데이터 비식별화 이슈와 기술 현황

정길준, *이흥노
광주과학기술원 전기전자컴퓨터공학부
e-mail : *jinpeg112@gist.ac.kr*, *heungno@gist.ac.kr*

Data De-identification Issues and Technology Status

Giljun Jung, *Heung-No Lee
School of Electrical Engineering and Computer Science
Gwangju Institute of Science and Technology (GIST)

Abstract

The use of Deep Learning (DL) is extending steadily for many companies to solve optimization problem in various areas. In addition, as the companies have more data, they can get more accurate solutions. Fortunately, the development of some industries such as Internet of Things (IoT) has become active recently, and each person makes lots of data. However, when the companies use data for analysis, they must be careful not to infringe the right of other people. We aim to summarize how data can be utilized without the infringement.

I. 서론

딥러닝(Deep Learning)은 최근 많은 분야에서 데이터를 분석하거나 문제를 풀어내는 방법 중 하나로 널리 쓰이고 있다. 딥러닝에는 학습을 위해 많은 데이터를 필요로 하고, 이 수요가 빅데이터(Big Data)라는 분야를 더욱 관심 받도록 만들었다[1].

미국의 IT 시장조사기관인 IDC가 2012년 발표한 보고서[2]에 따르면 디지털 데이터의 양은 2년마다 2배씩 증가하여 2020년에는 약 40ZB의 데이터가 쌓일 것으로 예측되었다. 빅데이터의 규모가 커질수록 다양한 이점도 있지만 데이터 사용에 더욱 주의를 기해야 한다. 특히, 개인정보를 포함하는 데이터는 일반 개인정보보호법에 의하면, 그 활용에 있어서 안전하게 활용될 수 있도록 개인정보 보호 조치를 필요로 한다[3].

본 논문에서는 데이터 비식별화 이슈와 관련하여 개인정보 보호를 위해 개인정보가 포함된 데이터를 비식별 조치하는 과정과 그 방법을 요약하고, 그 기술의 현황 및 한계와 빅데이터 산업이 활발하게 발전하기 위해서 어떠한 방향으로 나아가야 하는지를 제시한다.

II. 본론

데이터 비식별 조치는 한국인터넷진흥원(KISA)에 권고된 개인정보 비식별 조치 가이드라인[4]에 따르면, 다음과 같이 4가지 단계로 수행된다. 사용할 데이터에 개인정보가 들어있는지 확인하는 사전 검토 단계, 개인 식별을 가능하게 하는 요소를 처리하는 비식별 조치 단계, 비식별 조치가 적절하게 이루어졌는지 평가하는 적정성 평가 단계, 재식별을 방지하는 사후관리 단계가 있다.

1단계인 사전 검토 단계에서는 개인정보 보호법[5] 등을 참조하여 각 데이터가 개인정보를 포함하는지 확인하고, 개인정보를 포함하지 않는다면 빅데이터 분석 등에 바로 사용 가능하다. 하지만 데이터가 개인정보를 포함한다면 다음 단계로 넘어가야 한다.

2단계인 비식별 조치 단계에서는 식별자는 그 자체로, 속성자는 다른 자료와 결합 시 특정 개인을 식별할 수도 있기 때문에 식별자와 속성자에 대한 조치가 선행되어야 한다. 데이터에 포함된 식별자는 HIPAA 프라이버시 규칙[6]을 참고하여 삭제하고, 속성자도 원칙으로는 삭제해야 한다. 하지만, 반드시 사용하여야 하는 식별자 혹은 속성자라면 엄격한 비식별 조치 후에 활용할 수 있다. 비식별 조치 방법에는 가명처리(Pseudonymization), 총계처리(Aggregation), 데이터 삭제(Data Reduction), 데이터 범주화(Data Suppression), 데이터 마스킹(Data Masking) 기법 등이 있고 복합적으로 활용할 수 있다. 각 기법은 다양한 세부기술로 구현 가능하며, 데이터의 이용 목적에 맞춰 적절하게 선택하여 활용한다.

