

3者間対話上での相手のモデル発見に至る推論構造の分析

ヒューマンエージェントインタラクション研究室
筑波大学 助教 大澤博隆

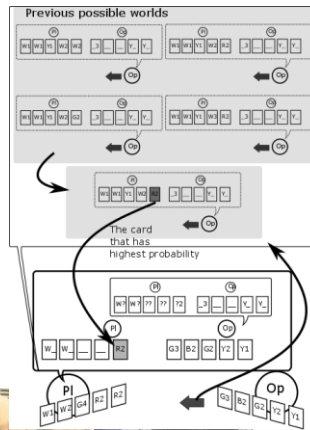
(共同研究者: 国立情報学研究所 教授 佐藤健)



メンバー

- 大澤 博隆(筑波大)
 - ヒューマンエージェント
インタラクション
 - ロボット
 - ゲームシミュレーション

- 共同研究者: 佐藤 健
(NII)
 - 論理学
 - 法学



共通項：コミュニケーション研究



課題：3者間対話の複雑さ

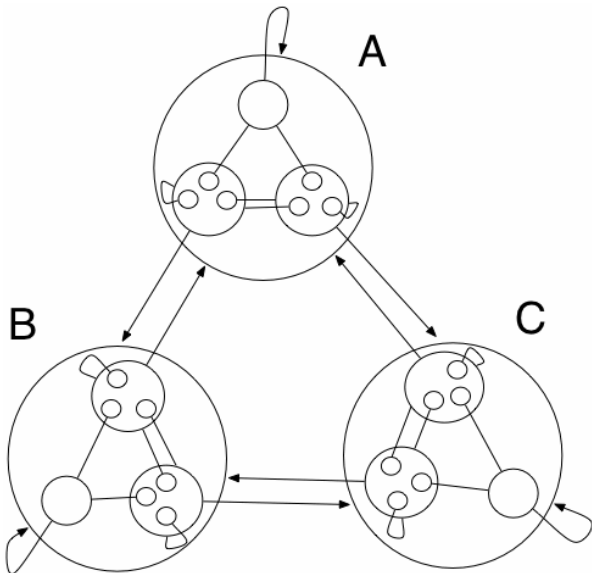
- 2者間対話: 協調か対立か
- 発話は相手への直接的な説得として働く

課題：3者間対話の複雑さ

- 3者間対話:「3人いれば社会ができる」
 - 第三者への説得の利点の発生
 - 「ある発話」が第三者に与える副次的影響の考慮

3者間対話のエージェント応用

- 「対話」だけでないエージェント
 - 漫才エージェント
 - 教育エージェント
 - いじめと共感



課題としての人狼

- コミュニケーションゲーム「人狼」
 - スパイ発見ゲーム
 - 村人の中に隠れた人狼を議論と処刑を用いて当てる
 - 村人: 毎晩一人人間を処刑
 - 人狼: 毎晩一人村人を襲撃
 - 占い師: 毎晩一人対象が人間か狼か見分ける
- 人狼を解く人工知能を作る
 - 人間の代わりに振る舞うエージェント
 - コンテストの形式で行う(チェス・将棋・ロボサッカーと同様)

日本経済新聞
2015年9月14日 (月)

Web刊 速報 ビジネスリーダー マーケット マネー テクノロジー ライフ スポーツ 映像 朝

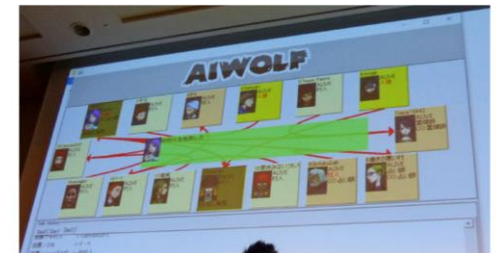
トップ ニュースプラス IT & 経営 eライフ 新技術 & 科学 環境・エネルギー スタートアップ

テクノロジー > ニュースプラス > Tech Frontline > 記事

Tech Frontline [フォローする](#)

