케이션 서비스와 의료(medical) 애플리케이션 서비스의 수는 약 14만 개에 달한다[1]. 최근 연구에 따르면, 스마트폰 소유자의 절반은 건강 정보를 얻기 위해 자신의 기기를 사용하고 있으며, 스마트폰 소유자의 5분의 1은 건강과 관련된 애플리케이션 서비스를 사용하고 있다[6].

이러한 폭발적 성장에도 불구하고 적절한 하위 카테고리가 제공되지 않고 있다. 관련 서비스의 수가 폭발적으로 증가하는 상황에서, 소비자가 본인의 요구 사항에 맞게 적절하게 서비스를 선택하기 위해서는 적절한 하위 카테고리가 제공되는 것이 필수적이다. 뿐만 아니라 서비스 제공자 관점에서도 서비스의 최신 동향을 파악함으로써 신규 서비스를 개발하는 데 중요한 단서를 얻을 수 있다.

이에 본 연구는 애플리케이션 서비스 플랫폼인 구글 플레이 스토어(Google Play)를 활용하여 헬스케어와 관련된 애플리케이션 서비스를 수집하고 유형을 분석하여 하위 카테고리를 제안하고자 한다.

4.1 데이터 수집

애플리케이션 서비스 데이터 수집은 python 3.6.4를 통해 크롤링 프로그램을 개발한 뒤 이를 활용하여 구글 플레이 스토어(Google Play)에서 데이터 수집을 진행하였다.

헬스케어 관련 서비스를 수집하기 위해서는 크게 두 가지 방법이 가능하다. 첫 번째 방법은 헬스케어 관련 카테고리를 파악하여 각 카테고리

리 내의 모든 서비스를 수집하는 것이고, 두 번째 방법은 헬스케어 관련 키워드를 통해 해당 키워드를 포함하는 서비스를 수집하는 것이다.

먼저 첫 번째 방식을 살펴보면, 구글 플레이 스토어(Google Play)에서 헬스케어와 관련된 카테고리인 건강 및 운동(health&fitness), 라이프스타일(lifestyle), 의료(medical)가 존재했다. 하지만 라이프스타일(lifestyle) 카테고리에는 헬스케어 관련 애플리케이션 서비스뿐만 아니라 패션 및 스타일(fashion) 애플리케이션 서비스와 같이 헬스케어와 관련이 없는 서비스도 포함하고 있음을 확인하였다. 따라서 헬스케어와 관련된 카테고리를 선형적으로 지정하여 데이터를 수집하지 않고, 헬스케어 관련 키워드를 연구자가 선정하고 이를 검색하여 애플리케이션 서비스를 수집하는 두 번째 방식을 택하였다.

본 연구에서는 <Table 2>와 같이 헬스케어 관련 키워드를 선정하였으며, 애플리케이션 서비스의 설명(Description)에 해당 키워드가 포함되는 서비스를 수집하였다. 각 키워드 검색 결과 당 250개의 애플리케이션 서비스가 노출되어 총 2,250개의 애플리케이션 서비스 데이터를 수집하였다.

<Table 2> Healthcare Keywords

Keywords
Health & Fitness
Healthcare
Medical
Diagnosis
Medical services
Health
Diet
Disease prevention
Treatment

4.2 데이터 전처리

데이터 수집 이후, 키워드를 입력하여 검색된 결과 내에 중복된 애플리케이션 서비스가 노출되었을 가능성을 고려하여 중복제거를 수행하였고, 수집된 애플리케이션 서비스의 카테고리 중 자동차(auto & vehicle) 카테고리에 해당하는 애플리케이션 서비스는 헬스케어와 관련이 없다고 판단하여 삭제하였다. 또한 애플리케이션 서비스의 이름 및 설명(Description)이 영어가 아닌 경우 분석 데이터에서 제외하였다. 그 결과 1,584개의 애플리케이션 서비스 데이터가 최종적으로 선정되었다. 본 연구에서 활용된 1,584개의 애플리케이션 서비스는 평균적으로 166.19개의 단어를 포함하고 있으며 (중앙값: 231.5단어), 1사분위수는 128.75단어, 3사분위수는 370.25 단어를 포함하고 있다. 일반적으로 특허나 제품 사양서는 특정한 기술 및 제품에 대해 매우 방대한 정보를 포함하고 있는 기술문서로, 길이가 매우 길고 많은 정보를 담고 있다. 본 연구에서 사용된 애플리케이션 서비스 기술문서는 특허나 제품사양서보다는 짧은 길이를 가지고 있지만, 일반적으로 서비스 트렌드 분석에 고객 리뷰, 트위터 및 페이스북 포스트와 같은 짧은 텍스트가 많이 사용되고 있는 것을 감안할 때, 서비스 트렌드를 분석하는 데 있어 무리가 없을 수준의 충분한 길이를 가지고 있다고 볼 수 있다.

애플리케이션 설명(Description)에 해당하는 텍스트 데이터는 정형화된 데이터로 변환하기 위해 먼저 불용어(stopwords)를 제거하고 단어 토큰화(tokenization) 작업을 수행하였다. 불용어(stopwords) 제거 및 단어 토큰화(tokenization) 작업은 python 3.6.4의 Scikit-learn 서브

패키지인 `feature_extraction.text` 패키지를 활용하여 수행하였다.

4.3 문서 유사도 매트릭스

다음으로 애플리케이션 서비스 유사도 매트릭스를 구축하기 위해 먼저 TF-IDF 매트릭스를 구축하였다. 이때 전처리 단계에서 활용했던 python 3.6.4의 Scikit-learn 서브 패키지인 `feature_extraction.text` 패키지를 동일하게 활용하였다. 그 결과 1,584×22,114 크기의 매트릭스가 생성되었으며 이 매트릭스를 활용하여 애플리케이션 서비스 유사도 매트릭스를 구축하였다. 애플리케이션 서비스 유사도 매트릭스는 코사인 유사도(cosine similarity)를 바탕으로 계산되었으며 python 3.6.4의 Scikit-learn 서브 패키지인 `metrics.pairwise` 패키지를 활용하였다. 애플리케이션 서비스 유사도 매트릭스는 <Table 3>과 같다.

4.4 기초적 분석

네트워크 분석을 수행하기에 앞서 구글 플레

이 스토어(Google Play)에서 수집한 1,584개의 애플리케이션 서비스들의 카테고리 양상을 살펴보기 위해 기초적 분석을 수행하였고 결과는 <Table 4>와 같다.

