

빅데이터 분석을 통한 개인정보자기결정권 영향 요인 모델 연구

이연수, 이욱한*

숙명여자대학교, *숙명여자대학교

pluto2002p99@sookmyung.ac.kr, *uhlee@sookmyung.ac.kr

A Study on Factors Affecting the Right to Control Personal Information through Big Data Analysis

Lee Yeon Su, *Lee Uk Han

Sookmyung Women's University, *Sookmyung Women's University

요 약

본 논문은 빅데이터의 시대 및 코로나 사태를 경험함으로써 개인정보의 활용 및 개인정보 보호의 필요성이 커짐에 따라 개인정보자기결정권 영향력이 커짐에 주목하였다. 이에 빅데이터 분석을 통하여 개인정보자기결정권에 영향을 미치는 요인 모델을 제시하는 데에 목적이 있다. 따라서 개인정보자기결정권에 대해 살펴본 후, 빅데이터 분석을 활용한 개인정보자기결정권 영향 요인 모델을 제안하고자 한다.

I. 서 론

4차 산업혁명의 시대는 데이터의 시대이다. 사회에서 개인정보는 ‘자원’으로서 경제, 사회 등 다양한 분야에서 기능하게 된 한편, 한편 개인의 신분을 나타냄으로서 ‘인격’으로의 성격을 나타내기도 한다. 이렇듯 개인정보는 데이터 시대에 융합된 정체성을 갖고 사회 전반에서 활용된다.

사회가 고도로 발전하고 진화할수록 개인정보는 개인 데이터로서의 이용 가치가 커지게 될 것인 한편, 국가적 재난 상황에서 개인정보의 활용 가능성이 커짐에 따라 개인정보를 형성하는 개인의 권리 보호도 중요하게 될 것이다. 개인정보의 활용 및 보호가 중요시된 것으로 현재의 COVID-19를 들 수 있다. 우리나라는 코로나19 사태의 대응과정에서 신속한 정보공개로 통한 방역시스템의 구축은 세계적으로 주목받았다. 그러나 이와 같은 주목의 시선에는 부정적인 시각도 있었다. 정보공개 시스템 실행의 이면에는 국가인권위원회의 권고에도 불구하고 확진자의 성별, 나이, 이동경로 등 개인을 특정하거나 유추할 수 있는 정보가 공개되었고, 방역에 필요한 수준을 넘어서 정보내용 공개도 있었다.[1] 이처럼 국가에 의한 개인정보의 처리는 시민들의 더 많은 자유의 보장과 안전에 기여할 수 있지만 인권에 대한 위협이 될 수 있다. 이러한 점에서 민주주의의 이념과 법치국가의 원리에 입각하여 국가의 감시행위를 효율적으로 통제할 수 있는 법적 장치가 필요하다.[2] 개인정보자기결정권은 스스로에 대한 정보 노출을 억제하는 데만 중심이 있는 것이 아닌 스스로에 대한 정보의 흐름을 적극적으로 형성하기도 한다. 따라서 빅데이터 분석을 통한 개인정보자기결정권 영향 요인 모델을 제안함으로써 개인정보자기결정권의 실효적 보장에 기여하고자 한다.

II. 본론

1. 개인정보자기결정권의 개념

개인정보자기결정권은 정보주체가 자신에 관한 정보에 대한 조사·수집·보관·처리·이용 등의 행위에 대해 동의 또는 거부할 수 있는 권리를 핵심적인 내용으로 하며, 경우에 따라 자신에 관한 정보에 관해 통제력을 행사

하기 위하여 자신의 개인정보에 접근하여 열람하거나 그 정보의 정정·삭제·차단·처리정지 등을 요구하는 권리로 현출될 수 있다고 해석된다.[3] 따라서 이와 같은 개인정보자기결정권의 핵심은 정보주체인 개인이 자신의 의사에 따라 개인정보에 대해 조사·수집·보관·처리·이용하는 것에 대해 스스로가 동의 또는 거부할 수 있는 권리를 보장하는 것이다. 즉, 개인정보에 대한 정보주체의 통제권의 인정이 개인정보자기결정권의 핵심이라고 볼 수 있다.

