

빅데이터 분석결과와 실증조사 결과의 비교

김도관¹ · 신성윤^{2*}

Comparing the Results of Big-Data with Questionnaire Survey

Do-Goan Kim¹ · Seong-Yoon Shin^{2*}

¹Division of Information and Electronic Commerce, Wonkwang University, Iksan 54538, Korea

²Department of Computer Information Engineering, Kunsan National University, Kunsan 54150, Korea

요 약

스마트폰 보급의 확산과 데이터 저장 및 분석 기법의 발전은 빅데이터 관련 산업을 미래의 유망 산업으로 탈바꿈하게 만들었다. 마케팅 분야에서는 소셜 데이터를 분석하여 소비자의 니즈를 파악하고, 효과적인 마케팅의 수단으로 활용하고 있다. 빅데이터 분석이 불가능했던 시대에는 소비자를 이해하기 위해서는 소수의 소비자를 대상으로 하는 조사 및 실험에 의존할 수밖에 없었으며, 이러한 전통적인 시장조사 방법은 현재도 활용되고 있다. 빅데이터 분석과 전통적인 조사방법 모두 고객을 이해하는 중요한 방법이지만, 두 가지 방법을 통해 도출된 결과가 소비자의 트렌드에 대하여 유사한 시사점을 주는지는 확인할 필요가 있다. 이러한 점에서 본 연구에서는 화장품 브랜드를 대상으로 소셜 데이터 분석 결과와 소비자를 대상으로 하는 설문조사의 결과를 비교하고자 하였다. 연구 결과 두 가지 방법 모두 유사한 시사점을 제공하는 것으로 나타났다.

ABSTRACT

The rapid diffusion of smart phones and the development of data storage and analysis technology have made the field of big-data a promising industry in the future. In the marketing field, big-data analysis on social data can be used for understanding the needs of consumers as an effective and efficient marketing tool. Before the age of big-data, companies had relied upon the traditional methods such as questionnaire survey and marketing test in which a small number of consumers had participated. The traditional methods have still been used. Although both of big-data analysis and traditional methods are useful to understand consumers. It is need to check whether the results from both include similar implications. In this point, this study attempts to compare the results of big-data analysis with that of questionnaire survey on some cosmetics brands methods. As the results of this study, both results of big-data analysis and questionnaire survey include similar implications.

키워드 : 빅데이터, 설문조사, 마케팅, 화장품, 브랜드

Key word : Big-Data, Questionnaire Survey, Marketing, Cosmetics, Brand

Received 30 October 2016, Revised 31 October 2016, Accepted 06 November 2016

* Corresponding Author Seong-Yoon Shin(E-mail:s3397220@kunsan.ac.kr, Tel:+82-63-850-6281)

Department of Computer Information Engineering, Kunsan National University, Kunsan, 54140, Korea

Open Access <http://dx.doi.org/10.6109/jkice.2016.20.11.2027>

print ISSN: 2234-4772 online ISSN: 2288-4165

©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.
Copyright © The Korea Institute of Information and Communication Engineering.

I. 서론

최근 스마트폰 보급의 확산과 소셜 서비스 활용의 증가로 인하여 일반 대중들의 창조해낸 엄청난 양의 데이터들이 네트워크상에 존재하게 되었다. 이러한 일반 대중들이 만들어낸 데이터는 빅데이터 분석을 통해 사회를 이해하는 소중한 자원으로 활용되고 있으며, 기업 측면에서는 빅데이터가 하나의 경영자원으로서 소비자 행동과 시장의 변화를 예측하여 기업의 전략에 활용되고 있다[1].

21세기의 석유에 비유되는 빅데이터는 과거에 기술적 한계로 인하여 불가능했던 것들이 데이터 저장기술과 분석 기술의 발전이 토대가 되어 유용한 산업적 가치뿐만 아니라 미래 유망 산업으로 떠오르고 있다[2]. 빅데이터는 단어가 의미하는 바와 같이 대용량의 데이터를 의미하지만, 기업 내적으로는 기업 내부에 흩어져 존재하던 다양한 데이터의 통합을 의미하기도 하며, 공공분야에서는 각각의 기관들이 보유하고 있는 데이터의 통합을 통한 데이터로의 발전을 의미하기도 한다. 또한 소셜 서비스 활용을 통해 일반 대중들이 생산해낸 엄청난 양의 데이터들을 의미하기도 한다.

