

데이터 발생시의 관계성과 상호작용성이 빅데이터 분석에 미치는 영향 연구

이정선*, 안진호**

*울지대학교 장례지도학과, **(주)아이디이노랩

A Study on the Effect of Relationship and Interaction at the Time of Data Generation on Big Data Analysis

Lee, Jung Sun, Ahn, Jin ho

Ulji University, IDINNOlab corp

E-mail : jslee@ulji.ac.kr, pibuchi@gmail.com

요약

본 연구는 빅데이터 기반의 데이터 분석의 품질의 영향 요인에 관한 연구이다. 빅데이터의 중요성이 증가하면서 무분별한 적용이 문제가 될 수 있다. 경우에 따라서는 빅데이터 적용이 적합하지 않은 경우가 발생하는 데 그에 대한 원인을 데이터 발생시점에 국한하여 연구를 진행하였다. 구체적으로 사람과 사람, 사람과 사물간의 관계성과 상호작용성이 강한 경우에서 발생하는 데이터가 빅데이터 분석의 품질에 영향을 미치는 것으로 검증되었으며, 이에 대한 대안으로는 민속지학(ethnography)과 사용자경험(User eXperience)에서 주로 사용하고 있는 관찰조사를 통한 데이터 수집 방식이 필요하다.

1. 서론

Big data의 활용가치는 데이터 안에 숨어있는 가치있는 정보의 의미가 궁금하기 때문이다. 여기서 주목해야 할 부분은 숨어있는 정보가 반복되는 패턴에 숨어있다면 다행이지만, 그렇지 않은 경우가 문제가 된다.

피쳐폰이 대세이던 시절에 아이폰을 중심으로 스마트폰의 대중화에 대한 패턴이 그 당시 Big data 분석에서 나타나지 않았다. 글로벌 컨설팅회사들과 당시 대부분의 피쳐폰 제조회사의 데이터 분석에서 나타나지 않았다. 이것은 데이터 분석기술의 문제가 아니라, 유의미한 패턴으로 나타나야 할 데이터들이 그 모집단 안에 나타나지 않았다는 것을 의미한다.

본 연구는 빅데이터 기반의 데이터분석에 있어

서의 관계성과 상호작용성의 영향에 관한 연구이다.

이에 본 연구에서는 그 모집단 설정의 오류 발생의 원인이 사람과 사람, 사람과 사물간의 관계와 상호작용이 강한 경우에 발생하는 데이터의 측정 및 해석은 그 강도가 강한 경우에는 User eXperience 와 ethnography 관점으로 수집된 데이터를 모집단에 포함하는 것이 중요하다는 것을 Big data의 수집시점에서 그 의미를 파악하여 최적의 방향성 도출을 위한 탐색을 시도한다.

2. 본론

2.1 Relationship의 의미

Relationship은 둘 이상의 사람, 사물, 현상 따위가 서로 연결되는 성질을 의미하는 것으로서, 분리되

거나 독립적이고 불연속적인 것과 대비되는 말이다.

Relationship은 단지 나와 타자와의 관계만을 일컫는 것이 아니라, 나와 자기 자신, 타자, 자연, 지구, 우주에 이르기까지 생물과 무생물 모두를 포함하여 모든 것과 관련이 있고, 서로 연결되어 있는 성질을 의미한다(강선보, 2008).

본 논문에서는 Relationship의 범위를 인간과 인간, 인간과 사물간의 관계로 한정하여 해석하고자 하는 데이터의 특성이 Relationship의 강도에 따라서 디지털화된 Big data분석 방식만으로 해석하는 것이 적합한지 아닌지의 의미를 살펴보고자 한다.

Relationship의 강도가 데이터 해석에 있어서 중요한 요인일 수 밖에 없는 이유는 다음과 같다. 인간은 다른 동물들에 비해 가장 무기력한 상태로 태어난다. 그렇기 때문에 출생의 순간부터 생명유지에 필요한 기본적인 조건들을 누군가의 도움으로 해결해야 하는 나약하고 의존적 존재이다. 이러한 불리한 신체적 조건을 가지고 약육강식의 환경 속에서 생존할 수 있는 유일한 방법은 서로 힘을 합치는 협동적인 생활방식을 택하는 것이다. 신체적 나약함이라는 인간의 생물학적 조건은 인간을 타인과의 관계 속에서 살아가야 하는 사회적 존재로 만들어 왔다. 인간은 이러한 과정을 통해 타인의 존재가 반드시 필요하며 타인과의 상호의존적 관계가 중요한 가치임을 의식적, 무의식적으로 습득하게 된 것이다(Fromm.E.1963, 이성태, 2006).

