

빅 데이터를 이용한 소셜 미디어 분석 기법의 활용

Utilization of Social Media Analysis using Big Data

이병엽*, 임종태**, 유재수**

배재대학교 전자상거래학과*, 충북대학교 정보통신공학부 및 충북BIT연구중심대학육성사업단**

Byoung-Yup Lee(bylee@pcu.ac.kr)*, Jong-Tae Lim(jtlim@chungbuk.ac.kr)**,
Jaesoo Yoo(yjs@chungbuk.ac.kr)**

요약

빅 데이터를 활용한 분석 방법은 빅 데이터를 처리 할 수 있는 기술 기반으로 발전되어 오고 있다. 많은 IT 리서치 기관들이 빅 데이터를 통한 새로운 분석의 패러다임을 예견하고 있고, 또한 IT 벤더들을 중심으로 빅 데이터 처리를 위한 표준 기술들을 제시하고 있다. 빅 데이터는 IT 기기 및 환경의 발달과도 상호 연관적이고 소셜 미디어를 주축으로 기존에 예측하지 못하는 비정형화된 데이터들을 정형화 하여, 이에 따른 다양한 분석, 예측 및 최적화에 초점이 맞추어 발달 하고 있다. 과거의 분석 기법은 정형화된 데이터를 기반으로 데이터 마이닝, OLAP, 통계 분석등을 통한 의사결정 도구로서 사용되어 왔다. 하지만 최근 빅 데이터를 이용한 새로운 분석의 패러다임을 통해 분석기법의 다양화, 비정형 데이터 분석 등 새로운 형태의 기반 기술발전과 다양한 형태의 데이터를 통한 새로운 분석을 통해 통찰력을 높일 수 있다. 더욱이 고성능의 컴퓨팅 환경들의 발달과 표준화된 대용량 데이터 처리 기술 발달이 향후 조금 더 다양한 형태의 분석 패턴을 만들어 갈 것이다. 따라서 본 논문은 빅 데이터를 통해 분석 가능한 다양한 기법을 알아보고, 기존의 데이터 마이닝 분석 기법을 통한 소셜 미디어의 분석 형태에 대한 활용 및 분석방안을 제시 하였다.

■ 중심어 : 빅 데이터 | 데이터마이닝 | 분산처리 |

Abstract

The analysis method using Big Data has evolved based on the Big data Management Technology.

There are quite a few researching institutions anticipating new era in data analysis using Big Data and IT vendors has been sided with them launching standardized technologies for Big Data management technologies. Big Data is also affected by improvements of IT gadgets IT environment.

Foreran by social media, analyzing method of unstructured data is being developed focusing on diversity of analyzing method, anticipation and optimization. In the past, data analyzing methods were confined to the optimization of structured data through data mining, OLAP, statics analysis. This data analysis was solely used for decision making for Chief Officers. In the new era of data analysis, however, are evolutions in various aspects of technologies; the diversity in analyzing method using new paradigm and the new data analysis experts and so forth. In addition, new patterns of data analysis will be found with the development of high performance computing environment and Big Data management techniques. Accordingly, this paper is dedicated to define the possible analyzing method of social media using Big Data. this paper is proposed practical use analysis for social media analysis through data mining analysis methodology.

■ keyword : Cluster DBMS | High Availability | Grid |

* 본 논문은 2012년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임
(2012R1A1A2A10042015)

접수번호 : #130108-004

접수일자 : 2013년 01월 08일

심사완료일 : 2013년 01월 25일

교신저자 : 유재수, e-mail : yjs@chungbuk.ac.kr

I. 서론

소셜 서비스의 이용과 개인당 스마트 기기 보유량이 늘어남에 따라 데이터가 기하급수적으로 증가하는 대용량의 데이터의 시대가 도래 했다. 이러한 데이터 홍수 속에서 우리가 흔히 논하는 빅 데이터는 단순한 스토리지 서비스나 데이터 분석만을 의미하는 것이 아니라 빅 데이터의 핵심은 대량의 다양한 데이터를 빠르게 검색하고 분석해 경제적인 가치를 이끌어 나는데 있다. 기존 금융, 통신 사업을 비롯해 의료, 농업, 국방, 교통 정보 등 전 세계 다양한 분야에서의 무한한 빅 데이터 활용 가능성을 접하고 있다. 또한 빅 데이터를 이용한 분석과 해석으로 전염병과 자연재해의 이동경로를 파악하는 등 데이터 자원의 무궁무진하게 분석, 예측 할 수 있게 되었다. 또한 기업 환경에서는 방대한 정보 속에서 의사결정에 필요한 의미 있는 정보의 발견, 분석 능력이 기업 비즈니스의 핵심 경쟁력으로 부상하고 있으며 이는 분석 도구와 온라인상에서의 분석(Business Intelligence)플랫폼 서비스에 대한 수요 증가로 이어지고 있다. 위키피디아의 빅 데이터는 다음과 같이 정의되고 있다[3]. “빅데이터란 기존 데이터베이스 관리도구의 데이터 수집, 저장, 관리, 분석의 역량을 넘어서는 대량의 정형 또는 비정형 데이터 세트 및 데이터로부터 가치를 추출하고 결과를 분석하는 기술을 의미한다.” 즉 [그림 1]과 같이 기존에 다루던 수준을 뛰어 넘는 대규모의 자료를 뜻하는데 이와 같은 빅 데이터의 분석을 위한 IT의 발전은 최근 들어 그 속도를 더하고 있어 빅 데이터의 분석이 현실화되어 가는 과정에 있다.

