

# 人狼ゲームにおける明示的推定理由の分析

小村 友希 坂本 航

## 要旨

近年、囲碁や将棋などの完全情報ゲームでは、工知能が人間に勝利するなど、目覚ましい研究成果が報告されている。また、不完全情報ゲームを対象とした研究に対しても多くの注目が集まっている。代表的な不完全情報ゲームとして、多人数コミュニケーションゲームの一つである人狼ゲームがあげられる。人狼ゲームでは、各プレイヤーには占い師や霊能者などの役職が割り振られ、役職独自の能力を駆使しながらゲームをプレイする。また会話を通じ、自分の役職隠すための騙し合いや他プレイヤーの信頼獲得を行いながら所属陣営の勝利を目指すこととなる。

人狼ゲームを対象とした人工知能、すなわち人狼 AI の構築を考えた場合、精度の高い判断や行動に加え、その根拠を明示化し、人間へとフィードバックできることが望ましい。これらのことを背景に、本研究では、役職推定タスクを対象とし、ツリーモデルの簡略化を通じて推定モデルの解釈と明示的な推定理由の分析を行う。具体的には、2018 年人狼知能大会決勝戦のログデータを対象とし、回帰木アンサンブルを通じた解釈可能モデル構築手法である `defragTrees` と `inTrees` を用いた分析を行った。