

# 건강추천시스템(HRS) 연구 동향: 인용네트워크 분석과 GraphSAGE를 활용하여\*

장하림

경희대학교 일반대학원  
빅데이터응용학과  
(gkfua00@khu.ac.kr)

유지수

경희대학교 일반대학원  
빅데이터응용학과  
(7slwm7@khu.ac.kr)

양성병

경희대학교 경영대학 경영학과  
& 빅데이터응용학과  
(sbyang@khu.ac.kr)

.....

현대사회는 정보통신기술 및 빅데이터 기술의 발전으로 누구나 인터넷을 통해 손쉽게 방대한 데이터를 얻고 활용할 수 있는 시대로, 양질의 데이터를 수집하는 능력을 넘어 수많은 정보 속에서 올바른 데이터만을 선별하는 능력이 더욱 중요해지고 있다. 이러한 기조는 학계에서도 이어지고 있는데, 축적되는 연구물 속에서 양질의 연구를 선별하여 올바른 지식 구조를 형성하기 위해, 다양한 연구 분야에서 체계적 고찰(systematic review) 및 비체계적 고찰(non-systematic review)과 같은 문헌연구(literature review)가 수행되고 있다. 한편, 코로나19 팬데믹 이후 의료산업에서도 그동안 합의에 이르지 못했던 원격의료의 제한적으로나마 허용되고, 인공지능 및 빅데이터 기술이 응용된 건강추천시스템(health recommender systems: HRS)과 같은 새로운 의료서비스가 각광을 받고 있다. 하지만, 실무적으로 HRS가 미래 의료산업 발전을 이끌 중요한 기술로 평가받고 있음에도 불구하고, 학술적인 문헌연구는 다른 분야에 비해 매우 부족한 실정이다. 더불어 HRS는 학제적 성격이 강한 융합 분야임에도 불구하고, 기존의 문헌연구는 비체계적 고찰과 체계적 고찰 방법만을 주로 활용하여 이뤄졌기 때문에, 다른 연구 분야와의 상호작용이나 동적인 관계를 유추하기에는 한계가 존재한다. 이에, 본 연구에서는 인용네트워크 분석(citation network analysis: CNA)을 활용하여 HRS 및 주변 연구 분야의 전체적인 네트워크 구조를 파악하였다. 또한, 이 과정에서 최신 논문이 인용 관계가 잘 나타나지 않는 문제를 보완하기 위해 GraphSAGE 알고리즘을 적용함으로써, HRS 연구에 있어 ‘recommender system’, ‘wireless & IoT’, ‘computer vision’, ‘text mining’ 등과 같은 연구 분야들의 중요도가 높아지고 있음을 파악하였으며, 이와 동시에 개인화(personalization) 및 개인정보보호(privacy) 등과 같은 새로운 키워드가 주요 이슈로 등장하고 있음을 확인하였다. 본 연구를 통해 HRS 연구 커뮤니티의 구조를 파악하고, 관련된 연구 동향을 살펴봄, 미래 HRS 연구 방향을 설계함에 있어 실질적인 통찰을 제공할 수 있을 것으로 기대한다.

**주제어** : 건강추천시스템, 헬스케어, 연구 동향, 인용네트워크 분석, GraphSAGE

.....

논문접수일 : 2023년 2월 16일    논문수정일 : 2023년 3월 4일    게재확정일 : 2023년 3월 5일  
원고유형 : Regular Track    교신저자 : 양성병

## 1. 서론

제4차 산업혁명 이후 현대사회는 정보통신기술 (information & communications technology: ICT)의 발달과 함께 사람과 사물, 사물과 사물이 연결된

최첨단 시대로 진입하고 있다(손진아, 2020). 동시에 컴퓨터 성능의 향상과 데이터 처리 기술의 발전, 보건의료빅데이터개방시스템 등 데이터센터의 설립으로 누구나 방대한 데이터를 얻고 활용할 수 있게 되었다(장요한, 2022). 그러나, 정보의

\* 본 논문은 연구재단 4단계 BK21 사업으로부터 지원받은 연구임.  
이 논문은 2022년 대한민국 정부(과학기술정보통신부)와 한국연구재단의 지원을 받아 연구되었음(NRF-2022K2A9A2A11097154).

양이 급증함에 따라, 수많은 데이터 속에서 필요한 것만을 선별하고 올바르게 처리하는 능력이 새롭게 요구되고 있기도 하다(조인동, 김남규, 2011). 이는 학술적 분야에서도 마찬가지로, 쏟아지는 연구물 속에서 양질의 정보만을 선별하고 이를 통해 올바른 지식구조를 파악하는 것이 더욱더 중요한 업무로 받아들여지고 있다(Sterling & Montemore, 2022).

서지학(bibliography)적 관점에서, 학술적 지식 구조는 완전히 새로운 정보로부터 생겨나는 것이 아니라, 기존의 지식을 기반으로만 건설되는 특징을 갖고 있다(Choi et al., 2011). 이에 따라, 다양한 연구자들은 체계적 고찰(systematic review) 및 비체계적 고찰(non-systematic review)과 같은 문헌연구(literature review)를 통해 기존의 연구물들을 살펴보고 해당 학술 분야의 지식구조를 파악하기 위한 연구들을 수행하고 있다(김현정, 안형식, 2015; Scheuren, 2004). 이러한 문헌연구들은 해당 학술 분야의 연구자들 간 지식공유를 가능케 하고, 견고한 지식구성 체계 구축에 기여하며, 학술적 발전을 이끄는 주축이 되고 있다(이동호 등, 2012). 더불어, 최근에는 다양한 주제와 방법론을 결합한 학제적 접근(interdisciplinary approach)이 증가하고 있고, 주변 연구 분야와의 상호작용도 함께 살펴봐야 할 필요성이 커지고 있어, 문헌연구의 중요성은 더욱더 강조되고 있는 실정이다(강은경 등, 2022).

한편, 인공지능 및 빅데이터 기술과 스마트기기(smart devices)의 발전으로 의료산업에서도 건강 추천시스템(health recommender systems: HRS)과 같이 컴퓨터 과학이 융합된 기술 분야가 태동하고 있다(문장섭, 2021). 이러한 HRS 연구는 PubMed 사이트 기준 2014년 11건에서, 2022년 140건으

로 그 수가 폭발적으로 증가하고 있으며(NCBI, 2023), 이에 따라 HRS 분야에서도 해당 학술 분야의 지식구조를 체계적으로 파악하고자 하는 다양한 연구가 진행되고 있다(De Croon et al., 2021). 그러나, HRS 분야의 문헌연구 수는 다른 건강 분야의 연구에 비해 매우 부족한 실정이며, 대부분의 선행연구가 체계적 혹은 비체계적 고찰 방법론을 활용한 연구가 대부분인 관계로(<표 1> 참조), 인접한 연구 분야와의 상호작용이나 동적인 관계를 분석함에 있어 한계가 명확히 존재한다(Leppink & Perez-Fuster, 2019).

이에, 본 연구에서는 인용네트워크 분석(citation network analysis: CNA)을 기반으로 HRS 연구 커뮤니티의 전체적인 구조를 파악하고, 동향을 분석해 보고자 한다. 한편, CNA를 활용할 경우, 최근 게재된 논문들의 인용 관계가 약하게 나타나는 단점이 존재하는데, 이를 해결하기 위해 그래프 인공신경망(graph neural network: GNN)의 일종인 GraphSAGE 알고리즘을 적용하여 이러한 문제를 보완하고자 하였다. 이를 통해 HRS 연구 분야의 과거부터 현재까지를 폭넓게 조망하고, 학제적인 관점에서 HRS 연구를 둘러싼 다양한 연구 분야 간 상호영향을 분석해 보고자 하였다. 결과적으로, 본 연구는 (1) 기존의 건강(health) 관련 분야에 비해 연구가 적었던 HRS 분야의 문헌연구를 보완하고, (2) 인용네트워크 분석을 활용하여 기존의 HRS 문헌연구가 살펴보지 못했던 동적인 관계를 파악하며, (3) 기존의 인용네트워크 분석에서 활용이 적었던 GNN 알고리즘을 적용하여 HRS 분야의 미래 동향을 예측함으로써, 후속 연구자들에게 통찰력을 제공할 수 있을 것으로 기대한다.

## 2. 이론적 배경

### 2.1 헬스케어 산업

최근 인구 증가, 출산율 감소, 고령화 심화 및 의료비 증가 등 다양한 사회문제가 불어지며, 건강 분야에 대한 사람들의 관심이 증가하고 있다(Etemadi et al., 2023). 국내의 경우, 2014년 이후 국내총생산(gross domestic product: GDP) 대비 의료비 비중이, OCED(organisation for economic co-operation and development) 평균 0.1% 포인트 성장하는 동안 4.8% 포인트 상승하였으며, 2021년 기준 경상의료비의 직접 부담 비율은 30.2%로, OECD 평균인 19.8%를 훨씬 상회하고 있는 것으로 나타났다(김기태, 2022). 게다가, 이현복과 이수연(2019)의 시나리오에 의하면, 2060년에는 노인의료비의 규모가 GDP 대비 약 5%에 육박할 정도로 상승하여, 인구고령화에 따른 의료비 지출이 더욱 증대할 것으로 예측되고 있다.

한편, 스마트폰과 연동된 웨어러블 디바이스(wearable device)가 확대되며, 일상에서 생성되는 라이프 로그(life log) 데이터가 끊임없이 수집되고 있다(정혜실, 2014). 특히, 국내에서는 건강보험심사평가원에서 국민의 약 98%에 달하는 의료데이터를 체계적으로 수집하고 있는 등, 강력한 건강 빅데이터가 구축되고 있다(Kim et al., 2017). 이와 동시에, 코로나19 팬데믹 이후 그동안 합의에 이르지 못했던 원격의료의 제한적으로나마 허용되며, ICT 기반 스마트 헬스케어 서비스 또한 태동하기 시작했다(오의금, 2020). 이러한 사회의 변화에 따라 의료서비스도 기존의 공급자 중심에서, 소비자가 중심이 되는 정밀의료의 시대로 접어들고 있다(문장섭, 2021).

