

마이데이터 환경에서 개인의 전자 건강/의료 데이터 활용을 위한 데이터 거래모델

오현택*, 양진홍**

Personal Health Record/Electronic Medical Record Data Trading Model for Medical My Data Environments

Hyeon-Taek Oh*, Jin-Hong Yang**

요약 개인정보를 포함한 다양한 데이터를 활용하기 위해서는 정보주체의 권리를 보장하며 데이터 활용을 할 수 있는 모델이 필요하다. 정보주체를 고려하는 개인정보 활용 패러다임의 변화는 마이데이터를 기반으로 하는 다양한 데이터 활용 모델을 만들어내고 있으며, 다양한 개인정보가 생겨나고 다뤄지는 의료분야에서도 이러한 움직임이 진행되고 있다. 이번 논문에서는 마이데이터 기반 개인 건강/의료 데이터 활용 생태계가 확산되었을 때, 생겨날 수 있는 이해관계자들 간의 데이터 거래모델을 제안하였고 다양한 실험 결과를 기반으로 한 수학적 모델링을 통해 정보주체와 데이터 활용자 모두의 요구사항을 만족할 수 있는 개인 건강/의료 데이터 거래모델이 성립할 수 있음을 보였다.

Abstract Today, data subjects should be considered to utilize various personal data. To support this paradigm, the concept of “My Data” has proposed and has realized in various industrial sectors, including medial sectors. Based on the concept of the medical My Data, this paper proposes a personal health record (PHR) and an electronic medical record (EMR) data trading model. Particularly, this paper proposes a system model to support the medical My Data environment and relevant procedure among stakeholders for PHR/EMR data trading that ensures the rights of data subjects. Based on the proposed system model, this paper also proposes various mathematical models to analyze the behavior of stakeholders and shows the feasibility of the proposed data trading model that satisfies the requirements of both data subjects and data consumers.

Key Words : Data Subjects, Data Trading, Electronic Medical Record, Medical My Data, Personal Health Record

1. 서론

4차 산업혁명 시대가 열리게 되면서, 다양한 환경에서 생성되는 데이터를 활용하는 것은 이제 온라인/오프라인 비즈니스 환경에서 필수 불가결한 요소가 되었다 [1]. 특히, 서비스를 제공하고 사용자들의 서비스 사

용 이력 등의 정보를 수집하고 가공하여 다시 서드 파티(Third party)들에게 제공하거나 판매하는 데이터 브로커 (Data broker)들은 우리나라를 포함하여 미국, 유럽 등 전 세계적으로 출현하였다. 대표적인 예시로는 사회 관계망 서비스(Social Network Service;

This Paper was supported by Institute of Information & communications Technology Planning & Evaluation & grant funded by the Korea government (MSIT) (No: 2018-0-00261, GDPR Compliant Personal Identifiable Information Management Technology for IoT Environment) in 2020.

*Intelligent Tech Team, KAIST Institute for IT Convergence(oht6423@kaist.ac.kr)

**Corresponding Author : Department of Healthcare IT, INJE University (jinhong@inje.ac.kr)

Received June 15, 2020

Revised June 16, 2020

Accepted June 24, 2020

SNS)들의 발전과 함께 활성화된 온라인 광고시장 환경이 있다. SNS의 보급으로 개인정보가 수집할 수 있는 환경이 점점 확대되면서, 데이터 브로커들은 수집된 개인정보(혹은 개인 데이터)를 자신들의 이익을 창출하기 위해 활용하였으며, 개인정보의 활용에 관한 문제점들이 밝혀지게 되었다.

구체적인 데이터 브로커들의 개인정보 활용에 관한 실태는 다양한 보고서들을 통해 밝혀졌다 [2-4]. 데이터 브로커들이 개인정보의 대상이 되는 정보주체의 의지와는 상관없이 다양한 비즈니스에서 무분별하게 개인정보를 활용하는 것으로 나타나면서 개인정보 활용에 대한 정보주체의 주권을 보장하는 움직임이 나타나게 되었다 [5]. 가장 대표적인 예시는 전 세계에서 처음으로 데이터 주권을 명시적으로 보호한 유럽의 일반 개인정보보호 규정(General Data Protection Regulation: GDPR)이다 [6]. GDPR을 시작으로 각국에서는 정보주체의 권리 (예: 개인정보 활용의 명시적 동의, 개인정보 열람권, 개인정보 삭제권 등)을 고려하게 되었다. 이에, 현재 개인정보 생태계는 크게 개인정보를 보호하는 방식과 개인정보를 거래하는 방식으로 나아가고 있으며, 데이터의 잠재적인 가치가 높아지는 상황에서 많은 국가에서는 개인정보의 보호와 활용을 동시에 취할 수 있는 방향으로 개인정보 관련 법/규제가 변경되고 있다.

대한민국 또한 세계적인 추세를 반영하여 정보주체의 데이터 주권을 확보하면서 데이터 경제를 활성화하기 위해 변화하고 있다. 특히, 2020년 1월 데이터 3법(개인정보보호법, 정보통신망법, 신용정보법) 개정안 [7]이 통과되는 것을 계기로, 개인정보에 관한 정보주체의 권리를 보장해주면서 개인정보를 더욱 활용할 수 있는 통로가 생기게 되었다. 이러한 배경에서 생겨난 “마이데이터” 개념은 개인정보의 주체인 개인이 자신의 개인정보 생명주기(생성, 관리, 활용, 파기 등)를 직접 관리하면서, 필요할 때 자신의 동의하에 제삼자에게 자신의 데이터를 제공하는 개념으로, 의료·금융·공공·물류 등 다양한 분야에서 각 분야의 데이터 활용 활성화를 위해 적용하려는 움직임이 있다 [1].

의료분야의 경우, 건강/의료 데이터의 정보주체는 진료를 받은 개인이지만, 현재의 의료체계에서는 자신

의 건강/의료 데이터에 대한 권리를 주장할 방법이 부족한 상황이다. 그래서 실제로 각 의료기관에서 생성된 개인 건강/의료 데이터는 개인이 원하는 상황에 바로 활용되지 못하는 것이 현실이며, 각 의료기관의 정책에 따라 개인의 건강/의료 데이터가 제대로 활용되지 못한 채 폐기되는 것이 현실이다. 이렇게 폐기되는 데이터들을 활용하지 못하는 것은 의료비를 부담하는 개인, 건강보험공단, 보험사 등 다양한 이해관계자들에게 추가적인 비용 등의 형태로 부담이 되고 있다.

