

金子 知適

東京大学大学院情報学環
准教授

思考する AI とのコミュニケーションの実現

§ 1. 研究成果の概要

AI の思考や判断の論拠を人間が理解できるようにするためには、AI の思考記録である探索の記録(巨大グラフ)を、該当分野の専門家が通常行う自然言語での説明と対応付け、それを元に質問応答を行うシステムに接続することが必要である。

画像認識の分野でニューラルネットワークの判断根拠を提示する技術がある。これを囲碁や将棋の AI で用いられるニューラルネットワークに応用して、どこを見て着手を判断したかを提示する技術を 2018 年度より開発中であり、一部、判断根拠の可視化に成功するなど、より分かりやすい手法を研究中である。引き続き、強化学習のための適切なデータ作成や、より説明が分かりやすい AI になるように学習過程での改良などを進めていく。