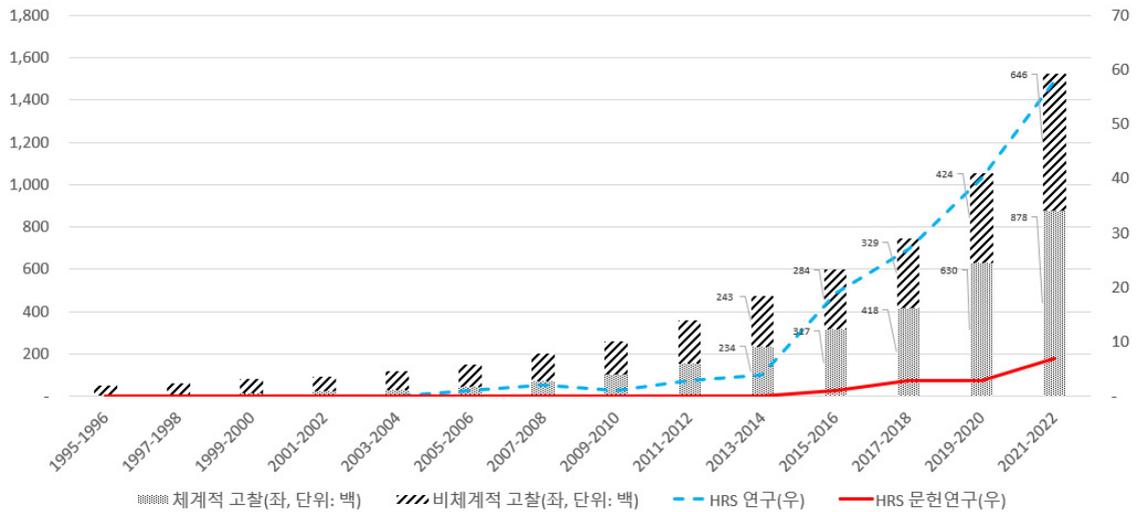
이처럼 의료 시장의 규모가 커짐에 따라, 헬스

케어 산업은 연평균 25%씩 성장하며 가파른 상승세를 이어가고 있다(김형철, 홍길중, 2022). 미래의 의료 기술로는 인공지능과 로봇공학 기술을 이용한 치료가 연구되고 있으며(채수미, 2020), 빅데이터를 이용한 질병 진단 기술이나 실시간 생체정보 측정, 분석 기술이 유망 기술로 거론되고 있다(배영임, 신혜리, 2020). HRS는 이러한 요구에 부응하는 대표적인 학술 분야로, 기존의 정보 기술을 응용하였던 의료전문가시스템(medical expert system)이나, 임상 의사결정지원시스템(clinical decision support system)과는 다르게 이용자의 생체정보를 활용한 개인맞춤형 건강서비스를 제공한다는 점에서 이전의 기술과는 차별화된다.

<그림 1>은 PubMed 사이트에 게시된 HRS 연구 및 HRS 문헌연구의 수와 함께 건강 분야의 체계적 고찰 연구와 비체계적 고찰 연구의 수를 그래프로 나타낸 것이다(NCBI, 2023). HRS와 관련한 연구는 2014년 기준 11건에서 2022년 기준 140건으로 그 수가 계속해서 증가하고 있으나, HRS 분야를 다루는 문헌연구는 2016년부터 2022년까지 총 13건에 불과해 다른 건강 분야의 문헌연구에 비해 현저히 적은 상황이다. 따라서, 본 연구에서는 HRS 분야에서의 인공네트워크 분석을 수행함으로써, 기존의 건강 분야 문헌연구에 비해 부족한 HRS 분야의 문헌연구를 보충함과 동시에, HRS 분야를 폭넓게 조망함으로써 후속 연구자를 위한 현재 및 미래 연구 동향을 제시하고자 한다.

### 2.2 건강추천시스템(HRS)

추천시스템은 정보 검색(information retrieval)과 경영과학(management science) 분야에 근간을 두고 있는 기술로, 정보의 홍수 속에서 개인에게 필요한 정보를 선별하는 알고리즘을 연구하는



〈그림 1〉 건강 분야 문헌연구 및 건강추천시스템(HRS) 관련 연구 출판물의 수

분야이다(Kamran & Javed, 2015). 추천시스템은 기본적으로 협업필터링(collaborative filtering: CF) 및 콘텐츠기반 필터링(content-based filtering: CBF)과 같은 필터링 기법들로부터 발전되기 시작하여(Ricci et al., 2011), 현재에 이르러서는 행렬 분해(matrix factorization), 인공신경망(artificial neural network: ANN) 및 그래프기반 신경망 등 다양한 알고리즘과 함께 여러 실무 분야에서 효과적으로 응용되고 있다(Wang et al., 2019).

HRS는 이러한 추천시스템 알고리즘을 건강 분야에 적용한 것으로, 사용자에게 신뢰할 수 있는 건강 정보를 제공함으로써 이용자의 건강을 증진시키고, 의학적 판단을 보조하는 기술을 의미한다(Wiesner & Pfeifer, 2014). HRS가 이전의 기술과 분명하게 구분되는 특징은 사용자에게 일반적인 의학적 지식이나 정보를 제공하는 것이 아니라, 개별 사용자의 특징을 반영한 개인맞춤형 정보를 제공한다는 점이다(Tran et al., 2021). 이러한 이유로 HRS 기술은 면밀하게 검

증된 의학적 지식뿐만 아니라 개인건강기록시스템(personal health record system)과 같이 체계적인 개인건강 정보수집 기술이 추가적으로 요구되고 있다(Wiesner & Pfeifer, 2014).

Wiesner and Pfeifer (2014)에 따르면, HRS 분야는 의료, 임상, 의약 등과 같은 건강전문가를 위한 추천과 식단, 생활 습관, 운동, 건강행동 등 개인의 건강에 영향을 주는 추천으로 구분된다. De Croon et al.(2021)은 일반인을 대상으로 하는 HRS 분야의 체계적 고찰 연구를 수행한 결과, 생활습관(lifestyle), 영양소, 일반적 건강정보, 특정한 건강관련 추천 등 네 가지 영역의 연구들이 존재하고 있음을 파악하였다. 여기서 나아가, Etemadi et al.(2023)의 HRS 관련 최신 체계적 고찰 연구에서는 영양소 정보, 명상 정보, 치료 계획 정보, 진단 정보, 질병 예측, 육체적 운동, 의사/병원 탐색, 적합한 의료서비스 탐색 등의 더욱 다양한 세부 연구 분야를 제시하고 있다. <표 1>에서는 HRS의 선행 문헌연구들을 체계적 고찰

〈표 1〉 건강추천시스템 분야의 선행 문헌연구

문헌연구 유형	저자(출판연도)	연구내용
비체계적 고찰 (non-systematic review)	Wiesner and Pfeifer(2014)	건강추천시스템에 대해 정의하고, 범위를 설정하는 등 연구 분야를 확립함.
	Cappella et al.(2015)	추천시스템의 발전을 살펴보고, 건강메시지 서비스에 적용될 수 있는 부분을 탐색함.
	Kamran and Javed(2015)	헬스케어 분야에 적용되고 있는 추천시스템 알고리즘에 대해 살펴보고, 이를 건강 분야에 적용하기 위한 추천시스템을 제안함.
	Calero Valdez et al.(2016)	건강추천시스템 논문 작성을 위한 3단계의 프레임워크를 제안함.
	Mohamed et al.(2017)	건강추천시스템 연구물을 살펴보고, 추천시스템 알고리즘이 갖고 있는 문제와 함께 헬스케어 분야에서의 특수한 문제들을 제시함.
	Schäfer et al.(2017)	추천시스템의 발전에 대해 살펴보고, 헬스케어와 같이 위험부담이 큰 분야에서 해결해야 할 문제들을 제시함.
	Sharma et al.(2019)	헬스케어 4.0 시대에 발맞춰, 인간 중심의 헬스케어 시스템과 건강추천시스템의 적용에 대해 살펴봄.
	Su et al.(2020)	건강추천시스템의 최신 관심 분야, 방법론, 평가방법, 이후 해결 과제 등을 분석함.
	Yue et al.(2021)	추천시스템의 세 가지 기술을 살펴보고, 건강분야에 적용할 수 있는 다섯 가지의 응용시나리오를 제시함.
체계적 고찰 (systematic review)	Sezgin and Özkan(2013)	건강추천시스템에 대한 최초의 체계적 고찰 연구를 수행함.
	Ferretto et al.(2017)	모바일 앱을 이용한 건강추천시스템에 대한 체계적 고찰 연구를 수행함.
	Hors-Fraile et al.(2018)	주제범위 문헌고찰(scoping review)을 통해 건강추천시스템의 긍정적 효과에 대한 증거를 제시함.
	Cheung et al.(2019)	사용자와 추천의 상호작용에 초점을 맞춰 어떤 방식으로 추천이 제공되는지를 분석함.
	Pincay et al.(2019)	분류 프레임워크를 통해 건강추천시스템 문헌들을 여러 가지 하위 분야로 분류함.
	Çelik and Elçi(2020)	건강 3.0 시대에 발맞춰, 개인맞춤형 건강추천시스템의 기술 및 응용, 성공적인 사례들을 살펴봄.
	De Croon et al.(2021)	일반인 대상으로 하는 건강추천시스템에 대해 분석함.
	Tran et al.(2021)	추천대상에 따라 어떠한 추천시스템 알고리즘이 사용되는지 분석함.
	Etemadi et al.(2023)	건강추천시스템의 현재 위치에 대해 분석하고, 해결해야 할 문제점을 제시함.

과 비체계적 고찰로 나눈 뒤, 각 연구의 내용을 요약하여 제시하였다.

본 연구에서는 AMiner Dataset에서 제공하는 데이터를 이용하여 인용네트워크 분석을 진행한

뒤, 기존의 체계적 고찰이나 비체계적 고찰 연구로는 파악할 수 없었던 HRS 분야의 연구 네트워크를 조망하고, 현재 및 미래의 연구 동향에 대해 살펴보고자 한다.

### 2.3 네트워크 분석

네트워크 분석은 네트워크를 구성하는 대상을 노드(node)로 하고, 이 대상들 간의 연결관계인 엣지(edge)를 바탕으로 전체적인 네트워크 구조를 파악하는 방법이다(정유경, 2020). 네트워크 분석에 주로 쓰이는 지표 중 하나는 중심성(centrality)으로, 전체 네트워크에서 특정 노드가 중심에 위치하는 정도를 의미하며, 다양한 방법으로 측정할 수 있다(박준형, 곽기영, 2013). 대표적 중심성 지표는 (1) 노드에 따라 연결된 링크 수를 측정하는 연결중심성(degree centrality), (2) 노드와 다른 노드들 사이에 중계자 역할의 정도를 측정하는 매개중심성(betweenness centrality), (3) 한 노드에서 다른 노드에 도달할 때 필요한 최소 단계를 측정하는 근접중심성(closeness centrality), 그리고 (4) 노드 간 연결중심성에서 얼마나 영향력 있는 노드와 연결되어 있는지에 따라 가중치를 두어 측정하는 위세중심성(eigenvector centrality) 등으로 나눌 수 있다(정중희, 김종우, 2013).

한편, 학술지 인용네트워크 분석은 학술지 간의 인용 관계를 네트워크적으로 표현하여 그 관계를 분석하는 기법으로, 네트워크 시각화의 정량적인 측정을 위해 계량서지학에서 자주 활용되어 왔다(Otte & Rousseau, 2002). 좀 더 구체적으로, 학술지 인용네트워크 분석은 (1) 거시적 수준에서 전체 과학 분야의 구조 분석(Bassecoulard & Zitt, 1999), (2) 중위적 수준에서 특정 학문 분야의 지식구조 분석(Ding et al., 2000), (3) 미시적 수준에서 특정 학술지와 관련된 다른 학술지들 간의 관계 분석(Calero Medina & van Leeuwen, 2012) 등에 활용될 수 있다. 노드는 분석 대상으로서 키워드, 논문, 학술지 등의 학술 정보 또는 저자, 기관, 국가 등의 학술 주체로 다양하게 정의

될 수 있으며(Noyons, 2001), 엣지는 노드들 간의 관계로서, 인용(citation), 동시인용(co-citation), 공동저술(co-authorship), 동시출현단어(co-word), 동시분류(co-classification) 등 다양한 계량서지분석 지표를 통해 정의될 수 있다(White & McCain, 1997).

이러한 학술지 인용네트워크 분석은 학문 분야의 다학제적 특성을 파악하는데 유용하게 활용될 수 있어(Leydesdorff, 2007), 기술경영(Lee, 2015), 정보과학(Borgman & Rice, 1992), 물리화학(Leydesdorff, 1994), 나노생물학(Rafols & Meyer, 2010), 나노기술(Zhou & Leydesdorff, 2007) 등 융합학문의 학술적 구조 파악에 효과적으로 활용되어 왔다. 하지만, 건강분야와 추천시스템이 융합된 HRS 연구 분야에서 학술지 인용네트워크 분석을 활용한 연구는 많이 이뤄지지 못하고 있는 실정이다.

