

국방분야 빅데이터 분석의 활용가능성에 대한 고찰

김성우 · 김각규[†] · 윤봉규
국방대학교 운영분석학과

A Study on a Way to Utilize Big Data Analytics in the Defense Area

Seong-Woo Kim · Gak-Gyu Kim · Bong-Kyu Yoon

Department of Operations Research, Korea National Defense University

■ Abstract ■

Recently, one of the core keywords in information technology (IT) as well as areas such as business management is big data. Big data is a term that includes technology, personnel, and organization required to gather/manage/analyze collection of data sets so large and complex that it becomes difficult to manage and analyze using traditional tools. The military has been accumulating data for a long period due to the organization's characteristic in placing emphasis on reporting and records. Considering such characteristic of the military, this study verifies the possibility of improving the performance of the military organization through use of big data and furthermore, create scientific development of operation, strategy, and support environment. For this purpose, the study organizes general status and case studies related to big data, traces back examples of data utilization by Korean's national defense sector through US military data collection and case studies, and proposes the possibility of using and applying big data in the national defense sector.

Keywords : Big Data, Big Data Analytics Tools, 3V, Cheap Data

1. 서 론

빅데이터는 최근 우리사회의 산업, 과학, 예술분야에서 공공부문과 민간부문을 불문하고 모든 영역에 변화를 몰고 오고 있는 가장 주목받는 키워드(Key Word)이다. 2010년 2월 The Economist의 특집호 'The data deluge[33]'에서 데이터의 홍수라는 표현으로 현실 세계의 데이터양의 방대함을 기사화하면서 그 관심이 커졌고, 2011년 5월 맥킨지 글로벌 인스티튜트가 발표한 보고서 '빅데이터 : 혁신, 경쟁, 생산성을 위한 차세대 프론티어(Big data : The next frontier for innovation, competition, and productivity) [29]'에서 최초로 키워드에 '빅데이터'란 용어를 사용하면서 이제는 이 시대를 대표하는 유행어이자 고유명사화 되었다. 실제로 2012년 12월 IDC(International Data Corporation) 보고서에 의하면 2012년 한 해 동안 인류가 쏟아낸 데이터 양은 2.8 ZB(Zetta Bytes, 10^{12} Giga Bytes)이며, 2020년에 이르면 대략 40 ZB가 될 것이라고 발표하였다[17]. 그리고 관련 시장규모 역시 매년 40% 성장해 2012년 50억 달러, 2013년 100억 달러, 2017년에는 530억 달러에 이를 것으로 전망하고 있다. Gartner[17]는 2013년 전략적 기술 트렌드 10선에서 앞으로 기업은 비정형 데이터 및 외부 데이터 소스에 집중하게 될 것이라고 전망하였다. 또한 하둡(Hadoop)과 NoSQL의 발전에 따라 소셜 네트워크를 통한 빅데이터의 활용이 증대되고, 클라우드, 패키지 분석 등의 빅데이터 트렌드가 가속화될 것으로 전망하고 있다.

실제로 민간부문은 데이터가 증가하였고, 이를 활용할 수 있는 도구들도 많아졌다. 정부, 공공기관, 민간 기업들은 이제 데이터 없이 그 조직을 운영하기 힘들어진 시대가 되었다. 군사분야 또한 과거에는 병력의 수나 총알과 포탄 보유량 등 전쟁 당사국의 군사력이 전쟁의 승패를 좌우하는 결정적인 역할을 한다고 인식하였으나, 이제는 전장에서 수집되는 적 상황분석, 지형정보판단 등 각종 체계와 장비 등을 통해 획득한 정보데이터를 분석하여 지휘관의 의사결정 능력을 향상시키고 그 결과 전쟁에서 유리한 위치

를 보장받게 하는 분석 능력이 전쟁의 중요한 요소로 주목받고 있다[15]. 어느 때보다 많은 데이터를 보유하고 있는 한국군도 폭발적으로 증가하는 데이터의 관리와 활용을 위해 관심을 갖게 되었다. 북한을 비롯하여 동북아 주변국 간의 군비경쟁이 날로 심각해지고 있는 현실에서 주위에 널려 있는 수많은 데이터로부터 누가 더 빨리 유용한 통찰력과 지식을 찾아내 활용하느냐도 정부와 민간기업이 그러하듯 주변국가와 소리없는 안보경쟁에서 우위를 차지하게 되는 밑바탕이 될 수 있다. 하지만 국방분야의 이러한 관심은 민간분야의 추세를 따라가지 못하고 있는 것이 사실이다. 일부 수집해 놓은 데이터를 활용하여 새로운 가치와 통찰력(Insight)을 얻기 위한 노력이 시도되고는 있지만, 아직까지 빅데이터에 대한 인식수준이 미흡하여 국방분야에서 생산 및 유지되는 데이터가 빅데이터 분석을 위한 양질의 데이터가 되지 못하고, 빅데이터를 분석하기 위한 IT 기술과 전문 인력 역시 부족한 게 현실이다. 따라서 본 연구의 목적은 군에서 빅데이터 활용을 통한 가치 창출 흐름의 속도를 높이자는 취지에서 빅데이터의 이해도를 높이고, 국내외 사례 및 현황에 대한 분석을 통해 국방분야에서 빅데이터 분석의 적용 가능성을 탐색해 보는 것이다.

본문의 구성은 제 2장에서 빅데이터의 정의와 빅데이터의 특성, 빅데이터의 부각이유에 대하여 알아보고, 제 3장에서는 빅데이터 분석을 위한 기반 기술과 분석방법을 중심으로 빅데이터 활용사례를 공공분야와 민간분야로 나누어 정리해 보았다. 제 4장에서는 선진군의 데이터 수집 및 관리, 분석 후 활용하기 위한 체계적인 절차와 노력 등을 소개하고, 한국군의 빅데이터 활용실태를 분석 후 이를 토대로 국방분야의 효율성 제고와 가치창출을 유도할 수 있는 빅데이터 활용의 방향성을 제시하고자 한다.

2. 빅데이터의 개념과 등장배경

2.1 빅데이터 vs. 기존 데이터 분석

현재 많은 전문기관 및 각종 연구에서 빅데이터에

대해 소개와 정의를 하고 있으나, 이러한 보고서에서 빅데이터의 의미는 거의 비슷하다. 먼저 MacKinsey [29]에 따르면 ‘빅데이터란 일반적 데이터베이스 SW가 저장, 관리, 분석할 수 있는 범위를 초과하는 규모의 데이터’라고 정의하였고, IDC[17]는 ‘다양한 종류의 대규모 데이터로부터 낮은 비용으로 가치를 추출하고 데이터의 초고속 수집, 발굴, 분석을 지원 하도록 고안된 차세대 기술 및 아키텍처’라고 하였으며, Gartner[18]는 ‘빅데이터를 향상된 시사점과 더 나은 의사 결정을 위해 사용되는 비용 효율이 높고, 혁신적이며, 대용량, 고속 및 다양성을 특성으로 가진 정보 자산’으로 보았다. 한국 정보화 진흥원[1]에 따르면 문자 그대로 굉장히 큰 양의 데이터를 전제하며, 이때의 ‘데이터’란 처음부터 특정 형식과 목적에 맞춰 작성된 구조화된 데이터만이 아니라, 비구조화 데이터인 SNS 메시지나 구글검색, 블로그, 이미지, 동영상 등이 분석과 활용의 대상으로 포함된다는 것이라 정의하였다. 또한 광의의 의미의 빅데이터는 좁은 의미의 빅데이터 개념에 인적자원, 조직, 데이터 처리·축적·분석기술(하둡, NoSQL, 기계학습, 통계분석 등)을 합친 개념이라고 기술하였다. 삼성경제연구소에서는 2010년 ‘빅데이터’란 기존의 관리 및 분석 체계로는 감당할 수 없을 정도의 거대한 데이터의 집합을 지칭하면서 대규모 데이터와 관련된 기술 및 도구(수집·저장·검색·공유·분석·시각화 등)도 빅데이터의 범주에 포함된다고 기술하였다. 이 밖에도 수없이 많은 전문기관, 각종 보고서 및 글에서 빅데이터를 정의하고 있다. 내용을 종합해보면 빅데이터는 ‘기존 처리방식을 적용하기 어려운 크고 복잡한 성격의 데이터로써 대용량 데이터를 분석해 가치 있는 통찰과 지식을 얻고자 하는 모든 노력들의 포괄적 개념’이라고 정리할 수 있겠다.

이상에서 살펴본 빅데이터 정의의 공통점을 보면 대용량의 정형·비정형 데이터로 요약할 수 있다. 그러나 대용량의 정형·비정형 데이터는 이전에도 존재해왔다. 다만 현재 빅데이터에 대한 관심이 집중되고 있는 것은 ‘데이터 수집 비용’과 ‘분석된 정보의 가치’ 두 가지 측면에서 기존의 데이터 분석

과 차이가 있다는 것이다.

우선 ‘데이터 수집 비용 측면’을 살펴보자. 정보화가 진행됨에 따라서 대부분의 일상이 우리가 인식하고 있던 그렇지 않던 간에 정보기와 상호작용하는 빈도가 높아졌다. 특히, 스마트폰의 사용으로 대부분의 사람들은 하루 종일 뭔가를 입력하고 다른 사람이 입력한 정보를 소비한다. 이 과정에서 자연스럽게 데이터가 축적되고 있으며 이렇게 축적된 데이터는 원래 의도를 가지고 축적한 데이터가 아니므로 기회비용이 없거나 매우 낮으므로 데이터를 축적하기 위한 비용이 거의 들지 않는다. 마치 서울에서 부산까지 출장을 가는데 옆에 있는 동료가 서류 한 장 전달해 달라고 할 때 크게 고민하지 않고 수락할 수 있는 것과 같다. 낮은 비용으로 데이터가 축적될 수 있기 때문에 대용량의 데이터가 축적이 가능해졌으며, IT 기술의 발전에 따른 연산비용의 감소는 데이터를 저장하고 유지하는 비용도 감소시켜 빅데이터가 출현할 수 있는 자연스러운 기반을 제공했다.

다음으로 ‘정보의 가치 측면’에서 고객이나 시장의 행동 패턴에 대한 예측이 창출하는 가치가 과거에 비해 상대적으로 높아졌다. 특히 넷플릭스(Netflix)나 아마존(Amazon)과 같이 온라인 비즈니스가 기업의 핵심 시장이 되는 Web 2.0 기반의 기업이 성장함에 따라서 고객 정보와 이를 기반으로 한 시장에 대한 분석적 예측의 가치는 충분히 입증되었으며 기존의 기업들도 이를 경쟁적으로 따라하고 있는 실정이다. 그 결과 분석적 예측이 기업의 핵심 경쟁요소로 변화하고 있으며 중요성이 높아지고 있다. 분석적 예측의 중요성은 분석된 정보의 가치의 증가로 자연스럽게 이어지며 빅데이터의 축적과 활용을 가속화시키고 있다.

