

個人の傾向を考慮したプロ野球投球データからの 反実仮想説明集合の抽出

武藤 暖

要旨

深層学習のような機械学習手法の発展により、機械学習モデルが医療の現場やマーケティングによる施策などに応用され始めている。このことから、機械学習モデルの予測根拠や判断基準を人間が理解可能な形で提示することができる説明可能 AI 技術の重要性が高まっている。その中でも反実仮想説明法は、機械学習モデルによる予測結果に対して、結果状態を覆すために必要な説明変数の変化量を算出し、例として提示する説明手法であり、具体的な目標や行動をイメージすることができるため注目を集めている。

本研究では、反実仮想説明の生成に Diverse Counterfactual Explanations (DiCE) と呼ばれる手法を用い、プロ野球投球データを対象とした反実仮想説明の選択・抽出手法を提案する。具体的には、打席ごとの結果を予測する LSTM モデルを構築し、そこから DiCE を用いて大量の反実仮想説明を生成する。次いで、提案する抽出基準を適用し、少数の反実仮想説明を抽出する。提案する抽出基準には、正確性・適合性・典型性・実現容易性の4つの因子を考慮する。4つの因子を用いて各反実仮想説明を数値化した後、主成分分析に基づき総合力が高いものを抽出する。

提案する抽出手法の有用性を検証するため、有識者へのアンケートによる定性評価と、因子の基本統計量に対する定量評価を行った。その結果、提案した因子を考慮した反実仮想説明の方が、状況や投手個人の傾向を考慮した上で受け入れやすいとされ、提案手法の有用性が確認された。