

영상 데이터 익명화를 위한 고려사항

성민경, 이강원, 한주연, 심기창*, 김동례*, 조민용**, 김태훈**

한국정보통신기술협회, *(주)이지서티, **(주)딥핑소스

[mksung, blong116, hanjy]@tta.or.kr, [gcshim, drkim]@easycerti.com,

[minyong.cho, pete.kim]@deepingsource.io

Issues for anonymizing image/video data

Sung Min Kyoung, Lee Kangwon, Han Ju Yeun, Shim Gi Chang*, Kim Dong Rye*,

Minyong Cho**, Tae-hoon Kim**

Telecommunications Technology Association, EASYCERTI CO., Ltd.*, Deeping Source Inc.**

요 약

데이터 활용이 늘어남에 따라 개인정보 보호를 위해 익명화된 데이터 사용 대한 관심도 증폭되고 있다. 특히, 다양한 촬영 기기에서 생성된 영상 데이터 활용 수요가 많은데, 현재 발간된 개인정보 가명·익명처리 가이드라인/안내서는 대부분 테이블 데이터로 대표되는 정형데이터를 대상으로 한 내용만 다루고 있다. 영상 데이터는 정형데이터와 다른 특징을 가지고 있으므로 익명화 시 추가로 고려해야 하는 사항이 필요하다. 본 논문은 영상 데이터 익명화 시 고려해야 하는 사항을 기존 발간된 가명·익명처리 가이드라인을 기반으로 분석 및 제시하였다.

I. 서 론

영상, 이미지 등 비정형데이터는 자율주행, 객체인식 등 인공지능 학습 용도로 널리 활용되고 있다[1]. 데이터의 중요성 강조됨에 따라 정부 차원에서 인공지능 학습용 데이터 구축사업을 진행하고 있고 있으며[2], 이 과정에서 데이터에 포함된 개인정보보호를 위해 비식별처리를 수행하도록 하고 있다. 비식별처리를 위해 개인정보보호위원회를 비롯하여 각 중앙부처에서는 가명·익명정보처리 가이드라인 등 비식별 처리를 위한 지침을 마련하고 있으나[3-7] 대부분 정형데이터에 초점이 맞춰진 내용이며, 영상 등 비정형 데이터에 비식별처리 절차와 기준은 모호한 상황이다. 본 논문에서는 기존 가이드라인의 절차를 기반으로 영상 데이터를 익명화 시 고려해야 하는 사항을 분석한다.

II. 본론

1. 기존 가이드라인

각 정부 부처는 가명·익명정보 처리를 위해 가이드라인 및 안내서를 발간하였으며, 현황은 표1과 같다.

문서명	발간부서	발간일
가명정보 처리 가이드라인	개인정보보호위원회	2022.04.
교육분야 가명·익명정보 처리 가이드라인	교육부	2022.07.
보건의료 데이터 활용 가이드라인 개정(안)	보건복지부	2022.05.
금융분야 가명·익명처리 안내서	금융감독원	2022.01.
공공분야 가명정보 제공 실무안내서	행정안전부	2021.01.

표 1 . 부처별 가명·익명정보 처리 가이드라인/안내서

기존 가이드라인의 가명·익명처리 등 비식별 처리 절차는 표2와 같다.