의료분야의 마이데이터 사업들은 환자가 여러 병원에서 진료받은 진료기록이나 자신의 건강기록 정보를 자신이 관리할 수 있는 건강/의료 데이터 관리 플랫폼으로 내려받고, 개인의 동의에 따라 다른 기관들이 개인 건강/의료 데이터를 활용할 수 있는 형태로 진행하고 있다.

이러한 배경을 바탕으로, 이번 논문에서는 마이데이터 환경에서 개인의 전자 건강/의료 기록 정보 활용을 위한 데이터 거래모델을 제안한다. 특히, 개인 건강/의료 데이터를 사용자 동의에 기반하여 저장하고 활용할 때, 해당 정보를 활용하고자 하는 활용처에서 관련 비용을 포함하여 부담하고, 개인정보 활용에 동의한 정보주체에게 유·무형의 보상을 제공하는 모델을 통해 정보주체의 권리를 보장하면서 개인 건강/의료 데이터를 활용할 수 있는 것을 보인다.

2. 선행연구

이번 장에서는 마이데이터 개념의 기술적 바탕이 되는 정보주체들을 위한 개인정보 저장소의 개념과 개인 정보와 관련된 정보주체의 데이터 공유/판매 의지 및 이를 기반으로 한 데이터 거래모델들의 선행연구를 살펴본다.

2.1 개인정보 저장소

마이데이터 환경에서 개인정보 제공자의 권리 강화를 위해서 기술적으로 제안된 개념이 개인정보 저장소(Personal Data Store; PDS)[8]이다. 개인정보 저장소 서비스를 통해 정보주체는 안전하고 체계적인 방식으로 주요한 개인정보를 저장, 관리 및 배포할 수 있

다. 이때, 정보주체는 목적에 맞는 개인정보 저장소에 자신의 개인정보를 저장하게 된다. 개인정보를 저장할 때, 데이터 제공자의 권리를 보호하고 정보주체의 데이터 통제권 행사를 도와주는 다양한 방식이 적용될 수 있다. 개인정보 저장소는 정보주체의 모든 데이터에 대한 접근 포인트 역할을 하게 되며, 개인정보에 관한 모든 요청은 개인정보 저장소를 통해 이루어지게 된다. 이러한 구조를 취함으로써, 개인들은 자신의 개인정보가 활용되려는 시점에 자신의 의사에 따라 해당 데이터의 공유 여부를 결정할 수 있으며, 자신의 개인정보가 어떻게 활용되고 관리되고 있는지 쉽게 파악할 수 있는 장점이 있다. 이러한 개념을 플랫폼화한 것 중 한 가지 방식이 마이데이터 플랫폼으로 볼 수 있다.

2.2 정보주체의 개인정보 공유/판매 의지

건강/의료 데이터를 공유하거나 판매하는 것에 대한 정보주체들의 인식에 대하여 알아보기 전에, 이번 절에서는 먼저 일반적인 개인정보를 공유하거나 판매하는 것에 대한 정보주체의 인식을 알아본다.

정보주체가 자신의 개인정보를 자신의 이익에 견주어 제공하거나 판매하는 것을 ‘개인정보 제공 의지 혹은 판매 의지 (willing to -share/-sell personal data)’라고 하며, 의료분야를 포함한 다양한 분야에서 연구되고 있다 [9]. 설문 조사마다 절대적인 수치는 조금씩 다르지만 (특히, 정보주체의 특성 (연령, 수입, 성별, 국적 등)에 따라 다른 분포를 보임), 자신의 개인정보가 활용되는 것을 신경 쓰지 않는 그룹 (프라이버시 무관심자; Privacy Unconcerned)과 자신의 이익과 개인정보 제공에 따른 손익을 계산하는 그룹 (프라이버시 실용주의자; Privacy Pragmatist)의 합이 약 70~80% [9] 정도로 나타나며, 이익의 양과 상관없이 개인정보를 제공하지 않겠다는 그룹 (프라이버시 근본주의자; Privacy Fundamentalist)도 약 20~30% [9] 정도로 되는 것으로 나타났다. 의료분야에서도 연구목적이나 추가적인 건강/의료 서비스 등을 제공하는 가정하에, 다양한 특성(예: 성별, 인종, 연령, 연 수입, 중병 정도, 교육수준 등)을 가진 참여자들의 개인 건강/의료 데이터의 제공 의지가 높아지는지 낮아지는지에 관한 통계분석이 많이 이루어지고 있다 [10,11].

개인정보 활용에 동의하는 그룹들도 공유/판매되는 개인정보의 종류에 따라 그 가치를 다르게 생각하게 된다. 정보주체의 입장에서 개인정보의 가치는 개인정보가 제공되었을 때 얼마나 자신의 프라이버시와 직접 연결되는지 여부 (프라이버시 민감도)와 관련이 있는데, 예를 들어, 정보주체 입장에서는 웨어러블 디바이스 같은 것으로 수집된 심박수 등과 같은 건강정보보다는 의료 서비스를 통해 생성된 의료정보가 더 프라이버시 민감도가 높으며 그에 따라 가치를 더 높게 받아야 한다고 생각하는 것이다 [12]. 이러한 경향성은 개인정보 활용 동의를 얻는 과정에서 프라이버시 민감도가 높은 데이터를 취득하게 될 때, 정보주체가 수용할 수 있는 더 좋은 서비스나 보상을 제공해야 하는 것을 의미한다.

2.3 개인정보 거래 모델

일반적인 데이터 거래를 포함하여, 개인정보의 특성을 고려한 다양한 거래모델이 연구되고 있다. 특히, 정보주체의 입장을 고려한 연구는 공학 분야에서는 소수로 진행되고 있다. Z. Su 등은 사이버-물리 시스템에서 수집되는 크라우드소싱(crowdsourcing) 기반의 센서 데이터들을 소유주(정보주체)들로부터 인센티브 기반으로 구입하는 방식을 제안하였다 [13]. 또한, Tian 등은 블록체인의 스마트 계약을 활용한 계약 기반의 개인정보 거래방식을 제안하였으며, 해당 연구에서 판매자(정보주체)는 데이터 판매의 이익과 데이터 프라이버시 유출에 따른 비용을 고려하여 거래를 진행하였다 [14].

