



빅데이터 분석의 자동화와 정보보호

임준 연구위원

- 최근의 데이터 분석은 개인화와 자동화 경향을 보이고 있는데, 이러한 경향은 정보 활용 및 보호와 관련한 새로운 과제를 제시하고 있음
 - 분석 대상이 그룹에서 개인으로 세분화되는 개인화 경향은 다양한 맞춤형 서비스의 제공이라고 하는 긍정적 측면도 있지만 부당한 차별의 가능성이 높아지는 부정적 측면도 있음
 - 한편, 분석에 있어서 인간의 역할은 감소하고 기계의 역할은 증가하는 자동화는 개인화에 의한 부당한 차별의 문제를 더욱 복잡하게 만들 수 있음
 - 왜냐하면 데이터 분석이 자동화되면서 알고리즘이 과거에 비해 상당히 복잡해져 어떠한 논리에 의해 결과가 나왔는지를 알아내는 것이 점점 더 어려워지기 때문임
- 국내 정보보호 관련 법규인 「개인정보 보호법」과 「신용정보의 이용 및 보호에 관한 법률」에는 아직 개인정보 분석의 자동화와 관련된 명시적인 조항이 존재하지 않음
- 반면, 유럽의 일반정보보호규칙(General Data Protection Regulation)에는 개인정보 분석의 자동화와 관련된 명시적인 조항이 존재함
 - 제22조에 의하면, 정보주체는 자동화 처리에만 근거한 결정이 법적 효과를 가져오거나 자기에겐 심각한 영향을 미칠 경우 이러한 결정에 따르지 않을 권리를 가짐
 - 단, ① 정보주체와 정보관리자 간의 계약 체결이나 이행을 위해 필요한 경우, ② EU법이나 회원국법에 의해 허용되는 경우, ③ 정보주체의 명시적인 동의에 의한 경우에는 정보주체의 권리를 제한할 수 있음
 - 한편, 정보주체의 반대권을 제한할 수 있는 첫 번째와 세 번째의 경우에는 정보관리자가 정보주체의 권리를 보장하기 위한 적절한 장치를 마련하여야 함
- 인공지능 등 디지털 기술의 발전 속도를 감안할 때, 국내 보험산업에서도 가까운 장래에 데이터 분석에 있어서 기계의 역할이 증대될 것으로 예상되므로, 개인정보 분석의 자동화와 관련된 법규의 정비가 필요할 것으로 보임
 - 이러한 법규 정비에 있어서 보험산업의 특수성을 어떻게 반영하느냐가 주요 연구과제가 될 것으로 판단됨

1. 검토배경



- 최근 빅데이터 분석의 특징 가운데 하나는 분석 대상의 세분화로, 그룹 전체의 평균적 행태 예측에서 개인별 특성 및 행태 예측으로 발전하고 있음
 - 개인별 행태 분석은 다양한 맞춤형 서비스의 제공이라고 하는 긍정적 측면도 있지만, 다른 한편으로는 부당한 차별 가능성이 높아지는 부정적 측면도 있음
- 이러한 개인별 행태 분석의 부정적 측면과 관련된 이슈는 빅데이터 분석의 또 다른 특징이라고 할 수 있는 자동화에 의해 더 복잡해지고 있음
 - 여기서 자동화는 분석에 있어서 인간의 역할은 점차 감소하고 기계의 역할이 커지는 것을 의미하는데, 예를 들어, 기계학습(machine learning)의 경우 컴퓨터가 스스로 예측모형을 계속 개선해 나감
 - 데이터 분석이 자동화되면서 발생하는 문제점 가운데 하나는 알고리즘이 과거에 비해 상당히 복잡해져서 어떠한 논리에 의해 결과가 나왔는지를 파악하는 것이 점점 더 어려워진다는 점임¹⁾
 - 이전까지의 데이터 분석은 잘못된 결과가 나왔을 경우 그 원인을 추적해서 왜 그런 결과가 나왔는지 알아내는 것이 상대적으로 용이했음
- 이와 같은 데이터 분석의 개인화와 자동화 경향은 정보 활용 및 보호와 관련한 새로운 이슈를 제기함
 - 자동화 분석에 의해 개인별로 차별화된 의사결정이 이루어지는 경우 이러한 분석을 어디까지 허용할 것인가 그리고 이러한 분석으로부터 개인을 보호하기 위해 어떤 장치를 마련해야 하는지 등이 이슈로 제기될 수 있음
- 본고에서는 데이터 분석의 개인화 및 자동화와 관련된 이슈를 보험산업과 관련하여 살펴보고 시사점을 도출하고자 함

