데이터 과학을 통한 전통기업 밸류업

최문경, Ph.D.
자산운용팀, 고용노동부
세종특별자치시 한누리대로 422
정부세종청사 11동 508호
cmk1280@korea.kr

강형구
Handa Partners (Founder)
한양대학교 경영대학
서울특별시 성동구 왕십리로 222
hyoungkang@hanyang.ac.kr
(Corresponding)

Extended Abstract

데이터를 활용하는 기업과 그렇지 않은 기업간 가치의 차이는 갈 수록 증가하고 있다. 특히 기업의 초미시 의사결정과 중장기전략에 있어서 인공지능 등 데이터 과학은 큰 가치를 창출할 수 있다. 초미시데이터(IoE, real-time, hyperpersonalization, etc.)를 이용한 의사결정이 중요해지면서 방대하고 복잡한 정보를 신속하게 처리할 필요가 있다. 여기서 인간은 인공지능 등 데이터 기반 기계학습의 상대가 되지 않는다. 매크로, 중장기전략에 있어서 대체로 기업인들은 기업행동이론(BTF: a behavioral theory of the firm)이나 행동경제학에서 논의되는 오류를 범한다. 데이터 기반 의사결정시스템은 규칙에 근거한 의사결정으로 조직과 개인의 오류를 진단/극복할 수 있다. 그리고 산업적/학술적인 성과/정보 등을 지속적으로 업데이트하여 의사결정에 공헌할 수 있다. 하지만 많은 기업에서는 조직이나 집단 수준의 메소(meso) 데이터와 일별, 주별, 월별 시계열 데이터를 이용하여 예측 모형을 만들고 단기 전략에 적용을 시도하고 있다. 이러한 분야는 데이터 과학이 가장 남용되는 분야 중 하나다. 업계의 전문지식이나 게임이론, 산업조직론, 심리학, 행동경제학 등 경제/경영 분야의 알려진 이론과 성과에 기반을 두고 인과관계(identification)을 고려하지 않으면 데이터 기반 전략은 성공을 거두기 어렵다. 데이터 과학이 기업의 밸류업(value up)에 크게 공헌을 할 수 있는 분야가 위험관리다. 특히 비정형 빅데이터를 이용하여 선행적인 위험관리(forward-looking risk management)의 잠재력이 크다. 고용노동부 등 일부 조직에서 이러한 시도를 하고 있는 점은 매우 바람직하다. 그러나 전반적으로 대부분의 국내 조직들은 의사결정과정에서 데이터 과학의 장점을 충분히 활용하고 있지 못하다고 판단된다. 기업들은 데이터가 가장 큰 가치를 창출할 수 있는 영역에 역량을 집중해야 한다. 그리고 쉽다는 이유로 메소 데이터 분석에 데이터 과학을 남용하지 않아야 한다. 이를 위해서는 업계의 경험과 인과관계 파악(Angrist & Pischke, 2008)을 위한 경제/경영에 대한 이론을 바탕으로 기술이 아닌 현업의 문제에 집중하는 것이 바람직하다(*Strategy, not technology, drives digital transformation*: Kane et al., 2015). 흥미롭게도 경영/경제 분야의 클래식한 이론인 아키텍처 혁신(Henderson & Clark, 1990), 기업행동이론(Cyert & March, 1963), 지식기반이론(Cohen & Levinthal, 1990; Kogut & Zander, 1992)은 데이터를 이용한 기업의 가치향상에 귀중한 시사점을 줄 수 있다. 첫째, 아키텍처 혁신은 기업이 데이터를 이용한 혁신전략 수립에 직접적인 시사점을 준다. 기업은 데이터를 혁신을 위하여 활용해야 하는데 이때 아키텍처 혁신은 데이터와 기계학습 기술 사용에 모두 적용된다. 둘째, 기업행동이론(a behavioral theory of the firm; Cyert & March, 1963)의 주요 컨셉인 나이티안 불확실성 (Knightian uncertainty)과 이해관계자 갈등(stakeholder conflict)에 주목하되 나이티안 불확실성은 사업기회 (entrepreneurship), 이해관계자 갈등은 공유가치창출(shared value creation)과 관련이 있고 이를 조합하면 데이터 활용에 관한 네가지 전략을 도출할 수 있다. 셋째, 지식기반이론(KBV: knowledge-based view)에 의하면 경쟁조직이 따라하기 힘든 암묵적 지식과 이를 흡수하고 처리하는 역량이 기업의 지속가능한 경쟁우위(sustainable competitive advantage)를 결정한다. 따라서 데이터를 지식화해야 하는데 지식기반이론은 이에 대한 방안들을 제시한다.

핵심어: 데이터, 혁신, 밸류업, 아키텍처 혁신, 기업행동이론, 지식기반이론, 흡수역량, 비정형데이터, 위험관리, 고용노동부.

서론

데이터에 기반한 기업들과 그렇지 않은 기업들의 가치(valuation) 차이는 갈수록 커지고 있다. 2020 5월 Apple, Microsoft, Alphabet, Amazon, Facebook은 S&P500의 20%를 차지하는 수준[1]이 되었고 이는 수동적 투자자들(passive investors)에게까지 위험 요인(risk factor)이 될 정도다. 코로나-19 이후 대형 기술 회사들의 비중은 더욱 커지게 되고 위기를 거치면서 최종적인 승자로 인식되고 있다.[2] 대형 기술 기업들 스스로도 코로나-19를 오히려 기회로 보고 이를 적극적으로 활용하고 있다.[3] 이처럼 데이터를 어떻게 활용하느냐에 따라 기업의 성공이 좌우되고 있다. 다른 말로 하면 비록 침체되거나 저평가된 기업이라도 데이터를 잘 활용하여 무형자산(intangible) 혹은 이를 넘어 전략적 자산(strategic resources)으로 발전시키고 이를 시장에 인식시키면 기업의 가치를 획기적으로 높일 수 있다(value up)는 의미가 된다. 예를 들어 같은 금융기관이지만 은행과 핀테크 기업의 가치 비율(valuation ratio)는 극단적인 차이가 난다. 만약에 은행이 핀테크 기업들 처럼 데이터를 활용하면 엄청난 밸류업이 발생할 것이다.

이는 밸류업(value up)을 비즈니스 모델로 삼는 private equity (PE), venture capital (VC) 등 뿐만 아니라 기업과 투자자 그리고 경제의 성장동력을 찾는 정부에게도 시사점을 준다. 실제로 연구자가 수행한 인터뷰 결과 몇몇 PE는 데이터 밸류업을 핵심 투자 모델(PE's investment model)로 삼고 이를 바탕으로 적극적인 펀드레이징도 하고 있다. 앞서가는 VC들도 데이터를 투자모형(VC's company-making model)과 가치평가에 명시적으로 도입하고 있다. 이는 투자자들을 위해서나 국가 경제를 위해서나 매우 바람직한 시도라고 판단된다. 하지만 밸류업을 과학적으로 수행하기 위해서는 데이터를 이용해서 어떻게 밸류업을 할지에 대한 구체적인 학술적인 연구나 실용적인 가이드라인이 필요하다. 연구자가 파악한 바로는 이에 대한 체계적인 연구도 실무적인 프레임워크도 존재하지 않는다. 본 연구는 이러한 중요한 연구와 실무의 공백을 학계 최초로 해결하고자 한다. 따라서 본 연구는 PE, VC, 기업, 투자자, 정부에게 직접적인 시사점을 줄 것이다.

