

빅 데이터를 활용한 의료정보 질 관리

조영복* · 우성희** · 이상호*

*충북대학교, **한국교통대학교

The Management of Medical Information Quality Utilizing Big Data

Young-bok Cho* · Sung-Hee Woo** · Sang-Ho Lee

*Chungbuk National University

E-mail : bogicho@cbnu.ac.kr

요 약

오늘날 IT 기술의 지속적인 발전과 국민의 평균 수명이 연장되면서 의료 서비스의 질도 중요한 관심사가 되고 있다. 개인의 일상에서 생성되는 트윗 빅 데이터를 분석해서 의료정보의 질 관리 도구로 사용한다. 분석된 빅 데이터는 근거중심의학에 기반한 향상된 의료정보를 제공한다. 또한 만성질환의 추적관찰이 가능하고 유병자의 추가적인 합병증 발생률을 감소시킬 수 있다. 따라서 효율적인 질병 치료와 질병 예방이 가능하게 된다.

ABSTRACT

Today, the quality of medical service has become a major concern because that sustainable development of IT technology and extending people's life expectancy. This paper, it is used as a tool for the medical information quality management that analyze tweets big data form generated by individual's daily . The result of the analyze big data offers improvement medical information based evidence based medicine. Also it has been possible for a trace observation of chronic disease and can reduce additional other complications of patients. Therefore, effective treatment of disease and prevention is possible.

키워드

u-Healthcare Service, big data, quality management

1. 서 론

최근 소셜 네트워크에 대한 관심도가 증가하면서 많은 사람들이 개인 트위터를 사용한다. SNS(Social Networking Service)나 트위터(tweeter)는 동영상 및 음성, 텍스트 등의 비정형 데이터를 생성한다. SNS에서 생성되는 정보들은 서비스의 특성상 사용자의 주관적인 의견이나 개인 생활에 대한 내용을 단어를 이용해 표현한다.

한편 전 세계적으로 의학 기술의 발전과 함께 질병의 예방, 진단과 치료 등 관심이 증가하고 또한 우리나라 국민의 평균 기대수명이 증가하면서 함께 증가한 만성질환은 장기간 지속되어 의료산업 발달과 의료서비스 이용의 급격한 증가에 결정적 영향을 미치고 있는 것이 현실이다.

보건의료서비스는 질병으로 인한 치료의 의미뿐만 아니라 만성적인 질병에 대한 예방의 개념

을 포함한다. 예방적 보건의료서비스는 질병이 발생하는 것을 예방하는 모든 서비스로 질병 발생 이전에 주로 일반적인 건강상태의 향상을 위해 제공되는 건강증진 서비스를 비롯하여 특정 발생을 예방할 수 있는 모든 서비스를 의미한다.

모든 국민은 건강하게 오래 살기 위해 양질의 의료서비스를 제공받기를 희망하고 이것은 곧 “의료 질 향상(Medical Quality Improvement)”을 의미한다. 따라서 이 논문에서는 개인 트위터를 기반으로 일상에서 발생하는 실시간 빅 데이터 정보를 분석하여 보건의료서비스의 질을 향상시킨다.

이 논문의 구성은 2장에서는 관련연구로 근거중심보건의료와 의료의 질에 대해 기술한다. 또한 빅 데이터에 대해서 간략히 기술하고 3장에서는 빅 데이터를 기반으로 의료정보서비스 질 향상 방안을 제안한다. 마지막으로 4장에서는 결론으로

구성한다.

II. 관련연구

2.1 근거중심의학과 의료의 질

근거중심의학(Evidence Based Medicine :EBM)이란 최상의 연구근거(the best research evidence)를 의사의 숙련도(clinical expertise)와 환자의 가치 (patient's unique values and circumstances)에 접목시킨 것이다. 최상의 연구 근거는 진단검사의 평가 연구, 예후인자의 예측력 평가 연구 및 치료, 재활, 예방서비스 효능을 평가하는 연구 등을 모두 포함하는 다양한 임상관련 연구에서 얻어지는 결과를 의미한다. 의사의 숙련도란 환자의 건강상태를 살펴 진단을 내리고, 환자가 받을 치료로 인한 편익과 위험을 예측하며, 치료에 대한 환자의 선호와 기대치를 파악하고, 여기에 임상경험과 기술을 적절하게 조화시킬 수 있는 능력의 정도를 의미한다. 마지막으로 환자의 가치란 치료과정에서 환자가 갖게 되는 처치에 대한 선호도 및 관심, 치유에 대한 기대치를 의미한다[1,2].

