

개인정보 비식별 조치와 결합 개선을 통한 핀테크 시장 활성화

오원겸* · 박대우*

*호서대학교 벤처대학원

Fintech Industry Invigoration by the De-identification and Linkage Reform of Personal Information

Won-Gyeom Oh* · Dea-woo Park*

*Hoseo Graduate School of Venture

E-mail : wgoh7147@fsec.or.kr, prof_pdw@naver.com

요 약

국내 빅데이터 산업 활성화를 위해 2016년 6월 정부부처 합동으로 개인정보 비식별 조치 가이드라인을 발간하였다. 이 가이드라인은 현행 개인정보보호 관련 법령 하에서 개인정보를 보호하면서 사업자의 빅데이터 분석을 보다 활성화하고자 하려는 목적이 있다. 이 가이드라인은 국내의 엄격한 개인정보보호법 하에서 전자정부 선진국을 위한 정보환경을 조성하고자 하는 것이다. 다만, 이러한 가이드라인의 내용 중 비식별 조치 절차, 제도 및 정보집합물 결합 방법에 불합리한 부분이 있다면 빅데이터 산업 활성화의 걸림돌이 될 수 있다. 따라서 본 논문에서는 가이드라인의 비식별 조치 및 결합 방법 및 절차 중 개선이 필요하다고 생각되는 적정성 평가, 비식별 조치 지원 및 결합 임시대체키 생성 방법을 살펴보고 각각에 대해 해결 방안을 제시한다. 그리고 이러한 해결 방안이 어떻게 핀테크 산업 활성화에 도움이 되는지 알아본다.

ABSTRACT

The Korean government published the personal information de-identification guideline on June 2016, which were made by related government ministries. The guideline's objective is that the invigoration of Korean bigdata industry on personal information protection under the current laws. However, if there is some unreasonable method or process in the guideline, it can be an obstacle to bigdata analysis. This article will review the guideline to find defects in methods and processes of de-identification evaluation, de-identification support and data-linkage and then propose the best solutions to improve them. Lastly, this article will mention how these solutions can invigorate Fintech industry.

키워드

개인정보 비식별 조치, 적정성 평가, 정보집합물 결합, 핀테크

I. 서 론

2016년 6월 한국 정부는 개인정보 비식별 조치 가이드라인을 발간하였다. 이 가이드라인은 개인정보를 보호하면서 빅데이터 산업도 활성화하기 위한 방법의 하나이다.

국내의 개인정보보호법이 엄격하게 시행되고 있고 개인정보 이용에 따른 문제로서 비식별 조치 절차, 제도 및 정보집합물 결합 방법이 존재하지 않아 빅데이터 산업 활성화가 되지 않았다.

이러한 문제점을 해결하기 위해서 가이드라인

이 발간되었다. 가이드라인에 따라 비식별 조치한 개인정보를 개인정보가 아니라고 추정하고 있으며, 비식별 조치는 사전검토, 비식별 조치, 적정성 평가 및 사후 관리의 4단계로 이루어지고 있다.

비식별 조치 4단계 중 중요 단계를 살펴보면 2 단계인 비식별 조치는 다양한 비식별 조치 방법으로 비식별 조치를 수행할 수 있다. 3단계인 적정성 평가는 외부 전문가가 과반수 포함된 최소 3인 이상의 평가단이 k-익명성 등 프라이버시 보호 모델을 이용하여 사업자의 비식별 조치에 대해 적정하게 비식별 조치가 되었는지 평가한다.

가이드라인에서는 비식별 조치 지원 및 정보집합물 결합 지원을 위해 전문기관 제도를 도입하였다. 정보집합물 결합 시에는 2개 이상의 사업자가 각각 자신의 데이터를 비식별화 한 후 전문기관에 전달하고 전문기관은 해당 데이터의 결합을 수행 후 결합된 데이터를 이용할 사업자에 전달한다. 결합된 데이터를 받은 사업자는 적정성 평가를 수행하여 해당 데이터의 비식별 수준을 점검하여야 한다.

본 연구에서는 개인정보 비식별에 관련된 연구를 살펴본 후 가이드라인의 비식별 조치 절차, 제도 중 개선이 필요한 부분을 조사한다. 개선이 필요한 부분이 도출되면 그에 대한 최적의 해결 방안을 제시하고자 한다.

