

의료 빅데이터 처리 및 관리 시스템 설계

이승진*, 신영록*, 박준영*, 허의남*

*경희대학교 컴퓨터공학과

e-mail: {seungjin, shinyr, parkhans, johnhuh}@khu.ac.kr

Design of the Medical Bigdata Processing and Management System

Seung-Jin Lee*, Young-Rok Shin*, Jun-Young Park*, Eui-Nam Huh*

*Dept of Computer Engineering, Kyung-Hee University

요 약

최근에는 네트워크가 진화하고 데이터 처리기술이 발달하여 디지털 데이터가 활성화되면서, 기존 데이터 처리 방식으로 감당하기 힘든 규모의 데이터인 빅데이터가 매일 생산되고 있다. 이러한 대규모 데이터는 분석 및 관리를 하는데 어렵고 시간이 많이 걸리지만, 분석을 함으로써 새롭고 유용한 많은 정보를 얻을 수가 있다. 이처럼 빅데이터 분석을 통해 얻어지는 정보가 기존 분석 방식에서 얻어지는 정보와 다른 새로운 정보이기에 많은 산업분야에서 빅데이터 처리에 대한 관심이 많아지고 있다. 이러한 흐름에 따라, 의료분야에서도 빅데이터를 효율적으로 처리 및 관리하기 위한 시스템 구축을 시도하고 있다. 즉, 기존에 정형화 되어 있는 의료 데이터를 분석하여 얻는 정보에 비정형화 되어있는 의료 데이터를 추가하여 새로운 정보를 도출하려 시도하고 있다. 하지만, 여러 병원에서 서로 호환이 가능한 의료 빅데이터 처리 및 관리 시스템을 사용하기 위해서는 명확한 의료 빅데이터 처리 및 관리에 대한 요구사항과 기능정의가 필요하다. 이에 본 논문에서는 의료 빅데이터 처리 및 관리를 위한 요구사항과 기능정의를 하고 의료 빅데이터 처리 및 관리 시스템 구조를 구축하고자한다.

1. 서론

최근 사회에 다양한 산업분야의 데이터 크기가 폭발적으로 증가하는 추세이다. 이러한 이유로는 인터넷의 확산과 더불어 데이터를 처리하는 분산 처리 기술의 발달로 고효율의 컴퓨팅 성능이 가능하게 되어서이다.[1-3]

기존 데이터 크기보다 데이터의 크기가 감당하기 힘든 정도인 빅데이터로 진화하면서, 기존 데이터 처리 기술로는 데이터 처리 시간이 너무 길어지고 제대로 처리되지 않는 경우가 발생하기 시작하였다. 이러한 빅데이터를 처리하기 위해 빅데이터 처리 기술이 개발되었다. 새롭게 개발된 빅데이터 처리 기술은 기존 데이터 처리 기술가 RDBMS로 관리되는 테이블에 저장된 정형 데이터만을 처리하는 방식에서 특정한 형식이 없이 존재하는 비정형 데이터까지 처리하는 기술이다.[1-7]

새롭게 개발된 빅데이터 처리 기술로 데이터를 분석 처리함으로써 특정한 틀에 고정되는 정형데이터에서만 정보를 얻는 것보다 많고 유용한 새로운 데이터를 비정형데이터를 분석함으로써 얻게 된다. 의료분야에서도 빅데이터를 효율적으로 처리하여 새롭고 유용한 정보를 얻으려면, 의료 빅데이터에 맞는 빅데이터 처리 및 관리 기술 개발과 이에 따른 시스템을 구축해야한다.[1-7]

의료분야에서 비정형데이터는 CT나 환자의 상태를 촬영한 영상 데이터, 환자의 위치 데이터, 병원이나 환자의 집

에서 인터넷에서 입력한 SMS, 검색어와 같은 문자 데이터, 환자 스스로 자신의 상태를 얘기하거나 보호자가 얘기하는 비정형의 다양한 데이터 등이 있다.[2-5]