정보의 가치는 높아지고 정보를 생산하기 위한 기초 데이터의 생산·저장·처리 비용은 낮아져서 누구라도 데이터를 이용하여 높은 수준의 가치 창출이 가능해진 것이 기존의 데이터 분석과 빅데이터의 차이점이다. 또한 누구나 분석이 가능하므로 데이터를 이용한 분석이 경쟁우위가 아니라 경쟁을 위해서 필요한 기본적인 능력으로 자리매김해 가고

있는 것 또한 기존의 데이터 분석과의 차이점이다.

불과 몇 년 전만해도 비즈니스 인텔리전스(Business Intelligence, BI)라는 이름으로 많은 기업들이 경영효율화를 위하여 데이터를 활용해 왔다. 예를 들어 미국의 대형 슈퍼마켓 체인 Wal-Mart에서는 시간당 약 100만 건의 트랜잭션을 처리하며 약 2.5 페타바이트의 데이터를 자사의 데이터 웨어하우스(Data Warehouse)에 축적하여, 생성되는 데이터를 분석해 재고와 가격 설정을 최적화하는 등의 노력을 하였다. 그렇다면 빅데이터와 BI 분석의 차이점은 무엇인가? ‘빅데이터의 특성’ 차원에서 차이점을 살펴보면, 첫째, 기존 매출이나 재고량 같은 수치 데이터, 즉 구조화된 데이터가 아니라 클릭스트림(Click Stream¹⁾) 데이터나 소셜 미디어의 데이터 혹은 센서 데이터처럼 기존의 관계형 데이터베이스 위주와는 다른 구조화되지 않은 데이터(비정형 데이터) 영역의 분석에 더 관심을 집중하고 있다는 것이다. 둘째, 데이터양도 기존보다 클 뿐만 아니라 데이터의 다양성과 데이터의 발생 빈도 면에서도 차이가 있다. 셋째, 실시간 발생한 많은 양의 데이터를 실시간 처리하는 능력이 있다. 마지막으로 가장 큰 차이점은 과거 일반 기업들이 간과했던 로그 데이터 등과 같이 버려지는 데이터로부터 빅데이터를 적극 분석함으로써 새로운 가치를 창출해 냈다는 점이다. 이러한 속성을 종합하여 미국 IT 분야의 시장조사 및 연구 컨설팅 업체인 Gartner[18]는 이른바 빅데이터의 특성을 ‘다양한 종류로(Variety) 빠르게 흐르는(Velocity) 막대한(Volume) 데이터’, 즉 ‘3V’로 정립하였다.

‘3V’에서 Volume을 나타내는 ‘데이터 규모’란 빅데이터라 불리기 위해서 일단 데이터 집합의 크기가 커야 한다는 것이다. 그러면 어느 정도나 큰 데이터를 ‘빅데이터’라 부를 것인가? 앞에서 제시한 빅데이터의 정의로 본다면 기존 기술로는 관리할 수 없는 데이터양으로 대략적으로 적게는 수 테라

바이트(Terrabyte, TB)에서 많게는 수 페타바이트(Petabyte, PB) 정도 크기의 데이터 집합을 지칭하는 것이 일반적이다[3, 6]. 그러나 단순히 큰 규모만으로 빅데이터라 할 수 없을 것이다. Variety, 즉 ‘다양성’이 있어야 한다는 것이다. 과거에는 데이터들이 특정 형식에 맞춰 잘 정리된 정형적, 또는 구조적 데이터들로 구성되어 있다면 최근 생성되는 데이터들은 미리 형식을 하나하나 정해놓기 어려운 것이 더 많으며, 실례로 일상생활에서 이용하는 데이터들을 생각해 보면 인터넷 뉴스기사, 블로그나 온라인 커뮤니티 게시판의 게시물, YouTube 등에 올라와 있는 동영상, 팟캐스트, 음악, 사진 등 데이터의 범위를 한정할 수 없을 정도라 하겠다. IDC는 발표한 연구보고서에서 비정형, 또는 비구조적 데이터가 차지하는 비율이 90% 이상이 될 것이라 전망하였다[16]. 또한 데이터 발생 빈도와 갱신 빈도로 표현하는 Velocity(속도)도 빅데이터의 특징을 보여주는 중요한 특성이다. 불과 20여 년 전만 해도 데이터 통신망의 설치 및 사용을 위해서는 매우 비싼 요금을 지불하였으나 이제는 가정마다 초당 100메가바이트 이상의 유선 네트워크 장비가 있고, LTE 이상급 무선 네트워크도 보편화 되어가고 있다. 이와 같이 우리 사회는 데이터 생성, 이동 및 처리 속도가 눈부시게 빨라지고 있다.

3V로 정의되는 빅데이터의 특성은 데이터 수집/저장/처리 비용의 감소로 설명할 수 있다. 낮은 비용으로 정보가 수집되므로 데이터가 특정한 목적으로 재가공된 것이 아니고 원래 정보처리를 목적으로 수집/저장된 데이터가 아니므로 다양한 형태로 존재할 수밖에 없다. 이런 특징이 Variety를 설명할 수 있다. 또한 낮은 비용으로 생산되는 데이터이므로 대용량의 데이터가 생성될 수 있으며 이런 특징이 Volume으로 나타난다. Velocity는 생성되는 속도와 관련되는 것으로 Velocity가 클수록 Volume이 증가하는 시간이 짧아지므로 이를 Volume과 동일한 개념으로 생각할 수 있으며, 독립된 개념으로 생각하더라도 저장과 처리비용의 감소에서 비롯된 것이므로 빅데이터의 특성은 처리 비용 감소에 근원적인 뿌리를 가지고

1) 클릭스트림(Click Stream) : 한 사람이 인터넷에서 보내는 시간 동안 방문한 웹사이트를 기록한 것.

있다. 따라서 빅데이터를 한마디로 요약하면 값싼 데이터, 곧 'Cheap Data'인 것이다.

2.2 빅데이터 분석의 등장 배경

Anderson[7]은 지금을 'an era of massively abundant data', 즉 매우 엄청난 양의 데이터를 갖는 시대로 정의하면서 어떤 사건에서 생성된 많은 양의 데이터를 통해 얻은 상관관계가 과거 일부 데이터를 통해 얻은 이론에서 설명할 수 없는 것들을 설명할 수 있다고 하였다. 다시 말하면, 빅데이터 분석이 지금까지 대부분의 문제해결을 일부 데이터로부터 전체를 예측하는 샘플링 기반의 귀납적 모델방식에서 범하기 쉬운, 예외 케이스를 놓치는 문제로부터 해방시킬 수 있다는 것이다. 빅데이터가 과거 이론을 대체할 수 있을 만큼 효과가 크다는 것이다. 그 근거로 수많은 번역 데이터를 토대로 통계적인 상관 규칙을 이용한 구글의 번역 시스템이 다른 번역 시스템보다 우수한 것을 사례로 제시하고 있다. 또한 Graham[21]은 사람들이 관심을 갖고 있는 사회의 질문에 대하여 언젠가 빅데이터의 충분한 양으로 모두 대답할 수 있는 시점이 올 것이라 하였다.

빅데이터가 관심을 끌게 된 배경의 첫 번째는 '데이터의 폭발적 증가'이다. 2007년부터 전 세계적으로 생성된 디지털 정보량이 사용가능한 저장 공간(Available Storage)을 초과하기 시작했다[33]. Mackinsey[29]에 따르면 전 세계 데이터 매년 40% 증가하고 있으며, 미국 MIT에서는 모바일 기기, 온라인 상거래, 소셜 네트워크 서비스(SNS) 등에서 하루에 250 TB 분량의 비정형 데이터가 생성되고 있다고 발표하였다. 이해를 돕기 위해 데이터의 크기를 대략 가능하면, 한 스푼에 가득 담긴 모래알의 양을 1메가바이트(MB, Mega Byte) 데이터라고 하였을 때, 1기가바이트(10^9 bytes)는 생수통 절반을 채울 수 있는 모래의 양이며, 1테라바이트(10^{12} bytes)는 85m² 국민 주택 아파트에 10cm 깊이의 모래의 양이며, 1제타바이트(10^{21} bytes)는 한반도 모든 백사장 모래 합을 1,000배에 해당한다[6].

이러한 빅데이터가 발생하게 된 이유는 크게 컴퓨팅 가격으로 대표되는 데이터를 처리/분석할 수 있는 '인프라 비용의 감소'와 '데이터를 수집하는 비용의 감소'로 구분할 수 있다. 인프라 비용의 감소는 데이터 저장 및 처리비용의 하락, 컴퓨팅 파워의 증가에 따른 처리 시간의 감소로 설명할 수 있다. 데이터 수집비용의 감소는 SNS, 그림자 정보, 사물간 인터넷(Internet of Things) 등으로 인해 시장이나 설비의 상호 작용과 관련된 데이터들이 거의 아무런 비용 없이 지속적으로 생성되고 있는 현상으로 설명할 수 있다.

두 번째 빅데이터 부각이유는 '실시간 취합 및 분석이 가능한 기술 환경의 도래'이다. 과거에는 일정 규모의 데이터만 처리가 가능했고 실시간 빅데이터 처리는 불가능했으며, 분석에 상당한 비용과 시간이 소요되었다. 하지만 현재는 클라우드 기술의 등장으로 대용량의 데이터 분석·저장이 실시간 가능해졌고 대규모 데이터에서 특징과 패턴을 추출하는 데이터 분석기술, 다양한 데이터를 기존의 특징·패턴과 조합하는 기술이 실제 활용 가능해진 것이다.

세 번째는 '비즈니스 및 경쟁 환경 변화'이다. 과거의 추세를 토대로 소비자의 행동을 예측하는 통계분석에 의한 기존 마케팅은 급격하게 변화하는 시장 환경을 반영하기에는 어렵다. 특히 인터넷과 SNS 발달에 따라 소비자의 기호가 빠른 속도로 변화하는 최근의 시장 환경을 기존의 수단으로는 대응하기 어렵다. 특히, 경기침체, 환경의식 고조, SNS의 인기상승, 소비자 영향력 증대, 글로벌화 등으로 사업 환경이 급변하면서 소비자들이 기업의 마케팅 정보보다는 SNS를 통한 타 소비자들에 제공하는 정보를 더 신뢰하기 시작하였기 때문이다. 따라서 기업들은 실시간적인 개별 소비행동의 분석이 필요하게 되었으며, 이는 빅데이터 분석을 통해 가능하다. 이처럼 빅데이터는 기존에는 불가능했던 의미 있는 시장 정보를 실시간으로 도출하고 소비자의 취향과 행동 등 다양한 변화를 실시간으로 분석하여 공급자가 신속히 의사 결정을 할 수 있게 함으로써 산업구조와 경쟁양상을 변화시키고 있으며, 아마존이

나 넷플릭스 같은 인터넷 기반 기업을 필두로 빅데이터의 활용을 경쟁우위를 넘어 시장 생존을 위해 반드시 갖추어야 하는 요소로 변화시키고 있다. 앞서 설명한 빅데이터 분석의 등장배경을 정리하면 <표 1>과 같다.

