

의료 빅 데이터를 활용한 서비스 제공 프레임워크 설계

신봉희¹, 전해경^{2*}

¹인천대학교 컴퓨터공학부 교수, ²YM-나을텍 선임연구원

Design of Service Provision Framework using Medical Big Data

Bong-Hi Shin¹, Hye-Kyoung Jeon^{2*}

¹Professor, Dept. of Computer Science & Engineering, Incheon National University

²Senior Researcher, YM-NaeulTec

요약 본 논문에서는 의료용 빅 데이터를 활용하여 비즈니스와 연계하여 새로운 서비스를 창출하기 위한 프레임워크를 설계하였다. 단순한 데이터 분석 단계를 나타내는 것이 아니라 데이터의 활용 목적을 명확히 하고, 이에 대한 분석을 수행하여 그 속에서 가치를 추출하고 실제 사업이나 서비스를 운용할 때까지의 과정을 설계한다. 설계된 프레임워크는 기본 아키텍처, 사회 시스템 모델까지 커버할 수 있도록 하였다. 설계된 프레임워크를 참조하여 사회 시스템에 적용될 수 있도록 디자인하였으며, 기본 데이터로는 의료용 빅 데이터를 중심으로 하였다. 의료용 기본 데이터를 적용한 프레임워크 설계로 여러 의료용 사업 제휴 및 서비스 창출을 실현할 수 있을 것으로 기대하고 있다.

주제어 : 융합, 빅 데이터, 서비스창출, 프레임워크, 의료

Abstract In this article, we have presented a framework, designed to create new services for businesses, which use large sets of medical data. It is not a simple data analysis step, but it clarifies the purpose of data utilization, analyses it, extracts value from it, and designs a process from actual business or service to an operation. The designed frame work covers the basic architecture and social system model. It was designed, using basic data, which was focused on large sets of medical data, and to be applied to a social system with reference to the designed framework. We are looking forward to create various medical business alliances and services applying the designed framework to the available sets of basic medical data.

Key Words : Convergence, Big Data, Creation of service, Framework, Medical

1. 서론

1.1 연구 배경 및 목표

더 이코노미스트(2010)는 빅 데이터를 활용하여 전 세계에 닥친 환경, 에너지, 식량, 의료문제를 어느 정도 해결할 것으로 예상하고 있고, 가트너(2011)는 빅 데이터가 미래 기업의 사업력을 좌지우지할 Crude Oil 이라고 하

였으며, 맥킨지(2011)는 빅 데이터를 어떻게 활용하는가에 따라 기업과 공공사업 분야의 경쟁력 확보와 생산선의 향상, 신규 사업 발견의 차이가 생길 것으로 보고 하였다. 또한, 맥킨지(2011)는 빅 데이터를 활용하여 미국의 의료분야에서 연간 약 3,000억 달러, 유럽 공적분야에서 연간 약 2,500억 달러의 경제적 효과가 있을 것으로 보며 우리나라는 약 10조의 정부예산지출을 감소시킬 것

*This work was supported by the Incheon National University Research Grant in 2018.

*Corresponding Author : Hye-Kyoung Jeon(jhk7010@nate.com)

Received December 13, 2018

Revised January 7, 2019

Accepted February 20, 2019

Published February 28, 2019

으로 예상하고 있다. 일본 총무성(2012)은 빅 데이터의 활용이 빠르게 진행된다면 여러 사업에서의 가치 창출과 공공적 비용의 절감이 있을 것으로 보았다[1].