II. 관련 연구

2.1 빅데이터 분석

'12년에 가트너는 Volume(대용량), Velocity(실시간 변경), Variety(비정형)의 3Vs를 가진 데이터를 빅데이터라 하였다[1]. 예측적인 데이터 분석(Predictive Data Analytics)은 이러한 빅데이터를 이용하는 대표적 과학으로 데이터에서 특정 패턴을 찾아내고 예측하는 것이다[2]. 특정 패턴 검색을 자동으로 하는 머신러닝이 예측적인 데이터 분석의 대표적인 방법이라 할 수 있다.

김원필은 정보보안 트렌드를 분석하기 위해 IEEE 논문에 대한 정보를 이용하였고 빅데이터 등 전망이 밝은 기술들을 찾아내었다[3].

최도현의 1명은 Naive Bayes Algorithm 등을 이용한 빅데이터 분석으로 사용자의 거래 성향을 분석하였다[4].

이와 같은 연구에서 알 수 있듯이 빅데이터 분석은 무의미해 보이는 데이터에서 새로운 정보를 찾아내는 것으로 미국 등 선진국에서는 이러한 빅데이터 분석으로 새로운 수익 및 산업을 창출해 내고 있다.

2.2 개인정보보호법

국내의 개인정보 보호 관련 법령은 전세계에서 가장 엄격하다고 볼 수 있다. 차상욱은 빅데이터 산업을 활성화하기 위해 선진국의 개인정보보호 제도 변화를 참고하여 개인정보의 활용과 보호의 적절한 균형점을 찾아야 한다고 하였다[5].

2.3 비식별 조치 및 프라이버시 모델

김동국의 1명은 개인정보 관련 법률 기반 개인 식별정보 파악, 빅데이터 활용 단계별 비식별화 적용 방법 및 비식별화 기술 등에 대해서 연구하였다[6].

개인정보를 비식별 조치하는 방법은 여러 가지가 있으며 대표적으로 이용되는 방법은 데이터

범주화와 마스킹 등이다. 이러한 비식별 조치가 적절하게 되었는지를 검증하는 프라이버시 모델로는 k-익명성, l-다양성, t-근접성 등이 있으며 Sweeney의 k-익명성이 가장 보편적으로 이용되고 있다[7]. k-익명성은 개인정보가 포함된 데이터에서 준식별자를 비식별화 하여 데이터내 같은 준식별자를 갖는 레코드가 최소 k개 이상 존재하게 하는 방법이다. 이러한 k-익명성을 통한 검증은 개인정보가 포함된 데이터에 대해 특정 기준 이상의 개인정보 보호 수준을 보장하게 한다.

국내에서는 강진영외 1명이 k-익명성과 l-다양성을 만족시키는 구체적인 알고리즘을 연구하였다[8].

비식별 조치에 관한 연구는 국내뿐만 아니라 해외에서도 많지 않다. 왜냐하면, 데이터 분석 선진국들이 각국의 빅데이터 산업 활성화를 위해 빅데이터 분석에 중점을 두고 있기 때문이라고 생각된다. 예를 들면, 미국의 경우 의료 데이터 외에는 비식별 조치를 할 의무가 없다.

III. 비식별 조치와 결합에 관한 분석

3.1 적정성 평가 수행 시기 분석

가이드라인에서는 비식별 조치 시 적정성 평가를 비식별 조치시마다 수행하도록 하고 있다. 하지만 빅데이터 분석 프로세스(CRISP-DM: Shearer, 2000)에 따르면 빅데이터 분석 시 필요에 따라 수시로 분석에 필요한 데이터를 수집하여야 한다고 하였다. 이러한 빅데이터 분석 프로세스에서 외부전문가를 2명 이상 포함한 평가를 최소 2회 수행해야 하는 가이드라인의 적정성 평가 방법은 과도하다고 할 수 있다.

3.2 전문기관의 비식별 조치 지원 분석

개인정보 비식별 조치의 경우 대기업, 중견기업은 수행하는데 큰 무리가 없을 수 있다. 하지만 선진국의 사례를 볼 때 페이팔(Paypal), 페이스북(Facebook) 등 빅데이터 분석이 사업 비중에서 큰 핀테크 업체는 초기에 소규모 벤처기업으로 출발하였고 이러한 핀테크 업체들에게 가이드라인의 비식별 조치는 본연의 업무인 빅데이터 분석보다 더 큰 업무가 될 수 있다. 이러한 문제점을 해소하기 위해 가이드라인에서는 전문기관에 비식별 조치 지원을 할 수 있게 하였으나 개인정보보호법 상 전문기관에게 개인정보를 처리할 법적 근거를 부여하지 않아 지원의 한계가 분명히 존재한다.