현재 마케팅 분야에서는 소셜 서비스 상에 존재하는 일반 사용자들이 생산해낸 다양한 데이터들을 소셜 비즈니스 인텔리전스(Social Business Intelligence)를 통해 소비자의 니즈(needs)를 파악하고 있으며[3], 역으로 소셜 서비스를 통하여 마케팅을 위한 유용한 도구로 활용하고 있다.

빅데이터 분석이 불가능했던 시대에는 기업들은 소비자를 이해하기 위하여 설문조사 및 실험과 같은 직접 조사 방법에 의존할 수밖에 없었다. 그러나 이러한 전통적인 시장조사방법들은 조사 설계자의 의도 및 편견이 개입될 수 있으며, 소수의 소비자를 대상으로 하며, 시간과 비용이 많이 소요된다는 단점을 가지고 있다. 이에 비해 소셜 서비스에 대한 빅데이터 분석은 일반 대중이 생성한 데이터를 사후에 분석하는 것이기 때문에 조사과정의 편견과 외부 영향으로부터 자유롭다. 그러나 소셜 데이터에 대한 빅데이터 분석의 단점은 분석 방법을 기존의 분석 서비스를 제공하는 곳의 데이터 분석 방법에 의존할 수밖에 없기 때문에 연구자가 직접적으로 데이터를 확보하지 못할 경우 분석의 폭이 제한될 수 있다.

이러한 이유에서 현재 많은 기업들은 소비자를 이해하는 방법에 있어서 빅데이터 분석과 전통적인 조사방법을 병행하여 활용하고 있다. 만약 두 가지 방법을 통하여 나타난 결과가 유사하거나 동일한 시사점을 내포하고 있다면 빅데이터 분석의 가치가 상승될 수 있지만, 그 반대의 경우 빅데이터 분석결과가 실제 시장의 전개 상황과 다르게 나타난다면 기업들 입장에서 빅데이터의 활용에 의구심을 가지게 될 것이다.

이러한 점에서 본 연구에서는 빅데이터 분석 결과와 전통적인 설문조사를 통한 결과가 동일한 시사점을 제공 여부의 확인을 주목적으로 하고 있다. 이러한 연구 목적을 위해 본 연구에서는 블로그와 트위터에서 가장 높은 빈도를 보인 화장품 브랜드를 선정하고, 이들 브랜드에 대한 소셜 네트워크 상에서의 여론을 분석하여, 인지도와 선호도의 순위를 제시하였다. 그 이후 브랜드들에 대한 인지도와 선호도에 대한 설문조사 결과와 비교하고자 하였다.

II. 빅데이터 활용연구

일반 대중에 의해 생성되는 소셜 서비스 상의 방대한 양의 데이터들은 방대한 텍스트에서 의미가 있는 정보를 추출하여 다른 정보와 연계하는 텍스트 마이닝 기술과 텍스트에 대한 긍정, 부정, 중립과 같은 선호도를 파악하는 오피니언 마이닝 기술, 소셜 네트워크의 연결구조 및 연결강도들 분석하는 소셜 네트워크 분석, 페이스북이나 트위터에서 관측되는 대상 집단에 대한 군집 분석 기술 등이 활용될 수 있다[4].

빅데이터 활용 관련 연구에서는 소셜 빅데이터를 마이닝 기반 기술을 통해 분석하여 주요 이슈에 대한 분석 보고서를 자동으로 생성하게 하는 시스템에 대한 연구도 이루어졌으며[5], 빅데이터 기술의 도입 및 활용을 위한 요인들에 관한 연구에서는 빅데이터의 특성요인으로 예측성, 관리성, 지원성, 경쟁성 등이 선정되어 요인에 대한 중요도 분석이 이루어지기도 하였다[6].