1) 빅토르 마이어 쉐이버거 · 케네스 쿠키어(2013), 『빅데이터가 만드는 세상』, 21세기 북스

2. 국내현황



■ 여기서는 ① 언더라이팅, ② 고객관리(이탈방지), ③ 지급심사(보험사기)의 세 분야를 대상으로 데이터 분석의 개인화 및 자동화 측면에서 국내 보험산업의 수준을 살펴보고자 함

■ 언더라이팅의 경우, 빅데이터 분석의 발전과 함께 제기될 수 있는 이슈 가운데 하나는 자동화임

- Aggour et al.(2006)²⁾ 자동 언더라이팅의 발전단계를 아래 표에서 보듯이 3단계로 구분하였는데, 현재 국내 보험산업의 경우에는 2.5세대 정도에 해당된다고 볼 수 있음
 - 처리 가능한 질병의 범위는 2세대보다 넓으나, 아직 자연어 처리에 있어서는 다소 한계가 있는 상황임

〈표 1〉 자동 언더라이팅 시스템의 발전단계

구분	특징
1세대	• 질병이 없는 가입자의 신청서(applications with no medical impairments) 처리
2세대	• 고혈압이나 당뇨와 같은 주요 질병 가입자의 신청서(applications with some major impairments) 처리
3세대	• 처리 가능한 질병 범위의 확대 • 자연어 처리 기능 추가 • 완전한 자동화가 어려운 경우 인간 언더라이터(human underwriter)를 보조하는 기능 추가

자료: Aggour et al.(2006)

■ 고객관리와 관련해서는 이탈이 예상되는 고객을 예측하고 적절한 조치를 통해 이들을 유지하는 데 빅데이터 분석이 활용될 수 있음

- 이 경우 데이터 분석의 핵심은 마케팅 비용은 최소화하면서 고객 유지효과는 최대화할 수 있는 타겟 고객 군을 어떻게 골라내느냐 하는 데 있음
- 초급수준의 빅데이터 분석에서는 고객유지를 위한 마케팅 활동을 하면 남아 있을 고객을 단순히 예측하게 되는데, 이들 가운데에는 마케팅 노력을 하지 않더라도 남아 있을 고객이 포함되어 있음

²⁾ Aggour, Kareem S., Piero P. Bonissone, William E. Cheetham, and Richard P. Messmer(2006), "Automating the Underwriting of Insurance Applications", *AI Magazine*, Volume 27 Number 3, pp. 36~50

- 고급수준의 빅데이터 분석에서는 마케팅 활동을 하지 않더라도 남아 있을 고객은 제외한 그룹, 즉 고객유지 마케팅을 하면 남아 있지만 하지 않으면 떠날 고객 군만을 선별해냄³⁾
 - 국내 보험산업의 경우 이탈방지를 위한 빅데이터 분석에 있어서 아직 정밀하게 타깃 고객 군을 선별해낼 수 있는 수준에 이르지 못하는 상황임
- 보험산업에 있어서 빅데이터 분석이 유용하게 활용될 수 있는 또 다른 분야는 보험사기로 국내의 경우 다양한 노력들이 현재 진행되고 있음
- 예를 들어, 금융감독원의 경우 자동차 고의사고 다발자, 허위·과다 입원환자, 허위·과다 입원 조장병원의 3개 유형에 대해 상시감시시스템을 구축하였음⁴⁾
 - 고질적 보험사기 3개 유형에 대해 상시감시지표를 마련하고, 지표를 토대로 보험사기 혐의자를 유의, 심각, 위험의 3개 등급으로 분류
 - 자동차 고의사고 다발자의 경우에는 전체 사고건수, 보험금 수령규모, 연간 집중사고 건수, 미수선 수리비 비중, 외제차 사고건수 등 22개 지표 마련
 - 허위·과다 입원환자의 경우에는 보험계약 건수 및 체결회사 수, 최근 5년간 입원 횟수, 주요 질병 입원 횟수 등 12개 지표 마련
 - 허위·과다 입원 조장병원의 경우에는 장기 입원환자 비율 등 9개 지표 마련
 - 국내 보험사기적발시스템에서 고도화가 필요한 영역 가운데 하나는 보험금 지급이전 단계에서의 보험사기 모니터링임
 - 현재는 보험금 지급이전 단계의 모니터링이 주로 개별 회사 차원에서 지급심사 담당자의 경험법칙(rule of thumb)에 의해 이루어지고 있는데, 보험산업 차원에서 빅데이터 분석을 통해 모니터링이 이루어지게 되면 보험금 누수 방지효과를 제고할 수 있음
- 한편, 보험산업의 개인정보 분석 및 정보보호와 관련된 법규에는 「개인정보 보호법」과 「신용정보의 이용 및 보호에 관한 법률」 등이 있는데, 아직 개인정보 분석의 자동화와 관련된 명시적인 조항이 존재하지는 않음⁵⁾