비슷한 방식으로 앞서 소개한 개인정보 저장소 개념과 개인정보 민감도(개인 데이터 공개에 따른 자신의 프라이버시가 밝혀질 정도)를 고려하여, 정보주체 입장에서 어느 정도 자신의 개인 데이터를 공유하거나 판매할지 결정하는 방법들이 연구되고 있다. T. J. Li와 S. Raghunathan은 프라이버시 민감도를 낮음과 높음 두 가지 카테고리로 구분하고, 개인정보 판매 정도를 결정하는 방식을 제안하였다 [15]. J. Parra-Arnau는 정보주체의 입장에서 공개하는 데이터의 조합에 따라 밝혀질 수 있는 공개자의 프라이버시 침해 정도를 고려한 (프라이버시 민감도를 고려한) 개인정보 판매 방

식을 제안하였다 [16]. 해당 연구에서는 개인정보의 프라이버시 민감도를 상/중/하 3가지 카테고리로 구분하였다. 정보주체는 개인정보의 프라이버시 민감도에 따라 상/중/하 카테고리로 분류하며, 자신이 선택한 개인 데이터를 공개함에 따라 실제 자신이 특정될 위험도가 얼마나 높아지는지와 자신의 개인정보를 판매할 때의 이익을 판단하여 판매여부를 결정하게 된다.

또한, 본 연구팀에서는 정보주체가 가지는 다양한 프라이버시 민감도의 개인정보를 거래할 때, 정보주체가 생각하는 적정가격과 데이터를 구매하고자 하는 서드 파티들의 수익모델들을 다양한 통계적 실험을 기반으로 수학적 모형을 진행하였다 [17,18]. 해당 연구에서는 데이터 구매자들이 정보주체의 권리를 보장하는 선에서 비용을 지불하면서 비즈니스를 진행하여도 기존의 비즈니스모델을 유지하면서 충분히 수익을 낼 수 있다는 것을 보였으며 결과적으로 데이터 브로커를 포함한 개인정보 생태계 이해관계자 모두의 요구사항을 만족하면서 개인정보 활용 비즈니스가 성립할 수 있음을 보였다.

이번 논문에서는 이러한 연구결과들을 바탕으로 마이데이터 환경에서 개인 건강/의료 데이터 활용을 위한 데이터 거래모델에 관하여 다룬다.

3. 개인 건강/의료 데이터 활용을 위한 데이터 거래모델

이번 장에서는 마이데이터 환경에서 개인의 건강/의료 데이터 활용을 위한 데이터 거래모델을 제안한다. 이를 위해, 먼저 개인 건강/의료 데이터 거래 환경에서의 이해관계자들과 그들의 관계를 알아보기 위한 모델을 설명하고, 그들의 관계에서 생기는 개인/의료 데이터 생성부터 활용까지 일련의 절차를 제안한다. 이러한 모델을 바탕으로 주요한 이해관계자인 개인 건강/의료 데이터의 활용 동의 여부를 결정하는 정보주체와 데이터를 실제 확보하여 활용하는 데이터 활용자들의 행동을 수학적 모형을 통해 분석하고, 마이데이터 기반의 개인 건강/의료 데이터 활용을 위한 데이터 거래모델이 성립할 수 있음을 보인다.

3.1 마이데이터 기반 개인 건강/의료 데이터 활용 환경

이번 절에서는 먼저 제안하는 개인 건강/의료 데이터 활용 환경을 설명하기 위해 주요 이해관계자들과 그들의 특성 및 상호작용에 관하여 설명한다. 그림 1은 마이데이터 환경 기반 개인 건강/의료 데이터 활용 환경을 나타낸다. 이번 논문에서 제안하는 환경에서는 크게 5가지의 이해관계자(정보주체, 데이터 제공자, 데이터 관리 플랫폼, 마이데이터 플랫폼, 데이터 사용자)가 존재한다.

- 정보주체: 생성되는 건강/의료 데이터 관련 정보주체로 건강/의료 데이터 제공자와의 상호작용을 통해 개인 건강/의료 데이터를 생성하게 된다. 또한, 생성된 데이터의 활용/위탁관리 등에 관한 동의의 주체로써 개인 건강/의료 데이터의 활용은 정보주체가 동의한 범위에서만 일어난다.

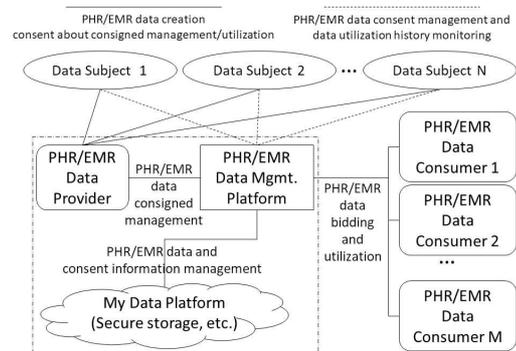


그림 1. 마이데이터 기반 개인 건강/의료 데이터 거래 환경

Fig. 1. PHR/EMR data trading environment with the concept of My Data

- 건강/의료 데이터 제공자: 건강/의료 서비스를 제공함으로써 정보주체들과 관련된 건강/의료 데이터를 생성한다. 생성된 데이터에 관하여 일차적으로 동의를 받는 이해관계자이며, 건강/의료 데이터 관리 플랫폼에 건강/의료 데이터 관리를 위임한다. 또한, 건강/의료 데이터 활용을 통해 얻은 이익을 기반으로 정보주체들에게 유·무형의 보상(예: 더

좋은 건강/의료 서비스 등을 제공함으로써 그들의 참여를 촉진하는 역할도 담당한다.

- 건강/의료 데이터 관리 플랫폼: 데이터 제공자로부터 개인 건강/의료 데이터를 전달받아 위탁관리하는 주체로 정보주체들에게는 개인 건강/의료 데이터 활용 내역 열람 및 동의 관리 등의 기능을 제공한다. 이와 동시에, 데이터 제공자 및 데이터 활용자 간의 건강/의료 데이터 거래를 중계하며, 데이터 거래 시 보안 및 규제준수, 정보주체의 동의 관리 등을 대행하는 역할을 담당한다.
- 마이데이터 플랫폼: 정보주체의 권리를 보장하면서 개인 건강/의료 데이터를 관리해주는 데이터 저장소 및 보안체계 등으로 이루어진 플랫폼으로 개인 건강/의료 데이터 생명주기(lifecycle) 뿐만 아니라 동의 정보의 생명주기도 동시에 관리한다.
- 건강/의료 데이터 이용자: 다양한 정보주체들과 데이터 제공자들의 상호작용으로 생성된 건강/의료 데이터를 확보하여 활용하고자 하는 이해관계자로, 데이터를 확보하였을 때 자신이 얻을 수 있는 이익과 데이터를 확보하는데 필요한 비용을 저울질하여 건강/의료 데이터 플랫폼이 중계하는 거래 시장에 참가한다.