데이터 활용과 관련된 유명한 케이스가 GAP이다. GAP은 미국에 본사를 둔 글로벌 의류 및 액세서리 리테일러다. GAP은 Old Navy, Banana Republic 등 브랜드로도 유명하다. 미국을 상징하는 브랜드지만 2000년대 이후 침체기를 겪는다. 특히 ZARA로 유명한 Inditex, 유니클로 등이 GAP을 압도하게 된다. Art Peck은 GAP이 이러한 침체기에 있을때 CEO로 취임했다. Art Peck은 취임 초반 크리에이티브 디렉터들을 대량 해고한다. 크리에이티브 디렉터 들은 패션 트렌드를 예측하고 이에 기반하여 봄 신상품, 가을 신상품 등을 위한 디자인 패션하우스의 theme일 결정하는 사람들이다. 패션업계의 꽃인 인력이다. Art Peck은 더이상 크리에이티브 디렉터나 디자이너 들의 감에 의존하지 않고 데이터를 이용하여 패션에 관한 의사결정을 하자고 주장했다. 패션산업에서 일종의 머니볼(Lewis, 2004)을 시도한 셈이다. 당연하게도 GAP은 물론 패션업계가 반발하게 된다. 패션은 창의성 생명인데 아무것도 모르는 무뢰한이 패션과 시장을 죽일거라는 비판이 많았다. 사실 이러한 회의와 비난은 스포츠의 머니볼에도 있었고 조직에서 새로운 시도를 할때 항상 직면하는 비판이다. 혁신을 하자고 하면서도 막상 시도하면 조직, 산업과 현실에 대해서 아무것도 모른다는 비판을 흔히 한다. 그렇다면 Art Peck의 혁신은 성공했을까?

2016/01/01 - 2020/06/08까지 Inditex, Gap, SPDR의 주가를 비교해보면 Gap의 데이터 전략이 확실한 성공을 거두었는지 아직 판단하기는 힘들다. Art Peck은 2019년 11월 성과부진의 책임을 지고 사퇴를 했으니 성공이라고 하긴 어려울 것 같다. 그렇다면 Art Peck의 데이터 전략이 실패한 이유는 무엇일까? 애당초 패션 트렌드의 예측을 데이터 과학으로 하는게 무리였을까? 아니면 데이터 과학 역량이 없어서 일까? 이러한 판단을 하기 위해서는 데이터 밸류업에 대한 과학적인 프레임워크가 필요하고 본 논문은 이를 해결하고자 한다.

많은 관심에도 불구하고 기업들이 데이터를 활용하는데 있어서 오해도 많은데 본 연구자들은 이 역시 과학적인 프레임워크가 없어서라고 판단한다. 첫째, 데이터가 가장 큰 가치를 창출할 수 있는 분야에서 오히려 데이터가 적어도 국내에서는 제대로 활용되지 않고 있다. IoE (internet of everything) 등에서 생성되는 초미시 데이터와 이를 활용한 의사결정의 경우 즉각적으로 데이터 과학이 도움을 줄 수 있다. 그러나 이 분야에서 몇몇 대형 기술기업 등을 제외하면 오히려 인공지능 등 데이터 과학이 잘 활용되지 않고 있다. 둘째, 데이터 구조의 근본적인 문제로 고도의 기계학습이 논리적으로 적용되기 어려운 부분에서 오히려 기계학습이 남용되고 있다. 셋째, 수학적으로 정의된 데이터 기반 알고리즘이 오히려 인간의 직관적인 의사결정에 비해서 투명하게 운영될 수 있으나 이에 대한 이해와 활용이 적다. 예를 들어

[1] https://www.economist.com/leaders/2020/05/02/big-tech-is-thriving-in-the-midst-of-the-recession
[2] https://www.economist.com/leaders/2020/04/04/big-techs-covid-19-opportunity
[3] https://www.nytimes.com/2020/06/13/technology/facebook-amazon-apple-google-microsoft-tech-pandemic-opportunity.html

인공지능의 불투명성(black box)에 대한 비판이 많지만 대다수 기업들의 내부 의사결정 프로세스에서의 불투명성에 비할 바가 아니다. 넷째, 데이터가 제대로 활용되기 위해서는 전통적인 경제, 경영에 대한 이해는 물론 기업행동이론(Cyert & March, 1963), 행동경제학적 사고가 중요하다. 특히 데이터가 충분하지 않거나 불완전한 상황에서는 경제/경영 이론이 매우 중요해진다. 그러나 단순한 데이터 마이닝으로 오버피팅(overfitting)이 만연하고 당연히 이에대해 실망스러운 결과들이 나오고 있는 것이 현실이다. 다섯째, 비정형 데이터를 이용한 비정형 리스크 관리는 기업에 큰 도움이 될 수 있으나 활용이 미미하다. 이러한 문제점을 해결하기 위해서 아키텍처 혁신, 기업행동이론, 지식기반이론에 근거한 대안을 제시한다. 이 논문은 이러한 문제점들을 체계적으로 분석하고 이에 대한 해결책을 제시하고 이를 바탕으로 어떻게 기업이 밸류업을 할 수 있을지 논의한다.

전통 기업의 데이터 활용에 관한 이슈들

초미시데이터

IoE 등에서 생산되는 실시간(real-time), 고빈도 (high frequency data), 초개인화(hyper-personalization 데이터를 편의상 초미시 데이터라고 하자. 초미시 데이터를 이용한 초미시 의사결정은 이미 기계학습 등 데이터 과학으로 상당히 대체되었다. 이를 활용하는 기업과 그렇지 못하는 기업들간 격차도 갈수록 심해지고 있다. 그렇다면 왜 초미시 데이터에서 인공지능과 같은 데이터 과학 역량이 중요한가? 그 이유는 첫째, 데이터가 충분하다는 점이다. 예를 들어 고객 행동이나 시장 상황에 대한 실시간 데이터를 수집하면 딥러닝 등 기술을 활용할 수 있게 된다. 데이터는 기계학습의 성과에 결정적인 영향을 미친다.

둘째, 데이터 기반 알고리즘은 인간과 비교도 안되게 빠른 속도로 의사결정을 할 수 있다. 경매로 대체된 온라인 광고시장은 밀리세컨드 단위에서 의사결정이 이루어진다. 경쟁자보다 더 빠르고 더 실시간으로 촘촘하게 의사결정을 내릴 수록 기업의 이익이 커진다. 여기에서 인간은 데이터 기반 알고리즘의 상대가 될 수 없다.

셋째, 정보처리 역량의 차이다. 뉴스, 웹사이트, 사회관계망(SNS) 등에서 새로운 정보가 생성되었다고 하자. 이를 이용하여 이익을 거두기 위해서는 정보를 습득하고, 해석하고, 관련된 의사결정 알고리즘을 만들어야 한다. 필요한 경우 그 실행도 알고리즘에의해 이루어진다. 데이터 과학자들은 스크레이핑(scraping)을 통하여 정보를 습득하고, 자연어처리기법(NLP: natural language processing)으로 정보를 해석하고, 이미 프로그래밍된 알고리즘으로 실시간에 의사결정을 할 수 있다. 이러한 일련의 절차를 인간이 체계적으로 수행하며 기계를 상대로 경쟁우위를 거두는 것은 어렵다. 넷째, 알고리즘의 경우 빠른 속도로 오류를 파악하고 역시 빠른 속도로 스스로를 업데이트 할 수 있다. 이 역시 인간이 따라가기 힘들다. 인간이 공부하는 속도보다 지식이 발전하는 속도가 훨씬 빠르다.

하지만 안타깝게도 일부 기술기업을 제외한 국내 대다수 기업들에게 초미시 데이터를 확보하고 활용하는 역량은 매우 부족한데 그 이유는 다음과 같다. 첫째, 초미시 데이터를 확보하는 것이 어렵다. 국내외 대형 기술회사들이나 일부 금융기관을 제외하고는 해당 데이터를 구매하기도 어렵다. 비즈니스 모델이 데이터 확보와 연관되지 않으면 이를 별도로 구해야 한다. 그러나 이는 너무 비싸거나 법적/제도적 장벽 때문에 아예 불가능한 경우가 많다. 게다가 이미 성공을 거두고 데이터를 흡수하는 기업들이 데이터를 내놓을리도 없다. 둘째, 설령 데이터를 확보하더라도 이를 분석할 수 있는 고급 인력을 구하기도 어렵다. 이코노미스트 기사에 의하면 역량있는 인공지능 관련 인력을 확보할 수 있는 자원을 가진 조직은 대형기술회사나 헤지펀드 정도라고 보고하기도 했다.[4] 셋째, 데이터와 인력을 구해도 이를 보관하고 처리하는 비용이 너무 많이 든다. 다음은 역시 이코노미스트 기사에서 발췌했다: “Jerome Pesenti, Facebook’s head of AI, says that one round of training for the biggest models can cost “millions of dollars” in electricity consumption.[5]”

엄청난 가치에도 불구하고 초미시 데이터에 관한 이상의 문제들을 개별 기업에서 특히 중소기업이나 스타트업 수준에서 해결하기는 매우 어렵다. 결국 현업의 전문가들의 견해를 바탕으로 어떤 초미시 데이터를 활용할 것인지에 관하여 선택과 집중을 해야 한다. 초미시데이터는 수집, 보관, 관리는 비용이 많이 들지만 분석하기는 오히려 쉬운 경향이 있다. 잘 정리된 대량의 데이터 때문에 적당한 기계학습 모형을 선택해도 당장 활용할 수 있고 눈에 띄는 결과를 쉽게 얻을 수 있기 때문이다. 이처럼 초미시 데이터 관련 국내 기업들의 활용에 어려움이 있고, 데이터 과학 역량도 부족한데, 조직 수준의 메소(meso)

[4] https://www.economist.com/technology-quarterly/2020/06/11/businesses-are-finding-ai-hard-to-adopt
[5] https://www.economist.com/technology-quarterly/2020/06/11/the-cost-of-training-machines-is-becoming-a-problem

데이터에서는 정반대의 상황이다. 이하 섹션에서 상술한다.

