

# 효과적인 빅데이터분석 기획 접근법에 대한 융합적 고찰

남수현\*, 노규성\*\*

한남대학교 경영정보학과\*, 선문대학교 경영학부\*\*

## A Study on the Effective Approaches to Big Data Planning

Su Hyeon Namn\*, Kyoo-Sung Noh\*\*

Dept. of MIS, Hannam University\*

Dept. of Business Administration, Sunmoon University\*\*

**요약** 빅데이터분석은 조직의 문제해결을 위한 융합적 수단이다. 효과적인 문제해결을 위해서는 문제의 형태, 데이터의 유형 및 존재여부, 데이터 분석역량, 분석을 위한 기반정보기술의 수준 등 다양한 요인을 융합적으로 고려하여 문제해결의 접근법이 결정되어야 한다. 본 연구에서는 기획 접근법으로 논리적인 하향식 접근법, 데이터기반의 상향식 접근법, 그리고 문제해결 환경의 불확실성을 극복하기 위한 프로토타이핑 접근법 등 세 가지 유형을 제안한다. 특히, 이 유형 중에서 창의적 문제해결과 상향식 접근법이 어떤 연관성을 갖는지 살펴본다. 또한 데이터 거버넌스와 데이터 분석역량을 융합적으로 고려하여 조직의 빅데이터분석의 소성과 관련한 주요 전략적 이슈를 도출한다.

**주제어** : 빅데이터분석 기획, 데이터 융합, 하향식 접근법, 상향식 접근법, 프로토타이핑 접근법, 데이터 거버넌스

**Abstract** Big data analysis is a means of organizational problem solving. For an effective problem solving, approaches to problem solving should take into account the factors such as characteristics of problem, types and availability of data, data analytic capability, and technical capability. In this article we propose three approaches: logical top-down, data driven bottom-up, and prototyping for overcoming undefined problem circumstances. In particular we look into the relationship of creative problem solving with the bottom-up approach. Based on the organizational data governance and data analytic capability, we also derive strategic issues concerning the sourcing of big data analysis.

**Key Words** : Big data planning, Data Convergence, Top-down approach, Bottom-up approach, Prototyping approach, Data governance

### 1. 서론

최근 빅데이터 융합 분석에 대한 관심은 조직에서 발생하는 데이터의 규모, 다양성, 속도 등에서 과거와의 차

이, 빅데이터 분석을 가능하게 하는 데이터 처리 및 분석 기술 등의 진전으로 높아가고 있다. 이는 기업 내부, 기업과 기업 간 또는 기업과 고객 사이에 발생하는 모든 거래 데이터가 인터넷을 기반으로 수집되고 저장되는 데이터

\* The research was supported by the 2014 Hannam University Research Fund.

Received 5 November 2014, Revised 8 December 2014

Accepted 20 January 2015

Corresponding Author: Kyoo-Sung Noh (SunMoon University)

Email: ksnoh@sunmoon.ac.kr

© The Society of Digital Policy & Management. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

가 존재하기 때문에 가능하다. 더욱이 소셜네트워크를 중심으로 발생하는 관계 데이터와 사물인터넷에서 자동으로 포착되는 데이터의 양과 다양한 데이터 형태는 과거의 데이터처리기술로는 불가능한 새로운 융합기반을 요구하고 있다. 이런 관점에서 하둡, R 등 최근의 빅데이터 관련 기술에 대한 일반의 높은 관심은 이와 같은 기술의 이해가 마치 빅데이터분석의 전부인 양 오해를 할 수 있는 위험이 있다. 왜냐하면, 빅데이터의 본질적 목적은 빅데이터로부터 전략적으로 효용이 높은 정보 또는 지식을 추출하고 이를 경영활동에 활용함으로써 조직의 경쟁력을 높이는 것이기 때문이다.

빅데이터를 분석하는 많은 기업들이 다양한 분석이 가능한 솔루션 도입, 빅데이터 분석 전문가들을 채용함으로써 빅데이터 분석 시 통찰력을 증대시키려는 노력을 경주하고 있다. 그러나 아직까지 많은 기업들은 빅데이터에 대한 올바른 기획 접근법을 수립하고 적절히 활용하고 있지는 못하고 있다. 이는 빅데이터에 대한 인식과 이해 부족 및 적절한 빅데이터 분석에 대한 지식 부족에 그 원인이 있다고 할 수 있다.

이에 본 연구는 빅데이터분석 기획에 대한 접근법을 제시하고 전략적 이슈를 발굴하여 빅데이터를 추진하는 많은 기업들에게 올바른 분석 기획 수립 및 실행을 통해 각 기업의 빅데이터 효과성을 제고하도록 하는 데에 목적이 있다.