한편, 학술지 인용네트워크 분석은 인용 관계가 두드러지게 나타나지 않는 경우(예: 최근 게재된 논문), 분석의 정확도가 떨어지는 문제가 지적되어 왔는데, 본 연구에서는 Wang(2013)이 제시한 방법론을 따라 아직 인용 관계가 강하게 형성되지 못한 최근 4년 이내(2018년-2021년)의 자료에 그래프 인공지능망 방법론의 일종인 GraphSAGE 알고리즘을 적용하여 HRS 네트워크 분석에 활용하고자 한다. 이를 통해, 기존의 인용 관계뿐만 아니라 키워드, 저널명, 발행 연도 등을 함께 고려함으로써, 인용 관계가 약하게 나타나는 자료의 경우에도 보다 정확한 분류가 가능하게 하였다.

### 2.4 그래프 인공지능망

그래프 인공지능망(GNN)은 Scarselli et al.(2009)에 의해 제안된 개념으로, 그래프의 노드와 엣지로 표현될 수 있는 정보들에 인공지능망의 방법을

적용한 알고리즘을 칭한다. 이러한 GNN은 기존의 알고리즘으로 파악할 수 있는 정보의 유사도를 파악할 수 있을뿐만 아니라, 데이터 간의 연결 관계나, 위상학적 구조(topology) 등을 포착하는데 유용하게 활용될 수 있다(Perozzi et al., 2014). 따라서, GNN은 지식그래프(knowledge graph)나 물리시스템(physical system), 단백질 구조 등 그래프로 구성된 정보를 처리하는 데 있어 기존의 알고리즘보다 월등한 성능을 보이는 것으로 알려져 있다(Wu et al., 2022).

한편, 학술 분야에서도 공저자 관계, 사제 관계(advisor-advisee) 등을 이용한 지식그래프를 구축하기도 하는 등 기존의 인용 관계만 이용하던 접근 방법을 넘어 다양한 정보를 활용하고자 하는 시도들이 나타나고 있다(Liu et al., 2020). 연구자는 이렇게 구성된 그래프를 통해 노드 분류(node classification), 연결 예측(link prediction), 군집분석(clustering), 시각화(visualization) 등과 같은 분석을 수행할 수 있다(Goyal & Ferrara, 2018). Xiao et al.(2019)은 가장 널리 알려진 GNN 알고리즘 중 하나인 GraphSAGE를 이용하여 그래프의 노드를 분류하는 작업을 수행하였으며, Pham and Dang(2021)은 GNN의 가장 기본적인 알고리즘인 GCN(graph convolution network)과 Node2Vec을 이용하여 의생명과학 분야 네트워크의 연결 예측을 수행한 바 있다.

이같은 GNN 알고리즘은 기존의 노드 간 연결만을 통해 네트워크를 분석하던 방법을 넘어, 노드의 고유한 정보를 함께 고려한다. 특히, GraphSAGE 알고리즘은 이웃 노드의 상호작용 정보와 노드 고유의 속성 특성을 활용함으로써, 노드를 분류하는 데 높은 정확도를 보이는 것으로 알려져 있다(Xiao et al., 2019). 이에, 본 연구에서도 GraphSAGE를 HRS 연구 네트워크 분석에 적용하였는데,

GraphSAGE는 GCN의 성공 이후 발전된 알고리즘 중 하나로(Pham & Dang, 2021), 추론적 학습(transductive learning)을 하는 GCN과 달리 귀납적 학습(inductive learning)을 활용하여 예측성능을 더욱 높인 알고리즘으로 평가받고 있다(최건영, 홍참길, 2020). 이 알고리즘은 네트워크를 구성하는 노드가 동질적인 성격을 띠고 있는 동종 네트워크(homogeneous network)에서 주로 활용될 수 있으며(정이태, 안현철, 2020), 핵심 수식은 다음의 수식 (1) 및 (2)와 같다.

$$h_N^k = \text{AGGREGATE}_K(h_u^{k-1}, \forall u \in N(v)) \quad (1)$$

$$h_v^k = \sigma(W^k \cdot \text{CONCAT}(h_v^{k-1}, h_{N(v)}^k)) \quad (2)$$

GraphSAGE는 먼저 타겟이 되는 이웃 노드들의 정보들을 AGGREGATE 함수를 통해 규합하여  $h_N^k$ 를 형성한 뒤, 이전 반복 회차의 타겟 정보  $h_v^{k-1}$ 와 CONCAT하여 활성화 함수를 지나 현재 회차의 타겟 정보  $h_v^k$ 를 갱신한다(Hamilton et al., 2017). 이 과정에서 모델은 정보를 규합하는 방식에 대해 학습하게 되고, 이렇게 학습된 모델을 통해 연결을 예측하거나, 노드를 분류할 수 있다(Wu et al., 2022). 본 연구는 이러한 GraphSAGE 알고리즘을 이용하여 인용 관계가 비교적 약하게 나타나는 최신 논문들을 분류함에 있어 보다 정확한 분류를 수행하고자 한다.

### 3. 연구방법

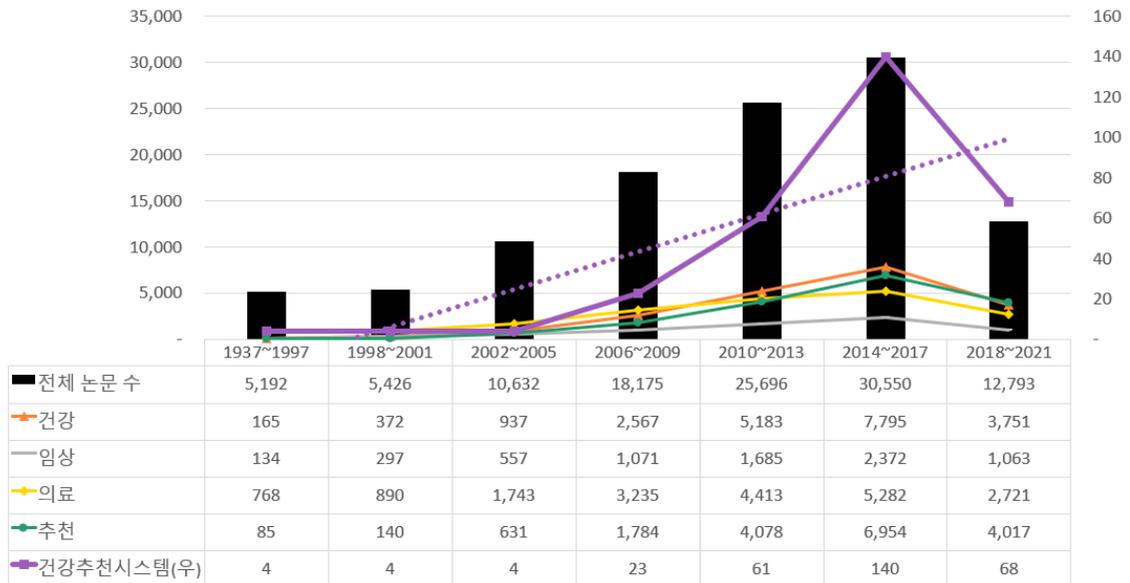
#### 3.1 데이터 수집

본 연구는 AMiner Dataset에서 제공하는 ‘Citation Network 공공데이터’를 이용하였다. 해당 데이터는

서지학 사이트의 논문들을 이용하여 구성된 데이터로, 2021년 5월 14일에 업데이트된 DBLP Dataset V.13에는 총 5,354,309개의 논문과 48,227,950개 인용 관계가 포함되어 있다. 이후 중복을 제거한 뒤, 해당 데이터에 구성된 ‘아이디’, ‘논문 제목’, ‘연도’, ‘피인용수’, ‘저널명’, ‘키워드’, ‘논문분야’, ‘참고문헌’ 정보를 추출하였고, 참고문헌 정보가 존재하지 않는 결측치를 제거하여 총 3,678,763개의 논문 정보를 구성하였다. 이후 파이썬(Python) 3.8.10 버전을 활용하여 논문 제목 정보로부터 건강 추천시스템 및 건강 관련 논문만을 선별하였다. 논문의 선별은 기존의 문헌연구를 참고하여, 건강(health, healthy, healthcare), 의료(medical) 및 임상(clinical, clinic)과 함께 추천(recommend, recommender, recommender system, recommendation)이 제목에 포함된 논문을 추려내는 방식으로 진행하였다(Celik & Elci, 2020; De Croon et al., 2021; Etemadi et

al., 2023). 이후 추려진 논문의 이웃 노드들을 파악하여 자료에 포함하였으며, HRS와 가장 연관성이 높은, 건강(health), 의료(medical), 임상(clinical), 추천시스템(recommender system) 논문들을 자료에 추가하였다. 해당 과정을 통해 최종적으로 108,464개의 논문과 836,286개의 인용 관계를 선별하였다.

본 연구에서 수집된 연도별 게재 논문의 수는 <그림 2>와 같다. 수집된 논문은 파이썬의 Networkx 라이브러리와 Gephi 0.10.1 버전을 이용하여 인용 네트워크 그래프를 만들고 시각화하였고, 이후 파이썬의 StellarGraph, Sklearn, Tensorflow Keras 라이브러리를 이용하여 GraphSAGE 알고리즘을 구현하여 인용 관계가 약한 2018~2021년도 자료를 분류하고, HRS를 구성하는 하위 네트워크(component)들의 세부 분야를 파악하였다.



<그림 2> 건강추천시스템 및 관련 분야 논문의 수

### 3.2 데이터 전처리

본 연구의 전처리 과정은 앞 절에 설명된 과정을 통해 형성된 그래프에 파생변수를 추가하는 작업으로 진행되었다. 먼저 L1 열을 만들어 건강추천시스템 논문으로 분류된 데이터를 '1'로 표시하였다. 이후 L2 열을 만들어 L1 논문이 인용하고 있는 데이터를 '1'로 표시하였다. 이와 함께, L1 논문을 인용하고 있는 데이터도 '1'로 표시하였다. 같은 방식으로 L2 논문을 이용하여 L3 열을 만들어 논문을 표시하였다. 이후, L4 열을 만들어 건강, 의료, 임상, 추천시스템 관련 논문들을 '1'로 표시하였다. 마지막으로, 건강, 의료, 임상, 추천시스템 열을 만들어서 해당 키워드가 논문 제목에 포함된 자료들을 '1'로 표시하였다. 이렇게 표시된 데이터와 함께, 저널 정보, 키워드 정보, 연구 분야 정보를 파이썬의 Pandas 라이브러리를 이용하여 원핫인코딩(one-hot encoding)을 수행하였다. 이러한 과정을 거쳐, 최종적으로 108,464개의 노드와 443개의 열로 이루어진 행렬을 구성하였다.

### 3.3 인용네트워크 분석

본 연구에서는 HRS 분야 전체 네트워크의 학제적 관계를 파악하기 위해 Gephi의 그래프 통계 분석을 이용하여 평균 연결정도(average degree), 지름(diameter), 평균 거리(average path length), 결집계수(clustering coefficient)를 파악하였다. 이후, 전체 네트워크의 모듈러리티(modularity)를 분석하여 네트워크의 군집성을 파악하고, 하위 네트워크를 분류하고자 하였다(박은희, 2021). 이후, 분류된 그래프를 시각화하였고, 하위 네트워크의 키워드 정보, 연구 분야 정보, 저널명 정보를 통합하였다. 이를 통해 먼저 HRS 주변 네트워크

들의 연구 분야와 그 비율을 파악하고자 하였다.