<표 1> 빅데이터 분석의 등장 배경

배경	내용
비용 감소 → 데이터 증가	<ul style="list-style-type: none"> 인프라 비용 감소 <ul style="list-style-type: none"> 데이터 저장 및 처리비용의 하락 컴퓨팅 파워 증가에 따른 처리 시간의 감소 데이터의 수집 비용 감소 <ul style="list-style-type: none"> 시장이나 설비의 상호작용과 관련된 데이터들이 추가비용 없이 지속적 생성 ex) SNS, IoT(Internet of Things)
분석기술 환경의 도래	<ul style="list-style-type: none"> 대용량 실시간 처리기술 등장 <ul style="list-style-type: none"> 클라우드 기술, Hadoop, NoSQL 등 그 밖의 다양한 데이터의 특징과 패턴을 알아내는 분석기술
비즈니스 환경 변화	<ul style="list-style-type: none"> 기업의 산업구조와 경쟁양상의 변화 <ul style="list-style-type: none"> 인터넷과 SNS 발달에 따른 기업의 소비자에 대한 실시간 기호 분석 및 의사결정이 요구 인터넷 기반 기업의 시장변화에 대한 빠른 대응으로 다수의 성공 사례 가시화 ex) 아마존(Amazon), 넷플릭스(Netflix) 등

3. 빅데이터 분석기술 및 활용사례

3.1 빅데이터 분석기술

빅데이터 분석이 기존 데이터 분석과 차이점은 무엇일까? 실제 빅데이터 분석기술은 기존 데이터 분석방법 면에서 큰 차이를 보이지 않는다. 다만 빅데이터의 특성인 대규모 자료를 처리하기 위해 분산 처리방식인 맵리듀스(MapReduce)²⁾를 구현한 하둡(Hadoop)³⁾ 등과 NoSQL 데이터베이스와 같은 기반

기술이 필요하다. 하둡은 ‘데이터양’을, NoSQL 데이터베이스는 ‘다양성’을 대표할 수 있는 데이터 처리기반이며, 이와 더불어 ‘속도’는 스트림 데이터 처리기술이 필요하다. 실제로 빅데이터와 연계된 다양한 신기술들이 이미 시장에 영향을 미치고 있으며, 대표적인 신기술들은 빅테이블(Big Table), 카산드라(Cassandra), 데이터 웨어하우스 및 분석 어플라이언스, 분산 시스템(Distributed System), 구글 파일 시스템(Google File System), Hadoop, HBase 등이 있다. 이와 같은 빅데이터 처리 기반기술이 동반된다면 데이터로부터 유용한 의미를 끌어내기 위한 분석기법들은 현재도 통계학, 경영학 등 많은 분야에서 활용 중인 기계학습(Machine Learning), 데이터 마이닝 기법을 활용할 수 있다. 최근에는 소셜미디어 등 비정형 데이터가 증가하면서 텍스트 마이닝, 오피니언 마이닝, 소셜 네트워크 분석, 로그 데이터 마이닝 등이 큰 주목 받고 있다[5].

기계학습이란 인공지능의 한 분야로써 인간의 학습능력을 컴퓨터가 동일하게 학습할 수 있도록 하는 알고리즘과 기술을 개발하는 분야를 말한다[20]. 기계학습을 통해 수신한 이메일이 스팸인지 여부를 구분하고, 인식된 화상에 대하여 유사한 사물이나 사람 등을 구분하며, 기기고장 예측, 일기예보 등 그 응용범위는 매우 넓다. 데이터 마이닝은 대규모로 저장된 데이터를 분석하고 그 속에 숨어 있는 항목 사이의 상호관계와 패턴을 찾아내는 것을 말하며 [23], 다른 용어로 KDD(Knowledge Discovery in Databases)라고도 일컫는다(Wikipedia). 데이터 마이닝은 데이터 분석을 통해 ‘군집화(Clustering)’나 ‘분류(Classification)’, ‘신경망 네트워크(Neural Network)’, ‘회귀분석(Regression)’, ‘결정트리(Decision Tree)’ 등 대량의 데이터 속에서 지식과 패턴을 찾아낼 수 있다.

2) 맵리듀스(MapReduce) : 분산 시스템 상에서 대용량 데이터 세트를 처리하기 위해서 google이 소개한 소프트웨어 프레임워크(Framework)[14].

3) 하둡(Hadoop) : 분산 시스템 상에서 대용량 데이터 처리 분석을 지원하는 오픈소스 소프트웨어 프레임워크로써, 하둡의 상용 배포판으로 클라우더라의 CDH, IBM의 InfoSphere BigInsights, 데이터스텍스의 Brisk 등이 있음[3].

비정형 데이터의 내재된 가치를 효과적으로 알아내는 텍스트 마이닝은 사람들의 언어로 작성된 비정형 텍스트에서 자연어처리⁴⁾ 기술을 이용하여 유용한 정보를 추출하거나, 연계성 파악, 분류 혹은 군집화, 요약 등 빅데이터에 숨겨진 의미 있는 정보를 발견하는 기법이다. 오피니언 마이닝은 문서나 소셜미디어의 텍스트 문장을 대상으로 자연어처리 기반 감성분석 기술을 적용하여 사용자의 의견과 감정을 분석하는 것을 말하는데, 주로 상품에 대한 평가나 서비스에 대한 소비자의 반응, 사회적 이슈 등에 대한 여론을 살피는데 활용되고 있다. 소셜 네트워크 분석은 특정한 관계의 패턴을 가지는 인간이나 집단의 집합을 의미하는 사회 연결망(Social Network)의 구조 또는 형태의 특징을 도출하여 네트워크의 특성을 설명하거나 네트워크를 구성하는 단위의 행위를 설명할 수 있는 방법이다[36]. 그 밖의 Mackinsey[29]에서 소개한 다양한 빅데이터 분석기법은 <표 2>에 정리하였다.

<표 2> 빅데이터 분석기법

분석기법	
• A/B testing	• Association rule learning
• Classification	• Cluster Analysis
• Crowdsourcing	• Data fusion and data integration
• Ensemble learning	• Genetic Algorithms
• Natural language processing(NLP)	• Neural networks
• Network analysis	• Optimization
• Pattern recognition	• Predictive modeling
• Regression	• Sentiment analysis
• Signal Processing	• Spatial analysis
• Statistics	• Supervised learning
• Simulation	• Time series analysis
• Unsupervised learning	• Visualization

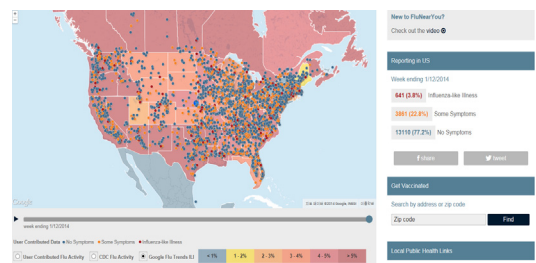
출처 : Mckinsey Global Institute[29], "Big data : The next frontier for innovation, competition, and productivity."

4) 자연어 처리 : 컴퓨터를 이용하여 사람 언어의 이해, 생성 및 분석을 다루는 인공지능 기술.

3.2 빅데이터 활용사례

최근 빅데이터 분석의 위력을 보여준 사건이 있었다. 지난 2013년 4월 15일(현지 시간) 미국 매사추세츠 주에서 진행된 보스턴마라톤 대회의 결승점 근처에서 두 차례의 폭발사건이 발생하였다. 사건 직후 미 정보기관과 분석전문가들은 약 10테라바이트(TB) 분량의 방대한 데이터 자료를 3일 만에 분석하여 범인을 검거하였다. 이 사례는 빅데이터 분석의 가치와 중요성을 보여준 대표적 사건으로 최근 이와 같은 빅데이터를 활용한 사례는 각종 뉴스, 인터넷, 기업홍보 자료 등을 통하여 쉽게 접할 수 있다. 본 연구에서는 빅데이터 활용사례에 대하여 공공분야와 민간분야로 나누어 주로 학술연구나 이와 대등한 연구결과를 토대로 유형별 정리하였다.

먼저 공공분야에서 Ginsberg et al.[19]는 매년 수만 명의 호흡기 질환자를 발생시키고, 약 25~50만 명 사망자가 생기는 계절성 독감에 대하여 연구하였다. 이전에는 면역체계가 없었으나, 새로운 변종이나 사람간의 전염 등으로 인해 전 세계적으로 유행할 수 있는 독감에 대하여 빅데이터 분석을 통해 예방 방안을 제시하였다. 인터넷 검색엔진 구글을 통해 독감관련 검색자의 위치 분포와 수적 증가 등에 대해 컴퓨터 인터넷 주소를 역추적하는 방법으로 독감징후를 감지함으로써 장차 질병 유행 방향을 예측하였다.



자료제공 : <https://fluinearyou.org/>.

[그림 1] 정부가 제공하는 독감 지도(Flu Map)

Clarke et al.[10]은 지금까지 생물학에서 수행한 최대 규모의 분산 데이터 수집 및 분석 프로젝트의 하나인 1000 게놈 프로젝트(Genomes Project)를

수행하였다. 이 1000 개놈 프로젝트를 통해 인간 유전자 정보를 공유하여 그 다양성을 분석함으로써 난치병/불치병 및 새로운 질병에 대한 빠르고 정확하게 진단하여 치료하는 연구를 하였다. 영국 법무부 [24]는 국민을 보호하고 재범을 감소시키며, 피해자와 대중을 위한 보다 효과적이고 투명한 법률 운영체계를 제공하기 위해 노력하고 있다. 그 사례로써 범죄자의 데이터로부터 OASys(Offender Assessment System)를 통해 범죄유형과 장차 행동에 대해 숨겨진 패턴을 찾아 범죄를 예측하고 있다. OASys는 400만 건 이상의 범죄기록이 들어 있으며, 이 데이터 내에는 교육, 인간관계, 재무관리, 소득, 라이프스타일, 약물, 알코올 남용 등 자료가 포함되어 관리되고 있다. 또한, 영국정부는 전력사용량의 효율적인 관리를 위해 2014년까지 1,000만 대, 2020년까지 가정과, 중소기업을 대상으로 5,400만 대의 스마트 계량기를 설치할 계획이다. 스마트 계량기란 전력 사용량 측정이 디지털화되어 있고, 양방향 통신 기능과 그 밖의 관리 기능을 갖춘 고성능 전력 계량기이다. 센트리카(Centrica)는 무상으로 스마트 계량기를 배포하여 1일 48회 SAP의 'HANA'라는 인메모리(in-memory) 데이터베이스를 활용하여 소비자의 전력 사용량 및 사용 패턴을 분석하여 전력수요를 예측하고 있다 [22]. 또한 빅데이터 분석을 활용하여 2010 영국의 총선 결과를 예측하기 위해 영국의 정치 정보 사이트 Tweetminster가 Twitter를 통해 집계한 결과가 ICM 및 TNS-BMRB 결과와 유사한 결과를 제시하기도 했다[34]. Wolfers[38]는 미국 대학농구 경기 4만 4천 건 이상의 데이터 분석을 통해 라스베이거스 도박자들의 예상 스프레드⁵⁾가 6점 이하일 때 매우 정확하다는 사실로부터 스프레드가 6~12점일 경우 선수는 팀의 승리에 영향을 주지 않는 한도 내에서 점수 조작을 하고 있다는 것을 증명하기도 하였다.