개인정보 보호위원회	교육부	보건복지부	금융감독원	행정안전부
1. 사전준비	1. 사전준비	1. 사전준비	1. 사전준비	1. 사전준비
-	1.1 사전협의	-	-	1.1 신청서 접수
-	1.2 계획수립 및 운영	-	-	1.2 적합성 검토
-	1.3 정보집합물 구성	-	-	1.3 적합성검토결과 통보
-	1.4 관리체계점검 및 준비	-	-	-
2. 가명처리	2. 위험성 검토	2. 가명처리	2. 가명처리	2. 가명처리 단계
2.1 대상선정	2.1 가명처리 대상항목 및 명세서 검증	2.1 대상선정	2.1 위험도 측정	2.1 대상선정
2.2 위험도 측정	2.2 위험성 검토	2.2 위험성 검토	2.2 가명처리 수준측정	2.2 위험도 측정
2.3 가명처리 수준정의	2.3 검토결과보고서 작성	2.3 가명처리 방법 및 수준정의	2.3 가명처리	2.3 가명처리수준정의
2.4 가명처리	-	2.4 가명처리	-	2.4 가명처리수행
3. 적정성검토 및 추가처리	3. 가명처리	3. 적정성 검토 및 추가처리	3. 적정성 검토	3. 적정성검토 및 추가처리
-	3.1 가명처리수준 정의표 작성	-	-	3.1 적정성 검토
-	3.2 가명처리 수행	-	-	3.2 추가처리
-	3.3 특이정보 검증	-	-	-
4. 활용 및 사후관리	4. 적정성 검토	4. 사후관리	4. 활용 및 사후관리	4. 활용 및 사후관리
-	5. 안전한 관리	-	-	-

표 2 . 가이드라인별 가명·익명처리 절차

부처별 가이드라인의 세부절차는 도메인마다 차이가 있으나 공통적인 절차는 다음과 같이 요약할 수 있다(그림1).



그림 1. 가명/익명처리 절차 요약도

2. 영상 데이터의 특징

영상 데이터는 객체인식(얼굴인식, 문자인식 등), 상황감지(자율주행, 이상행위 탐지 등), 모션인식(신체 동작인식, 재활보조 기술 등) 등 인공지능 기술 활용을 위해 다양하게 이용되고 있다[8]. 그러나 영상에는 개인의 민감한 정보가 포함될 수 있기 때문에 익명화 등을 통해 프라이버시 침해 문제를 해결해야 한다. 영상 데이터 익명화는 영상 내에서 익명처리 할 영역을 탐지하는 기술과 탐지된 부분을 익명화하는 기술로 구성된다. 또한, 익명화된 데이터의 품질 유지를 위해 유용성과 안전성에 대한 기준설정도 필요하다. 예를 들어, 사람의 얼굴을 익명화할 때, 얼굴 윤곽 이상으로 너무 크게 익명화 영역을 탐지하면 필요 이상으로 익명화가 수행되어 데이터 유용성이 떨어진다. 또한, 얼굴 익명화 시 사람이 봤을 때, 누군지 알아볼 수 있게 익명화를 수행하거나 역가공을 통해 원본이 드러날 수 있다면 안전성이 떨어진다.

3. 가이드라인/안내서 기반 영상 데이터 익명화 시 고려사항

본 절에서는 기존 가이드라인/안내서의 절차를 기반으로 하여 영상 데이터 익명화를 위한 고려사항을 기술한다. 영상 데이터 익명화는 원본 데이터와 비식별 대상 데이터 사이의 맵핑 정보에 대한 내용이 없으므로 영상 데이터에 대한 내용은 익명처리로 기술한다. 자세한 내용은 표3으로 정리하였다.

단계	고려사항
① 대상선정	<ul style="list-style-type: none"> 정형 데이터의 목적(통계작성, 과학적 연구, 공익적 기록보존)이 아닌 인공지능 학습으로 활용될 대상 정형 데이터와 다르게 항목(이름, 주민번호, 주소 등)으로 구분되지 않고 목적에 맞는 위치, 시간대, 수집 환경 등을 고려하여 대상이 선정되어야 함
② 위험도 측정	<ul style="list-style-type: none"> 데이터가 수집된 주변환경, 이미지 자체에 대한 식별 위험성 요소가 고려되어야 함 비정형 데이터의 위치정보의 식별 가능성이 고려되어야 함
③ 익명처리 수준정의	<ul style="list-style-type: none"> 비식별대상의 부가정보 노출 유무 정의 필요 (비식별대상이 얼굴인 경우 성별, 연령 등이 부가 정보가 될 수 있으며, 번호판인 경우 차종, 사업자 유무 등이 부가정보가 될 수 있음) 비정형 데이터의 위치 노출로 인해 재식별 리스크가 높다고 판단되는 경우, 해당 이미지 데이터를 삭제하거나 해당 영역 비식별 처리 필요
④ 익명처리 수행	<ul style="list-style-type: none"> 정형 데이터처럼 특정 규칙을 통한 일괄적인 익명처리가 불가능하기 때문에 별도의 프로그램을 통해 익명처리 수행 필요
⑤ 적정성 검토	<ul style="list-style-type: none"> 원본 데이터와 익명처리된 데이터의 비교를 통한 유용성 분석 필요 프로그램과 수작업을 통한 안전성 및 유용성 분석 필요 프로그램을 통해 익명처리가 적용되지 않은 비식별대상 확인 프로그램 및 수작업을 통한 재식별 가능성 확인 비식별대상이 아닌 기타 개인정보 및 민감한 정보가 포함되어 있는지 검토 필요

