기업의 빅데이터 적용방안 연구 - A사, Y사 빅데이터 시스템 적용 사례 -

Study on the Application Methods of Big Data at a Corporation - Cases of A and Y corporation Big Data System Projects -

이 재 성¹ 홍 성 찬^{2*} Jae Sung Lee Sung Chan Hong

요 으

지난 수년간 스마트 폰 같은 스마트 기기의 빠른 확산과 함께 인터넷과 SNS 등 소셜 미디어가 급성장함에 따라 개인 정보와 소비 패턴, 위치 정보 등이 포함된 가치 있는 데이터가 매 순간 엄청난 양으로 생성되고 있으며, M2M (Machine to Machine)과 loT (Internet of Things) 등이 활성화되면서 IT 및 생산인프라 자체도 다량의 데이터를 직접 생성하기 시작했다. 본 연구는 기업에서 활용할 수 있는 빅데이터의 대표적 유형인 정형 및 비정형 데이터의 적용사례를 고찰함으로써 데이터 유형에 따른적용 영역별 파급효과를 알아본다. 또한 일반적으로 알려져 있는 비정형 빅데이터는 물론 정형빅데이터를 활용하여 실제로 기업에 보다 나은 가치를 창출할 수 있는 방안을 알아보는 것을 목적으로 한다. 이에 대한연구 결과로 빅데이터의 기업내 활동이 나아갈 수 있는 지향점으로써 내・외부에서 발생하는 정형데이터와 비정형 데이터를 적절히 결합함으로써 분석의 효과를 극대화 할 수 있음을 보여 주었다

☞ 주제어 : 빅데이터, 정형, 비정형 데이터, 하둡

ABSTRACT

In recent years, the rapid diffusion of smart devices and growth of internet usage and social media has led to a constant production of huge amount of valuable data set that includes personal information, buying patterns, location information and other things. IT and Production Infrastructure has also started to produce its own data with the vitalization of M2M (Machine-to-Machine) and IoT (Internet of Things). This analysis study researches the applicable effects of Structured and Unstructured Big Data in various business circumstances, and purposes to find out the value creation method for a corporation through the Structured and Unstructured Big Data case studies. The result demonstrates that corporations looking for the optimized big data utilization plan could maximize their creative values by utilizing Unstructured and Structured Big Data generated interior and exterior of corporations.

🖙 Keyword : Big data, Structured, unstructured data, Hadoop

1. 서 론

빅데이터는 기존의 데이터 수집, 저장, 관리, 그리고 분석 역량을 넘어서는 대량의 데이터 세트를 의미하며, 기존의 관계형 데이터와 비교하여 양, 속도, 다양성 및 복잡성에서 그 차이를 볼 수 있다[1]. 데이터에는 정형화 된 데이터와 비정형화된 데이터가 있는데, 최근에 논의 되고 있는 빅데이터는 정형화된 것이든 아니든 상관없이 엄청난 양의 데이터를 말한다. 빅데이터에 대한 정의는 다양하지만, 기업적인 측면에서 빅데이터를 기업의 효과 적인 전략 도출에 필요한 상세하고 높은 빈도로 생성되는 다양한 종류의 데이터로 정의할 수도 있다[2].

구글, 페이스북, 아마존, 야후 들의 인터넷 채널들이 빅데이터를 기반으로 소설 분석의 효과를 입증하면서 많 은 기업이 빅데이터에 주목하고, 빅데이터 기반의 새로 운 비즈니스 모델 발굴을 추진하면서 빅데이터 시장이 본격화 되고 있다.

본 연구에서는 빅데이터 시장의 발전에 착안하여 기업의 내 외부에서 발생한 정형(Structured), 비정형(Unstructured) 빅데이터를 기업이 경쟁상황을 파악하고 가치 창출에 활용하기 위해 빅데이터 기술을 적용하였을 때 나타나는

LG CNS Co., 24, Yeoui-daero, Yeongdeungpo-gu, Seoul, Korea
 Division of Information and Telecommunications, Hanshin Univ., Yangsan-dong, Osanshi, Gyeonggi-do, Korea

^{*} Corresponding author (schong@hs.ac.kr)

[[]Received 4 December 2013, Reviewed 10 December 2013, Accepted 22 December 2013]

[☆] 이 논문은 한신대학교 학술연구비 지원에 의하여 연구되 었음.

효과를 검증하고 기업 내부 정형 빅데이터 처리의 개선 효과와 '마케팅'이라는 구체적인 기능에서의 비정형 데 이터 활용 방안을 제시한다.