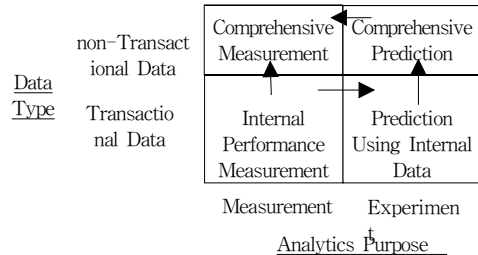
논문의 서술은 다음과 같다. 2장에서는 빅데이터분석 프레임워크와 기존 정보시스템계획과의 비교를 통하여 빅데이터분석 기획 접근법에 대한 근거를 마련한다. 이어 3장에서는 빅데이터분석 추진 환경을 기반으로 분석 기획 접근법의 세가지 유형, 즉 하향식 접근법, 상향식 접근법, 프로토타이핑 접근법을 중심으로 내용을 기술한다. 4장에서는 전략적 이슈를, 그리고 5장에서는 결론이 전개된다.

## 2. 빅데이터분석 프레임워크

본 장에서는 빅데이터분석 프레임워크와 기존의 정보시스템계획과의 차이점을 근거로 빅데이터분석에 대한 융합적 접근법에 대한 필요성을 기술한다. 먼저 빅데이터분석은 데이터의 유형과 분석의 목적에 따라 [Fig. 1]

과 같은 2x2 프레임워크로 정리한다. 이 프레임워크의 기본 체계는 Parise et al.[11]에 의해 제안되었는데, 빅데이터분석을 위한 접근법이나 기회발굴을 위한 유용한 도구로 사용될 수 있다고 판단된다.

이 프레임워크는 사전에 분석 목적이 명확하게 정의될 수 있는지의 여부와 데이터의 원천이 조직내부의 정형화된 운영 데이터인지에 따라 분석의 접근이 달라야 함을 의미한다. 즉, 데이터 유형은 조직의 운영 데이터베이스로부터 생성되는 거래처리 데이터와 소셜네트워크 사이트 등에서 발생하는 비거래처리 데이터로 구분된다. 한편, 데이터분석의 목적은 조직의 과거 성과에 대한 측정과 미래의 전략계획 도출을 위한 실험목적으로 구분하고 있다.



[Fig. 1] Virtuous cycle framework for Big data analytics

Source: P. Parise, B. Iyer, and D. Vesset, 2012 with modification

이 프레임워크의 설명은 다음과 같다. 첫째, “내부성과 측정 (Internal Performance Measurement)”은 측정관리가 필요한 성과지표를 중심으로 빅데이터로부터 사전에 정의된 질의나 다차원분석을 통하여 의사결정자에게 주로 그래픽을 이용한 대시보드 형식으로 빅데이터 분석결과가 제공된다. 이 영역의 활용은 균형성과표에 효과적이고, 계획 대비 성과를 이용한 문제인식 단계에 주로 활용될 수 있다.

둘째, “내부데이터추정 (Prediction Using Internal Data)” 영역은 조직의 정형화된 운영 데이터를 이용하여 특별질의(ad-hoc)나 가설검정 등에 사용이 가능하다. 예를 들면, 고객의 세그먼트이전과 각 세그먼트에 적합한 광고 메시지의 선택이나, 효과적인 교차판매 제품의 선택과 관련한 분석을 제공할 수 있다. 기존의 데이터마이닝을 통하여 도메인 지식에 기반한 기회 발굴에 많이 활용되는 영역이다.

셋째, 거래처리 이외의 소셜네트워크나 사물인터넷 환경에서 생성되는 다양한 형태의 비정형데이터까지도 포함하여 측정을 하는 “포괄적측정 (Comprehensive Measurement)”을 할 수 있다. 예를 들어, 사전에 정해진 지표와 관련하여 소셜네트워크에서 발생하는 데이터를 이용하여 사용자들의 반응을 측정하는데 빅데이터 분석이 활용되는 영역이다. 측정 대상의 지표로서는 브랜드 인지도, SNS공간에서의 활동과 상호작용 정도, 추천이나 구전의 정도 등을 이용한 마케팅 활동의 평가를 들 수 있다 [11].

넷째, “포괄적추정 (Comprehensive Prediction)”은 고객이 남긴 상품 관련 후기, 아이디어, 리뷰 등과 같은 비정형데이터를 이용 및 감성분석을 통하여 [11] 신상품개발과 같은 전략적 의사결정에 빅데이터 분석결과를 실험적으로 반영할 수 있게 지원해 줄 수 있는 영역이다.