노인 인구의 증가와 고혈압, 당뇨와 같은 만성질환자의 증가로 인해 의료비에 대한 부담과 의료 서비스의 접근성과 질에 대한 문제가 제기 되면서 여러 국가에서 IT와 의료 기술을 접목한 u-Health 도입을 서두르고 있다. u-Health는 의료비 감소 등의 사회 및 경제적 비용의 절감 효과와 공공의료서비스 및 예방관리 보건 등의 국가 정책적 효과를 기대할 수 있다[2]. u-Health는 의료분야에서 많은 변화가 있을 것으로 예상된다. 통신기술과 Sensing 기술의 발전으로 u-Health 단말기나 Smart TV 등을 통해 의사와 원격으로 진료가 가능한 의료 서비스를 이용할 수 있고 개인의 건강정보를 저장하는 전자의무기록(EHR)을 이용하여 환자 개개인의 건강 상태를 즉시 확인할 수 있다. u-Health는 다양한 Raw 생체 정보를 수집하기 위해 다양한 센서들의 네트워크가 필요하다. 스마트 센서들이 수집한 의료 정보 등은 여러 형태로 분석 및 처리되어 개인 의료 정보 저장소에 저장되고 의료진에게 전송되어 이용될 수 있다. 최근 u-Health 의료 서비스를 이용하면서 생산되는 의료 정보 관련 빅 데이터의 관리와 활용에 대한 논의가 활성화 되고 있다[3].

의료용 빅 데이터는 일반적인 빅 데이터가 가진 용량, 다양성, 신속성, 정확성이라는 특징에 복잡한 계층성, 불확실성이 더해진 것이다. 또한 일반 빅 데이터 분야에서는 인과 관계가 아닌 상관관계를 중시하는 반면, 의료용 빅 데이터 분야에서는 상관관계에서 도출된 결과를 의사 결정에 연결 할 때 신중한 판단을 필요로 하는 경우가 많다. 특히 의료적 개입이 잘못되었을 경우에 사람의 생사에 치명적 문제를 발생시킬 수 있기 때문이다[4,5].

본 논문에서는 이러한 빅 데이터를 활용하여 새로운 비즈니스와 서비스 창출을 위한 프레임 워크를 제안한다. 프레임 워크의 프로세스와 기본 아키텍처를 2장과 3장에서 고찰하고, 4장에서 결론에 대해서 논한다.

2. 데이터 활용

지금까지의 데이터 관리는 일정 품질의 데이터 집합을 필요할 때 필요한 만큼만 적정한 비용을 사용하여 적시 적소에 공급하기 위한 관리 활동이었다. 클라우드 컴퓨팅의 발전에 따라 데이터 발생원의 중심이 조직 외부로 옮겨 개별 조직에서 통합 관리 할 수 없게 되고 있는

가운데 방대한 데이터에서 미래를 예측하고 개별적으로 도움을 주는 정보를 제공하기 위해 기존의 축적된 데이터 분석 기술뿐만 아니라 데이터 통합 환경과 데이터 모델링 연구가 진행되어 왔다[6-9]. 대량으로 축적된 데이터는 일반적으로 원본 데이터만으로 얻을 수 있는 가치가 적기 때문에 적절한 기술을 기반으로 분석하고 해석하여 새로운 가치를 창출할 수 있어야 한다. 이 장에서는 건강, 의료 영역을 예로 데이터를 활용에 대해 정리하고 3장에서 의료용 빅 데이터를 활용하여 비즈니스에 적용하기 위한 프레임 워크에 대해서 정의한다.

2.1 데이터분석을 통해 얻을 수 있는 가치

방대한 데이터에서 미래를 예측하기 위해서는, Fig 1에서처럼 먼저 주목해야 할 결과와 수집된 데이터 중에서 관계를 가질만한 예측 인자를 확인하고, 관계식(상관관계와 인과 관계)를 정의하고, 예측 인자를 사용하여 결과를 설명할 필요가 있다. 그리고 그 관계식에 현재 데이터를 적용하여 미래의 상태를 예측하고 판단한다.

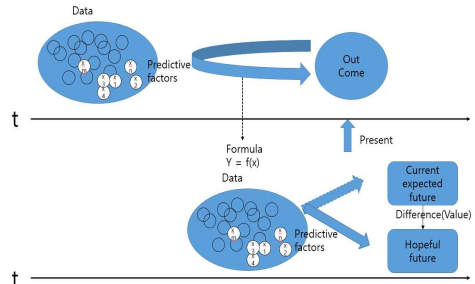


Fig. 1. Analysis of data and creation of value

만약 어떤 개입을 통해서 지금 예상되는 미래의 상태보다 좀 더 바람직한 미래의 상태를 알아낼 수 있다면 그 차이는 개인에게 가치가 될 수 있다. 즉, 분석과 예측, 그리고 판단을 하고 그에 따른 실제 행동을 취함으로써 새로운 가치가 창출될 수 있다.