본 연구에서 사용될 연구방법론은 실제사례를 중심으로 한 적용연구이다. 앞서 밝힌 바와 같이 기업에서 사용될 수 있는 정형 및 비정형 데이터의 대표적인 활용사례들을 분석함으로써 각 유형별로 적용될 수 있는 기술적요소, 이행방법, 기업이 느끼는 가치 및 적용 분야의 차이를 살펴보는 것이다. 빅데이터에 대한 논의는 활발하지만 아직 기업에서의 구체적인 활용사례가 그다지 많지않고, 특히, 국내에서의 활용 사례를 찾기가 쉽지 않은 상황이며, 연구의 단위를 개인이 아닌 기업으로 보는 연구의 대상이 될 수 있는 사례는 더욱 적다. 이에 따라 연구의 목적이 통계적 의미를 찾는 것이 아닌 기업 경영실무에서의 함의와 향후 활용성을 찾는 것이기 때문에, 양적 방법론이 아닌 질적 방법론인 사례연구, 그리고 기업내의 두 유형의 사례에 대한 분석 연구는 적절한 방법이라 할 수 있다.

2. 이론적 배경

2.1 빅데이터의 정의

빅데이터는 기존의 데이터 수집, 저장, 관리, 분석 역량을 넘어서는 대량의 데이터 세트를 의미하며 기존의관계형 데이터와 비교하여 양, 속도, 다양성 및 복잡성에 있어서 그 차이를 볼 수 있다[2]. 빅데이터의 정의는 다양하지만, 기업적인 측면에서 빅데이터를 기업의 효과적인전략 도출에 필요한 상세하고 높은 빈도로 생성되는 다양한 종류의 정형 또는 비정형 데이터로 정의할 수 있다[3].

맥킨지는 '빅데이터를 일반적인 데이터베이스 소프트웨어가 수집저장 관리 분석할 수 있는 범위를 초과하는 대규모의 데이터(McKinsey, 2011)'라고 정의하고 있으며, IDC는 빅데이터 기술을 '다양한 데이터로 구성된 방대한 양의 데이터로부터, 고속 캡처, 데이터 탐색 및 분석을 통해 경제적으로 필요한 가치를 추출할 수 있도록디자인 된 차세대 기술과 아키텍처(IDC, 2011)'라고 정의하고 있다[3]. 빅데이터를 특징짓는 가장 큰 부분은 기존기술로는 처리하기 어려운 정형 및 비정형 데이터가 다양한 형태로 혼재된 복잡도 높은 대용량 데이터를 신속하게 처리 가능하며 이를 기반으로 고급분석(Advanced Analytics)과 예측 등을 통한 새로운 차원의 서비스 창출

이 가능하다는 점이다[4]. 이렇듯 빅데이터는 방대한 규모(Volume), 빠른 처리 속도(Velocity), 다양한 형태(Variety), 새로운 가치(Value)의 '4V'로 정의할 수 있다[5]. 빅데이터의 특성과 효과를 구체적으로 살펴보면 (표 1)과 같다.

(표 1) 빅데이터의 특성과 효과 (Table 1) Properties and effects of big data

빅데이터 특성	효과
대규모 (Huge Scale)	 기술 발전으로 데이터 수집, 저장, 처리 능력 향상 현실계 데이터를 기반으로 한 정교한 패턴분석 가능 데이터가 많을수록 유용한 데이터, 전혀 새로운 패턴의 정보를 찾아낼 수 있는 확률 증가
현실성 (Reality)	 우리사회 일상에서의 데이터 기록물의 증가 등 현실정보, 실시간 정보의 축적이 급증할 전망 개인의 경험, 인식, 선호 등 인지적인 정보 유통 증가
시계열성 (Trends)	현시점 뿐만 아니라 과거 데이터의 유지로 시계열적인 연속성을 갖는 데이터의 구성 과거, 현재, 미래 등 시간 흐름상의 추세분석 가능
결합성 (Combination)	의료, 범죄, 환경, 안보 등 타 분야, 이종 데이터의 결합으로 새로운 의미의 정보 발견 실제 물리적인 결합 이전에 데이터의 결합을 통한 사전 시뮬레이션, 안전성 검증 분야 발전 가능

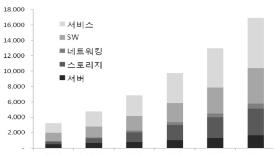
데이터 발생 원천의 증가와 빅데이터를 처리할 수 있는 기술의 발전이라는 두 가지 요인이 일반적으로 논의되는 빅데이터의 등장 배경이다. 데이터 발생 원천의 증가는 각종 센서의 발달과 상대적 가격 하락으로 인한 적용의 증대, 스마트폰을 비롯한 스마트 기기의 증가, 그리고 소셜미디어의 확산을 의미하고. 하둡(Hadoop)을 비롯한 빅데이터 처리 기술의 발전이 빅데이터가 시장에서가치를 주목 받기 시작한 기본적인 요건을 제공해 주었다[5].