메소(meso) 데이터와 단기 전략

메소 데이터는 그룹이나 조직 수준의 행동 데이터다. 주가 등 극히 일부를 제외하고는 집계를 해야하는 특성상 초미시 데이터 만큼 실시간으로 생성하기는 매우 비싸다. 흔히 볼 수 있는 일별, 주별, 월별로 생성되는 성과나 행동 데이터를 생각하면 된다. 기업은 메소데이터를 이용하여 주로 단기전략 수립에 활용하려고 시도한다. 그런데 적어도 국내에서는 메소데이터를 이용한 데이터 과학이 "데이터 마이닝"의 형태를 띄면서 가장 오용되고 있는 분야로 보인다. 초미시 데이터 관련된 산업의 실정과 정반대 상황이다.

이유는 간단하다. 관련 데이터를 상대적으로 쉽게 구하고 분석할 수 있기 때문이다. 기초통계학을 배운 누구라도 간단한 알고리즘을 만들 수 있다. 기계학습을 학교에서 배우고 기업의 성과데이터를 구해서 쉽게 간단한 분석을 할 수 있다. 실제로 수업에서 숙제로 관련 작업을 출제하기도 한다. 언뜻 쉬워보이니 관련 회사들도 많이 등장했다. 수많은 데이터 컨설팅 회사들이 다루는 데이터들도 이러한 단기 메소 데이터들인 경우가 많다. 그런데 여기에는 심각한 문제가 존재한다.

어떤 기업 전략의 성과를 예측하기 위한 데이터가 있고, 그 데이터의 크기는 N = T*K로 표현하자. N이 관측치의 숫자(# of observations)를 나타내는데 N이 크면 빅데이터라고 한다. N은 T와 K의 곱인데 K는 횡단면 변수의 숫자, T는 데이터의 시간별 관측치 즉 시계열의 길이(length)다. 학자들은 T에 관심을 갖는 경우가 많다. 그러나 업계에서 빅데이터라고 하면 K가 큰 경우가 많다. 예를 들어 기업의 회계 데이터는 물론 인터넷에서 스크레이핑으로 수집한 텍스트, 사진 등 비정형데이터를 포함하면 기업에 관한 변수의 숫자인 K는 빠른 속도로 늘어난다. 하지만 이를 이용하여 전략의 성과를 예측하는 것은 전혀 다른 문제다. 전략의 성과를 예측하기 위한 정보가 많을수록 문제를 해결하기는 커녕 심화시킬 수 있다. 차원의 저주(curse of dimensionality) 때문이다. 예를 들어 몇 십만개의 동전을 던지다 보면 유행이나 패션 트렌드를 계속 맞추는 동전은 존재할 수 밖에 없다.

시장이나, 기업, 조직이 아닌 개인의 경우는 상대적으로 문제해결이 쉬운 편이다. 왜냐하면 비슷한 개인의 데이터를 많이 모아서 해결할 수 있기 때문이다. 그러나 여전히 민감한 개인의 데이터를 얼마나 수집할 수 있느냐의 문제가 남는다.

결국 이론을 활용할 수 밖에 없다. 즉 수십년간 학술적으로 정립된 성과를 활용해야 한다. 이는 데이터의 한계를 이론과 직관으로 극복하려는 노력이다. 기업이 데이터 과학을 사용하는 이유는 사실 인과관계에 관심이 있기 때문인 경우가 많다. A라는 전략을 쓰면 B에는 무슨 일이 일어날까? 같은 문제들이다. 인과관계(학자들은 identification problem이라고 부르는 문제)를 파악하기 위해서는 특별한 데이터 분석 기술이 필요하다. 이는 단순히 데이터가 많다고 해결될 수 있는 문제가 아니다. 아무리 데이터가 많아도 적절한 기술을 사용하지 않으면 인과관계가 아닌 상관관계를 파악할 수 있을 뿐이다. 그리고 인과관계를 파악하기 위하여 어떤 기술을 사용할지는 필연적으로 이론, 예를 들어 업에 대한 직관이나 검증된 학술적 성과에 대한 지식이 필요하다. 이론에 대한 고려없이 무조건 딥러닝을 하는 컨설팅 업체나 데이터 과학자들을 경계할 필요가 있다. 이들의 행위는 데이터 마이닝에 불과하며 대체로 해로운(mostly harmful) 방법이다. 이와 반대로 대체로 무해한(mostly harmless) 분석을 위해서는 특별한 기술과 함께 가설을 수립할 수 있는 업계의 전문성이 필요하다(identification; Angrist & Pischke, 2008).

인간의 행동에 관심이 있다면 행동경제학이나 심리학 등, 회사라면 미시경제학, 게임이론, 전략 등, 시장의 변화라면 산업조직론 등에 관한 이해가 중요할 것이다. 기업은 메소데이터를 이용한 인과관계에 관심이 있다면 단순 통계전문가가 아니라 데이터 분석의 경험이 있는 경제/경영학자, 또는 업계 전문가에게 의뢰를 해야 할 것이다. 그리고 인과관계를 어떤 기술로 어떻게 파악할(identify) 것인지 집요하게 질문해야 한다.

거시데이터와 중장기 전략

기업의 중장기전략에는 시나리오 분석이 들어갈 수 밖에 없고 이를 위해서는 트렌드와 같은 거시데이터를 활용하게 된다. 그런데 메소 데이터에서 딥러닝 등을 적용하는 것이 데이터 문제로 힘들다면 논리적으로 거시 데이터는 더 어렵고 이는 사실이다. 많은 기업에서 관심을 가지고 있는 유가(oil price)의 경우 일별 데이터를 20년을 모아도 5,000여개의 관측치에 불과하다. 그렇다면 수만개의 계수를 추정해야 하는 딥러닝을 사용하는 것이 불가능하다. 물론 초단위 또는 실시간 거래 데이터를 활용하면 가능하지만 이런

데이터를 분석하여 기업의 전략, 특히 중장기 전략을 수립하는 경우는 별로 없다.

그러나 인공지능과 같은 데이터 기반 알고리즘은 중장기 전략수립에 의외로 도움을 줄 수 있다. 행동경제학에 의하면 중장기 의사결정에서 사람들은 비합리적인 편향을 보이는 경향이 있다. 특히 먼 미래의 일일수록 의사결정을 제대로 하지 못한다(예: planning fallacy, hyperbolic discounting, optimism bias). 따라서 데이터에 기반한 알고리즘은 대부분의 사람들에게 좋은 행동경제학적 처방이 될 수 있다. 예를 들어 학술적으로 검증된 이론에 기반한 의사결정 시스템을 활용할 수 있다. 특히 인간은 감이나 그날 그날의 단편적인 정보만 보고 주먹구구식으로 의사결정을 하는 경향(availability heuristics)이 있는데 이는 중장기 플래닝에 치명적인 악영향을 준다.

요약해보자. 데이터 기반 알고리즘에 기업행동이론과 행동경제학의 성과를 반영하여 조직과 인간의 실수를 지적할 수 있게 하고 귀찮은 최적화 문제를 풀게할 수 있다. 여기에 최신의 학술적 성과와 업계의 주요 정보를 지속적으로 업데이트하여 자동으로 분류하여 보여줄 수 있게 한다. 이를 바탕으로 인간과 기계는 더 나은 의사결정을 할 수 있다. 중장기 플래닝은 기계도 어렵지만 인간에게는 더 어렵다. 따라서 기계가 큰 도움이 된다.