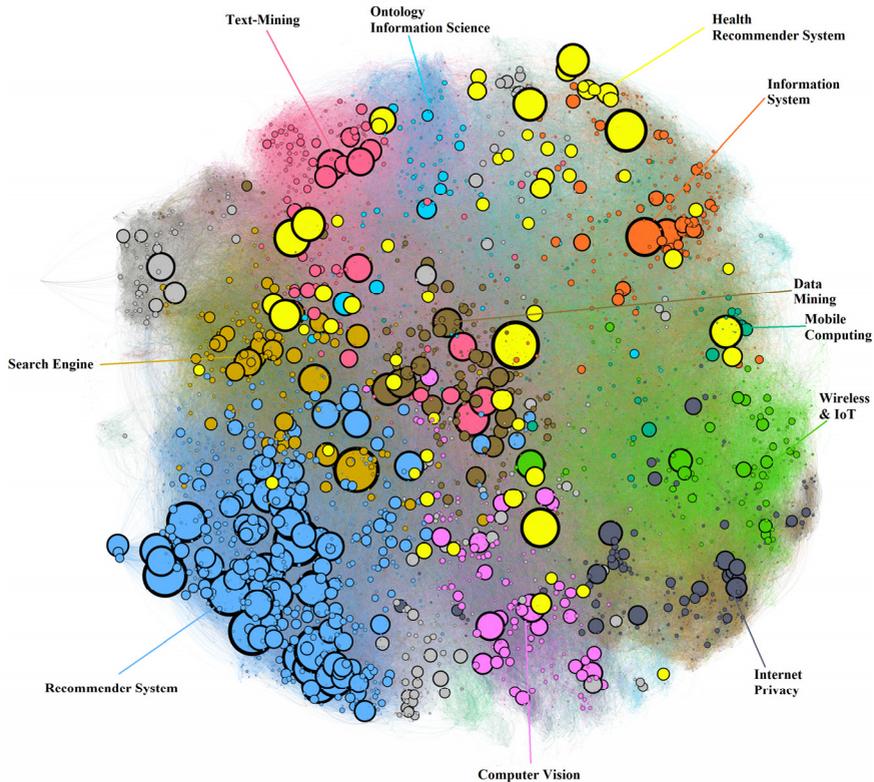
HRS 주변 연구 네트워크를 세부적으로 살펴보기 위해서 전체 논문을 4개년도 기준으로 나누어, 1937-1997(기간1), 1998-2001(기간2), 2002-2005(기간3), 2006-2009(기간4), 2010-2013(기간5), 2014-2017(기간6), 2018-2021(기간7) 등 총 일곱개 기간으로 분류하여 분석을 시행하였다. 다만 '기간1'의 경우, 전체 논문을 포함하기 위해 첫 연구가 게재된 이후 시점부터 1997년까지로 기간을 넓게 잡았으며, '기간7' 자료의 경우, 논문의 특성상 인용이 나타나기까지 평균적으로 3년 정도의 기간이 필요하다는 Wang(2013)의 연구 결과에 따라, 모듈러리티를 이용한 해석에서 제외하였다. 실제로 '기간7'의 경우, 평균 연결정도가 1보다 작은 0.968이며, 전체 연구를 통틀어서도 평균 피인용 수가 9.34로, '기간1'의 593.59에 비해 압도적으로 낮은 경향을 보이는 것으로 나타났다. 따라서, 본 연구에서는 인용 관계만을 이용하는 모듈러리티를 대신하여 GraphSAGE 알고리즘을 이용해 '기간7'의 하위 네트워크를 파악하고자 하였다.

이후 본 논문에서는 '기간1'부터 '기간6'까지의 네트워크 시각화 자료와 노드 수, 링크 수, 평균 연결정도, 지름, 평균 거리, 평균 피인용 수, 결집계수(clustering coefficient) 등을 계산하였다. 또한, 연도별로 주변 연구 분야 상위 10개를 표시하여, HRS 연구가 어떤 연구 분야와 연관되어 있는지를 시간의 흐름에 따라 제시하였다. 이와 동시에, 주변 연구 분야의 연도별 추이를 함께 살펴봄으로써, HRS 주변 연구가 시간에 따라 어떤 양상을 보이는지를 확인하였다. 마지막으로, HRS와 주변 연구 분야의 키워드 및 주요 학술지를 시간에 따라 살펴봄으로써, 연구 흐름의 큰 가닥을 찾아 제시하고자 하였다.

### 3.4 GraphSAGE를 통한 네트워크 분류

본 연구에서는 인용네트워크 분석을 활용하여 HRS 관련 연구 동향을 세부적으로 살펴보고자 하였다. 다만, ‘기간7’의 HRS 및 주변 연구 분야를 파악함에 있어 해당 데이터는 최신 정보를 담고 있음에도 불구하고, 발행 후 3년 이내의 논문들은 적은 인용 관계가 나타나기 때문에 인용네트워크 분석의 정확도가 떨어질 수 있다(Wang, 2013). 이에 따라, 본 연구에서는 그래프의 위상학적 구조와 함께 노드의 정보들을 함께 고려할 수 있는 GraphSAGE 알고리즘을 적용하여 ‘기간7’ 연구물들을 분류함에 있어 정확성을 보완하

고자 하였다. 분석 과정은 먼저, 전체 데이터를 ‘기간1’부터 ‘기간6’까지, 그리고 ‘기간7’로 각각 구분하여 나누었고, 전자를 학습데이터, 후자를 귀납적 분류(inductive classification)의 테스트 데이터로 이용하였다. 학습 과정의 활성화 함수로는 Relu를 사용하였고, 최종적으로는 Softmax 함수를 사용하였다. Optimizer로는 Adam Optimizer를 사용하였으며, GraphSAGE의 Aggregation 함수로는 MeanAggregator를 사용하였다. 학습 과정에서 25%의 데이터를 검증 데이터(validation data)로 이용하였고, 이 때 교차 엔트로피 손실(cross entropy loss)는 0.8765, 정확도(accuracy)는 74.55%로 나타났다.



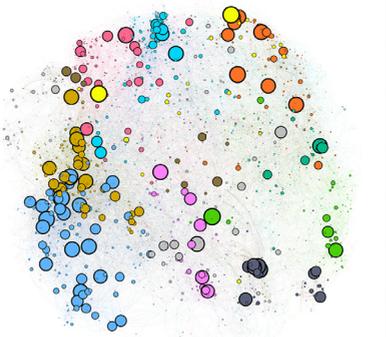
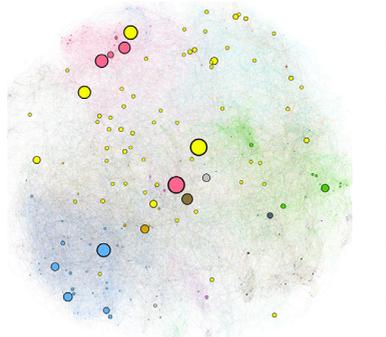
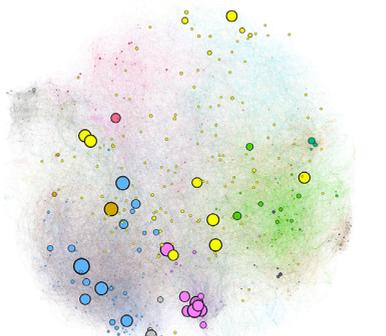
〈그림 3〉 전체 네트워크 시각화

〈표 2〉 HRS 주변 하위 네트워크의 키워드 및 연구 분야

순위	비율	논문 수	키워드	연구 분야
1	14.43%	15,648	Recommender System	Recommender System
2	11.07%	12,002	Health Informatics, Electronic Health Record	Health Care, Information System
3	9.29%	10,080	Health Care, Internet of Things, Wireless Sensor Networks	Health Care, Wireless
4	8.27%	8,973	Feature Selection, Support Vector Machine, Neural Network	Data Mining, Pattern Recognition, Mathematical Optimization
5	7.57%	8,210	Feature Extraction, Image Segmentation, Image Classification	Computer Vision, Pattern Recognition, Facial Recognition System
6	7.17%	7,772	Natural Language Processing, Bioinformatics	Annotation, Text Mining
7	5.73%	6,214	Mobile Computing, Ubiquitous Computing, Sensors	Mobile Computing, Internet Privacy
8	5.41%	5,871	Semantic Web, Description Logic	Ontology, Information Science, World Wide Web
9	5.38%	5,868	Information Retrieval, Data Mining	Web Search Query, Social Media, Search Engine
10	4.85%	5,836	Data Privacy, Cloud Computing	Internet Privacy, Authentication, Cryptography

〈표 3〉 인용네트워크 분석 결과

기간	전체 네트워크	범주	값	주변 연구 분야
~1997		노드 수	5,192	Data Mining(18.16%) Expert System(14.75%) Search Engine(10.42%) Information System(9.9%) Ontology(7.11%) Computer Vision(7.11%) Internet Privacy(5.2%) Recommender System(3.99%) Text Mining(3.74%) Mobile Computing(3.47%)
		링크 수	14,359	
		평균 연결 정도	2.776	
		지름	16	
		평균 거리	4.702	
		평균 피인용 수	594	
		결집 계수	0.104	
1998~2001		노드 수	6,264	Data Mining(15.24%) Information System(13.56%) Expert System(9.64%) Search Engine(8.31%) Recommender System(7.85%) Ontology(7.39%) Computer Vision(7.34%) Text Mining(6.01%) Internet Privacy(5.2%) Mobile Computing(4.02%)
		링크 수	10,293	
		평균 연결 정도	1.643	
		지름	14	
		평균 거리	3.841	
		평균 피인용 수	432	
		결집 계수	0.071	

기간	전체 네트워크	범주	값	주변 연구 분야
2002~2005		노드 수	10,632	Information System(13.94%) Data Mining(11.43%) Recommender System(9.95%) Ontology(9.63%) Computer Vision(7.28%) Text Mining(7.25%) Expert System(6.42%) Search Engine(6.25%) Wireless & IoT(5.58%) Internet Privacy(4.94%)
		링크 수	16,017	
		평균 연결 정도	1.506	
		지름	11	
		평균 거리	2.587	
		평균 피인용 수	306	
		결집 계수	0.062	
2006~2009		노드 수	18,175	Information System(13.2%) Recommender System(11.72%) Ontology(8.84%) Data Mining(8.58%) Wireless & IoT(8.28%) Computer Vision(7.02%) Text Mining(6.94%) Search Engine(6.36%) Mobile Computing(5.36%) Expert System(5.35%)
		링크 수	26,536	
		평균 연결 정도	1.46	
		지름	13	
		평균 거리	3.129	
		평균 피인용 수	157	
		결집 계수	0.055	
2010~2013		노드 수	25,696	Recommender System(14.22%) Information System(11.59%) Wireless & IoT(10.81%) Data Mining(7.35%) Computer Vision(7.34%) Text Mining(7.12%) Mobile Computing(6.92%) Internet Privacy(5.66%) Ontology(5.12%) Search Engine(4.83%)
		링크 수	39,714	
		평균 연결 정도	1.546	
		지름	12	
		평균 거리	3.079	
		평균 피인용 수	99	
		결집 계수	0.051	
2014~2017		노드 수	30,550	Recommender System(17.08%) Wireless & IoT(11.87%) Information System(9.48%) Computer Vision(7.55%) Text Mining(7.55%) Mobile Computing(6.47%) Data Mining(6.26%) Internet Privacy(5.65%) Search Engine(4.58%) Expert System(3.21%)
		링크 수	53,128	
		평균 연결 정도	1.739	
		지름	18	
		평균 거리	3.848	
		평균 피인용 수	64	
		결집 계수	0.049	