우리나라에서는 1997년 세계가정의학회 아시아 태평양학술대회에서 Smith Richard에 의해 ‘임상진료지침 및 근거중심의학’이라는 강연이 소개되면서 각종 논문이나 서적 등을 통해 EBM에 대한 연구가 진행되었으며, 의학 분야 이외에도 ‘근거중심’의 ‘Evidence based’라는 용어를 다양한 학문분야에 적용하여 근거중심보건의료(Evidence based Health care:EBH), 근거중심간호(Evidence based nurse:EBN), 근거중심 심혈관의학(Evidence based Cardiovascular Medicine:ECM) 등으로 확대되었다.

2008년 보건복지부와 대한의학회가 서로 협력하여 임상진료지침 정보센터를 설립하는 등 임상진료지침의 개발에 박차를 가하면서 EBM의 중요성이 한층 더 대두되기 시작하였다. 특히 의료기관의 평가, 건강보험 적정성평가제도 등 의료의 질 평가 정책들이 도입되면서 EBM에 대한 인식이 점차 확산되었으며 임상 의사들의 새로운 관심의 대상으로 떠오르고 있다. 의료분야에서 환자의 권리 신장 및 제한된 보건의료 자원의 적절한 활용 필요성과 더불어 EBM이 현재 보건의료가 안고 있는 문제를 해결 할 수 있는 하나의 방법론으로 각광을 받게 하는 요인들로 작용하고 있다”고 하여, EBM은 이제 선택이 아닌 필수요소로, 어떻게 시행하느냐가 문제로 다가오고 있음을 강조하였다.

한편 근거중심의학과 질 개선(quality improvement) 모두 근거와 실무의 간극을 줄이려고 한다는 점에서 공통적 특성이 있지만 기본 원

리와 활동방식에 차이가 존재하고 있다.

이처럼 중재효과에 대한 정보를 주는 연구근거(research evidence)를 “global knowledge” 혹은 “global evidence”로 부르기도 한다. 반면에 질 개선 영역에서는 임상적 진료과정 혹은 보건의료체계의 변화를 추구하고, 반복적으로 나타는 체계적 오류를 발견하고 개선하는 것이 요구되며, 여기에는 진료과정이나 보건의료체계에 대한 정보인 과정적 지식(process knowledge)도 필요하다. Glasziou 등(2011)은 이와 같은 관계를 다음과 같이 제시하였다. [그림 1] 근거중심보건의료가 보건의료계의 주요한 흐름으로 정착되면서 임상진료지침도 근거중심적 방법을 통해 개발되기 시작했다. 앞서 언급한 체계적 고찰, 메타분석, 비용효과성 평가 등은 질병의 예방, 치료, 진단 및 재활에 필요한 중재를 권고하는데 있어 주요한 근거가 된다.

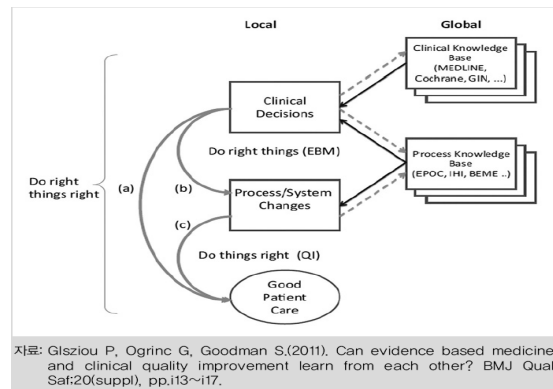


그림 1. 근거중심의학과 질 개선의 관계

2.2 빅 데이터 기술

빅 데이터(Big Data)는 소셜 미디어의 성장과 스마트폰 등 다양한 휴대용 모바일 장치에서 생성되고 사용되는 지식의 확산등으로 주목받기 시작했다. 그러나 이렇게 광범위한 범주 내에서 정확하고 신속하게 원하는 데이터를 얻는 것이 점점 어려워지고 있다는 문제점을 갖는다. 따라서 현재 제공되는 기술보다 더 효과적인 데이터 저장, 검색, 분류, 처리, 분석 방법이 요구되고 있는 것이다.