비정형 데이터와 위험관리

데이터가 기업에 가장 큰 가치를 창출할 수 있는 분야 중의 하나가 위험관리다. 위험관리는 사실 의사결정 프로세스에서 가장 중요하다고 볼 수도 있다. 많은 기업들이 위험관리를 하고 있다고 하지만 리스크를 예측하고 이를 경보하는 조기경보시스템(early warning system)을 활용하는 경우는 드물다. 미중 패권경쟁, 기후변화, 양극화 등 불확실성에 대해서 어떻게 대응을 하고 있는가? 특히 정량화 되기 어려운 불확실성, 즉 질적불확실성 혹은 나이티안 불확실성과 관련해서는 업계가 익숙하지 않다. (질적) 불확실성 관리라는 말은 존재하지도 않는다. 그렇다면 어떻게 질적불확실성에 관한 조기경보시스템을 만들 수 있을까?

직관적으로 생각해보자. 미래의 위험과 불확실성을 파악하려면 어떻게 해야 할까? 일단 업계의 주요 자료들은 물론 Financial Times (FT)같은 신문이나 Economist 같은 잡지를 열심히 보고 업계의 주요 커뮤니티를 관찰해야 한다. 그러나 현실적으로 경영진이 FT하나라도 제대로 볼 시간이 없다. FT도 제대로 못보는 휴먼에게 업계의 주요 동향은 물론 전 세계 주요 미디어의 동향을 빠른 속도로 파악하고 그 톤을 분석해주는 인공지능의 존재는 큰 도움이 될 수 있다.

특히 기업이 확보 가능한 비정형 데이터를 체계적으로 관리할 필요가 있다. 코로나-19 이후는 비정형리스크, 질적불확실성의 시대다. 이제까지 대부분의 회사에서는 질적불확실성과 비정형리스크에 대한 체계적인 위험관리가 거의 없었다고 해도 과언이 아니다. 하지만 코로나-19는 경제 전반에 걸쳐서 질적 불확실성이 얼마나 큰 영향을 미치는지를 여실히 보여주었다. 코로나-19는 업계에서 흔히 다루는 리스크와는 다른 나이티안 불확실성(Knightian uncertainty)이다. 나이티안 불확실성은 unknown unknowns으로 인식되기도 한다. 뉴욕타임스에서는 칼럼니스트인 Thomas L. Friedman이 코로나-19 사태를 unknown unknowns로 표현하기도 했다.[6]

Unknown unknowns와 같은 불확실성을 경제학에서는 질적불확실성, 혹은 나이티안 불확실성 (Knightian uncertainty;Keynes, 1921; Knight, 1921)이라고 부른다. 나이티안 불확실성은 Keynes (1921)에도 잘 설명되어 있듯이 의사결정자가 의사결정과정에서 발생하는 위험을 정량적으로 묘사할 수 없을때 발생한다. 나이티안 불확실성은 어떤 이슈가 매우 모호하거나 복잡하여 사업에 관한 결과를 예측하기는 커녕 가능한 상태의 확률적 분포를 묘사하기 조차 힘들때 발생한다. 이하 Keynes (1937)의 설명이다.

> By "uncertain" knowledge, let me explain, I do not mean merely to distinguish what is known for certain from what is only probable. The game of roulette is not subject, in this sense, to uncertainty... Or, again, the expectation of life is only slightly uncertain. Even the weather is only moderately uncertain. The sense in which I am using the term is that in which the prospect of a European war is uncertain, or the price of copper and the rate of interest twenty years hence, or the obsolescence of a new invention, or the position of private wealth owners in the social system in 1970. About these matters there is no scientific basis on which to form any calculable probability whatever. We simply do not know. Nevertheless, the necessity for

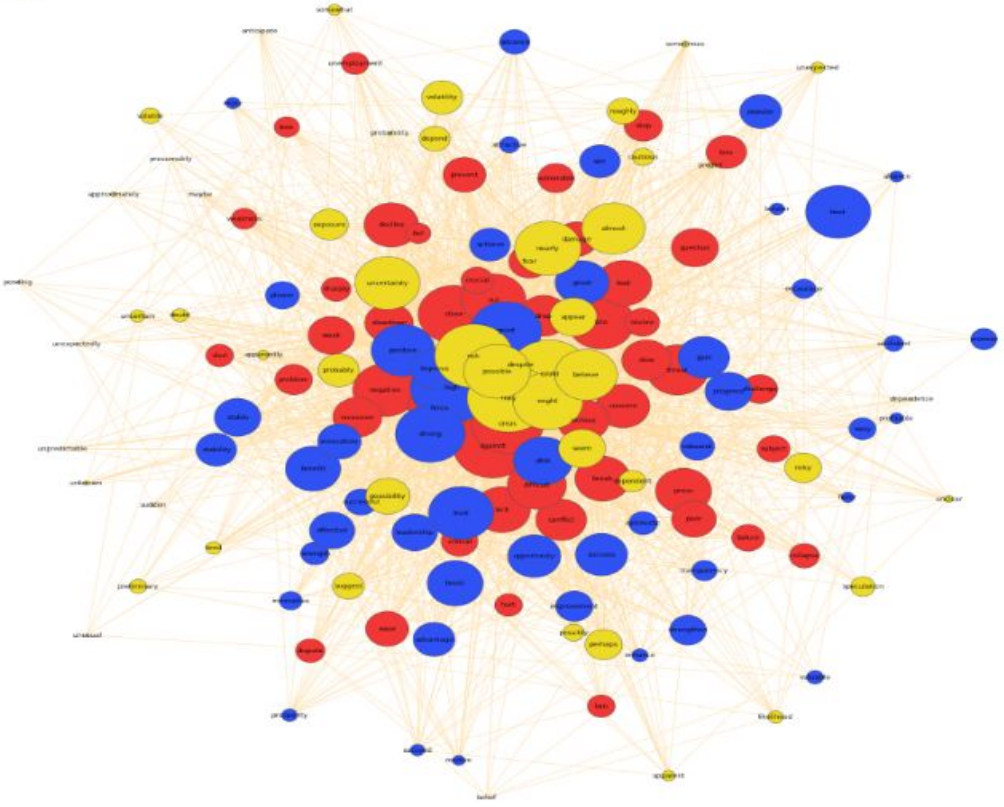
[6] https://www.nytimes.com/2020/03/17/opinion/coronavirus-trends.html

> action and for decision compels us as practical men to do our best to overlook this awkward fact and to behave exactly as we should if we had behind us a good Benthamite calculation of a series of prospective advantages and disadvantages, each multiplied by its appropriate probability, waiting to be summed. (Keynes, 1937; 213–214)

논리적으로 비정형위험, 질적불확실성을 파악하기 위해서는 비정형 데이터가 필요할 수 밖에 없다. 질적불확실성은 정의상 측정이 어려운, 확률분포를 특정하기 어려운, 비정량적인 불확실성이다. 따라서 전통적인 정량적인 데이터로 예측과 관리가 어려울 수 밖에 없다. 여기서 고용노동부의 사례는 매우 유용한 시사점을 준다. 고용노동부는 비정형데이터를 이용하여 위험을 선제적으로 파악하려는 시스템을 1년이 넘게 운영해오고 있다. 고용노동부는 해당 시스템을 이용하여 코로나-19 위험을 2개월 전에 선제적으로 파악한 바가 있다. 아래는 해당 시스템에서 어떻게 코로나-19를 경보했는지 예를 보여준다.

2020. 01월 글로벌 미디어들은 한국의 경제금융상황에 대해 긍정적인 평가와 부정적인 평가가 혼재되어있다. 12월 말 대비 긍정의 비중이 약 2.00 %p 감소하고 부정의 비중이 약 1.96 %p 증가하여 상대적으로 부정의 비중이 크게 증가하였다. 불확실성의 경우 약 0.04%p 증가하였다.

연구진의 분석에 의하면 아래 semantic network에서 보이는 부정과 불확실성의 증가는 코로나바이러스와 연관되어 있을 가능성이 크다. 만약 아래 semantic network의 패턴이 코로나바이러스와 관련이 되어 있는게 맞다면 코로나바이러스가 금융 및 경제 관련 변수 들에서 부정과 불확실성을 증가시키는 방향으로 영향을 끼칠 가능성을 상정한다.