이번 논문에서 고려하는 개인 의료/데이터 관리를 위한 마이데이터 모델은 데이터 제공자, 데이터 관리 플랫폼 및 마이데이터 플랫폼이 하나의 큰 조직으로 이루어져 있고, 정보주체 입장에서는 자신의 데이터 권리를 모두 보장받는 형태로 서비스가 제공됨을 가정한다. 이러한 모델을 통해, 데이터 제공자 및 데이터 관리 플랫폼은 개인 건강/의료 데이터의 외부 유출 등에 관한 보안 및 개인정보 규제준수 등을 위한 위험 부담은 줄이면서, 정보주체들에게는 마이데이터 플랫폼이 조직 밖에 있을 때와 같은 정도의 서비스를 제공할 수 있다.

또한, 이번 논문에서 고려하는 데이터 거래모델은 건강/의료 데이터의 개인정보 민감도가 높고, 독점적으로 활용했을 때 그 가치가 높아지는 특성을 고려하여 데이터 활용자들과 데이터 관리 플랫폼 간의 데이터 경매 모델로 가정한다. 경매에 나온 데이터의 접근

권을 독점하는 환경에서 데이터 이용자 간의 행동 모델을 위해 비크리 경매(Vickrey auction)로 알려진 밀봉 입찰 기반 2차 가격 경매 방식(낙찰자가 두 번째로 높은 입찰가격을 내는 방식)을 가정한다 [19].

이러한 환경을 기반으로 개인 건강/의료 데이터 거래를 위한 이해관계자들의 상호작용 흐름은 그림 2와 같다. 전제적인 흐름은 개인 건강/의료 데이터의 생성 및 동의 정보 수집 단계와 개인 건강/의료 데이터의 활용 단계로 나뉜다. 이번 논문에서는 데이터의 활용에 관하여 초점을 맞추고 있어 정보주체의 권리 중 하나인 삭제권 등과 관련된 개인 건강/의료 데이터의 파기 절차의 상세는 고려하지 않으며, 데이터 활용 동의가 더는 유효하지 않게 되었을 때 데이터가 자동으로 파기되는 것으로 가정하였다.

데이터의 생성 및 동의 정보 수집 단계에서는 먼저, 정보주체와 데이터 제공자 간의 상호작용으로 개인 건강/의료 데이터가 생성된다. 처음 생성된 개인 건강/의료 데이터는 데이터 제공자가 일시적으로 관리하며, 데이터의 활용도를 높이기 위하여 정보주체에게 개인 건강/의료 데이터 활용 및 위탁관리에 관한 제안을 할 수 있다. 이때, 데이터 제공자들은 데이터를 활용했을 때 가능한 유·무형의 적절한 보상(예: 주치의 기반 가족력 관리 서비스 등)을 포함한 제안을 하며, 정보주체는 그러한 제안을 기반으로 자신의 개인 건강/의료 데이터 활용에 동의하게 된다. 정보주체의 동의를 얻은 데이터는 데이터 관리 플랫폼으로 전달되고, 데이터 관리 플랫폼은 마이데이터 플랫폼을 활용하여 개인 건강/의료 데이터 및 그에 관한 동의 정보를 관리하게 된다. 이와 동시에, 정보주체에게 데이터 관리를 확인하고 제어할 수 있는 수단을 제공하여, 정보주체의 신뢰를 확보한다.

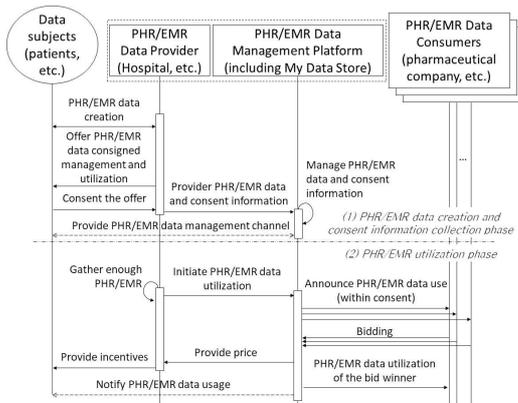


그림 2. 마이데이터 기반 개인 건강/의료 데이터 거래 절차
Fig. 2. PHR/EMR data trading procedure for medical My Data environment

이렇게 수집된 개인 건강/의료 데이터를 활용하기 위하여, 먼저 수집절차를 통해 충분한 모집단의 수를 확보한다. 충분한 모집단이 확보되면, 개인 건강/의료 데이터 활용을 위해 데이터 관리 플랫폼을 통해 데이터 활용자들에게 어떠한 내용의 데이터인지를 공고한다. 이때, 공고되는 데이터는 모두 정보주체의 활용 동의를 받은 것으로만 한정한다. 공고를 받은 데이터 활용자들은 공고된 개인 건강/의료 데이터의 종류와 특성 그리고 해당 데이터의 분석을 통해 자신이 얻을 수 있는 이익과 데이터의 확보에 필요한 비용을 고려하여 입찰에 참여한다. 데이터 관리 플랫폼은 데이터 활용자들의 입찰 내용을 확인하고, 낙찰자를 선정한다. 선정된 낙찰자는 비용을 지불하고, 활용 동意的 범위와 플랫폼과의 계약 내용을 기반으로 개인 건강/의료 데이터를 활용할 수 있는 권리를 얻게 된다. 또한, 이렇게 지불된 비용은 데이터 제공자에게 전달되며, 데이터 제공자는 이러한 비용을 기반으로 정보주체들에게 데이터 활용에 대한 유·무형의 보상을 제공하게 된다. 정보주체의 입장에서는 개인 건강/의료 데이터 활용에 관련된 절차와 데이터 관리 플랫폼과 데이터 이용자 간의 데이터 활용 계약과 같은 정보주체의 권리와 관련된 사항은 언제든지 관리 플랫폼이 제공하는 채널을 통해 파악할 수 있으며, 필요에 따라 관련 정보를 받을 수 있다.