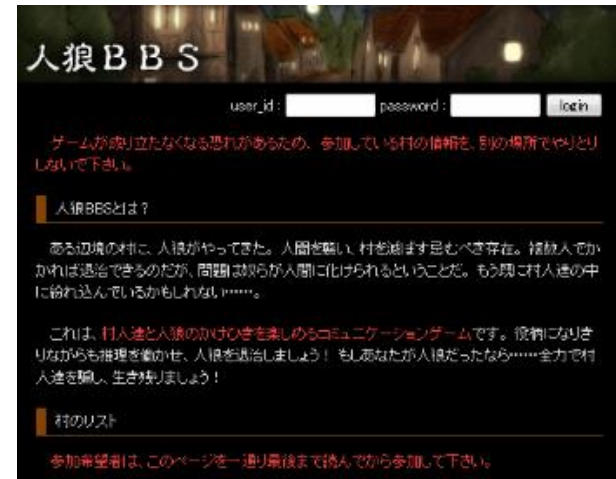
電腦将棋はもう古い「ウソ見破れ」AI最前線
2015/9/13 3:30 (1/2ページ)

人工知能(AI)の進捗度合いを測る「物差し」の一つとして、ながら国内では「将棋」が使われてきた。だが、研究対象としての将棋は終わりを迎えつつある。将棋に代わり、AI分野で注目されているゲームが「人狼」だ。最新ゲームAIの実像と人狼の魅力に迫った。



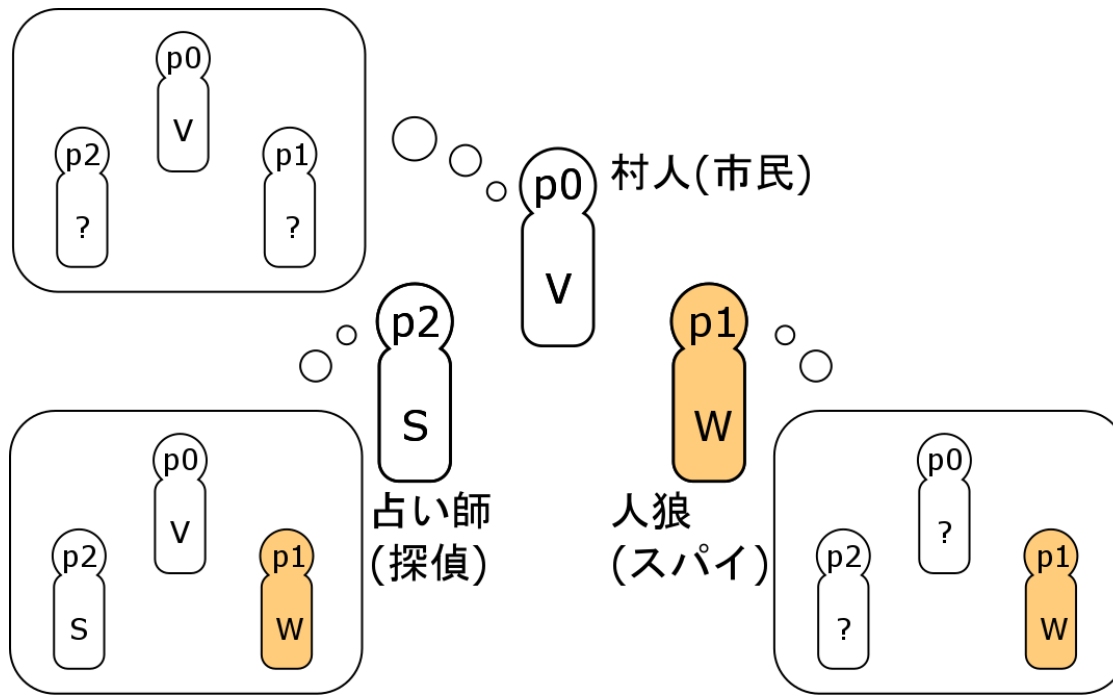
人狼の歴史

- 人狼の歴史
 - 1986年: ソ連で誕生
 - Mafia(Russia/USSR)
 - 2001年: 標準化
 - Are you werewolf? (USA)
 - 2002年～: 役職の誕生
 - Lupus in Tabula (Italy)
 - The Werewolves of Miller's Hollow (France)
 - 2003年～: オンライン人狼
 - 人狼BBSとそのクローン
 - 匿名化・長期化・人狼側のディスカッション
 - 2012年～: 魅せる人狼
 - 人狼演劇(TLPT)
 - テレビ放送・映画等