헬스케어 애플리케이션 수집 결과 건강 및 운동(health & fitness), 의료(medical) 카테고리의 비율의 합이 전체의 약 83%로 가장 높은 비율을 차지하였다. 그러나 헬스케어 애플리케이션 서비스는 건강 및 운동(health & fitness), 의료(medical) 카테고리 이외에도 25개 카테고리에 걸쳐 다양하게 퍼져있음을 알 수 있었다. 특히 교육(education) 카테고리는 세 번째로 높은 비율을 차지하고 있는 카테고리로 나타났다. 도서(books & reference) 카테고리나 비즈니스(business) 카테고리도 교육(education) 카테고리가 뒤를 이었다.

4.5 네트워크 분석

다음으로 애플리케이션 유사도 매트릭스를 활용하여 네트워크 분석을 수행하였다.

본 연구에서는 헬스케어 애플리케이션 서비스 유형을 세분화하여 하위 카테고리로 제공하

<Table 3> Application Service Similarity Matrix

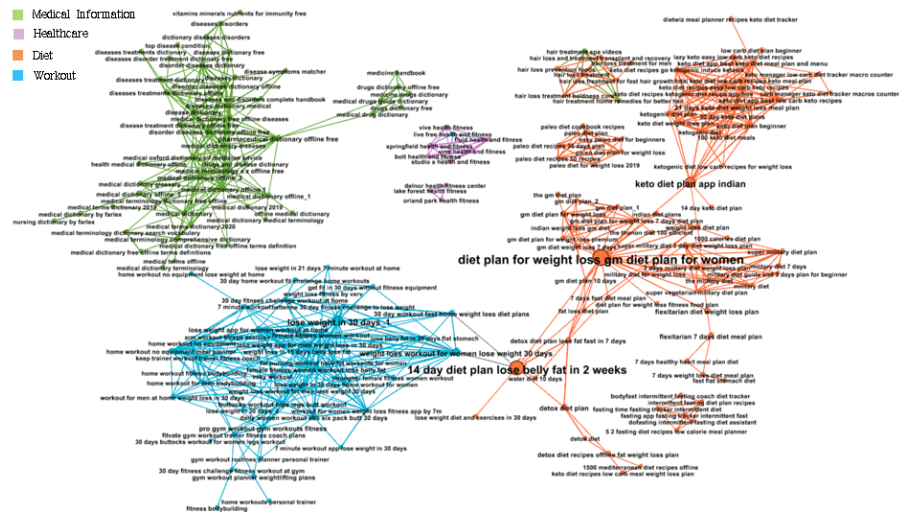
	D_1	D_2	D_3	D_4	...	D_{1581}	D_{1582}	D_{1583}	D_{1584}
D_1	1.000	0.211	0.102	0.128	...	0.033	0.017	0.019	0.043
D_2	0.211	1.000	0.213	0.353	...	0.017	0.002	0.031	0.040
D_3	0.102	0.213	1.000	0.126	...	0.008	0.009	0.032	0.020
D_4	0.128	0.353	0.126	1.000	...	0.004	0.001	0.011	0.010
...	...	...	...	...	...	...	...	...	...
D_{1581}	0.033	0.017	0.008	0.004	...	1.000	0.008	0.051	0.013
D_{1582}	0.017	0.002	0.009	0.001	...	0.008	1.000	0.029	0.020
D_{1583}	0.020	0.031	0.032	0.011	...	0.051	0.029	1.000	0.091
D_{1584}	0.043	0.040	0.020	0.010	...	0.013	0.020	0.091	1.000

〈Table 4〉 Summary of Healthcare Application Services Category

Category	# of samples	%	Category	# of samples	%
Beauty	5	0.3	Lifestyle	18	1.1
Books & Reference	32	2	Medical	527	33.3
Business	31	2	News & Magazines	7	0.4
Casual	2	0.1	Parenting	1	0.1
Comics	1	0.1	Productivity	11	0.7
Communication	6	0.4	Shopping	1	0.1
Education	88	5.6	Simulation	7	0.4
Educational	8	0.5	Social	2	0.1
Entertainment	3	0.2	Sports	1	0.1
Events	1	0.1	Tools	9	0.6
Finance	4	0.3	Travel & Local	2	0.1
Food & Drink	17	1.1	Trivia	2	0.1
Health & Fitness	795	50.2	Weather	1	0.1
House & Home	2	0.1			
Total			1,584		

기 위해 네트워크 내 커뮤니티를 탐지하는 모듈성(modularity) 분석을 수행하였다. 이를 위해 본 연구에서는 Gephi 0.9.2 프로그램을 사용하였다. 각 네트워크 노드(node)는 헬스케어 애플리케이션 서비스를 의미하여, 각 노드(node)와 연결되어있는 엣지(edge)는 애플리케이션 서비스 유사도를 나타낸다. 모듈성 분석은 Gephi를 통해 수행되었으며, Louvain 알고리즘에 기반하여 수행하였다. 그 결과 1,584개의 애플리케이션 서비스에 대하여 총 4개의 커뮤니티를 탐지하였다. 먼저 전체 네트워크의 양상을 알아보기 위해 이를 도식화하였으며 가시성을 위해 cut-off 값을 0.5로 설정하였다. 즉, cosine similarity 값이 0.5 이상인 것만 네트워크에 표현하였다. cut-off 값이란 해당 기준값보다 큰 엣지(edge)만 네트워크에 나타나도록 설정하는 값을 의미한다. 또한 네트워크에서 노드(node)와 텍스트(text)의 크기는 매개중심성(betweenness centrality) 값을 적용하여 나타내었다. 매개중심성은 일반적인

로 노드들 간 최단 경로로 계산된다. 이는 직접 연결되지 않은 노드 간 관계를 매개하는 수준을 나타내고 있으며, 본 연구에서는 각 서비스가 서로 다른 애플리케이션 서비스들을 매개하는 수준을 나타낸다고 볼 수 있다. 본 연구의 목적이 애플리케이션 서비스 네트워크의 구조를 바탕으로 하위 카테고리를 설계하는 것이기 때문에 각 클러스터 내부의 서비스 간 관련 수준을 극대화하여 보여주기 위해 매개중심성을 택하였다. 그 결과 <Figure 2>와 같은 네트워크 분석 결과를 얻었으며, 총 4개의 커뮤니티 구조를 탐지하였다. 유사도를 바탕으로 탐지된 4개의 서비스 클러스터에 대한 정보는 <Table 5>와 같다. 또한 <Figure 2> 전체 네트워크는 모듈성(modularity) 분석 수행 이후 cut-off 값을 0.5로 설정한 후 도식화하였기에 엣지(edge) 값이 0.5 이하는 표현되지 않았다. 이와 같은 이유로 <Figure 2>에서는 모든 애플리케이션 서비스 노드(node)들이 표현되지는 않았다.