3.2 개인 건강/의료 데이터 거래를 위한 주요 이해관계자 수학적 모델링

이번 절에서는 앞서 설명한 시스템 모델을 기반으로 데이터 거래와 관련된 다양한 분야의 실험적 데이터를 기준으로 수학적 모형을 진행한다. 먼저, 경매 방식의 건강/의료 데이터 거래모델의 주요 참여자인 데이터 활용자들의 행동 패턴 분석을 위해, 데이터 활용자가 낙찰확률을 모델링하고 데이터를 확보했을 때 얻을 수 있는 데이터의 기대가치와 그에 따른 이익에 관한 수학적 모델링을 진행한다. 그에 더해, 데이터를 확보하기 위해서는 정보주체의 동의가 필요하므로, 데이터 활용자에게서 제시되는 비용을 기준으로 정보주체의 개인정보 활용 동의를 통한 참여 비율의 변화가 어떻게 바뀌는지를 알아보고, 실제 데이터 활용자가 부담하는 비용에 관한 수학적 모델링을 진행한다. 이번 모델에서는 개인 건강/의료 데이터 경매에 참여하는 정보주체의 수를 N 으로 전체 데이터 이용자 수를 M 으로 표시하고, m 은 각 데이터 이용자를 지정하는 인덱스로 사용한다.

3.2.1 데이터 활용자의 낙찰확률 모델

데이터 활용자들은 개인 건강/의료 데이터를 확보하기 위해 경매에 참여하므로, 데이터를 확보하기 위해서는 경매에서 낙찰을 받아야 한다. 온라인 광고시장 환경에서 실제 데이터를 활용하여 경매 모델[20]을 연구한 W. Zhang 등에 따르면, 경매에서 이길 확률 W 는 다음과 같이 모델링할 수 있으며,

$$W(b_m) = \frac{b_m}{\omega + b_m} \quad (95)$$

이때, ω 는 시장 경쟁률 상수이고, b_m 은 m 번째 데이터 활용자의 입찰 금액이다.

이번 논문에서는 이러한 낙찰확률 모델을 통해 3.3 절에서 데이터 활용자의 관점에서 예상하는 이익 분석을 진행한다.

3.2.2 정보주체의 활용 동의 비율에 따른 비용함수 모델

정보주체가 개인정보를 제공하느냐 제공하지 않느냐를 결정하는 방식은 얻을 수 있는 이익 혹은 보상을 기준으로 결정할 것이다. 즉, 개인정보의 제공으로 얻을 수 있는 이익이 자신의 기준점 이하라면 개인정보를 제공하지 않을 것이고, 기준점보다 이상이라면 개인정보를 제공할 것이다. 정보주체 한 명의 개인정보 제공 의지가 어떻게 이분법적으로 결정되기 때문에 정보주체 한명 한명에 대하여 모델을 하는 것보다는, 비슷한 조건의 정보주체 집단을 대상으로 데이터나 보상 조건들을 변경했을 때 어떻게 정보주체들이 행동하게 되는지에 대한 분석이 더 효과적이다. 2018년 V. Benndorf와 H.-T. Normann은 50명의 개인정보 제공자를 대상으로 다른 가치를 가지는 데이터 종류(프라이버시 민감도 기준)와 제시되는 가격에 따라 전체 집단에서 어느 정도의 비율이 실제 개인정보 판매 의사를 보였는지 실험하였다 [21]. 그림 3(실제 실험 결과를 재구성한 그래프)을 보면 제시되는 가격이 높아질수록, 데이터의 가치가 낮을수록 더 많은 정보주체들이 데이터를 제공할 의사를 밝힌 것으로 알 수 있다.

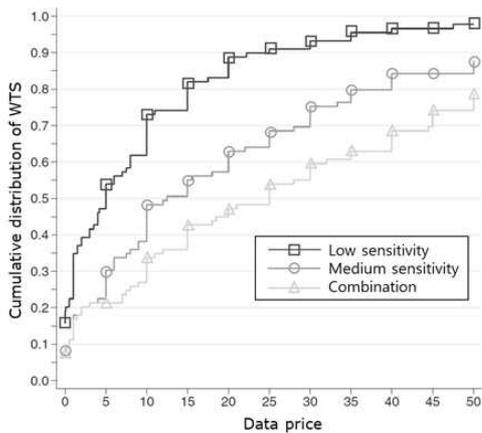


그림 3. 개인정보 프라이버시 민감도와 받는 보상에 따른 개인정보 제공 비율
 Fig. 3. Willingness-to-share personal data with respect to privacy sensitivity and incentives

이러한 실험 결과를 바탕으로 이번 논문에서는 주어

지는 보상을 b_m , 개인정보의 프라이버시 민감도를 ρ 라고 했을 때, 전체 정보주체의 집단에서 개인정보를 제공하는 비율 $P(b_m)$ 을 다음과 같이 정의하였으며,

$$P(b_m) = 1 - \frac{1}{1 + (1 - (1 - \rho)b_m)} \quad (2)$$

이때, $\rho(\rho \in (0, 1))$ 는 프라이버시 민감도를 나타내는 상수이다. 민감도 값이 작을수록 상대적으로 가치가 낮은 데이터를 의미하고, 값이 클수록 상대적으로 가치가 큰 데이터를 의미한다. 모델을 고려할 때, 결과 값이 $[0, 1]$ 에 있으며, 비슷한 개형의 곡선을 근사할 수 있는 다양한 함수를 적용할 수 있으며 [17], 이번 논문에서는 이후 분석의 단순화를 위하여 해당 모델을 고려하였다.

실제 정보주체가 동의할 비율을 예측할 수 있음에 따라, 데이터 확보에 필요한 비용을 계산할 수 있게 된다. 정보주체의 모집단의 수를 N 이라고 할 때, 데이터 활용자의 비용함수 $C(b_m)$ 은 다음과 같이 정의할 수 있으며,

$$C(b_m) = b_m \cdot NP(b_m), \quad (3)$$

이때, b_m 은 한 정보주체의 데이터 확보에 필요한 비용, $NP(b_m)$ 은 실제 확보한 데이터에 관련된 정보주체의 수이다. 이번 논문에서 b_m 은 정보주체에게 제공되는 유·무형의 보상 비용, 데이터 제공자 및 데이터 관리 플랫폼의 유지에 들어가는 비용을 포함한다.

3.2.3 데이터 활용자가 확보한 데이터에서

얻을 수 있는 정보량과 수익함수 모델

데이터 활용자 관점에서 일정 비용을 지불하면서 확보할 수 있는 데이터를 기반으로 다양한 분석한 결과를 도출하고 이러한 정보를 통해 새로운 수익을 내는 환경을 가정했을 때, 수익함수 모델링 관점에서 고려해야 하는 것은 2가지가 있다. 첫 번째는 확보한 데이터에서 얼마나 새로운 정보를 찾아낼 수 있는지, 그리고 두 번째는 얻을 수 있는 정보의 가치에 대비하여 얼마

나 많은 수익을 낼 수 있는가이다.