³⁾ Eric Siegel(2013), *Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die*, Wiley

⁴⁾ 금융감독원(2017. 2. 14), “「보험사기 예방 3중 레이더망」으로 고질적 보험사기 적발”, 브리핑 자료

⁵⁾ 국내 보험산업 관련 정보보호 규제현황은 [부록] 참조

3. 유럽사례



■ 유럽의 경우에는 개인정보 분석의 자동화와 관련된 명시적인 조항이 존재함

- 2016년 일반정보보호규칙(General Data Protection Regulation, 이하 ‘GDPR’) 제3장 정보주체의 권리에 관한 조항들 가운데 하나가 프로파일링(profiling)을 포함한 자동화된 개별적 의사결정(automated individual decision-making)에 관한 것임⁶⁾
 - ‘프로파일링’이란 자연인과 관련되는 일정한 개인적인 측면을 평가하기 위하여, 특히 그 자연인의 업무성취, 경제적 상황, 건강, 개인적 선호, 관심, 신뢰성, 행동, 위치나 이동에 관한 측면을 분석하거나 예측하기 위하여 그들 정보를 이용하는 것을 포함하는 자동화된 처리의 형태임⁷⁾

■ GDPR 제22조에 의하면, 정보주체는 자동화 처리에만 근거한 결정이 법적 효과를 가져오거나 자기에 심각한 영향을 미칠 경우 이러한 결정에 따르지 않을 권리를 가짐⁸⁾

- 그러나 다음의 경우에는 자동화된 의사결정에 대해 정보주체가 반대할 권리를 가지지 못함
 - 정보주체와 정보관리자(data controller) 간의 계약의 체결이나 이행을 위해 필요한 경우
 - EU법이나 회원국법에 의해 허용되는 경우
 - 정보주체의 명시적인 동의에 근거한 경우
- 정보주체가 반대권을 가지지 못하는 세 가지 경우 가운데 첫 번째와 세 번째 경우에는 정보관리자가 정보주체의 권리를 보장하기 위해 적절한 조치를 이행하여야 함
 - 예를 들어, 자동화 처리 결과에 대해 이의를 제기하거나 사람의 개입(human intervention)을 요구할 수 있는 권리를 보장

■ 현재로서는 데이터 분석의 자동화에 대한 대응책으로 정보주체에게 자동화 처리에 대한 반대권을 주는 정도인데, 아직 현실화된 것은 아니지만 보다 근본적인 대안 가운데 하나는 빅데이터 분석의 모니터링

⁶⁾ GDPR과 관련된 내용은 함인선(2016), 『EU개인정보보호법』에서 인용

⁷⁾ GDPR Article 4, (4) ‘profiling’ means any form of automated processing of personal data evaluating the personal aspects relating to a natural person, in particular to analyse or predict aspects concerning the data subject’s performance at work, economic situation, health, personal preferences or interests, reliability or behavior, location or movements

⁸⁾ GDPR Article 22, 1. The data subjects shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her

과 투명성 제고를 위한 새로운 유형의 전문 기술 개발과 제도의 도입임⁹⁾

- 쉐이버와 쿠키어(2013)는 공인회계사의 등장 과정과 비교하면서 빅데이터 분석의 적정성을 평가하는 새로운 전문가 집단이 출현할 것으로 예상하였는데, 그들은 그러한 전문가 집단을 ‘알고리즘리스트(Algorithmist)’라고 명명하였음
 - 20세기 초에 철도의 등장과 함께 이전과는 비교할 수 없을 정도로 회계 관련 정보의 양과 복잡성이 증가하였고, 이러한 발전을 정부규제 및 감독이 쫓아가지 못하면서 여러 회계조작 관련 스캔들이 발생하였음¹⁰⁾
 - 이로 인해 회계 투명성 제고에 대한 요구가 제기되었고, 그 과정에서 회계법인 등 회계 관련 전문가 집단이 출현하게 되었음