2.2 분석결과를 비즈니스에 연결하는 모델링

데이터 분석을 통해 가치를 창출하는 관계식을 정의할 경우 Fig. 2와 같이 비즈니스를 위한 모델링이 필요하다. 결과를 착안하여, 원본 데이터에서 예측인자 후보 변수를 데이터 집합으로 추출해야 한다. 그리고 데이터를 분석하여 인자와 결과의 관계를 발견하고 공식화 한다. 이것을 초기 모델이라고 부른다. 방대한 데이터에서 미

래를 예측하고 개별적으로 유용한 정보를 제공할 경우 초기 모델이 비즈니스에 적합하지 않는 것으로 여겨진다. 구체적 생각해 본다면 예측 인자가 되는 데이터값을 구할 때 비싼 측정 장비 등으로 필요 비용이 많이 드는 경우이거나 측정 자체에 며칠 혹은 몇 주의 시간이 필요한 경우를 들 수 있다. 이러한 경우는 간단한 측정 방법을 선택하고, 다른 지표로 대체하는 등 인자의 변화를 검토할 필요가 있다. 인자의 교체 시에는 예측/진단 결과의 정확성과 가치의 저하를 고려해야 한다. 실제 서비스는 각 개인에 대해 측정된 데이터와 모델링을 사용하여 제공하여 예측 결과와 진단 결과 등의 형태로 제공된다.

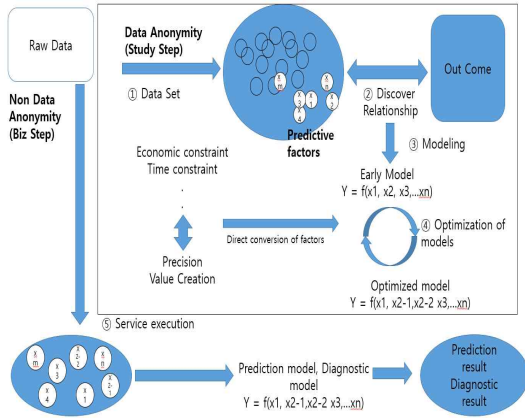


Fig. 2. Appropriate model for business

데이터집합을 추출해서 모델링 할 때까지의 단계를 연구단계, 모델링된 예측 결과와 진단결과를 제공하는 단계를 비즈니스 단계라고 부른다. 이 두 단계는 연구활동과 비즈니스 활동 등 활동 성격 차이 외에 처리 데이터를 익명화하는 것이 가능한가라는 차이도 있다. 연구 단계에서는 결과와 인자의 관계를 적절한 샘플 수의 데이터를 이용하여 구하는 것이 중요하며 각 샘플들의 구분이 필요하다. 하지만 개인에 대한 특정을 할 필요가 없다. 왜냐하면 익명화된 데이터에서 충분히 적용할 수 있으므로 상세한 개인 정보를 필요로 하는 수준은 아니기 때문이다.

2.3 데이터 활용 프로세스

지금까지의 고찰을 바탕으로 서비스 단계를 포함하여 데이터 활용 과정을 정리한 것이 Fig. 3이다.

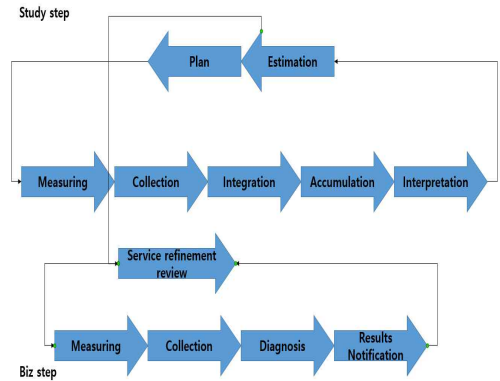


Fig. 3. Data utilization process

연구 단계에서 필요한 데이터는 계획에 따라 측정된 여러 데이터가 적절한 빈도, 시간에 수집 및 통합되고 축적된다. 새로운 기술과 서비스를 발견하고, 결과와의 관계의 유효성을 검증하기 위해 분석이 이루어지고 그 결과의 타당성 등을 평가함으로써 측정 항목 및 측정 방법의 개선을 검토한다. 계획에서 평가까지의 사이클을 적절한 횟수로 실시한 후 비즈니스 단계로 이행한다.