민간분야에서는 Vosen and Xchmidt[35]가 Google Trends의 소비관련 검색 결과가 MCSI(Michigan

5) 스프레드 : 도박에서 승률이 낮은 팀에게 주어지는 어드밴티지.

Consumer Sentiment Index)와 CCI(Consumer Confidence Index)의 결과가 거의 일치함을 확인하였다. Lavalle[28]은 소비자의 상품 구매를 위한 광고채널 예산 배분을 위한 분석기법을 통해 Best Buy 광고비의 최적 배분 안을 제시했다. Watson et al.[37]은 미국 최대 물류업체 UPS가 자사 배송트럭에 텔레매틱스(Telematics)를 장착하여 차량의 작동, 속도 등을 모니터링하여 이를 연료비 절감과 차량유지관리의 효율성 향상으로 연결시킨 사례를 보여주기도 한다. Jacobs[26]은 게임을 일반 웹사이트가 아닌

〈표 3〉 분야별/유형별 빅데이터 활용사례

구 분	내용(기존 ⇨ 변화)
공공분야	범죄 발생 후 범인추적 ⇨ 범죄 예측 예) 미국 샌프란시스코 범죄예방시스템
	주기적 유가 분석 ⇨ 실시간 유가 예보 예) 한국 석유공사 유사 예보 서비스
	예기치 않은 통신 장애, 통화불능, 복구 ⇨ 장애 감지, 실시간 자동화 복구 예) TELCO
	질병 발생 이후 표준 처방 ⇨ 질병예측, 개인화된 질병 치료 예) 미국 질병치료 체계
	범인 DNA 정보 확인 시간 소요, 해결 지연 ⇨ 1시간 이내 범인 DNA 분석 예) 미국 FBI 유전자 색인 시스템
	일정 주기 소비전력 확인 ⇨ 소비전력 실시간 모니터링 예) 영국 스마트그리드(Smart Grid)
	Sampling 및 Batch 기반의 탈세 방지 ⇨ 실시간 분석 사기 징후 검출 예) 미국 국세청 탈세 방지 시스템
민간분야	일관되게 적용되는 가격할인 전략 ⇨ 고객 구매정보, 취향 등 분석을 통한 고객 세분화 및 맞춤형 전략 추진 예) 미국 넷플릭스
	1년 단위 자동차 보험 할증 ⇨ 실시간 보험료 반영
	대량으로 일관되게 적용되는 광고 ⇨ 이용자의 가까운 미래 행동을 예측해 해당 상황에 맞는 광고를 선별하여 제공 예) 맞춤형 광고(아마존, eBook, 페이스북 등)
	자산의 주기별 운항정보, 수리/보수 정보 분석 ⇨ 자산 수요예측, 수리/보수 자동화를 통한 효율화 예) 일본 코마츠 콤투랙스(Komtrax) 시스템

facebook 플랫폼을 활용하여 사용자에게 피드백 함으로써 바이럴 계수를 향상시켜 수익을 증가시켰고, Davenport et al.[13]에 따르면 단계별 애널리틱스의 의사결정을 자동화하여 보험사기로 손실이 큰 회사들이 사기 가능성이 있는 경우를 찾는 시간을 95% 줄였고, 그 해결도 55%에서 88%로 개선시켰다. 그 밖의 빅데이터 활용사례를 분야별, 유형별로 구분하여 <표 3>에 정리하였다.

4. 국방분야 데이터 활용실태

효율적인 국방개혁 추진과 군사력 건설 및 운용을 뒷받침하기 위해서는 과거의 주관적이고 정성적인 의사결정체계에서 벗어나, 객관적이고 정량적인 근거자료에 의한 합리적인 의사결정체계로의 전환이 요구된다. 군은 특성상 자료를 축적하고 보관하는 것에 익숙하므로 빅데이터를 활용의 효과를 쉽게 얻을 수 있는 조직이다. 이런 맥락에서 한국군은 오래전부터 데이터 수집/관리 체계를 구축하려고 노력하였으나 그 (빅)데이터 활용 필요성에 대해 인식의 수준은 높다고 평가하기 어려운 수준이다.

이번 장에서는 미군의 체계적인 데이터 수집 및 관리현황을 알아보고, 빅데이터 활용을 위한 노력을 확인한다. 또한 한국군의 빅데이터 현황을 ‘저비용 데이터’, ‘분석기술’, ‘분석된 정보의 가치’의 세 가지 관점에서 살펴보고, 효율적인 빅데이터 활용을 위한 방안을 제시한다.

4.1 선진군의 데이터 관리 및 활용과 빅데이터

세계적으로 가장 많은 국방비를 사용하고 있는 나라는 미국으로 2013년 통계에 의하면 한해 국방비로 약 690조를 지출하고 있으며, 이는 세계 국방비의 41%이고, 나머지 상위 14개 국가의 국방비를 합친 것보다 많은 수치이다. 이러한 미국은 1970년대부터 데이터의 중요성을 인식하여 수집 및 관리하고 있으며, 1990년대 중반부터 산재되어 있던 각종 데이터 자료를 유기적으로 통합, 관리, 분석하여 결

프전 이후 많은 전쟁과 전투에서 그 효과를 보고 있다. 심지어 최근에는 모든 훈련 데이터뿐만 아니라 실제 전투에서 발생하는 데이터를 수집하여 무기 및 탄약효과, 취약성 분석 등을 통해 교육훈련이나 위게임, 실제 전술적, 전략적 규모 작전에 환류하고 있다.

미군은 체계적인 데이터 수집 및 관리, 분석을 담당하는 전담 조직을 설립하였는데, 그 대표적인 기관으로 미 육군의 교육사령부(TRADOC⁶⁾) 예하 전훈분석센터(CALL⁷⁾)와 물자사령부(AMC⁸⁾) 예하 물자체계분석연구소(AMSAA⁹⁾)가 있다. CALL은 육군, 합동작전, 정부기관 및 다국적군 등에 모든 영역의 군사작전을 지원하기 위해 전 세계 작전지역/전투훈련장/각종연습 및 실험 등을 통해 도출된 전술, 군사기술, 교훈 등을 수집, 분석, 통합, 분배하고 자료를 축적하는 임무를 수행한다. AMSAA는 미군의 전투역량 강화를 위한 분석 솔루션 제공을 담당하는 부서이다. 물자체계분석연구소의 체계 분석은 크게 물자체계 분석과 군수체계 분석으로 나눌 수 있으며, 각각의 역할은 <표 4>와 같다.

<표 4> 물자체계분석연구소(AMSAA)의 역할

물자체계 분석	군수체계 분석
• 시스템 성능 분석	• 보급체계 분석
• 개발/인증 체계 성능데이터	• M&S 체계지원능력 및 데이터 개발
• M&S 체계 개발 방법론	• FEDC 및 분석
• 기술과 위험 평가	• 비용분석
• M&S 목록 및 체계, VV&A 수행 대행	• 육군 RAM 표준 수행 대행
• 국방성 JTCG/ME 사업 관리	• 신뢰성 분석

출처 : U.S. Army Materiel Systems Analysis Activity Command Overview, 2010.

이들의 데이터 수집활동은 크게 능동적 수집과 수동적 수집으로 구분하는데 능동적 수집은 국립

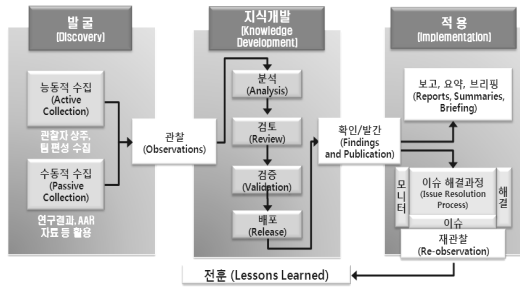
6) TRADOC : Training and Doctrine Command.

7) CALL : Center for Army Lessons Learned

8) AMC : Army Materiel Command.

9) AMSAA : Army Materiel Systems Analysis Activity.

훈련장을 비롯한 4개 훈련장에 관찰자를 상주시키고 대규모 작전 혹은 주요 훈련의 경우에는 팀을 편성하여 집중 수집을 하는 것을 말하며, 수동적 수집은 연구부서, 훈련장, 실제 작전 투입부대 등으로부터 사후검토, 연구결과 등의 전훈자료를 수집하는 것이다. 이렇게 수집된 데이터가 활용되기까지 [그림 2]의 절차를 밟아 검증하게 된다.



[그림 2] 훈련데이터 수집 절차

미군의 데이터 관리는 전 분야에서 이루어지고 있으나 전장분야, 군수분야, M&S 분야 등은 빅데이터 분석 이전부터 매우 체계적인 관리가 되어왔다. 먼저 전장분야는 전훈분석센터에서 주로 수집 및 관리하였으며, 전장과 관련된 데이터를 <표 5>에서와 같이 7개 분야(지휘통제, 인사, 정보, 작전, 군수, 민사작전, 연합전력의 데이터)로 구분하여 수집하고 있다.

군수분야에서 SDC¹⁰⁾는 미 육군의 주요 장비 성능 및 지원능력(performance and supportability)에 기초한 자료를 수집하는 프로그램으로써 수집된 데이터 분석결과는 장비의 획득, 운용, 운용인력 편성 등과 관련한 각종 업무에 사용된다. 또한 효율적이고 효과적인 야전 운영데이터 획득을 위해 대상체계를 대표하는 표본을 선택하여 정보를 수집하고 분석한다. 야전 운영데이터는 신규 개발 및 획득 장비나 중점관리 대상 장비에 대한 시스템의 문제점 및 개선방향

<표 5> CALL 수집데이터 분류

구분	내용
• 지휘통제 분야	개인 메모, 중요인물 브리핑, 지휘관 주도 작전, 지휘부 회의 기록 등
• 작전 분야	전투보고, 전투 평가, 작전 전훈, 작전 명령, 하위 명령, 표적 목록과 타격 수단, 전술 계획, 지휘관 고려사항, 등의 기록 등
• 인사 분야	인원자료, 수훈 병사/부대 추천, 사상자 보고, 지휘부 명단 등
• 정보 분야	지형 정보, 일별 정보, 전술 정보 팀 보고, 지형 분석, 적의 전투 명령과 인물 보고서, 전투 피해 평가 등
• 군수 분야	장비 상태 및 준비율, 보급 소요, 신 장비 운용이력, 전투 중 중요 장비의 파손/손실 등
• 민사작전 분야	지역주민성향, 시가지 평가, 지역 조사 보고, 일일/주간/월간 민군 전술 상황 보고 갱신 등
• 연합전력 분야	타 국가의 훈련, 부대, 인원, 작전, 장비, 군수 등 보고서

도출을 위해서 획득하며, 성능향상을 제시하거나 운용성의 최적화, 군수지원의 효율화, 차기무기체계 개발 시에도 적용하고 있다.

