금융정보학회 토론문

서울대 명예교수 이천표

data analytics 및 이를 효과적으로 단 시간 내에 시현할 수 있게 하는 AI는 매우 중요
이 둘은 4차 산업혁명에서도 핵심적 요소, 특히 핀테크 및 precision medicine에서 데이터를 자체 생산함은 물론 이런 것을 자력 투입하는 자력갱생 성격의 강화학습을 통해 진가를 발휘하고 있음

이런 것을 훌륭하게 수행할 수 있도록 하는 전제조건은 많은 데이터를 수집하고 그것을 의미있게 분석하는 data analytics의 능력을 갖추는 것
data analytics와 관련되어 데이터의 중요성을 강조하느라 요즈음의 상황을 big data 시대라고 지칭하기도 해

그런데 우리나라에서는 법적인 제약으로 데이터의 수집이 매우 어렵다는 주장이 특히 강해
이러한 주장은 즉각 이러한 법률인 개인정보보호법, 신용정보법, 정보통신망법 등을 개정해 데이터의 수집과 활용을 용이하게 해야 한다는 주장으로 연결되고 있음

그러나 법의 개정이나 제정은 쉽지 않아
현재 우리나라의 사정에서 국회선진화법이 법의 제정이나 개정을 어렵게 하고 있어
보다 근본적으로는 데이터의 수집을 어렵게 하는 이상의 법들을 개정하려면 어떻게 개정해야 할 것인지에 대한 개정안을 가지고 있어야 하는데 우리가 어떤 개정안을 가지고 있는지는 의문(미국의 특허법 개정 예 참조)
on-line 및 off-line 세상을 모두 아우르는 법제를 마련해야 하나 이런 것이 어떤 것이어야 하는지에 대한 정리된 대안이 없어 개정안을 확정하기는 결코 쉽지 않아, 이점은 우리는 말할 것도 없고 미국이나 EU 모두가 고심하는 문제임

법의 개정이 쉽지 않은 상황에서 타협안으로 나온 것이 '개인정보 비식별화 가이드라인'임
이 가이드라인은 행정안전부, 미래창조과학부, 금융위원회 등 부처의 협의를 통해 강구된 다음 공표된 후 시행이 보류된 것임, 이것은 이 가이드라인을 따르는 경우 개인정보를 침해하지 않은 것으로 간주하여 개인정보 수집과 이용을 용이하게 하려는 의도를 가진 것이며, 이 안에는 비식별화를 가능하게 하는 여러 기술적인 방도가 제시 활용되고 있음, 즉 여기에서는 기술적인 방책을 통해 개인정보의 수집 및 이용을 쉽게 하려는 의도와 노력이 숨어 있음
단 어떠한 기술적 방책을 쓰더라도 그런 것은 불완전한 것이기에 그것을 넘어 재식별화하려는 시도를 막는 것은 본질적으로 불가능하다고 하고 그로써 기술적 방도에 대해 별 기대를 가져서는 안 된다는 기술전문가의 주장도 없지는 않음

우리의 비식별화 가이드라인은 미국의 HIPAA(personal data privacy guideline)과 아주 유사함

이러한 가이드라인에 의한 비식별화 이외에 암호화를 통해서도 개인정보 침해를 예방할 수 있다고 여겨지고 있음, 즉 그것은 개인정보법 등을 우회할 수 있도록 하는 기술적인 묘책이 될 수 있을 것이라고 기대되고 있음

이런 가이드라인이 2016년 6.30일 공표되자마자 법학계 및 법업계에서 즉각 반발이 있었음 이 가이드라인은 그 성격상 행정명령에 불과한 것이기에 그 상위에 있는 법률인 이상 열거된 여러 법률을 능가하지 못한다는 점을 반대론은 지적했음, 따라서 가이드라인에 따라 어떤 행위를 하게 되면 행정명령 상위에서 개인정보 침해를 금하는 이상 열거된 법률을 위반하는 것으로 되어 즉각 법적 처벌을 받게 될 것이라고 사실상 협박하는 양상이 전개되었음

이러한 현재의 우리의 상황은 비식별화 등에 대한 패배주의에 싸여 있는 상황이라고 성격지울 수 있겠음(핀테크의 추진과정에서의 양상과 비슷)

이러한 상황에서는 데이터 활용을 통해 경제활동 사회활동 등 여러 분야에서 효율성을 제고하고 생산성을 높이려는 시도는 위축되고 불가능하게 될 것임, 그로써 빅 데이터 시대를 제대로 살아가지 못하게 되고 여러 측면에서 외국에 비해 낙후되는 것을 피할 수 없게 될 것임