빅 데이터 시장은 물리적 하드웨어로부터 시작해 인 프라 소프트웨어 에서 서비스 소프트웨어 부문으로 확장되고 있다. 특히 빅 데이터 의미 파악 및 이해 능력, 분석결과 등 총체적이고 직관적인 시각화를 위하여 애플리케이션을 제공하는 거대 플랫폼 시장으로 확대되고 있다. 이를 위해 기업에서는 내·외부의 사용자와 다양한 IT 기기를 통해 생성되고 수집되어진 대규모의 정형·비정형 데이터를 전략적으로 활용하기 위하여 통찰력을 발견하고 예측력을 제공하는 등 기업의 가치를 높일 수 있는 업무수행 과정 및 지원 플랫폼을 필요로 하

고 있고 나아가 분석기법에 대한 부분들도 새롭게 해석이 되어져야 하는 과제를 가지고 있다.

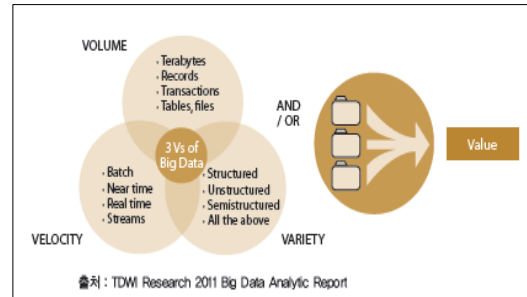


그림 1. Big Data 분석 리포트

2011년 국내 빅 데이터 시장은 2010년 대비 16.8% 성장한 2천 750억 원의 규모를 형성했으며, 2012년 국내 시장은 2011년 대비 13.1% 성장한 3천 110억 원의 규모를 형성하며 10% 이상의 고성장을 이어갈 것으로 예상되고 있다. 해외 시장은 BI, 분석 및 성과관리 애플리케이션에 대한 글로벌 투자가 2010년에는 무려 13.4%나 증가한 105억 달러로 조사되고 있다[3].

본 논문에서는 빅 데이터의 개념을 소개하고 빅 데이터가 가져온 새로운 분석 패러다임을 정리하였다. 본 논문의 구성은 다음과 같다. 2절에서는 빅 데이터의 기술 개발 동향을 소개하고 3절에서는 소셜 미디어 데이터를 활용한 데이터 마이닝 기법의 응용을 소개하며 4절에서는 결론을 제시하였다.

II. 빅 데이터 기술 동향

1. 빅 데이터 기술 개발 동향

오늘날 우리가 빅 데이터에 열광하는 이유는 과거에는 너무 커서 분석을 할 수 없었던 데이터의 분석이 이제는 가능해졌고 이것을 통해 새로운 가치(Value: 데이터의 가치)를 찾아내고 있기 때문이다. 데이터는 과거에 비해서 더 빨리 늘어나고(Volume: 데이터의 규모), 더 많은 형태(Variety: 데이터의 다양성)를 가지고, 실시간에 가까운 속도(Velocity: 데이터의 생성주기)로 생성되고 있는데, 오늘날 이러한 데이터를 이용한 빅 데

이터 분석 기술들이 현실화 되고 있다[9].

[그림 2]와 같이 빅 데이터의 기술 전망은 기업들에게 다양한 측면에서 새로운 변화와 도전을 요구하고 있지만, 이것은 곧 새로운 기회이기도 하다. 도전을 기회로 바꾸는 것은 차세대 비즈니스 분석가라고 불리는 '데이터 과학자(Data Scientist)를 통해서만 가능할 것이다. 데이터 과학자는 강력한 통계 지식을 보유하고, 대규모 데이터 세트로부터 정보를 생산하며, 비전문 분석가들에게 가치를 제공하는 능력을 갖추고 있어야 한다. 본 절에서는 빅 데이터를 분석하여 가치를 창출하기 위한 빅 데이터 분석의 기술적인 관점에 대해 살펴보고, 이것을 통해 어떤 방법으로 더 정확도 높고 적시성 있는 분석이 가능한지 알아보려고 한다.