〈Figure 2〉 Healthcare Application Network

먼저 가장 높은 비율을 차지하는 클러스터는 1번 클러스터(Medical Information)로 전체에서 약 40.59%를 차지하였고 그 다음으로 2번 클러스터(Healthcare)는 약 26.58%의 비율을 차지하였다. 3번 클러스터(Diet)와 4번 클러스터(Workout)가 그 뒤를 이었으며 각각 17.42%, 15.4%의 비율을 차지하였다. 전체 네트워크의 클러스터 유형을 파악하여 각 클러스터의 이름을 <Table 5>에 표기한 대로 정의하였다. 지금까지는 헬스케어 애플리케이션 서비스 유형의 전체적인 양상을 파악하였으며 다음으로 각 클러스터에 어떤 애플리케이션 서비스가 주로 존재하는지 클러스터별로 세세하게 살펴보고자 한다.

〈Table 5〉 Cluster Information

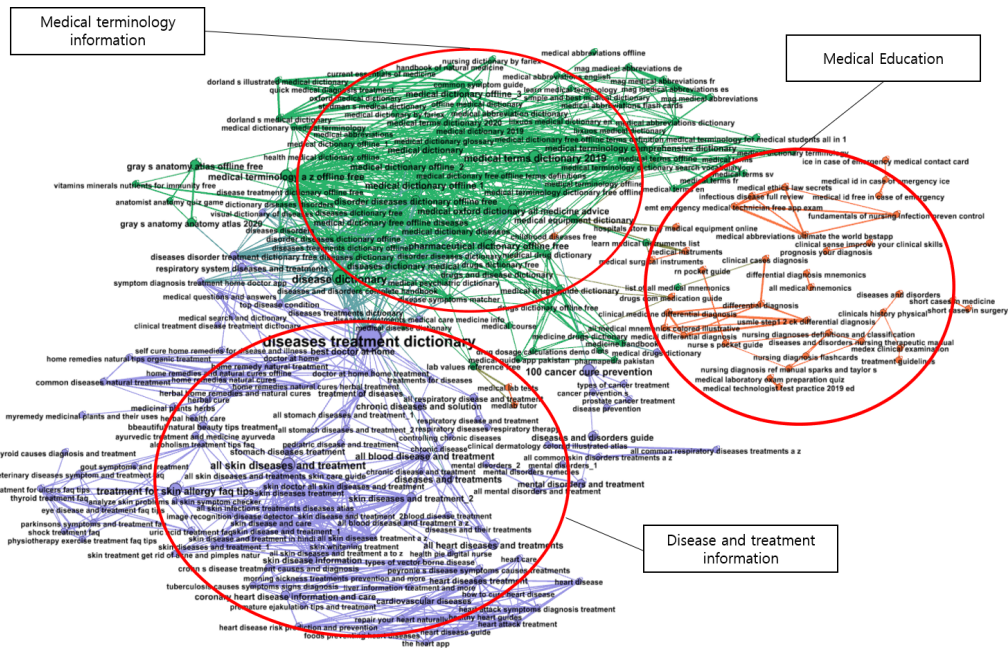
Clusters	# of application service	%
1 - Medical Information	643	40.59
2 - Healthcare	421	26.58
3 - Diet	276	17.42
4 - Workout	244	15.4

4.6 클러스터 분석 및 비교

각 클러스터마다 존재하는 애플리케이션 서비스의 유형을 더 자세하게 살펴보기 위해 2차 모듈성(modularity) 분석을 수행하였다. 이후 네트워크 도식화 과정에서 각 클러스터에 속해 있는 노드(node)와 텍스트(text)의 크기는 매개 중심성(betweenness centrality) 값을 적용하여 나타내었다.

4.6.1 의학 정보(Medical Information) 클러스터

먼저 전체 헬스케어 애플리케이션 서비스 중 가장 높은 비율(40.59%)을 차지하는 1번 클러스터를 살펴보자. 모듈성(modularity) 분석 결과 총 3개의 커뮤니티 구조를 탐지하였으며 네트워크 도식화 시 cut-off 값은 0.3으로 설정하였으며 결과는 <Figure 3>과 같다. 의학 정보 클러스터에는 의학적인 정보를 제공하는 애플리케이션 서비스들로 이루어져 있다. 첫 번째로 각종 질병 및 치료법에 대한 정보를 제공하



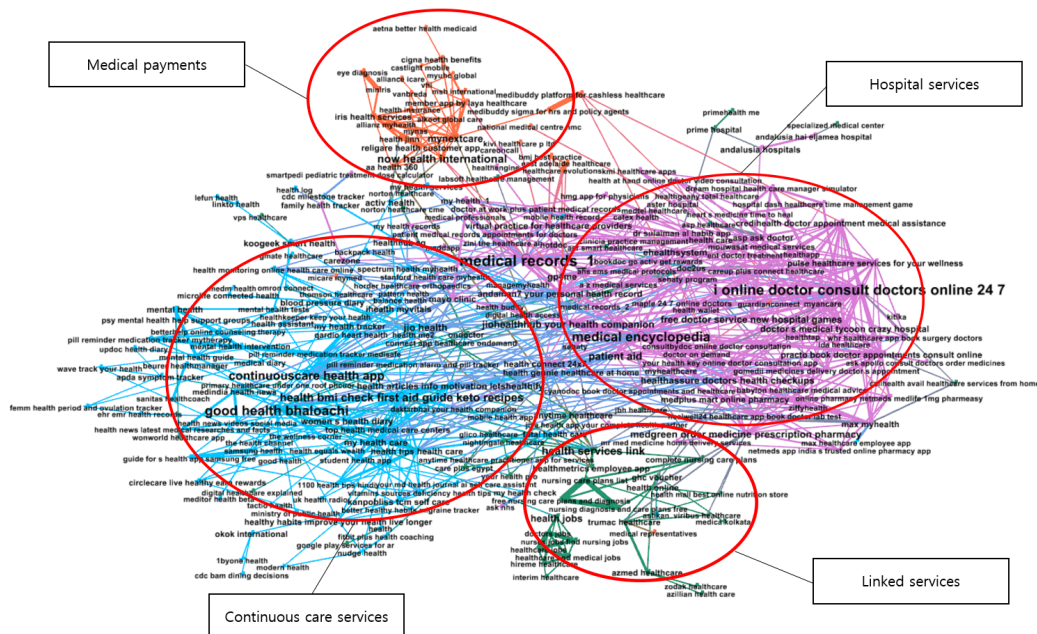
〈Figure 3〉 Medical Information Network(cutoff = 0.3)

는 서비스들이 존재했다. 해당 클러스터는 피부질환, 정신질환, 호흡기질환, 심장질환, 간 질환 등 다양한 질환에 대한 정보와 치료법에 대한 서비스를 포함하고 있다. 두 번째로 의학 및 약학에 관련된 수많은 용어들을 사전 형태로 제공하는 서비스가 존재했다. 또한 용어들의 축약 사전 존재하였으며 여러 가지 언어로 제공되는 서비스도 존재했다. 마지막 클러스터는 주로 의학 교육에 도움이 될 만한 서비스들이 존재했다. 방대한 의학 지식을 쉽게 습득할 수 있는 연상암기법이나 의학 퀴즈, 시험뿐만 아니라 임상 사례에 대한 정보를 제공하는 애플리케이션이 존재했다.