기업의 입장에서는 소설 데이터와 유사한 유형의 비정형 데이터뿐 만 아니라 기존의 제조 장비 등에서 발생하는 데이터, 기존의 정보시스템에서 발생하는 데이터 등 정형의 데이터도 빅데이터화 하고 있고, 이를 수집,

저장, 처리하는 것이 중요한 이슈로 떠오르고 있다. 뿐만 아니라 최근에는 정형과 비정형, 기업 외부에서 발생한 데이터와 내부에서 발생한 데이터를 결합하여 더 고도화 된 함의를 추출하려는 노력도 이루어지고 있다. 81.

2.2 글로벌 빅데이터 시장

IDC는 세계 빅데이터 시장이 2010년 32억 달러에서 2013년 97억 달러. 2015년 169억 달러 규모로 연평균 약 40% 성장할 것으로 전망하고 있으며 이는 전체 정보통 신기술 시장 성장률의 약 7배에 달하는 수치이다. 부문별 성장률을 보면 스토리지 부문이 가장 빠르게 성장할 것 이고, 2015년에는 소프트웨어와 서비스 부문이 빅데이터 시장의 약 66%를 차지할 것으로 전망된다. 빅데이터 시 장 전망을 살펴보면 (그림 1)과 같다. 실제로 구글, 페이 스북, 아마존, 야후 등의 소셜 미디어나 인터넷업체들이 빅데이터를 마케팅 도구로 활용하며 빅데이터 기반의 소 셜 분석(Social Analytics)의 효과를 입증하고 있으며 이러 한 성과를 지켜보던 기존 공공기관과 민간 기업들도 이 제 빅데이터와 고급분석이라는 명제 하에 빅데이터의 도 입과 활용을 서두르고 있다. 하둡(Hadoop)이나 NoSQL 등의 빅데이터 오픈소스 솔루션과 빅데이터가 강화된 상 용 솔루션 등이 엔터프라이즈 환경에 적용되는 사례가 증가되면서 빅데이터는 엔터프라이즈 솔루션으로 빠르 게 자리를 잡을 것으로 전망된다[9].



(그림 1) ^{2C13} 빅데이터 시장 전망(10) (단위: 백만달러) ²⁰¹⁵ (Figure 1) Big data market trend (10) (unit; million dollar)

2.3 기업의 빅데이터 추진 현황

글로벌 IT기업들은 기존 DW/BI 기술에 빅데이터 영역을 접목하거나 빅데이터 전문 기업을 인수 합병하며 빅데이터의 프로세스 전 과정을 아우르는 솔루션과 서비스

제공을 목표로 하고 있다. (표 2)는 IT 글로벌 기업의 빅데이터 추진 상황을 잘 보여주고 있다.

(표 2) IT 글로벌 기업의 빅데이터 추진 현황(11) (Table 2) Big data status of global IT companies(11)

회사이름	추진내용
111111111111111111111111111111111111111	세계적인DB업체 '하이페리온社'를 인수로 분석
	기술 확보
Oracle	오라클 Big Data Appliance(CDH탑재) 제품 출시
	Endeca, Exalytics 등 빅데이터 분석 솔루션 출시
	BI솔루션업체 'Vertica'와 기업용 검색엔진업체
	'Autonomy' 인수 Autonomy에서 제공하는정보처
HP	리 레이어와 버티카의 고성능 실시간 분석 엔진
	의 조합을 토대로 빅데이터 인프라 서비스 제공
3.5	Hadoop on Window, Hadoop on Azure 출시 예정
Microsoft	Hortonwork의 Hadoop 탑재
	분석용 데이터 저장 관리 업체 'Netezza', 데이
	터 통합 업체 'Essentaicl', 분석 솔루션 업체
IBM	'Cognus' 등 비즈니스 분석 관련 업체 인수
	빅데이터솔루션 : InfoSphere Biginsight(Hadoop),
	InfoSphere Streams
	'HPA(High Performance Analytics) 기반의 SEMMA
	방법론'제시
SAP	'IT+분석+비즈니스' 통합플랫폼 구현(SAS 빅데
52 H	이터 분석플랫폼 SolutionMAP)
	인 메모리 컴퓨팅 기반의 어플라이언스 HANA
	출시
	Enterprise Intelligence Platform보유
Palantir	미국정부 및 월가 금융사들 중심의 대형범죄분
	석, 사기분석, 재난구조 등의 사업 레퍼런스 확보
	데이터웨어하우징 및 비즈니스 인텔리전스(BI)
Teradata	전문 업체
	비정형데이터의 고급분석・관리솔루션업체 인
	수(Ester)
	'Ester MapReduce Platform'제시 데이터저장부터 관리, 분석까지 빅데이터와 관
EMC	한 모든 것을 제공하기 위해 Greenplum, Isilion
	등 빅데이터 솔루션업체 및 데이터 관련 다수
	중 빅데이터 출구선법제 및 데이터 관련 나구 업체 인수
	[日에 린]

글로벌 IT기업들이 국내 빅데이터 시장 진출을 선언하면서 국내 SI업체들도 관련 솔루션 출시와 빅데이터 사업 참여를 본격화하기 시작했다(표3). 자체 솔루션을 가지고 사업 적용 사례를 확대하며 역량을 강화하고 있는 일부 국내 IT기업을 제외하고는 다수의 국내 빅데이터 기업들은 상대적으로 경쟁력이 약한 편이다. 이를 극복하기 위해 국내 빅데이터 기업들은 자체 네트워크를 기반으로 한 경쟁력 강화를 모색하고 있다[12].