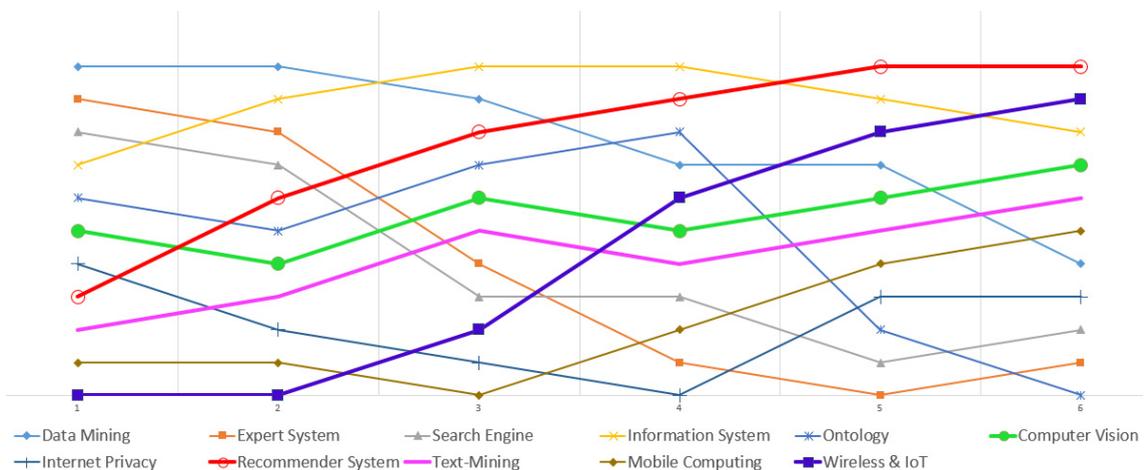
## 4. 연구결과

### 4.1 인용네트워크 분석 결과

본 연구의 분석 결과를 크게 두 부분으로 나누어 제시하였다. 먼저 전체 네트워크를 분석한 것으로, 네트워크의 노드는 총 108,464개이며, 전체 연결은 836,286개가 존재하였다. 네트워크의 평균 연결정도는 7.71이며, 네트워크의 지름은 44, 평균 거리는 8.509로 나타났다. 하위 네트워크를 파악하기 위해 모듈러리티를 분석하였고, 이를 통해 전체 네트워크에서 22개의 하위 네트워크가 존재함을 확인하였다. <그림 3>은 Gephi의 모듈러리티 분석을 이용하여 2018년 이전 상위 10개의 연구 영역을 색으로 구분한 뒤 전체 네트워크를 시각화한 자료이며, HRS 논문의 경우에는 최소 크기를 보정 후 노란색으로 표현하였다. 노드의 크기는 네트워크에서 내적 연결정도(in-degree)를 표현하고 있는데, 최소 크기는 1, 최대 크기는 1,000으로 표현되었다. 노드의 색은 Gephi를 통해 임의로 지정되었으며, 모듈러리티를 통해 구

분된 연구 영역을 표현하고 있다. <표 2>는 모듈러리티를 통해 분류된 하위 네트워크를 순위에 따라 제시하고 있다. 각 분야가 전체 네트워크에서 차지하는 비율과 포함된 논문의 숫자를 나타내었으며, 분야별 논문을 모두 모아 키워드, 연구 분야 정보를 추출하였다. 이후, 분야별로 상위에 속하는 키워드와 연구 분야 정보를 표에 나타내었다(중복 제외). 이를 통해 HRS 주변 연구 분야로는, ‘Recommender System’(14.43%), ‘Health care - Information System’(11.07%), ‘Health care & IoT’(9.29%), ‘Data Mining & Pattern Recognition’(8.27%), ‘Computer Vision’(7.57%), ‘Annotation & Text Mining’(7.17%), ‘Mobile Computing’(5.73%), ‘Ontology & Information Science’(5.41%), ‘Search Engine’(5.38%), ‘Internet Privacy & Authentication’(4.85%) 등이 존재함을 파악하였다. 본 연구에서는 이렇게 분류된 분야를 토대로 세부적 분석을 진행하였다.

다음으로, 본 연구에서는 HRS 연구의 동향을 파악하기 위해 전체 네트워크를 4개년도로 나누



<그림 4> 기간별(기간1 - 기간6) 주변 연구 분야의 비율 순위 변화

어 세부적인 분석을 수행하였다. 분석 결과는 <표 3>에 제시하였으며, 기간별 네트워크의 시각화 자료와 노드 수, 링크 수, 평균 연결 정도, 지름, 평균 거리, 평균 피인용 수, 결집계수 등을 표시하였다. 이 중 시각화 자료는 Gephi의 모듈러리티 분석을 이용하여 상위 10개의 연구 영역을 색으로 구분하고 기간에 맞추어 필터링하였다. 이후, 각 기간의 네트워크를 내적 연결정도에 따라 노드의 최소 크기를 1, 최대 크기를 1,000으로 하여 표현하였다. 또한, 연도별 주변 연구 분야의 비율을 순위에 따라 제시하였다. 이를 통해 시간 변화에 따른 주변 연구 분야의 변화 추이를 파악하려고 하였으며, 이러한 변화 양상을 <그림 4>를 통해 제시하고 있다. 단, HRS 분야의 ‘기간7’ 학술지 정보의 경우, 데이터의 수가 적어 제외하였다. <그림 4>는 ‘기간1’부터 ‘기간6’까지 주변 연구 분야의 비율 순위를 시각화한 것으로, HRS 주변 네트워크에서 어떤 분야의 숫자가 증감하는지 그 추이를 관찰할 수 있다. 전체적으로 보았을 때, 시간이 지남에 따라 노드와 링크의 숫자는 증가하는 경향을 보인다. 반면, 평균 연결정도는 시간에 따라 지속적으로 감소하다가, 2010년을 기점으로 다시 상승하는 모습을 보인다. 전체 분야에서의 평균 피인용 수와 결집계수는 시간이 지남에 따라 계속해서 감소하는 모습을 띤다. 주변 연구 분야를 살펴보면, Recommender System 분야는 8순위로 시작하였지만, 최근에는 1순위로 접어들었고, 반면, Expert System 분야는 2순위에 존재하였지만, 최근에는 10순위로 비교적 낮은 편에 속하고 있다. 마찬가지로, Search Engine 분야도 3순위로 시작하였지만, 9순위로 추락한 양상을 보이고 있음을 확인할 수 있다. 그러나, Mobile Computing 분야와 Wireless & IoT 분야는 각각 10순위, 11순위로 시작하여 최근 6순위, 2

순위로 상승하였으며, Computer Vision과 Text Mining 분야도 각각 6순위에서 4순위로, 9순위에서 5순위로 상승하는 추이를 보였음을 알 수 있다.

HRS 연구 분야와 주변 연구 분야의 키워드 및 주요 학술지 변화 양상을 살펴보면, 먼저 HRS 분야의 경우, 2010년도 이전에는 Information Technology, Decision Support System과 같은 키워드가 등장하다가, 2010년을 기점으로 Recommender System이 키워드로 등장하기 시작했음을 알 수 있다. 이 밖에도, Personalization 및 Privacy와 같은 키워드가 등장하며 개인화 및 정보 보호에 대한 관심이 증가하고 있음을 알 수 있다. 주요 학술지의 경우, 2010년 이전에는 Journal of the American Medical Informatics Association(JAMIA)가 주로 나타났으나, 이후에는 J. Medical System, BMC Med Inf. & Decision Making 등의 저널이 나타나는 것을 확인할 수 있다. 최종적으로, 1순위에 접어든 Recommender System 분야의 경우, 초기에는 Collaborative Filtering과 Information Retrieval이 키워드로 등장하다가, 최근에 들어서는 Information Filtering 및 Deep Learning이 키워드로 등장함을 알 수 있다. 주요 학술지의 경우, Communications of the ACM이나 Intelligent User Interfaces 등의 학술지에서부터 RecSys, Expert Systems with Application, Conference on Information and Knowledge Management 등의 학술지가 나타남을 알 수 있다. 이를 포함하여 나머지 분야에 대한 키워드의 추이를 <표 4>로 제시하였다. 해당 표에 정리된 ‘기간1’에서 ‘기간6’까지의 자료는 기존 모듈러리티로 구분된 자료를 사용하였으며, ‘기간7’의 경우에는 GraphSAGE를 통해 분류된 자료를 이용하여 키워드와 학술지 정보를 추출하였다.

〈표 4〉 HRS 및 주변 연구 분야 4개년도 별 키워드 및 학술지 분석

분야	기간	키워드	학술지
HRS	1	Health Care and Public Health, Software Systems, Information Technology	Journal of the American Medical Informatics Association
	2	Constructive Assessment	Journal of the American Medical Informatics Association
	3	Computerized Medical Records, Decision Support Systems	Journal of the American Medical Informatics Association
	4	Clinical Decision Support	Journal of the American Medical Informatics Association, MedInfo
	5	Medical Information Systems, Recommender System	J. Medical Systems, BMC Med. Inf. & Decision Making
	6	Collaborative Filtering, Medical Information Recommender Systems	J. Medical Systems, RecSys, BMC Med. Inf. & Decision Making
	7	Recommender System, Personalization, Privacy	-
Recommender System	1	World Wide Web, Information Retrieval	Communications of the ACM, User Model User-Adapted Interaction., Scientometrics
	2	Collaborative Filtering, Recommender system, Information retrieval	Intelligent User Interfaces(IUI), Communications of the ACM
	3	Recommender System, Collaborative Filtering, Data Mining	Web Intelligence, Knowledge Discovery and Data Mining, User Modeling
	4	Recommender System, Collaborative Filtering, Social Network	RecSys, Expert Systems with Application, World Wide Web
	5	Recommender System, Collaborative Filtering, Social Network	RecSys, Expert Systems with Application, Conference on Information and Knowledge Management
	6	Recommender System, Collaborative Filtering, Information Filtering	RecSys, UMAP, SIGIR
	7	Recommender System, Collaborative Filtering, Deep Learning	SIGIR, World Wide Web, CIKM
Information System	1	Information Technology, Information System	MIS Quarterly, Commun.ACM, Information Systems Research
	2	Health Care, Information Technology, Information System	JAMIA, MIS Quarterly, Commun.ACM
	3	Health Care, Information System, Information Technology	JAMIA, HICSS, AMIA
	4	Health Care, Information Technology, Information System	AMIA, AMCIS, Decision Support Systems
	5	Health Care, Health Informatic, Medical Information System	J. Medical Systems, BMC Med. Inf. & Decision Making, Computers in Human Behavior
	6	Health Care, Electronic Health Record, eHealth	BMC Med. Inf. & Decision Making, J. Medical Systems, HICSS
	7	Electronic Health Record, mHealth, Health Information Exchange	BMC Med. Inf. & Decision Making, J. Medical Systems, HICSS

