深層強化学習を用いたシチュエーション対話向け応答選択モデル

佐藤真　高木友博
明治大学

目次

Akitakaの開発方針

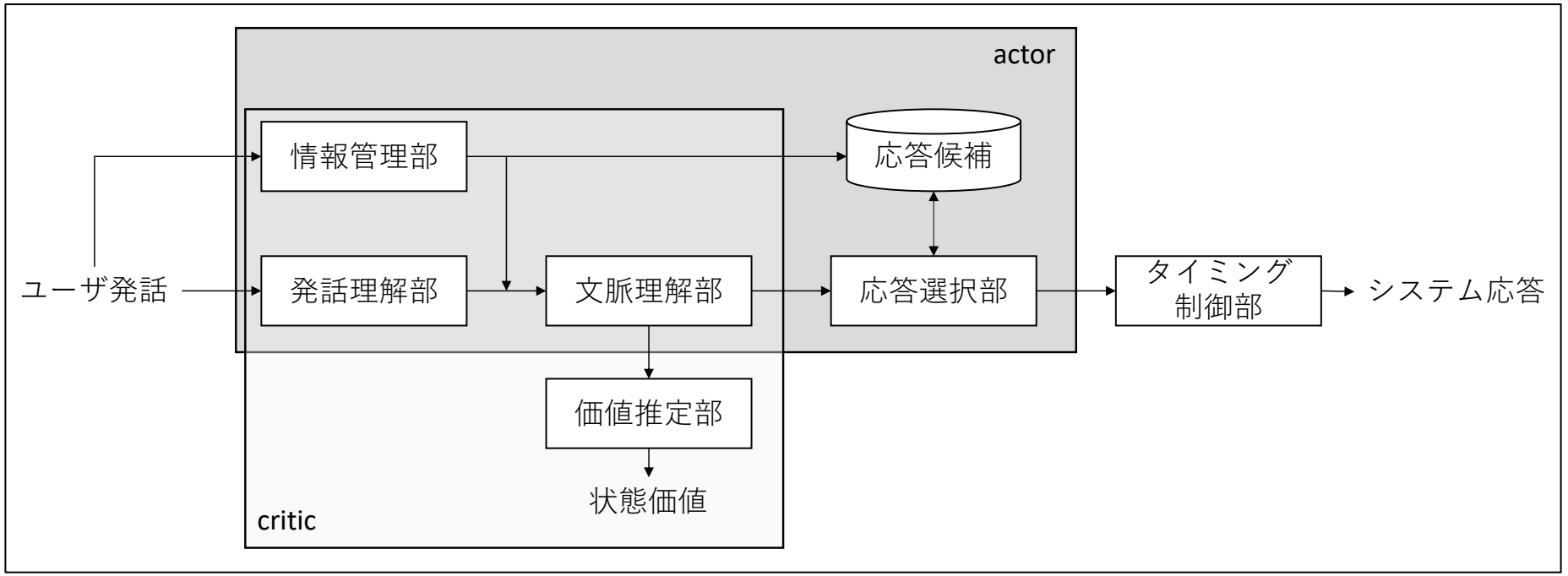
・評価基準である「人らしさ」の実現
  - 人らしい発話文であること
  - 文脈に沿っていること
  - 対話主導権の交代への対応

・機械学習手法の使用

・省データであること

Akitakaシステム概要

- 性別　　：男性
- 対話方式　　：機械学習を利用した応答選択
- 学習手法　　：教師ありによる事前学習 + 強化学習
- 利用した知識源: 人手で作成した対話データセット



Akitakaのシステム概要図

データセット

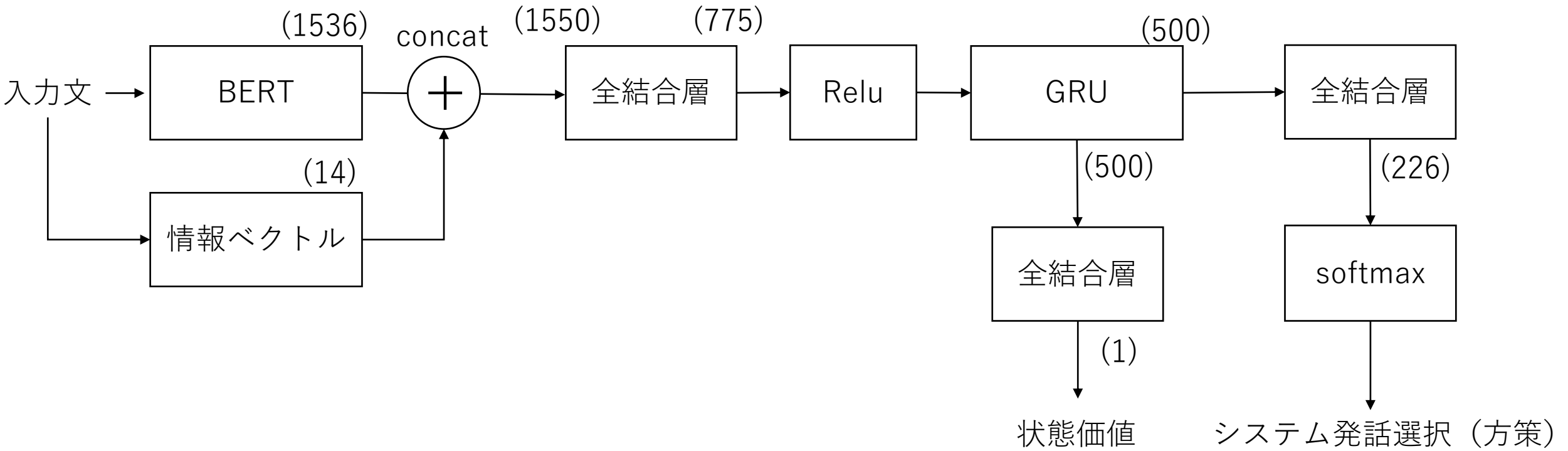
・課題に沿った対話データを49件作成
  - 研究室のメンバーや友人にご協力いただきました
  - 様々な話題や対話主導権の交代を含むように意識
  - データ収集用の用例ベース対話システムを構築・利用
    - 望ましくないシステム応答だった場合はユーザ自身で修正(1人2役)