빅데이터 산업의 활성화와 관련된 연구에서는 정부의 중장기 정책수립, 빅데이터 지원 기관 설립, 법 제도 개선, 개인 정보보호법 개선, 빅데이터 인식 개선 지원, 수요 발굴 지원, 전문인력 양성, 인프라시스템 지원, 데이터 유통 및 활용 지원, 데이터 품질관리, 기술 연구 및

발굴, 기반 기술 활성화와 같은 다양한 전략들이 논의되었다[7].

관광분야에서의 한 연구에서는 ‘캠핑’이라는 검색어를 통해 나타난 23만 건에 이르는 텍스트를 수집 정제하여 분석한 후 정부의 정책과 인식의 변화를 제안하기도 하였다[8]. 국내의 온라인 뉴스, 블로그, 카페, 소셜 네트워크 서비스, 게시판 등에 나타난 소셜 빅데이터를 의사결정 나무 분석기법을 적용하여 국내 빅데이터 수요공급에 대한 예측모형을 제안하기도 하였다[9]. 영화 산업과 관련하여 영화 흥행을 위한 빅데이터 활용효과에 대한 연구에서 오피니언 마이닝 처리과정을 통하여 영화의 흥행을 예측하는 척도로 활용할 것을 제안하기도 하였다[10].

이와 같이 빅데이터 연구는 기반기술의 발전을 위한 연구와 함께 분석을 통한 결과의 활용이 한정된 분야가 아닌 다양한 분야에서 활용될 수 있는 가능성들을 제기하는 활용 측면의 연구들도 많이 등장하고 있다. 이러한 점에서 과연 빅데이터가 시장의 현실을 그대로 보여주는 거울인지 아니면 시장의 현실과 다른 일부분의 모습을 보여주는 하나의 도구인지는 지속적인 검증이 요구된다.

III. 분석 방법

본 연구에서는 화장품 브랜드에 대한 인지도와 선호도가 소셜 네트워크 서비스에서 나타난 결과와 설문조사를 통한 결과와 일치하는지를 알아보기 위한 연구이다. 연구에 필요한 분석을 위해 트위터와 블로그의 내용에 대한 텍스트 마이닝과 오피니언 마이닝을 제공하는 S사이트를 활용하였다.

먼저 화장품 브랜드를 선정하는 방법은 트위터와 블로그에 나타난 ‘화장품’이라는 텍스트와 함께 등장하는 높은 빈도를 나타내는 브랜드를 선정하였다. 빅데이터 분석에서의 선정된 브랜드의 인지도와 선호도를 알아보는 방법으로는 특정 브랜드명이 가장 많이 언급된 브랜드 순서에 따라 인지도의 순위를 평가하였으며, 선호도는 S사이트에서 제공하는 오피니언 마이닝의 결과를 토대로 긍정, 부정, 중립의 빈도를 바탕으로 결정하였다. 여기서 제기될 수 있는 문제는 단순한 빈도가 인지도를 대변할 수 있는가?라는 문제와 브랜드에 대한 오

피니언 마이닝의 빈도 및 비율이 선호도를 나타낼 수 있는가?라는 문제가 제기될 수 있다. 사전적 의미에서는 분명하게 다를 수 있지만, 소셜 서비스에 나타난 데이터의 여론에 근거하여 많은 빈도의 언급은 높은 인지도를 상징하고, 특정 브랜드에 대한 긍정 부정의 여론은 제품에 대한 선호도를 간접적으로 표현 할 수 있기 때문에 분석 결과를 그대로 활용하였으며, 빅데이터 분석을 통해 나타난 결과를 설문조사를 통한 직접조사의 결과와 비교하고자 하였다.

설문조사를 통한 브랜드의 인지도와 선호도 평가는 앞서 빅데이터의 텍스트 마이닝을 통해 선정된 브랜드를 가지고 일반 대중들에게 브랜드의 인지도와 선호도에 대한 순위를 체크하도록 하였다. 각각에 기입된 순위를 그대로 집계하여 순위를 선정하였다.