본 프레임워크의 요체는 [Fig. 1]에 제시된 화살표이다. 조직은 분석대상의 데이터를 확장하고, 분석목적의 다양화와 고도화를 동시에 추구하여, 데이터분석 프로세스를 관리 대상으로 간주해야 한다는 것이다. 즉, “측정”의 경우, 의사결정자에게 주요성과지표(KPI)에 대한 시각화 도구를 이용하여 현 상태를 지속적으로 점검하고 문제를 조기에 발견하여 능동적으로 해결할 수 있는 기반을 제공할 수 있다. “실험”과 관련한 빅 데이터 분석은 아직 조직에서 모르고 있던 지식이나, 알려지지 않았던 새로운 사실을 발굴하는 의미있는 지식발굴 작업이고 그 결과를 이용하여 새로운 비즈니스 규칙을 발견하고 운영 시스템에 실제로 적용할 수 있는 기회를 제공한다. 또한 적용된 지식의 가치를 조직 내외부의 데이터를 이용하여 측정과 평가를 통하여 새로운 빅데이터 실험을 하는 선순환 구조로 연계(Fig. 1의 빅데이터 프로세스 참고)하여 조직의 분석역량과 결과적으로 경쟁력을 높일 수 있게 된다.

빅데이터분석 기획과 유사하게 많이 사용되는 용어로 정보시스템계획(Information Systems Planning, ISP)이 있는데, ISP는 조직에서 정보기술 또는 정보시스템을 전략적으로 활용하기 위하여 조직의 내외부 환경을 분석하여 기회나 문제점을 파악하고, 사용자의 요구분석을 통하여 정보시스템 도입을 위한 제반 절차를 말한다. 이들은 문제해결 방안을 제시한다는 점에서 동일하나, 해결방안의 유형과 특성 측면은 <Table 1>과 같이 차이점이

있다. 따라서 기존의 ISP 방법론을 그대로 빅데이터분석 기획 방법론으로 적용하는 것은 무리가 있다고 할 수 있다.

<Table 1> ISP vs Big data analysis

Characteristics	ISP	Bid Data
Planning Boundary	Unit/Integrated System	Module/Rule of a system
Problem solving basis	Process	Data
System development phase	Big bang	Incremental
Evaluation criteria	Quality and performance	Utility of discovered knowledge
Development duration	> months	< months
Project emergence	Regular basis	haphazardly
Life cycle	Mid/long term	sensitive to users' needs
volatility/flexibility	Low	High/agile
Major value	Efficiency	Experiment/ Analytics capability

### 3. 빅데이터분석 기획 접근법

빅데이터분석의 영역은 다양하게 정의될 수 있으나, 본 논문에서는 조직에서 문제해결 또는 경쟁력확보의 수단으로 빅데이터에 기반한 접근법을 사용할 때, 구체적으로 빅데이터분석 문제를 도출할 때까지, 즉, 비즈니스 요구사항을 빅데이터 문제로 변환하는 단계까지의 과정에 국한한다. 즉 데이터를 실제적으로 분석하는 과정은 고려하지 않는다. 이런 관점에서 우리가 조망하는 단계는 빅데이터분석 기획에 해당한다고 할 수 있다.

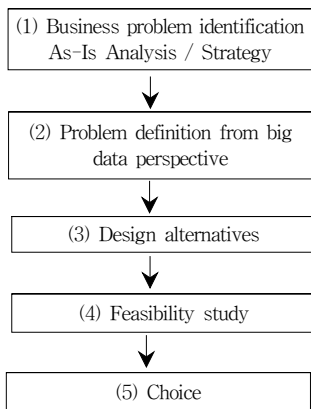
기획단계는 비즈니스 목표 달성 최적화를 위해 의사결정과 실행과정에 필요한 정보와 인사이트를 과학적 분석을 통해 제공하는 분석체계를 수립하는 것으로 일반적으로 인식되고 있다. 즉, 분석기획은 의사결정을 위한 분석기회를 발굴하고 요구사항을 구체화하여 필요한 분석과 데이터를 정의하는 접근법이다. 그러나 이와 같은 접근방법은 전통적인 하향식 문제해결방법으로 문제의 구

체적 정의가 가능하고, 필요한 데이터가 존재하고, 데이터를 분석할 수 있는 분석역량을 보유하고 있다는 전제 조건에서 가능하다. 만약 이러한 조건을 만족시키지 못한다면 이와 같은 전통적 접근방법은 효과적이지 못하다. 본 연구에서는 이에 대한 대안 또는 보완적 접근법으로서 사항식과 프로토타이핑 접근방법을 제안한다. 또한 빅데이터분석이 성공적으로 수행되기 위해서 조직에서 고려해야 할 전략적 이슈를 제시한다.

### 3.1 사항식 접근법 : 해법을 찾아라

사항식 접근법은 빅데이터 분석 기획과정을 단계화하여 각 단계에서 수행될 활동들을 명확히 정의하고, 그에 따르는 활동들을 수행하고 산출물을 점검하고 확인함으로써 성공적인 기획활동을 도모하도록 하는 방법론으로 정의할 수 있다.