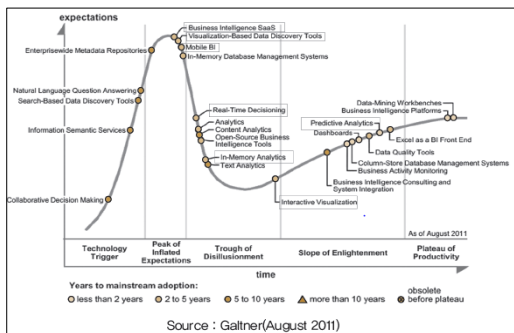


그림 2. 빅 데이터 기술의 전망(Source: Gartner 2011, Aug)

1.1 빅 데이터 솔루션 포럼(BIGSF) 구축 전략

국내 빅 데이터는 기존 BI 솔루션 벤더의 올랩(OLAP) 분석을 기반으로, 트위터 및 페이스북 등의 소셜 네트워크 데이터 분석을 결합한 정형 및 비정형 데이터 분석으로의 접근을 시도하고 해외에서도 글로벌 벤더들이 데이터 시장 주도권을 장악하기 위해 전문 솔루션 업체들에 대한 합병을 통해 관련 기술을 융합하려는 시도가 있다. 모바일 및 소셜 네트워크 서비스로 인한 방대한 비정형 데이터 분석에 특화되어 있는 소프트웨어 기업 및 전문 분석·통계 솔루션 기업들의 인수에 주력하고 있는 실정이다. 가트너 자료에 의하면 플랫폼에 관련한 기술들은 데이터 탐색을 빠르고 직관적으로 분석하는 인터랙티브 비주얼라이제이션(Interactive

Visualization), 인 메모리 분석, 실시간 결정, 모바일 의사결정(Mobile BI), 직관적 데이터 분석 툴(Visualization Based Data Discovery Tool), 그리고 BI SaaS 시장을 향후 2년에서 5년 내 활성화될 솔루션으로 분석하고 있다.

1.2 국내 기술개발 동향

빅 데이터 솔루션 포럼(Bigdata solution forums: BIGSF)의 통합 브랜드명인 싸이벨류(Cyvalue)는 대한민국의 선도 중소 소프트웨어 컨설팅 기업으로 국내뿐만 아니라 전 세계 기업시장에 내놓는 제품 라이선스 및 개발·컨설팅 서비스를 아우르는 빅 데이터 서비스 이름이기도 하다. 이는 비정형 데이터 수집 및 정형 데이터의 품질관리, 성능관리와 이를 저장·관리할 수 있는 오픈 소스형 데이터베이스, 하둡 관련한 서비스와 이를 사용자에게 효과적으로 전달할 수 있는 분석툴(OLAP), 비주얼라이제이션(Visualization) 도구 및 컴포넌트 등 빅 데이터용 통합 프레임워크 및 서비스를 제공한다. 또한 현황분석이나 모니터링 수준을 뛰어넘어, 업종별·업무별 예측 모델을 플러그인할 수 있도록 구성되어 기업에서 보다 쉽게 빅 데이터를 수용 하고 전략적인 의사결정 도구로 활용할 수 있으며 그 구성원은 다음과 같다[3].



그림 3. 빅 데이터 솔루션 포럼

빅 데이터 분석을 위해서는 빅 데이터 소스의 수집, 저장, 분석 및 의사결정을 위하여 분석된 내용을 최종적으로 모니터링, 대시보드, 소셜 네트워크 차트/그래프 및 리포팅을 통하여 보여주는 기능이 필요하며, 이

를 위해 각 기업의 보유기술을 적절하게 분담하여 빅 데이터의 각 영역을 담당하게 한다. 특히, 빅 데이터 분석에서 가장 중요한 부분은 정형 및 비정형 데이터의 인 메모리(In-Memory) 분석, 실시간 결정 및 모바일 의사 결정 툴(Mobile BD)의 구현이다[3].

결국 비즈니스 관점에서 가장 중요한 부분은 빅 데이터로부터 찾아낼 수 있는 가치가 될 것이며, 빅 데이터 분석은 의미 있는 가치를 찾아 내기위한 다양한 접근 방법이라고 정의 할 수 있다[4]. [표 1]에서 보는 바와 같이 빅 데이터 분석의 패턴은 과거의 정형화된 데이터 기반의 분석을 넘어서 새로운 IT 환경에서의 다양한 기법이 적용될 수 있는 빅 데이터 구현 기술, 분석의 기법들이 다양화 되게 활용 될 수 있다.