M&S 분야는 AMSAA 내의 합동기술조정단(JTCG/ME)¹¹⁾이 전담하여 무기체계 효과/취약성 자료를 전담하여 개발하고 있다. 수집된 데이터에 대해 크게 3가지 단계(데이터 검증(Data Verification), 데이터 확인(Data Validation), 데이터 인증(Data Certification)¹²⁾)에 걸쳐 인증함으로써 생산자와 사용자의 입장에서 자료의 신뢰성과 품질향상을 보장한다. 전담조직에 의해 데이터 분석을 위한 노력과 활동은 ‘합성환경 자료표현 및 교환명세(SEDRIIS : Synthetic Environment Data Representation & Interchange Specification)’와 ‘합동탄약효과교범(JMEM)’ 등과 같은 결과물으로써 각 분야에 환류되어 활용되고 있다. ‘합

10) SDC : Sample Data Collection, 미 육군 규정 AR 750-37(Sample Data Collection : The Army Maintenance Management System) 지침에 의거 특정장비를 대상으로 선별적인 야전 운용자료를 분석하여 필요한 정보를 분석하는 기법.

11) JTCG/ME는 Joint Technical Coordinating Group for Munitions Effectiveness의 약어로서 미국의 합동탄약효과교범(JMEM) 개발을 관장하는 조직.

12) VV&A(Verification, Validation and Accreditation)가 시뮬레이션 모델에 대한 검증과정이라면 VV&C(Verification, Validation and Certification)는 데이터에 대한 검증과정임.

성환경 자료표현 및 교환명세'는 합성환경 구축 및 교환에 관한 표준으로써 합성환경자료의 공유를 통해 M&S 체계간의 상호 운용성을 제고하고, 전장 표현 및 모의의 일관성을 보장함과 동시에, 환경자료에 대한 기관간의 중복개발 노력을 최대한 방지한다[2]. '합동탄약효과검'은 지침, 효과분석 운용 도구 및 방법론, 공대지, 지대지, 방공전(공중우세, 지대공, 함대공) 등의 다양한 형태의 효과도 및 특수작전 표적 취약성 자료, 그리고 공격무기/탄약별 물리적, 기능적 특성자료 등을 분석한 결과물이다. 이러한 자료들은 전투준비태세, 임무계획수립, 전술개발 및 위계임/훈련 등과 같은 전투수행업무와 임무분야 분석, 소요연구, 조달계획수립, 성능평가, 대안분석, 설계 개선/향상 및 생존성 향상 등과 같은 전력기획/계획 업무수행에 활용되고 있다.

또한 수집된 데이터의 분석 자료를 활용하기 위하여 구축된 미군의 데이터 서비스 시스템인 포털 웹사이트는 자료 표준화, 자료접근, 시간단축 및 지식 공유 등과 같은 4가지의 목적을 가지고 구축하였다. 자료 표준화는 육군의 분석관련 기관들에게 합의된 표준과 프로세스들을 제공하는 것이다. 현재 포털이 개발하여 제공하는 표준에는 표준 명칭 데이터베이스(SND : Standard Nomenclature Database)와 통상 표준파일양식(SFF : Standard File Formats)이라는 특성/능성자료 교환양식(C&P Data Interchange Formats) 등이 있다. 자료교환양식은 표준 명칭 데이터베이스와 미 육군 물자체계분석연구소의 성능자료를 위한 자료교환양식을 정의하고 있다. 모든 표준 명칭 데이터베이스 어휘들(플랫폼 이름, 무기체계 이름, 탄약이름, 센서이름 등)은 포털 사용자들에 의해 질의되며, 포털은 사용자들에게 표준 명칭 데이터베이스의 어휘들에 대한 변경제안을 허용하고 있다. 이러한 의사결정 및 분석시간 단축은 포털의 핵심목적에 해당하며 온라인을 통해 항상 신뢰성 있는 자료들을 사용자의 목적에 맞도록 제공하고 있다.

이밖에 미 국방부는 방대한 데이터를 다양한 분야에서 활용하고 있다. 먼저 '국가 안보분야'에서는 미 국방부(DoD, Department of Defense)와 국가

안보국(NSA, National Security Agency)이 테러와 관련해서 빅데이터를 활용한다. 미국은 지난 9.11 테러 이후 국가 안보와 관련된 부서의 통합과 다른 부서간의 유기적 정보공유를 통해 테러와 범죄방지를 위해 많은 양의 데이터를 수집, 분석 및 예측체계를 도입하였다. 공항 입국심사에서 테러리스트 용의자 색출을 위해 각 국가별 의심스러운 국제 테러리스트 정보를 'TIDE(Terrorist Identities Datamart Environment)'라는 데이터 베이스에 320테라바이트(TB) 분량으로 통합하여 관리하고 있으며, 입국심사 시 이 데이터를 바탕으로 초당 6만여 개의 개인 기록을 분석·처리하여 테러와의 관련 여부를 확인 후 입국 허가를 진행한다[27].

미 국가안보국은 잠재적 테러용의자에 대한 정보를 수집하기 위하여 인터넷 사용자의 메타데이터를 저장한다. NSA의 하루 인터넷을 통한 정보 접촉 양은 인터넷의 1일 트래픽 1,826페타바이트(10^{12} bytes)의 1.6%인 약 29.2×10^{12} 바이트이며, 이 중 0.025% 정도를 선택하고, 0.00004%를 분석한다고 한다. 수집된 메타데이터를 저장하는 일명 '마리나(Marina) 시스템'은 지목된 사람의 1년 동안의 삶을 재구성 할 수 있을 정도로 방대한 정보를 저장한다. 수집된 정보는 주로 전화나 이메일, 보험 가입정보, Facebook 프로필, GPS 정보, 재산목록 등이며 이러한 메타데이터를 통해 미국인들의 사회적 관계에 관한 정보까지 분석이 가능한 실정이다. 즉, NSA는 마리나를 통해 표적 대상에 대한 일상생활의 패턴을 찾고 테러 등 국가 안보에 도움이 되는 다양한 차트도 만들어 활용하고 있다[8].

또한 국가 안보와 관련하여 '폭탄 테러범 색출'에도 빅데이터를 활용하고 있다. 지난 2013년 4월 15일 14시 50분경 미국의 보스턴 마라톤 대회에서 폭발 사건이 발생하였다. 미국 보스턴 마라톤 대회 결승 지점에서 2개의 폭탄 1분 간격으로 폭발하였고, 이 테러로 부상 260여 명, 사망 3명의 희생자가 발생하였다. FBI는 범인을 찾아내기 위하여 주변 600여 대의 CCTV 데이터 및 당시 주변에서 있었던 Twitter 및 Facebook 대화 그리고 당일 현장 기자회견에 남은

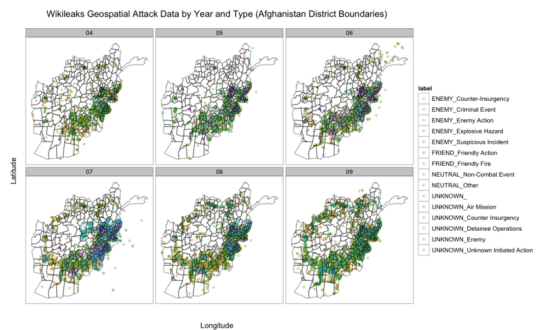
사람들의 통화로그 등 확보하였으며, 그 분량만 하더라도 약 10테라바이트(TB)였다. FBI는 수집된 자료를 바로 분석전문가에게 전달하였고, 전문가들은 수집된 멀티미디어 파일의 데이터 형식을 통일하여 빅데이터 분석을 시작하였다. 2일 후 폭발당시 사람들의 증언을 토대로 용의 추정인물이 흰색 야구모자, 밝은색 후드티, 검은 재킷을 입고 있었다는 유형분류가 끝났고, 폭발사건 3일 만인 4월 18일 범인은 검거되었다.

인사분야 중 ‘의료 서비스’ 관련해서는 미국의 퇴역 군인국(VA, Department of Veterans Affairs)은 퇴역 군인에게 맞춤형 의료 서비스를 지원하기 위해 빅데이터를 활용한다. 미군은 걸프전 이후 국방 의료 감시 시스템(DMSS, Defense Medical Surveillance System)을 도입하여 현역 군인 및 퇴역 군인, 참전 용사들의 의료 관련 데이터를 관리하고 있다. 이 시스템에는 군사작전 시 전개지역 환경, 병원관리에 위협이 되는 요소, 해당지역의 질병, 전투원들의 의료 기록 등을 포함한다[9]. 또한 국방 혈청 저장국(DoDSR, Department of Defense Serum Repository)은 의료 감시 목적을 위해 장병들로부터 혈청을 뽑아 보관 및 데이터화하여 관리하고 있다[32]. 최근 미 국방부와 함께 퇴역 군인국은 빅데이터 분석을 위해 DMSS와 DoDSR의 데이터를 토대로 2년간 25개 데이터 웨어하우스를 배치하여 퇴역군인의 전자 의료 기록(EHR, Electronic Health Records) 데이터를 분석하고, 2,200만 퇴역 군인들에게 진료 기관에서 개별 환자에게 맞춤형 진료할 수 있도록 지원하고 있다. 빅데이터 분석을 활용하여 페타바이트 수준의 임상적, 유전적 데이터 분석으로 보다 효과적인 의료 서비스 지원이 가능해 진 것이다[4, 30].

미군은 ‘잠재적 위협 방어’에도 빅데이터를 활용하는데, 미군 정보 분석가들에 의해 이메일, 휴대폰 사용기록, 언론기사, 동영상, 음성대화 등 방대하고 다양한 종류의 데이터를 분석해 물리적 공격뿐만 아니라 사이버 대전에도 대비하고 있다. 또한 미 육군, 공군 해군은 훈련과 정찰 동영상 분석, 이메일 분석에 주로 시맨틱(Semantic) 검색, 분석 툴을 사용하고 있다.