사항식 분석 기획 접근법은 현황분석을 통해서 또는 인식된 문제점 혹은 전략으로부터 기회사 문제를 탐색하고, 해당 문제에 대한 사용자의 요구분석, 해결방안 설계, 그리고 빅데이터 분석의 타당성 검토를 거쳐 최종적으로 빅데이터 분석 과제를 도출하는 과정으로 이루어져 있다. 이와 같은 절차는 [Fig. 2]와 같이 표현할 수 있다[8].



[Fig. 2] Big data analysis planning stages

#### 3.1.1 과제발굴 단계

이 단계는 빅데이터분석을 통해 고객서비스 및 의사결정 효율성 향상을 위한 기회식별이나 경영목표달성을 위해 해결해야 할 문제를 식별하는 단계이다. 식별된

문제의 예로는 고객 이탈현상 심화, 공장기계의 잦은 정지로 납기 지연 및 손실초래, 환경 악화로 온난화 등 기후변화 심각현상 등을 들 수 있다. 이와는 별개로, 다른 조직에서 어떤 문제에 빅데이터를 적용하였는지의 사례분석도 문제식별에 유용할 것이다.

#### 3.1.2 빅데이터 관점에서의 문제정의

식별된 비즈니스 문제를 빅데이터 문제로 변환하여 정의하는 단계이다. 문제의 정의 및 요구사항은 최종사용자(고객, 실무자, 정책결정자) 관점에서 이루어져야 한다. 발굴된 문제와 이에 대응하는 빅데이터 문제로의 변환은 <Table 2>와 같다[9].

<Table 2> Business problem vs Big data problem

Business problem	Big data problem
Increased customer churning	Identification of factors affecting churning, estimation of likelihood of churning
Sales loss due to machine breakdown	Finding signals leading to machine breakdown, identification of machine breakdown indicators
Unstable supply chain due to earth warming	Sensitivity of earth warming on supply chain

#### 3.1.3 해결대안 설계

이 단계에서는 정의된 빅데이터 문제를 해결하기 위한 다양한 방안이 모색된다. 동일한 빅데이터 문제라 해도 어떤 데이터 또는 분석시스템을 사용할 것인가도 고려되어야 한다. 즉, 기존 정보시스템의 단순한 보완으로 분석이 가능한지, 엑셀 등의 기존의 간단한 도구로 분석 가능한지, 또는 하둡 등 심화된 빅데이터 분석 도구를 통해 보다 체계적이고 심도있는 방안 등 여러 대안이 도출될 수 있다.

#### 3.1.4 타당성 분석

제시된 대안에 대한 평가를 위한 단계로 다음과 같은 다각적인 타당성 분석이 수행된다.

- (1) 경제적 타당성: 비용 대비 편익 분석으로 비용항목은 데이터, 시스템, 인력, 유지보수 등과 같은 분석 비용으로 구성된다. 편익으로는 분석

결과를 적용함으로써 추정되는 실질적 비용절감, 추가적 매출과 수익 등과 같은 경제적 가치로 산출된다.

- (2) 데이터 및 기술적 타당성: 빅데이터 분석에는 데이터 존재 여부, 분석 시스템 환경, 그리고 분석 역량이 필요하다. 효과적인 평가를 위해서는 비즈니스 지식과 기술적 지식이 요구되어 비즈니스분석가, 데이터분석가, 시스템엔지니어 등과의 협업이 필수적이다.

### 3.1.5 선택

여러 대안 중에서 평가과정을 거쳐 가장 우월한 대안을 선택하여 이를 프로젝트화하고, 프로젝트 계획의 입력물로 설정한다.

## 3.2 상향식 접근법 : 데이터가 말하게 하라

한 제약조사는 특허기간이 만료된 의약품 약 2,000 종류의 데이터를 분석, 상호 결합하여 새로운 의약품으로 개발하려는 시도를 하였다[7]. 이 시도에 의하면, 의약품 집합으로부터 두 개의 조합을 선택할 수 있는 방법은 백만 개 이상이다. 그 회사는 새로운 결합의 효과성을 검증하기 위하여 정보기술과 로봇기술을 통해 데이터를 분석하고 있다. 이러한 분석 사례는 하향식 문제해결 혹은 지식창출이 아니라 경험적인 과거 데이터를 무작정 결합하여 상향식으로 새로운 정보 혹은 지식을 얻고자 하는 새로운 분석 패러다임이라 할 수 있다.

여기에서는 전통적인 하향식 문제해결방식과 대비되어, 기업에서 보유하고 있는 다양한 원천으로부터 생성되고, 다양한 형태로 존재하는 데이터로부터 비즈니스 문제와 빅데이터 문제를 도출하고 통찰력과 지식을 얻고자 하는 상향식 접근방법을 기술한다.