분야	기간	키워드	학술지
Wireless & IoT	1	Packet Switching, Computer Networks, Discrete Event Simulation	IEEE Journal on Selected Areas in Communications, SIGCOMM, Communications of the ACM
	2	Knowledge-Based System, Land Mobile Radio, Wireless Network	Mobile Computing and Networking, MobiCom, INFOCOM
	3	Wireless Sensor Network, Ad Hoc Network	SenSys, INFOCOM, IEEE Transactions on Information Technology in Biomedicine
	4	Wireless Sensor Network, Protocol, Patient Monitoring	IJSN, IEEE Transactions on Information Technology in Biomedicine, Pervasive Health
	5	Wireless Sensor Network, Electrocardiography, Body Area Networks	IEEE Transactions on Information Technology in Biomedicine, Enterprise IS, J. Medical Systems
	6	Internet of Things, Cloud Computing, Big Data, Wireless Communication	J. Medical Systems, Wireless Personal Communications, IJDSN
	7	Internet of Things, Cloud Computing, Medical Services, Security, Monitoring	Wireless Personal Communications, J. Medical Systems
Data Mining	1	Pattern Recognition, Machine Learning, Data Mining	Machine Learning, Artificial Intelligence In Medicine, Pattern Recognition
	2	Data Mining, Classification, Neural Network	Artificial Intelligence In Medicine, KDD, Machine Learning
	3	Data Mining, Feature Selection, Neural Network	Artificial Intelligence In Medicine, KDD, Journal of Machine Learning Research
	4	Data Mining, Feature Selection, Support Vector Machine	Expert Syst.Appl., Inf.Sci., Pattern Recognition
	5	Data Mining, Feature Selection, Fuzzy Set Theory	Expert Syst.Appl., J. Medical Systems, Inf.Sci.
	6	Data Mining, Clustering, Classification	Expert Syst.Appl., Inf.Sci., J. Medical Systems
	7	Feature Extraction, Classification, Clustering	J. Medical Systems, Neural Computing and Applications, Multimedica Tools Applications
Computer Vision	1	Image Segmentation, Face Recognition, Computer Vision	IEEE, Pattern Recognition, CVRMed
	2	Face Recognition, Image Segmentation, Medical Images	MICCAI, International Journal of Computer Vision, FG
	3	Image Segmentation, Feature Extraction, Image Registration	NIPS, ISBI, IEEE
	4	Image Segmentation, Feature Extraction, Image Retrieval	CVPR, NIPS, IEEE
	5	Feature Extraction, Image Segmentation, Image Classification	CVPR, ACM Multimedia 2001, ISBI
	6	Feature Extraction, Deep Learning, Feature Extraction	CVPR, MICCAI, Multimedia Tools Appl.
	7	Deep Learning, Image Segmentation, Biomedical Imaging	Multimedia Tools Applications, J. Medical Systems, IEEE Journal of Biomedical and Health Informatics

분야	기간	키워드	학술지
Text Mining	1	Natural Language, Information Extraction, Information Retrieval	JAMIA, Journal of the American Medical Informatics, Computational Linguistics
	2	Natural Language Processing, Unified Medical Language System	JAMIA
	3	Natural Language Processing, Unified Medical Language System, Text Mining	AMIA, Journal of Biomedical Informatics, Bioinformatics
	4	Natural Language Processing, Algorithms, Text Mining	AMIA, BMC Bioinformatics, Journal of Biomedical Informatics
	5	Algorithms, Natural Language Processing, Bioinformatics	Journal of Biomedical Informatics, BMC Bioinformatics, AMIA
	6	Natural Language Processing, Electronic Health Record, Bioinformatics	Journal of Biomedical Informatics, CLEF, BMC Med. Inf. & Decision Making
	7	Natural Language Processing, Electronic Health Record, Deep Learning	BMC Med. Inf. & Decision Making, BMC Bioinformatics
Mobile Computing	1	User Interface, Awareness, Computer Support	CSCW, CHI, Commun.ACM
	2	Augmented Reality, Ubiquitous Computing, Health Care	CHI, Computer-Supported Cooperative Wo가 96, Agents
	3	Ubiquitous Computing, Health Care	CHI, CSCW, Personal and Ubiquitous Computing
	4	Health Care, Everyday Life, Mobile Computing	CHI, PervasiveHealth, UbiComp
	5	Health Care, Mobile Computing, Ubiquitous Computing	CHI, UbiComp, PervasiveHealth
	6	Miscellaneous, Design, Health Care	CHI, CSCW, PervasiveHealth
	7	mHealth, Mental Health, Internet of Things, Sensors	CHI, PervasiveHealth
Ontology	1	Knowledge Representation, Knowledge Base	Commun.ACM, J.ACM, Inf. Process. Manage.
	2	Description Logic, Knowledge Base, Semantic Web	VLDB, IEEE
	3	Semantic Web, Description Logic, Web Service	International Semantic Web Conference, WWW, SIGMOD Conference
	4	Semantic Web, Ontology, Description Logic	ESWC, Journal of Biomedical Informatics, International Semantic Web Conference
	5	Bioinformatics, Biomedical Research, Health Care	J. Biomedical Semantics, Journal of Biomedical Informatics, Expert Syst.Appl.
	6	Ontology, Semantic Web, Bioinformatics	J. Biomedical Semantics, BMC Bioinformatics, Journal of Biomedical Informatics
	7	Ontology, Knowledge Representation, Big Data	OTM, J. Data Semantics, BCB
Search Engine	1	Information Retrieval, Indexation, Query Expansion	SIGIR, Inf. Process. Manage., JASIS
	2	Information Retrieval, WWW, Search Engine	SIGIR, JASIS, Inf. Process. Manage.
	3	Information Retrieval, Query Expansion, Support Vector Machine	SIGIR, TREC, CIKM
	4	Information Retrieval, Data Mining, Search Engine	SIGIR, CIKM, WWW
	5	Information Retrieval, Data Mining, Sentiment Analysis	CIKM, WWW, SIGIR
	6	Sentiment Analysis, Information Retrieval, Twitter	TREC, CSCW, SIGIR
	7	Social Media, Sentiment Analysis, Machine Learning, Deep Learning	SIGIR, CHI, WI

분야	기간	키워드	학술지
Internet Privacy	1	Access Control, Data Privacy, Cryptography	CRYPTO, IEEE, EUROCRYPT
	2	Access Control	IACR Cryptology ePrint Archive, CRYPTO, ASIACRYPT
	3	Data Mining, Access Control, Data Privacy	IACR Cryptology ePrint Archive, ACM, IEEE
	4	Data Privacy, Data Mining, Health Care	IACR Cryptology ePrint Archive, IJNS, ACM
	5	Cloud Computing, Health Care, Data Privacy	J. Medical Systems, IACR Cryptology ePrint Archive, IEEE
	6	Cloud Computing, Privacy, Authentication	J. Medical Systems, Peer-to-Peer Networking and Applications, Journal of Medical Systems
	7	Cloud Computing, Privacy, Blockchain	J. Medical Systems, ICCCS, J. Ambient Intelligence and Humanized Computing

## 4.2 GraphSAGE 분석 결과

본 연구에서는 GraphSAGE를 이용하여 ‘기간 1’부터 ‘기간6’까지의 자료를 학습시킨 뒤, ‘기간 7’ 연구 분야의 분류를 수행하였다. 이를 통해 분류한 ‘기간7’의 자료와 모듈러리티를 통해 분류한 자료를 비교하여 <표 5>로 제시하였다. 그 결과를 살펴보면, 여러 부분에서 비슷한 경향성을 보이지만, GraphSAGE 알고리즘을 통한 분류에서 Text Mining 분야의 순위가 상승한 것을 발견할 수 있다. 이와 함께, 모듈러리티를 통한 분류에서는 볼 수 없었던 Data Visualization 분야가 나타나고 있음을 파악할 수 있으며, Mobile Computing 분야의 순위 또한 상승한 것을 볼 수 있다. 반면, Wireless & IoT, Internet Privacy, Data Mining 분야의 순위는 각각 2단계씩 떨어졌으며, Mathematical Optimization 분야는 다른 분야에 밀려 관찰되지 않았다.

<표 5> ‘기간7’ HRS 주변 연구 분야 분석

순위	모듈러리티 분류	GraphSAGE 분류
1	Recommender System	Recommender System
2	Wireless & IoT	Text Mining
3	Computer Vision	Computer Vision
4	Text Mining	Wireless & IoT
5	Information System	Information System
6	Internet Privacy	Mobile Computing
7	Data Mining	Data Visualization
8	Mobile Computing	Internet Privacy
9	Web Search Query	Data Mining
10	Mathematical Optimization, Decision Making	Web Search Query

## 5. 결론

### 5.1 연구요약

본 연구는 인용 관계를 통해 그래프 네트워크를 구성하고, 이를 통해 HRS 연구 분야와 주변 연구 분야들의 동적인 상호 관계를 기간별로 살펴보는 것에 초점을 맞추었다. 먼저 전체 네트워크를

살펴보면, HRS 연구는 ‘Recommender System’과 ‘Information System’ 분야와 가장 크게 연관된 것을 확인할 수 있었다. 뿐만 아니라, ‘Wireless & IoT’, ‘Data Mining’, ‘Computer Vision’, ‘Text Mining’ 분야와도 크게 밀접한 모습을 보이고 있으며, ‘Mobile Computing’이나 ‘Internet Privacy’와 같은 분야와도 연결이 관찰되었다. 이를 통해, HRS 연구는 단순하게 최신의 추천시스템 알고리즘을 받아들여 적용하는 것에 그치지 않고, 건강 정보를 수집하는 스마트기기나, 이렇게 수집된 비정형 데이터를 가공하는 분야까지 폭넓게 관련되어 있는 것으로 파악된다. 더불어 추천시스템의 거대한 성공 이전부터 정보과학 분야에서 큰 부분을 차지하고 있던 Information Systems 분야를 괄시하지 않고 살펴볼 필요성이 요구된다. 이는 HRS 이전에 존재하였던 Medical Expert System이나, Clinical Decision Support System과 같은 정보과학이 응용되었던 분야를 살펴봄으로써, 미래의 HRS 연구를 위한 아이디어를 찾을 수 있음을 시사한다. 마지막으로, HRS 연구자들에게는 고객의 정보를 안전하게 관리하는 분야와 추천하고자 하는 정보를 어떤 방식으로 사용자에게 전달할지에 대한 분야에 대해서도 고민할 필요성이 있을 것으로 판단된다.

좀 더 구체적으로, 시간에 따른 변화 추이를 살펴보면, HRS 연구의 경우, ‘기간1’에는 Software System과 Information Technology가, ‘기간3’에는 Computerized Medical Record와 Decision Support System이, ‘기간4’에는 Clinical Support System이 각각 주요 키워드로 관찰되었다. 이후 ‘기간5’부터 Recommender System이 키워드로 관찰되기 시작하였으며, ‘기간7’에 이르러서는 Personalization과 Privacy가 새로운 키워드로 떠오른 것을 확인할 수 있다. 이를 통해, HRS 연구의 본격적인 시

발점을 ‘기간5’에 해당하는 2010년 전후로 파악할 수 있다. 초창기의 HRS 연구는 개인에게 필요한 건강정보를 필터링하는 Medical Information Recommender System인 것으로 파악되며, 현재에 이르러서는 개인맞춤형 의료시스템과 이를 제공하기 위한 개인 정보관리 분야에도 초점을 맞추고 있음을 파악할 수 있다.

