

「新しい社会システムデザインに向けた情報基盤技術の創出」  
2016 年度採択研究者

2018 年度  
実績報告書

金子 知適

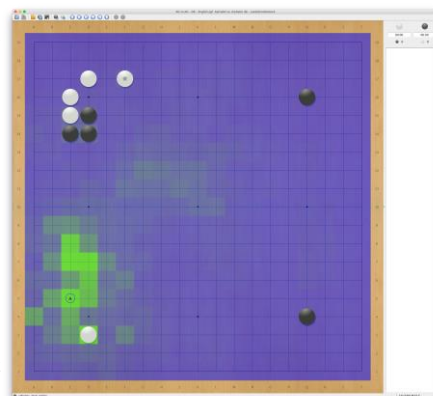
東京大学大学院情報学環  
准教授

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。

画像認識の分野において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が、本年度は、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の判断根拠を可視化する技術を開発し、一部可視化に成功した。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。



## § 2. 研究実施体制

①研究者:金子 知適 (東京大学大学院情報学環 准教授)

②研究項目

- ・囲碁将棋を題材とした強化学習で作られた AI と人の思考との関連の研究