システム発話間の整合性を重視した発話選択への深層強化学習の適用

黒田佑樹　武田龍　駒谷和範 (大阪大学　産業科学研究所)

研究概要

【大目標】

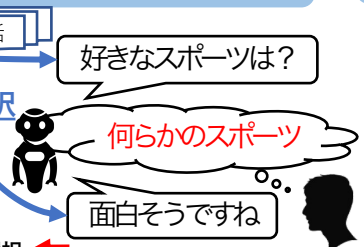
聞き役対話システムの実現

- システム発話列のみをコントロール

【小目標】

- 文脈的に適切な発話選択
  - ➢ システム発話間の整合性を重視した発話選択
    →以前Q学習によって実装
- 将来的により多くのユーザ状態を考慮したい
  - ➢ 表形式でQ関数を表現するのは困難(Q学習)
  - ➢ Q関数をニューラルネットで表現

深層強化学習



ユーザ発話は
ある程度予測可能

↓

システム発話の
整合した出力順を考える

従来: システム発話間の整合性を重視した発話選択

状態s: 38×2×3 = 228状態

発話内容(発話ID): 38状態 ← より細かく分類したもの　対話行為: 8状態

S: 競技は何をご覧になりますか？(指示語なし質問)
U: 野球ですね(心象: 高)

特定名詞の有無: 2状態　心象: 3状態

※発話内容(発話ID)の状態に関しては一部過去の内容も考慮して定義

システム

ユーザモデル

↓ 直前の交換の状態から次のシステム発話を決定

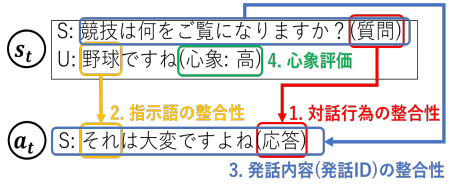
行動a: 35個

発話内容: 35個

S: それは大変ですよね(応答)

報酬の与え方

S:システム
U: ユーザ

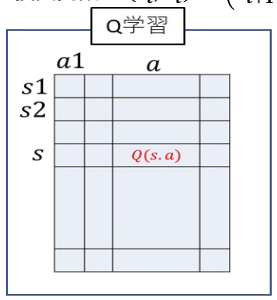


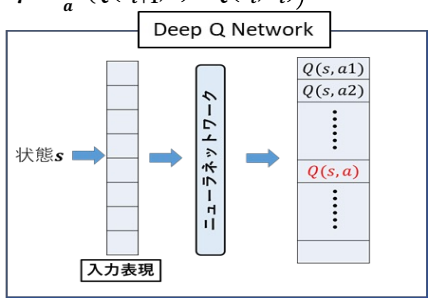
状態と行動の関係から報酬決定

1. 対話行為の整合性
   - 例; 質問→応答なら正の報酬
2. 指示語の整合性
   - 例: 指示語の対象があれば正の報酬
3. 発話内容の整合性
   - あらかじめ人手で設定
4. 心象
   - 例: 高いなら正の報酬

深層強化学習(DQN)を用いた実装

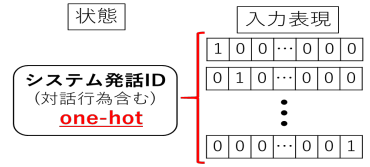
- Q学習: Q値を1ターンごとに更新
  - ➢ 更新式: $Q(s_{t+1}, a_{t+1}) = Q(s_t, a_t) + \alpha\left(r_{t+1} + \gamma * \max_a(Q(s_{t+1}, a) - Q(s_t, a_t)\right)$
    t: 現在のターン　$t+1$: 次のターン　γ: 割引率　α: 学習率
- DQN: 状態を入力，Q値を出力としたニューラルネットワークを学習
  - ➢ 損失関数: $E(s_t, a_t) = \left(r_{t+1} + \gamma * \max_a(Q(s_{t+1}, a) - Q(s_t, a_t)\right)^2$