사업 단계에서는 모델링된 서비스를 검토하고 서비스를 받는 개인에 대한 데이터 측정을 실시한다. 측정된 데이터를 수집하고 모델을 이용한 진단 판정 분석을 실시하고 그 결과를 이용자에게 제공한다. 또한 이용자의 의견을 통해 서비스 자체의 개선 등의 검토도 실시한다. 지금까지 데이터의 활용에 있어서는, 수집 / 통합 / 축적이 정보 기반이 담당하는 영역으로 인식되고 있으며, 일반적으로 데이터 센터가 그 역할을 담당하고 있었다. 그러나 이용자나 고객 관점의서비스 제공을 생각하면 위의 과정을 지원하는 기반(데이터 뱅크 기능)이 필요할 것으로 생각된다.

2.4 데이터 활용이 적용된 사회 시스템 모델

Fig. 4에서는 건강, 의료 분야를 소재로 데이터 활용에 관한 이해 관계자와 그들 각각이 누리는 가치에 대해 보여준다. 이것은 역학(Epidemiology)연구중심으로, 그 성과가 서비스로 확장되고 개인 가치가 환원되는 구조를 나타낸다.

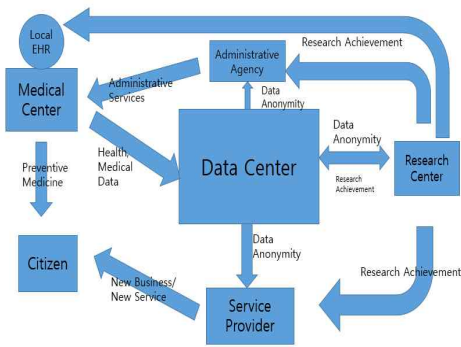


Fig. 4. Leverage data across health and healthcare

개인 (역학 연구에 참여하는 주민 등)은 의료기관에서의 진료 데이터나 건강 진단 결과 등의 건강, 의료 정보를 데이터 뱅크에 등록한다. 연구 기관은 데이터 뱅크보다 익명화된 정보를 입수하고, 역학 연구 등의 건강, 의료 관련 연구를 실시한다. 연구 기관의 성과는 새로운 기술로 의료 기관에 제공하는 다른 생활 습관에 관한 예측 모델 및 위험 평가 모델로서 서비스 제공자에게 제공한다. 그 성과는 의료기관 행정에서의 서비스와 비즈니스/서비스 제공자의 개인 건강 서비스로 주민에게 제공한다.

이처럼 데이터의 활용 방식은 업종을 넘어 서로 다른 업종의 사업자가 연계 된 비즈니스 시스템을 형성하고 사람과 사회에 가치를 제공한다. 시스템을 형성하고 비즈니스를 연결해 단일 사업자로는 할 수없는 새로운 비즈니스를 창출하기 위해서는 ICT((Information and Communications Technologies)를 잘 적용하는 것이 유효하다고 생각한다[10,11].

3. 데이터를 활용하기 위한 기반 (데이터 뱅크)

3.1 기능

데이터 뱅크는 주민 및 데이터를 수집하고 저장하는 기관 (기존 보건소 및 지역 의료기관 및 오픈 데이터)에서 정보와 샘플을 제공 받아 수집, 통합 데이터 센터 및 시료 보관 시설에 축적한다.

연구 기관 및 서비스 제공자가 데이터 뱅크에서 데이터 및 시료를 제공을 받을 때 데이터 뱅크에 의한 이용 심사를 통과하고 이용 목적이 적절하다고 판단되면 익명화된 데이터 및 시료를 얻을 수 있게 된다.

