개인정보 비식별화 역추적을 위한 머신러닝 활용에 대한 연구 제시

A Study on the Utilization of Machine Learning for De-identification Backtracking

저자
(Authors)

오예원, 김기천
Yewon Oh, Keecheon Kim

출처
(Source)

한국통신학회 학술대회논문집 , 2018.1, 83-84(2 pages)
Proceedings of Symposium of the Korean Institute of communications and Information Sciences , 2018.1, 83-84(2 pages)

발행처
(Publisher)

한국통신학회
Korea Institute Of Communication Sciences

URL

http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE07368563

APA Style

오예원, 김기천 (2018). 개인정보 비식별화 역추적을 위한 머신러닝 활용에 대한 연구 제시. 한국통신학회 학술대회논문집, 83-84

한국통신학회 2018년도 동계종합학술발표회

개인정보 비식별화 역추적을 위한 머신러닝 활용에 대한 연구 제시

오예원 (건국대학교), *김기천 (건국대학교)

iamoyw@konkuk.ac.kr, *kckim@konkuk.ac.kr

A Study on the Utilization of Machine Learning for De-identification Backtracking

Yewon Oh(Konkuk Univ.), *Keecheon Kim(Konkuk Univ.)

요 약

정보화 사회에서 데이터는 아주 중요한 요소가 되며, 동시에 그 정보를 수집, 처리, 보호하는 과정이 사회의 중요한 이슈가 되고 있다. 데이터가 방대하게 많아지는 것을 빅데이터(Big data)라고 부르며, 빅데이터를 활용 속에는 개인의 관련된 정보 수집도 필요하다. 하지만, 개인정보에는 민감정보가 포함되어 있으며, 개인은 프라이버시를 보호 받을 자격이 있기 때문에, 개인정보를 보호하고, 이를 비식별처리하는 것은 4차 산업혁명의 중요한 요소가 된다. 본 논문에서는 먼저 비식별화가 무엇이고, 비식별의 기술적인 방법을 이해하고, 우리나라의 개인정보 비식별 조치 관리의 취약점을 다시 한 번 되짚어보고 있다. 또한, k-익명성 알고리즘 중 하나인 Incognito를 깊이우선탐색(depth first search)을 토대로 머신러닝 알고리즘을 이용하여 역추적을 하고 이에 대한 위험성이 존재하는 지를 연구를 제시한다.

Ⅰ. 서 론

4차 산업 혁명이라는 단어와 함께 사물인터넷(Internet of things), 빅데이터(Big data), 그리고 인공지능(Artificial intelligence) 같은 단어들이 화두에 오르고 있다. 이 기술들을 처리하기 위해서는 방대한 데이터들이 필요하며, 그에 따른 해킹과 유출 및 오·남용 사고들에 대한 보안 이슈도 주목받고 있다. 특히, 정보통신의 발전으로 개인의 관련된 정보들의 수집·활용의 욕구가 증대되고 있다. 개인정보에는 민감정보를 포함하고 있기 때문에 개인정보 데이터를 비식별처리한 후에 공개하거나 배포해야 한다. 하지만, 비식별 처리된 정보라고 해서 완벽하게 안전하다고 하기는 힘들다. 왜냐하면, 다른 정보들과 결합하여, 재식별이 가능한 경우도 존재하기 때문이다.

본 논문에서는 개인정보보호에 이용되고 있는 비식별화를 조금 더 안전하게 강화하기 위해 개인정보 비식별 조치의 취약점을 다시 한 번 되짚어 본다. 더 나아가 재식별이 아닌 역추적에 대한 위험성이 존재하는 지를 머신러닝 알고리즘을 이용하여 알아보고 위험 관리의 중요성을 인식하기 위한 연구이다.

본 논문의 구성은 다음과 같다. 2장에서는 개인정보 비식별조치와 기술적 프라이버시 보호 모델 중 가장 많이 사용되고 있는 K-익명성에 대한 관련연구를 분석하고 있다. 3장에서는 개인정보 비식별화 역추적에 대한 위험성에 대해 기술하고, 마지막 4장에서 결론은 맺고 있다.

Ⅱ. 관련 연구

2.1 비식별화(De-identification)

비식별화 란, 수집(Collect), 사용(Use), 저장(Archive), 또는 공유(Share)되는 데이터로부터 개인을 식별하지 못하게 조치하는 방법이다. 미국 국립표준기술연구소인 NIST의 개인정보 비식별화 내부 보고서에 의하면 비식별화의 궁극적인 목표는 데이터를 어떤 개인과도 연결시킬 수 없도록 만드는 것이라고 정의내리고 있다[2]. 개인정보 비식별 처리방법으로는 가명처리, 총계처리, 데이터 값 삭제 및 데이터 마스킹 처리 등 18가지 다양한 방법들이 존재하고 있다. 이에 따른 기대효과로는 개인정보보호법 준수 하에 다양한 데이터를 필요한 조직에서 사용가능하며, 개인의 사생활도 보호받을 수 있다. 더 나아가 빅 데이터의 활용 및 도입 확산에 중요한 데이터 수집이 용이해 질 수 있다. 하지만, 우리는 비식별화가 이미 공개된 데이터 혹은 앞으로 추가될 데이터와 결합하여 재식별화 될 수 있다는 가능성과 이에 따라 발생 가능한 개인정보 유출 및 사건·사고를 방관해서는 안 되며, 대책을 마련해야 한다.