표 1. 빅 데이터 분석(과거 vs 빅 데이터 시대)[4]

분석속성	과거	빅 데이터 현재
데이터세트	사전정의	포괄적, 반복적
데이터속도	배치	실시간
데이터분석	과거분석	예측,시계열, 최적화
인프라형태	독립인프라	리소스 풀
어키텍처	최적화	분산/병렬/인메모리 연계
데이터크기	수십 GB	수십,수백 TB 이상
데이터종류	정형데이터	반정형, 비정형데이터
복잡도	특정주제	예측 및 최적화 관점의 요건
분석의 양	수십개의 모델	수천,수만개 이상의 세분화 모델
분석사용자	분석 전문가	데이터 시각화, 분석 모델링 자동화

1.3 국내 구축사례

현재 대기업 및 중소기업을 비롯해 모든 기업에서 가장 문제가 되고 있는 것이 설비의 돌발고장에 의한 재해와 이에 따른 심각한 경제적 손실이다[3]. 따라서 무엇보다 이를 미리 예방하는 것이 기업의 경쟁력을 높이는 방법이다. 돌발고장을 방지하기 위해 일부 기업에서는 고가의 외산 장비를 이용한 설비상태관리를 실시하고 있다. 그러나 중소기업의 경우 설비비가 고가이고 많은 비용이 소요되어, 설비에 이상이 발생할 경우 원인도 모르고 설비를 교체하거나 폐기하게 된다. 따라서 설비의 운전 상태를 파악하고 고장을 예측하여, 심각한 고장으로 진전되기 전에 대책을 세울 수 있는 설비관리 프로그램이 필요하다. 이는 발전소 내의 각종 시설에 대한 상태를 진단하여 시설장비의 수명을 사전에 예측

하여 위험을 예방하기 위한 시스템이 필요한 것이다. 이를 위해 ‘싸이벨류(Cyvalue)’ 를 통하여 시설장비의 수명 예측진단 시스템을 적용하였다[3]. 발전소를 비롯한 정유 및 석유 화학 등 대기업들은 고가의 외산 장비를 사용 하여 설비를 관리하고 있다. 그러나 고장 발생 정도만 알려 줄 뿐, 고장의 원인과 대책을 세우기 위해서는 별도의 전문 인력을 확보하거나 외부 전문 기관에 의뢰를 하고 있어 설비 고장에 따른 신속하고 정확한 대처는 미흡한 실정이다. 이를 위해서는 설비상태 평가 모듈, 설비결함 주파수 계산모듈, 설비결함, 진단모듈, 설비고장 예측모듈, 고장추적 및 대책모듈, 설비수명 예측모듈, 지식검색모듈, 데이터베이스 관리 모듈 등의 기능을 제공함으로써 명확한 고장 원인을 파악하고 적합한 대책을 세우게 되므로, 비용절감과 함께 설비가동률이 향상되며 기존의 고가 프로그램에 대한 수입 대체효과가 클 것으로 기대된다. 특히 발전소 및 석유화학 플랜트 등 고가의 외산 설비를 사용하면서 효과를 얻지 못하는 업종을 대상으로 적용이 가능하다.

2. 빅 데이터 분석 기법

빅 데이터 분석은 데이터의 생성속도와 다양성에 따라 금융 및 공공 분야의 사기 검출, 이벤트 기반 마케팅, 소셜 미디어 분석, 그리고 다양한 비즈니스 최적화 등 다양한 분야에 적용될 수 있다[9].

여기에서는 미국의 카탈리나(Catalina) 마케팅의 사례를 통해 빅 데이터 분석이 비즈니스에 어떻게 적용되는지를 설명하고자 한다. 카탈리나(Catalina) 마케팅에서는 2.5 Peta Bytes의 방대한 데이터를 전사 데이터 웨어하우스로 운영하고 있다. 이에 따르면 고객이 대형 할인마트의 매장 계산대에서 결제를 하는 순간 고객의 구매이력 데이터가 분석되어 이 고객을 위한 맞춤형 쿠폰이 인쇄된 영수증이 출력된다. 이때, 여러 명의 소비자가 동일 제품을 구매해도 각 구매이력에 따라 각각 다른 쿠폰이 출력된다. 고객의 개인별 구매 성향이 실시간으로 반영된 쿠폰 제시를 통해 구매를 유도하는 것이다. 그 결과 쿠폰 사용을 통한 추가 구매율이 기존 10%에서 25%로 크게 향상되었고 고객의 마케팅 반응율이 향상되어 카탈리나(Catalina)의 수익 증가로 이어

졌다[9]. 이러한 결과를 얻기까지 카탈리나(Catalina) 마케팅은 데이터웨어하우스(Data Warehouse(이후 DW)) 어플라이언스에 고성능 분석(High Performance Analytics(이후 HPA)) 기술을 적용하여 빅 데이터로부터 실시간으로 고객의 구매 성향을 예측하였다. 즉, 고객에게 맞춤형 쿠폰을 발행할 때 실시간으로 고객 행동에 대한 프로파일링을 수행하였으며, 기존에 45시간 걸리던 예측 모델에 대한 스코어링 시간도 불과 60초로 단축시켰다. 이렇게 카탈리나(Catalina) 마케팅은 빅 데이터를 활용하여 통찰력의 속도 향상을 이룸으로써 고객 밀착 서비스에 성공하였다[9].

