

報酬最大化 A G I のための 意思疎通機構の設計と プロトタイプ実装

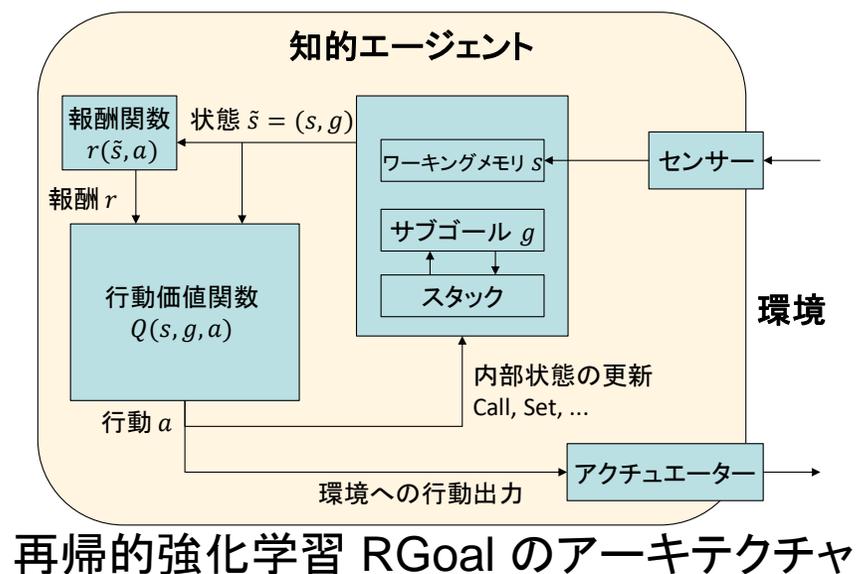
第21回 汎用人工知能研究会 (SIG-AGI)

2022-7-14

一杉裕志、中田秀基、高橋直人、竹内泉（産総研）、
佐野崇（東洋大学）

中期的な研究構想 [一杉 第18回汎用人工知能研究会 2021]

脳をヒントにして、再帰的強化学習、**プログラム合成**、生成モデルの3つを
中核技術とした脳型AGIアーキテクチャの構築を目指す

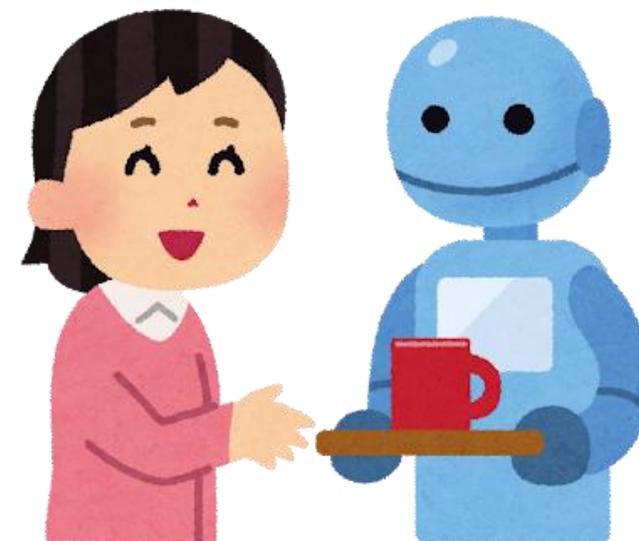


合成しやすく、かつ十分な表現力を持つ合成対象言語の設計が必要

→ 合成対象言語 Pro5Lang を設計中 [一杉+ 第21回汎用人工知能研究会 2022]

本発表の研究の動機

- 将来的には、自然言語を通じて人間と自在に意思疎通できる機械を作りたい。
- 前提とする仮説：
報酬最大化原理 “Reward is enough” [Silver+ 2021]
 - 言語活動を含む様々な知能は報酬最大化という目的から生じる。
- 合成対象言語 Pro5Lang を拡張し、エージェントどうしが意思疎通する機構を追加
- テストプログラムを書いて妥当性を検証中



Pro5Lang の特徴

[一杉+ 第21回汎用人工知能研究会 2022]