가이드라인과 관련되는 기술적 대비 이외에 더 감안해야 할 것이 유인제도(incentive system)를 활용하는 제도적 대비 방법임
이런 제도적 방법은 비식별화 또는 암호화한 데이터를 재식별화하고 복호화하려면 돈과 노력이 든다는 사정을 주목함, 이에 따라 누구나 언제나 마구 재식별화 내지 복호화하지는 않을 것이라고 인식함, 여기에 더하여 어떤 필요에 따라 시간과 돈을 들여 비식별화 또는 암호화를 했음에고 불구하고 그런 의도에 반하는 시도를 하는 경우에는 그런 시도를 사안별로 점검하여 징벌을 부과하기로 함, 이런 제도적 방식으로 비식별화 내지 암호화의 기술적 노력을 보완함

개인정보보호의 상위개념으로서 cyber security의 보호를 생각해 볼 수 있는데 그곳에서의 논의도 기술적인 방책으로 완벽을 기하는 것이 어렵다는 것을 수긍하고 그것의 한계를 유인제도를 도입해 보강하려고 하고 있으며 궁극적으로는 시장기구를 활용하도록 해 보호제도를 보완해야 한다고 결론내고 있음

비식별화 내지 암호화에 대한 우리나라에서의 인식과 대비되는 것으로서 미국의 사정을 참고할 필요가 있음
미국에서는 위의 HIPAA가 시행되고 있는 한편 빅 데이터 시대에 임해 각종 데이터의 적극적 이용을 선도하는 여러 기업에 의한 효율적 이용의 선례가 속속 나타나고 있음, 그 이외에 정부부처에서도 이를 보다 용이하게 하려고 제도를 바꾸는 정책변경까지 나타나고 있음

기업에 의한 선례로서는 많은 데이터를 가진 Google, Facebook, Amazon, Walmart 등의 데이터 분석 및 활용을 주목할 필요가 있음, 이들은 opt-out형식으로 제공받은 데이터를 target marketing이나 target advertizing등에서 잘 활용하고 있음, 이러한 활용 예에서는 자사가 공여받은 데이터뿐만 아니라 SNS등을 통해 구할 수 있는 사람들의 정치 사회 문화활동에 관한 자사 밖의 데이터도 함께 섞어 이용하는 것으로 알려지고 있음, 나아가 그러한 데

이터를 필요로 하는 제3자에게 그 필요에 맞게끔 가공하여 공여하고 중개하는 사실상의 data broker, data bank로서 변신하고도 있음, 이런 변화에 기승하여 데이터의 가공에 장기를 보이고 있는 소프트웨어 스타트업들도 나타나 data broker나 data bank의 역할을 하면서 Google등과 경쟁하고 있음

정부에 의한 정책변화의 예는 FCC(federal communication commission)가 많은 데이터를 가지는 통신기업들로 하여금 data broker로서 활동하는 것을 허용하려고 한다는 것에서 볼 수 있음, 데이터를 활용해 빅 데이터 시대를 이끌고 미국경제의 효율성을 높이고 있는 이상의 Google 등 기업들이 모두 FTC(federal trade commission)의 감독 하에 있는 비통신기업이라는 것을 인지하고 있는 FCC는 그런 기업들에 못지않게 많은 데이터를 가지고 있는 각종 통신기업들도 그러한 데이터를 활용해 미국경제에 기여하도록 촉구하고자 하여 이런 정책적 변화를 꾀하는 것이라 해석 됨, 이를 가능하게 하는 법은 현재 의회를 통과했고 대통령의 공포를 기다리고 있다고 보도되었음

오늘 날 사람들은 하루에도 십여 번 CCTV에 의해 촬영되고 있음, 이는 사실상 개인정보가 많이 노출되고 있다는 증좌임, 미국에서는 CCTV의 촬영내용을 사후적으로 사람이 확인해야 하는 과정을 거치지 않고 실시간으로 촬영내용을 수사기관 등에 전하고 있는 AICCTV도 이용되고 있음(우리나라에서는 이를 2020년까지 ETRI로 하여금 개발하게끔 계획하고 있음), 그런데 이런 촬영이 개인정보 침해가 된다면 많은 나라에서 불법행위가 공공연히 자행되고 있는 꼴임, 그런데 CCTV를 통한 무차별적 촬영을 개인정보의 침해라고 문제 삼는 나라가 있는가? 어떤 정보의 이용이 개인정보의 침해로 되는지의 여부가 반드시 분명하지 않은 이러한 사정은 앞으로 IoT(internet of things)가 보편화되어 현재보다 현저히 많은 센서가 다기한 정보를 취득해 전파하는 세상이 본격화되게 될 경우 더 빈번하게 나타날 수 있고 더 심각한 상황으로 될 수도 있음