'빅데이터 솔루션 포럼(BIGSF)'에서는 국산 중소 전문

S/W기업 간의 동맹 관계를 발판으로, 빅데이터 수집, 저장, 처리, 분석, 표현 등의 단계별로 각 기업의 특성화된 솔루션 및 서비스를 제공하는 것을 목표로 한다. 이를 위해 솔루션 간 연동, 표준화 작업 등을 기반으로 한 빅데이터 통합서비스인 '싸이밸류 얼라이언스(Cyvalue Alliance)'를 출시하고자 분주히 움직이고 있다. 그 외에도 BI 전문기업들이 모여 BI 플랫폼을 공동 개발하는 'BI포럼', 각계 전문가들이 모여 빅데이터 산업발전과 경쟁력 강화를 논의하는 '빅데이터 포럼', 공공분야에서의 빅데이터 가치창출을 도모하는 '빅데이터 국가 전략 포럼'등 산・학・연 차원에서의 다양한 활동들이 이뤄지고 있다. 소설 데이터 분석 서비스 분야의 경우, 가장 활발하게 관련기업들이 증가하는 반면, 제한적 기술 역량에 따른 서비스의 한계, 전문 인력의 부족, 출혈 경쟁 등에 대한 우려도 함께 커지고 있다[13].

(표 3) 국내 IT 기업의 빅데이터 추진 현황(14) (Table 3) Big data status of domestic IT companies(14)

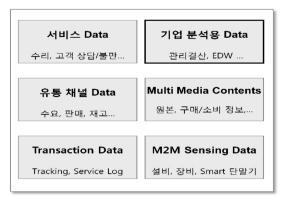
	•	
LG CNS	2012년 10월 오픈소스 하둡 기반의 자체 빅데이터 플랫폼 LG CNS Smart Big data Platform과 소설 미디어 분석 솔루션 Smart SMA 출시 빅데이터 전문가로 구성된 전담조직인 AA(Advanced Analytics)사업부문 출범하여 국내 적용 사례 확대 중 2013년에는 빅데이터 어플라이언스와 빅데이터 분석 솔루션 추가 예정	
삼성 SDS	삼성지놈닷컴을 통하여 유전자 분석 서비스 를 제공 유전자정보를 분석·관리하는 바이오인포매틱 스 사업 진행 오픈소스 기반 빅데이터 분석 플랫폼 개발 중	
SK C&C	실시간 데이터 분석 솔루션 스톰 국민연금관리공단 콘텐츠관리시스템(CMS) 구 축에 SNS 분석을 적용	
KT	유전자 분석 서비스인 게놈클라우드 제공 다양한 분야 기업이 쓸 수 있는 빅데이터 분 석 서비스인 유클라우드비즈맵리듀스 출시	
SK텔레콤	소설 모니터링/분석 솔루션 '스마트 인사이트 (Smart Insight)' 제공 독자적 NLP 기반의 텍스트마이닝, 네트워크 분석 지원 빅데이터 분석엔진 'T-MR'을 T클라우드 비즈 사용자들에게 시범 서비스 형태로 무료 제공	

3. 빅데이터 기술 활용 방안

3.1 하둡(Hadoop) 기반 정형 빅데이터 분산 병 렬처리 방식을 통한 기업 데이터 처리 성능 개 선 사례

글로벌 기업들이 데이터에 의한 경영이 점점 정착되 면서 수많은 데이터가 발생하고 저장하고 있다[15]. A사 는 전 세계110여 개 나라에 있는 사업장에서 수많은 데 이터들이 발생되고 있다. 외부 소셜미디어의 데이터와 는 달리, 이러한 경영관리용 데이터는 대부분이 정형으 로 만들어져 있다. A사는 2008년부터 전 세계 법인을 통 합하여 경영관리 시스템(ERP)이 운영되며 그 데이터량이 수백TB가 된다. 기존에 발생된 데이터량도 많지만 매년 늘어나는 데이터를 보관하고 처리하는 데 많은 시간이 소요되고, 일부는 원하는 데이터를 가공하기 어려워지고 있다. 아래 (그림 2)에서 보는 바와 같이 기업 내 빅데이 터는 여러 분야에서 발생된다. 대표적으로 제품 판매 후 수리, 고객 상담/불만 정보 등 애프터 서비스(After Service) 와 관련한 데이터, 수요, 판매, 재고 등 유통 채널 관련 데이터, 운송, 배달 등과 관련된 데이터, 그리고 결산, EDW(Enterprise Data Warehouse) 등의 정형 데이터가 있다.