4. 시사점



- 당장은 아니지만 인공지능 등 IT 기술의 빠른 발전을 감안할 때, 국내 보험산업의 경우에도 머지않은 장래에 데이터 분석에 있어서 기계의 역할이 증대할 것으로 예상되므로, 개인정보 분석의 자동화와 관련된 법규의 정비가 필요할 것으로 보임
 - 이러한 법규 정비에 있어서 보험산업의 특수성을 어떻게 반영하느냐가 주요 연구과제가 될 수 있음
 - 예를 들어, GDPR에서는 계약 체결 및 이행에 필요할 경우 정보주체의 자동화 처리 반대권을 제한할 수 있는데, 자동 언더라이팅을 계약 체결 및 이행에 불가결한 요소로 볼 것인지 여부도 하나의 이슈가 될 수 있음
 - 그리고 보험사기 혐의자 분석은 계약 체결 및 이행을 위한 데이터 처리로 보기 어려운 측면이 있는데, 그렇다면 보험사기의 경우에는 정보주체의 자동화 처리 반대권을 인정할 것인지, 만약 제한한다면 어떤 법에 어떤 방식으로 반영할 것인지 등에 대한 연구가 필요
- 단기적으로는 정보주체에게 자동화 처리 반대권을 부여하는 방안이 논의의 중심이 되겠지만 보다 장기적으로는 빅데이터 분석의 투명성 제고를 위한 전문기술과 제도에 대한 연구가 필요함 **kiri**

9) 빅토르 마이어 쉐이버 · 케네스 쿠키어(2013), 『빅데이터가 만드는 세상』, 21세기 북스

10) 공인회계사 등장의 역사적 배경은 제이콥 솔(2016), 『회계는 어떻게 역사를 지배해왔는가』 메멘토에서 인용

[부록] 국내 보험산업 관련 정보보호 규제현황



- 국내 보험산업의 개인정보 분석 및 정보보호와 관련된 법령에는 「개인정보 보호법」과 「신용정보의 이용 및 보호에 관한 법률」(이하 ‘신용정보법’)이 있음
 - 본고에서는 이들 법령의 조문 가운데 ① 정보의 처리와 ② 정보주체의 권리에 관한 내용을 중심으로 소개하고자 함
- 우선, 「개인정보 보호법」의 경우에는 제3장에 개인정보의 처리에 관한 내용이, 그리고 제5장에 정보주체의 권리 보장에 관한 내용이 담겨져 있음
 - 법 제15조에 의하면, 개인정보처리자는 수집 목적의 범위 내에서 이용할 수 있으며, 다음과 같은 경우에 개인정보를 수집할 수 있음
 - 정보 주체의 동의를 받은 경우
 - 법률에 특별한 규정이 있거나 법령상 의무를 준수하기 위하여 불가피한 경우
 - 공공기관이 법령 등에서 정하는 소관 업무의 수행을 위하여 불가피한 경우
 - 정보주체와의 계약 체결 및 이행을 위하여 불가피하게 필요한 경우 등
 - 민감정보와¹¹⁾ 고유식별정보의¹²⁾ 경우에는 별도의 규정을 두고 있는데, 법 제23조와 제24조에는 이들 정보에 대한 처리제한 규정이 담겨져 있음
 - 정보주체로부터 민감정보와 고유식별정보의 처리에 대한 별도의 동의를 받거나 법령에서 요구·허용하는 경우에 한해 처리가 가능함
 - 한편, 정보주체의 권리 보장을 위해 개인정보 열람요구권(법 제35조), 정정·삭제요구권(법 제36조), 처리정지요구권(법 제37조) 등을 두고 있음
 - 처리정지요구권의 경우 계약의 이행을 위해 개인 정보처리가 필요하거나 법률에 특별한 규정이 있으면 개인정보처리자는 정보주체의 처리정지 요구를 거절할 수 있음
- 「신용정보법」의 경우에는 제3장에 신용정보의 처리에 관한 내용이, 그리고 제6장에 신용정보주체의 보호에 관한 내용이 담겨져 있음

11) 민감정보에는 사상·신념, 노동조합·정당의 가입·탈퇴, 정치적 견해, 건강, 성생활 등에 관한 정보, 유전자검사 등의 결과로 얻어진 유전정보, 「형의 실효 등에 관한 법률」 제2조제5호에 따른 범죄경력자료에 해당하는 정보 등이 포함됨

12) 고유식별정보에는 주민등록번호, 여권번호, 운전면허의 면허번호, 외국인등록번호 등이 해당됨

- 법 제15조에 의하면, 신용정보회사 등은 수집·조사 및 처리의 목적을 명확히 하고, 목적 범위 내에서 개인신용정보를 수집·조사 및 처리하되 다음과 같은 경우에 수집 가능함
 - 정보 주체의 동의를 받은 경우
 - 법률에 특별한 규정이 있거나 법령상 의무를 준수하기 위하여 불가피한 경우
 - 정보주체와의 계약 체결 및 이행을 위하여 불가피하게 필요한 경우 등
- 한편, 신용정보주체의 보호를 위해 법 제38조와 제38조의 3을 통해 정보주체에게 신용정보 열람 청구권, 정정청구권, 삭제요구권 등을 인정하고 있음