개인정보보호를 지고지상의 가치라고 보고 그것에 매달려 나라 밖에서의 변화를 외면할 것이 아니라 우리도 빅 데이터 시대에 임해 사회 전체로서 데이터를 보다 광범위하게 모우는 한편 data sharing에 대해서도 눈을 떠야 하겠음, personal data와 private data를 구분할 줄 알아 보호의 대상을 부당하게 확대하지 말아야 할 것이고 demographic data와 behavioral data를 구분하여 후자의 이용에 대해서까지 쓸 데 없이 위축되는 어리석음을 범해서는 안 될 것임, 데이터의 부당한 악용은 막되 선용까지 회피하지는 않게끔 올바른 data governance의 시각을 어서 정립해야 하겠음, 무조건적으로 개인정보를 신성시할 것이 아니라 정보 이용의 윤리를 정립하고 그 틀 안에서 데이터를 최적으로 이용할 수 있도록 하는 제도를 가지도록 되어야 할 것임, 데이터의 부당이용을 징계하는 한편 데이터의 새로운 이용을 격려하는 제도를 정립해야 할 것이며 사소한 것이지만 data portability를 허용하는 방안도 용인되어야 좋겠음, 이런 고려를 모두 담는 기술적 대비책과 사회제도적 대비책을 가질 수 있어야 하겠음

구술이 서 말이라고 하더라도 꿰어야 보배라고 함, 1차적으로 많은 데이터가 마련된다 하더라도 2차적으로 그것을 분석하고 효과적으로 쓸 수 있도록 하는 알고리즘의 마련이 꼭 필요함, 가지고 있는 문제의식에 상응하는 알고리즘을 생산하거나 고안해 내는 능력을 갖출 수 있게 되어야 하겠음, 일부에서 무책임하게 open source algorithm을 가지고 알고리즘에 대한 필

요를 채울 수 있다고 말하는 것은 오늘날의 비즈니스를 좌우하는 알고리즘들이 어떠한 것인가를 반추해 볼 때 말도 안 되는 넌센스임을 알아야 함, real time data와 legacy data를 섞는 방법, business data와 engineering data와 SNS data와 기타를 융합하는 algorithm을 제작할 수 있어야 하겠고 그에 상응하는 다기한 AI bot을 지닐 수 있어야 하겠음, 이를 위해 어서 coding교육을 펼쳐야 하겠고 그로써 이른바 computational thinking을 할 수 있도록 되어야 하겠음

마지막으로 평소 알고 싶었던 기술적 차원의 질문 4가지

우선 비식별화와 관련해

미국에서 비식별화에 대해 우리처럼 민감하지 않은 이유는 오늘날 많은 IT 디바이스를 간단없이 쓰고 있는 개방사회에서 적당한 정도 또는 상당한 정도의 노력을 하기만 하면 비식별화한 데이터를 상당한 정도 재식별화하는 것이 가능하게 되어 있고 그로써 어차피 개인정보를 철저히 보호하는 것은 불가능하다고 하는 사실상 사태인식이 깔려 있기 때문은 아닐까? 반면 우리는 이러한 ICT 사회의 실상을 외면한 채 재식별화의 남용에 관한 극단적 사례를 끄집어내고 우려하면서 데이터의 효과적 활용에는 게으른 '부작위에 의한 작위'의 어리석음을 범하고 있지 않은가? 우리나라에 주민등록번호제도가 있다는 것을 핑계거리로 삼아 개인정보의 누출의 개연성을 과다하게 걱정하고 그 이면에서 그것을 선용하려는 노력은 부당하게 억제되고 있다고 볼 수는 없는가?

다음 암호화와 관련해

우선 암호화에 소요되는 코스트에 대해 알고 싶음, 암호화 코스트가 매우 높다면 이를 통한 기술적 대비에는 한계가 있을 것이기 때문임, 암호화 코스트는 기술발전과 더불어 차차 낮아져 왔는가?

10여년 전까지 미국은 최신예의 암호화기술의 수출을 금지하는 정책을 써온 것으로 알고 있는데 작금 암호화기술의 수출에 대한 미국의 정책은 어떠한가?

향후 양자컴퓨팅이 제법 쓰이게 되는 상황으로 되었을 때 암호화 기술의 이용에서는 어떠한 변화가 있을 것인가?