### 3.2.1 하향식 접근법의 문제

지금까지 가장 일반적으로 사용되고 있는 문제해결과 정으로는 Simon이 제안한 단계별 접근법이다. Simon의 논리적 접근법과 비교하여, Mintzberg & Westley [6]는 시각적 접근과 행위적 접근을 이용한 문제해결인 "Seeing first"와 "Doing first"를 제안하였다. 시각적 접근 관점에서는 시각적인 관찰을 통해 순간적으로 얻어지

는 통찰에 기반한 문제해결의 중요성을 강조한다. 통찰로부터 복잡한 문제가 창조적으로 해결될 수 있는 기회를 제공한다는 것이다. 행위적 접근 관점에서는 실제로 어떤 행동에 옮김으로서 거기에서 발생하는 결과 (데이터)에 입각하여 어떤 것이 작동되는지의 결과로 문제의 본질을 이해하게 되고, 그로부터 문제해결이 가능하다고 한다. 행위적 접근법은 불확실한 환경에서의 문제해결에 효과적일 수 있다고 주장된다. 이와 같은 시각적 접근 및 행위적 접근 관점과 비교하여 기존의 접근방법인 사고적 접근 (Thinking first)방법은 문제의 구조가 분명하고, 문제를 해결하고 해결책을 도출하기 위한 데이터가 의사결정자에게 주어져 있음을 가정한다. 그러나 이는 Gorry et al.[3]에서 이미 설명한 바와 같이, MIS 시스템은 문제의 구조화 수준을 높이는 역할을 수행함으로써, 만약 문제의 구조가 명확하다면 이는 더 이상 의사결정자의 고민거리가 될 수 없는 것이다. 따라서, 기존의 사고적 접근방법에 기반한 문제해결, 즉, why에 기반한 문제해결과정은 최근의 복잡한 경영환경에서 발생하는 문제해결을 위해서 적합성을 점차 잃고 있다고 판단된다.

### 3.2.2 Fact 기반의 문제해결

스탠포드 대학의 d스쿨의 모토인 디자인적 사고 (design thinking) 접근법[12]은 전통적인 분석적 사고와 디자인적 사고를 차별화시킴으로서 과거의 문제해결 방식의 문제점을 지적하고 있다. 즉, 창조적인 문제해결은 "이미 우리가 알고 있는 것" 즉 분석에 의한 접근법의 함정을 경고하고 있다. 대신 우리의 관점으로부터 현장관찰과 감정이입, 즉 대상의 관점으로서의 전환이 이루어져야 한다는 것이다. 즉 우리의 관점은 분석적으로 사물을 인식하려는 'why'를 강조지만, 이는 우리가 알고 있다고 가정하는 것에 오류가 매우 많을 수 있다는 문제가 있다. 이를 사물이 있는 그대로 인식하는 'what' 관점에서 보아야 한다는 것이다. 이는 Mintzberg et al.[6]의 시각적 접근과 행위적 접근법과 일맥상통 한다. 우리가 중심이 되어 관찰할 데이터의 유형을 식별하고 어떻게 수집하는 등의 접근이 아닌, 객관적으로 존재하는 데이터 그 자체를 관찰하고 실제적으로 행동에 옮김으로서 대상을 좀 더 잘 이해할 수 있을 것이다. 이와 같은 점을 고려하여 d스쿨에서는 감정이입을 특히 강조하고 있다. 또한 Seemann[13]은 혼합적 통찰 (hybrid insights)라는 개념

을 적용하여 'why'와 'what'의 조화를 강조하고 있다.

### 3.2.3 인과분석 vs 상관분석

일반적으로 상향식 데이터 분석은 비감독학습(unsupervised learning) 방법에 의해 수행된다. 비감독 학습은 데이터 분석의 목적이 명확히 정의되거나 특정 필드의 값을 구하는 것이 아니고, 데이터 자체의 결합, 연관성, 유사성 등을 중심으로 데이터의 상태를 표현하는 것이다. 데이터 마이닝 기법을 예로 들면, 장바구니분석, 군집분석, 기술통계 및 프로파일링 등이 이에 속한다. 참고로 이와는 반대로 명확한 목적하에 데이터분석을 실시하는 것은 감독학습(supervised learning)으로서, 분류, 추측, 예측, 최적화 등으로 사용자의 주도하에 분석을 실시하고 지식을 도출하는 것이 목적이다.

기존의 통계적 분석에서는 인과관계 분석을 위해 가설을 설정하고 이를 검증하기 위해 모집단으로부터 표본을 추출하고 그 표본을 이용한 가설검정을 실시하였다. 그러나 빅데이터 환경에서는 이와 같은 논리적인 인과관계 분석을 거치지 않더라도 상관관계분석 또는 연관분석을 통하여 다양한 문제해결에 도움을 받을 수 있다. 즉, 인과관계(Know-why)로부터 상관관계(Know-affinity) 분석으로의 이동이 빅데이터 분석에서의 주요 변화라고 할 수 있다.