2. 개인정보자기결정권 영향 요인 모델 제안

2.1 연구 설계

본 논문에서는 빅데이터 분석을 통해 개인정보자기결정권 영향 요인 모델을 제안하고자 한다. 이에 개인정보자기결정권에 관한 토픽에 따른 키워드를 분류 및 경향을 파악하기 위해 토픽모델링 분석과 개인정보자기결정권 상에 내재된 개인정보자기결정권에 대한 영향 요인을 파악하기 위해 회귀분석을 제안하고자 한다.

2.2 연구 방법

연구방법은 다음과 같다. 우선 연구의 데이터는 네이버(Naver), 다음(Daum), 구글(Google) 등의 대표적인 포털사이트에서 개인정보자기결정권을 키워드로 한 단어들을 수집한다. 데이터 수집 도구로써 텍스트롬(Textom) 서비스를 이용하고자 하며, 분석 채널은 뉴스에 한정한다. 자료의 분석 기간은 3년으로 2018년 1월 1일부터 2020년 12월 31일까지 연구 대상의 기간으로 한정한다. 분석 데이터 정보는 다음(표 1.)과 같다.

표1. Research Data

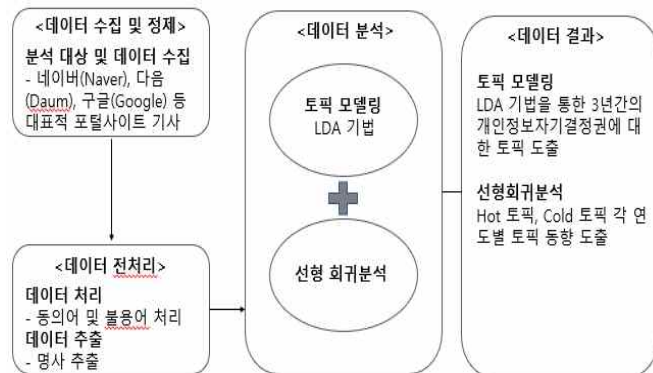
Target	Naver(news), Daum(news), Google(news)
Period	2018.1.1~2021.12.31
Tool	Textom, R Packages(3.6.1)
Key word	개인정보자기결정권

본 연구는 개인정보자기결정권 영향 요인 분석을 위해 토픽모델링과 선형회귀분석을 제안한다. 먼저 토픽모델링은 문서가 가지는 잠재적 의미와 주제들을 찾아 추론해 내는 방법으로, 이를 통해 대량의 문서들에서 출현할 확률이 높은 키워드를 도출하여 해당 토픽의 의미와 주제를 찾는 확률 모형이다.[4] 단어들의 집합들로 이루어진 문서의 경우 토픽의 집합이 확률적으로 내포되어 있으며 단어들 간의 분포에 대한 분석을 시행함으로써 구조를 파악해 문서 내의 주제를 파악한다. 이를 통해서 문서에서 중요하게 다루어지는 주제를 파악할 수 있는 것이다. 문서 내 토픽을 분석하는 기법으로 최근 주제별 단어의 확률을 기반으로 하는 LDA(Latent Dirichlet Allocation) 기법이 있으며 이는 기존 토픽모델링의 단점을 보완하여 이루어지는 분석 기법이라는 점에서 의미 있다. LDA 알고리즘 모델 기법은 관찰 되어지는 변수(observed variable)와 문서 내 감춰져 있는 주제 즉, 보이지 않는 변수(hidden variable)를 찾고 추론해준다.[4][5] 따라서 LDA 분석을 통해 전체 문서들의 집합을 기반으로 주제 도출, 문서 내에서 주제의 비율 그리고 주제에 포함된 단어의 확률을 알아낸다. 이는 R Packages(3.6.1)를 이용하는 것을 제안한다.