데이터 공급자에 대해서는 데이터 제공 업체와 이용 상황, 사용할 수 있는 서비스 등을 제공한다. 또한 데이터

이용자에 대해서는 데이터 뱅크가 보유한 데이터와 시료에 대한 카탈로그의 제공과 이용에 있어서의 컨설팅도 실시한다. 데이터 뱅크의 운영은 행정의 가이드라인에 근거한 가이드라인에 따라 진행되며 데이터의 수집, 통합, 축적에 강력한 보안 및 개인 정보 보호가 적용된다. 기능의 자세한 내용은 Table 1, 2와 같다.

Table 1. Data bank function1

Function	Function (Subdivision)	Summary
Data Provider Service	Matching	Information on the data providing destination is presented, and a range for providing data is set.
	Feedback	Present information on how the provided data is used.
	Basic survey	Based on a unified measurement standard, we collect basic information for data providers.
Collection, integration	Integration	For data provided from each data source, <ul style="list-style-type: none"> ▪ Name identification (integration of personal identification ID) ▪ Standardization of codes
	Data, curation	Correct the collected data at an appropriate level.Extract necessary items from the collected data.Exclude inappropriate items from the collected data.
Security, privacy protection	Data Anonymity	In the collected and integrated information, the ID for identifying an individual is converted into another system.This function is implemented by an organization independent of other function classification.Organizations conducting anonymization do not retain information other than IDs.
	Security	Access control on accumulated dataData leakage prevention (encryption, decentralization)

Table 2. Data bank function2

Function	Function (Subdivision)	Summary
Offer	Providing data catalog	Provide outline (basic statistics, multidimensional classification, catalog like information such as simple trend analysis) on data accumulated / managed.
	Usage review	We review the usage plan and judge whether the purpose of using the data is appropriate.
	Data provision	Provide necessary data / specimen based on usage plan after anonymizing it.
	Consulting on data usage	Provide advice and consultation to users on data plan that can be provided by platform and consistency of sample.
Accumulation	Accumulation	Collect and integrate collected data and samples in the data center and specimen repository.
Operation	Creation and operation of guidelines	Set and operate the guidelines necessary for the operation of the data bank.
	Internal audit	We internally audit whether the data bank is managed according to guidelines to be observed.
	External audit	We will audit by external organizations whether the data bank is managed according to guidelines to be observed.
	Ethics review	In the operation of data banks and user review, if the judgment from the ethical aspect is necessary, the ethics committee consisting of external experts will examine the contents.

3.2 지켜야할 정책들

데이터 뱅크의 운영에 있어서는 보안 및 개인 정보 보호 대책이 중요하다. 행정에서 제시되는 법령과 가이드

라인을 준수하는 형태로 데이터 뱅크로 운영 규칙, 가이드라인을 정의하고 데이터 공급자 데이터 이용자에게 제시한다. 법령, 가이드라인에 대해 정리하면 다음과 같다.

- 개인 정보 보호 및 개인 정보 보호 등을 중심으로 기능적인 면 이외에도 제도와 조직 설계 등이 필요하다.
- ICT 및 건강, 의료 영역에서는 법령, 가이드라인이 있고, 데이터 뱅크에서 취급하는 데이터에 따라 동일하게 준수해야 할 가이드라인을 정리하고 대응할 필요가 있다.

4. 결론

본 논문에서는 의료용 빅 데이터를 활용하여 비즈니스와 연계하여 새로운 서비스를 창출하기 위한 프레임워크를 설계하였다. 데이터의 활용 목적을 명확히 하고, 이에 대한 분석을 수행하여 가치를 추출하고 실제 사업이나 서비스를 운용할 때까지의 과정을 설계하였다. 설계된 프레임워크는 기본 아키텍처, 사회 시스템 모델까지 커버할 수 있도록 하였다.

향후 고령화 사회를 생각했을 때, ICT로 해결할 수 있는 과제는 다수 존재하는 것으로 보인다. 특히 대규모 데이터의 활용 결과를 ICT를 활용하여 이용자 관점과 고객 관점의 서비스에 연결하는 노력은 앞으로도 지속적으로 확장할 수 있으며 확장이 요구되고 있다고 생각한다. 농업 현장에서 CAS(Content Addressed Storage) 기술을 응용하면 수확한 농산물을 버리는 것은 일어나지 않는다. 또한 CAS 기술과 건강, 의료계의 노력과 기술을 융합, 의료 분야에서 한층 더 응용도 기대할 수도 있다.