한편, HRS 연구자가 살펴봐야 할 연구 분야로는, Recommender System, Wireless & IoT, Computer Vision, Text Mining 분야로 파악할 수 있다. 이들 분야는 ‘기간1’에 비해 ‘기간6’에서 HRS 주변 연구 분야에서 차지하는 비율이 가파르게 상승한 분야이며, 이들 분야는 양질의 건강 추천을 제공하기 위한 필수적인 기술로 파악할 수 있다. 다양한 연구자들에 의해서 강조되었듯, 올바른 추천을 위해서는 정확한 신체 정보가 필요하며 이를 수집하기 위한 필수적인 기술로 다양한 스마트기기가 언급되고 있다. 따라서, 양질의 HRS 개발을 위해서는 현재까지도 발전하는 추천시스템의 알고리즘뿐만 아니라, 스마트기기를 이용한 양질의 건강 데이터 수집 기술에도 집중해야 할 필요성이 있다고 볼 수 있다. 더불어, 이렇게 수집된 데이터를 가공하는 Computer Vision이나, Text Mining 기술 등 정량화하기 어려운 건강 정보를 올바르게 가공하는 역량 또한 요구된다고 볼 수 있다.

마지막으로, 해당 분야들의 기간별 키워드 변동을 살펴보면, Recommender System 분야는 초기의 World Wide Web과 Information Retrieval에서 Data Mining, Social Network, Information Filtering, Deep Learning의 순서로 변화가 나타난 반면, Wireless & IoT 분야는 초기에는 Packet Switching, Land Mobile Radio, Ad-Hoc Network 등과 같은 키워드가 나타났으나, 이후로 갈수록

Internet of Things, Cloud Computing, Big Data 등과 같은 키워드가 출현함을 확인할 수 있다. Computer Vision 분야와 Text Mining 분야에서도 최근 Deep Learning, Biomedical Imaging, Bioinformatics와 같은 키워드가 등장함을 확인할 수 있다. 이를 통해, 현재의 기술적 동향은 Deep Learning으로 빠르게 이동하고 있으며, Recommender System, Computer Vision, Text Mining 등과 같은 다양한 분야에 응용되고 있음을 알 수 있다. 따라서, HRS 분야의 연구자들은 여러 분야의 Deep Learning 알고리즘 발전에 민감하게 반응할 필요가 있으며, 발전하는 알고리즘을 빠르게 응용하는 것이 효과적인 HRS 개발에 있어서 주요한 열쇠가 될 것으로 판단된다.

## 5.2 시사점

본 연구의 시사점은 이론적 시사점, 실무적 시사점으로 나누어 살펴볼 수 있다. 우선 이론적 시사점으로는, 첫째, 본 연구는 기존에 수행되었던 HRS 관련 문헌연구와는 다르게, 인용네트워크 분석을 활용하여 연구를 진행함으로써, HRS 연구 네트워크를 둘러싼 전체적인 영향 관계를 조망할 수 있었다. 둘째, 본 연구는 HRS 주변 연구 분야를 4개년 단위로 나누어 키워드와 저널 정보를 살펴봄으로써, 주변의 동향을 전체적으로 살펴볼 수 있었다. 마지막으로, 본 연구는 GraphSAGE 알고리즘을 추가적으로 적용하여, 인용 관계가 약하게 나타나는 기간에서의 분류를 보다 정확하게 수행하고, 이를 통해 정확한 동향 정보를 제공했다는 점에서 의의를 찾을 수 있다.

한편, 본 연구의 실무적인 시사점은 크게 HRS 연구자와 건강 분야 종사자의 입장으로 나누어 살펴볼 수 있다. 우선, HRS 연구자 관점에서는

본 연구의 결과를 통해 학제적인 성격의 HRS 연구가 어떤 분야와 밀접한 연관이 있는지를 파악할 수 있으며, 이와 더불어 시간에 따른 키워드의 변화를 파악하여 연구의 전체적인 동향이 어떤 방향으로 나아가는지를 확인할 수 있었다. 이를 통해, 연구자들은 주변 기술의 발전을 전체적으로 알아볼 수 있고, 후속 HRS 연구의 방향성 또한 설계할 수 있게 된다. 즉, 효과적인 연구를 위해 단순히 HRS 연구 분야에만 집중할 것이 아니라, 이와 함께 발전하는 스마트기기, 비정형 데이터 처리, 추천시스템 알고리즘 등 다른 기술 관련 연구들을 함께 살펴봄으로써, 수준 높은 연구 성과를 창출하기 위한 아이디어를 얻을 수 있을 것으로 기대한다.

더불어, 건강 분야 종사자의 입장에서는 효과적인 건강서비스를 제공하기 위해 앞으로 필요한 실용적인 기술이 무엇인지를 본 연구를 통해 파악할 수 있다. 만약 HRS 연구 성과를 실제 서비스에 제공하고자 한다면, 이를 위해 필요한 기술들이 무엇이 있으며, 현재 어떠한 기술들이 발전하고 있는지를 체계적으로 파악할 수 있을 것이다. 이를 통해, 지금까지 제공되었던 서비스보다 더욱 효과적인 서비스를 제공할 수 있는 기반을 마련할 수 있을 것이며, 급변하는 의료서비스 환경에서 살아남기 위한 방향성을 파악할 수 있을 것으로 기대한다.

## 5.3 한계점 및 향후 연구방향

먼저 본 연구의 한계점은 다음과 같다. 첫째, 본 연구는 AMiner-DBLP V.13 Dataset에서 HRS와 연관성이 큰 것으로 파악되는 대략 10만 개의 논문을 통해 인용네트워크를 구성하고, HRS의 학제적인 구조와 동향을 관찰하고자 하였으나,

ACM, Google Scholar, PubMed 등의 대형 서지학 사이트의 데이터가 포함되어 있지 않기 때문에, 본 연구의 결과가 모든 HRS 연구 구조를 설명할 수 없다는 한계가 존재한다. 둘째, 본 연구는 기존 체계적 고찰 방법론을 따라 논문의 제목에 특정 키워드가 포함된 논문을 우선 선별하여 분석에 이용하였다. 그러나, 이 과정에서 키워드, 연구 분야, 저널명, 초록 등의 정보를 활용하지 못하였고, 이에 따라 HRS 관련 연구에 해당함에도 불구하고 제외되었을 가능성이 있는 논문들이 존재한다. 이처럼 연구자의 주관이 개입될 여지가 일부 있었음에도, 엄격한 전처리 과정을 통해 이를 완벽히 제거하지 못했다는 점을 한계점으로 들 수 있다. 마지막으로, 본 연구는 GraphSAGE와 같은 최신 알고리즘을 적용하여 분석의 정확도 향상을 시도하였다는 점에서 의의가 있다. 그러나, 데이터의 크기 문제로 미가공 데이터(raw data)를 이용하여 학습을 진행한 것이 아니라, 연구자의 판단하에 가공된 데이터 일부에 해당 알고리즘을 적용하였다는 점이 한계로 지적된다. 따라서, 후속 연구에서는 데이터의 수집과 전처리, 알고리즘의 적용에 있어서 더욱 정확하고 엄격한 접근이 요구되며, 이를 통해 향후 발전 가능성이 큰 HRS 분야에 대한 수준 높은 연구들이 지속해서 수행될 수 있기를 기대한다.

## 참고문헌(References)

### [국내 문헌]

- 김은경, 정연식, 양선욱, 권지윤, 양성병. (2022). MIS Quarterly 연구동향 탐색: 토픽모델링 및 키워드 네트워크 분석 활용. *지능정보연구*, 28(2), 207-235.
- 김기태. (2022). 한국과 유럽 8개국의 가구 의료비 지출 부담. *보건복지포럼*, 18 - 33.
- 김현정, 안형식. (2015). 체계적 고찰/메타분석 연구 비판적으로 바라보기. *The Korean Journal of Helicobacter and Upper Gastrointestinal Research*, 15(2), 73-79.
- 김형철, 홍길중. (2022). 강원도 디지털 헬스케어 산업 경쟁력 분석과 시사점. *The E-Business Studies*, 23(5), 177 - 191.
- 문장섭. (2021). 정밀의료 시대의 뇌전증 관리. *에필리아: 뇌전증과사회*, 3(2), 74-77.
- 박은희. (2021). 텍스트 네트워크 분석을 활용한 대학교양영어 연구주제 분석. *Multimedia-Assisted Language Learning*, 24(4), 214-239.
- 박준형, 광기영. (2013). 특허 인용 관계가 기업 성과에 미치는 영향. *지능정보연구*, 19(3), 127-139.
- 배영임, 신혜리. (2020). 코로나 19, 언택트 사회를 가속화하다. *이슈&진단*, 1-26.
- 손진아. (2020). 제4차 산업혁명 시대에 있어서 문화 산업의 발전조건. *한국자치행정학보*, 34(1), 19-35.
- 오의금. (2020). 포스트코로나 뉴노멀 시대의 간호의 전망과 대책. *Korean Journal of Adult Nursing*, 32(3), 221-222.
- 이동호, 김우수, 문성주, 정이상. (2012). 메타 데이터와 계량서지 데이터를 이용한 경영사학 연구특성 분석. *경영사연구(경영사학)*, 27, 97-117.
- 이현복, 이수연. (2019). 노인의료비 추계와 건강 보험에 대한 시사점. *보험학회지*, 117, 43-68.
- 장요한. (2022). 목절별 통행량으로 보는 지역별 특색 읽기 1: 통계청 SDC 통신모바일 인구 이동량 통계데이터를 중심으로. *국토*, 72-77.
- 정유경. (2020). 국내 문헌정보학 분야 학술지의

- 인용네트워크분석. *한국문헌정보학회지*, 54(4), 221-238.
- 정이태, 안현철. (2020). 이종 네트워크 구조를 활용한 HinSAGE 기반 추천시스템. *KMIS International Conference*, 294-295.
- 정중희, 김종우. (2013). UCC 추천을 위한 사회 네트워크 분석을 활용한 협업 필터링 기법 연구. *한국정보기술학회논문지*, 11(1), 185-195.
- 정혜실. (2014). 헬스케어 웨어러블 디바이스의 동향과 전망. *KHIDI Brief*, 115, 3-20.
- 조인동, 김남규. (2011). 소셜 네트워크와 데이터 마이닝기법을 활용한 학문 분야 중심 및 융합 키워드추천. *지능정보연구*, 17(1), 127-138.
- 채수미. (2020). 코로나바이러스감염증-19와 미래 질병 대응을 위한 과제. *보건복지 Issue & Focus*, 374, 1-8.
- 최건영, 홍참길. (2022). 약물 유인성 간염 예측을 위한 그래프 신경망 앙상블. *한국정보과학회 학술발표논문집*, 1244-1246.
- A., & Holzinger, A. (2016). Recommender systems for health informatics: State-of-the-art and future perspectives. *Machine Learning for Health Informatics: State-of-the-Art and Future Challenges*, 391-414.
- Cappella, J. N., Yang, S., & Lee, S. (2015). Constructing recommendation systems for effective health messages using content, collaborative, and hybrid algorithms. *The ANNALS of the American Academy of Political and Social Science*, 659(1), 290-306.
- Celik, E. D., & Elci, A. (2020). A survey on semanticized and personalized health recommender systems. *Expert Systems*, 37(4), e12519.
- Cheung, K. L., Durusu, D., Sui, X., & de Vries, H. (2019). How recommender systems could support and enhance computer-tailored digital health programs: A scoping review. *Digital Health*, 5, 1-19.
- Choi, J., Yi, S., & Lee, K. C. (2011). Analysis of keyword networks in MIS research and implications for predicting knowledge evolution. *Information & Management*, 48(8), 371-381.
- De Croon, R., Van Houdt, L., Htun, N. N., Štiglic, G., Vanden Abeele, V., & Verbert, K. (2021). Health recommender systems: Systematic review. *Journal of Medical Internet Research*, 23(6), e18035.
- Ding, Y., Chowdhury, G. G., & Foo, S. (2000). Journal as markers of intellectual space: Journal co-citation analysis of information retrieval area, 1987~1997, *Scientometrics*, 47(1), 55-73.
- Etemadi, M., Abkenar, S. B., Ahmadzadeh, A., Kashani, M. H., Asghari, P., Akbari, M., & Mahdipour, E. (2023). A systematic review of healthcare recommender systems: Open issues, challenges, and techniques. *Expert*