이러한 무의식속에 숨어있는 본성이 관계하는 Relationship이 강한 경우에는 드러나있는 데이터에서만 의미를 해석하는 것에 한계가 있다. 사람이 지금하고 있는 행동이나 행위가 내재된 욕구나 욕망을 그대로 표현하지는 않는다. 관계의 강도에 따라서 욕구와 욕망이 변하기에 현재 측정되어 있는 데이터에서 그 의미를 제대로 해석할 수 없다.

기준	유형
일차적 vs 이차적	본인 선택의사에 상관없이 부여된 관계인지, 가입 및 탈퇴가 자유로운 일시적 관계인지 선택 ¹⁾
수직적 vs 수평적	동등성이 없는 불평등한 수직적 관계인지, 자발적 상호작용의 가능성이 높

기준	유형
	은 수평적 관계인지 선택 ²⁾
실존적 vs 역할적	진실성이 있는 실존적 만남의 관계인지, 욕구충족을 위해 꾸며진 만남(스침)을 통한 역할적 관계인지 선택 ³⁾
협동적 vs 경쟁적	상대방에 대해 느끼는 감정과 태도에 따라 협동적 관계인지, 경쟁적 관계인지 선택

표 1 Relationship 유형

본 연구에서는 문헌연구를 바탕으로 하여 Relationship이란 사람과 사람, 사람과 사물 사이에서 심리적 · 정서적 · 신체적으로 연결되거나 관련되는 정도를 의미한다.

2.2 Interactivity의 의미

상호작용(interaction)은 인간이 주어진 환경에서 타인과 행하는 행위, 또는 인간과 사물 혹은 존재(entities) 하는 것들 사이에 주고받는 모든 행위로서 이러한 행위 가능성을 제공하는 대상을 상호작용적(interactive)이라 한다(Lombard, Snyder & Dutch, 2001)⁴⁾. 반면에 상호작용성(interactivity)은 주로 인간과 인간을 넘어서 인간과 사물간의 관계까지로 확장시켜 보는 것이다. 상호작용이 주로 대인 커뮤니케이션 영역에서 일어나는 인간의 행위를 표현하는 말로 사용된다면 상호작용성은 미래 기술과 서비스의 핵심 요소로 대인 간 상호작용을 넘어 미디어(media)나 디바이스(device) 등의 비인적 관계의 범위까지 확장될 수 있다(Pavlik, 1996).

Interactivity이란 사람과 사람, 사람과 사물 사이에서 행동이나 반응을 바꾸는 개인 또는 그룹사이의 사회적 행동의 동적인 변화 과정이다. 이 논문에서는 서로 영향을 주고받는 행위나 소통방식

1) 권석만(1997), 앞의 책, p.37-38참조

2) 최지민(2014), 관계성(Relationship)의 도덕교육적 의미 연구, 한국교원대학교 대학원

3) 이재창(1998), 『인간관계론』, 서울:문음사, pp.26-27.

4) Lombard, M. S. D., & Snyder-Dutch, J. J.(2001):"Interactive Advertising and Presence: A Framework". Journal of Interactive Advertising, 1(2), 581-592.

(communication)으로서 상호작용(Interaction)을 비인적 관계의 범위까지 확장한 것으로 정의한다.

본 연구에서는 기존의 빅데이터 연구에서 소외되었던 외부에서 생성되는 비정형데이터로서 관계성과 상호작용성 관점에서 모집단 산정의 의미를 연구하였다. 이러한 관계성과 상호작용성이라는 관점에서 데이터를 바라보면, 깊이 있게 보는 것 보다는 맥락적 관점에서 두껍게 바라볼 필요가 있기에 이러한 데이터를 thick data라고 정의하였다.