- 強化学習でプログラム合成するための、合成対象言語
- 論理型言語と機械語の特徴
 - 制約 1 : レジスタと連想記憶装置に証明済み命題のみが保存可能
→ メモリアクセスを制約
 - 制約 2 : データと行動ルールが固定長の整数ベクトル
→ 教師なし学習によるデータ構造とプログラムが獲得可能
- このような制約のもとでヒトのような会話が実現可能か？
→ POMDP や関連性理論との関係についても考察
我々のアプローチは有望そう

部分観測マルコフ決定過程 (POMDP) と Pro5Lang

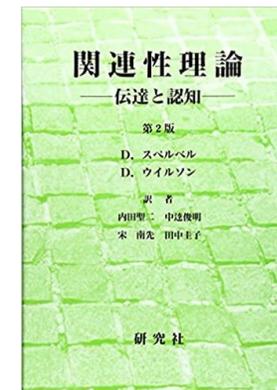
- 直接観測できない状態があるような状況を定式化
- 環境モデルがある場合は、信念状態を状態とみなせば MDP になり解ける。
belief MDP [Kaelbling+ 1998]
 - **信念状態** : エージェントが推定する環境の状態の確率分布
- POMDP は情報を得るための行動と世界を変える行動を統一的に扱う枠組み
 - エージェントは能動的に環境を観測したり [Kaelbling+ 1998]
他者に聞いたり [Williams+ 2007] する。
- Pro5Lang は POMDP を近似的に解いている。
 - 変数の値が unknown → その変数は信念状態において一様分布
 - 行動ルール集合 (プログラム) は belief MDP の行動価値関数を近似表現
 - サブタスク共有・時間抽象・状態抽象の3つの恩恵により学習を効率化

関連性理論 [Sperber and Wilson 1986]

- 言語学における語用論に属する理論
- 発話がいかに理解されるかということに関する理論
- (説明が定性的、独自用語が多くとっつきにくい)

- POMDP を用いた定式化 [Kaiwen+ 2022]

- 発話は主に利他的行為
- 話し手は、聞き手の信念状態の価値を高める目的で発話
- 聞き手は、その前提で発話者の意図を推論



関連性の定式化の一例 [Kaiwen+ 2022]

Kaiwen Jiang, Annya L. Dahmani, Stephanie Stacy, Boxuan Jiang, Federico Rossano, Yixin Zhu, and Tao Gao. What is the point? a theory of mind model of relevance. In Proceedings of the Annual Meeting of the Cognitive Science Society, Vol. 44, 2022. <https://escholarship.org/uc/item/7hz7p7f3>