상관관계는 두 데이터 집합간의 통계적 관련성을 측정한다. 상관성이 높다는 것은 하나의 데이터 값이 변화할 때, 다른 하나도 변할 가능성이 높다는 것을 의미한다. 반대로 상관성이 낮다는 것은 하나의 데이터 값이 변해도 다른 것에는 변화가 일어나지 않는다는 것이다. 상관성은 현상의 내부 원리를 알게 해주는 해결방안을 제시할 수 있고, 문제를 인식하는 유용한 대응물이 될 수도 있다. 두 데이터간의 상관관계가 다음과 같다고 하자. A와 B간의 상관관계가 높고, B와 C간의 상관관계가 높을 때, 우리는 A와 C가 결과적으로 상관관계가 높다고 할 수 있다. 또한 A와 B는 C의 프록시(대용물)로서 역할을 할 수 있음을 의미한다[5]. 또한 만약 데이터 A의 측정이 매우 힘들거나 높은 비용이 소요될 때, A를 대신하여 B가 사용될 수 있다.

한편, 연관분석이나 상관관계분석 결과에 대한 해석과 관련하여 일반적으로 다음과 같은 세 가지의 설명이 가능하다[1]. (1) 의미있는 규칙(Actionable Rules)이나 지

식으로 사용가능함. (2) 당연한 결과(trivial results)를 보여주는 것에 지나지 않음. (3) 설명이 불가능한 규칙(Inexplicable Rules)으로 적용할 필요가 없음. 그러나 3)과 같이 설명이 어렵다 하더라도 여전히 유용한 정보가 될 수 있다. 예를 들어, "주황색 차량은 차량 손상 비율이 다른 색 차량의 1/2에 지나지 않는다"라는 상관관계를 발견했다고 하자. 차량의 색상과 손상과는 어떤 명확한 인과관계를 찾기 어렵지만, 그래도 그 결과는 중요한 의미를 전달하고 있다[4].

### 3.3 프로토타이핑 접근법 : 시행착오로 찾아라

프로토타이핑 접근법은 사용자가 정보요구사항이나 데이터를 정확히 규정하기 어렵고, 데이터 소스도 명확히 파악하기 어려운 상황에서 일단 분석을 시도해보고 그 결과를 확인해가면서 반복적으로 개선해 나가는 방법을 말한다[10]. 하향식 접근방법은 문제가 정형화되어 있고, 문제해결을 위한 데이터가 완벽하게 조직에 존재할 경우에 효과적이다. 이와 반하여, 프로토타이핑 방법론은 비록 완전하지는 못하다 해도 신속하게 해결책이나 모형을 제시함으로써 이를 바탕으로 문제를 좀 더 명확하게 인식하거나, 필요한 데이터를 식별하고 구체화할 수 있게 함으로써 빅데이터분석에 유용한 방법론이다.

특히 빅데이터분석 환경에서 프로토타이핑의 필요성을 정리하면 다음과 같다.

- (1) 문제에 대한 인식 수준: 정보시스템 개발 환경에서와 마찬가지로 문제정의가 불명확하거나, 이전에 접해보지 못한 새로운 문제일 경우, 사용자는 프로토타입을 이용하여 문제를 구체화하는데 도움을 받을 수 있다.
- (2) 필요 데이터 존재 여부의 불확실성: 문제해결을 위해 필요한 데이터의 집합이 모두 존재하지 않을 경우, 그 데이터의 수집을 어떻게 할 것인지? 또는 그 데이터 대신에 다른 데이터로 대체할 것인지? 등에 대한 빅데이터 사용자와 분석가 간의 반복적이고 순환적인 협의 과정이 필요하다.
- (3) 데이터의 사용 목적의 가변성: 데이터의 가치는 사전에 정해진 수집목적에 따라 확정되는 것이 아니고, 그 효용이 지속적으로 변화할 수 있다. 오히려 빅데이터 환경에서는 데이터의 가치가 본질적인

사용목적으로부터가 아닌, 부수적인 사용목적에 의해서 달성된다는 것이다. 따라서 조직에서 보유 중인 데이터라 하더라도 데이터 사전에 기록된 기존의 데이터 정의를 재검토하여 데이터의 사용 목적과 범위를 확대할 수 있을 것이다. 예를 들면, 이동통신사에서 수집하는 사용자의 위치추적은 사용자의 호출을 효율적으로 처리하기 위한 원래의 목적으로부터, 사용자들이 특정 시간에 많이 모이는 장소가 어디인가의 정보를 이용한 다양한 애플리케이션에 활용이 가능하다[4].

## 4. 전략적 이슈

빅데이터분석에서 요구되는 방대한 양의 데이터, 다양한 형식의 데이터, 그리고 빠른 속도로 생성되는 데이터, 그리고 조직의 데이터 분석역량 등은 조직에서 빅데이터 프로젝트에의 접근성을 낮추는 요인으로 작용할 수 있다. 이와 같은 어려움을 극복하기 위해 다음과 같은 전략적 관점을 고려해야 한다.