미 공군 기지에서는 ‘에너지 소비절약’ 분야에서도 빅데이터를 활용하였다. 플로리다에 위치한 에글린(Eglin) 공군 기지에서는 100개의 건물에 20,000개 이상의 센서를 설치하여 전원스위치, 밸브, 온도 조절장치, 에어컨 등에서 에너지 소비에 대한 데이터를 수집하였다. 그 결과 일년간의 난방, 환기, 에어컨 사용 등에 대한 정보 데이터를 구축할 수 있었고, 구축된 정보를 활용하여 한 해 동안 250만 달러를 절약하였다. 에너지 비용 절감 외에 공군은 또한 장비가 유사한 장치 또는 장비의 과거사용 이력 등을 비교하여 장비정비에 빅데이터 분석을 활용하기 위해 계획 중에 있다[25].

미 국방분야에서 직접 활용한 사례는 아니지만 뉴욕 대학에서 박사과정 중인 드루 콘웨이(Drew Conway)는 2010년 위키리크스(Wikileaks)가 공개한 미 정부의 아프가니스탄 전쟁에 관한 76,900여 건의 미공개 문서인 ‘아프가니스탄 전쟁일지(Afghan War Diary)’를 분석하였다. 위키리크스에 저장되어 있는 테라바이트 급의 방대한 데이터들을 ‘R’프로그램을 통해 미국과 아프가니스탄 연합군의 병력 활동 동향을 식별하였다. 드루 콘웨이 먼저 아프가니스탄 주요 5곳을 적, 중립, 동맹지역으로 나눠 정보를 분류하고 각 지역에서 어떤 활동이 주로 일어나고 있는 패턴을 [그림 3]과 같이 분석했다. 정리된 정보는 탈레반의 활동이 어느 지역에서 많이 일어나는지, 미국과 동맹 맺은 지역이 어딘지를 쉽게 알아볼 수 있었고, 시간



출처 : <http://drewconway.com/zia/2013/3/26/what-data-visualization-should-do-simple-small-truth>.

[그림 3] 미국과 아프가니스탄 병력 활동 동향

흐름에 따라 아프가니스탄에서 전쟁 양상이 어떻게 변하고 있는지도 확인할 수 있다[4, 12].

이와 같이 국방분야와 관련해서 빅데이터의 활용은 일반 공공분야와 민간분야 만큼이나 다양하다.

2012년 3월 오바마 행정부가 빅데이터를 바탕으로 한 정부운영 전략(Big Data R&D Initiative)을 발의한 이후 미 국방부는 연간 2억 5,000만 달러를 투자해 더욱더 빅데이터 통한 새로운 가치를 창출하고자 노력하고 있다. 특히 미국 방위고등연구 계획국(DARPA, Defense Advanced Research Projects Agency)에서는 약 4년 동안 연간 2,500만 달러를 투자해 국방과 관련한 많은 양(Volume)의 데이터와 다양한 형식(Complexity)의 데이터를 동시에 통합하여 분석할 수 있는 컴퓨팅 기술과 소프트웨어 개발을 위해 XDATA 프로그램을 진행 중이다. XDATA가 완성되면 불완전한 데이터를 처리하고 시각화하는 알고리즘이 개발되어 자율시스템이 구축되고, 전투원이 다양한 임무들에 대해 스스로 군사행동을 추진할 수 있도록 도와주게 된다. 이 밖에 일상적으로 발생하는 다양한 데이터 속에서 국방 위협 요소를 발견·감시할 수 있는 ADAMS(Anomaly Detection at Multiple Scales), 군사 네트워크 내부의 데이터를 감시하고 사이버 공격을 사전에 차단하기 위한 CINDER(Cyber-Insider Threat) 프로그램 개발을 하고 있다. 또한 자연어로 구성된 텍스트를 해독하고 이를 토대로 의미 기반의 결과를 제시하는 Machine Reading 프로그램, 입력된 영상정보를 기초로 관련 정보를 추론하고 실제 및 허구의 사건을 예측할 수 있는 Mind's Eye 프로그램, 데이터 암호화와 관련된 프로그래밍 언어를 개발하기 위한 PROCEED(Programming Computation on Encrypted Data) 프로그램, 방대한 군사 동영상 콘텐츠를 빠른 속도로 검색·분석할 수 있도록 하는 VIRAT(Video and Image Retrieval and Analysis Tool) 프로그램 등이 빅데이터 시대를 맞이하여 미 국방부가 새롭게 추진하는 계획들이다[11, 31]. 미 국방분야에서 추진하고 있는 프로그램들은 일상생활에서 지금까지 자연스럽게 축적되어온 데이터들로부터 그 의미를 찾

아 새로운 가치를 창출하기 위하여 분석의 기술적 인프라를 구축하는 것이라고 볼 수 있다.

4.2 한국군의 국방분야 빅데이터 실태 및 활용 가능 분야

한국군은 그동안 각 분야에서 데이터를 수집하고, (빅)데이터를 활용하기 위한 나름의 시도를 하고 있지만, 빅데이터 활용의 필요성에 대해 피상적으로 인지해 본질과는 크게 이격되어 있는 것도 사실이다. 따라서 먼저 일반적으로 빅데이터 활용을 위하여 필요한 3가지 요소, 즉 '저비용 데이터', '분석기술', '분석된 정보의 가치' 관점에서 우리 국방분야의 현 실태를 살펴보고, 우리가 국방분야에서 활용 가능한 빅데이터 분야를 모색해본다.

4.2.1 한국군의 빅데이터 활용을 위한 현 실태

먼저 '저비용 데이터' 측면에서 국방분야를 평가해본다. 한국군은 조직의 특성상 자료를 축적하고 보관하는데 익숙하다. 또한 IT 분야의 기술발달과 함께 군도 역시 다양한 분야에서 수많은 자료를 축적할 수 있는 시스템을 갖추고 있다. <표 6>에서 처럼 현재 육군이 운용하고 있는 분야별 소프트웨어 현황만 보아도 군이 얼마나 다양한 데이터 축적을 위한 기반을 구축하고 있는지 알 수 있다.

<표 6> 육군 표준 소프트웨어 현황

구분	계	기획	예산	인사	의무
계	208	24	4	66	3
군수	수송/시설	동원	전자행정	전장관리	
25	3	2	61	20	

따라서 한국군은 빅데이터 특성에서 언급한 이미 구축되어 있는 시스템으로부터 저비용 데이터, 즉 Cheap Data를 얻기 위한 충분한 조건을 갖추고 있다. 그러나 한국군의 데이터 수집/관리 체계를 살펴보면 미군과 비교하였을 때 제도, 조직, 인력 및 예산 등 대부분이 미비한 현실이다. 그 결과 빅

데이터로써 분석할 가치가 있는 데이터를 방치 또는 폐기하고, 데이터가 있어도 전사 차원의 체계적인 관리와 공유가 이루어지지 못해 빅데이터로부터 통찰력(Insight)을 도출하지 못하고 있다. 그러므로 곳곳에 산적해 있을 데이터 양을 고려했을 때 전장관리, 군수 효율화, 인사운영개선, 정책수립, 부대운영계획 수립 등에 필요한 데이터의 적시적인 활용과 재사용에 한계가 있다. 또한 데이터를 통합 관리하고 주관하는 기관도 부재한 실정으므로 효율적인 데이터 수집/관리 체계 구축이 요구된다. 즉, 데이터의 양적인 면에서는 선진군과 비교하거나, 국내의 다른 정부부처 등과 비교하였을 때 충분히 방대한 양을 보유하고 있으나, 축적된 데이터에 가치를 부여하지 못하여 데이터 품질을 개선하기 어려운 실정이다.

빅데이터 ‘분석기술 측면’은 국방분야의 빅데이터 분석에 필요한 지식기반 및 분석 도구들의 도입이 미흡하고 인력도 부족한 상태이다. 여러 곳에 흩어져있는 데이터들을 동시에 통합적으로 관리할 수 있는 컴퓨팅 인프라를 하루빨리 구축하고, 빅데이터 분석을 위해서 기본적으로 하둡, NoSQL 등의 빅데이터 분석 인프라 기술을 갖추어야 하겠다. 더불어 다양한 통계처리, 데이터 마이닝, 텍스트 마이닝, 그래프 마이닝 등 다양한 분석 방법 및 기계학습 등을 적용할 수 있는 인력을 양성하여야 한다.

‘분석된 정보의 가치’ 측면에서 한국군은 아직 데이터를 의사결정 합리화의 장식품으로 오용하여 데이터에 근거한 사후 회고와 반성, 교훈 도출의 과정을 대부분 간과하고 있다. 또한 사실에 기반을 둔 의사결정보다는 지휘관이나 해당 분야 전문가의 직관에 의존한 의사결정을 중시하는 조직문화 역시 문제점이라고 할 수 있다. 미군은 새로운 의견제안 시 데이터에 근거한 분석인지를 질문하고, 데이터에 근거하고 있지 않다면 채택하지 않는다. 그만큼 분석된 정보의 가치를 중시하는 문화가 저변에 깔려있는 것이다. 구글, 아마존과 같은 인터넷 기업들이 기존의 기업들과 몇 가지 차별적 역량을 보이게 된 것도 이와 같은 분석된 정보의 가치를 알고, 의

사결정시 기업전략의 기반으로 데이터를 활용하였다는 것이다.

4.2.2 한국군의 빅데이터 활용가능 분야

미군의 활용사례나 공공분야, 민간분야의 사례와 같이 한국군도 다양한 분야에서 빅데이터를 활용한 새로운 가치 창출이 가능하다. 현재 군에서 운용중인 소프트웨어를 참고하여 빅데이터 활용이 가능한 체계를 전장관리 분야, 군수분야 등으로 나누어 <표 7>과 같이 정리해 보았다.

<표 7> 분야별 빅데이터로 활용 가능한 체계

구분	내용
전장관리	<ul style="list-style-type: none"> · 과학화 전투훈련(KCTC) · 전투지휘훈련(BCTP) · 창조 21 모델 · 화랑 21 모델 · 전투 21 모델 · 청해 모델 · 천자봉 모델 · 군사정보통합처리체계(MIMS) · 합동지휘통제체계(KJCCS) · 군사지리정보체계(MGIS) · 지상전술 C4I 체계(ATCIS) · 해군전술 C4I 체계(KNCCS) · 공군전술 C4I 체계 · 자동화방공통제체계
군수분야	<ul style="list-style-type: none"> · 장비정비정보체계(DELIIS) · 국방 물자정보체계 · 국방 탄약정보체계
인사분야	<ul style="list-style-type: none"> · 국방 인사정보체계 · 국방 의료정보체계 · 신인성검사 · NDEMIS(또는 구 DEMIS) · 연대행정업무 통합관리체계
기타	<ul style="list-style-type: none"> · 국방 수송정보체계 · 온-나라 시스템 · 국방 전자결재체계 등

<표 7>에서 예시한 것과 같이 각 분야별로 생산되고 있는 데이터를 빅데이터 분석을 통해 새로운 가치를 창출한다면 국방분야 효율화는 극대화 될 것이다.