発信者と受信者が共有する知識

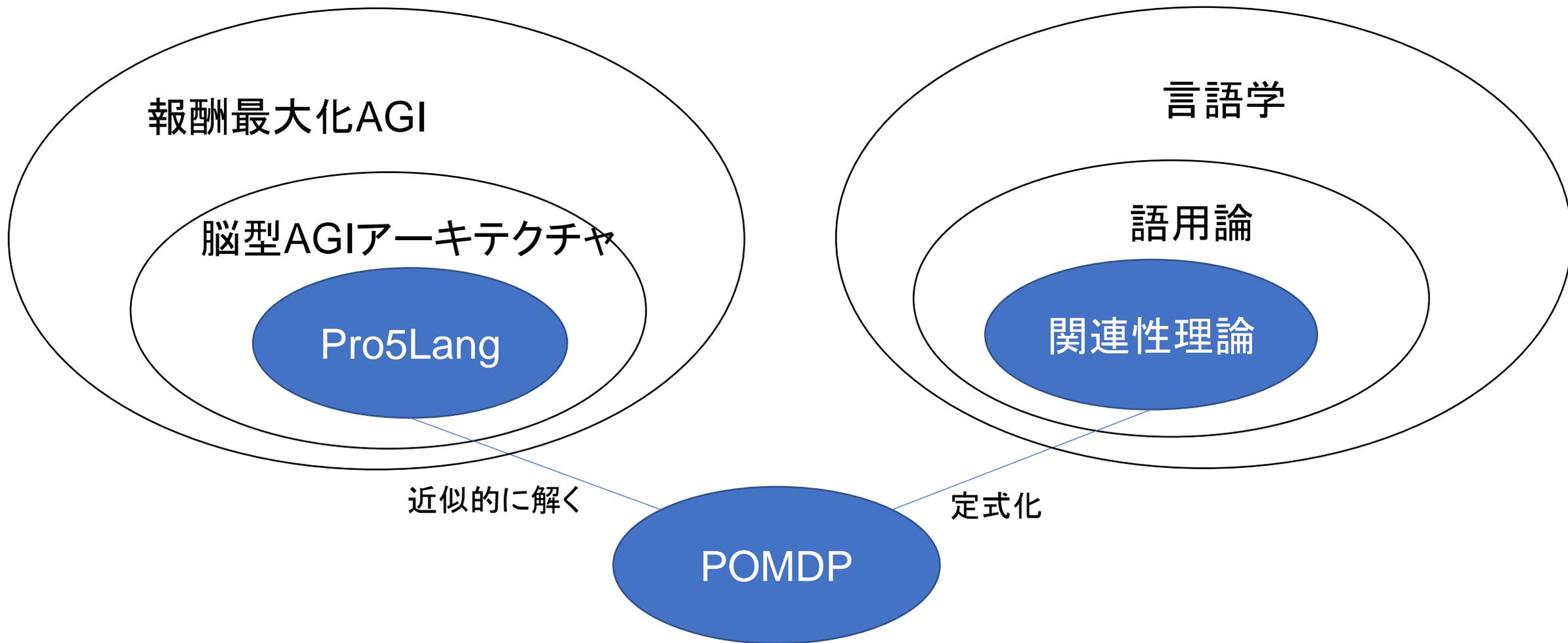
受信者の信念状態 b における行動 a の価値: $Q(b, a)$
信号を伝えなかったときの受信者の信念状態: b_{Rec}
その時の受信者の行動: $a_{Rec} = \underset{a}{\operatorname{argmax}} Q(b_{Rec}, a)$

発信者による信号の選択

発信者の信念状態: b_{Sig}
信号 u を受信後の受信者の信念状態: $b'_{Rec} = P(s|b_{Rec}, u)$ # 受信者の心のモデル
信号受信後の受信者の行動: $a'_{Rec} = \underset{a}{\operatorname{argmax}} Q(b'_{Rec}, a)$
信号 u の受信者にとっての効用: $U(u) = Q(b_{Sig}, a'_{Rec}) - Q(b_{Sig}, a_{Rec})$
信号 u を選択する確率: $P(u|b_{Sig}) \propto e^{\alpha U(u)}$

受信者による信号の解釈

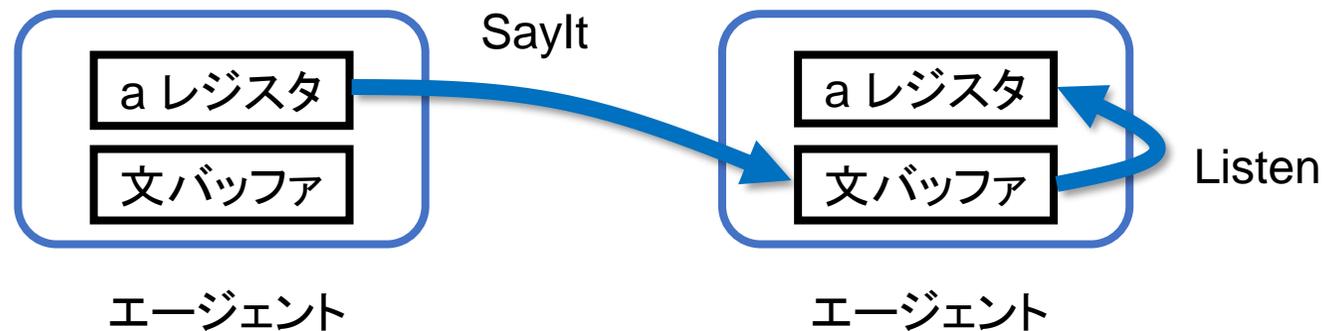
$P(b_{Sig}|u) \propto P(u|b_{Sig})P(b_{Sig})$ # 発信者の心のモデルを使って信号の意味を推論