위와 같은 잠재력에도 불구하고 A기관 이외에 비정형 데이터를 이용하여 위험관리를 하고 있는 사례는 극히 드문 것으로 보인다. 결론적으로 국내 기업들은 가장 큰 가치를 창출할 수 있는 데이터 활용에는 상대적으로 약하고 이에 반하여 밸류업 효과를 보기 어려운 부분에 데이터 과학을 남용하고 있는 것으로 판단된다. 다음 섹션에서는 기업이 어떻게 대응방안을 수립할 수 있을지 연구자의 제안을 설명한다.

기업의 데이터 기반 밸류업 전략 수립을 위한 프레임워크

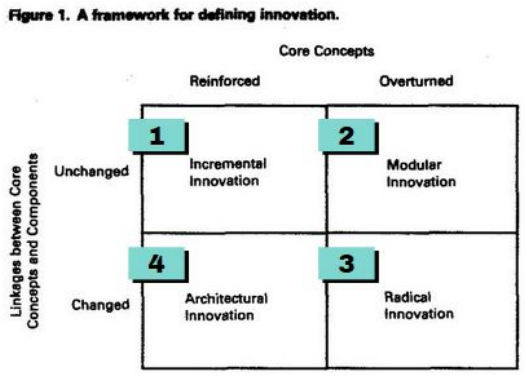
#	Considerations	Approaches	Key literature

1	Innovation strategy	Architectural innovation	Henderson & Clark (1990)
2	Organizational change	A behavioral theory of the firm (BTF)	Cyert & March (1963)
3	Sustainable competitive advantage	Knowledge-based view of the firm; absorptive capacity	Cohen & Levinthal (1990), Kogut & Zander (1992)

Architectural Innovation

데이터 과학 백그라운드가 부족한 현업 전문가들은 비정형 데이터, 기계학습 등 빠르게 발전하는 기술혁신에 이해하고 적용하는데 어려움을 겪고 있다(예: the innovator's dilemma, Christensen, 2013). 데이터 과학자들의 경우 현업의 가치제안(value proposition)을 이해하지 못하고 각종 시행 착오를 겪고 있다. 경영자들은 이미 인공지능에 회의를 느끼고 있다.[7] 특히 시장전략, 투자의사결정 등 경제/경영 이론/경험과 데이터 과학을 결합이 중요한 부분에서 그 문제는 더 심각하다(eg. causality vs correlation). 이 문제는 데이터 과학의 전문가도 현업의 전문가도 풀 수가 없다. 산업 경험, 이론과 실무를 겸비한 데이터 과학자들만이 문제를 해결할 수 있다. 그러나 이런 인력은 대형 기술회사나 헤지펀드 정도나 채용할 수 있을 것이다. 그래서 4차산업혁명 솔루션들을 기업에서 아이폰처럼 사용하도록 할 수 있는 UI/UX를 제공하는 기업이나 서비스가 크게 각광을 받을 것이다. 그러나 당장 그러한 서비스를 이용할 수 없거나 너무 비싼 상황이면 기업은 어떻게 해야 하는가? 이는 혁신가의 딜레마에서 설명하듯 대기업들에게도 중요한 문제일 수도 있다.

학자들은 혁신을 다음 그림과 같이 네가지로 분류하기도 한다. 경영학자들의 견해에 의하면 혁신을 추구하는 전략은 두가지 축이 있는데 이는 컨셉과 컨셉간 관계다. 그래서 한 축을 따라서 컨셉을 강화하거나 전복시키는 것이고, 다른 축을 따라 컨셉들간의 관계를 유지하거나 변화시킬 수 있다. 결국 네가지 유형의 혁신이 도출 되는데 이들이 컨셉 강화/관계 유지의 점진적 혁신, 컨셉 전복/관계 유지의 모듈 혁신, 컨셉 강화/관계 변화의 아키텍처 혁신, 컨셉 전복/관계 변화의 급진적 혁신이다.



Henderson, R. M., & Clark, K. B. (1990). Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms. *Administrative science quarterly*, 9-30.

1. Conventional meaning; conventional connection; better understanding through curation (Emerging economies)

2. Relationship unchanged; using updated and enhanced information; or different interpretation of existing information (Startups)

3. New architecture; new concepts; new system of knowledge (Tech giants)

4. Constant core meaning; different relation between meaning; reconfigure existing system (large Korean firms)

급진적 혁신의 경우는 일부의 대규모 기술회사나 대형 헤지펀드를 제외하면 대부분의 전통 기업들에게 지나치게 비싸거나 시간이 걸릴 수 밖에 없다. 점진적 혁신의 경우는 이미 대다수의 기업들이 시도하고 있지만 혁신가의 딜레마에 노출될 가능성이 높다. 혁신가의 딜레마는 기업이 타겟 시장이 원하는 혁신을 열심히 추구하다가 오히려 이 때문에 실패하는 상황을 경고하고 있다. 그렇다면 기업이 모듈혁신이나 아키텍처혁신 중 선택을 해야 하는데 아키텍처 혁신이 훨씬 효과적이다. 모듈혁신을 추구하기 위해서는

[7] https://www.economist.com/technology-quarterly/2020/06/11/businesses-are-finding-ai-hard-to-adopt

컨셉 강화, 즉 전복적인 데이터를 축적하거나 새로운 기계학습 기술을 개발에 초점을 두는 전략인데 이는 모두 어렵고 실용적이지 않다. 전복적인 데이터는 데이터 자체가 비즈니스 모델인 회사에 적합하고, 대다수 기업들은 새로운 기계학습 기술을 개발할 필요도 없다. 기계학습에 대한 새로운 논문이 지속적으로 나오고 있고, 논문에 사용된 코드 들은 Github, Gitlab 등에 포스팅하는 것이 최근 관행이다.

그렇다면 아키텍처 혁신이 대안이 된다. 새로운 기술을 개발하는 것보다는 Github 등에 있는 코드를 다듬어서(customizing) 가져다 쓰는 것이 중요하고 이는 경영학과 학부생도 수업을 들으면서 하고 있다. 결국 산업의 특수한 연구주제(research question)를 파악하는 것이 중요하고 이를 해결하기 위해서 Github의 소스코드를 커스터마이징 후 결합하고 기존 데이터를 향상시키고 조합하는 것이 현실적인 전략이다. 새롭고 혁신적인 데이터를 구하려고 노력하는 대신에 기존 데이터를 어떻게 조합하여 잘 활용할 수 있을지 먼저 연구하는 것이 중요하다. 이것이 아키텍처 혁신인데 아키텍처 혁신 마저도 쉬운 것이 아닌데 여기에 가장 큰 도전은 기술이 아닌 조직과 전략이다. 대부분의 기업은 조직과 관행의 문제로 이미 존재하는 데이터를 제대로 활용하지 못하고 있는데 이는 후술한다.

결국 아키텍처 혁신을 위해서 기업이 필요한 인력은 명확하다. 비싼 인공지능 전문가가 필요한 것이 아니다. 산업을 이해하고, 인과관계(identification problem; Angrist & Pischke, 2008)에 분석에 능숙한 연구자를 실무자들과 같이 작업하게 하여야 하는 것이다. 경험많고 노련한 현업 에이스의 도움을 받을 수 있다면, 이미 개발된 기계학습 툴을 다운로드 받아서 잘 쓸 수 있는 학부생 수준 역량으로도 큰 효과를 거둘 수 있다. 이를 위해서는 조직 자체도 아키텍처 혁신에 맞도록 구축되어야 하는데 이에 관해서는 기업행동이론(BTF: a behavioral theory of the firm)이 귀중한 시사점을 제공한다.

기업행동이론(BTF: a behavioral theory of the firm)

기업의 데이터 전략에 관한 양상은 매우 다양한데 이를 이해할 수 있는 관점이나 프레임워크가 학술적으로 부족한 것이 사실이다. 다양성의 원천과 패턴을 이해하지 못하면 이를 일반화할 수 없고, 그렇다면 데이터 전략에 대한 제안이나 전략을 개발하는 것도 어려워진다.