기준	구조
통제 (주도)	상호작용의 내용, 시기, 순서적 흐름 등에 대한 대안을 찾거나 변환하는 것에 있어서의 통제력의 강도
역할 교환	메시지 전달자와 수용자간의 역할을 주고받는 것의 강도
상호적 대화	A와 B가 대화할 때 B에 대한 A의 반응이 먼저 일어난 대화에서 B의 A에 대한 반응에 좌우되면서 순차적으로 발생하는 강도

표 2 Interactivity 3차원 구조

3. 연구가설 및 분석

문헌연구를 토대로 하여 사람과 사람, 사람과 사물간에서 관계성과 상호작용성의 강도와 그 안에서 수집된 데이터의 품질에 어떠한 영향을 미치는지를 판단하기 위하여 아래와 같은 가설을 설정하였다.

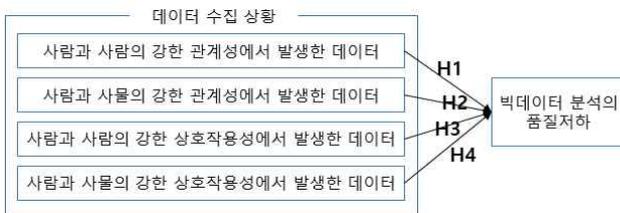


그림 1 연구모형

H1. 사람과 사람의 관계성이 강한 경우에서 발생하는 데이터는 빅데이터 분석의 품질을 저하시킬 것이다.

H2. 사람과 사물의 관계성이 강한 경우에서 발생하는 데이터는 빅데이터 분석의 품질을 저하시킬 것이다.

H3. 사람과 사람의 상호작용성이 강한 경우에서

발생하는 데이터는 빅데이터 분석의 품질을 저하시킬 것이다.

H4. 사람과 사물의 상호작용성이 강한 경우에서 발생하는 데이터는 빅데이터 분석의 품질을 저하시킬 것이다.

본 연구는 빅데이터의 수집, 가공, 분석에 관계하는 기업의 IT 및 데이터분석 담당자를 대상으로 설문을 진행하였다. 2019년 5월에서 8월 사이에 총 150명에게 온라인상에서 설문을 진행하였고, 답변의 일관성과 분석자료가 부적절한 부분을 제외하고 128부를 활용하여 통계 분석을 진행하였다.

변수	조작적 정의	측정, 수집이 어려운 원인
사람과 사람의 강한 관계성	데이터 측정, 수집이 어려운 사람과 사람사이의 관련되거나, 서로 연결되어 있는 정도가 강한 상태에서 발생하는 데이터	사람은 행동이나 행위에서 내재된 욕망과 욕구를 그대로 표현하지 않는 본성으로 기존 일반 데이터 측정 및 수집 방식으로는 어려움
사람과 사물의 강한 관계성	데이터 측정, 수집이 어려운 사람과 사람사이의 관련되거나, 서로 연결되어 있는 정도가 강한 상태에서 발생하는 데이터	
사람과 사람의 강한 상호작용성	데이터 측정, 수집이 어려운 사람과 사람 사이에서 서로 영향을 주고받는 행위나 소통방식이 강한 상태에서 발생하는 데이터	
사람과 사물의 강한 상호작용성	데이터 측정, 수집이 어려운 사람과 사물 사이에서 서로 영향을 주고받는 행위나 소통방식이 강한 상태에서 발생하는 데이터	
빅데이터 분석의 품질저하	인사이트를 도출하기 위해서 알고리즘과 수학적 처리 과정을 적용하여 해당 정보에 대한 결론을 도출하고 패턴을 찾기 위한 목적으로 데이터를 다루는 활동의 품질 저하	

표 3 조작적 정의

연구가설의 검증을 위해 SPSS에서 측정항목의 신뢰도, 타당도를 측정하였다. 연구가설의 검증을 위해 SPSS프로그램에서 측정항목의 신뢰도와 타당성 부분을 검정하였다. 각 요인들을 구성하는 항목의 신뢰성을 탐색적 요인분석과 크론바하 알파값으로 측정하였다. 가설검정에 사용된 5개의 변수들의 크론바하 알파값이 모두 0.8 이상이었기에 상

호 관련성과 방향성을 측정하기 위한 아래와 같은 상관분석을 실시하였다.