提案した意思疎通機構は関連性理論をより一般化した、
ヒトの対話の計算論的モデルの有望な候補では？

提案する意思疎通機構

- Pro5Lang に新たに追加した行動プリミティブ：
 - Listen
 - エージェントの**文バッファ**に値がなければ、何もしない待つ。
 - 文バッファに値が書き込まれていればその内容を a レジスタにコピーする。
 - SayIt
 - a レジスタに置かれている発話準備内容を発話相手の文バッファに書き込む。



注:

- 音声や文字列ではなく、発話の内部表現を直接やりとり
- プリミティブは脳では前頭前野の下位の領野が実行すると想定

テストプログラム

対話のシナリオ

Alice はハサミの場所が知りたいと思う。
Alice はハサミの場所を誰が知っているか考える。
Alice は Bob が知っていると思い出す。
Alice は「Bob, ハサミはどこにある?」と聞く。
Alice は返事を待つ。
Bob は「ハサミは Room1 にある」と返事する。
Alice はハサミは Room1 にあると知る。

```
Alice の行動ルール:
taskGoal(w(a(Now, PLS, Scissors, Exists, 0, 0)));
k(e(That01, c(Now, Here, Bob, Knows, 0, 0),
    c(Now, Wh, Scissors, Exists, 0, 0)));

// ハサミがどこにあるか知っている状態。
StateN g1 = w(a(Now, PLS, Scissors, Exists, 0, 0));
// 「ハサミがどこにあるかを誰が知っているか」を知っている状態。
StateN g2 = w(a(That01, c(Now, Here, PLS, Knows, 0, 0),
    c(Now, Wh, Scissors, Exists, 0, 0)));

// サブルーチン g1
rule(w(), g1, call(g2));
rule(w(a(That01, c(Now, Here, z, Knows, 0, 0),
    c(Now, Wh, Scissors, Exists, 0, 0))),
    g1,
    set(a(That02, c(Now, Here, Alice, WantsToSayTo, z, 0),
        c(Now, Wh, Scissors, Exists, 0, 0))));
rule(w(a(That02, c(Now, Here, Alice, WantsToSayTo, z, 0),
    c(Now, Wh, Scissors, Exists, 0, 0))), g1, SayIt);
rule(w(a(That02, c(Now, Here, Alice, SaidTo, z, 0),
    c(Now, Wh, Scissors, Exists, 0, 0))), g1, Listen);
rule(w(a(That02, c(Now, Here, __, SaidTo, Alice, 0),
    c(Now, x, Scissors, Exists, 0, 0))),
    g1,
    set(a(Now, x, Scissors, Exists, 0, 0)));

// サブルーチン g2
rule(w(), g2, recall(e(That01, c(Now, Here, PLS, Knows, 0, 0),
    c(Now, Wh, Scissors, Exists, 0, 0))));
rule(w(e(That01, c(Now, Here, z, Knows, 0, 0),
    c(Now, Wh, Scissors, Exists, 0, 0))),
    g2,
    set(a(That01, c(Now, Here, z, Knows, 0, 0),
        c(Now, Wh, Scissors, Exists, 0, 0))));

Bob の行動ルール:
rule(w(), w(), Listen);
// 聞かれたことに即座に返答。
// 「ハサミの場所はどこ」と聞かれたら「Room1 にある」という。
rule(w(a(That02,
    c(Now, Here, Alice, SaidTo, Bob, 0),
    c(Now, Wh, Scissors, Exists, 0, 0))),
    w(),
    set(a(That02,
        c(Now, Here, Bob, WantsToSayTo, Alice, 0),
        c(Now, Room1, Scissors, Exists, 0, 0))));
rule(w(a(That02,
    c(Now, Here, Bob, WantsToSayTo, __, __),
    c(__, __, __, __, __, __))), w(), SayIt);
```

テストプログラム

対話のシナリオ

Alice はハサミの場所が知りたいと思う。
Alice はハサミの場所を誰が知っているか考える。
Alice は Bob が知っていると思い出す。
Alice は「Bob, ハサミはどこにある?」と聞く。
Alice は返事を待つ。
Bob は「ハサミは Room1 にある」と返事する。
Alice はハサミは Room1 にあると知る。