본 연구는 기존 이론을 통합하여 관점을 제시하고자 한다. 이를 위해서 기업행동이론(이하 BTF: a behavioral theory of the firm; Cyert & March, 1963)을 적용한다. BTF에는 두가지 중요한 개념이 존재하는데 이들은 나이티안 불확실성(Knightian uncertainty)과 이해관계자 갈등(stakeholder conflict, controversy)이다.[8] 우리의 판단으로는 사회혁신에서 인공지능의 활용을 위해서는 나이티안 불확실성이 높은 상황에서는 창업가정신(entrepreneurship)이 중요하고(Knight, 1921) 이해관계자 갈등이 높은 상황에서는 공유가치(shared value) 창출이 중요하다(Porter & Kramer, 2011). 이를 확장하면 다음과 같다.

Organizational strategy for value-up on data science

	Knightian Uncertainty Low	Knightian Uncertainty High
Stakeholder Conflict Low	Approach (ABM: agent-based modeling): Strategically implement ABM Example: digital twin to enhance existing projects	Approach (Entrepreneurship): Find mutually beneficial innovation and business opportunities to address uncertain issues Example: AI and social values
Stakeholder Conflict High	Approach (Stakeholder social capital): Strategically yield to community on some issues, while building social capital in the process	Approach (Institution-based view; social contract): Exercise nonmarket leadership in designing social contract Example: human-machine teaming vs

[8] BTF(a behavioral theory of the firm, 기업행동이론)은 카네기 학파의 대표적인 연구이면서도 경영학 분야의 클래식 중 하나다. 경영학과 조직이론에 대한 귀중한 시사점으로 가득차있는데 여기에서 모든 것을 다룰 수는 없고 BTF의 핵심개념인 나이티안 불확실성, Knightian uncertainty (Keynes, 1921; Knight, 1921)와 이해관계자 갈등에 초점을 맞춘다.

	Example: data ownership	technological unemployment

나이티안 불확실성의 높음 또는 낮음, 그리고 이해관계자 갈등의 높고 낮음에 따라 데이터 전략에 관하여 네 가지 접근방법을 도출할 수 있는데 이들은 ABM (agent-based modelling), 창업/신규사업(entrepreneurship), 이해관계자 자본(stakeholder social capital), 제도기반 전략(institution-based view; Peng et al., 2009), 더 넓게는 사회적계약(social contract)이다. 이 관점은 불확실성과 이해관계자 관점에서 철저히 문제에 중심을 두고 관련 기술을 동원하는 것을 강조한다. 그리고 이는 디지털 혁신은 기술이 아니라 전략이 결정한다는 관점 -- "Strategy, not technology, drives digital transformation" (Kane et al., 2015) --과 부합한다.

Low stakeholder conflict and low Knightian uncertainty

ABM (agent-based model)이란 정량적 모델링 기술인데 개별 주체와 그들간의 상호작용을 모델링하거나 추정한 후 다양한 상황을 시뮬레이션하여 그 영향을 파악하는 방법이다(Railsback & Grimm, 2019). 나이티안 불확실성과 이해관계자 갈등이 모두 낮다면 퀀트 모델링의 장점은 부각되고 단점은 덜 문제가 된다. 특히 ABM이 데이터 과학을 기업에 적용하는 좋은 접근법으로 보인다. 나이티안 불확실성이 낮으면 비록 위험이 높더라도 가능한 상태(state)나 시나리오를 파악하고 확률적으로 묘사 할 수 있다. 이에 반하여 불확실성이 높으면 정량적인 정보를 입력해야 하는 데이터 기반 의사결정 시스템의 특성을 고려할때 시스템을하는 것이 까다로와진다. 이를 위해서는 정성적인 정보를 파악할 수 있는 비정형데이터(alternative data)등 구하기 어렵고, 구할 수 있더라도 비싸고, 처리하는데도 무거운 컴퓨팅 파워가 수반되는 자료를 이용해야 한다. 이에 반하여 위험(risk)은 정량적이므로 그 높고 낮음에 상관없이 이를 해석하여 관련 데이터를 모으고 수집하고 분석하는 과정이 훨씬 쉽다.

한편 이해관계자 갈등이 낮으면 이해관계자들의 전략적 행동을 모델링할 필요가 낮아진다. 첫째, 전략적 행동을 모델링하는 것은 단순히 데이터가 많다고 되는 것은 아니라는 점을 고려할 필요가 있다. 예를 들어 페이스북 사용자들이 자신이 남들에게 보여주고 싶은 정보(예: 행복한 사진)만 전략적으로 보여준다면 페이스북의 데이터를 많이 모으면 모을 수록 모형의 편향은 심해질 것이다(Stephens-Davidowitz & Pabon, 2017). 둘째, 갈등이 높으면 이해관계자를 모델링할 필요는 높아지는데 반하여 이해관계자들의 협조를 구해서 데이터를 구하는 것은 어려워진다. 셋째, 갈등이 높으면 모델링 자체도 어려워진다. 딥러닝 등 대표적인 데이터 기술은 최적화 알고리즘이다. 그런데 이해관계자 갈등이 높다는 것은 목표 불일치(goal incongruence)를 의미하고 이렇게 되면 무엇을 최적화해야 하는지도 모호해진다.

기업에서 인공지능을 사용하는 이유는 역시 비용을 줄이면서도 과학적으로 현황파악과 의사결정을 하기 위해서일 가능성이 높다. 관심이 있는 이슈의 현황을 편리하게 파악해야 하고, 해당 정보를 바탕으로 사업의 목표를 달성하기 위하여 어떤 옵션을 고려해야 하는지 조사하고, 각 옵션별 예상 성과를 예측해야 한다. 이를 위한 가장 좋은 방법은 사업과 관련있는 객체와 환경을 모델링하는 것이다. 모형이 완성되고 실시간으로 데이터를 수집하며 모형을 업데이트해 나갈 수 있다. 이렇게 되면 모형만 봐도 현황을 파악할 수 있다. 의사결정 과정에서는 사업에 관한 다양한 옵션을 모형에 반영하여 어떤 결과가 나오는지 보면 된다. 즉 모형을 통해서 사회혁신 사업을 어떻게 추진할 것인지에 관하여 시뮬레이션을 통해서 결정하는 것이다. 이러한 방법은 ABM이라고 부를 수 있는데 ABM은 사회학, 경영학, 네트워크 이론 등에서 활발히 활용되고 있다(Macy & Willer, 2002; Bergenti et al., 2013; Harrison et al., 2007; Brandon et al., 2020). ABM이 완성되면 이를 조직학습의 수단으로 삼아 ABM 포맷에 맞게 데이터를 수집하고 분석하는 조직관행도 만들 수 있다. 이를 통해 기업의 경쟁력도 향상된다. ABM은 인공지능과 개념적으로도 실무적으로도 밀접한 관계가 있다(Jennings, 2000; Klügl & Bazzan, 2012). 이상의 논의를 요약하면 불확실성과 이해관계자 갈등이 모두 낮은 상황이라면 기업이 ABM 접근방법을 사용하는 것이 효과적이라는 명제를 도출할 수 있다.

Low stakeholder conflict and high Knightian uncertainty

이해관계자 갈등이 낮다면 논리적으로 공동가치(shared value)를 만들 여지가 존재한다는 뜻이다. 그렇다면 기업 들도 조직의 목적을 달성하기 위하여 공동가치 창출 가능성(Porter & Kramer, 2011; Crane et al., 2014)을 이용할 수 있다는 의미다. 한편 불확실성 높으면 인공지능 등 혁신기술을 사용하여 기회를

만들거나 찾아낼 가능성도 커진다. 불확실성이 없으면 기회도 없기 때문이다. “High risk, high return”과 비슷하게 생각할 수 있다. 따라서 기업들은 당연히 혁신기술을 활용한 새로운 기회를 포착하도록 노력해야 한다. 이를 종합하면 불확실성은 높고 이해관계자 갈등이 높은 상황에서는 ‘공동가치로부터 혁신을 창출해내는’ 창업가 정신(entrepreneurship)이 데이터를 활용하는 적절한 전략이 된다. 이는 창업가의 정의와도 부합한다. 창업가는 보통의 사람들과 비교하여 기꺼이 ‘정성적(qualitative)’ 불확실성을 받아들이고 그 과정에서 새로운 기회를 찾아내는 사람이다(Knight, 1921).