다음으로 토픽모델링 분석 결과를 기반한 개인정보자기결정권에 영향을 미치는 요인을 분석하기 위해 선형회귀분석을 이용한다. 토픽에 대한 전체 분석 기간 동안의 비중에 대한 추세를 분석하여 비중의 증가 추세를 보이는 개인정보자기결정권 Hot 토픽과 감소 추세를 보이는 개인정보자기결정권 Cold 토픽에 대한 분석을 실시한다. 각 토픽의 추세를 증가와 감소로 나눠서 판단은 기준은 다음과 같다. 분석 데이터는 3년 간의 연도별 토픽의 비중 평균값을 활용하며, 연도는 독립변수, 개인정보자기결정권에 관한 토픽에 대한 비중값을 종속변수로 활용한다. 이를 통해 개인정보자기결정권 영향 요인에 대한 회귀계수 값을 산출하여 추이를 분석한다. 회귀 계수 값이 양수의 형태를 가질 경우 Hot, 음수의 형태를 가지면 Cold로 분류하여 토픽의 추이를 도출한다.

따라서 2018년부터 2020년인 3년간의 포털사이트 기사 데이터를 바탕으로 토픽 모델링 중 LDA 기법을 통해 개인정보자기결정권에 관한 주요 토픽을 추출 후 분류한다. 이후 각각의 토픽에 대한 3년 간의 추이 분석을 통하여 개인정보자기결정권의 기간별 증가(Hot) 주제와 빈도가 줄어드는 감소(Cold) 주제를 밝혀 선형회귀분석을 제안한다. 분석 설계 모델은 다음(그림 1.)과 같다.

그림 1. 개인정보자기결정권 영향 요인 분석 설계 모델



III. 결론

본 논문에서는 개인정보자기결정권에 대한 영향 요인 모델을 LDA 기법을 통한 토픽모델링 및 선형 회귀분석을 결합하여 제안하였다. 이와 같은 제안 모델을 통하여 개인정보자기결정권에 대한 영향 요인을 도출해냄으

로써 개인정보자기결정권에 대한 영향 요인을 분석하고, 나아가 개인정보자기결정권에 대한 이해 및 향후 개인정보자기결정권의 개념 방향에 대해서 제시해 볼 수 있다는 점에서 의미가 있을 것이다.

다만 데이터에 있어 3년 간의 뉴스 데이터만으로 연구를 설계 했다는 점에서 한계가 있을 것이다. 이에 개인정보자기결정권에 대한 영향 요인을 파악하기 위한 키워드를 다양화하고 그 기간과 수집 범위를 확대한다면 보다 의미 있는 연구가 될 것으로 기대한다.

참 고 문 헌

- [1] 조소영, “코로나19 시대의 개인정보자기결정권의 보호-감염병 예방법상의 개인위치정보 공개규정을 중심으로-”, 헌법학연구 제27권 제2호, pp. 179-205, 2021
- [2] 권건보, “정보통신기술의 발달과 한국의 프라이버시권 관련 주요 쟁점”, 헌법재판연구, Vol.8, No.1, pp 121-153, 2021.
- [3] 권건보, “개인정보의 헌법적 기초와 과제”, 저스티스 통권 제144호, pp 17, 2014.
- [4] Bei, D. M. “Probabilistic topic models Communications of the ACM”, 55(4), pp. 77-84, 2012.
- [5] 이원성, 손소영, “공간빅데이터 연구 동향 파악을 위한 토픽모형 분석”, 대한산업공학학회지, 제41권, 제1회, pp. 64-73, 2015.
- [6] Bei, D. M, Ng, A Y., & Jordan, M. I., “Latent dirichlet allocation”, Journal of Machin Learning Research, 3, 993-1022, 2003.