빅 데이터 분석이 최근에 이슈가 되는 이유는 빅 데이터를 분석하기 위해 여러 가지 자원의 활용을 최적화하여 적시에 활용 가능하고 좀 더 정확한 분석 정보를 제공해야 한다는 니즈가 있기 때문이다[4]. 이러한 니즈가 전통적인 분석 방법의 패러다임을 바꾸고 있다고 해도 과언이 아니다. 데이터 저장과 관리의 측면에서 본다면, 어플라이언스 데이터웨어하우스와 함께 오픈소스 기반의 분산 데이터 저장 기술인 하둡(Hadoop)이 기존 데이터베이스로 관리하기 힘든 규모와 성격의 데이터를 처리하기 위한 기술로 주목받고 있다. 데이터 분석의 관점에서 주목해 볼 만한 사실은 분석의 가치를 극대화 할 수 있는 고급 분석(Advanced Analytics) 영역이 고성능 분석 기술과 밀접하게 결합되면서 본격적으로 빅 데이터 분석에 대응하고 있다는 점이다[4][10].

여기서 얻을 수 있는 가장 큰 장점은 분석 가능한 데이터 크기의 한계를 극복했다는 점이다. 이것은 샘플링에서 발생했던 표본오차를 줄일 수 있으며, 차원이 늘어남에 따라 기하급수적으로 늘어나는 계산 량에 효과적으로 대응할 수 있음을 의미한다. 따라서 수억 건 이상의 관측치와 수만 개 이상의 변수를 가진, 그 동안 다루기 어려웠던 데이터를 별다른 처리 없이 기존의 분석 작업에 활용할 수 있게 된다는 것은, 불가능했던 분석이 가능해진다는 의미일 뿐 아니라 분석 시간의 단축으로 인해 모델의 정확도를 높이는 주요한 요소가 될 것이다.

분석 라이프 사이클의 단축의 측면에서 전통적인 분석 프로세스에 따르면 데이터를 준비하고 탐색하고, 모

델을 개발하고, 운영 환경에 적용하는 데 수주에서 수개월 이상의 많은 시간이 소모되곤 하였다. 그러나 빅 데이터 분석 환경에서는 모델의 생성에서 적용에 이르는 전체 분석 라이프 사이클에 고성능 분석 기술을 적용함으로써 분석의 효율성을 획기적으로 높일 수 있게 된다.

예를 들어, 분석가는 모델링 단계에서 매우 빠르게 수천 개 이상의 변수로부터 유의 변수를 선정하거나, 다양한 모델링 방법을 비교하여 모델의 정확도를 높일 수 있게 된다. 또한 복잡한 마이닝 프로세스 대신 자동화된 방법(Rapid Predictive Modeling)으로 예측 모델을 생성함으로써 모델 생성 시간을 몇 분 이내로 단축시키고, 이렇게 생성된 분석 모델을 고성능 어플라이언스 데이터웨어하우스에 자동으로 적용하여(Deploy) 즉시 운영업무에 반영할 수 있다. 따라서 고성능 분석 툴을 이용할 경우 분석 모델의 생성과 적용 전반에 걸쳐 과거에는 생각하기 어려웠던 실시간 분석 및 운영(Operational Analytics) 프로세스가 가능해진 것이다 [2].

I장에서 언급한바 인터넷이나 소셜 네트워크, 모바일과 같은 개인화된 환경에서 쉴 새 없이 쏟아져 나오는 데이터의 폭발적 증가는 빅 데이터라는 용어를 만들어 낸 주원인이라고 할 수 있다. 따라서 비정형 데이터 및 소셜 네트워크 분석을 통한 고객 DNA 완성을 이룰 수 있는 기반을 제공하고 있다. 이와 동시에 과거 정형화된 데이터만을 분석의 대상으로 했던 분석 방법들도 비정형 데이터 영역으로 빠르게 확장되고 있다[2][4].

많은 기업들이 그 동안 적지 않은 마케팅 비용을 투자하면서도 정작 고객의 솔직한 생각을 읽는 데 어려움을 겪어 왔다. 따라서 기업들이 고객들의 진솔한 생각이 나타나 있는 인터넷과 소셜 네트워크 분석에 적극적인 관심을 가지는 것은 매우 자연스러운 현상이라 할 수 있을 것이다. 정형 데이터를 기반으로 한 고객들의 기본 정보와 오프라인 행동 정보에 비정형 데이터(고객의 커뮤니티 정보 포함)를 추가하면 고객 DNA가 완성된다. 이것을 활용하면 제한적인 분석 인사이트의 한계를 벗어날 수 있을 것이다.