최근 들어 빅 데이터의 활용이 사회적 이슈가 되면서 중요성이 점점 더 증가하고 있다. 빅 데이터의 효과적인 활용은 새로운 지식 생산이 가능하고 사회경제적 가치를 창출할 수 있기 때문이다. 현재 해외 기업들의 다양한 빅 데이터 활용이 이루어지고 있는데 민간 분야뿐 아니라 정부를 포함한 공공 부문에서도 빅 데이터를 활용하기 위해 노력하고 있다. 맥킨지는 의료, 소매, 제조, 개인 위치정보 이외에 공공 분야도 빅 데이터 활용 사례로 소개했다.

의료계에서 빅 데이터가 활용되는 사례들도 점점 늘어나고 있다. 영국의 국가건강서비스(National Health Service)에서는 전국의 약국과 병원의 처방 데이터를 수집하여 국민 건강에 대한 예측을 수행하고 CPRD(Clinical Practice Research Data link)라는 사이트를 통해 다양한 데이터를 연구자들에게 제공하고 있다. 현재 이용 가능한 데이터는 1차 및 2차 의료 기관으로부터 수집된 질병 등록 자료를 모두 포괄하고 있으며, 인구학적, 사회경제적 변수가 포함되어 있다. 이렇듯 영국 의료계는 빅 데이터 활용이 자리를 잡아가고 있다.

연구자들은 현재 유행하고 있는 질병의 발생 장소 및 전염 속도, 주요 질병의 분포, 연도별 증가 등에 대한 통계치를 확보하여 최종적으로 효율적이고 신속한 질병관리가 가능해지게 해준다. 질병 예측과 예보의 측면에서도 다양한 빅 데이터 활용 사례가 존재한다. 미국의 존스 홉킨스 대학에서는 소셜 미디어인 트위터를 이용하여 질병 예보 시스템을 개발했으며, 이는 인플루엔자부터 알레르기까지 다양한 종류의 질병 추적이 가능한 기술을 구현하게 되었다. Seton Health Care Family와 IBM 공동개발 솔루션은 연간 200만 명 환자의 진료 정보를 분석, 추적하여 환자가 미래에 겪을 수 있는 질환, 증상을 예측하였다. 더불어 IBM은 심혈관 질환 예측을 통해 심근경색 발병 위험을 줄이는 솔루션도 개발할 수 있었다. 국내에서도 해외 사례들을 바탕으로 빅 데이터를 의료계에서 활용하고자 기술적인 도입이 이루어지고 있다.

III. 빅 데이터를 기반으로 의료서비스의 질 관리

오늘날 의료산업은 첨단 IT 기반의 언제 어디서나 사용 및 접근이 가능한 인간중심의료로 진화하고 있다. 스마트 기기를 이용한 다양한 미디어 콘텐츠들 통해 과거 치료중심의료 서비스에서 질병의 예방 및 관리에 초점을 맞춘 인간중심의 서비스로 패러다임이 변화하고 있다. 스마트 기기를 통한 원격진료는 의료서비스에 있어 과거 문제시 되었던 소외지역의 접근성 저하, 병원중심의 서비스 제공 등의 수동적 서비스에서 발전하여 지능화된 서비스 기술력이 급속도로 발전하고 있다.

<표2>는 의료서비스 발전에 영향을 미친 IT 기술을 정리한 것이다.

빅 데이터 분석 활용의 효과는 우선 경쟁 환경의 이해가 선행되어야 한다. 예를 들어 정보가 발생하는 소셜 네트워크의 구조와 정보전달 패턴의 파악이 매우 중요하다. 또한 트위터들의 소셜 네트워크 구조를 파악하여 소셜 미디어에 나타나는 정보의 경로를 분석하여 다양한 커뮤니티 구조를 파악하는 것이 중요하다.

