미국, 빅데이터 시대의 윤리와 프라이버시에 대한 심층 논의 본격화

- ◆ IAF(The Information Accountability Foundation), 빅데이터 분석을 위한 "윤리 4단계 프로젝트" 추진
- ◆ Randal Scott King, 자가 진단이 가능한 "데이터의 윤리적 조건 충족 판단을 위한 5대 질문" 제시
- ◆ NSF(National Science Foundation), "보안과 개인정보 침해 방지" 두 가지 모두를 달성할 방법론과 기술 제시

개요

o **미국 연방 정부**는 빅데이터 산업의 활성화를 위해 민간 비즈니스를 중심으로 하는 **다양한 빅데이터 연구 및 프로젝트를 활발히 진행**

- 美 국립과학재단, 국립보건원, 방위고등연구계획국 등 6개 부처 합동으로 **'빅데이터 R&D 이니셔티브(Big Data R&D Initiative)'** 진행
- 미국 IT R&D 프로그램 **NITRD**와 **OSTP**을 중심으로 **빅데이터 신규 프로젝트 진행**

· NITRD : Networking Information Technology R&D, 네크워킹 및 정보기술 연구개발
· OSTP : Office of Science and Technology Policy, 과학기술정책국

o 빅데이터 기술의 발전 및 시장 확대로 **기업이 수집·보유한 데이터 활용에 대한 윤리적 문제**와 **데이터 관리에 대한 보안 기술요건**에 대한 논의 등 다양한 **프라이버시 이슈가 제기**

- 회원 가입이나 어플리케이션 다운로드 및 업데이트, 경품 추천 등을 통해 개인정보 수집동의 약관이 남발되는 등의 여러 개인정보 문제가 발생
- 빅데이터 시대 개인정보를 침해하지 않으면서 개인에게 이득이 되는 서비스를 제공하는 것이 가능한가에 대한 프라이버시 논의 활발
- 이에 따라, **기업이 수집한 데이터의 관리 방법 및 보안 기술 수준에 대한 논의**가 필요해졌으며, **데이터를 다루는 윤리적인 마인드와 이를 체계화 시켜줄 윤리 프레임워크**에 대한 다양한 의견이 제시되고 있는 등 **본격적인 빅데이터 시대를 맞이하기 위한 보안과 윤리에 대한 대비가 필요**한 시점

◈ **미국 IAF(The Information Accountability Foundation)에서 진행 중인 '빅데이터 윤리 프로젝트'와 NSF(National Science Foundation)의 'Big Data – Security with Privacy' 보고서 주요내용을 살펴본 후 시사점 도출**

빅데이터 활용 시 고려해야 할 윤리적 측면에 관한 국가차원의 연구와 논의 진행

o 미국 인포메이션 거버넌스 기관인 The Information Accountability Foundation(이하 IAF)는 4단계의 빅데이터 윤리 프로젝트 진행

- IAF는 인포메이션 거버넌스를 기반으로 다양한 컨설팅 및 연구 활동을 하는 공공 기관으로 개인의 프라이버시에 대한 권리를 지키면서 데이터가 이끄는 혁신을 가능하게 하기 위해 정부, 공공 기관, 기업, 시민과 협력하는 등 데이터 보호를 위한 기관

o 「빅데이터 분석을 위한 통합적인 윤리 프레임(A Unified Ethical Frame for Big Data Analysis - Big Data Ethics Project)」이라는 주제로 4가지 파트의 윤리 프로젝트를 시작, 2014년 10월 첫 번째 보고서 발간

- 4 단계의 프로젝트는 2014년부터 2015년 동안 진행할 예정으로, 빅데이터 윤리에 관한 논의를 확장 및 구체화하고 실질적인 적용을 확산하는 것이 최종 목표

· Part 1 : '통합 윤리 프레임' 이라는 주제 하에 분석을 위한 데이터 활용 단계에서 생각해 봐야 할 5가지의 가치(Beneficial, Progressive, Sustainable, Respectful, Fair)를 제시

· Part 2 : 빅데이터 분석을 시행할 때 생각해 봐야 할 윤리적 문제에 대한 질의표를 구성하는 등 '질의응답 프레임워크' 작업이 진행

· Part 3 : '논의의 집행' 단계로 국가 기관이나 비즈니스 분야에 실질적인 적용을 해보는 것에 의미

· Part 4 : '산업 분야에서의 질의응답 모델'을 구축하는 단계로 다양한 산업 군에서의 활용 예시를 확산하는 단계

o 빅데이터를 위한 윤리 프레임워크와 관련한 민간 차원의 논의도 진행 중

- 빅데이터와 분석 분야의 기술 컨설턴트인 Randal Scott King은 5가지의 윤리 프레임워크를 제시하였으며, 빅데이터 프로젝트 수행에 앞서 자신이 제시한 5가지 질문을 통해 기업(혹은 기관)이 활용하고자 하는 데이터에 대한 윤리적 조건이 충족되었는가를 판단 가능