4.6.2 헬스케어(Healthcare) 클러스터

다음으로 헬스케어 클러스터를 살펴보자. 해당 클러스터는 전체 애플리케이션 서비스 중

두 번째로 높은 비율(26.58%)을 차지한다. 모듈성(modularity) 분석 결과 총 4개의 커뮤니티 구조를 탐지하였으며 cut-off 값은 0.15로 설정하였다. 헬스케어 클러스터의 결과는 〈Figure 4〉와 같다. 헬스케어 클러스터에는 주로 통합적인 건강관리와 관련된 서비스들이 존재했다. 첫 번째로 지속적인 건강관리를 할 수 있도록 도와주는 측정 및 추적 기능을 가진 서비스들이 존재했다. 사용자의 일상 활동과 습관을 기록하고 분석하는 것이다. 예를 들어, 다양한 활동에 대한 칼로리를 측정하거나 심박 수, 체질량지수(BMI), 수면 및 스트레스 등 건강에 기여하는 일상생활의 다양한 측면을 측정하고 추적하여 개인의 건강관리에 도움을 주고자 하는 서비스들이 존재했다. 이러한 유형의 서비스들은 건강 및 웰빙(well-being) 목적으로만 사용되며 질병 또는 기타 질환의 진단이나 치료, 완



〈Figure 4〉 Healthcare Network(cutoff = 0.15)

화 및 예방으로 사용되지는 않는 서비스들이다. 두 번째로 서비스 및 인적자원을 연결시켜주는 애플리케이션 서비스가 존재했다. 예를 들어 의사 및 간호사 인력이 필요한 곳으로 일자리를 연결해주거나 헬스케어 서비스가 필요한 곳에 서비스를 제공하고 관리할 수 있는 애플리케이션이 존재했다. 세 번째로 환자-의사 간 진료와 관련된 서비스들이 존재했다. 예를 들어, 온라인으로 진료 예약을 하거나 의료 상담을 할 수 있는 서비스, 개인의 의료 기록을 기록해둘 수 있는 서비스, 처방된 약을 사고 배달받을 수 있는 서비스가 존재한다. 위와 같은 서비스들은 대체로 병원 내원 시 필요한 절차들 중 일부를 모바일 서비스를 통해 해결할 수 있어 시간을 절약할 수 있다는 장점이 있다. 또한, 불가피하게 병원까지 갈 수 없는 사람들에게 전화 상담을 통한 원격 진료 서비스를 이용

함으로써 쉽고 편하게 건강관리를 할 수 있다. 마지막으로 건강보험 및 병원비 청구서를 모바일을 통해 지불할 수 있는 서비스가 존재했다.

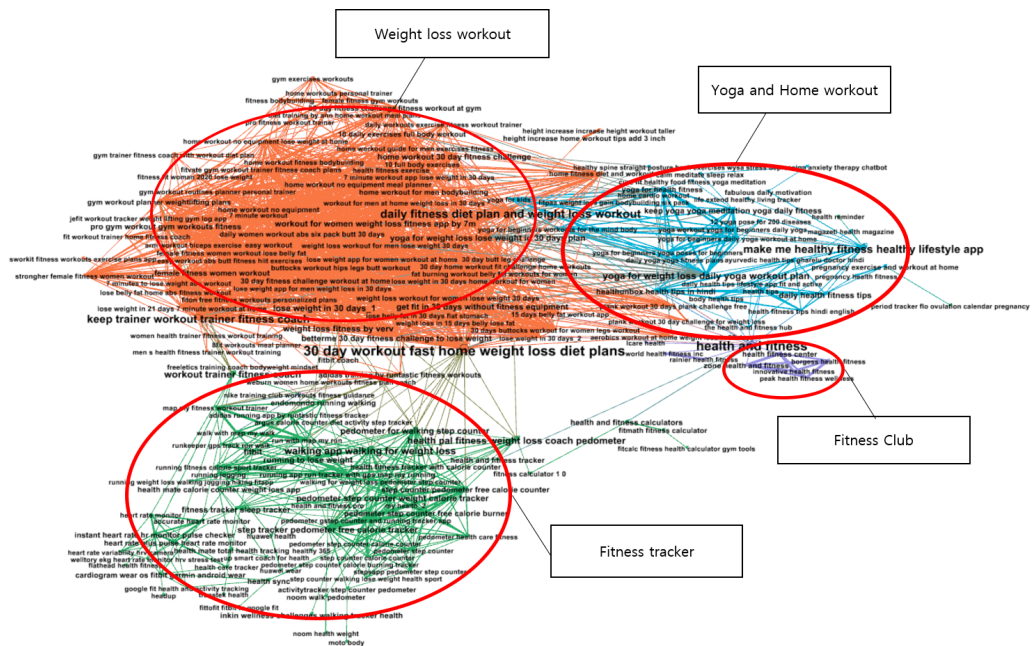
4.6.3 다이어트(Diet) 클러스터

다음으로 다이어트(Diet) 클러스터를 살펴보자. 전체 애플리케이션 서비스 중 세 번째로 높은 비율(17.42%)을 차지하고 있다. 모듈성(modularity) 분석 결과 총 4개의 커뮤니티 구조를 탐지하였으며 cut-off 값은 0.3으로 설정하였다. 다이어트 클러스터는 다이어트에 도움이 될 수 있는 서비스들을 다양하게 포함하였으며 결과는 〈Figure 5〉와 같다. 첫 번째 다이어트에 도움이 될 수 있는 측정 관련 애플리케이션 서비스가 존재했다. 예를 들어 체중을 기록하고, 섭취한 음식의 칼로리를 계산하여 목표



(ketogenic) 다이어트와 관련된 서비스가 존재했다. 케토제닉 다이어트 플랜, 케토제닉 다이어트 요리법과 관련된 서비스들이 주를 이루고 있었다. 다이어트 방법 중 하나인 케토제닉 다이어트와 관련된 서비스가 하나의 클러스터를 형성할 만큼 큰 비중을 차지하는 것을 알 수 있었다.