### 4.1 데이터 거버넌스

빅데이터분석에 필요한 데이터의 가용성과 데이터의 품질은 빅데이터분석의 성과를 결정하는 요인이다. 따라서 적기에 지원받기 위한 데이터 거버넌스 체계의 확립이 필수적이다. 데이터 거버넌스의 역할은 조직의 데이터관리에 대한 총괄적 책임을 지는 조직으로 데이터관리에 따르는 방침, 절차, 표준, 통제, 실행계획 등을 수립한다. 데이터 거버넌스의 궁극적 목적은 빅데이터 분석을 통한 조직의 문제해결역량의 향상에 있고 이를 위해서 다음과 같은 요구조건을 충족시켜야 한다. (1) 데이터의 가시성(visibility)을 보장하여 조직구성원들이 필요한 데이터에 유연하게 접근할 수 있어야 한다. (2) 데이터 품질, 데이터 무결성, 데이터 일관성, 데이터 최신성, 데이터 보안, 개인정보 보안 등을 통하여 조직원들의 데이터 활용 및 신뢰도를 높일 수 있어야 한다. (3) 데이터관리 규칙 준수여부 및 책임소재를 명확하게 하기 위하여 데이터 소유자(owner), 데이터 관리자(steward), 데이터 책임자, 사용자 등과 같은 데이터 이해관계자의 식별이 필요하다. 또한 데이터의

수명주기와 관련한 데이터의 생성, 처리 및 가공, 저장, 사용자에게 데이터 제공, 데이터 폐기 등과 관련한 역할을 정의해야 한다.

또한 현재 보유하고 있는 조직의 데이터에 대해 (1) 이미 보유하고 있는 데이터는 어떤 것이 있으며, (2) 데이터의 사용용도 (1차적/2차적)가 무엇인지, (3) 어떤 경로로 수집되었으며, (4) 어디에 저장되어 있으며, (5) 어떤 품질 수준이며, 그리고 (6) 데이터가 유용하게 사용되기 위해서는 어떤 전처리 작업이 필요한가 등에 대한 평가가 이루어져야 한다.

### 4.2 빅데이터 분석역량

빅데이터분석을 위해서는 데이터의 가용성 및 거버넌스와 더불어, 빅데이터 분석역량이 중요하다. 빅데이터분석 역량은 다음과 같이 나눌 수 있다[4]. (1) 데이터기술역량 단계: 무엇인 일어났는지를 데이터로부터 아는 역량으로 데이터의 요약, 시각화, 상관관계, 기술통계의 작성 등과 관련한 역량이 포함된다. (2) 진단역량 단계: 어떤 문제나 이벤트가 왜 발생했는지 그 원인을 찾아낼 수 있는 역량이다. 즉 문제와 관련한 데이터를 단순히 기술하는 수준이 아니고, 인과관계를 밝혀낼 수 있는 역량이 해당된다. (3) 예측역량 단계: 과거의 문제로부터 예견되는 미래는 무엇인가를 제시할 수 있는 역량을 말한다. (4) 처방역량 단계: 내가 바라는 미래의 상태를 만들기 위해서 어떤 처방을 내릴 수 있는가에 대한 능력이 해당된다.

빅데이터 분석역량 단계는 각 조직에서 보유하고 있는 현행분석역량(as-is) 대비 요구분석역량(to-be) 간의 차이를 측정하여 조직의 빅데이터분석에 대한 전략적 의사결정을 할 수 있는 수단으로 사용될 수 있다. 만약 요구분석역량이 현행분석역량에 비해 현저히 높을 경우에는 조직에 필요한 데이터가 있더라도 빅데이터 분석을 할 수 없게 된다. 이럴 경우, 분석역량의 인적자원 채용 혹은 분석과제의 아웃소싱 방안을 강구할 수 있다. 따라서 조직의 빅데이터 분석역량의 수준은 데이터 보유 수준과 함께 빅데이터 프로젝트의 성공적 수행 여부를 결정짓는 핵심 요인이라 할 수 있다.

Hayashi[4]가 제시한 1단계의 역량은 주로 빅데이터 자체의 처리를 목표로 하는 역량으로서 기술적 성격이 강하다. 그러나 수준이 높아질수록 기술적 역량

자체보다는 비즈니스 역량의 중요성이 강조된다. 따라서 낮은 수준의 역량은 일반 재화(commodity)와 같이 아웃소싱이 상대적으로 용이하지만, 조직의 핵심역량을 결정하는 것은 아니다[2]. 조직의 데이터 거버넌스 수준과 분석역량과의 관계를 통해 각 영역에서 취할 수 있는 전략을 살펴보면 <Table 3>과 같다.