1) 데이터 가치 모델

데이터양이 많아질수록 얻을 수 있는 정보량이 많아질 확률이 높아진다. 이러한 현상은 특히 확보된 데이터양 대비 기계학습이나 인공지능 모듈들의 정확도가 높아지는 다양한 예시로 확인할 수 있다[17]. 이러한 연구들의 공통점은 데이터의 양이 많아지면 많아질수록 실제로 얻어지는 정확도 상승량은 적어지는 한계효용의 법칙이 데이터 분석에서도 적용되는 사실이다. 이러한 실험들을 기반으로 제안하는 데이터 가치 모델은 데이터양이 많아질수록 큰 값을 가지면서, 한계효용의 법칙을 적용할 수 있는 형태로 모델링을 진행하였다. b_m 의 비용을 활용하여 얻을 수 있는 데이터 가치 $I(b_m)$ 은 다음과 같이 정의할 수 있으며,

$$I(b_m) = \log(1 + NP(b_m)), \quad (4)$$

이때, N 은 정보주체 모집단의 수, $P(b_m)$ 은 실제 b_m 의 가격에서 동의하는 정보주체의 비율을 나타낸다.

2) 데이터 활용자 수익모델

도출한 데이터 가치 모델을 기반으로 여러 가지 분석을 한다고 가정했을 때, 실제 얻을 수 있는 수익을 모델링 할 수 있다. 건강/의료 데이터는 개인정보 민감도가 높고, 이번 논문에서 제안하는 거래 시장 환경에서는 낙찰받은 데이터를 독점적으로 활용할 수 있으므로 데이터 가치에 따른 활용 이익이 점점 지수함수적으로 커진다고 가정하였다. 이러한 가정을 기반으로 정의한 데이터 활용자 수익함수 $R(b_m)$ 은 다음과 같으며,

$$R(b_m) = \alpha_m e^{I(b_m)}, \quad (5)$$

이때, α_m 은 이익상승률 상수이다.

3.3 데이터 활용자의 이익 분석

이번 절에서는 제안한 데이터 활용자의 수익함수와 비용함수 모델을 기반으로 이익함수를 모델링하고, 제안된 비크리 경매 방식에 기반한 개인 건강/의료 데이터 경매 모델에서 각 데이터 활용자가 자신의 이익을 최대화하기 위해 취하는 전략을 도출한다. 이를 위하여 먼저 데이터 활용자의 이익함수 $F(b_m)$ 을 정의한다. 데이터 활용자 관점에서 낙찰되었을 때는 수익과 비용이 발생하지만, 낙찰되지 않았을 때는 수익과 비용이 발생하지 않으므로 다음과 같이 정의할 수 있으며,

$$F(b_m) = W(b_m)(R(b_m) - C(b_m)), \quad (6) \\ + (1 - W(b_m)) \times 0$$

이때, b_m 은 입찰액, $W(b_m)$ 은 낙찰확률, $R(b_m)$ 은 수익함수, 그리고 $C(b_m)$ 은 비용함수이다.

데이터 활용자 관점에서 자신의 수익을 최대화하기 위해서는 주변 경쟁자들의 행동전략과 경매로 나오는 데이터의 가치 등을 종합적으로 고려하여 입찰 전략을 잘 세우는 것이 중요하다. 하지만, 비크리 경매 방식에서는 경쟁자들의 행동전략과 상관없이 데이터 활용자 자신이 가장 큰 이익을 낼 수 있는 가격입찰 (truthfully bidding)을 하는 것이 가장 효율적인 전략이라는 것이 증명되어 있다 [18]. 따라서, 데이터 활용자의 소신껏 입찰액을 작성하는 것이 최대이익을 보장하므로, 이익함수 값을 최대로 할 수 있는 b_m 을 찾는 것이 목표가 된다.

$$F(b_m) = \frac{b_m}{\omega + b_m} \left\{ \alpha(1 + N(1 - \frac{1}{1 + (1 - \rho)b_m})) - b_m N(1 - \frac{1}{1 + (1 - \rho)b_m}) \right\} \\ s.t. \quad \omega \geq 1, \quad 0 < \rho < 1, \\ \alpha_m > 0, \quad N > 0, \quad b_m > 0. \quad (7)$$

이익함수 F 는 볼록함수 (convex function)이므로 1차 도함수 값이 0이 되는 (즉, $dF(b_m)/db_m = 0$) b_m^* 을 찾을 수 있다. 미분을 통해 위 조건들을 만족하

는, 이익을 최대화할 수 있는 입찰가 b_m^* 을 구하면, 다음과 같다.

$$b_m^* = \frac{\alpha_m}{2N} \left\{ N + 1 + \sqrt{\frac{4N}{\alpha_m \rho} + (N + 1)^2} \right\}. \quad (8)$$

이러한 결과를 통해 데이터 활용자는 자신의 수익이 나는 정도와 실제 획득하려는 개인 건강/의료 데이터의 양 (즉, 참여하는 정보주체의 수) 그리고 데이터의 프라이버시 민감도 정도에 따라 입찰 금액을 바로 결정할 수 있게 된다. 특히, 데이터 활용자로서는 입찰 금액을 정할 때, 수익이 나는 정도(α_m)와 데이터의 양 (N)이 주요하므로 획득할 수 있는 데이터가 자신의 사업 등에 얼마나 도움이 되는지에 따라 데이터 경매 시장 참여 여부를 결정할 수 있다.

4. 성능 분석

이번 장에서는 3장에서 제시하고 도출된 다양한 모델을 검증하고, 개인 건강/의료 데이터 경매 시장의 주요 파라미터들 (α_m, ρ, N)이 실제 데이터 활용자의 입찰 금액에 어떻게 영향을 주는지 알아본다.