영상 데이터를 분석함으로써 수치 데이터에서 발견하지 못한 새로운 정보를 눈이 아닌 분석 기술로 알아낼 수 있다. 또한, 기존에 환자들이 의사의 일방적 지시에 따르기만 하는 수동적인 환자 유형에서 벗어나, 의료서비스 선택과 치료방법 결정, 질병관리 등에 적극적으로 참여하려고 하는 유형으로 변화하고 있다. 이러한 환자의 유형 변화로 환자가 인터넷에서 검색한 내용이나 SMS 내용을 검색어에 대한 분석과 특정 병에 대한 조회 수, SMS 내용을 분석하여 유행하려하는 병이나 특정 병에 증상 검색으로 병을 진단할 수도 있다. 이처럼 새로운 정보를 얻을 수 있는 비정형 데이터 분석을 하기위한 의료 빅데이터 처리 및 관리 시스템을 구축할 필요가 있다.

본 논문에서는 의료 빅데이터 처리 및 관리를 위한 요구사항과 기능정의를 하고 정형 데이터 분석뿐만 아니라, 비정형 데이터 분석이 가능한 의료 빅데이터 처리 및 관리 시스템 구조를 구축하고자한다.

본 논문의 구성은 다음과 같다. 2장에서는 빅데이터 처리의 중요성과 빅데이터 처리기술 종류에 대해서 설명한다. 3장에서는 본 논문에서 제안하는 의료 빅데이터 처리 및 관리를 위한 요구사항과 기능정의를 하고 의료 빅데이터 처리 및 관리 기술을 포함한 시스템 구조를 설명한다. 마

지막으로 4장에서는 결론과 향후연구 방향에 대해 기술한다.

2. 관련연구

2.1. 빅데이터 처리의 중요성

빅데이터 처리는 방대한 데이터의 분석을 통해 미래를 추론하고 예측하는 핵심적 기반을 제공해 준다. 이러한 예측 정보는 경제적 이익과 기존 데이터 처리 방식에서 얻을 수 있는 가치를 극대화 할 수 있다.

빅데이터를 처리하여 활용하는 것은 여러 산업분야의 혁신을 촉진하고 기업 간 경쟁력을 강화하며, 여러 산업분야의 생산성 향상에 기여하고 있다. 또한, 빅데이터 활용 시각 분야의 예산 절감 및 삶의 질 향상이 가능해진다. 예를 들어, 오바마 정부의 'Pillbox 프로젝트'로 연간 5000만 달러 비용 절감과 독일의 연방 노동기관에서 빅데이터 활용 맞춤형 고용으로 3년간 백억 유로 비용 절감이 있다. 이처럼 빅데이터 처리로 인한 사회적 이익을 창출할 수 있기 때문에 빅데이터 처리가 중요하다.[1-5]

의료분야의 경우에는 기존 정형화 데이터 검색에서 틀에 얽매이지 않은 자연어 검색으로 의료 연구 및 환자 치료 검색 활성화와 기존 치료데이터 검색 시간 단축을 통한 치료 효율성이 증가한다. 또한, 틀이 없는 검색 방식으로 데이터 검색 정확성 개선될 수 있고 형식이 없는 환자의 의견이나 의사의 의견을 자연어 검색으로 검색하고 분석하여 기존보다 상세한 진료나 환자의 상태 예측이 가능해진다. 이러한 빅데이터 처리 활용으로 인한 이점을 위해 빅데이터 처리는 반드시 활용할 필요가 있다.

2.2. 빅데이터 처리기술 종류

빅데이터 분석 기법은 네 가지로 구분하여 볼 수 있다. 먼저 Text Mining은 다양한 텍스트로부터 숨겨진 패턴을 찾고 다른 정보와의 연계성을 파악하여 카테고리들을 찾아내는 등 이를 토대로 잠재적인 유용한 정보를 제공한다. 이러한 텍스트 분석을 하기 위해서는 대용량 언어자원과 통계적, 규칙적 알고리즘이 사용된다. 두 번째로 Opinion Mining은 정형/비정형의 텍스트 상에서 사용자 의견이나 선호도에 대한 정보를 추출하는 기법으로 상품의 대한 긍정, 중립, 부정의 선호도 판별하여 상품에 대한 판매량 예측 등에 활용된다. 세 번째는 Cluster Analysis는 비슷한 특성을 가진 개체들을 분석하여 유사 특성의 그룹을 발굴하는데 사용된다. 즉, 공통 취미를 가진 집단을 분석하여 가치 있는 정보를 추출하는 것이 목적이다. 마지막으로 Social Network Analytics는 소셜 네트워크 연결 구조를 바탕으로 중심이나 허브 역할을 하는 사용자를 찾는 데 활용하는데, 모니터링이나 관리를 통해 마케팅에 활용한다.[1, 4-6]