3. 빅 데이터 분석 활용

3.1 빅 데이터 분석력

빅 데이터의 분석 활용은 환경 변화를 신속하게 감지하고 대응하는 역량이 중요해지면서, 기업 내, 외부에 축적된 빅 데이터의 중요성이 부각되고 있다. 기존에도 경영정보의 중요성은 강조되어 왔으나, 대부분 기업 활동과 관련된 내부 정보에 집중이 되어 한정적인 분석으로 활용되어 왔다. 신용정보회사와 금융회사가 개인 금융거래 기록을 분석해 신용등급을 산정하고 마케팅과 리스크 관리에 활용하는 것이 대표적인 사례로 들 수 있고 또한 제조업에서도 경리, 영업, 재고, 고객관리 등을 위한 전사적 자원관리, 고객 관계 관리 등의 시스템에 활용 될 수 있다. 빅 데이터는 웹사이트 방문기록, 온라인 검색통계, 소셜 미디어 소통 기록 등의 빅 데이터에는 기존의 데이터로는 파악하기 어려운 환경변화와 소비자의 니즈를 반영하는 잠재적인 정보들이 무궁무진 하다. 검색 및 포털 사이트, 소셜 미디어에는 실시간으로 각종 이슈가 등장하고 있으며, 소비자들의 선호와 행동정보가 여과 없이 노출 되어 지고 있기도 하다 [5].

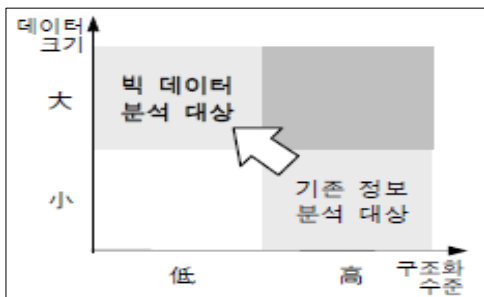


그림 4. 빅 데이터 분석과 기존 경영정보 분석의 차이점[9]

[그림 4]에서 볼 수 있듯이 기존의 경영정보의 분석은 일정한 규칙에 따라 정제된 데이터를 주로 취급하여 분석되어 왔고[9], 웹사이트 방문기록, 소셜 미디어의 소통정보를 망라한 빅 데이터는 양식이 제각기 다른 비정형화된 형태의 구조적 특성을 가지고 있다. 빅 데이터 분석은 구조화 수준이 낮고 매우 방대한 데이터를 다루므로 진보된 분산처리 기술과 통계적인 기법이 필

요 하다.

3.2 소셜 네트워크 빅 데이터 분석 활용 효과

빅 데이터 분석 활용의 효과는 우선 경쟁 환경의 이해가 선행되어야 한다. 예를 들어 정보가 발생하는 소셜 네트워크의 구조와 정보전달 패턴의 파악이 매우 중요하다. 또한 소비 활동에 영향을 주고받는 소비자들의 소셜 네트워크 구조를 파악하여 효과적인 마케팅을 위한 기반 정보를 확보하고, 소비자들은 기업이 제공하는 정보보다 프로슈머, 커뮤니티의 동료 등이 제공하는 정보를 더욱 신뢰하는 환경의 이해사 필요하다. 또한 소셜 미디어에 나타나는 정보의 경로를 분석하면 잠재적 소비자 군과 이들이 소속되어 있는 다양한 커뮤니티 구조를 파악하는 것이 가능하다. [그림 5]에서와 같이 매우 복잡해 보이는 소셜 네트워크 구조 내부에는 정보를 보다 긴밀하게 주고받는 많은 커뮤니티가 중첩됨을 알 수 있고, 각 커뮤니티의 성격과 결속 유형에 대한 정보가 중요함을 알 수 있다[6].

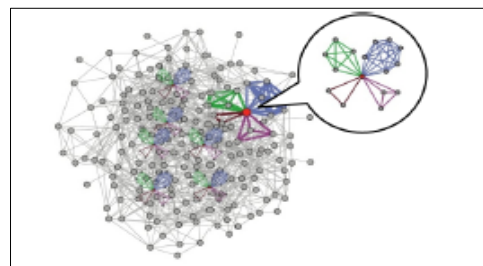


그림 5. 소셜 네트워크의 커뮤니티 구조

커뮤니티의 결속 형태를 통한 분석의 포인트는 다음 [표 2] 와 같이 활용 될 수 있다[8]. 첫째 정보의 방싯력이 우수한 오피니언 리더 및 커뮤니티와 이들을 활용하는 정보 발신 채널에 마케팅 노력을 집중함으로써 효과를 극대화 할 수 있다. 둘째, 최근 유행하고 있는 소셜 커머스 업체들은 소셜 미디어의 전파력을 마케팅에 적극 활용 할 수 있다.

소셜 미디어의 정보 분석 방법을 분류해 보면 크게 세 가지로 요약 될 수 있다 첫째, 기술 통계 부분이다. 트위터 팔로어의 변화, 페이스북 리류의 개수 등 소셜

미디어에서 일어나는 활동과 트렌드를 보여주는 다양한 통계정보를 산출해 제공할 수 있다 두 번째, 네트워크 분석이다. 소셜 미디어에서 활동하는 사람들 사이의 연결 관계를 보여주고 상호 영향의 강도를 측정하여 정보의 흐름을 구조적으로 파악하고 분석 할 수 있다. 셋째, 텍스트 마이닝 부분이다. 통상의 방법으로는 파악하기 어려운 소셜 미디어 콘텐츠의 주제 내용, 어조 등을 판별 분석을 할 수 있다[5].