먼저, 제시한 데이터 활용 동의 비율 $P(b_m)$ 에 관하여 분석한다. 그림 4는 같은 정보주체의 수 (N)을 기준으로 다양한 프라이버시 민감도를 가지는 개인 건강/의료 데이터를 다룰 때, 제시되는 입찰 금액에 따른 정보주체의 참여 비율을 나타낸다 (그림 3과 비슷한 개형을 나타내는 것을 알 수 있다). 또한, 프라이버시 민감도가 낮은 개인 건강/의료 데이터 (예: 심박수 같은 웨어러블 디바이스에서 얻을 수 있는 건강 데이터 등)일수록 낮은 가격일 때 많은 참여율을 가지며, 프라이버시 민감도가 높은 데이터 (예: 실제 진료기록 등)일수록 높은 가격을 제시해야만 확보할 수 있는 것으로 나타났다.

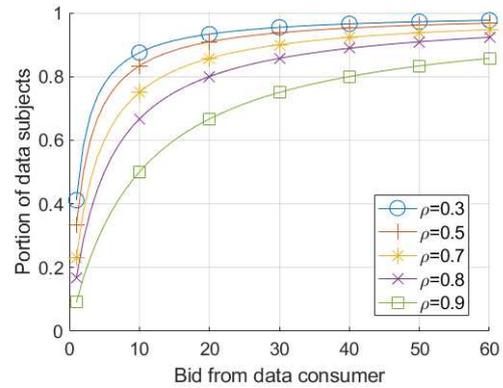


그림 4. 데이터 프라이버시 민감도 (데이터 가치)에 따른 정보주체의 개인 건강/의료 데이터 제공 비율
Fig. 4. Willingness-to-share PHR/EMR of data subjects with respect to data privacy sensitivity

다음은 데이터 활용자의 입찰 금액에 따라 얻을 수 있는 이익 $F(b_m)$ 에 관하여 분석한다. 그림 5는 같은 정보주체의 수 (N)와 수익률 (α_m)를 기준으로 다양한 프라이버시 민감도를 가지는 개인 건강/의료 데이터를 입찰했을 때, 얻을 수 있는 이익의 변화를 나타내었다. 다루는 데이터 종류에 따라 이익의 절대량을 그 변화도는 비슷한 것을 알 수 있다. 일정 입찰가격까지는 데이터양의 증가에 따라 얻을 수 있는 이익이 더 큰 상태가 지속되어 이익이 늘어나지만, 데이터양이 계속 증가할수록 한계효용의 법칙이 적용되어 얻을 수 있는 수익에 비해 데이터 획득을 위한 전체 금액이 증가하므로 데이터 활용자의 이익이 줄어들게 된다.

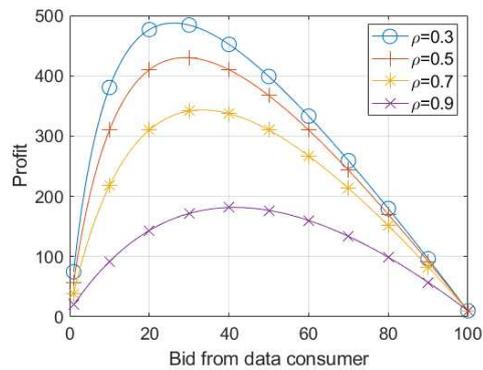


그림 5. 데이터 활용자의 입찰가격에 따른 이익 분포
Fig. 5. The trends of profit with respect to bid from a data consumer

그림 5에서는 이익함수에 관한 기본적인 개형을 확인하기 위하여 개인 건강/의료 데이터의 프라이버시 민감도(데이터 종류별 가치)가 수익률 변수(α_m)에 영향을 주지 않는 것을 가정하여 분석하였다. 이에 결과적으로, 적은 입찰가로 데이터양을 많이 얻을 수 있는 프라이버시 민감도가 낮은 데이터를 다룰 때 이익이 큰 것으로 분석되었다. 하지만, 일반적으로 가치가 높은 건강/의료 데이터를 활용했을 때 더 큰 수익을 기대할 수 있을 것이므로, 그 관계를 더 알아낼 수 있다면 더 정확한 분석이 가능하게 될 것이다.

마지막으로 데이터 활용자 관점에서 자신의 수익률(α_m)과 정보주체 수(N)에 따라 들어가는 비용을 고려한 최대의 이익을 낼 수 있는 입찰금액(b_m^*)의 경향성을 알아본다. 그림 6은 같은 프라이버시 민감도를 가지는 데이터를 다루는 환경에서 정보주체의 수와 데이터 활용자의 수익률에 따라 최대이익을 낼 수 있는 입찰금액의 분포를 나타낸다. 데이터 활용자는 자신이 얻을 수 있는 수익률이 높을수록 많은 데이터를 확보하기 위해 최대 입찰 금액이 높아지는 경향을 보였다. 정보주체의 수가 많을수록 더 많은 정보량을 얻을 수 있지만, 한계효용을 법칙에 따라 얻을 수 있는 정보량이 조금씩 줄어든다. 이러한 결과로 정보주체의 수가 늘어날수록 한 사람에게 제공되는 금액은 조금씩 줄어들게 된다. 단, 실제 데이터를 제공하는 정보주체의 수가 늘어나므로 전체적인 비용 자체는 늘어난다.

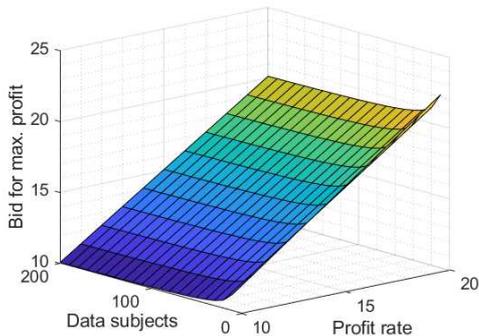


그림 6. 정보주체의 수와 데이터 활용자의 수익률에 따라 최대이익을 낼 수 있는 입찰 금액 분포
 Fig. 6. The trends of bids for maximum profit of data consumers with respect to data subjects and profit rates

이러한 분석을 통해 데이터 활용자 관점에서는 확보하고자 하는 개인 건강/의료 데이터의 프라이버시 민감도(혹은 정보주체가 느끼는 데이터의 상대적인 가치)와 자신이 확보한 데이터 분석을 통해 얻을 수 있는 수익을 고려하여 입찰해야 하는 것으로 알 수 있었다. 특히, 이번 논문에서 자세하게 고려되지는 않았지만, 데이터 활용자가 제공하는 비용에는 데이터 저장·관리 비용뿐만 아니라 정보주체들에게 제공되는 유·무형의 보상을 위한 비용(예: 개인 맞춤형 의료 서비스 등)도 포함하므로 데이터 관리 플랫폼 및 데이터 제공자 관점에서는 자신들이 부담해야 하는 비용을 고려하여 최소 입찰 금액을 정하는 것 또한 중요하게 작용할 것이다.