미군의 빅데이터 활용사례에 맞추어 한국군이 적용 가능한 빅데이터 활용 분야를 제시해보면, 먼저 ‘전장관리’ 측면에서 육군은 과학화 전투훈련(KCTC)

데이터 활용이 가능하다. KCTC는 실제 전장과 가장 유사한 환경에서 두 부대 간의 쌍방교전훈련을 위해 27종의 마일즈 장비와 238종의 과학화 체계 장비를 갖추고 있다. 이들 장비로부터 발생한 각 종 신호는 중계소 등의 통신망을 통해 훈련통제본부(EXCON)로 전송되며, 전송된 데이터는 ORACLE 기반의 데이터 웨어하우스(Data Warehouse, DW)에 저장된다. 또한 이동비디오 차량을 통한 영상정보와 유·무선 통신장비를 통한 감청정보 등의 교전 간 발생하는 교전정보가 저장·관리되고 있는데 그 저장 양은 약 36.6 TB(Tera Bytes, 테라바이트) 정도이다. 따라서 빅데이터 분석을 통해 축적된 데이터로부터 다양한 가치를 창출할 수 있는데, 첫째, 전장상황에서 전투원의 행동패턴을 파악할 수 있으며, 전쟁 승리를 위한 훈련 방안을 제시하고 한국지형에 맞는 각종 교리를 검증하는데 활용할 수 있다. 더불어 국방 M&S 체계의 모의논리를 보강하는 역할도 가능하다. 둘째, 각개 병사의 생존시간에 영향을 미치는 요소(교육훈련, 복무개월 수, 직책, 개인 특성 등)를 분석하여 생존성을 높이기 위한 방안을 찾을 수 있으며, 더 나아가 분대, 소대, 중대, 그 이상의 부대단위의 생존성을 보장하기 위한 훈련, 사격술, 체력단련 등의 방안을 찾는 데도 활용 가능 할 것이다. 셋째, 현재 운용하고 있는 교육훈련체계에 대한 발전방안도 제시가 가능하다. 예를 들어 사격술 훈련의 거리가 현재 100, 200, 250미터를 조준하여 3자세로 표적을 맞추게 되어있는데, 한국지형과 같이 산악이 많고, 도시지역 작전을 하는 상황에서 사격 거리나 자세 등을 재검증하고 싶을 때 KCTC 훈련에서 훈련인원들의 사망 시 사격자와 피격자의 거리, 위치, 자세, 상황 등을 축적된 데이터의 분석을 통해 검증할 수 있을 것이다.

또한 군사정보통합처리체계(MIMS)와 각 군의 C4I체계 등에서 수집하고 있는 방대한 데이터를 활용하여 전장상황을 예측하고, 최적의 대응방안도 찾을 수 있다. 현재 군사정보통합처리체계(MIMS)는 육군이 보유하고 있는 TOD/RASIT, UAV, 인 간정보, ES/EA와 해·공군의 ES, AGS, P3-C,

RF-4C로부터 북한군의 정보(비행기지 위치정보, 보유항공기, 장착가능 무장 등) 및 한국군 자산(기지별 가용 항공기, 비상대기 전력 등)을 각각 파악하고 있으며, 특히 공군 C4I체계의 전장상황 DB 내에는 중앙방공통제소(MCRC, Master Control and Reporting Center)로부터 받은 항적위치정보(피아 식별, 무장현황, 이동경로 등)를 저장·관리하고 있다. 따라서 북한군의 비행기지 정보(기지별 보유 항공기, 항공기별 장착 가능 무장)와 과거 항적 활동정보(기지별 예상이동 경로, 항적의 예상 이동 속도)로부터 빅데이터 분석을 통해 기지별, 항공기별로 항적상에서 이동 속도와 이동경로의 패턴 등의 비행종합정보를 제공함으로써 전장상황을 예측할 수 있는 시스템을 체계화할 수 있다. 그 결과 평소와 다른 패턴이나 속도의 움직임이 식별되면 자동적으로 경보 및 상황전파를 하도록 하는 시스템 구축이 가능하고, 더 나아가 상황별 한국군이 보유하고 있는 자산 중 가장 적합한 항공기나 기타 무기체계를 추천하도록 할 수 있다.

‘인사분야’ 중 미군이 시행하고 있는 퇴역 군인을 위한 의료 서비스와 마찬가지로 한국군도 기존의 데이터 활용을 통해 적용가능하며, 빅데이터를 활용하면 그 이상의 의료 서비스도 제공가능하다. 우리 군은 매년 현역 간부 11.6만 명(2014. 3. 현재)을 대상으로 정례신검을 통해 혈액, 소변, 생화학 검사 등을 실시하고 있으며, 각 군병원 및 각 군 본부에서는 이 데이터를 관리하고 있다. 2013년도부터는 51.7만 명(2014. 3. 현재)의 현역병사에 대해서도 상병진급 시 ‘상병건강검진’을 시행하고 있는데 간부 정례신검과 같이 모든 인원에 대한 임상정보(혈액, 소변, 생화학 검사 등)를 포함하며, 국군 의무사령부 서버인 NDEMIS(또는 구 DEMIS)에 데이터를 관리하고 있다. 현재 여러 곳에 분산되어 관리되는 군 간부 및 병사들의 의료정보를 한 곳에 통합시켜 빅데이터 분석을 실시한다면 진료 시 환자 중심의 진료가 가능하며, 유발 가능한 질병에 대해서도 예측 가능하여 병원관리에 상당한 도움이 될 것이다. 뿐만 아니라 육·해·공군별 부대의 특성과 부대의 지역 차이에서 나타나는

질병, 그리고 계절적 요인에 따라 유행하는 질병을 예측하고, 이에 대한 효과적인 의료 지원도 가능할 것이다. 또한 해외 파병인원에 대한 파병 전·후 신검 정보에서 파병지역의 발생가능 풍토병과 같은 질병, 복귀 후 파병 지역에 따라 나타날 수 있는 예상 심리치료 등의 예측도 가능하며, 그 결과 필요한 약품을 사전에 확보하고 관련 심리상담 인원을 배치할 수 있다. 군에서 보유하고 있는 방대한 임상적, 유전적 의료기록에 대한 분석은 장차 군 전역자들에게도 군 생활 동안 해당업무, 근무지역, 근무환경 등의 개인별 특성을 분석하여 효과적인 의료 서비스를 지원 하는데 빅데이터 활용이 도움이 될 것이다.

‘군수분야’는 일본 코마츠 기업의 사례와 같이 현재 운용 중인 국방 장비정보정보체계(DELIIS, Defense Logistics Integrated Information System)의 데이터를 활용하여 장비의 수요와 예측, 장비유지관리 등의 효율성을 높일 수 있다. 장비정보정보체계는 육군 본부로부터 편성부대까지 하나의 통합된 웹(Web) 환경을 기반으로 데이터 센터를 운영하여 장비, 수리부속 및 정비재료, 공구관리정보 등을 축적하고 있다. 현재도 일부 데이터를 바탕으로 수리부속 및 재료 현황이나 장비정비 현황을 파악하고 있으나, 빅데이터 분석기법을 활용하여 특정 훈련이나 부대활동 이후 많이 발생하는 정비소요 및 부품소요 등을 예측하거나, 부대 장비의 운영환경별 사전 예방정비를 제안하거나 필요 부품을 사전 판단할 수 있다. 또한 지역별 어떤 장비가 가동 시간이 늘어나고 감소하는지 파악하여 수요 동향을 예측하고, 재고와 보충량을 적절히 제어할 수 있으며, 운행 일자별, 운행 거리/시간별 유류소비량을 파악하여 유류 보충시기를 예측할 수 있다. 또한 항공기나 함정 등 장착된 GPS 위치정보와 운행 정보 수집장비로부터 운행패턴을 분석하고 에너지 소비량을 파악하여 효율적인 무기체계 운용방안을 조언해 줌으로써 에너지 소비 효율을 높일 수 있다. 더불어 장비에 대한 수요를 예측하고 수리 및 보수에 대한 일정 판단을 자동화하여 이에 소모되는 행정소요를 줄일 수 있을 것이다. 그 밖에 미군의 공군기지 에서 시행하였던 소비전력 사용량 분석 사례와 같이

한국군도 각 부대별, 각 기지별 에너지 소비전력 모니터링을 통해 에너지 소비 효율화를 달성할 수 있다.

기타 활용 가능한 방안으로는 각종 센서(R/D, 전

〈표 8〉 한국군 분야별 빅데이터 활용방안 요약

분야	내용
전장관리	1. KCTC 데이터 활용 1) 공격작전의 승리요인 분석 → 방법 : 효과측정도구 개발, 다변량 회귀분석, PLS, Group Lasso 적용 등 2) 부대 전투력 운용패턴 분석 → 방법 : 데이터 푸리에 변환 후, Self Organizing Map 방법적용 등 3) 전투원의 생존분석 → 방법 : Cox비례위험모형, 로지스틱 분석 등 4) 사격훈련 효과검증 → 방법 : LDA, QDA, 군집분석 등 2. 군사정보통합처리체계(MIMS), C4I 체계 활용 1) C4I 데이터를 통한 이동패턴분석 2) MIMS 분리분석 → 방법 : 1)과 2)로부터 대응 시나리오 최적화 분석, Simulation 등
인사분야	1. NDEMIS(또는 구DEMIS), 국방 의료정보 체계 1) 맞춤형 의료서비스 분석 → 방법 : ① 흠어진 데이터 통합 ② 증상별 or 신체부위별 or 유발환경별 데이터 분류 분석 : 분류분석(Decision Tree 등) ③ 계절성 질병 예측 : 시계열 분석 등 2) 해외파병자 의료관리 예측 → 방법 : 맞춤형 의료서비스 분석과 유사
군수분야	1. 장비정보정보체계(DELIIS) 1) 장비수요 및 장비유지 예측 → 방법 : DELIIS를 활용 장비의 고장 원인 분석 : 다변량 회귀분석, PLS 등 2) 계절성 or 부대운영 환경별 정비소요, 부품소요 예측 : Markov-Chain 방법, 재고모형 예측 방법 등 2. 해군전술자료처리체계((KNTDS) or 해군 C4I(KNCCS)), 국방 물자정보체계 1) 에너지 수요/관리 예측 → 방법 : 장비 운행일자, 운행거리, 속도 vs. 유류소비량 분석 (시계열 분석, 대기행렬 모형 등)

자파), 플랫폼(함정, 항공기 등) 등에서 수집되는 정형/비정형 데이터를 활용하여 적의 행동패턴을 사실적, 통계적으로 분석함으로써 적의 의도 및 장차징후를 사전에 파악할 수 있을 것이다. 그 밖에 방위사업청 개청(2006년)이후 체계적으로 축적한 데이터를 바탕으로 사업관리에서 획득사업의 지연을 초래하는 부정당 업체에 대하여 국내 조달업체의 계약이행 현황을 분석하여 사전에 식별이 가능할 것이며, 병사들의 군수품 보급에 있어서도 지연되지 않고 적시/적소에 보급될 수 있을 것이다. 방위사업청은 미국의 방위고등연구 계획국이 미 국방부와 별도로 XDATA, ADAMS와 같은 빅데이터 관련 프로젝트를 계획하고 실행하는 것을 주시하여 보다 미래 지향적이고 우리 실정에 맞는 빅데이터 활용 방안을 모색하는 것도 필요하다.