표 2. 커뮤니티 결속 형태와 활용 포인트

유형	특성	사례	분석 포인트
집단형	공통의 활동이나 목표, 가치 관으로 강하게 연결되어 있거나 멤버간 교류 적음	애플의 총성고객 철인 3종 경이 참가자	공통 가치에 호소
그물형	동일한 니즈 또는 보완 관계의 니즈를 가진 멤버끼리 1:1로 직접교류	페이스북회원	입소문, 추천 강조
허브형	카리스마가 있는 핵심 인물로 결속	팬클럽	핵심인물 영향 력이용

3.3 빅 데이터와 인공지능 기법

과거 의사결정의 유용한 도구로 사용되어온 데이터 마이닝의 인공지능 분석 기법은 의사결정 과정에 부여되는 직관적인 데이터를 이용해 이를 객관적으로 검증함으로써 의사결정의 정확성을 예측 할 수 있었다[1]. 또한 복잡하고 불확실한 환경, 많거나 적은 정보가 만들어내는 혼란 속에서 의사결정자들은 상당부분을 오랜 경험에 의해 만들어진 직관에 의존하는 경향이 많았다.

빅 데이터와 인공지능의 기법의 결합은 인간의 많은 경험과 학습을 통해 통찰력을 결정 하듯이, 많은 데이터를 이용해 컴퓨터 모형의 판단 능력을 강화시키는 다양한 기법들이 발전해 왔다. 예를 들어 인공 신경망(neural network) 기법은 인간 두뇌에서 신경세포들이 학습을 반복하면서 결합 강도를 조정하여 문제해결 능력을 획득하는 과정으로 요약 될 수 있다. 또한 다중지성(swarm intelligence) 기법은 개미, 꿀벌 등 지능이 낮은 개체들이 상호 소통하면서 집단적으로 고도의 문제를 해결하는 과정을 응용하여 의사결정을 최적화해

나가는 과정이다. 따라서 빅 데이터를 이용한 다양한 기법중 인공지능 기법을 이용한 분석은 기존의 데이터 마이닝을 통한 의사결정의 최적화 단계를 조금 더 모집단의 새로운 패러다임을 통해 과학적 분석을 통한 의사 결정에 반영이 될 수 있을 것이다.

3.4 빅 데이터와 연관규칙 기법

연관규칙은 데이터 마이닝 기법에서 흔히 사용되는 기법으로서 한 항목 그룹과 다른 항목 그룹 사이에 존재하는 강한 연관성을 찾아내는 기법이다. 주로 검색 알고리즘에 대한 연구가 많이 수행되고 있으며 현 시점에서 가장 많이 활용이 되고 있는 분석 기법이기도 하다. 본 논문에서 제시하는 연관규칙을 활용 방안은 [표 3]과 같이 크게 세 가지로 분류할 수 있다.

표 3. 연관규칙을 활용한 분석 기법

구분	분석 응용
개인분석	개인의 관심도에 따른 마케팅의 활용 스마트 기기를 데이터를 통한 퍼스널 마케팅 라이프 트래킹 분석활용
검색 분석	관심 키워드 검색에 따른 실시간 추천 소비자의 검색 연관을 통한 수요 조사 및 예측 시대적 패러다임 연관 분석
소셜 미디어 분석	집단형, 그물형, 허브형 관계에 따른 연관규칙 분석 기업 이미지 마케팅의 연관분석 수요 응대를 위한 소셜 미디어 키워드 분석

첫째, 개인분석의 측면이다. 퍼스널 빅 데이터의 측면에서 다양한 니즈들을 반영한 결과들을 연관성 있게 추천할 수 있고, 두 번째 검색엔진을 통한 관심 분석의 연관성에 대한 발견 분야 이다. 세 번째, 소셜 미디어의 활용이다. 본 논문에서 언급한 바 소셜 미디어의 다양한 형태, 집단형, 그물형 그리고 허브형에 따라 연관된 규칙의 활용 부분들을 제언 하고자 한다. 연관규칙을 이용한 소셜 미디어의 빅 데이터의 분석의 응용 방법은 다음과 같다. 소비자가 검색엔진을 통해 수집된 데이터를 빅 데이터에 축적하고, 이를 이용해 제품개발 단계에서부터 고객의 성향을 반영한 제품 개발이 가능하다. 또는 라이프 트래킹 측면에서의 활용을 살펴보면 건강이나, 라이프 사이클에 필요에 다양한 연관성이 있는 부분들을 개인의 성향에 맞게 추천할 수 있는 부분으로 활용될 수 있다.