마지막으로 운동 클러스터를 살펴보자. 모듈성(modularity) 분석 결과 총 4개의 커뮤니티 구조를 탐지하였으며 cut-off 값은 0.2로 설정하였다. 분석 결과는 <Figure 6>과 같다. 첫 번째로 운동과 관련된 추적 애플리케이션 서비스가 존재했다. 해당 서비스들은 주로 운동 중 걸음 수를 기록하거나 심박 수를 측정하여 소모된 칼로리를 알려주는 기능을 포함하고 있다. 두 번째 클러스터는 체중 감량을 위한 운동들



〈Figure 6〉 Workout Network(cutoff = 0.2)

이 다양하게 존재했다. 여성을 위한 체중 감량 운동 애플리케이션, 복근 운동을 위한 애플리케이션, 다양한 운동법을 제공하는 애플리케이션 등을 포함한다. 세 번째로는 요가운동과 홈 트레이닝과 관련된 애플리케이션 서비스들이 존재했다. 이와 더불어 임산부를 위한 요가 및 홈 트레이닝 애플리케이션 서비스도 존재했다. 마지막으로 살펴볼 클러스터는 멤버십 위주로 운영되는 애플리케이션 서비스가 존재했다. 이에 해당되는 서비스들은 모바일 애플리케이션을 통해 멤버십 회원들이 지역 헬스장을 이용하거나 다양한 운동 프로그램을 신청하고 수강할 수 있는 서비스를 제공하고 있다. 이를 모두 종합하면 헬스케어 내 서비스에 대한 유형화 결과는 <Table 6>과 같다.

다음으로 각 카테고리의 특성이 차이가 있는지를 파악하기 위해 리뷰(review) 수, 설치

(installation) 수, 사용자 평점(ratings)의 평균을 클러스터 별로 비교하였다. 리뷰(review) 수와 설치(installation) 수의 경우 데이터값의 편차가 큰 점을 고려하여 로그(log)를 적용하여 계산하였으며 분산분석(ANOVA)을 통하여 클러스터별 차이가 유의미하다는 것을 검증하였으며 이를 <Table 7>로 나타냈다.

이후 클러스터별 리뷰(review) 수의 평균, 설치(installation) 수의 평균, 사용자 평점(ratings)의 평균에 대한 그래프로 도식화하였으며 <Figure 7>과 같다. 리뷰(review)의 경우 운동(workout) 클러스터의 리뷰(review) 수 평균이 제일 높은 것으로 나타났다. 수집된 헬스케어 애플리케이션 서비스 중 운동(workout)과 관련된 애플리케이션 서비스가 차지하는 비율은 <Table 5>에서 알 수 있듯이 약 15.4%로 가장 비율이 낮다. 이처럼 운동(workout) 애플

리케이션 서비스의 비율이 가장 낮음에도 불구하고 리뷰(review) 수의 평균이 다른 클러스터보다 높은 이유는 서비스 이용자들이 피드백을 제공하며 적극적으로 서비스를 이용하기 때문이라고 생각할 수 있으며 운동(workout) 클러스터에 해당하는 애플리케이션 서비스 설치(installation) 수에 대한 높은 평균이 이를 뒷받침하고 있음을 <Figure 7>를 통해 알 수 있다.

다이어트 클러스터의 경우 약 17.42%로 전체 애플리케이션 서비스 중 낮은 비율을 차지하는 편에 속하지만 이에 비해 설치(installation) 수와 사용자 평점(ratings)의 평균은 매우 높은 편인 것을 알 수 있었다. 운동 클러스터와 마찬가지로 제공되고 있는 서비스 수의 비례 이용률이 높고 피드백이 활발하다는 것을 알 수 있다. 또한 전체 서비스 중 약 40.59%로 가장 높은

<Table 6> Healthcare Service Categorization Results

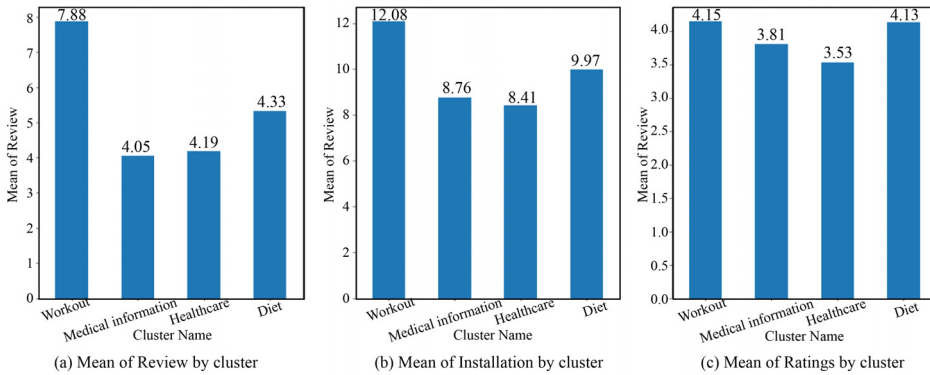
Category	Sub category	Main keyword	%
Medical Information	sub1 - Disease and Treatment information	disease, treatment, symptoms, heart, remedies, blood, diagnosis, cancer, diabetes, system, pain, body, cure, disorder, treatment	45.87
	sub2 - Medical Education	clinical, information, contents, emergency, patient, free, students, mnemonics, nursing, professionals, reference	34.95
	sub3 - Medical Terminology information	dictionary, app, search, offline, information, words, abbreviation, records, drug, disease	19.19
Healthcare	sub1 - Continuous care services	health, track, care, help, blood, medical, tips, time, healthcare, access, pressure, free, weight, life, family, mental, keep	33.09
	sub2 - Linked services	healthcare, service, quality, access, plan, administration, products, provider, patients, pet, support, hospital, time, get	27.9
	sub3 - Hospital services	doctor, app, online, healthcare, hospital, access, medicine, record, book, appointment, available, lab, consult, order	23.95
	sub4 - Medical payments	app, view, access, mobile, policy, feature, insurance, claim, location, check, contact, benefit, account, card	15.06
Diet	sub1 - Diet tracker	weight, diet, app, food, health, loss, plan, water, meal, body, calorie, lose, track, nutrition, tracker, gain	38.06
	sub2 - Diet Plan	diet, plan, day, fat, meal, lose, paleo, recipes, low, detox, body, high, protein, menu, follow, feature	38.06
	sub3 - Diet tips	tips, diet, natural, hair, symptoms, help, information, life, causes, remedies, growth	13.43
	sub4 - Keto diet	diet, keto, recipes, carb, low, ketogenic, weight, plan, fat, easy, body, free, chicken, ketosis, traker	10.45
Workout	sub1 - Fitness tracker	health, tracker, fitness, weight, rate, track, running, data, walking, get, step	36.78
	sub2 - Weight loss workout	workout, fitness, weight, app, fat, home, exercise, body, lose, training, belly	35.54
	sub3 - Yoga and Home workout	yoga, health, tips, fitness, weight, workout, daily, loss, improve, meditation, free, eye, sleep, women	17.36
	sub4 - Fitness Club	fitness, workout, goals, class, track, gym, class, activities, personal, view, challenges, wellness	10.33