네 가지 빅데이터 분석 기법을 기반으로 구축된 빅데이터 처리기술은 빅데이터 분석 기법을 이용하여 어떤 방식으로 분석, 처리 할 것인지를 포함하는 기술로서 하둡,

NoSQL 등이 있다. 하둡은 분산 시스템 상에서 대용량 데이터 처리 분석을 지원하는 오픈소스 소프트웨어 프레임워크로 정형, 비정형 빅데이터 분석이 가능하여 페이스북, 야후, 아마존에서 하둡 기반의 검색 서비스로 사용하고 있다. NoSQL은 관계형 데이터베이스처럼 일관성과 유효성에 중점을 두지 않고 분산가능성에 중점을 둔 비관계형 데이터베이스로 Cassandra, Hbase, Big Table, Dynamo, Riak, CouchDB, MongoDB 등이 존재한다.[1, 4-6]

3. 본론

의료 빅데이터 처리 및 관리 시스템의 구조도를 설계하기 위해서 본 시스템에 필요한 요구사항과 기능정의는 다음 <표 1>과 같다.

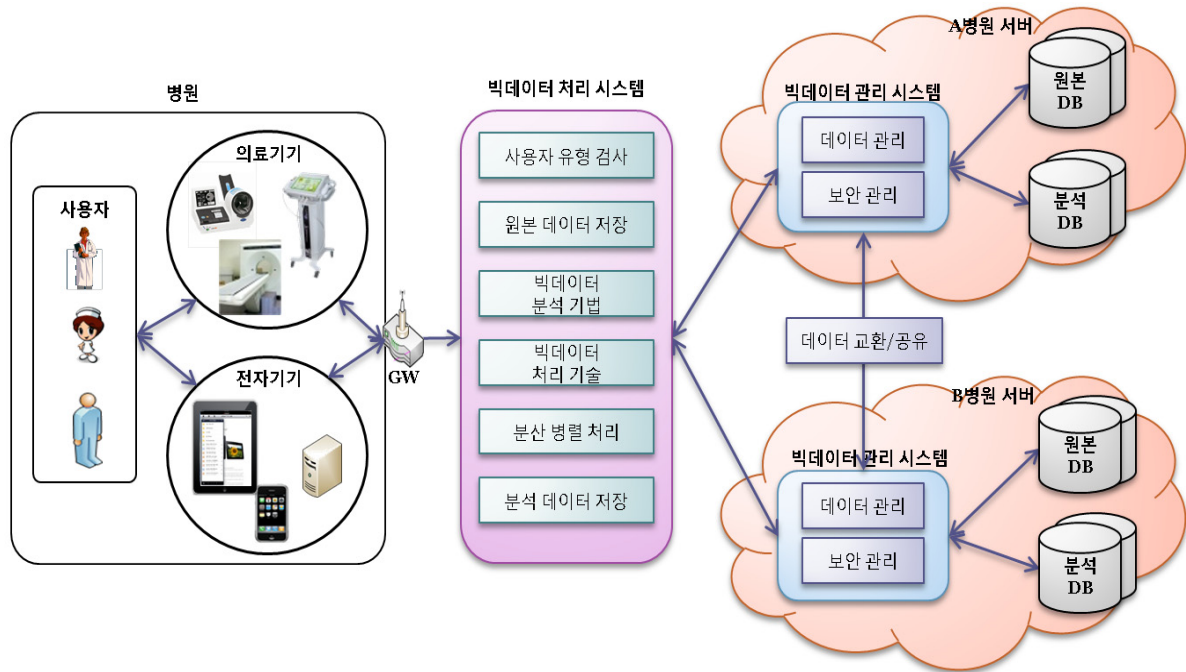
<표 1> 의료 빅데이터 처리 및 관리 요구사항과 기능정의

기능	요구사항
데이터 수집/전송	-서버와 기기간 연결
데이터 유형 분류	-데이터 유형 분류 및 필터링
데이터 분석	-데이터 유형별 분석 엔진 -데이터 분산 처리
데이터 저장/관리	-원본 데이터 분산 저장 -분석 데이터 분산 저장
웹 기반 데이터 열람	-웹과 기기간 호환성 -열람용 응용 SW
데이터 교환/공유	-교환/공유용 응용 SW

데이터 수집/전송 기능은 병원 내에 존재하는 MRI, CT, 혈압 측정기, 초음파 측정기와 같은 다양한 의료 장비와 PC, 노트북, 스마트폰과 같은 일반 전자기기로부터 측정되거나 입력되는 의료 데이터를 수집하여 서버로 전송하는 기능으로 서버와 기기 간에 의료 데이터 송수신을 위한 연결에 필요한 기능이다.

데이터 유형 분류 기능은 서버로 전송되는 의료 데이터 유형에 따라 얻을 수 있는 정보가 달라질 수 있기 때문에 활용해야하는 빅데이터 분석 기법이 다르기에 필요한 기능이다. 예를 들어, 환자의 본인 상태 인지 내역이나 의사의 상담 또는 의견 내역은 정형/비정형 데이터가 섞여있으므로 Opinion Mining 기법을 활용한다. 