5. 결론

앞으로 개인정보를 포함한 다양한 데이터를 활용하기 위해서는 이제 정보주체의 권리를 보장하면서 적절한 보상을 제공하며 데이터 활용을 할 수 있는 모델이 필요하다. 정보주체를 고려하는 개인정보 활용 패러다임의 변화는 마이데이터를 기반으로 하는 다양한 데이터 활용 모델을 만들어내고 있으며, 다양한 개인정보가 생겨나고 다뤄지는 의료분야에서도 이러한 움직임이 진행되고 있다. 이번 논문에서는 마이데이터 기반 개인 건강/의료 데이터 활용 생태계가 확산되었을 때, 생겨날 수 있는 이해관계자들 간의 데이터 거래모델을 제안하였고 다양한 실험 결과를 기반으로 한 수학적 모델링을 통해 개인 건강/의료 데이터 거래 모델이 성립할 수 있음을 보였다. 이번 논문에서는 마이데이터 기반 데이터 거래 모델의 성립 가능성에 관한 직관을 얻기 위해 이해관계자 행동에 관한 다양한 수학적 모델링이 간소화된 형태로 고려되었다. 앞으로 의료분야 특성에 맞는 다양한 통계나 연구결과를 바탕으로 더욱 정교한 모델링을 통해 마이데이터 기반 개인 건강/의료 데이터 생태계를 뒷받침할 수 있는 다양한 데이터 활용 모델들을 연구할 수 있을 것이다.

REFERENCES

[1] KISDI, "A Study on the Internet Policy Framework in the era of 4th Industrial

- Revolution”, Report 2018-0-00373, Oct. 2018.
- [2] World Economic Forum, Rethinking Personal Data: A New Lens for Strengthening Trust, 2014.
- [3] E. Ramirez, J. Brill, M. K. Ohlhausen, J. D. Wright, and T. McSweeney, “Data Brokers: A Call for Transparency and Accountability”, *Federal Trade Commission*, May 2014.
- [4] A. Rieke, H. Yu, D. Robinson, and J. von Hoboken, “Data Brokers in An Open Society”, *Open Society Foundations*, November 2016.
- [5] M. Little, “Personal Data Futures: The Disrupted Ecosystem - How user-control of personal data could ‘game-change’ Big Data and reconfigure the Internet”, *OVUM*, January 2014.
- [6] Regulation (EU) 2016/679 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ 2016/L 119/1, 27 April 2016.
- [7] Y. Lee, “A Study on the Revision Trend of Data 3 Act”, *The Journal of Comparative Private Law*, vol. 27, no. 2, pp. 423-463, May 2020.
- [8] pde.cc. Personal data ecosystem consortium - empowering people with their personal data. [Online]. Available: <https://pde.cc/>
- [9] P. Kumaraguru and L. F. Cranor, “Privacy Indexes: A Survey of Westin's Studies,” in *CMU-ISRI-5-138*. Institute for Software Research International, School of Computer Science, Carnegie Mellon University, December 2005.
- [10] M. Karampela, S. Ouhbi, and M. Isomursu, “Connected Health User Willingness to Share Personal Health Data: Questionnaire Study”, *J Med Internet Res*, vol. 21, no. 11, November 2019.
- [11] E. F. Franklin, H. M. Nichols, L. House, J. Buzaglo, and K. Thiboldeaux, “Cancer Patient Perspectives on Sharing of Medical Records and Mobile Device Data for Research Purposes.” *Journal of Patient Experience*, May 2020.
- [12] Ponemon Institute, “Privacy and Security in a Connected Life : A Study of European Consumers”. *Trend Micro and Ponemon Institute*, 2015.
- [13] Z. Su, Q. Qi, Q. Xu, S. Guo, and X. Wang, “Incentive scheme for cyber physical social systems based on user behaviors,” *IEEE Transactions on Emerging Topics in Computing*, 2018.
- [14] Tian, J. Li, W. Li, B. Ramesh, and Z. Cai, “Optimal contract-based mechanisms for online data trading markets,” *IEEE Internet of Things Journal*, vol. 6, October 2019.
- [15] X.-B. Li and S. Raghunathan, “Pricing and disseminating customer data with privacy awareness,” *Decision Support Systems*, vol. 59, pp. 63-73, 2014.
- [16] J. Parra-Arnau, “Optimized, direct sale of privacy in personal data marketplaces,” *Information Sciences*, vol. 424, pp. 354-384, 2018.
- [17] H. Oh, S. Park, G. M. Lee, H. Heo and J. K. Choi, “Personal Data Trading Scheme for Data Brokers in IoT Data Marketplaces,” *IEEE Access*, vol. 7, pp. 40120-40132, 2019.
- [18] H. Oh, S. Park, G. M. Lee, J. K. Choi and S. Noh, “Competitive Data Trading Model With Privacy Valuation for Multiple Stakeholders in IoT Data Markets,” *IEEE Internet of Things Journal*, vol. 7, no. 4, pp. 3623-3639, April 2020.
- [19] Vijay Krishna, “Auction Theory”, *Academic Press*, 2002.
- [20] W. Zhang, S. Yuan, and J. Wang, “Optimal real-time bidding for display advertising”, *In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '14)*. ACM, New York, NY, USA, 2014.
- [21] V. Benndorf and H.-T. Normann, “The willingness to sell personal data,” *The Scandinavian Journal of Economics*, vol. 120, no. 4, pp. 1260-1278, 2018.

저자약력

오 현 택 (Hyeontaek Oh)



- 2012년 2월: 한국과학기술원 전산학 (학사)
- 2014년 2월: 한국과학기술원 전기 및 전자공학 (석사)
- 2020년 2월: 한국과학기술원 전기 및 전자공학 (박사)
- 2020년 2월 ~ 현재: 한국과학기술원 IT융합연구소 지능화기술연구팀 팀장

〈관심분야〉 ICT트러스트, 마이데이터, 개인정보 생태계

양 진 홍 (Jinhong Yang)



- 2017년 2월: KAIST 정보통신공학 박사
- 2017년 2월~2018년 1월: HECAS 기술책임(CTO)
- 2017년 10월~현재: 한국과학기술원 IT융합연구소 겸직교수
- 2018년 3월~현재: 인제대학교 헬스케어IT 학과 조교수

〈관심분야〉 데이터 컴플라이언스, 마이데이터, 헬스케어 데이터 활용