目的を持った発話

発話計画

心の理論

通信の同期

発話理解

Alice の行動ルール:

```
taskGoal(w(a(Now, PLS, Scissors, Exists, 0, 0)));
k(e(That01, c(Now, Here, Bob, Knows, 0, 0),
    c(Now, Wh, Scissors, Exists, 0, 0)));

// ハサミがどこにあるか知っている状態。
StateN g1 = w(a(Now, PLS, Scissors, Exists, 0, 0));
// 「ハサミがどこにあるかを誰が知っているか」を知っている状態。
StateN g2 = w(a(That01, c(Now, Here, PLS, Knows, 0, 0),
    c(Now, Wh, Scissors, Exists, 0, 0)));

// サブルーチン g1
rule(w(), g1, call(g2));
rule(w(a(That01, c(Now, Here, z, Knows, 0, 0),
    c(Now, Wh, Scissors, Exists, 0, 0))),
    g1,
    set(a(That02, c(Now, Here, Alice, WantsToSayTo, z, 0),
        c(Now, Wh, Scissors, Exists, 0, 0))));
rule(w(a(That02, c(Now, Here, Alice, WantsToSayTo, z, 0),
    c(Now, Wh, Scissors, Exists, 0, 0))), g1, SayIt);
rule(w(a(That02, c(Now, Here, Alice, SaidTo, z, 0),
    c(Now, Wh, Scissors, Exists, 0, 0))), g1, Listen);
rule(w(a(That02, c(Now, Here, __, SaidTo, Alice, 0),
    c(Now, x, Scissors, Exists, 0, 0))),
    g1,
    set(a(Now, x, Scissors, Exists, 0, 0)));

// サブルーチン g2
rule(w(), g2, recall(e(That01, c(Now, Here, PLS, Knows, 0, 0),
    c(Now, Wh, Scissors, Exists, 0, 0))));
rule(w(e(That01, c(Now, Here, z, Knows, 0, 0),
    c(Now, Wh, Scissors, Exists, 0, 0))),
    g2,
    set(a(That01, c(Now, Here, z, Knows, 0, 0),
        c(Now, Wh, Scissors, Exists, 0, 0))));
```

Bob の行動ルール:

```
rule(w(), w(), Listen);
// 聞かれたことに即座に返答。
// 「ハサミの場所はどこ」と聞かれたら「Room1 にある」という。
rule(w(a(That02,
    c(Now, Here, Alice, SaidTo, Bob, 0),
    c(Now, Wh, Scissors, Exists, 0, 0))),
    w(),
    set(a(That02,
        c(Now, Here, Bob, WantsToSayTo, Alice, 0),
        c(Now, Room1, Scissors, Exists, 0, 0))));
rule(w(a(That02,
    c(Now, Here, Bob, WantsToSayTo, __, __),
    c(__, __, __, __, __, __))), w(), SayIt);
```

Alice の行動ルール集合（プログラム）の疑似コード

g1 = <ハサミは PLS にある> // PLS は unknown 以外の値、サブゴール: 「ハサミがどこにあるか知る」

rule(_, g1, call(<ハサミがどこにあるかを PLS が知っている>))

rule(<ハサミがどこにあるかを z が知っている>, g1, set(<自分は z にハサミがどこにあるか聞きたい>))

rule(<自分は z にハサミがどこにあるか聞きたい>, g1, SayIt)

rule(<自分は z にハサミがどこにあるか聞いた>, g1, Listen)

rule(<_ が自分に「ハサミは x にある」と言った>, g1, set(<ハサミは x にある>))

レジスタの状態のパターン
(対話の状態遷移を制御)

サブゴール

アクション

このプログラムの拡張性

g1 = <ハサミは PLS にある>

rule(_, g1, call(<ハサミがどこにあるかを PLS が知っている>))

誰かに聞く以外の選択肢：
自分で考える、自分で探す、など。

rule(<ハサミがどこにあるかを z が知っている>, g1, set(<自分は z にハサミがどこにあるか聞きたい>))

rule(<自分は z にハサミがどこにあるか聞きたい>, g1, SayIt)