표 1. 의료서비스 발전에 영향을 미치는 IT기술

ICT기술	영향
가격-성능 개선	의료장비의 저렴화
이동통신	접근성, 즉시성의 향상
인터넷	네트워크 헬스케어의 활성화
소형화	센서, 모니터 등의 소형화
영상진단단계	MRI, CT, OCT
의료데이터 관리	EMR, PHR, & Healthcare Big Data
융합기술	Pharmaceutical & Devices

빅 데이터를 분석하기 위해서는 오픈소스 기반의 분산 데이터 저장 기술인 하둡(Hadoop)이 기존 데이터베이스로 관리하기 힘든 규모와 성격의 데이터를 처리하기 위한 기술로 주목받고 있다.

이 논문에서는 개인의 소셜 미디어를 기반으로 한 의료 질 향상 도구로 활용하기 위해 트윗들의 연관된 규칙의 활용 부분들을 제안 한다. 연관규칙을 이용한 소셜 미디어의 빅 데이터의 분석의 응용 방법은 다음과 같다. 트윗들의 검색엔진을 통해 수집된 데이터를 빅 데이터에 축적하고, 이를 이용해 환자군들의 일상생활에서의 특이점을 분석하고 라이프 트래킹 측면에 건강이나, 라이프 사이클에 필요에 다양한 연관성이 있는 부분들을 개인의 성향에 맞게 분석하여 개인의료정보 레코드에 활용한다.

3.1 하둡 기반의 SNS 분석

Hadoop 시스템은 대용량 데이터의 분산 저장 및 신속한 처리를 위해 다수의 컴퓨터를 네트워크로 연결하여 하나의 시스템과 같이 사용할 수 있도록 구성한다. 트위터가 작성한 트윗을 분석하여 특정 단어나 패턴을 추출하여 핵심키워드로 분류한다. Map-Reduce 모듈은 word count 모듈을 통해 로그파일의 텍스트 문장을 각각 key와 value로 구성한다.

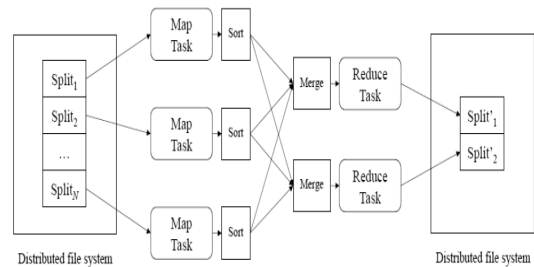


그림 2. 맵리듀스의 실행

word count는 map-reduce 분석을 통해 웹 로그 및 쿼리 로그에서 어떤 단어의 빈도를 확인할 수 있다. R 오브젝트를 기반으로 각 트윗의 단어의 빈도를 계산하고 rhocollect 함수를 통해 개별단어와 빈도수를 hadoop 시스템으로 전송한

다.

위에서 1차적으로 계산된 reduce task를 단어(key)와 value(빈도)의 누적 합을 계산하고 최종 결과 값을 HDFS에 분산 저장 한다. R에서는 대용량 패키지를 분석한다. [그림 3]은 리듀스 함수의 키-값 쌍을 입력으로 받아 최종 결과를 출력하는 의사코드이다.

```

Map(key: null, value: t = (a1, a2, ..., an, m) ∈ D)
for each q = (l1 ≤ A1 ≤ u1, l2 ≤ A2 ≤ u2, ..., ln ≤ An ≤ un, SUM(M)) ∈ Q such that
    l1 ≤ a1 ≤ u1, l2 ≤ a2 ≤ u2, ..., ln ≤ an ≤ un do
    emit (q.id, m)

Reduce(key: q.id, value_list: [m1, m2, ...])
    emit (q.id, m1 + m2 + ...)
    
```

그림 3. 의사코드

각 트윗을 통해 개인의 일상 정보를 근거로 유병자들의 생활패턴과 치료 패턴 및 약품 복용에 대한 정보를 의사들은 얻을 수 있다. 또한 보균자의 경우 질병 질병발생률을 최소화 할 수 있도록 추적감시가 가능하며 예방의학 및 근거중심의학에 매우 효율적이다.