IV. 분석 결과

트위터와 블로그에서 나타난 빅데이터의 분석 대상은 2016년 10월 한 달동안 언급된 ‘화장품’ 관련 브랜드에 대한 메시지를 그 대상으로 하고 있다.

먼저 ‘화장품’이라는 키워드와 동시에 언급된 브랜드를 통하여 4개의 브랜드를 선정하였다. 4개의 브랜드는 P, I, A, E 브랜드가 선정되었다. ‘화장품’이라는 키워드와 동시에 언급된 브랜드의 빈도수는 표 1에 제시하였다. 표 1에서 언급된 Rank는 단순히 ‘화장품’과 동시에 언급된 빈도의 순위이며 특정 브랜드에 대한 언급 빈도를 나타내는 것은 아니다.

Table. 1 Selected Cosmetics Brands

Rank	Brand	Frequency
1	P	5,307
2	I	4,605
3	A	1,527
4	E	1,123

선정된 4개의 브랜드를 다시 키워드 빈도 분석을 실시한 결과 E브랜드가 가장 높은 빈도를 보이는 것으로 나타났으며, 그 다음으로 A, I, P 순서로 나타났다. 표 2에 제시된 단순한 빈도 결과에서 나타나는 추가적인 사항으로는 브랜드에 따라 의존하는 소셜 서비스가 차

이가 존재한다는 것이다. 예를 들어 E, A, I 브랜드는 트위터에서 언급되는 빈도가 높으면 반면에 P브랜드는 트위터 보다 블로그에서 언급되는 빈도가 높게 나타났다. 이러한 결과는 각각의 브랜드가 마케팅 활동을 어떤 소셜 서비스에 집중할 것인지에 대한 가이드라인을 제시해줄 수 있을 것이다.

Table. 2 Frequency of Cosmetics Brands

Rank	Brand	Twitter	Blog	Total
1	E	41,717	3,210	44,927
2	A	15,077	923	16,000
3	I	7,334	434	7,768
4	P	1,126	5,329	6,455

다음에는 특정 브랜드에 대한 긍정/부정/중립의 메시지의 유형을 분석하는 오피니언 마이닝의 결과를 표 3에 제시하였다. 특정 브랜드와 관련하여 언급된 긍정/부정의 어휘들을 15개까지 선정하여 그 합계의 결과를 제시하였다. 표 3의 Rank는 긍정/부정을 나타내는 텍스트의 빈도순으로 나타내지 않고 긍정의 비율이 가장 높은 순으로 순위를 선정하였다. 그 결과 P브랜드가 가장 긍정적인 텍스트가 포함된 것으로 나타났으며, 그 다음으로 E가 높게 나타났다.

Table. 3 Opinion Mining Results on Cosmetics Brands

Rank	Brand	Messages		Total
		Positive	Negative	
1	P	9,730	643	10,373
2	E	13,690	3,646	17,336
3	I	625	1,416	2,041
4	A	1,342	3,142	4,484

블로그와 트위터에 나타난 화장품 브랜드에 대한 텍스트 마이닝과 오피니언 마이닝 결과 언급횟수에 따른 브랜드 인지도와 선호도의 결과를 표 4에 제시하였다. 인지도의 순위는 E브랜드가 제일 높게 나타났으며, 그 다음으로 A, I, P 순으로 나타났다. 긍정과 부정의 텍스트의 비율을 통해 나타난 선호도에 대한 결과는 P 브랜드가 가장 높게 나타났으며, 그 다음으로 E, I, A 브랜드 순으로 나타났다.

Table. 4 Cosmetic Brand Awareness and Preference from Big-Data Analysis

Brand	Rank	
	Awareness	Preference
E	1	2
A	2	4
I	3	3
P	4	1

빅데이터 분석에 사용된 4개의 브랜드를 대상으로 하여 156명을 대상으로 하여 브랜드 인지도와 선호도에 대한 조사를 실시하였다. 설문 방법은 해당되는 4개의 브랜드를 제시하고 인지도와 선호도의 순위를 기재하도록 하였다. 그 결과는 표 5에 제시하였다.

