

빅 데이터 분석 기술 현황 및 향후전망

한지선* 윤성열* 박석천**
*가천대학교 전자계산학과
**가천대학교 컴퓨터공학과
e-mail: scpark@gachon.ac.kr

Current Status and Future Prospects of Big Data Analysis Technology

Ji-Seon Han* Sung-Yeol Yun* Seok-Cheon Park**
*Dept of Computer Science, Gachon University
**Dept of Computer Engineering, Gachon University

요 약

데이터를 대량 생산하는 스마트 기기가 대거 등장한 요즘 빅데이터는 대용량 데이터를 분석하여 가치 있는 정보를 추출하고 이 지식을 바탕으로 위기에 대응하거나 변화를 예측하는 정보기술이다. 그러나 아직 빅데이터에 대해 분류하거나 분석하는 기술의 연구가 미비하다. 따라서 본 논문에서는 빅 데이터의 개념, 배경, 분석기술 현황에 대해 분석하고 빅 데이터 유형 분석 및 분석 기술의 향후 전망을 제시한다.

1. 서론

현재 수 많은 데이터가 대용량화 되면서 빅 데이터에 대한 관심이 증가하고 있다. 그러나 빅 데이터의 정의와 범위에 대한 다양한 시각차가 존재하기 때문에 지금 소셜, 기업, 금융, 통신, 안보, 의료 등 여러가지 분야의 데이터를 이용하여 사용되는 빅 데이터에 대한 확실하고 구체화된 분류 기준과 기술들 그리고 서비스 유형에 대한 분석이 필요하다.[1]

따라서 본 논문에서는 빅 데이터의 개념, 배경, 분석기술 현황에 대해 분석하고 빅 데이터 유형 분석 및 분석기술의 향후 전망을 제시한다.

2. 관련연구

2.1 빅데이터 개념

빅 데이터란 일반적인 데이터베이스 소프트웨어로는 관리하기 어려운 정도의 큰 데이터베이스 규모로서 현재로는 수십 TB에서 향후 PB, EB 정도 크기의 대용량 데이터를 의미한다.[2] 사용자 집단을 위해 허용 경과시간 내에 데이터를 캡처하고, 관리하고 처리하는 범용 하드웨어 환경 및 소프트웨어 도구의 허용범위를 넘어서는 데이터를 말한다. 최근 빅데이터는 대용량 데이터의 수집, 저장, 분석, 체계

*가천대학교 전자계산학과

**가천대학교 컴퓨터공학과 교수(교신저자)

화를 위한 도구, 플랫폼, 분석기법 등을 포괄하는 용어로 변화되었다[3]. 그림 1은 빅데이터의 정의를 나타낸다.



(그림 1) 빅데이터 정의

2.2 빅데이터 분석 기술현황

빅 데이터는 지금까지 처리할 수 없었던 대용량의 비정형 데이터를 처리하는 기술로서, 데이터를 기반으로 한 지능형 서비스를 구현하는 기반 기술이다.

분석 기술 현황에는 비/반정형 텍스트 데이터에서 자연어 처리 기술에 기반하여 유용한 정보를 추출, 가공하는 것을 목적으로 하는 텍스트 마이닝, 소셜미디어 등의 정형/비정형 텍스트의 긍정, 부정, 중립의 선호도를 판별하는 기술인 오피니언 마이닝이 있다.

그리고, 수학의 그래프 이론에 뿌리를 둔 소셜 네트워크 분석, 비슷한 특성을 가진 개체를 합쳐가면서 최종적으로 유사 특성의 군을 발굴하는데 사용되는 군집 분석 등이 있다[4].

3. 빅데이터 분석기술 및 향후전망

3.1 텍스트 마이닝

텍스트 마이닝은 비/반정형 텍스트 데이터에서 자연어 처리 기술에 기반하여 유용한 정보를 추출, 가공하는 것을 목적으로 하는 기술이다.

텍스트 마이닝 기술을 통해 방대한 텍스트 문치에서 의미 있는 정보를 추출해내고, 다른 정보와의 연계성을 파악하며, 텍스트가 가진 카테고리를 찾아내는 등 단순한 정보 검색 그 이상의 결과를 얻어낼 수 있다. 컴퓨터가 인간이 사용하는 언어(자연어)를 분석하고 그 안에 숨겨진 정보를 발굴해 내기 위해 대용량 언어 자원과 통계적, 규칙적 알고리즘이 사용되고 있다. 주요 응용분야로 문서분류, 문서군집, 정보추출, 문서요약등이 있다.

3.2 평판분석

텍스트 마이닝의 관련 분야로는 오피니언 마이닝 혹은 평판분석이라고 불리는 기술이 있다. 오피니언 마이닝은 소셜미디어등의 정형/비정형 텍스트의 긍정, 부정, 중립의 선호도를 판별하는 기술이다. 오피니언 마이닝은 특정 서비스 및 상품에 대한 시

장규모 예측, 소비자의 반응, 입소문 분석등에 활용되고 있다. 정확한 오피니언 마이닝을 위해서는 전문가에 의한 선호도를 나타내는 표현/단어 자원의 축적이 필요하다.

3.3 소셜 네트워크 분석

수학의 그래프 이론에 뿌리를 둔 소셜 네트워크 분석은 소셜 네트워크 연결구조 및 연결강도 등을 바탕으로 사용자의 명성 및 영향력을 측정하여, 소셜 네트워크 상에서 입소문의 중심이나 허브 역할을 하는 사용자를 찾는데 주로 활용된다. 이렇게 소셜 네트워크 상에서 영향력이 있는 사용자를 인플루언서라고 부르는데, 인플루언서의 모니터링 및 관리는 마케팅 관점에서 중요하다고 할 수 있다.