他の聞き方の検討：
「ハサミちょうだい」「ハサミの場所知らない？」「誰かハサミを使った？」など。

rule(<自分は z にハサミがどこにあるか聞いた>, g1, Listen)

rule(<_ が自分に「ハサミは x にある」と言った>, g1, set(<ハサミは x にある>))

他の応答への対処：
「きのう Room2 で見た」
「自分は知らないけど Carol が知っているかもしれない」など。

他の解釈の検討：
間違っているかもしれない、ウソをついているかもしれない、など。

まとめと今後

- 合成対象言語 Pro5Lang に意思疎通機構を追加
 - テストプログラムを動かして設計の妥当性を検討中
 - 発話計画も発話理解も合成したプログラムが行う
 - 柔軟かつ合目的的な振る舞いになる見込み
 - 推論結果や対話履歴が大容量のエピソード記憶に保存
 - 内容豊かな会話になる見込み
- POMDP や関連性理論との関係について考察
 - 提案機構は、より一般的なヒトの対話の計算論モデルになる可能性
- 今後さらに必要な拡張：
 - 複雑な環境下での対話
 - タイムアウト、割り込み、メモリへのポインタの機構
 - ニューラルネットのような非記号的な情報処理による補助

以上

関連性理論

語用論のすべて — 生成文法・認知
言語学との関連も含めて —
今井 邦彦 他, 開拓社, 2021.



- p. 87 「（情報が）関連性を持つ3つの場合をまとめると以下のようになります。」
 - a. その人自身にとって確信が持てなかった想定（≡考え）がその情報によって確信へと変わる（強化される）場合
 - b. その人の想定が間違っていたことがその情報によって明らかになったため、その想定を廃棄する場合
 - c. その情報がコンテキスト的含意を持つ場合
（「情報が関連性を持つ＝聞き手の信念状態の価値が高まる」？）
- p. 91 **関連性原理I**
 - 「人間の認知は関連性を最大にするように働く性質を持つ。」
 - 認知とは「様々な想定を持っている状態、想定を増加・改善させたいという要求、想定を増加・改善する場合の頭の働き」。
（「ヒトの脳は信念状態の価値を高める方向に動作する。」？）
- p. 96 **関連性原理II**
 - 「すべての意図明示的伝達行為は、それ自身が最適な関連性を持つことを当然視している旨を伝達している。」
 - 「つまり、発話をする事自体、「私の話を聞けば、解釈に unnecessary な労力をかけることなく、関連性のある情報が得られますよ」ということを伝えていることになるのです。それを表したものが、次の関連性原理IIなのです。」
（「発話は利他的行為である。」？）

関連性の定式化の一例 [Kaiwen+ 2022]

Kaiwen Jiang, Anya L. Dahmani, Stephanie Stacy, Boxuan Jiang, Federico Rossano, Yixin Zhu, and Tao Gao. What is the point? a theory of mind model of relevance. In Proceedings of the Annual Meeting of the Cognitive Science Society, Vol. 44, 2022. <https://escholarship.org/uc/item/7hz7p7f3>

発信者と受信者が共有する知識

受信者の信念状態 b における行動 a の価値: $Q(b, a)$
信号を伝えなかったときの受信者の信念状態: b_{Rec}
その時の受信者の行動: $a_{Rec} = \underset{a}{\operatorname{argmax}} Q(b_{Rec}, a)$

発信者による信号の選択

発信者の信念状態: b_{Sig}
信号 u を受信後の受信者の信念状態: $b'_{Rec} = P(s|b_{Rec}, u)$ # 受信者の心のモデル
信号受信後の受信者の行動: $a'_{Rec} = \underset{a}{\operatorname{argmax}} Q(b'_{Rec}, a)$
信号 u の受信者にとっての効用: $U(u) = Q(b_{Sig}, a'_{Rec}) - Q(b_{Sig}, a_{Rec})$
信号 u を選択する確率: $P(u|b_{Sig}) \propto e^{\alpha U(u)}$

受信者による信号の解釈

$P(b_{Sig}|u) \propto P(u|b_{Sig})P(b_{Sig})$ # 発信者の心のモデルを使って信号の意味を推論