2.2 프라이버시 보호 모델

비식별화에 대한 기술로써 프라이버시 보호 모델은 K-익명성(k-anonymity), L-다양성(l-diversity), T-근접성(t-closeness), 그리고 ε-차분 프라이버시(ε-differential privacy) 이렇게 4가지가 존재한다. 이 모델들은 비식별화 정도에 대한 계량분석 방법으로 많이 쓰이며, 적정성 평가 단계에서 평가 대상이 되기도 한다. 본 관련연구에서는 k-익명성(k-anonymity)에 대한 취약점을 더 자세하게 살펴보고 있다.

A. K-익명성

공개된 데이터에 대한 연결공격(Linkage attack)을 방어하기 위해 제안된 프라이버시 보호 모델로 2002년에 L.Sweeney가 제안한 모델이다[3]. 여기서 연결공격이란, 이미 공개된 데이터들이 결합을 통해서 개인의 민감 정보가 노출되는 경우를 말한다. 다시 말해, k-익명성은 주어진 데이터에서 같은 값이 적어도 k개 이상 존재하도록 수정하고, 다른 레코드와 구별되지 않는 적어도 k-1개 이상의 레코드를 갖도록하여 프라이버시를 보호하는 방법이다[1]. 그러나 데이터가 추가·삭제되는 경우, k값이 커진다면, 익명성은 높아질 수 있으나, 데이터의 정보가 손실되는 위험성이 존재하기 때문에 데이터의 용도 및 성격을 잘 구분할 필요가 있다.

다음 표1~3은 k-익명성의 예시이다[1]. (표1)공개된 의료데이터에서 "13053"의 지역코드를 갖은 28살 남자가 전립선염을 앓고 있다는 정보를 얻었다면, 선거인명부 데이터(표2)를 통해 같은 지역코드와 연령, 성별을 연결시켜 보았을 때, '김민준'이 전립선염을 앓고 있는 환자라는 가능성이 매우 커진다.

표 1 공개 의료데이터

구분	지역코드	연령	성별	질병
1	13053	28	남	전립선염
2	13068	21	남	전립선염
3	13068	29	여	고혈압
4	13053	23	남	고혈압

표 2 선거인명부

구분	이름	지역코드	연령	성별
1	김민준	13053	28	남
2	박지훈	13068	21	남
3	이지민	13068	29	여
4	최현우	13053	23	남

이러한 사례가 연결공격이며, 이를 보안하고자 표3는 4개(k값)의 항목을 비식 처리한 것을 볼 수 있다.

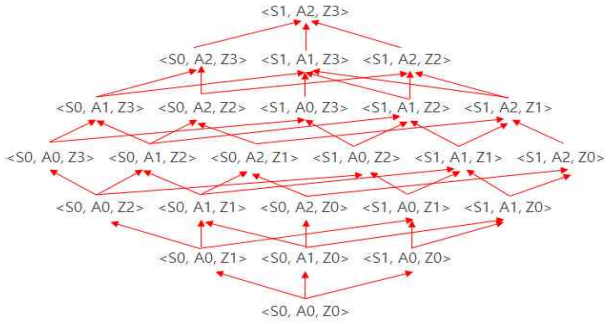
표 3 선거인명부

구분	지역 코드	연령	성별	질병
1	130**	〈30	*	전입선염
2	130**	〈30	*	전립선염
3	130**	〈30	*	고혈압
4	130**	〈30	*	고혈압