III. 결 론

빅 데이터를 이용한 분석 기법은 기존의 데이터 마이닝을 통한 다양한 분석기법, 기계학습, 인공지능, 연관 규칙, 회귀분석 등으로 정형화된 데이터를 토대로 분석 기법이 발전되어 왔고, 현실점에서 가장 의사결정을 과학적으로 검증하는 도구로 활용되어 오고 있다. 하지만 서론 부분에서 논한 것처럼 소셜 미디어의 발달, 그리고 스마트 IT기기의 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 바꾸어줄 획기적인 대용량 데이터 처리기술 입에는 틀림이 없다. 따라서 본 논문은 소셜 미디어에 주안점을 두고 기존의 분석 방법들을 어떻게 적용가능한지에 대한 방법을 제시하고자 하였다. 빅 데이터는 비정형 대용량의 데이터를 처리하는 문제를 떠나서 반드시 데이터 과학에 초점을 맞추어야 하고, 통계 모델, 텍스트 마이닝 기법, 더 나아가 심리분석의 전문적인 부분들이 추가 되어야 한다. 향후 연구 과제를 통하여 기존의 빅 데이터를 처리 할 수 있는 하둡 기반의 분산처리 기술을 통하여 비정형 데이터를 정형화 해 나가는 구조에 대한 부분들과, 비정형 데이터에 대한 분석을 위한 프레임워크에 대한 연구와 기법의 다양화를 통한 BI 측면에서의 활용도, 그리고 데이터웨어하우스의 새로운 통합의 형태들에 대한 부분들도 지속적으로 연구가 되어야 하고, 본 논문을 통해 빅 데이터를 분석하기 위한 기존의 기법들의 적용을 활용하기를 기대 한다.

참 고 문 헌

[1] 강현철, *고객관계관리를 위한 데이터마이닝 방법론*, 자유아카데미, 2006.
 [2] 김근태(SAS Korea), *"Big Data 분석을 위한 기업의 Big Analytics 환경 변화"*, 정보처리학회지, 2012.

[3] 정철호, *국내 솔루션 기반 빅 데이터 구축전략 및 사례*, *정보 속으로*, Vol.76, SEP-OCT, pp.31-37, 2012.
 [4] 최병정, 김혜진, 김자호, 진서훈, 빅 데이터 시대의 CRM을 위한 데이터 분석, *Entrue Journal of Information Technology*, Vol.11, No1, pp.19-27, 2012.
 [5] 채승병, *SERI 경영노트*, 제167호, 2012(10).
 [6] Y. Y. Ahn, *Link communication reveal multiacale complexity in metworke*, *Nature*(466), pp.761-764, 2010(8).
 [7] Doug Henschen, *"Cover Story: Big Data"*, *Infor-mationWeek The business value of technology*, 2010.
 [8] S. Fournier and L. Lee, *Getting Brand Communication Right*. *Harvard Business Review*, pp.105-111, 2009(4).
 [9] Gryman, G, *Tapping into power of Big Data*. *Technology Forecase*, pp.4-13, 2010(3).
 [9] P. Carter, *"Big Data Analytics: Future Archi-tectures, Skills and Roadmaps for the CIO"*, White pa-per, IDC sponsored by SAS. 2011.
 [10] SAS, *In-Memory Analytics for Big Data*, White paper, 2012.

저 자 소 개

이 병 엽(Byoung-Yup Lee)

종신회원



- 1991년 2월 : KAIST 전산학과 (공학사)
- 1993년 2월 : KAIST 전산학과 (공학석사)
- 1997년 2월 : KAIST 경영정보 공학(공학박사)

- 1993년 1월 ~ 2003년 2월 : 대우정보시스템 차장
- 2003년 3월 ~ 현재 : 배재대학교 전자상거래학과 부 교수

<관심분야> : XML, 지능정보시스템, 데이터베이스 시스템, 전자상거래학

임 중 태(Jongtae Lim)

정회원



- 2009년 2월 : 충북대학교 정보통신공학과 공학사
- 2011년 2월 : 충북대학교 정보통신공학과 공학석사.
- 2011년 3월 ~ 현재 : 충북대학교 정보통신공학과 박사과정

<관심분야> : 데이터베이스 시스템, 시공간 데이터베이스, 위치기반 서비스, 모바일 P2P 네트워크

유 재 수(Jaesoo Yoo)

중심회원



- 1989년 2월 : 전북대학교 컴퓨터공학과(공학사)
- 1991년 2월 : KAIST 전산학과(공학석사)
- 1995년 2월 : KAIST 전산학과(공학박사)

- 1995년 3월 ~ 1996년 8월 : 목포대학교 전산통계학과 (전임강사)
- 1996년 8월 ~ 현재 : 충북대학교 정보통신공학부 및 컴퓨터정보통신연구소 교수
- 2009년 3월 ~ 2010년 2월 : 캘리포니아주립대학교 방문교수

<관심분야> : 데이터베이스시스템, 빅데이터, 센서네트워크 및 RFID, 소셜네트워크서비스, 분산객체컴퓨팅, 바이오인포매틱스 등