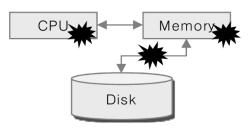
기업 내 Big Data 영역



(그림 2) 기업 빅데이터 영역(15) (Figure 2) Corporation big data domain(15)

그 중 수익성 분석은 기업 내에 발생된 거의 모든 데이터가 원천으로 평균 전사 데이터의 20% 정도가 된다. A사는 a제품의 경우 1월에 약 2시간 소요되던 수익성 분석 작업이 12월에는 8배 가까지 약 16시간이 소요되었다. 이는 시스템 처리 능력의 한계로 데이터 품질 향상을 위

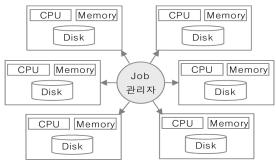
한 배부 정교화 작업의 횟수가 제한되고, 또한 상세 계 정, 상세 배부처 단위 수익성 분석이 어려워 획기적인 개 선 작업이 필요하였다.



데이터 처리 시, Disk I/O, Memory, CPU의 병목현상이 발생하여 성능 저하 발생

(그림 3) 기존 결산 시스템 구성 (Figure 3) Legacy formation of the closing account system

A사는 수익성 정보를 활용한 분석 관점 및 의사 결정 포인트가 변화되어 수익성 데이터의 정도 향상을 위해 수행하는 정교화 작업 회수가 점차적으로 증가하고 있으나 기존 시스템의 성능 한계로 적절한 서비스를 제공키어려운 환경으로 변화 되었다. 따라서, 기존 시스템의 처리 속도를 개선하기 위해 대표적인 2가지 빅데이터 처리기술을 검증하고자 하였다. 수익성 분석 작업 중 두 가지업무를 나누어, 업무 특성에 맞도록 제품별/모델별 판관비 배부 작업은 하둡(Hadoop) 기반의 플랫폼을 구축하고,나머지 배부 결과 인터페이스(Interface) 형성 및 요약 작업은 고도 병렬 그리드 데이터베이스 기기(DB machine)기반의 플랫폼을 구축하여 실증 검증(POC, Proof Of Concept)을 하였다.



(그림 4) 하둡(Hadoop) 기반 분산병렬 처리 구성 (Figure 4) Hadoop-based distributed parallel processing structure

A사의 제품별/모델별 판관비 배부 작업을 대상으로 검증 작업(POC)을 수행하였으며 아래 (그림 5)와 같이 성능 측면에서 획기적인 개선이 증명되었다. 특정 제품 하나를 대상으로 작업을 수행할 시 90%의 속도 개선이 있었고, 5개 제품을 동시에 수행하였을 경우에도 85%의 개선 효과가 나타났다. 비즈니스 로직(Logic) 처리와 데이터 정합성에도 전혀 문제가 없었으며, 하둡 기반으로 플랫폼을 변경할 경우에는 기존의 프로그램을 새로 개발하여야 하나 구현 용이성에서는 큰 문제가 없는 것으로 확인되었다.

PoC 요약

목적	• LG CNS의 SBP (<i>Hadoop 기반</i>) 적용을 통한 성능개선 수준 확인		
대상 시스템	● 연결수익성시스템(Actual CPS)		
대상 업무	• 판매 관리비의 모델별 잔액 배부 (SAS의서 수행되는 업무) • Hadoop 구현 기능 배부Factor 형성 모델별 잔액 당월금액 재형성 내부 대상 금액 행성		
Data (2011년 12월)	 1차 PoC 단독작업 성능검증 냉장고 사업부(Size 비중: 17%) 2차 PoC 동시작업 성능검증 냉장고, LTV, CAC, RAC, 세탁기(80%) 		

<u>PoC 결과</u>

결산 수행 시간 절감 및 배부 상세화 기능 검증

PoC 검증 항목		SAS	SBP	비고
성능	단독 작업 수행 (냉장고 사업부)	171분	18분	90% 개선
	동시 작업 수행 (상위 5개사업부)	275분	43분	85% 개선
Б.	Biz Logic 처리	가 능	가능	
Biz	데이터 정합성	-	원천 Data 동일	
구현	Migration 고려사항	Hadoop 기반으로 재개발이 필요		

(그림 5) 하둡(Hadoop) 기반 빅데이터 플랫폼의 기업 결산 시스템 적용 결과

(Figure 5) Results of Hadoop-based big data platform for the closing accounts system

3.2 정형 데이터 분석을 통한 대리점 이상거래 감지사례

Y사 내부 및 외부 전문 컨설팅 업체의 예상에 따르면, 2012년 한 해에 사기(Fraud)에 의한 예상 손실이 1,000억원대에 이르는 것으로 판단되었다. 또한, LTE 활성화 이후 명의 도용 의심 접수 건수가 폭증하고 있으며 이상 영업 관련 불법/편법이 확대되고 있어 예상 손실액도 증가될 것으로 전망되어 이상 영업에 대한 대응 체계를 신속히 고도화하여 대응하는 것이 사기 비용(Fraud cost)을 줄이고 고객 신뢰를 확보하기 위해 긴요한 것으로 판단되었다.