3.3 정보집합물 결합용 임시대체키 생성 문제점

가이드라인의 핵심은 정보집합물 결합을 통해 빅데이터 산업 활성화의 도모이다. 이러한 정보집합물 결합 시에는 2개 기관의 비식별된 데이터를

결합하기 위한 매칭키인 임시대체키를 생성하여야 한다. 하지만, 임시대체키 생성 시 법적인 문제로 고유식별정보 또는 이와 유사한 공개된 고유한 값을 사용하지 못하고 성명, 전화번호, 우편번호 등을 이용해 키를 생성해야 한다. 이와 같이 키를 생성할 경우 2개 기관의 정보가 달라 결합이 제대로 되지 않을 수 있으나 전문기관에 결합할 데이터를 보내 다수의 매칭이 실패한 결합결과를 확인하는 것 외에 현재로서는 뾰족한 해결 방법이 없다.

IV. 비식별 조치와 결합에 관한 개선 방안

4.1 빅데이터 분석 프로세스에 기반한 적정성 평가 수행 주기 변경

빅데이터 분석 프로세스에 따라 분석 할 경우 수시로 데이터 수집을 하여야 한다. 그러므로 데이터 스키마가 유사하고 데이터 변경이 자주 발생하는 데이터를 분석하는 경우 가이드라인의 적정성 평가 수행 주기를 데이터 분석 시작 시점과 모델 개발이 완료된 종료 시점 2회만 수행하여 빅데이터 분석의 효율성을 향상하여야 한다. 모델 개발 중에는 실제 개인정보가 재식별이 된다 할지라도 그러한 재식별에 의해 개인의 프라이버시를 침해할 가능성이 거의 없으며 개인정보보호 위반은 모델 개발 후 실제 해당 모델에 따라 마케팅 등의 업무 수행 시 발생할 것이다. 그러므로 빅데이터 분석을 저해하는 과도한 적정성 평가 수행 횟수를 최대한 줄여 빅데이터 분석의 부담을 줄여줘야 빅데이터 분석의 가시적인 성과가 보다 용이하게 나올 수 있을 것이다.

다만, 빅데이터 분석 시작 시점에서의 적정성 평가로 도출된 비식별 조치 수준으로 모델 개발이 힘들다고 판단될 시에는 적정성 평가를 재수행하여 비식별 조치 수준을 낮출 수 있는 방안이 있는지 검토하여야 할 것이다.

4.2 전문기관의 비식별 조치 지원 강화 및 데이터 브로커 양성

전문기관이 벤처기업 등 소규모 회사의 비식별 조치를 지원하기 위해 개인정보보호법 등 관련 법률로부터 개인정보를 제공받아 처리할 수 있는 권한을 받아야 비식별 조치가 원활이 이뤄질 수 있다. 이러한 경우 전문기관에서 개인정보보호법 위반 사례가 발생해 프라이버시 침해가 될 수 있다. 하지만, 이러한 침해 가능성에 대한 대책도 법률에 명시하고 정부에서 정기적으로 점검을 한다면 충분히 해결될 수 있을 것으로 판단된다.

또한, 이러한 전문기관의 적극적인 지원은 향후 데이터 활용 선진국과 같이 데이터 브로커의 밀거름이 될 수 있다. 액시엄사(Acxiom)와 같은 데이터 브로커는 데이터 활용의 기반시설이라고 할 수 있다. 이러한 데이터 브로커 없이 데이터

융합을 성공한 사례는 현재까지 국내외에 거의 없다.

그러므로 현재의 공공기관들로 구성된 전문기관에게 적극적인 비식별 조치 지원 권한을 우선 부여하고 일정 기간 동안 빅데이터 활성화 효과 및 개인의 프라이버시 침해요소를 모니터링 한 후 민간 전문기관에게도 비식별 조치 지원 권한을 부여해 국내 빅데이터 산업을 활성화 하여야 할 것이다.