비율을 차지하는 의학 정보(medical information) 클러스터는 애플리케이션 서비스 수가 많은 것에 비해 리뷰(review) 수 평균, 설치(installation) 수 평균 그리고 사용자 평점(ratings) 평균에 대해선 운동이나 다이어트 클러스터에 비해 낮은 편이었다. 이는 제공되는 서비스는 많지만 실질적으로 서비스를 이용하는 이용자 수는 적으며 만족도도 낮다고 해석

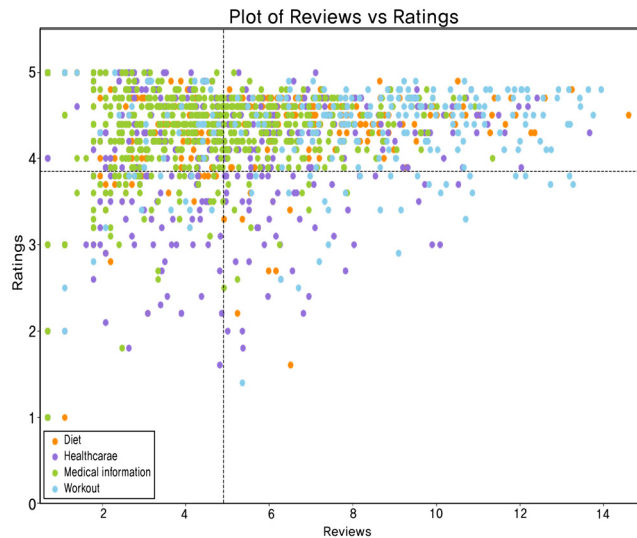
될 수 있다.

〈Table 7〉 ANOVA Results for Variables

Variables (mean)	F-statistic	p-value
Review	117.981	6.06e-69
Ratings	16.825	9.25e-11
Installation	109.904	1.247e-64



〈Figure 7〉 Mean Review, Installation, Ratings by Cluster



〈Figure 8〉 Distribution of Number of Reviews and user Ratings in Healthcare Application Services

마지막으로 리뷰(review) 수와 사용자 평점(ratings)에 따른 애플리케이션 서비스의 분포를 알아보기 위한 분석을 진행하였으며 결과는 <Figure 8>과 같다. 리뷰(review) 수의 평균(4.9)과 사용자 평점(ratings)의 평균(3.85)을 축으로 사분면을 나누어 헬스케어 애플리케이션 서비스의 분포를 나타내었다. 오른쪽 위 방향을 시작으로 1사분면, 2사분면, 3사분면, 4사분면이라고 정의하였으며 각 사분면에 대한 애플리케이션 분포는 <Table 8>과 같다.

먼저 의학 정보를 제공하는 애플리케이션 서비스의 약 78%가 1, 2사분면에 걸쳐 주로 분포하고 있으며 특히 2사분면에 가장 많이 존재했다. 리뷰(review) 수는 낮지만 그에 비해 사용자 평점(ratings)은 평균 이상인 서비스들이 많이 존재함을 의미했다. 또한 헬스케어 클러스터에 포함되는 애플리케이션 서비스들은 2, 3사분면에 걸쳐 약 62%의 서비스가 존재함을 알 수 있었다. 그 중에서도 2 사분면에 제일 많이 존재하였는데 의학 정보 애플리케이션 서비스와 마찬가지로 리뷰(review) 수는 낮지만 그에 비해 사용자 평점(ratings)은 평균 이상인 서비스들이 존재한다고 해석할

수 있다. 다이어트 클러스터에 해당하는 애플리케이션의 경우, 1사분면에 약 49%의 서비스가 분포하고 있다. 이는 현대인의 건강관리 중 체중 감량을 빼놓을 수 없을 만큼 사람들의 관심이 높은 것을 고려하였을 때 다이어트와 관련된 애플리케이션의 서비스 대부분은 리뷰(review)와 사용자 평점(ratings)이 모두 높다는 것을 알 수 있다. 마지막으로 운동 관련 애플리케이션의 경우 다이어트 애플리케이션 서비스와 마찬가지로 1사분면에 약 67%의 서비스가 분포하고 있으며 관련 서비스가 활발하게 이용되고 있음을 알 수 있었다. 본 연구에서 사용한 리뷰-평점 간 매트릭스 도식화의 경우 전체 평균값을 기준으로 사분면을 나누어 분포를 파악하였다. 이는 전체 헬스케어라는 분야 안에서 어떠한 유형의 서비스 사용자들이 활발한 피드백을 주고 있는지를 전반적으로 확인하기 위함이다. 하지만 각 유형별로 전체 평균값이 많이 다를 수 있기 때문에, 각 유형 내의 개별 서비스 특성을 파악하기 위해서는 각 유형의 리뷰, 평점 평균값을 사용해 유형별로 별도 매트릭스를 그려서 파악하는 것이 도움이 될 수 있다.

<Table 8> Distribution of Application Services for Each Quadrant

Cluster	Quadrant1	Quadrant2	Quadrant3	Quadrant4	Total
Medical information	222	275	135	11	643
	35%	43%	20%	2%	100%
Healthcare	112	138	123	48	421
	27%	33%	29%	11%	100%
Diet	134	93	39	10	276
	49%	34%	14%	3%	100%
Workout	163	30	22	29	244
	67%	12%	9%	12%	100%
Total	631	536	319	98	1584

4.7 시사점

본 사례연구에서는 헬스케어 서비스에 대한 네트워크 클러스터링을 통해 2단계의 계층적 카테고리를 제시하고, 각 하위 카테고리의 특성을 분석하였다.

사례연구 결과 전체 헬스케어 애플리케이션 서비스는 총 4개의 클러스터를 형성하였으며, 각 클러스터는 각각 의학 정보 클러스터, 헬스케어 클러스터, 다이어트 클러스터, 운동 클러스터로 나타났다. 이는 전체 헬스케어 관련 서비스의 구조가 크게 네 가지로 구성될 수 있기 때문에 앱 서비스 제공자는 1차적인 서비스 성격을 의학, 헬스케어, 다이어트, 운동으로 구분하는 것이 바람직하다는 것을 의미한다.