기존의 이상관리 시스템은 부정 영업 발생을 사후에 발견하고, 관리하는 시스템이었다. 이러한 관리는 일회성해결은 되나, 동일 문제가 반복적으로 발생할 리스크 (Risk)가 존재하고, 변형/신종 이슈에 대한 대응 속도에한계가 있었다. 이를 극복하기 위해 분석 메커니즘, 활용 데이터, 그리고 시스템 구조를 변경하여 시스템을 고도화하였다. 명의도용, 대출사기, 소액결재 후 재판매 등이상 상황에 대한 추정 규칙(Rule)을 보강하고, 가입자/단말기 위치 등 대용량 데이터를 활용도록 하였으며, 이를위해 시스템 구조도 하둡 클러스터(Hadoop cluster)를 활용한 분석 시스템으로 변경하였다.

또, 중장기적으로는 네트워크의 사용 데이터(log data) 및 소셜미디어 등 외부 비정형 데이터도 포함하는 차세대 이상 관리 시스템을 구축하여 의심 패턴을 상시 분석할 수 있도록 할 예정이다.



(그림 6) 차세대 이상관리 시스템 예시 (Figure 6) Example of the Next-generation warranty detection system

Y사는 빅데이터를 활용하여 구축한 이상관리 시스템

을 활용하여 2013년 한 해에만 전년에 비해 430억원 이 상의 수수료 절감 효과를 기대하고 있다. 95,000여건에 달하던 이상 개통이 19,000 여 건으로 대폭 줄어 들 것으 로 예상하고 있다.

정성적 효과로는 실시간 이상 정후를 제공함에 따라이상 영업 조기 감지와 부당 영업에 대한 거울 효과로 사전 예방 효과가 있을 것으로 보고 있다. 또한, 불량 거래점 조기 인지 및 예방 효과로 건강한 영업 채널 구축이가능할 것으로 판단하고 있다.

3.3 소셜 비정형 데이터 분석을 통한 마케팅 전략 수립

A사는 2011년에 자사의 노트북 컴퓨터 마케팅 전략을 휴대성과 디자인의 우수성을 강조하는 방향으로 진행하였다. 2012년 노트북 컴퓨터 마케팅 전략을 수립하기 위해 현재까지 추진된 전략의 적정성 검토가 필요하게 되어, 소셜미디어상의 문서를 분석하여 시장에서의 기존 전략의 적정성은 물론, 경쟁사와 장단점을 비교하고 그에 대한 대응 전략을 수립하고자 하였다. 2011년 주요 온라인 사이트 및 소셜미디어를 대상으로 3개 브랜드의 노트북에 대해 총 308,402건의 비정형 데이터를 취합하여 항목별 긍정과 부정 수준을 분석하고 그 결과를 반영한마케팅의 새로운 전략을 수립하였다.

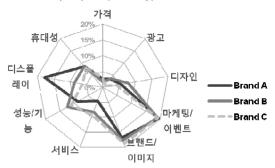
분석 결과 A브랜드는 디자인 항목과 휴대성 항목에 대해 긍정 반응이 높았고, B브랜드는 마케팅/이벤트 항목과 디스플레이 항목에 높은 긍정 반응을 보였다. 서비스 항목에 대해서는 A>B>C브랜드 순으로 높은 부정 반응을 보였다. 3개사모두 서비스에 대한 부정 반응이 많았으나, 서비스 질의 향상이 고객에게 차별화된 서비스를 지원할 수 있는 항목이라는 것도 발견되었다.

각 사의 제품 시리즈별 Buzz 총량을 분석해 보면, A사의 a, b 브랜드 모두 디스플레이에 대한 Buzz가 많고, 특히 휴대성에 대한 마케팅 효과가 b 모델로 가면서 나타나고 있음을 알 수 있었다. 경쟁사인 B, C사는 각각 브랜드와 마케팅에 대한buzz량이 많음을 알 수 있었다. 상세내용은 아래 표와 같다.

A브랜드는 디자인 항목과 휴대성 항목에 대해 집중 마케팅을 수행하고 있었으므로 자사의 의도대로 휴대성 과 디자인에 대해 동일하게 긍정 여론이 조성되었음을 확인하였다. A사는 디스플레이 항목의 높은 Buzz가 있었 으나 부정적이지도 않았지만 긍정적인 Buzz도 많지 않음

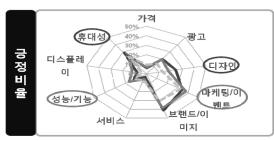
을 확인하였다. A사는 디스플레이에 대한 높은 Buzz 량을 긍정화 시키기 위한 마케팅 전략 타스크포스(Task Force)를 발족하였으며, 서비스 항목에 대한 높은 부정 비율을 감소시키기 위해 서비스 마케팅 활동을 기획하였다.