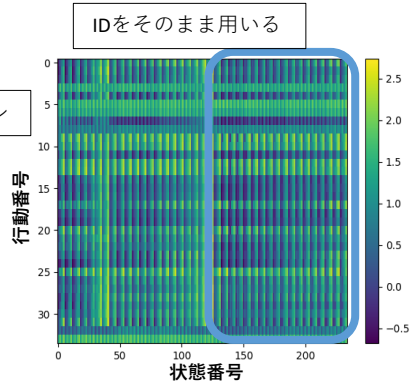




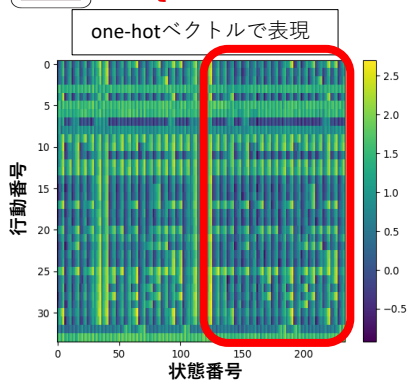
入力表現

- システム発話ID(対話行為含む)
  →IDをそのまま用いる×
  - 0~37の整数番号を0~1に正規化
  →one-hotベクトルで表現○
- ユーザ発話の特定名詞の有無
  →one-hotベクトルで表現
- ユーザ心象
  →離散値で表現





×異なる状態でも同じようなQ値
=同じような行動(発話)ばかり選択



○状態ごとにQ値が異なる
=状態に応じた行動ができる
実際，Q学習の場合と近いQ値

実験: DQNを用いた従来手法の再現

実験目的

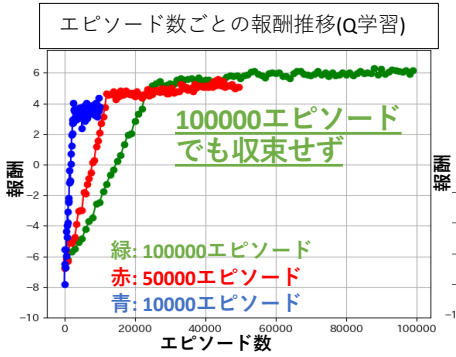
- システム発話間の整合性を重視した発話選択をDQNによって再現できるか
  - ➢ Q学習DQN各々の学習結果を用いて10交換の対話を10回行い，破綻数で比較
- Q学習を用いた場合とDQNを用いた場合で学習過程に違いはあるか

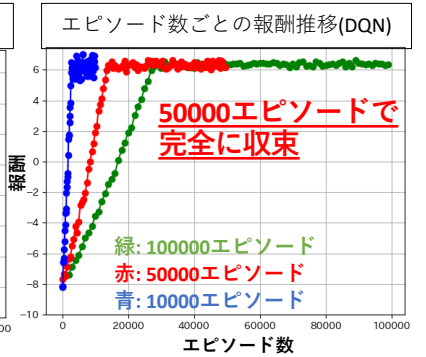
実験条件

- 選択対象の発話: 雑談対話コーパスHazumi1902から70発話程度抜粋
  - ➢ 発話内容に応じてID付けして状態と行動に使用
    →状態: 38状態　行動: 35個 にそれぞれ分類してID付け
- 学習環境: ユーザモデルを用いる
  - ➢ 選択システム発話に対してコーパス収集時のユーザ発話を返す
  - ➢ 10交換で1エピソード

十分な学習に必要なエピソード数

- 十分に学習できたとは: 報酬が収束した(上がり切った)状態
  - ➢ 報酬が上がらなくなるエピソード数を探る
- 探索方法: epsilon-greedy法
  - ➢ epsilonは初期値1．エピソード数の4分の1で0.1に収束するように減衰．





対話例と破綻数による評価

	破綻でない	破綻
Q学習	94	6
DQN	95	5

同じくらいの破綻数
&
同じようなQテーブル

DQNで適切に学習し，再現できた

S:システム
U: ユーザ

S1: 競技は何をご覧になりますか？
U1: 野球を見ます
S2: どういった所が好きなんですか?
U2: データが豊富なところです
S3: もう少し詳しく教えてください
U3: 打率とか，出塁率とかですね

↓ システム発話選択(S4)

OK: そうなんですか．
　一度独自に調べてみたいと思います．
破綻: それは残念です．