Table. 5 Cosmetic Brand Awareness and Preference from Questionnaire Survey

Brand	Rank	
	Awareness	Preference
E	1	1
A	2	4
I	3	2
P	4	3

빅데이터 분석에서의 인지도와 선호도에 대한 조사 결과와 설문조사를 통한 조사결과를 비교해 보면 인지도는 두 가지 방법의 결과가 정확하게 일치하는 것으로 나타났다. 그러나 선호도에서는 일부 순위에 차이가 있는 것으로 나타났다. 선호도에 있어서 두 방법으로 부터의 결과 차이에 가장 큰 역할을 한 브랜드는 P 브랜드이다. 따라서 P 브랜드에 대한 추가 조사를 위해 일부 설문 응답자를 대상으로 그 이유를 묻거나, 접근 가능한 범위에서 P 브랜드의 내용을 검색하였다.

그 결과 P 브랜드는 화장품뿐만 아니라 다양한 생활용품을 주요 상품으로 하고 있으며, 화장품 브랜드로는 저가형 상품에 속해 있다는 것을 알 수 있었다. 즉 화장품 보다는 동일한 브랜드명을 가지고 있는 다른 상품에 대한 내용이 소셜 서비스의 내용에 반영되었다고 판단할 수 있다.

V. 결 론

소비자를 이해하기 위한 시장조사 방법은 전통적인 방법과 빅데이터 분석을 통한 두 가지 방법 모두 유용한 방법이 될 수 있다. 그러나 각각의 방법 모두 중요한 단점을 가지고 있다.

전통적인 설문조사와 같은 직접조사의 경우 많은 시간과 비용을 필요로 하며, 조사 설계자의 편견이 개입되어 결과가 왜곡될 수 있다. 또한 제한적인 대상을 상대로 조사를 하기 때문에 적은 수의 조사 참여자들이 과연 모든 일반 대상의 여론을 대표할 수 있는가에 대한 의문점을 남기게 된다.

반면에 빅데이터 분석을 통한 조사의 경우, 조사 설계자의 편견으로부터 자유로울 수 있으며, 불특정 다수의 의견을 바탕으로 결과가 도출되기 때문에 일반 대중의 여론을 보다 정확히 설명한다고 할 수 있다. 그러나 빅데이터를 통한 분석은 특정 문제만을 해결하고자 설계된 빅데이터와 분석도구가 존재하지 않는 이상 제한된 방법에 의해 도출된 결과를 통해 간접적으로 이해한다는 단점이 있으며, 때로는 수백만 건에 이르는 빅데이터의 내용에 대한 심도 깊은 내용 분석이 불가능하기 때문에 부적절한 결과의 해석에 노이즈가 포함될 수 있다[11].

본 연구결과에서 나타난 바와 같이 특정 브랜드의 마케팅 믹스로 인하여 결과에서 차이가 나타났다. 또한 오피니언 마이닝의 경우 특정 브랜드에 대한 부정의 의미가 아닌 경우, 예를 들어 “내 피부가 좋지 않다. 그래서 A라는 브랜드를 사용하였다”라는 메시지가 있다면 A라는 브랜드의 부정적인 여론의 빈도가 한 개 증가하게 된다.

현재의 분석 기술들에 의해 제공되는 빅데이터 분석 서비스는 아직까지 인간의 뇌의 수준을 따라잡지 못하고 있으나, 인공지능과 같은 컴퓨팅 기술의 발전은 이러한 제한점들을 점차 해결해 나갈 것으로 예상된다. 그럼에도 불구하고 화장품 브랜드에 대한 두 가지 분석의 결과는 약간의 차이는 존재하지만 대략적으로 일치하는 것으로 나타났다. 두 가지 방법에 사용된 측정 방법이 일치하지 않기 때문에 자료에 대한 표준화는 불가능하다. 그러나 두 가지 방법에서의 결과에서 의미하는 시사점은 일부 조사과정에서 노이즈로 작용한 요인을 제거하면 거의 유사한 결과를 도출하였다는 것이다.