■ 브랜드 별 속성 비중 % (언론 Only)



■ 브랜드 별 속성 총량 (언론 Only)

	Brand A	Brand B	Brand C
가격	1,463	2.056	840
광고	1,882	4,436	186
디자인	3.084	9.651	1,175
마케팅/이벤트	10,700	23,336	5,464
브랜드/이미지	9,193	21,748	5,345
서비스	2,510	10,088	2,628
성능/기능	4,818	15,242	3,134
화면	9,810	10.649	1,761
휴대성	4.707	10,578	2,256



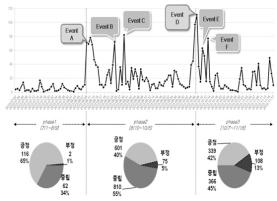


(그림 7) 노트북 Buzz 비정형 데이터 분석 결과 (Figure 7) Big data analysis result of notebook social buzz data

3.4 소셜 비정형 데이터 기반 광고효과 분석

A사는 최근 급속하게 발달하고 있는 빅데이터 처리 기술을 활용하여 인터넷상의 소셜미디어(Social Media)라할 수 있는 카페(Café), 블로그(Blog), 뉴스(News), 게시판 (BBS), 에스엔에스(SNS), 댓글(Reply) 등 외부의 데이터와 내부 고객의 목소리(Voice of Customer, VOC)를 총괄적으로 수집하고 분석하여 광고의 적절성과 개선 방안을 도출키로 하였다.

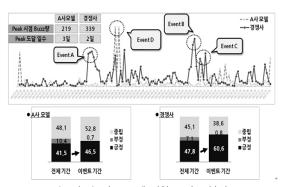
A사는 2012.7.1일부터 2012.11.18일까지 약 5개월간의 소셜 미디어상의 버즈를 수집하여 분석하였다. 자사의 버즈는 물론, 경쟁사의 버즈도 동시에 수집하여 광고의 영향도를 평가하였다. A사는 자사 광고 모델에 대한 비정형 데이터 분석 결과 Buzz의 양 자체는 증가하고 있으나 시간 경과에 따라 모델과 제품의 부정 반응이 동시에 증가하고 있음을 발견하였다. 특별한 이벤트가 없었던 초기 한달 간에는 긍정비율이 65%, 부정 비율은 1%에 그쳤다. 그러나 모델의 이벤트가 증가하면서 긍정이 40~42%로 줄어든 반면, 부정은 5%~13%까지 급속도로 증가 하였다. 상세 내용은 아래 (그림 8)에 표시되어 있다.



(그림 8) 광고 모델 Buzz량 분석 및 긍부정 변화 추이 (Figure 8) social buzz analysis of the advertising model and positive/negative trends

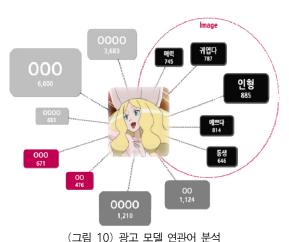
A사는 경쟁사의 광고 모델에 대한 버즈량을 분석하였다. 피크(Peak) 시점의 버즈량과 피크 도달 일수를 비교하였으며, 유사한 이벤트에 대한 광고 모델의 영향력을 분석하였다. 분석 결과 피크 시점의 버즈량이 A사는219개였으나 경쟁사는 339개였다. 피크 도달 일수도 A사는3일이 소요되었으나 경쟁사는 2일이 소요되었다. A사모델의 전체기간 긍정 비율은 41.5%로 경쟁사의 47.8%

에 비해 다소 열세에 있었으나, 이벤트 기간에는 46.5%에 비해 경쟁사는 60.6%로 그 격차가 더 벌어져 상대적으로 경쟁사 모델이 더 높은 영향력이 있는 것으로 판단되었다. 상세 내용은 아래 (그림 9)와 같다.



(그림 9) 광고 모델 영향도 비교 분석 (Figure 9) Comparison results of the effectiveness of advertising models

A사는 원인 분석을 위해 모델과 연관된 언어를 분석 하였다. '인형같다', '귀엽다', '예쁘다', '동생같다', '매력 적이다'라는 의견이 다수였다.