High stakeholder conflict and low Knightian uncertainty

불확실성은 낮지만 갈등이 높은 상황에서는 이해관계자 자본이 중요하다. 기업은 사회적 자본(social capital) 특히 이해관계자 자본(stakeholder capital)을 축적해야 한다. 그 이후 축적된 이해관계자 자본을 바탕으로 사회혁신 사업과 인공지능을 결합하는 것이 적합한 접근법이다. 여기서 이해관계자 자본은 이해관계자들간에 형성된 사회적 자본이라고 보면 된다(Dorobantu et al., 2017, 2013; Henisz et al., 2014).

불확실성이 낮은 상황에서 이해관계자 갈등은 매우 뚜렷하게 보이게 된다. 따라서 기업들은 이해관계자들 요구에 일정부분 응할 수 밖에 없다. 하지만 무조건적인 헛된 양보가 아닌 양보과정에서 이해관계자 자본을 축적하는 기회를 찾아야 한다. 이해관계자 자본도 전략적으로 구성해야 하는데 전략적으로 미덕(virtues)을 선택하고 이를 바탕으로 사회혁신 조직의 특성(character)이나 평판을 개발(develop)해야 한다.

이 과정에서 낮은 불확실성이 도움이 되는데 그 이유는 기업이 이해관계자 갈등의 주요 논점과 그 원천을 파악할 수 있고 이를 바탕으로 이슈 리스트를 만들 수 있기 때문이다(Kang et al., 2018). 그리고 조직과 이해관계자들은 리스트의 이슈들을 함께 해결해 나가는데 이 과정에서 이해관계자 자본을 축적하고 민감한 데이터도 활용할 수 있다. 이는 기술보다 전략이나 문제에 초점을 두어 오히려 데이터의 잠재력을 더 높일 수 있다(Kane et al., 2015).

이해관계자와의 관계에서 형성되고 최적화된 이해관계자 자본은 향후 기업 경쟁력에 큰 도움이 된다. 왜냐하면 축적된 이해관계자 자본과 미덕은 전략문헌에서 논한 바와 같이 조직의 핵심역량(core competence)이나 자산(strategic resource)이 되기 때문이다(Blyler & Coff, 2003; Coff, 1999). 그리고 이해관계자 자본을 이루는 미덕의 경우 위험관리 역할도 한다(Godfrey, 2005; Godfrey et al., 2009; Koh et al., 2014; Minor & Morgan, 2011; Shiu & Yang, 2017). 사업이 실패하고 이해관계자들에게 손해를 끼치는 상황이 발생하더라도 이해관계자들이 조직의 긍정적인 특성(character)을 감안하여 조직에 관용을 베풀 가능성이 커지기 때문이다.

High stakeholder conflict and high Knightian uncertainty

불확실성과 이해관계자 갈등이 모두 높다면 제도적/사회계약 접근법이 효과적일 수 있다. 애당초 제도나 사회계약론은 불확실성과 갈등이 높은 상황에서 사회 이슈들을 해결하고 특정 행위, 조직 혹은 개인의 정당성을 설명하기 위하여 등장했다. 예를 들어 ‘bellum omnium contra omnes’ (the war of all against all)와 같은 심각한 이해관계자 갈등 상황에서 국가라는 사회계약이 탄생(Hobbes, 1651)하고, 무지의 장막(veil of ignorance)과 같이 매우 불확실성이 큰 상황에서 최소수혜자(the least advantaged)의 편익(benefit)을 주는 사회계약을 도출(Rawls, 1971)하는 방식이다. 사회계약론의 역사를 보면 더욱 명확하다. 사회계약론은 17세기부터 19세기초까지가 Hobbes, Locke, Rouseau 등에 의하여 전성기인데 이 기간 동안 유럽에서 혁명과 같은 사회적 갈등과 혼란(e.g. The Dual Revolutions (Hobsbawm, 1962))을 설명하고 해결하기 위한 지배적인 이론으로 등장했다. 미국도 비슷한데 미국독립전쟁 이후 사회계약론은 미국 정부 수립에 중요한 교의(doctrine)로 대두되고 이는 미국 독립선언에서도 잘 드러난다.

제도중의와 사회계약론은 이해관계자 사이에서 조직이 어떻게 생성되고 정당화 되는지 연구한다. 따라서 불확실성과 이해관계자 갈등이 큰 상황에서 기업에 적합한 접근법이다. 사회계약론은 권위(authority), 주권(sovereignty)를 강조하는데 이를 본 연구의 맥략에 적용하면 비시장영역에서의 리더쉽(nonmarket leadership)이라고 볼 수 있다. 따라서 기업이 제도적인 접근법을 적용할 때 이를 달성할수 있는 비시장역량이 중요할 것이다. 특히 정치적인 권위와 제도적인 발전이 따라오기 힘든 빠르게 변하는 혁신

분야에서 기업은 재빠르게 이해관계자의 이해와 동의를 구하면서 제도를 만들어내고 권위의 공백, 자연상태(state of nature)보다 더 나은 상태를 달성할 수 있도록 하며 비시장 리더십을 달성해야 한다.

지식기반이론

지식기반이론(Cohen & Levinthal, 1990; Kogut & Zander, 1992)에 의하면 경쟁기업이 따라하기 어렵나 쉽게 대체되지 않는 암묵적 지식(tacit knowledge)과 이를 활용하는 역량에 의하여 기업의 지속가능한 경쟁우위(sustainable competitive advantage)가 결정된다. 데이터 과학이 그 자체가 목적이 아니라면 논리적으로 데이터는 지식이 되어야 하고 이를 바탕으로 기업이 경쟁우위를 가질 수 있어야 할 것이고 그래야 밸류업이 될 것이다.

데이터를 경쟁우위의 원천 지식으로 바꾸는 방법과 관련하여 지식기반이론은 흡수역량(Cohen & Levinthal, 1990)을 강조한다. 흡수역량(absorptive capacity)이란 기업이 정보를 습득하여 익히고 이익을 창출하기 위해 적용하는 일련의 과정을 수행하는 과정에서의 경쟁력이다. 흡수역량의 결정요소는 이전 지식(prior knowledge), 유인구조(incentive structure), 조직의 관행 (routine), 사회적 네트워크 (social network) 등이 강조된다. 거칠게 이야기하자면 지식기반이론은 공부잘하는 기업이 경쟁우위를 갖는 다는 것인데 기업의 공부 역량은 지금까지 무엇을 공부했는지(prior knowledge), 왜 공부를 해야하는지에 관한 유인이 분명한지(incentive structure), 공부는 어떻게 하고 있는지(routine), 공부는 누구한테 배우고 누구와 같이 하는지(social network) 등에 의해서 결정된다는 것이다. 데이터 기반 기업의 밸류업이란 지식기반이론에서 보면 어떻게 공부잘해서 시험 잘보는 기업을 만드는가에 관한 문제다. 이에 대하여 각각 자세히 논의해보자.

첫째, 이전지식: 이전 지식이 흡수역량을 결정한다는 것은 다른 말로 현재 회사가 어떤 데이터를 가지고 있느냐가 향후 데이터 수집과 처리 역량에 중요한 영향을 미친다는 뜻이다. 많은 기업들이 데이터가 충분하지 않다고 주장하는데 생각보다 좋은 데이터를 보유한 경우가 많다. 본인들이 가진 데이터를 제대로 파악도 못하고 (firms don't know "who knows what") 활용도 못하고 있는 것이다. 이는 다음 요소인 유인구조와 관련된다. 참고로 이전지식의 중요성은 지식기반이론에서 이야기하는 조합역량(combinative capabilities; Kogut & Zander, 1992)과도 관련이 된다. 조합역량은 조직이 내부와 외부에서 접근 가능한 기존 지식을 결합하여 새로운 스킬을 습득하는 역량이다.

둘째, 유인구조: 유인구조 관련 특히 문제가 되는 경우는 데이터 담당자들이 데이터를 공유하는데 비협조적이고 적대적인 경우가 많다는 점이다. 데이터를 공유해봐야 업무을 침해 당할 우려가 있다. 게다가, 자신들이 관리하고 분석하던 데이터를 박사급 혹은 전문 연구자들에게 보여주는 것에 큰 부담을 느끼는 것은 당연하다. 이런 상황은 반드시 타개되어야 데이터 과학으로 기업 밸류업이 가능하다. 다른 말로 데이터 밸류업의 첫걸음은 데이터를 누가 어떻게 가지고 있는지 파악하고 그들이 데이터를 공유하고 전문가와 분석하는데 협조할 수 있도록 해야한다.