Ⅲ. 제안사항

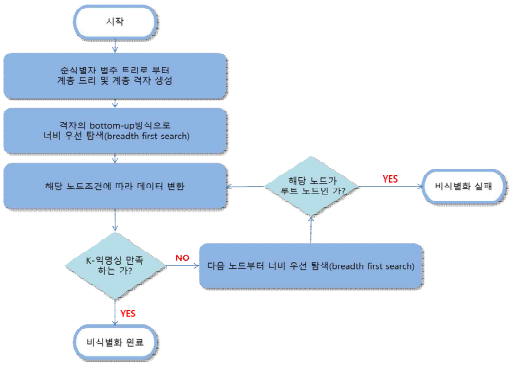
본 논문에서는 이러한 개인정보 비식별 조치에 대한 취약점을 분석하고 앞으로 해야 하는 위험 관리에 대해서 제시한다. 우리나라는 『개인정보 보호 비식별 조치 가이드라인을 보고를 했고, 비식별를 하기 위해서는 적정성 평가를 해야 한다고 제시하고 있다. 하지만, 이 적정성 평가는 k-익명성에만 국한 되어있으며, 다른 프라이버시 모델이나, 앞으로 나올 다른 기술에는 적용이 불가능하도록 되어있다. 또한 적정성 평가가 된 비식별 데이터임에도 불구하고, 다른 데이터와 결합하여 재식별의 위험에 노출될 가능성이 있다. 앞으로 빅데이터의 발전에 따라 더욱 안전한 개인정보 보호 비식별 조치를 위해서는 단순히 가이드라인이 아닌, 좀 더 체계적이고 구체적인 법률과 표준화가 필요한 실정이다.

앞서 말한 재식별에 대한 위험 대응방안은 많은 사례들이 연구되고 있다. 본 논문에서는 재식별이 아닌 k-익명성 알고리즘을 역추적을 하면 위험성이 존재하는 지에 대한 연구방향을 제시한다. k-익명성 알고리즘의 가장 대표적으로는 Incognito 알고리즘이 있다. 이 알고리즘은 데이터를 일반화(generalization)방법으로 준식별자 값들을 변환하여 추론하는 방식이며, 일반화는 계층트리(hierarchial tree) 자료구조를 이용하고 이에 따른 [그림 1]준식별자 계층 격자(hierarchial lattice)로 표현한다.



〔그림1〕 계층 격자(hierarchial lattice) 예시

이에 따라 탐색하는 순서도는 [그림2]을 따른다.



〔그림2〕 계층격자 탐색 순서도

계층격자 탐색에서 원본데이터인 루트노드부터 시작하여 너비 우선 탐색(breadth first search)을 했다면, 역추적은 비식별화 된 데이터부터 시작하여 거꾸로 깊이 우선 탐색(depth first search)을 수행한다. 깊이 우선 탐색을 하면서 노드의 조건에 따라 데이터를 변환하고, 변환된 데이터가 루트노드인지를 확인한다. 해당 데이터가 루트노드일 경우, 이 데이터는 원본데이터(Row data)이며, 역추적을 통해서 비식별된 데이터에서 원본 데이터를 찾아낼 수 있게 된다.

이 때, 각 노드마다 비식별한 방법들이 다양하게 있을 수 있기 때문에 데이터를 변환하는 과정에서 각 노드가 사용한 비식별 조치 방법에 따른 머신러닝 모델링을 구현하여, 이용한다. 예를 들어, 데이터 범주화 기법을 사용한 노드에는 클러스터링을 이용하여 모델링을 하고, 데이터 마스킹 처리기법을 사용한 노드에는 의사결정트리 알고리즘을 통해 최대한 효과적으로 데이터 역추적을 한다. k-익명성의 구조가 간단하기 때문에 지도학습방법을 이용하여 간단하게 처리할 수 있을 것으로 기대되며, 학습된 기계를 통해 다른 프라이버시 보호 모델의 역추적까지 가능할 것이라고 예상된다.

Ⅳ. 결론

개인정보보호를 위한 개인정보 비식별 조치에 대한 전반적인 취약점과 비식별된 데이터 역추적이 머신러닝기반 알고리즘을 통해 했을 때, 새로운 위험 요소가 될 수 있는지에 대한 가능성을 연구하였다. 전반적으로 우리나라에서는 개인정보 비식별 조치에 대한 가이드라인만 있고, 표준화나 구체적인 법은 없기 때문에 개인정보를 안전하게 보호하기 어렵다.

본 논문에서는 머신러닝을 이용한 개인정보 비식별화 데이터를 원본데이터로 역추적하여 새로운 개인정보 비식별조치의 취약점을 발견하는데 제시를 하고 있다. 향후 연구로는 k-익명성뿐만 아니라 다른 프라이버시 보호 모델(l-다양성,t-근접성)알고리즘에 대한 역추적도 머신러닝을 통해서 연구해보고, 예상되는 취약점에 대한 보안방법까지 도출해 낼 것이다.

ACKNOWLEDGMENT

이 연구는 2017년도 산업통상자원부 및 산업기술평가관리원(KEIT) 연구비 지원에 의한 연구임('10078261')

참 고 문 헌

[1] 미래창조과학[2016], 개인정보 비식별 조치 가이드라인

[2] NIST Special Publication 800-1882(2ndDRAFT),De-identifying government datasets,2016

[3] 개인정보보호위원회, '개인정보의 비식별화 처리가 개인정보 보호에 미치는 영향에 관한 연구", 2015