문장이 아닌 한 두 개의 단어로 이루어진 SMS나 검색어는 Text Mining 기법을 사용하고 x-레이와 초음파검사, MRI 촬영 등에서 나오는 사진이나 영상 데이터는 이미지 처리 기법을 사용하는 것과 같이 의료 데이터 유형에 따라 사용되는 데이터 분석 기법이 달라진다. 이렇게 분류된 데이터를 데이터 분석 기능에서 데이터 유형별에 따라 맞는 분석 기법을 활용하여 데이터는 분석하는데, 보다 빠른 분석을 위해 분산 처리된다.

데이터 저장/관리 기능은 원본 의료 데이터와 분석된 데



(그림 1) 의료 빅데이터 처리 및 관리 시스템 구조도

이터를 저장하기위해 필요한 기능이다. 원본 의료 데이터와 분석된 데이터는 일반적으로 중요도가 높은 의료 데이터로 보안을 위해 각 데이터가 분산되어 저장할 수 있는 기능이 요구된다.

웹 기반 데이터 열람 기능은 분석된 데이터를 PC나 노트북, 스마트 폰과 같은 웹에 접속이 가능한 전자기기를 이용하여 언제 어디서든 편리하게 열람할 수 있도록 필요한 기능으로 열람용 응용 SW와 웹과 기기간의 호환성이 요구된다.

데이터 교환/공유 기능은 동일한 시스템을 사용하는 다른 병원과 필요한 의료 데이터를 웹이나 P2P방식으로 데이터를 교환하거나 공유할 경우에 필요한 기능으로 열람용 응용 SW와 별도로 교환/공유용 응용 SW가 요구된다.

(그림 1)은 <표 1>에 의료 빅데이터 처리 및 관리를 위한 요구사항과 기능정의를 반영한 의료 빅데이터 처리 및 관리 시스템 구조도이다. 시스템 구조도의 병원 내에서 측정이나 직접 입력되어 생성되는 의료 데이터를 생성하는데 관련이 있는 사용자와 의료/전자기기 종류는 다음 <표 2>에서 보여준다.

<표 2> 사용자 및 의료/전자기기 예

구분	종류
사용자	의사, 간호사, 간호조무사, 전문의, 환자, 보호자 등
의료기기	혈압/초음파/키/몸무게 측정기, X-레이/CT/MRI 촬영기 등
전자기기	스마트 폰, 일반 PC, 노트북, PDA, 스마트 PC, 태블릿 등

(그림 1)의 의료 빅데이터 처리 및 관리 시스템 구조도에서 보이는 것처럼 먼저 사용자들이 의료기기와 전자기기로 측정하거나 입력한 데이터들은 병원 게이트웨이(GW)를 통해 빅데이터 처리 시스템으로 전송된다.