#### [국외 문헌]

- Bassecoulard, E., & Zitt, M. (1999). Indicators in a research institute: A multi-level classification of scientific journals, *Scientometrics*, 44(3), 323-345.
- Borgman, C. L., & Rice, R. E. (1992). The convergence of information science and communication: A bibliometric analysis, *Journal of the American Society for Information Science*, 43(6), 397-411.
- Calero Medina, C. M., & van Leeuwen, T. N. (2012). Seed journal citation network maps: A method based on network theory, *Journal of the American Society for Information Science and Technology*, 63(6), 1226-1234.
- Calero Valdez, A., Ziefle, M., Verbert, K., Felfernig,

- Systems With Applications*, 231, 118823.
- Ferretto, L. R., Cervi, C. R., & de Marchi, A. C. B. (2017). Recommender systems in mobile apps for health a systematic review. In *Proceedings of the 12th Iberian Conference on Information Systems and Technologies (CISTI)*, 1-6. IEEE.
- Goyal, P., & Ferrara, E. (2018). Graph embedding techniques, applications, and performance: A survey. *Knowledge-Based Systems*, 141, 78-94.
- Hamilton, W. L., Ying, R., & Leskovec, J. (2017). Inductive representation learning on large graphs. In *Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017)*, 1-11.
- Hors-Fraile, S., Rivera-Romero, O., Schneider, F., Fernandez-Luque, L., Luna-Perejon, F., Civit-Balcells, A., & de Vries, H. (2018). Analyzing recommender systems for health promotion using a multidisciplinary taxonomy: A scoping review. *International Journal of Medical Informatics*, 114, 143-155.
- Kamran, M., & Javed, A. (2015). A survey of recommender systems and their application in healthcare. *Technical Journal (University of Engineering and Technology)*, 20(4), 111-119.
- Kim, J. A., Yoon, S., Kim, L. Y., & Kim, D. S. (2017). Towards actualizing the value potential of Korea Health Insurance Review and Assessment (HIRA) data as a resource for health research: Strengths, limitations, applications, and strategies for optimal use of HIRA data. *Journal of Korean Medical Science*, 32(5), 718-728.
- Lee, H. Y. (2015). Uncovering the multidisciplinary nature of technology management: Journal citation network analysis using the relatedness index. *Scientometrics*, 102(1), 51-75.
- Leppink, J., & Perez-Fuster, P. (2019). Social networks as an approach to systematic review. *Health Professions Education*, 5(3), 218-224.
- Leydesdorff, L. (1994). The generation of aggregated journal citation maps on the basis of the CDROM version of the Science Citation Index. *Scientometrics*, 31(1), 59-84.
- Leydesdorff, L. (2007). Betweenness centrality as an indicator of the interdisciplinarity of scientific journals. *Journal of the American Society for Information Science and Technology*, 58(9), 1303-1319.
- Liu, J., Ren, J., Zheng, W., Chi, L., Lee, I., & Xia, F. (2020). Web of scholars: A scholar knowledge graph. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2153-2156.
- Mohamed, H. A., George, O. O., & Ronald, W. M. (2017). A survey of context-aware healthcare recommender systems. *International Journal of Science and Research*, 6(2), 962-967.
- National Center for Biotechnology Information (NCBI). (2023). PubMed. MEDLINE and PubMed Central, February 3, 2023, from <https://pubmed.ncbi.nlm.nih.gov/>
- Noyons, E. C. M. (2001). Bibliometric mapping of science in a policy context. *Scientometrics*, 50(1), 83-98.
- Otte, E., & Rousseau, R. (2002). Social network analysis: A powerful strategy, also for the information sciences. *Journal of Information Science*, 28(6), 441-453.
- Perozzi, B., Ai-Rfou, R., & Skiena, S. (2014). Deepwalk: Online learning of social representations. In

- Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 701-710.
- Pham, C., & Dang, T. (2021). Link prediction for biomedical network. In *Proceedings of the 12th International Conference on Advances in Information Technology*, 1-5.
- Pincay, J., Teran, L., & Portmann, E. (2019). Health recommender systems: A state-of-the-art review. In *Proceedings of the 6th International Conference on eDemocracy & eGovernment (ICEDEG)*, 47-55. IEEE.
- Rafols, I., & Meyer, M. (2010). Diversity and network coherence as indicators of interdisciplinarity: Case studies in bionanoscience, *Scientometrics*, 82(2), 263-287.
- Ricci, F., Rokach, L., Shapira, B., & Kantor, P. B. (2011). Recommender systems Handbook. In F. Ricci, L. Rokach, B. Shapira, & P. B. Kantor. *Recommender systems handbook*. Springer US.
- Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., & Monfardini, G. (2009). The graph neural network model. *IEEE Transactions on Neural Networks*, 20(1), 61-80.
- Schafer, H., Hors-Fraile, S., Karumur, R. P., Calero Valdez, A., Said, A., Torkamaan, H., Ulmer, T., & Trattner, C. (2017). Towards health (AWARE) recommender systems. In *Proceedings of the 2017 International Conference on Digital Health*, 157-161.
- Scheuren, F. (2004). *What is a survey*. Alexandria: American Statistical Association.
- Sezgin, E., & Ozkan, S. (2013). A systematic literature review on Health Recommender Systems. In *Proceedings of the 2013 E-Health and Bioengineering Conference (EHB)*, 1-4. IEEE.
- Sharma, D., Singh Aujla, G., & Bajaj, R. (2019). Evolution from ancient medication to human-centered Healthcare 4.0: A review on health care recommender systems. *International Journal of Communication Systems*, e4058.
- Sterling, J. A., & Montemore, M. M. (2022). Combining citation network information and text similarity for research article recommender systems. *IEEE Access*, 10, 16-23.
- Su, J., Guan, Y., Li, Y., Chen, W., Lv, H., & Yan, Y. (2020). Do recommender systems function in the health domain: A system review. *arXiv e-prints*, arXiv:2007.13058.
- Tran, T. N. T., Felfernig, A., Trantner, C., & Holzinger, A. (2021). Recommender systems in the healthcare domain: State-of-the-art and research issues. *Journal of Intelligent Information Systems*, 57(1), 171-201.
- Wang, J. (2013). Citation time window choice for research impact evaluation. *Scientometrics*, 94(3), 851-872.
- Wang, H., Zhao, M., Xie, X., Li, W., & Guo, M. (2019). Knowledge graph convolutional networks for recommender systems. In *Proceedings of the World Wide Web Conference*, 3307-3313.
- White, H. D., & McCain, K. W. (1997). Visualization of literatures. *Annual Review of Information Science and Technology*, 32, 99-168.
- Wiesner, M., & Pfeifer, D. (2014). Health recommender systems: Concepts, requirements, technical basics and challenges. *International Journal of Environmental Research and Public Health*, 11(3), 2580 - 2607.
- Wu, S., Sun, F., Zhang, W., Xie, X., & Cui, B. (2022). Graph neural networks in recommender systems: A survey. *ACM Computing Surveys*, 55(5), 1-37.

- Xiao, L., Wu, X., & Wang, G. (2019). Social network analysis based on graph. In *Proceedings of the 12th International Symposium on Computational Intelligence and Design, ISCID 2019*, 196-199.
- Yue, W., Wang, Z., Zhang, J., & Liu, X. (2021). An overview of recommendation techniques and their applications in healthcare. *IEEE/CAA Journal of Automatica Sinica*, 8(4), 701-717.
- Zhou, P., & Leydesdorff, L. (2007). The citation impacts and citation environments of Chinese journals in mathematics, *Scientometrics*, 72(2), 185-200.

Abstract

## Research Trends of Health Recommender Systems (HRS): Applying Citation Network Analysis and GraphSAGE

Haryeom Jang\* · Jeesoou You\* · Sung-Byung Yang\*\*

With the development of information and communications technology (ICT) and big data technology, anyone can easily obtain and utilize vast amounts of data through the Internet. Therefore, the capability of selecting high-quality data from a large amount of information is becoming more important than the capability of just collecting them. This trend continues in academia; literature reviews, such as systematic and non-systematic reviews, have been conducted in various research fields to construct a healthy knowledge structure by selecting high-quality research from accumulated research materials. Meanwhile, after the COVID-19 pandemic, remote healthcare services, which have not been agreed upon, are allowed to a limited extent, and new healthcare services such as health recommender systems (HRS) equipped with artificial intelligence (AI) and big data technologies are in the spotlight. Although, in practice, HRS are considered one of the most important technologies to lead the future healthcare industry, literature review on HRS is relatively rare compared to other fields. In addition, although HRS are fields of convergence with a strong interdisciplinary nature, prior literature review studies have mainly applied either systematic or non-systematic review methods; hence, there are limitations in analyzing interactions or dynamic relationships with other research fields. Therefore, in this study, the overall network structure of HRS and surrounding research fields were identified using citation network analysis (CNA). Additionally, in this process, in order to address the problem that the latest papers are underestimated in their citation relationships, the GraphSAGE algorithm was applied. As a result, this study identified ‘recommender system’, ‘wireless & IoT’, ‘computer vision’, and ‘text mining’ as increasingly important research fields related to HRS research, and confirmed that ‘personalization’ and ‘privacy’ are emerging issues in HRS research. The study findings would provide both academic and practical insights into identifying the

---

\* Department of Big Data Analytics, Kyung Hee Univeristy

\*\* Corresponding author: Sung-Byung Yang

Department of Business Administration/Big Data Analytics, Kyung Hee University  
26 Kyungheedaero, Dongdaemun-gu, Seoul 02447, Republic of Korea  
Tel: +82-2-961-9548, E-mail: sbyang@khu.ac.kr

structure of the HRS research community, examining related research trends, and designing future HRS research directions.

**Key Words** : Health Recommender System, Healthcare, Trend Analysis, Citation Network Analysis, GraphSAGE

Received : February 16, 2023 Revised : March 4, 2023 Accepted : March 5, 2023

Corresponding Author : Sung-Byung Yang

## 저 자 소개



**장하림**

경희대학교 일반대학원 빅데이터응용학과 석사과정에 재학중이며, 주요 관심분야는 머신러닝, 딥러닝, 의료데이터 분석 등이다.



**유지수**

경희대학교 일반대학원 빅데이터응용학과 석사과정에 재학중이며, 주요 관심분야는 금융데이터 분석, 머신러닝, 인공지능 등이다.



**양성병**

경희대학교 경영학과/빅데이터응용학과 교수로 재직 중이며, 주요 관심분야는 빅데이터 분석, 온라인 리뷰, 고객관계관리, 지식경영, 온라인 커뮤니티, 전자상거래, 스마트 투어리즘 등이다. MIS Quarterly, Information Systems Research, Computers in Human Behavior, Tourism Management, International Journal of Information Management 등 우수 저널에 다수의 논문을 게재하였다.