구체적으로 의학 정보 클러스터에는 3개의 하위 클러스터가 존재하였으며 질환 및 치료법에 대한 정보, 의학 용어에 대한 정보 그리고 의학 교육과 관련된 서비스들을 포함하고 있다. 헬스케어 클러스터의 경우 4개의 하위 클러스터가 존재하였으며 주로 통합적으로 건강관리를 할 수 있는 서비스와 헬스케어 서비스 연계, 의료 진단 관련 서비스 등 개인 건강 관리에 도움이 될 수 있는 확장된 서비스 형태를 포함하고 있음을 확인할 수 있었다. 다이어트 클러스터는 4개의 하위 클러스터를 형성하였으며 다이어트 계획 서비스와 체중 감량 시 도움이 될 수 있는 측정, 추적기능의 서비스를 포함한다. 마지막으로 운동 클러스터는 4개의 하위 클러스터를 형성하였으며 주로 체중 감량을 위한 운동법과 심박 수, 걸음 수 등을 측정하여 운동 중 소모되는 칼로리를 계산하여 제공하는 서비스가 존재한다.

본 연구의 결과를 토대로 서비스 제공자 및

서비스 개발자는 다음과 같이 헬스케어의 하위 카테고리를 계층적으로 체계화하고, 각 카테고리의 특성을 파악하여 적합한 서비스를 개발할 수 있을 것이다.

- 의학정보 - 질병 및 치료, 의학 교육, 의학 용어사전
- 헬스케어 - 일상적 헬스케어, 의료서비스 연계, 진료 및 예약, 지불 및 청구
- 다이어트- 다이어트 측정 및 추적, 다이어트 계획, 다이어트 팁, 케토제닉 다이어트
- 운동 - 운동 추적, 체중감량 도우미, 요가 및 홈트레이닝, 피트니스 클럽 서비스

특히 각 카테고리를 재정의함과 더불어 각 카테고리의 특성을 명확히 분석함으로써, 서비스 개발에 있어 중요한 단서를 얻을 수 있다. 예를 들어 다이어트 관련 서비스와 운동 관련 서비스가 차지하는 비율은 전체에서 낮은 편에 속하지만 리뷰(review) 수의 상관없이 평균 이상의 사용자 평점(ratings)을 받는 것으로 나타나 실질적으로 이용자 사이에서 활발하게 이용되고 있는 것으로 분석되었다. 이는 서비스 개발자 입장에서 다이어트 및 운동에 관련된 애플리케이션 서비스의 개발은 타 서비스에 비해 비교적 안정적이고 활발한 운영이 가능한 것으로 해석될 수 있어, 이와 관련한 차별적 서비스를 개발한다면 좋은 서비스 대안이 될 수 있을 것으로 생각할 수 있다.

소비자 관점에서는 본 계층적 체계화를 통해 현재 적절한 분류체계가 없이 제공되고 있는 애플리케이션 서비스를 훨씬 쉽고 체계적으로 탐색할 수 있어 시간과 노력을 절약할 수 있을 것으로 기대된다. 뿐만 아니라 이렇게 서비스 카테고리가 체계적으로 제공된다면, 소비자 역

시 새로운 서비스 유형이 등장하는 경우 새로운 하위 카테고리를 통해 새로운 서비스를 인지하고 쉽게 접근할 수 있을 것으로 기대된다. 예를 들어 다이어트 카테고리가 그 동안 세 가지로 제공되다가 최근 케토제닉 다이어트라는 새로운 하위 카테고리가 추가되는 경우, 소비자 입장에서 새로운 서비스의 유형을 쉽게 인지하고 활용함으로써 서비스의 접근성을 높일 수 있을 것으로 기대할 수 있다.

5. 결 론

본 연구는 소비자 및 개발자 접근성을 높이기 위해 애플리케이션 서비스의 하위 카테고리의 필요성을 제기하고 방법론을 제안하여 이를 사례연구에 적용하였다. 먼저 수집된 헬스케어 애플리케이션 서비스를 바탕으로 네트워크 분석을 수행하였다. 헬스케어와 관련된 서비스의 세부 유형을 파악하기 위해 모듈성(modularity) 분석을 수행하였고 이를 네트워크로 도식화하였다. 이후 탐지된 각 클러스터마다 존재하는 서비스의 유형을 더 자세히 살펴보기 위해 2차 모듈성(modularity) 분석을 수행하였다. 각 클러스터가 유의미한지 알아보기 위해 리뷰(reivew) 수, 설치(installation) 수, 사용자 평점(ratings) 데이터를 활용하여 분산분석(ANOVA)을 수행하였고 클러스터별 리뷰(reivew) 수, 설치(installation) 수, 사용자 평점(ratings)의 평균을 비교하는 기초분석을 수행하였다. 마지막으로 리뷰(review) 수와 사용자 평점(ratings)에 따른 애플리케이션 서비스 분포도를 도식화하여 각 클러스터마다 지니고 있는 특징을 파악하였다.

본 연구는 애플리케이션 서비스 플랫폼 내에 존재하는 서비스 유형 세분화에 대한 필요성을 인식하고 플랫폼 내 하위 카테고리를 제공하고 있지 않은 헬스케어 분야 서비스에 적용하여 분석하였다는 점에서 연구의 의의가 있다. 그럼에도 불구하고 본 연구는 다음과 같은 한계점을 가지고 있다. 먼저 본 연구는 구글 플레이 스토어(Google Play) 페이지에서 노출되는 애플리케이션 서비스의 최대 수가 고정되어 있어 헬스케어와 관련된 모든 애플리케이션 서비스를 수집하지 못했다는 한계가 있다. 헬스케어 관련 카테고리를 지정하여 수집 범위를 한정하는 경우 노출되는 서비스의 수는 본 연구에서 키워드 검색을 통해 페이지당 노출되는 최대 250개의 수보다 더 적다. 하지만 그럼에도 불구하고 이는 헬스케어 관련 서비스 모두를 포함한다고 보기 힘들며 여전히 페이지당 노출되는 상위 250개로 한정된다는 한계가 존재한다. 둘째, 본 연구는 healthcare, medical, diagnosis, diet, disease, treatment 등 헬스케어 서비스를 대표하는 몇 가지 키워드를 바탕으로 서비스를 수집하였다. 하지만 해당 키워드의 대표성 수준을 고려할 때, 유사어 등을 사용하여 보다 확장된 검색 쿼리를 활용하면 더 많은 수의 데이터를 수집할 수 있을 것으로 기대되며, 보다 범용적 측면에서 접근할 수 있는 기반이 될 수 있을 것으로 기대된다. 셋째, 네트워크 분석의 하나인 모듈성(modularity) 분석을 통해 클러스터를 형성하였지만 이를 평가할 수 있는 지표가 부족하다는 한계가 있다. 추후 연구에서는 클러스터의 결과를 정성적, 정량적 관점에서 평가할 수 있는 지표를 고려하여 더 논리적인 결과를 도출할 수 있을 것으로 생각된다.