IV. 결 론

의학은 순수한 의학 지식 생산에서 질병의 치료 진단예방에 이르기까지 기초와 응용이 융합된 학문이다. 현대 의학은 전문적인 지식의 경계로 나누어 점점 더 세분화되어 가고 있다. 분야 간의 정보와 지식의 공유는 점점 힘들어져 갔고, 상호 간의 협력이 저하된 상태에서 새로운 의학 문제들이 대두되고 된다. 약물반응의 다양성과 독성이 빈번하게 신약개발의 발목을 잡고 있다. 계속해서 항암제는 개발되고 있으나 여전히 항암제의 반응율은 높지 않고 이로 인해 환자의 존엄 문제는 계속해서 제기되고 있다. 뿐만 아니라, 많은 사람들의 생활과 밀접한 관련이 있는 건강보험제도는 의학 사회의 성격과 특성을 변화시켰고, 국민들의 신뢰 문제나 의료수가 문제 등으로 연결되었다. 이제는 의학만의 힘으로는 결코 해결할 수 없는 문제들이 번번이 발생하게 된 것이다.

학문의 세분화와 전문화는 지식생산의 효율은 높일 수 있으나 개별학문 각각의 고립을 초래하여 새로운 방향의 지식 생산에 방해가 되는 면도 있다. 그 해법 중의 하나가 빅 데이터일 것이다. 빅 데이터를 이용한 분석 기법은 기존의 데이터 마이닝을 통한 다양한 분석기법, 기계학습, 인공지능, 연관규칙, 회귀분석 등으로 정형화된 데이터를 토대로 분석기법이 발전되어 왔고, 현 시점에서 가장 의사결정을 과학적으로 검증하는 도구로 활용되어 오고 있다. 하지만 서론 부분에서 논한 것처럼 소셜 미디어의 발달, 그리고 스마트 IT

기기의 발달에서 실시간으로 대량의 데이터를 유발 하는 IT 환경의 패러다임에서 조금 더 기존의 데이터 마이닝을 통한 분석 기법들을 정형화 되지 않은 데이터들을 비즈니스적인 활용하고 분석하는데 초점이 맞추어져 있다. 빅 데이터의 분석은 무궁무진한 데이터 분석의 패러다임을 바꾸어 줄 획기적인 대용량 데이터 처리기술 입에는 틀림이 없다.

이와 같이 분석된 빅 데이터의 결과를 기반으로 의료 서비스의 질 향상에 톨로 사용한다. 질병관리를 위한 다양한 측면을 위한 U-헬스케어 서비스를 위해 유병자의 경우 추가적인 합병증이나 질병의 악화를 방지하고 보균자의 경우 질병 발생원인을 생활 밀착형 의료서비스를 제공한다면 의학에서 말하는 근거중심의료서비스를 통한 진정한 질병치료와 질병예방이 가능할 것이다.

참고문헌

- [1] Straus, Sharon E., et al., *Evidence-Based Medicine - How to Practice and Teach EBIM*, 3th ed. Elsevier, London, 2005.
- [2] 임상진료지침 정보센터(서울), <http://www.guideline.or.kr/>
- [3] 김용빈(2013), 빅데이터 활용에 있어서 개인 정보보호 문제점 및 개선방안(PIMS 활용) = *Problems of Personal Information Protection in Big Data Utilization and an Improvement Method Using PIMS*, 강원대학교 산업대학원, 컴퓨터정보통신공학과 학위논문(석사), 2013.
- [4] 이강만, "생물정보의 빅데이터 처리", 한국정보기술학회, *한국정보기술학회지* 10(3), 45-49, 2012.
- [5] 박병주 외, 『근거중심 보건의료, 고려의학, 서울, 2009.』
- [6] 김승환, 유비쿼터스 헬스케어를 위한 생체신호 모니터링 기술, *IT Soc Magazine*, 2013.
- [7] 김성수, 의료 ICT 융합산업의 효과적 비즈니스 모델 구축 방안, *KT경제경영연구소*, 2013.
- [8] Foto N. Afrati, Jeffrey D. Ullman, "Optimizing Multiway Joins in a Map-Reduce Environment," *IEEE Transactions on Knowledge and Data Engineering*, Vol.23, No.9, pp.1282-1298, 2011.
- [9] Pramod Bhatotia, Alexander Wieder, Rodrigo Rodrigues, Umut A. Acar, Rafael Pasquini, "Incoop: MapReduce for Incremental Computations," *In Proceedings of SOCC'11*, 2011.