<데이터에 대한 윤리적 조건 충족을 판단하는 5가지 질문>

① 우리가 보유한 데이터는 익명인가, 또는 특정인을 설명 하는가?
② 우리는 어떠한 방법으로 이 데이터를 얻게 되었는가? 개인이 우리에게 데이터를 주었나, 혹은 다른 데이터로부터 추론되어 얻어진 것인가, 혹은 구입한 것인가?
③ 이 데이터 사용은 일부 지역(다른 관할권)에서 불법으로 간주 될 수 있나? 그것은 불법적인 의혹이나 혐의가 있는 것이 아닌가?
④ 가장 중요한 문제로, 데이터의 주인은 우리가 이 방법으로 데이터를 사용하고 알고 있다면 불편해 할 것인가?
⑤ 우리는 다른 사람이 우리에 대해 많은 것을 알고 있다는 사실을 받아들일 수 있는가?

윤리적 프레임워크를 기술적으로 뒷받침해 줄 보안 기술에 대한 정부 주도 논의 진행

o MIT 공대에서 NSF(National Science foundation, 국립과학재단) 주도로 진행된 '빅데이터 보안 및 프라이버시에 대한 워크숍'에서 빅데이터 프라이버시 유지를 위한 다양한 보안 기술 요건을 논의(2014. 09)

- 보안과 개인정보보호 간의 간극을 좁히면서 보안을 강화하기 위한 연구 방향 및 관련 기술 요소 제시
- 논의 결과를 정리하여 「Big Data – Security with Privacy」 보고서 발간(2014. 10. 16)
- 보고서는 향후 워크숍 참가자들에 의해 지속적인 논의와 피드백이 이루어짐에 따라 지속적으로 개선될 것이며, 마지막 워크숍 리포트는 2015년 2월에 완성될 예정

o 「Big Data – Security with Privacy」 보고서를 통해 제시된 주제인 '보안을 달성하면서도 개인정보의 침해를 야기하지 않는 방법론 및 기술'은 다음과 같음

프라이버시를 보존하는 데이터 매칭, 공동(협업) 데이터 마이닝, 생체 인증 등 개인정보보호 강화를 위한 다양한 보안 기술 요건을 제시

(1) 프라이버시를 보존하는 데이터 매칭

- 기존에는 데이터 변환과 벡터 공간 매핑을 기반으로 하며 다자간 컴퓨팅의 조합(SMC, Secure Multiparty Computation), 데이터 Sanitization 접근법, k-anonymity 등을 통해 대용량 데이터 세트를 다루기 위한 확장이 가능했음
- 앞으로는 위에 제시된 다양한 보안 기술을 조합하기 위한 보안 분석 및 증명이 활발할 것

(2) 프라이버시를 보존하는 공동(협업) 데이터 마이닝

- 전통적인 데이터마이닝은 모든 데이터를 수집한 중앙 집중식 대형 데이터 웨어하우스에서 수행하나 데이터가 다른 조직에 속하는 경우에 중앙의 모든 데이터를 수집하는 것은 개인정보보호 및 기밀 유지 문제를 보유
- 향후에는 자신의 데이터 세트를 유지하면서 각각의 데이터 세트에서 데이터를 노출하지 않고 글로벌한 데이터 마이닝 결과를 얻을 수 있는 분산된 협업 접근법이 필요하며 이를 위해서는 다음과 같은 기술 요소가 필요

① 각각의 데이터 세트에 대해 아무것도 학습하지 않고 두 당사자 간 의사 결정 트리를 구축하는 기술
② 연관 규칙, 클러스터링, k-nearest neighbor classification에 대해 특화된 공동의 프라이버시 보존 기술

(3) 프라이버시를 보존하는 생체 인증

- 생체 인증에 대한 기존 접근 방식은 인증 시 사용자에 의해 제공되는 템플릿을 기 등록 된 사용자 생체 인식 템플릿에 매칭 시키는 방식이 요구
- 최근의 접근법은 지각간섭기술(perceptual hashing techniques), 분류기술(classification techniques), 제로지식증명프로토콜(ZKPK, zero-knowledge proof of knowledge)을 조합하여 사용. 이를 통해 사용자들이 다양한 서비스 제공자와 상호 작용해야하는 분산 환경 하에 축적되는 민감한 생체 정보를 강력하게 보호