셋째, 관행: 데이터 관련 관행이 잘못된 경우가 많다. 특히 데이터의 테이블 형식 등이 체계적으로 관리되지 않는 경우가 많다. 데이터 테이블이 잘 정리되어 있지 않으면 API로 데이터를 공유하는 것이 의미가 없어지고 그렇다면 제대로된 데이터 과학을 할 수 없다. 이러한 포멧의 문제는 둘째로 언급한 유인구조와 연관이 있다. 조직내에서 사용할 데이터의 테이블 표준 포맷 정의는 데이터를 관리할 IT인력이 하는 경우가 많다. 그러나 그렇게 하면 안된다. 테이블과 변수 정의는 데이터를 가장 많이 사용할 현업부서에서 정의해야 한다. 그렇게 해야 데이터가 조직내에서 공유가 된다. 현업부서(프런트)에서 향후 API로 연결될 수 있는 표준 테이블과 변수를 정의해야 그들이 사용할 수 있다. 만약 현업부서에서 역량이 없다면? 그렇다면 경영/경제 학자들에게 어떻게 패널 데이터(panel data) 테이블을 정의하면 되는지 자문을 구하면 된다. 비슷한 유형의 문제나 데이터를 다룬 논문을 찾아서 논문 저자를 섭외하는 등 관련 분야의 전문가를 섭외하면 된다. 프런트가 정의한 테이블과 변수에 따라 IT 부서에서 데이터를 정리하고 관리하면 향후 협업도 훨씬 자연스럽게 이루어질 수 있다. 이후 API를 통해서 조직 내의 각 부서가 표준 테이블로 정의된 데이터를 활용하게 할 수 있고 경우에 따라서 데이터 거래소 등에 팔 수도 있을 것이다.

결론

기업들은 밸류업 잠재력이 큰 데이터를 제대로 활용하지 못하고, 잠재력이 낮은 분야에 데이터 과학을 남용하는 경우가 많다. 초미시데이터(IoE, IoT data)를 이용한 초미시 의사결정, 거시수준의 데이터를

활용한 중장기 전략 수립, 비정형 데이터를 이용한 위험관리에서 데이터 과학의 잠재력은 크지만 국내 기업들의 활용은 낮은 편이다. 그 대신 기업들은 상대적으로 구하기 쉽고 잘 정리되어 있다는 이유로 메소데이터 분석에 집중하는 경향이 있고 여기서 나온 결과를 단기 전략 수립에 사용하려는 시도를 하는 경향이 있다. 그러나 단기전략 수립에는 딥러닝 등 기술로만으로는 유용한 시사점을 찾을 수 없다. 데이터 구조의 문제는 물론 인과관계를 파악하는 것이 중요하고 이를 위해서 경제/경영학에 정립된 방법들을 사용해야 한다. 이를 위해서는 관련 업계 경험과 이론이 중요하다. 따라서 기업들은 등한시되고 있는 분야에 역량을 투입하고, 인과관계 파악에 전문성이 있는 연구자를 섭외해야하고, 순수 데이터 전문가(예: 순수 인공지능 연구자) 보다는 이론과 실무를 갖춘 관련 응용 연구자들과 데이터 분석에 협력해야 한다.

본 연구는 기업의 데이터를 이용한 밸류업 관련 세가지 관점을 강조한다: 아키텍처 혁신, 기업행동이론, 지식기반이론. 첫째, 데이터와 관련 기술의 개별 컨셉보다는 관계를 재정립하여 혁신을 추구하는 아키텍처 혁신으로 기술 중심이 아닌 현업의 문제와 전문성을 중심으로 혁신전략을 수립해야 한다. 둘째, 기업행동이론에서 강조하는 나이티안 불확실성과 이해관계자 갈등을 중심으로 문제를 정의한 후 ABM, 신규사업 기회, 이해관계자 자본, 제도적 관점의 데이터 전략 중 선택하여 동원해야 한다. 마지막으로 지식기반이론 관점에서 데이터를 지속가능한 경쟁우위의 원천으로 변화시켜야 한다. 이를 위해서는 기존 데이터, 유인구조, 데이터 관련 관행을 점검해서 흡수역량을 강화해야 한다.

참고문헌

Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Blyler, M., & Coff, R. W. (2003). Dynamic capabilities, social capital, and rent appropriation: Ties that split pies. *Strategic Management Journal*, *24*(7), 677–686.

Christensen, C. M. (2013). *The innovator's dilemma: When new technologies cause great firms to fail*. Harvard Business Review Press.

Coff, R. W. (1999). When competitive advantage doesn't lead to performance: The resource-based view and stakeholder bargaining power. *Organization Science*, *10*(2), 119–133.

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, 128–152.

Crane, A., Palazzo, G., Spence, L. J., & Matten, D. (2014). Contesting the value of "creating shared value." *California Management Review*, *56*(2), 130–153.

Cyert, R. M., & March, J. G. (1963). A behavioral theory of the firm. *Englewood Cliffs, NJ*, *2*(4), 169–187.

Godfrey, P. C. (2005). The relationship between corporate philanthropy and shareholder wealth: A risk management perspective. *Academy of Management Review*, *30*(4), 777–798.

Godfrey, P. C., Merrill, C. B., & Hansen, J. M. (2009). The relationship between corporate social responsibility and shareholder value: An empirical test of the risk management hypothesis. *Strategic Management Journal*, *30*(4), 425–445.

Henderson, R. M., & Clark, K. B. (1990). Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms. *Administrative Science Quarterly*, 9–30.

Hobbes, T. (1651). *Leviathan (Project Gutenberg eBook of Leviathan, 2009)*. http://www.gutenberg.org/etext/3207

Hobsbawm, E. (1962). The age of revolution: Europe 1748-1848. *London: Weidenfeld and Nicholson*.

Kane, G. C., Palmer, D., Phillips, A. N., Kiron, D., & Buckley, N. (2015). Strategy, not technology, drives digital transformation. *MIT Sloan Management Review and Deloitte University Press*, *14*(1–25).

Kang, H.-G., Woo, W., Burton, R. M., & Mitchell, W. (2018). Constructing M&A valuation: How do merger evaluation methods differ as uncertainty and controversy vary? *Journal of Organization Design*, *7*(1), 2.

Keynes, J. M. (1921). *A treatise on probability*. Courier Corporation.

Keynes, J. M. (1937). The general theory of employment. *The Quarterly Journal of Economics*, *51*(2), 209–223.

Knight, F. H. (1921). *Risk, uncertainty and profit*. Courier Corporation.

Kogut, B., & Zander, U. (1992). Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization Science*, *3*(3), 383–397.
Koh, P.-S., Qian, C., & Wang, H. (2014). Firm litigation risk and the insurance value of corporate social performance. *Strategic Management Journal*, *35*(10), 1464–1482.
Lewis, M. (2004). *Moneyball: The art of winning an unfair game*. WW Norton & Company.
Minor, D., & Morgan, J. (2011). CSR as reputation insurance: Primum non nocere. *California Management Review*, *53*(3), 40–59.
Peng, M. W., Sun, S. L., Pinkham, B., & Chen, H. (2009). The institution-based view as a third leg for a strategy tripod. *Academy of Management Perspectives*, *23*(3), 63–81.
Porter, M. E., & Kramer, M. R. (2011, January 1). Creating Shared Value. *Harvard Business Review*, *January–February 2011*. https://hbr.org/2011/01/the-big-idea-creating-shared-value
Rawls, J. (1971). *A theory of justice*. Harvard university press.
Shiu, Y.-M., & Yang, S.-L. (2017). Does engagement in corporate social responsibility provide strategic insurance-like effects? *Strategic Management Journal*, *38*(2), 455–470.

목차

서론

전통 기업의 데이터 활용에 관한 이슈들
- 초미시데이터
- 메소(meso) 데이터와 단기 전략
- 거시데이터와 중장기 전략
- 비정형 데이터와 위험관리

기업의 데이터 기반 밸류업 전략 수립을 위한 프레임워크
- Architectural Innovation
- 기업행동이론(BTF: a behavioral theory of the firm)
- 지식기반이론

결론

참고문헌