빅데이터 처리 시스템으로 사용자로부터 전송된 데이터는 데이터 저장 및 처리를 위해 데이터 유형을 분류한다. 유형에 따라 분류된 데이터는 병원 서버에 원본 데이터베이스(DB)에 저장하기위해 전송되며, 빅데이터 처리 시스템은 유형에 맞는 빅데이터 분석 기법과 처리 기술로 분산 병렬 처리하여 데이터를 분석한다. 분석된 데이터는 병원 서버에 분석 DB에 저장하기위해 전송된다.

병원 서버에 전송된 원본 데이터와 분석된 데이터는 각 원본 DB와 분산 DB에 데이터 보안을 위해 데이터가 분산되어 저장되고 관리된다. 이렇게 관리되는 데이터들은 동일한 시스템을 사용하고 있는 다른 병원 서버와 웹 기반 교환/공유용 응용 SW를 통해 데이터를 교환하고 공유가 가능하다. 또한, 사용자는 열람용 응용 SW를 통해 데이터를 열람할 수 있다. 사용자를 구분하여 열람 할 수 있는 데이터를 제한하기위해 서버에서 빅데이터에 대한 보안 관리가 필요하다.

4. 결론

본 논문에서는 의료 빅데이터 처리 및 관리를 위한 요구사항과 기능정의를 통해 의료 빅데이터 처리 및 관리 시스템을 설계하였다.

본 시스템은 하둡을 기반으로 의료 빅데이터를 하나의 시스템에서 모든 유형의 데이터를 분석 및 관리하도록 하였다. 동일한 처리 순서로 모든 유형의 데이터가 분석 처리되고 관리되기 때문에 새로운 유형의 데이터도 바로 분

시스템에 적용하고 분석하여 활용할 수 있다. 또한, 본 시스템을 사용하는 병원 간에 교환/공유용 응용 SW를 통해 빅데이터 및 분석된 정보를 교환하거나 공유가 가능하여 이동이 필요한 데이터를 효율적으로 옮길 수 있다.

향후 실제 병원에 의료 빅데이터 처리 및 관리 시스템을 구축하여 환자보호자와 환자, 간호사와 의사로부터 수집되는 데이터를 분석한다. 또한, 수집된 데이터에서 분석될 수 있는 환자의 의견 패턴, 의사의 진단 패턴 등의 다양하고 의미가 있는 정보를 웹 서비스로 제공하도록 구현한다.

ACKNOWLEDGMENT

“본 연구는 지식경제부 및 정보통신산업진흥원의 대학 IT 연구센터 육성지원 사업의 연구결과로 수행되었음”
(NIPA-2013-(H0301-13-2001))

참고문헌

- [1] 이명진, 김우주, “빅 데이터를 위한 고급분석 기법과 지원 기술” *Entrue Journal of Information Technology*, vol.11 no.1, pp.45-56, 2012
- [2] 고숙자, 정영호, “국민건강 미래예측 시스템 구축 방안”, *보건복지포럼*, 통권 제193호, pp.43-52, 2012.11
- [3] 박기진, “빅데이터 Eco System”, *대한산업공학회* ie 매거진, 제19권 제3호, pp.41-47, 2012.9
- [4] 최성, 우성구, “빅데이터 정의, 활용 및 동향”, *정보처리학회지*, 제19권 제2호, pp.10-19, 2012.3
- [5] 박호진, “빅데이터와 인 메모리 컴퓨팅 기술”, *정보처리학회지*, 제19권 제2호, pp.45-53, 2012.3
- [6] 조성우, “BigData 시대의 기술”, *KT종합기술원*, pp.1-8, 2011.10
- [7] 이미영, “클라우드 기반 대규모 데이터 처리 및 관리 기술”, *전자통신동향분석*, 제24권 제4호, pp.41-54, 2009.8