3.4 군집분석

군집 분석은 비슷한 특성을 가진 개체를 합쳐가면서 최종적으로 유사 특성의 군을 발굴하는데 사용된다.

예를 들어 트위터 상에서 주로 사진/카메라에 대해 이야기하는 사용자군이 있을 수 있고, 자동차에 대해 관심있는 사용자군이 있을 수 있다. 이러한 관심사나 취미에 따른 사용자군을 군집분석을 통해 분류할 수 있다.

3.5 빅데이터 분석기술 분석

빅데이터의 분석기술은 텍스트마이닝, 평판분석, 소

<표 1> 빅데이터 분석기술

기술	주요사항	특징	활용
텍스트 마이닝	방대한 텍스트에서 의미 있는 정보추출하여 다른 정보와의 연계성 파악	구조화되지 않은 자연어로 이루어진 텍스트에서 의미를 찾아냄	문서분류, 문서군집, 정보추출, 문서요약
평판분석 (오피니언 마이닝)	데이터의 긍정, 부정, 중립의 선호도를 판별	전문가에 의한 선호도를 나타내는 표현/단어 자원의 축적이 필요	특정 서비스 및 상품에 대한 시장규모 예측, 소비자의 반응, 입소문 분석등에 활용
소셜 네트워크 분석	소셜 네트워크 연결구조 및 연결강도등을 바탕으로 사용자의 명성 및 영향력을 측정	네트워크 내 조밀한 연결 구역을 명명하는 작업을 한다.	입소문의 중심이나 허브 역할을 하는 사용자를 찾는데 활용
군집분석	관측 개체를 몇 개의 그룹으로 나눔으로써 대상 집단에 대한 효율적인 활용을 가능케 하기 위함	비슷한 특성을 가진 개체를 합쳐가면서 유사 군을 발굴	트위터 상에서 특정 취미 사용자군

설 네트워크 분석, 군집분석 등이 있는데 이는 표1과 같이 기술별 특징을 분석하였다.

4. 빅데이터의 향후전망

향후 빅 데이터는 클라우드 컴퓨팅과 소셜 미디어의 연결을 통해 2012년 이후 그 활용 움직임이 가속화 되고, 비정형 데이터를 활용하여 개개인에 최적화된 맞춤형 서비스를 제공할 수 있을 뿐 아니라, 정확도가 높은 예측분석이 가능할 것으로 전망된다.

빅 데이터에 관련한 정부의 새로운 정책들도 발표되고 있다. 2011년 11월 정부는 데이터를 정책에 활용하는 빅 데이터를 활용한 스마트 정부 구현 방안을 발표했다. 이 방안에 따르면, 정부는 공공과 민간에서 축적한 광범위한 지식 정보를 누구나 활용할 수 있도록 국가지식정보 플랫폼과 민간지식정보 유통을 촉진하는 기능의 사이버 지식정보 거래소를 2013년부터 단계적으로 구축하고, 범국가적 지식정보 활용 촉진에 필요한 법, 제도 개선과 안전한 지식정보 공유를 강화할 예정이다.

이렇듯 빅 데이터에 정확도가 높은 예측분석이 가능할 것이라고 전망되고 있고, 정부 역시 공공과 민간에서 축적한 광범위한 지식 정보를 누구나 활용할 수 있도록 국가적으로 지원을 할 예정이므로 향후 빅 데이터의 분석기술 뿐만 아니라 다양한 분야에서도 매우 적극적으로 활용이 가능할 것으로 예상된다.

5. 결론

빅 데이터는 개인과 기업에서 가치 있게 활용될 수 있을 뿐만 아니라 국가 신 성장 동력으로서의 가치가 높아지고 있다. 그러나 빅데이터의 분류기준과 기술들이 체계적으로 분류되어 있지 않기 때문에 본 논문에서는 빅데이터 분석 기술현황 및 향후전망에 대해 연구하였다.

이를 위해서 빅데이터 개념, 빅데이터 분석 기술현황, 빅데이터 유형 분석 및 향후 전망에 대해 분석하였다. 빅데이터 유형분석에서는 텍스트마이닝과 평판분석, 소셜 네트워크 분석, 군집분석 등이 있고, 이런 분석 기술들은 빅데이터 종류별로 상이하다.

빅데이터는 앞으로 여러 분야에서 중요한 화두가 될 것이다. 수 많은 비정형화된 데이터를 사용자의 니즈를 맞춰 분석하여 더 좋은 결과를 만들어내기 위한 중요한 밑천이 될 것은 분명하다.

향후에는 빅데이터에 대한 예측 분석의 정확도가 더욱 더 높아질 것이고, 국가적인 지원도 높아질 것

이므로 빅데이터의 발전 가능성은 높다.

본 논문의 연구를 기반으로 하여 국내 빅 데이터의 분석기술의 차이점을 도출하고 체계화하는 연구를 해 보고자 한다.

참고문헌

- [1] Philip Carter, "CIO를 위한 미래지향적 아키텍처 기술 그리고 로드맵", 2011
- [2] Merv Adrian, "이제는 빅 데이터이다", Teradata Magazine, Teradata Corporation, 2011
- [3] 손민선, 문병순, "빅데이터 시대 갈라파고스가 되지 않으려면", LG 경제연구원, 2012
- [4] 조성우, "BigData 시대의 기술", 중앙연구소 Intelligent Knowledge Service, KT 종합기술원, 2011