지금까지 살펴본 한국군의 빅데이터 활용가능 분야를 정리하면 <표 8>과 같다. 여기서 제시한 분석방법은 빅데이터 분석방법 중 일부이며, 분석대상과 목적에 따라 다양하게 적용할 수 있다. 뿐만 아니라 <표 7>에서 소개한 기타 한국군이 보유하고 있는 데이터도 아래와 같은 방법으로 활용이 가능하겠다.

5. 결론

빅데이터는 기존의 데이터 분석 방식과 비교하여 ‘데이터 수집비용이 낮아’지고, ‘분석한 정보의 가치가 증대’하였다는 점에서 차이가 있었다. 그리고 3V로 정의되는 빅데이터의 특성이 결국 데이터 수집, 저장, 처리 시 비용의 감소로부터 기인하였다는 점에서 빅데이터는 ‘Cheap Data’라는 개념을 적용해 보았다. 이러한 빅데이터가 등장하게 된 배경을 보면, 우리 사회에 데이터 수집 및 관리를 위한 비용이 감소하였고, 더불어 실시간 취합 및 분석이 가능한 기술 환경이 도래하였으며, 마지막으로 비즈니스 및 경쟁 환경 변화가 발생하였기 때문이라는 것을 알 수 있었다.

본 연구에서는 공공부문과 민간부문에서 다양하게 빅데이터로부터 유용한 통찰력과 지식을 찾아내

고 있는 사례를 정리하였으며, 막대한 군사비를 지출하며 방대한 양의 정보자산을 수집 및 분석하고 있는 미군은 과연 어떻게 (빅)데이터를 관리하는지 자료수집체계에서부터 활용까지 살펴보았다. 따라서 이번 연구는 사회의 전반적인 빅데이터 추세와 맞물려 우리 국방분야에서도 빅데이터를 활용하여 새로운 가치창출을 위한 노력의 필요성을 강조하였다.

사실 한국군은 조직의 특성과 과학화 군의 추구에 힘입어 많은 양의 저비용 데이터를 보유하고 있다. 그러나 이 많은 데이터를 동시에 통합적으로 관리할 수 있는 컴퓨팅 인프라 및 분석기술이 부족하고, 분석된 정보의 가치를 간과하는 조직문화가 남아있는 것이 문제라 할 수 있겠다. 따라서 공공부문과 민간부문, 그리고 선진군의 빅데이터 활용 사례를 잘 연구하여 지금까지 구축해 놓은 군의 각종 시스템과 연동시킨다면 국방분야의 빅데이터 활용은 무수히 많은 분야에서 가능성을 찾을 수 있을 것이다. 다만 빅데이터 활용을 위하여 앞서 문제점으로 제시한 분석기술의 부족한 역량을 키우는 것과 조직에서 정보의 가치를 중요시 여기는 분위기를 확산하는 것이 전제되어야 한다. 이러한 과제가 해결된다면 국방분야의 빅데이터 활용을 통한 새로운 가치창출은 가시화되고, 그 결과 조직의 생산성이 높아질 것이며, 신뢰성이 향상되어 궁극적으로 국방분야에 대한 로열티 및 만족도를 높여 국방력 증진에 도움이 될 것이다.

참 고 문 헌

- [1] 김현곤, “빅데이터 시대 전망과 국가 미래전략”, 한국정보화진흥원, 2012. 10.
- [2] 손미애, “합성환경 표현 및 교환명세(SEDRIS) 소개”, 국방정책연구, 제59권(2003), pp.99-127.
- [3] 시로타 마코토, 빅데이터의 충격, 한빛미디어, 2013.
- [4] 윤미영, 권정은, “빅데이터로 진화하는 세상”, 한국정보화진흥원, 2012. 5.
- [5] 조성우, “Big Data 시대의 기술”, KT경제경영연

- 구소, 2011.
- [6] 함유근, 채승병, “빅데이터, 경영을 바꾸다”, 삼성경제연구소, 2013.
- [7] Anderson, C., “The end of theory : The data deluge makes the scientific method obsolete,” *WIRED*, 2008 http://www.wired.com/science/discoveries/magazine/16-07/pb_theory. Accessed : 18 March, 2014.
- [8] Ball, J., “NSA stores metadata of millions of web users for up to a year, secret files show,” *The Guardian*, 30 September 2013.
- [9] Bascetta, C.A., “VA and Defense Health Care : Progress Made, but DoD Continues to Face Military Medical Surveillance System Challenges,” GAO-02-337T, Jan. 24, 2002.
- [10] Clarke, L., X. Zheng-Bradley, R. Smith, E. Kulesha, C. Xiao, I. Toneva, B. Vaughan, D. Preuss, R. Leinonen, M. Shumway, S. Sherry, P. Flicek, and The 1000 Genomes Project Consortium, “The 1000 Genomes Project : data management and community access,” *Nature*, Vol.9(2012), pp.459-462.
- [11] Chen, H., R.H.L. Chiang, and V.C. Storey, “Business Intelligence and Analytics : From Big Data to Big Impact,” *MIS Quarterly*, Vol.36, No.4(2012), pp.1165-1188.
- [12] Conway, D., “What Data Visualization Should Do : Simple Small Truth,” 2013, <http://drewconway.com/zia/2013/3/26/what-data-visualization-should-do-simple-small-truth>. Accessed: 18 March, 2014.
- [13] Davenport, T.H., J.G. Harris, and R. Morison, “Analytics at Work: Smarter Decisions, Better Results,” *Harvard Business Press*, Massachusetts, 2010.
- [14] Dean, J. and S. Ghemawat, “MapReduce : simplified data processing on large clusters,” *Communications of the ACM*, Vol.51, No.1 (2008), pp.107-113.
- [15] Edwards, John, “Why the Military Loves Big Data,” *BigDATA REPUBLIC*, 2014, http://www.bigdatarepublic.com/author-.asp?section_id=2759&doc_id=257612. Accessed : 18 March, 2014.
- [16] Gantz, J. and D. Reinsel, “Extracting Value from Chaos,” *IDC IVIEW*, 2011, <http://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf>. Accessed : 18 March, 2014.
- [17] Gantz, J. and D. Reinsel, “THE DIGITAL UNIVERSE IN 2020 : Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East,” *IDC IVIEW*, 2012, <http://idcdocserv.com/1414>. Accessed : 18 March, 2014.
- [18] Gartner, “Top 10 Strategic Technology Trends for 2013,” *Special Report*, 2012, <http://www.gartner.com/technology/research/top-10-technology-trends/>.
- [19] Ginsberg, J., M.H. Mohebbi, R.S. Patel, L. Brammer, M.S. Smolinski, and L. Brilliant, “Detecting influenza epidemics using search engine query data,” *Nature*, Vol.457(2009), pp.1012-1014.
- [20] Goldberg, D.E. and J.H. Holland, “Genetic Algorithms and Machine Learning,” *Machine Learning*, Vol.3, No.2-3(1988), pp.95-99.
- [21] Graham, M., “Big data and the end of theory?,” *The Guardian*, 2012. <http://www.guardian.co.uk/news/datablog/2012/mar/09/big-data-theory>. Accessed : 18 March, 2014.
- [22] Gungor, V.C., D.Sahin, T. Kocak, S. Ergut, C. Buccella, C. Cecati, and G.P. Hancke, “Smart Grid and Smart Home,” *IEEE Industrial Electronics Magazine*, Vol.12(2012), pp.18-34.
- [23] Han, J., M. Kamber, and J. Pei, “Data Mining : Concepts and Techniques,” *Elsevier*, San

- Francisco, CA, 2006.
- [24] IBM Business Analytics, "IBM SPSS Predictive Analytics Aids Offender Management," 2011.
- [25] Insinna, V., "Air Force Uses Big Data to Glean Energy Saving," *National Defense*, July, 2013.
- [26] Jacobs, Melinda, "Click, click, click. Zynga and the gamification of clicking," *The Italian Journal of Game Studies*, Vol.1(2012), pp.1-6.
- [27] Krouse, W. and B. Elias, "Terrorist Watchlist checks and Air Passenger Prescreening", CRS Report for Congress, 2009.
- [28] Lavallo, S., M. S. Hopkins, E. Lesser, R., Shockley, and N. Kruschwitz, "Analytics : The New Path to Value," *MIT Sloan Management Review*, 2010.
- [29] Manyika, James, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers, "Big data : The next frontier for innovation, competition, and productivity," *McKinsey Global Institute*, 2011, http://www.mckinsey.com/mgi/publications/big_data/pdfs/MGI_big_data_exec_summary.pdf. Accessed : 18 March, 2014.
- [30] Obenshain, M.K., "Application of Data Mining Techniques to Healthcare Data," *Infection Control and Hospital Epidemiology*, Vol.25, No.8(2004), pp.690-695. <http://www.jstor.org/stable/10.1086/502460>. Accessed: 18 March, 2014.
- [31] OSTP, Fact Sheet : Big Data Across the Federal Government, March 29, 2012.
- [32] Rubertone, M.V. and J.F. Brundage, "The Defense Medical Surveillance System and the Department of Defense Serum Repository : Glimpses of the Future of Public Health Surveillance," *American Journal of Public Health*, Vol.92, No.12(2002), pp.1900-1904.
- [33] The Economist, *The Data Deluge*, 25 Feb. 2010, pp.2-16.
- [34] The Guardian, <http://www.theguardian.com/media/pda/2010/may/13/twitter-tweetminster-election>, 2010.
- [35] Vosen, S. and T. Schmidt, "Forecasting Private Consumption : Survey-Based Indicators vs. Google Trends," *Journal of Forecasting*, Vol.30(2011), pp.565-578.
- [36] Wasserman, S. and K. Faust, "Social Network Analysis : Methods and Applications," Cambridge University Press, 1994.
- [37] Watson, R.T., M.-C. Boudreau, S. Li, and J. Levis, "Telematics at UPS : En route to energy informatics," *MISQ Executive*, Vol.9, No.1(2010), pp.1-11.
- [38] Wolfers, J., "Point Shaving : Corruption NCAA Basketball," *American Economic Association*, Vol.96, No.2(2006), pp.279-283.