또한 이용자/지역 밀착형의 관점에서 일상적인 쇼핑에서부터 의료 및 건강, 거기에 정내 에너지 관리의 원스톱화등을 실현하는 지역 통합 서비스로도 확장시킬 수 있을 것이다. 중장기적인 면에서도 차세대 기술로서 재생의료 및 센서 기술을 접목할 수 있고 예방의학적인 면에서도 의료와 음식, 농업 등으로 확장하여 스마트한 건강 도시 실현을 목표로 할 수 있다. 향후 여러 분야에 프레임워크를 적용하여 새로운 비즈니스 활동도 창출할 수 있으리라 기대된다.

REFERENCES

[1] K. S. Kim.(2015.08.12). *Little and often fills the purse*.
www.ablenews.co.kr

[2] Y. S. Jeong. (2015). U-healthcare Service Management Scheme for Big Data of Patient Infomation. *Journal of Convergence Society for SMB, 5(1)*, 1-6

[3] T. M. Song.(2013). Korea's Health and Welfare Big Data Trend and Utilization Plan. *Science and Technology Policy Institute, 23(3)*.

[4] J. S. Hwang. (2013). Big data, how to balance privacy and social values. *The Journal of Digital Policy & Management, 11(11)*, 143-153.

[5] M. G. Song. (2016). The suggestion of new big data platform for the strengthening of privacy and enabled of big data. *Journal of Digital Convergence, 14(12)*, 155-164.

[6] Y. Lee, C. H. Song & C. R. Park. (2016). A Study of Guidelines for Coverage in Dangerous Areas. *The Case of Japan. Japanese Cultural Studies, 58*, 227-252.

[7] J. B. Yi, C. K. Lee & K. J. Cha. (2015). An Analysis of IT Trends Using Tweet Data. *Journal of Intelligence and Information Systems, 21(1)*, 143-159.

[8] D. W. Kim. (2013). Big Data Use Cases of the Sector. *Dong-A University Business Research Center, 34*, 39-52.

[9] J. J. Hong. (2017). The Improvement Plans of Legal Framework and Regulations of IoT Healthcare Service. *KANGWON LAW REVIEW, 50*, 801-837.

[10] O. H. Kwan, S. H. Nam & C. H. Lee. (2001). A Study on Practical Analyzing and Improving Disaster Management Organization of Korean Government. *Korea Institute of Fire Science and Engineering, 15(1)*, 127-138.

[11] S. H. Oh & S. H. Shin. (2016) Correlation Analysis of Gyeongju Earthquake Waveform and Structural Damage Scale. *JOURNAL OF THE ARCHITECTURAL INSTITUTE OF KOREA Structure & Construction, 32(12)*, 33-44.

신봉희(Shin, Bong Hi)

[정회원]



- 1977년 : 인하대학교 전자공학과 공학사
- 1981년 : 인하대학교 전자공학과 공학 석사
- 1995년 : 단국대학교 전자공학과 공학 박사
- 2010년 ~ 현재 : 인천대학교 컴퓨터공학부 교수
- 관심분야 : 마이크로 프로세서, 임베디드시스템, 사물인터넷
- E-Mail : bhshin@inu.ac.kr

전혜경(Jeon, Hye Kyoung)

[정회원]



- 1995년 2월 : 인하대학교 일문과(문학사)
- 1999년 8월 : 인하대학교 정보공학과(공학석사)
- 2002년 9월 ~ 2005년 8월 : 인하대학교 컴퓨터정보학과 박사수료
- 2009년 4월 ~ 2015년 2월 : 이스트림 선임연구원
- 2016년 3월 ~ 현재 : YM나올텍 선임연구원
- 관심분야 : 상황인식, 센서네트워크, 유비쿼터스, 사물인터넷
- E-Mail : jhk7010@nate.com