信頼される AI の基盤技術
2020 年度採択研究者

2020 年度
年次報告書

原　聡

大阪大学産業科学研究所
准教授

機械学習モデルとユーザのコミュニケーション：モデルの説明と修正

§1. 研究成果の概要

1つ目の研究項目「“説明”のための関連性指標の開発と性能検証」について、「指標の設計」および「性能の評価指標の検討」に取り組んだ。「指標の設計」では複数の関連性指標およびデータ表現ベクトル間での“類推”の性能差を調査するための検証実験を行った。結果、関連性指標としては「コサイン類似度」が、表現ベクトルとしては「損失関数のパラメータ勾配」が有効であることがわかった。「性能の評価指標の検討」ではまず実験の準備として複数の関連性指標で“類推”の結果を可視化し、手法間で提示されるデータの傾向に違いがあることを確認した。