요인	1	2	3	4	5
1. 사람과 사람의 강한 관계성의 데이터	1				
2. 사람과 사물의 강한 관계성의 데이터	.735	1			
3. 사람과 사람의 강한 상호작용성의 데이터	.725	.695	1		
4. 사람과 사물의 강한 상호작용성의 데이터	.676	.677	.680	1	
5. 빅데이터 분석의 품질저하	.700	.732	.715	.684	1

표 4 변수간의 상관관계 분석(**p<0.01)

위의 표를 참고하면 변수간의 상관분석에서 모든 상관계수들의 값은 0.6 이상이며, 유의도에서도 0.01이상으로 유의하다. 이와 같은 전제로 독립변수와 종속변수의 회귀분석을 통한 가설 설정의 의미가 있다고 판단하였다. 이는 변수간의 방향이 설정한 가설의 방향과 적합하기에 사용 항목들은 규범타당성이 타당하다.

종속 변수	독립 변수	표준 오차	베타	t	유의 확률	공차
빅데이터 분석의 품질저하	(상수)	.279		2.644	.000	
	사람과 사람의 강한 관계성 데이터	.115	.084	2.799	.000	.353
	사람과 사물의 강한 관계성 데이터	.104	.491	2.833	.000	.291
	사람과 사람의 강한 상호작용성 데이터	.072	.187	2.372	.000	.304
	사람과 사물의 강한 상호작용성 데이터	.104	.177	2.791	.000	.413
R= .696, R제곱= .444, 수정된 R제곱= .412 Durbin-Watson=1.827, F=36.656, P=.000						

표 5 회귀분석 결과

데이터 발생시의 강한 관계성과 상호작용성 요인이 빅데이터의 품질저하에 유의미한 영향을 줄 것이라는 가설을 검정한 결과 H1에서 H4까지 모두 유의한 영향을 미치는 것으로 가설이 검정되었다.

3. 결론

본 연구는 빅데이터의 중요성이 커지는 시점에 관련 연구와 검정이 부족한 데이터 발생시점에 관한 연구를 진행하였다. 결론적으로 데이터 발생시의 관계성과 상호작용성이 강한 경우에는 일반적으로 진행되고 있는 빅데이터 분석방식이 적합하지 않을 수 있다는 사실을 알게 되었다.

데이터를 수집, 가공하는 방식에 있어서 최적화된 수단이 인류학자들이 다른 문화의 일상생활에 들어가 가까이에서 관찰하고, 기록하는 연구 과정 방식으로 다른 문화를 상세히 기록하고 설명하는 방식인 ethnography와 사용자가 어떤 시스템, 제품, 서비스를 직/간접적으로 이용하면서 느끼고 생각하게 되는 지각과 반응, 행동 등에 관한 총체적 감정과 기억을 중점적으로 분석하는 User eXperience적 접근이라 할 수 있다. 이 두 가지 방법은 관찰대상인 사람을 중심으로 사회적, 문화적, 경제적 관점의 다양한 맥락을 관찰하여 데이터를 수집하고, 해석하는 방식이다.

향후 민속지학(ethnography)와 사용자경험(User eXperience)에서 주로 사용하고 있는 관찰조사를 통한 데이터 수집 방식의 필요성에 대한 심도있는 연구가 데이터사이언스적 관점에서 필요할 것으로 보인다.

[참고문헌]

- [1] 강선보 외(2008), 「21세기 인성교육의 방향설정을 위한 이론적 기초연구」, 『교육문제연구』
- [2] 이재창(1998), 『인간관계론』, 서울:문음사
- [3] Fromm.E.(1963), The art of loving, New York:Bantam Book
- [4] 이성태 (2006), 『인간관계론』, 서울:양서원
- [5] 최지민(2014), 관계성(Relationship)의 도덕교육적 의미 연구, 한국교원대학교 대학원
- [6] Lombard, M. S. D., & Snyder-Dutch, J. J.(2001):“Interactive Advertising and Presence: A Framework”. Journal of Interactive Advertising, 1(2), 581-592.
- [7] Pavlik, J. V. (1996). New media technology: Cultural and commercial perspectives. Allyn & Bacon.