References

- [1] App Annie, Mobile Status Report 2019, 2019.
- [2] Basole, R. C., and Karla, J., "On the Evolution of mobile platform ecosystem structure and strategy," *Business & Information Systems Engineering*, Vol. 3, No. 5, pp. 313-322, 2011.
- [3] Blondel, V. D., Guillaume, J. L., Lambiotte, R., and Lefebvre, E., "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, Vol. 2008, No. 10, 2008.
- [4] Chun, Y., Chang, S., and Woo, T., "Classification of smartphone game based on mechanics," *Journal of Korea Game Society* Vol. 12, No. 6, pp. 15-23, 2012.
- [5] Lavid, B. L. D., and Kuflik, T., "Functionality-based clustering using short textual description," In *Proceedings of the 2013 International Conference on Intelligent user interfaces - IUI '13*. ACM Press, 2013.
- [6] Fox, S., and Duggan, M., *Mobile Health* 2012.
- [7] Genc-Nayebi, N., and Abran, A., "A systematic literature review: Opinion mining studies from mobile app store user reviews," *Journal of Systems and Software*, Vol. 125, pp. 207-219, 2017
- [8] Geum, Y., Lee, H., Lee, Y., and Park, Y., "Development of data-driven technology roadmap considering dependency: An ARM-based technology roadmapping," *Technological Forecasting and Social Change*, Vol. 91, pp. 264-279, 2015.
- [9] Huang, A., "Similarity measures for text document clustering," In *Proceedings of the Sixth New Zealand Computer Science Research Student Conference (NZCSRSC2008)*, Christchurch, New Zealand, 4, pp. 9-56. 2008.
- [10] Huh, J., Sung, J., Sohn, J., Lee, H., Chung, Y., and Paik, E., "Technology trends of mobile appstore," *Electronics and Telecommunications Trends (ETRI)*, Vol. 3, No. 25, 2010.
- [11] Islam, R., Islam, R., and Mazumder, T., "Mobile application and its global impact," *International Journal of Engineering & Technology (IJEST)*, Vol. 10, No. 6, pp. 72-78, 2010.
- [12] Jeon, B., and Ahn, H., "A collaborative filtering system combined with users' review mining: Application to the recommendation of smartphone apps," *Journal of Intelligence and Information Systems*, Vol. 21, No. 2, pp. 1-18, 2015.
- [13] Kajanan, S., Pervin, N., Ramasubbu, N., Datta, A., and Dutta, K., "Takeoff and Sustained Success of Apps in Hypercompetitive Mobile Platform Ecosystems: An Empirical Analysis," *ICIS*, 2012.
- [14] Kim, J., "Windows of opportunity in the mobile market, Application Store," *LG Business Insight*, Vol. 8, No. 19, pp. 17-35, 2009.

- [15] Kim, J., Park, Y., Kim, C., and Lee, H., "Mobile application service networks: Apple's app store," *Service Business*, Vol. 8, No. 1, pp. 1-27, 2014.
- [16] Kim, G., and Park, K., "A data-driven approach to categorization of mobile applications and analysis of application attributes affecting user evaluation," *Journal of the Korean Institute of Industrial Engineers*, Vol. 46, No. 2, pp. 156-163, 2020.
- [17] Kim, H., Lee, H., and Choi, S., "An Exploratory Study on the Determinants of Mobile Application Purchase," *The Journal of Society for e-Business Studies* Vol. 16, No. 4, pp. 173-195, 2011.
- [18] Lee, G., and Kim, J., "The design of the visual interface for the Mobile app-focusing on the internet bookstore," *Journal of Digital Design*, Vol.14, No.1, pp. 617-627, 2014.
- [19] Lee, H., Seol, H., Min, H., and Geum, Y., "The identification of new service opportunities: A case-based morphological analysis," *Service Business*, Vol. 11, No. 1, pp. 191-206, 2017.
- [20] Liu, C., Zhu, Q., Holroyd, K. A., and Seng, E. K., "Status and trends of mobile-health applications for iOS devices: A developer's perspective," *Journal of Systems and Software*, Vol. 84, No. 11, pp. 2022-2033, 2011.
- [21] Newman, M. E., "Modularity and community structure in networks," *Proceedings of the national academy of sciences*, Vol. 103, No. 23, pp. 8577-8582, 2006.
- [22] Oh, H., and Min, B., "App design style and usability analysis for smartphone application - focusing on the iphone and app-store," *Journal of the Korea Contents Association*, Vol. 10, No. 12, pp. 129-136, 2010.
- [23] Oh, Y., and Park, S., "Improving app categorization via noise filtering for app description," *The Korean Institute of Information Scientists and Engineers*, pp. 802-804, 2015.
- [24] Robertson, S. E., "The probability ranking principle in IR," *Journal of Documentation*, Vol. 33, No. 4, pp. 294-304, 1977.
- [25] Salton, G., and Buckley, C., "Term-weighting approaches in automatic text retrieval," *Information Processing & Management*, Vol. 24, No. 5, pp. 513-523, 1988.
- [26] Singhal, A., "Modern information retrieval: A brief overview," *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, Vol. 24, pp. 35-43, 2001.
- [27] Tiarawut, S., "Mobile technology: Opportunity for entrepreneurship," *Wireless Personal Communications*, Vol. 69, No. 3, pp. 1025-1031, 2013.
- [28] Vakulenko, S., Müller, O., and Brocke, J. V., *Enriching iTunes App Store categories via topic modeling*, 2014.
- [29] Xu, Q., Ibrahim, G., Zheng, R., and Archer, N., "Toward automated categorization of mobile health and fitness applications," In

- Proceedings of the 4th ACM MobiHoc workshop on Pervasive wireless health-care - MobileHealth '14. ACM Press, 2014.
- [30] Yoon, H., "Types and features of popular educational applications for young children - focusing on android app store," The HCI Society of Korea, pp. 481-484, 2014.
- [31] Yoon, Y., and Pahk, S., "Network embedding based mobile application clustering," The Korean Institute of Information Scientists and Engineers, pp. 622-624, 2019.

저 자 소 개



하소희

2019년

2019년~현재

관심분야

(E-mail: newhap.in.here@ds.seoultech.ac.kr)

서울과학기술대학교 산업정보시스템 전공 (학사)

서울과학기술대학교 일반대학원 데이터사이언스학과
(석사과정)

Data mining, Machine Learning



금영정

2004년

2013년 8월

2014년~2020년

2020년~현재

관심분야

(E-mail: yjgeum@seoultech.ac.kr)

KAIST 산업공학 (학사)

서울대학교 산업공학 (박사)

서울과학기술대학교 글로벌융합산업공학과 조교수

서울과학기술대학교 글로벌융합산업공학과 부교수

Technology Management, Data-driven innovation,

Service Engineering, Business Model Innovation