3단계인 적정성 평가 단계에서는 외부전문가를 포함한 평가단을 구성하여 비식별 조치된 데이터가 개인이 식별될 우려를 지니는지 엄격하게 평가한다. 적정성을 평가할 때 프라이버시 보호 모델의 k-익명성을 활용하며, 필요하다면 추가적인 평가모델을 복합적으로 사용한다. 부적정한 평가결과가 나온다면 개인정보처리자는 평가단의 의견을 반영하여 추가로 비식별 조치를 수행해야 한다. 평가가 종료된 정보집합물은 빅데이터 분석 등에 활용할 수 있다.

4단계인 사후관리 단계에서는 평가가 종료된 데이터라 하더라도 데이터 활용 과정에서 식별이 가능해지는 경우를 감시하고 막는다. 첫째로, 비식별 조치된 데이터가 유출되지 않도록 관리적, 기술적 보호조치가 필요하다. 둘째로, 비식별 조치된 데이터는 내부 혹은 외부 요인이 변화하면서 재식별 가능해질 수도 있기 때문에 주기적으로 그 가능성을 모니터링하여야 한다. 셋째로, 비식별 조치된 데이터를 제3의 기관에 제공하거나 위탁계약을 하는 경우에도 재식별 가능성이 있으므로 계약서에 위험관리에 대한 내용을 포함시킨다. 마지막으로, 재식별된 데이터는 즉시 활용을 멈추고 파기하거나 다시 비식별 조치를 거쳐 활용할 수 있다.

Ⅲ. 결론

4차 산업혁명 이후 수많은 데이터가 만들어지고 저장되고 있다. 하지만 그 모든 데이터가 유용하지는 못하며, 유용한 데이터라 할지라도 개인정보 등 활용 전에 처리해야 할 문제점이 존재한다. 데이터를 활용함에 있어 항상 조심하고, 주의를 기울여야 한다. 개인정보를 포함하는 데이터는 비식별 조치를 통해 빅데이터 분석 등에 이용할 수 있다.

하지만, 그 한계도 분명 존재한다. 개인정보 비식별 조치는 현행법에 맞춘 가이드라인에 의한 방법 중 하나이므로 추후 법안이 개정되거나 새로운 알고리즘 혹은 기술의 발달로 인하여 적정했던 비식별 조치도 부적절하게 여겨질 수도 있다. 그러므로, 활발한 빅데이터 산업으로 나아가려면 꾸준히 데이터의 개인정보에 대한 법안과 EU의 GDPR 등을 살피며, 더욱 효율적이고 효과적인 비식별화 기술을 개발할 필요가 있다.

Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIP) (NRF-2018R1A2A1A19018665)

참고문헌

[1] M. Najafabadi, F. Villanustre, T. Khoshgoftaar, N. Seliya, R. Wald and E. Muharemagic, "Deep learning applications and challenges in big data analytics", *Journal of Big Data*, 2015.

[2] John Gantz and David Reinsel, "THE DIGITAL UNIVERSE IN 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East", *IDC iVIEW: IDC Analyze the future*, 2012.

[3] https://gdpr-info.eu/

[4] "개인정보 비식별 조치 가이드라인", *kisa.or.kr*, 2018. [Online]. Available: https://www.kisa.or.kr/jsp/common/downloadAction.jsp?bno=282&dno=3&fseq=1 [Accessed: 22-May-2019]

[5] "국가법령정보센터 | 법령 > 본문 - 개인정보보호법", *Law.go.kr*, 2017. [Online]. Available: http://www.law.go.kr/%EB%B2%95%EB%A0%B9/%EA%B0%9C%EC%9D%B8%EC%A0%95%EB%B3%B4%EB%B3%B4%ED%98%B8%EB%B2%95 [Accessed: 23- May- 2019].

[6] "Summary of the HIPAA Privacy Rule", *HHS.gov*, 2002. [Online]. Available: https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html [Accessed: 23- May- 2019].