研究手法

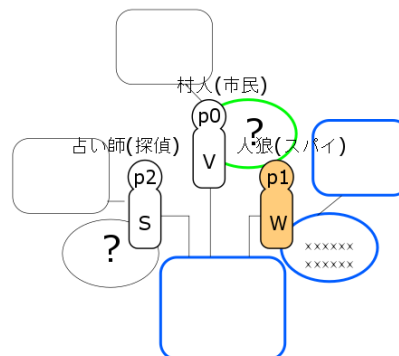
- 3人の人狼セットの展開を探索・分析する
 - 村(市民), 狼(スパイ), 占(探偵)
 - 探偵はスパイの場所を知っている
 - スパイは探偵の場所を知らない
 - 市民は探偵もスパイも知らない



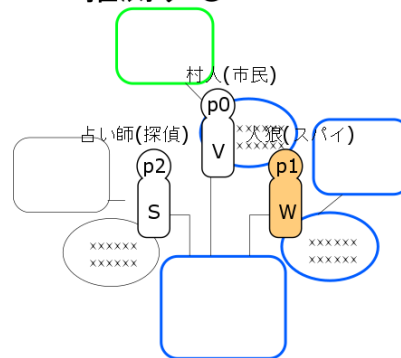
探索手法

- プロトコルを制限し、数え上げる
 - Xはaである、という発言に制限
- ルールセットをどこまでいれるかにしたがって勝ち負けが変化

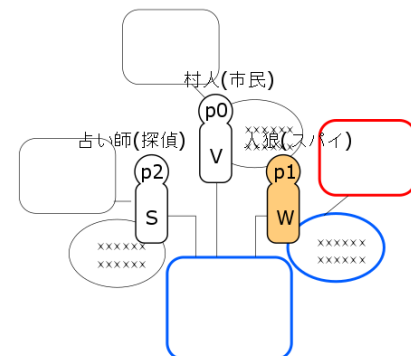
帰納：
ルールセットがわかってない状態でルールセットを推測する



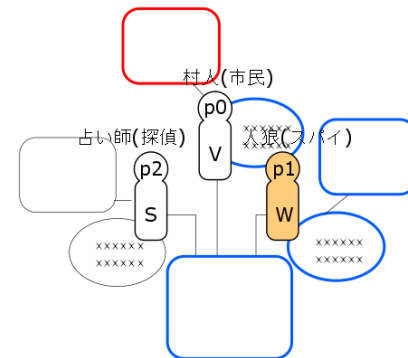
アブダクション：
共有知識＋自分の信念から相手の内部状態（ロール含む）を推測する



演繹：
ルールセットがわかってる状態で自分の内部状態および行為を変更する



説得：
共有知識＋自分の信念から、相手にして欲しい行為をさせる



探索例

* talk

Player 0 (villager) talks "Player 1 is villager."

Player 1 (wolf) talks "Player 1 is villager."

Player 2 (seer) talks "Player 1 is seer."

Player 2 (seer) talks "Player 1 is wolf."

Player 1 (wolf) talks "Player 0 is villager."

Player 0 (villager) talks "Player 1 is wolf."

* vote

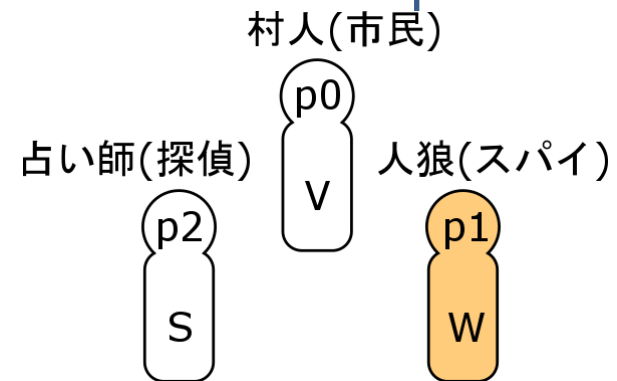
Player 0 (villager) voted to Player 2

Player 1 (wolf) voted to Player 0

Player 2 (seer) voted to Player 1

Villagers draw.

Wolves draw.



rate, win, lose, draw, invalid

0.629, 4094472, 2412190, 205128, 713288210