(Figure 10) Related words analysis of advertising models

그러나, A사의 브랜드에 대한 이미지는 (그림 11)과 같이 다르게 나타났다. '1위', '화려하다', '우아하다', '여성적', '고급'이라는 이미지가 크게 자리잡고 있음을 발견하였다. 결국, 상세 연관어 분석 결과 제품과 모델 컨

셉이 불일치 함을 확인하고 광고에 노출되는 모델의 이 미지를 변경하게 되었다. 기존의 귀엽고 깜찍한 이미지 의 광고를 고급스럽고 전문가의 이미지를 강조하는 방향 으로 수정되었다



(그림 11) 제품 연관어 분석

(Figure 11) Related words analysis of the product

4. 결 론

본 연구를 통해 소설데이터 및 장비에서 발생하는 로그데이터 등의 비정형데이터 외에 기업이 기존에 보유・활용하고 있던 정형의 데이터 중에도 빅데이터가 있음을확인하였고, 이를 기업의 가치 창출에 활용하기 위해 빅데이터 기술을 적용하였을 때 나타나는 효과를 검증할수 있었다. 또한, 현재 빅데이터의 실무적 활용 영역에서화두가 되고 있는 내・외부의 정형, 비정형 빅데이터의 연계를 통한 빅데이터 활용 사례도 대리점 이상 영업 조기 발견 시스템에서 확인되었다.

지금까지 학계의 연구나 업계에 회자되는 빅데이터 관련 내용들은 현황과 관련된 자료정리의 수준을 벗어나지 못하는 것이 대부분이었고, 인용된 사례도 앞서 언급한 바와 같이 피상적이며 단편적인 것이 대다수를 차지하였다. 본 연구의 공헌점은 이러한 기존 연구의 한계를 극복하기 위해 실제 기업에서 활용되는 빅데이터의 사례를 심도 깊게 다루어 그 함의를 도출하였으며, 사례의 내용에서도 비정형 뿐 아니라 정형 데이터의 사용까지 확대하여, 빅데이터의 활용 영역에 대한 기존의 편협 된 시각을 개선하는데 도움을 줄 수 있도록 한 것이다. 또한, 향후 빅데이터의 기업 내 활동이 나아갈 수 있는 지향점으로써 내・외부에서 발생하는 정형데이터와 비정형데이터를 적절히 결합함으로써 분석의 효과를 극대화 할

수 있음을 보여주는 사례를 소개하였고, 이를 통하여 빅데이터 처리를 통한 기술적 체계(architecture) 또한 전통적인 관계형 데이터베이스와 빅데이터 처리를 위한 하둡(Hadoop) 등의 신기술이 융합된 복합적 기술체계(Hybrid Technical Architecture)가 필요함을 보여 주었다.

참고문 헌 (Reference)

- [1] Ji-Sook Kim, Practical use and analytic method of big data, Korea Univ., 2013, p4
- [2] Ha Yeon editorial department, Market prospect of bigdata and DBMS, 2012, Ha Yeon
- [3] Database White Paper, Big data analytics trends, 2013, Korea Database Agency
- [4] Thomas M. Philip, Sarah Schuler-Brown, Winmar Way, A Framework for Learning About Big Data with Mobile Technologies for Democratic Participation: Possibilities, Limitations, and Unanticipated Obstacles, 2013, Technology, Knowledge and Learning Volume 18
- [5] A new possibility and matching strategy for big data, 2011.11, National Information Society Agency
- [6] Mckinsey, Big Data: The Next Frontier for Innovation, Competition, and Productivity(2011)

- [7] Database White Paper, Big data analytics trends, 2013, Korea Database Agency
- [8] Ling Liu. Computing infrastructure for big data processing, 2013, Frontiers of Computer Science Volume 7
- [9] Challenges and Opportunities with Big Data, A community white paper developed by leading researchers across the United States, 2010, www.cra.org/ccc/files/docs/init/bigdatawhitepaper.pdf
- [10] Worldwide Big Data Technology and Service Market Forecast, IDC, 2011
- [11] Press report(ITworld, 2013 Big data market trends 5 Big data companies, 2013)
- [12] Database White Paper, Big data analytics trends, 2013, Korea Database Agency
- [13] Database White Paper, Big data analytics trends, 2013, Korea Database Agency
- [14] Press report(LG CNS Press release, LG CNS first big data integrated solution 'Smart Big Data Platform', 2012)
- [15] Chris Yiu. The Big Data Opportunity, 2012, Policy Exchange
- [16] Advanced Analytics Insight Report, LG CNS Advanced Analytics Business Unit

① 저 자 소 개 ①



이 재 성 1985년 고려대학교 산업공학과(학사)

2009년~현재 한신대학교 정보통신학부(박사과정) 2001년~2007년 LG CNS 중국 법인장(상무)

2008년~현재 LG CNS 사업부장(전무)

관심분야 : 정보시스템, 빅데이터

E-mail: jsunglee@lgcns.com



홍 성 찬

1983년 고려대학교 통계학과(학사) 1990년 게이오대학 관리공학과(석사)

1994년 게이오대학 관리공학과(박사)

1994년~1995년 LG CNS 컨설팅사업부 책임연구원 1995년~1997년 상명대학교 정보처리학과 전임강사

2010년~2011년 한국인터넷정보학회 회장

1997년~현재 한신대학교 정보통신학부 교수

관심분야 : 정보시스템, 빅데이터 E-mail : schong@hs.ac.kr

_ -----