<Table 3> Data governance vs Analytics capability

Data Governance Capability Gap*	Existence of Data governance	
	Data not available	Data available
+	Data infra required such as data warehouses	In-house analytics
-	Need to upgrade analytic capability	Outsourcing or Recruiting human resources

\* Capability Gap = Current analytic capability - Required analytic capability

빅데이터분석과 관련한 주요 장애물은 빅데이터 프로젝트 수행 중에 발생하는 데이터 처리기술의 미숙이 전부는 아니다. 많은 조직은 자기들이 안고 있는 문제를 잘 알고 있고, 데이터도 보유하고 있지만, 그 데이터를 어떻게 탐색하고 활용할 것인가에 대한 빅데이터 프로젝트로 만들어내는 그 자체에 어려움이 많다. 즉 데이터 분석기술 그 자체보다는 데이터를 어떻게 전략적으로 활용할 것인가에 대한 데이터분석 로드맵의 도출 역량이 중요시 되고 있다[8].

## 5. 결론

본 연구는 빅데이터분석에서 기획부문에 초점을 맞추어 빅데이터분석 기획 접근법으로서 하향식, 상향식, 그리고 프로토타이핑 방법론을 제시하였다. 창의적 문제해결이 강조되는 경영환경에서 상향식과 프로토타이핑 접근방법은 비즈니스 전략 또는 ISP 수립에 익숙한 기존의 분석가들에게 새로운 도전을 제시할 수 있다고 판단된다. 그러나 본 연구는 문헌연구를 토대로 접근법을 제시하여 현장적용시의 실무적 장애나 문제를 실증적으로 연구하지 못한 한계를 갖는다. 향후

빅데이터 기획 프로젝트 사례가 축적되어 실증적 연구가 이어지기를 기대한다.

## ACKNOWLEDGMENTS

The research was supported by the 2014 Hannam University Research Fund.

## REFERENCES

- [1] M. Berry and G. Linoff, *Data Mining Techniques*, 2nd Ed., Wiley, 2004.
- [2] D. Fogarty, D and P. Bell, Should you outsource analytics?, *MIT Sloan Management Review*, 50(2), 41-45, Winter, 2014.
- [3] G. Gorry and M. Morton, A framework for management information systems, *Sloan Management Review*, 30(3), 49-61, 1989.
- [4] A. Hayashi, Thriving in a Big Data World, *MIT Sloan Management Review*, 50(2), 35-39, 2014.
- [5] J. Lee, *Big Data : A revolution that will transform how we live, work and think*, Book21 Publishing, 2013.
- [6] H. Mintzberg and F. Westley, Decision Making: It's not what you think, *Sloan Management Review*, 42(3), 89-93, 2001.
- [7] *New York Times*, Old Drugs In, New Ones Out, June 30, 2007.
- [8] *New York Times*, In Big Data, Shepherding Comes First, 2014. 12. 14.
- [9] K. Noh, Lecture note for the Practical Guide to Big Data at Gumi City Government, 2014.
- [10] K. Noh and N, Cho, *Management Information Systems*, Cytec Media, 2010.
- [11] P. Parise, B. Iyer, and D. Vesset, Four Strategies to Capture and Create Value from Big Data, *Ivey Business Journal*; 76 (4), 2012. Accessed through <http://iveybusinessjournal.com/topics/strategy/four-strategies-to-capture-and-create-value-from-big>



-data#.VKeMUslMVCo.

- [12] J. Park, Creative Confidence, ChungRim Publishing, 2014.
- [13] J. Seemann, Hybrid insights: Where the quantitative meets the qualitative, Rotman Magazine, 57-61, Feb 2013.

### 남 수 현(Namn, Su Hyeon)



- 1982년 2월 : 고려대학교 통계학과 (경제학사)
- 1988년 8월 : Texas Tech Univ 경영대학원 (경영정보학 석사)
- 1996년 5월 : Rutgers Univ 경영대학원 (경영정보학 박사)
- 1996년 9월 ~ 현재 : 한남대학교 경영정보학과 교수

- 관심분야 : 네트워크이론 활용, 빅데이터분석, 지식관리
- E-Mail : namn@hnu.kr

### 노 규 성(Noh, Kyoo Sung)



- 1984년 2월 : 한국외대 경영학과 (경영학사)
- 1995년 8월 : 한국외대 대학원 경영정보학과(경영정보학 박사)
- 2003년 ~ 2010년 : 中國 延邊科學技術大學 兼職教授
- 1997년 ~ 현재 : 신문대학교 경영학부 교수

- 2004년 ~ 현재 : 한국디지털정책학회 회장
- 2012년 ~ 현재 : 스마트융합학술전국연합 의장
- 관심분야 : 디지털정책&스마트융합, 디지털경제민주화, 창의기반 경영혁신, 빅데이터
- E-Mail : ksnoh@sunmoon.ac.kr