소비자를 이해하기 완벽하게 이해할 수 있는 방법은 아마도 존재하지 않을 수 있다. 그러나 여러 가지 방법들이 결합되어 유의미한 결과를 도출한다면 소비자를 이해하는데 더욱 가깝게 다가 갈 수 있을 것이다. 이러한 점에서 본다면 빅데이터 분석은 소비자를 이해하는데 있어서 유용한 도구임이 본 연구결과에서 확인되었고 볼 수 있다.

향후 연구에서는 특정 상품이나 브랜드를 대변하는 속성의 키워드를 중심으로 관련 텍스트 마이닝과 오피니언 마이닝을 할 경우 보다 유용한 조사방법이 될 것으로 판단된다.

ACKNOWLEDGMENTS

This paper was supported by Wonkwang University in 2016.

REFERENCES

- [1] J. C. Koh, K. U. Lee, J. Y. Jeong, and K. S. Kim, “Correspondence strategy for big data's new customer value and creation of business,” *Journal of Korea Safety Management & Science*, vol. 14, no. 4, pp. 229-238, Dec. 2012.
- [2] S. K. Lee, “A review of big data analysis based on marketing perspective,” *Korea Journal of Business Administration*, vol. 28, no. 1, pp. 21-35, Jan. 2015.
- [3] M. J. Lee, and W. J. Kim, “Advanced analytics and technology for big data,” *Entrue Journal of Information Technology*, vol. 11, no. 1, pp. 45-56, Special Issue. 2012.
- [4] S. J. Shin, “SNS using big data utilization research,” *The Journal of the Institute of Internet, Broadcasting and Communications*, vol. 12, no. 6, pp. 267-272, Dec. 2012.
- [5] J. Heo, C. H. Lee, H. J. Oh, Y. C. Yoon, H. K. Kim, Y. H. Jo, and C. Y. Ock, “Web science: automatic generation of issue analysis report based on social big data mining,” *KIPS Transaction on software and data engineering*, vol. 13, no. 12, pp. 553-564, Dec. 2014.
- [6] Y. K. Jung, M. S. Suk, and C. J. Kim, “A study on the success factors of big data through an analysis of

- introduction effect of big data,” *Journal of Digital Convergence*, vol. 12, no. 11, pp. 241-248, Nov. 2014.
- [7] S. D. Yoo, K. D. Choi, and S. Y. Sin, “Characterizing business strategy in a new ecosystem of big data,” *Journal of Digital Convergence*, vol. 12, no. 4, pp. 1-9, Apr. 2014.
- [8] M. H. Jang, and Y. I. Yoon, “Research into changes in government policies and public perceptions on camping via analyses of big data,” *Tourism Study*, vol. 31, no. 1, pp. 91-112, Feb. 2016.
- [9] J. S. Kim, E. J. Kwon, and T. M. Song, “Big data supply demand prediction using social big data, applying data mining decision making tree,” in *Proceeding of the 2014 KMIS Spring conference*, Chunan: Korea, pp. 708-714, 2014.
- [10] J. W. Kim, “big data's effect on its practical use for box-office success of films: focusing on the process of management of opion mining on 'Roaring Currents,” *Journal of the Korea Entertainment Industry Association*, vol. 9, no. 2, pp. 81-90, Jun. 2015.
- [11] D. G. Kim, “The reflection of religions in big data and the usage of big data,” *Won-Buddhist Thought & Religious Culture*, vol. 64, no. 2, pp. 191-219, Jun. 2015.



김도관(Do-Goan Kim)(제1저자)

원광대학교 정보전자상거래 학부 조교수
※관심분야 : MIS, E-Business, Big-Data Analysis



신성윤(Seong-Yoon Shin)(교신저자)

군산대학교 컴퓨터정보학과 교수
※관심분야 : 영상처리, 컴퓨터 비전, 가상현실, 멀티미디어