4.3 정보집합물 결합용 임시대체키 생성 시 전문기관 지원

정보집합물 결합을 위한 임시대체키 생성 시 최대한의 매칭을 위해 전문기관에게 2개 기관의 데이터를 검사하여 매칭이 제대로 이뤄질 수 있는지 여부를 판단할 수 있는 법적 권한을 부여하여야 정보집합물 결합이 제대로 될 수 있다. 임시대체키 생성 지원 시에는 2개 기관에서 민감정보를 제외한 성명, 전화번호, 우편번호 등 분석 대상이 되는 준식별자를 최대한 전송하여 전문기관이 어떠한 준식별자를 이용하여야 매칭이 최대한 잘 이뤄질 수 있는지 판단하게 하여야 한다.

다만, 이러한 방법은 고유식별정보를 이용한 결합과 비교하여 분명 한계가 존재하므로 향후 미국 액시엄사(Acxiom)의 AbiliTec Key와 같이 데이터 결합을 위한 특수 매칭키를 전문기관별로 만들 수 있게 하는 것이 필요할 것이다.

4.4 핀테크 시장 활성화 방안

앞서 살펴 본 빅데이터 분석 프로세스에 기반한 적정성 평가 수행 주기 변경, 전문기관의 비식별 조치 지원 강화 및 데이터 브로커 양성 및 정보집합물 결합용 임시대체키 생성 시 전문기관 지원은 표1과 같이 핀테크 산업 활성화에 도움이 될 수 있다.

Table. 1 Fintech industry invigoration effects by each reform

Reform details	Invigoration effects
De-identification evaluation time change based on the bigdata analysis process	- Save Fintech companies' evaluation cost, time and invest more time on bigdata analysis - Reduce burden on collection of updated or slightly different data from initial data
Strengthen specialized agency' s de-identification support	- Save Fintech companies' de-identification cost - Analyze better utilization de-identified data
Specialized agency' s support on temporary key creation for data-linkage	- Make perfect matching of 2+ data possible - Produce better Fintech's analysis result through linkage

빅데이터 분석으로 수익을 내야 하는 소규모의 영세한 핀테크 업체에게 비용 절감 및 조금이라도 활용성이 높은 데이터는 매우 중요하다. 본 논문의 개선 방안은 표1과 같이 핀테크 업체의 비식별 조치 및 결합 시 도움이 주어 핀테크 시장을 활성화할 수 있다.

V. 결 론

정부의 빅데이터 산업 활성화를 위한 개인정보 비식별 조치 가이드라인이 국내 빅데이터 산업에서 중요한 발전 계기를 마련한 것은 분명하다. 다만, 시행초기인 만큼 개선의 여지가 분명 존재하며 본 연구에서는 그러한 개선이 필요한 부분을 분석하고 이를 해소할 방안을 제시하였다. 또한, 이러한 개선 방안이 어떻게 핀테크 산업 활성화에 효과가 있는지도 살펴 보았다.

향후 가이드라인에 따라 비식별 조치 및 정보 집합물 결합을 수행한 사례가 만들어지면 가이드라인의 방법, 절차에서 추가적으로 개선이 필요한 부분이 나타날 것이며 이러한 문제들을 해결할 연구가 이뤄져야 할 것이다.

참고문헌

- [1] Peter Zadrozny and Ragha Kodali, *Big Data Analytics Using Splunk*, 1st ed, New York, NY: Apress, 2013.
- [2] John D. Kelleher, Brian Mac Namee and Aoife D'Arcy, *Fundamentals of Machine Learning for Predictive Analytics*, 1st ed, Cambridge, MA: The MIT Press, 2015.
- [3] W. P. Kim, "Analysis of Global Research Trend on Information Security," *J. Korea Inst. Inf. Commun. Eng.*, vol. 19, no. 5, pp.1110-1116, May 2015.
- [4] D. H. Choi et al., "The Application Method of Machine Learning for Analyzing User Transaction Tendency in Big Data environment," *J. Korea Inst. Inf. Commun. Eng.*, vol. 19, no. 10, pp.2232-2240, Oct. 2015.
- [5] S. Y. Cha, "A study on collision and tasks between personal information protection legislation and the use of big data" *Hang Yang Law Review*, vol. 27, no. 1, pp.315-359, May 2016.
- [6] D. K. Kim and H. Lee, "The De-identification Trend of Personal Information based on Big Data" *Review of Korean Society for Internet Information*, Vol 16, no 2, pp.15-22, Dec. 2015.
- [7] L. Sweeney. "k-anonymity: a model for protecting privacy" *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, vol 10, no 5, pp.557-570, 2002.

- [8] J. Y. Kang, "A study on the K-anonymity and L-diversity for data Privacy", MS. D. dissertation, Kookmin University, Republic of Korea, 2008