* ZKPK : 사용자의 생체 정보에서 비트 문자열을 추출 → 비트 문자열은 추가적인 분류와 변환 과정을 거침 → 확실한 암호화를 위해 난수를 사용하여 처리 → 원래 입력된 사용자의 생체 인식 정보가 전혀 드러나지 않는 식별 토큰 생성 → 향후 사용자 인증 시에 ZKPK 프로토콜 사용

데이터 내의 개인정보 보호를 위한 다목적 최적화 프레임워크 제시

· 응용 프로그램 도메인에 따른 유틸리티 정의, 프라이버시 리스크에 대한 정의, 주어진 프로토콜 상에서의 컴퓨팅, 스토리지, 통신 등의 비용 문제 정의가 필요

(1) 유틸리티 극대화, 위험과 비용 제약이 주어짐 : 이러한 조건은 제한적인 특정 개인정보보호 위험이 특히나 중요할 때 적합한 시나리오

(2) 개인정보보호 위협 최소화, 유틸리티와 비용 제약이 주어짐 : 일부 시나리오에서는 (예를 들어, 의료), 유틸리티의 상당한 저하가 허용되지 않을 수 있음. 이 설정에서 프로토콜의 매개 변수 값은(예를 들어, 차동적인 개인 정보 보호 같은) 우리에게 주어진 유틸리티 제약 내에서 개인 정보 보호를 최대화 하는 방법으로 선택되어짐. 일부 시나리오에서는, 모든 제약 조건을 만족시킬 수 있는 파라미터 설정이 없을 수도 있음

(3) 비용 최소화, 유틸리티와 위험 제약이 주어짐 : 일부의 경우, (예를 들어, 암호화 프로토콜)에서 모든 유틸리티 및 비용 제약 조건을 만족시킬 수 있는 가장 저렴한 프로토콜을 허용하는 프로토콜 매개 변수 설정을 발견 할 수 있음

빅데이터 보안 문제와 관련해 발생할 수 있는 다양한 문제를 예상하여 연구 과제와 종합적 접근 방법 제시

(1) 여러 데이터 기밀 기술 및 메커니즘 증 가장 주목되어 연구되고 있는 것은 '액세스 제어 시스템 및 암호화'로 빅데이터 적용을 위해서는 다음의 접근 방법 필요

- ① 다양한 액세스 제어 정책의 병합, ② 빅데이터에 대한 권한 관리와 사용 권한을 자동 부여, ③ 이종의 멀티미디어 데이터에 대한 액세스 제어 정책 시행, ④ 빅데이터 저장소에 액세스 제어 정책 시행, ⑤ 접근 제어 정책을 자동으로 설계, 개선 및 관리

(2) 예상치 못한 정보를 추출하기 위한 다양하고 큰 데이터 세트의 상호 연관성을 찾아내는 '데이터 상관관계 기술'을 위한 다음의 이슈 및 연구 방향 제시

- ① 추출 제어 및 무엇을 추출·사용·공유할지 점검 가능한 기술, ② 개인의 프라이버시와 주민의 프라이버시 모두를 지원, ③ 효율적이고 확장 가능한 프라이버시 강화 기술, ④ 데이터 프라이버시 정책의 유용성, ⑤ 데이터 서비스 수익 창출을 위한 접근, ⑥ 데이터 공개, ⑦ 데이터 품질에 대한 프라이버시 함축, ⑧ 위험 모델, ⑨ 데이터 소유권, ⑩ 인간적 요인, ⑪ 데이터 수집, 공유 등 데이터 라이프사이클 프레임워크

시사점

o 빅데이터 분석·활용이 효율성의 관점뿐만 아니라 윤리적 차원에서 논의될 필요성 제기

- 빅데이터 산업을 촉진하기 위한 공공 및 민간의 다양한 데이터 활용이 장려되지만 그와 함께 프라이버시 보호와 윤리적 요건을 갖추는 것이 필수

- 이에 미국 정부는 빅데이터 관련 정책을 만드는데 있어 산업 활성화 정책뿐만 아니라 빅데이터 시대의 프라이버시와 윤리에 대한 문제에 집중하기 시작하였으며, 기술 워크숍 진행, 자금 지원, 의견 청취 등 실질적인 액션을 취하고 있음

- 국내에서도 기업이 개인 정보를 다룰 때 필요한 보안 기술과 윤리적 측면에 대해 정부와 민간 차원의 이슈 제기 및 논의 등을 통해 균형적인 정책제안 필요

출 처

1. NSF Workshop, 「Big Data - Security with Privacy」
2. The Information Accountability Foundation, 「A Unified Ethical Frame for Big Data Analysis」
3. O'REILLY, Ethics of Big Data(번역본: 윤리적인 빅데이터 사용을 위한 정책 가이드)
4. Linked in, Randal Scott King, An Ethics Framework for Big Data