

信用スコアに潜むバイアスに どう備えるか

個人の信用度を数値化した信用スコアリングサービスが社会的に認知されてきている。一方で、信用スコアに性差別的な傾向があるとの指摘がなされており、アルゴリズムに社会的差別を固定化するようなバイアスが潜んでいることが問題となりつつある。アルゴリズムの活用には「入り口」から「出口」までの一貫した監査を整備するなどの対応が必要である。

アルゴリズムによる「性差別」？

近年、様々な事業者による与信の自動化、いわゆる信用スコアリングがサービスとして提供されている。最も有名かつ大規模なものはアント・フィナンシャルグループによる芝麻信用だろう。こうした普及とともに、特定の信用スコアに性差別的な傾向があるのではないかと指摘が出てきている。

一つは2019年に米アップルがゴールドマン・サックスと共同でサービスを開始したApple Cardの利用限度額が性別によって歪められているのではないかと指摘である。これは2019年11月にアップルの共同創業者であるスティーブ・ウォズニアク氏によるTwitterへの投稿が耳目を集めた。同氏によれば、夫妻は口座もクレジットカードも共同である（つまり夫妻の間で保有資産額に差はなかった）にも関わらず、Apple Cardの利用限度額に同氏と夫人では実に10倍の差があったと投稿したのである。

また日本でも、みずほ銀行とソフトバンクが共同で行っている信用スコアサービスであるJ.Scoreに、「年収や職業など他の条件が同じでも、性別を男性から女性にするだけでスコアが下がる」との指摘が寄せられていた。J.Scoreではこの指摘を受けて、性別の影響を弱めるようアルゴリズムの修正を行ったと伝えられている¹⁾。

これらの信用スコアは、過去の融資履歴や口座の入出金状況などの基本的な金融取引記録に加え、日常の活動に関する様々なデータを活用することで、これまでは把握できなかった「与信評価」を可能にするものである。そして、この「評価」を行うのは膨大なデータから与信

に関する情報を抽出し数値化するアルゴリズムである。このアルゴリズムに性差別的傾向があるのではないかが問題となっている。

欧米で進む 「アルゴリズムによる差別」規制

欧米では、このようなアルゴリズム（AIと言い換えてもいいが）による評価が、社会的差別や偏見を固定化させてしまうことに対する懸念が強まっている。

EUでは、EU一般データ保護規則（GDPR）22条で、「データ主体に対して法的効果（または同様の重大な影響）を及ぼす場合」において、「完全な機械化・自動化されたプロファイリングのみによる決定を禁止」している（ただしAIを用いたプロファイリング自体は禁止していない）。また、同規制21条では自動化されたプロファイリングに対してデータ主体が異議を述べる権利も明記されている²⁾。

またアメリカでは1970年代に制定された「信用機会均等法（Equal Credit Opportunity Act : ECOA）」によって、与信判断において性別、年齢や既婚・未婚といった婚姻状況、人種や肌の色、宗教、出身国といった社会的属性による差別を禁止している。この信用機会均等法の権限は非常に強く、広告や勧誘においても差別的な取扱いは厳しく制限されている。前述のApple Cardのケースでは、利用者からの指摘を受けてニューヨーク州当局が同法違反の可能性の調査に乗り出した。

アルゴリズムに潜むバイアス

アルゴリズムによる「差別」が生じる要因は、大きく

NOTE

- 1) 「採点される人生 (3) 未完成の審査AI どこまで頼る 迷う使い手」日本経済新聞 (2019/4/26)。
<https://www.nikkei.com/article/DGKKZO44220160V20C19A4MM8000/>
- 2) GDPRによるプロファイリングのガイドラインについては個人情報保護委員会の「自動化された個人に対する意思決定とプロファイリングに関するガイドライン」を参照。
https://www.ppc.go.jp/files/pdf/profiling_guideline.pdf
- 3) ただ、最近の研究では男女を完全に分けた与信モデルを作成することで、女性の信用スコアが上昇するという結果も報告されている。
"Gender-differentiated Credit Scores: Bridging the Gender Gap in Access to Credit"
<https://medium.com/center-for-effective-global-action/gender-differentiated-credit-scores-bridging-the-gender-gap-in-access-to-credit-87e040318cdb>
- 4) アルゴリズムの監査については、以下の記事も参照。
「いま、アルゴリズムには「監査」が求められている」WIRED.jp (2019/2/4)。
<https://wired.jp/2019/02/04/business-fair-audit-algorithm/>

3つあると考えられる。それらの要因の中にはアルゴリズムそのものではなく、その手前のデータに問題があるケースもある。

まず一つ目は実際にアルゴリズムが「性別」などの社会的属性を与信評価に利用している場合である。この場合、アルゴリズムから性別などのデータを利用しないような変更が必要となる³⁾。

二つ目の要因はさらに複雑である。Apple Cardのケースでは、カードの発行元であるゴールドマン・サックスは「(利用限度額を審査する) アルゴリズムは性別をデータとして入力していない」と反論した。しかし、仮にアルゴリズムが「性別」を利用していなくても、その他のデータから「性別」を推測できる場合、アルゴリズムはやはり「性別」による差別と同様の結果を導き出す可能性がある。例えば「職業」がアルゴリズムの入力データとして利用されている場合、女性に多い職業（看護師や小学校教諭など）や、男性に多い職業（技師、経営者など）といった情報から容易に男女が識別できてしまい、与信判断に性差別的バイアスが混入してしまう可能性がある。

そして最後の要因が「選択バイアス（セレクション・バイアス）」である。これは「分析対象とする集団全体」に対して、「実際の分析対象」が特定の層を多く（または少なく）選択してしまう場合のバイアスである。信用スコアの場合、信用スコアを積極的に利用しようとする層は、一般的には「高い信用スコアが得られることが期待できる層」が主体となるだろう。これらの層の情報に基づいてアルゴリズムを最適化した場合、信用スコアが高い層により有利となるようなバイアスがアルゴリズムに生じる可能性がある。

アルゴリズムの「監査」が求められる

信用スコアに代表されるアルゴリズムによる与信の自動化それ自体は与信コストの低下をもたらし、金融機会の拡充をもたらすイノベーションである。しかし、アルゴリズムによって社会的差別や偏見が再生産されることがあってはならないというのが世界的なコンセンサスである。そのためには、アルゴリズムの「入り口」と「出口」における「監査」の役割が重要になるだろう。

具体的には、アルゴリズムに利用する入り口の「データ」に関しては、社会的差別につながる変数が含まれていないかを（該当する法規制に準拠しているかも含めて）確認することが求められる。また得られた信用スコアという出口について、意図せざるバイアスが生じていないかを確認することも必要になるだろう。

このような「監査」を行うには相当程度の統計学の知識と、企業倫理に関する専門性が必要である。専門家の育成とあわせて、具体的な監査項目の検討を進めることが今後重要になってくるだろう⁴⁾。

信用スコアは担当者個々人の主観が混じらない客観的な信用基準になる可能性を秘めている。しかしそれを作り利用するのは人間である。この点を肝に命じておく必要があるだろう。

Writer's Profile



柏木 亮二 Ryoji Kashiwagi

金融イノベーション研究部
上級研究員
専門はIT事業戦略分析
focus@nri.co.jp