⑥ 추가처리	<ul style="list-style-type: none"> 프로그램을 통한 익명처리가 적용되지 않은 비식별 대상에 대해 추가 익명처리 적용 필요 기타 개인정보 및 민감한 정보에 대해 추가 익명처리 적용 필요
⑦ 활용 및 사후관리	<ul style="list-style-type: none"> 인공지능 학습용 데이터로 활용 인공지능 학습용으로 제공되기 전, 일정기간 동안 내부적으로 활용하면서 추가 익명처리를 진행하고 외부 공개 외부에서 접근이 가능한 데이터인 경우 무분별한 데이터 사용이 제한되도록 사용이력 등을 기록할 필요가 있음
+기타사항	<ul style="list-style-type: none"> 비정형 데이터는 정형 데이터 대비 용량이 크기 때문에 컴퓨팅 성능에 대한 익명처리 시간 및 품질에 대한 이슈가 있을 수 있음 익명처리를 수행하기 전 수집된 비정형 데이터들의 크기(픽셀 등)이 너무 작거나 크지 않아야 함 인공지능 학습용으로 제공되는 경우, 제공 유효기간을 설정하거나 주기적인 추가 익명처리 등을 통해 데이터를 갱신할 필요가 있을 수 있음 비정형 데이터는 익명정보 결합을 지원할 수 없음 개인정보보호법 개정안 적용 후에는 고정형 장비(CCTV 등)뿐 아니라 이동형 장비(드론 등)의 촬영 영상에 대한 기술적 익명처리 조치도 필요함

표 3. 영상 데이터 익명화 시 고려사항

III. 결론

본 논문에서는 비정형 데이터인 영상 데이터의 익명화 시 고려사항을 기존 가명·익명처리 가이드라인/안내서의 절차를 기반으로 분석하였다. 영상 데이터를 익명화하고 활용하기 위해서는 영상 데이터의 수집환경, 데이터 자체가 가지는 특징(피사체와 거리, 각도, 화질 등), 데이터를 활용할 환경 등 다양한 측면의 고려가 필요하다. 향후 연구로는 익명화 된 영상 데이터의 유용성, 안전성을 정량적으로 측정할 수 있는 모델, 기법 개발 등이 있다.

ACKNOWLEDGMENT

이 논문은 2021년 과학기술정보통신부의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No.2021-0-00634 ‘대용량 정형 데이터 대상 개인정보 가명·익명처리 자동화 및 안정성 검증 기술개발’, No.2021-0-00393 ‘영상 등 멀티미디어 데이터의 온전한 AI 학습활용을 보장하는 복원 불가형 개인식별정보 익명처리 핵심 기술 개발’)

참 고 문 헌

- [1] 안성원, 빅데이터의 특성과 동향, 소프트웨어정책연구소, 2016.
- [2] AI-Hub, <https://www.aihub.or.kr/>
- [3] 개인정보보호위원회, “가명정보 처리 가이드라인,” 2021.
- [4] 교육부, “교육분야 가명·익명정보 처리 가이드라인,” 2022.
- [5] 보건복지부, “보건의료 데이터 활용 가이드라인 개정(안),” 2022.
- [6] 금융감독원, “금융분야 가명·익명처리 안내서,” 2021.
- [7] 행정안전부, “공공분야 가명정보 제공 실무안내서,” 2021.
- [8] 영상 데이터 익명화 기술 및 평가방안, 한국정보화진흥원, 2019.