・事前学習や応答候補群作成に利用
  - 一貫性を感じさせるため、口調統一等のクリーニング
  - 応答候補は226文

モデル構成

・BERTには京都大学黒橋・褚・村脇研究室様が公開されている
事前学習済みモデルを使用

・情報抽出にはルールと日本語NLPライブラリのGINZAを併用



深層強化学習

・学習アルゴリズムはPPO(Proximal Policy Optimization)を使用

・ユーザシュミレータの作成と報酬関数の設計は困難
- 人手(一名)によってユーザ発話作成、報酬付与判断
  - 報酬付与は文脈に沿っているか,社会性に反していないかで判断(0.1 or 0)

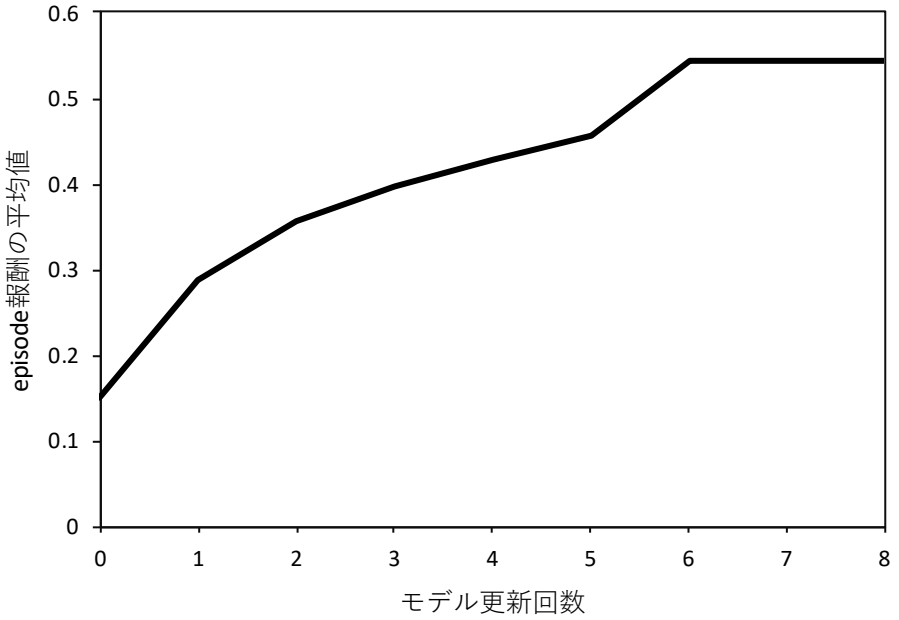
・望ましい応答であれば報酬を与えて対話継続、望ましくなければ終了
- 望ましい応答かに関わらず15ターンで対話を終了
- 1558 episode の対話を行い, 合計 5385 step(ターン)を収集

深層強化学習

・作成したデータセットで事前学習(15 epoch)
  - 対話履歴から次に実際に行われた発話を予測する教師あり学習
  - 方策(actor)部分のみ

・方策部分は事前学習した重み、
  それ以外は LeCun normal で初期化

・学習環境
  - optimizer : RAdam
  - 更新間隔　: 6回更新するまで 512
                以降 768 step



ルールによるマスク

・学習した方策だけでは対話破綻が発生

- 状況に応じて、ルールにより応答候補を制限

・ルールは３種

- 使用に適さない候補に対するマスク：そもそも利用に適さない(候補全体 226→219)

- 状況に適さない候補に対するマスク：情報ベクトルを参照。32候補対象

- 使用した候補に対するマスク　　：158 候補は再使用不可
  それ以外は2ターン使用不可

返信タイミング管理

・入力に対して即座に応答するのは人らしいとは言い難い

・ [“応答文の形態素数※”/2 + 5 + random(2, –2)] 秒分応答を遅延
- 入力時間考慮　　: “応答文の形態素数※”/2
- 内容理解考慮　　: 5
- ランダム性付与 : random(2,-2)

※: 2020年12月03日訂正
　システムの動作に変更はありません

評価

・平均 3.93/5点　シチュエーショントラック予選3位

・半数以上が高評価だが、対話破綻による低評価も目立つ

表 1: クラウドワーカによる評価分布

評価	件数	割合 (%)
5	18	34.0
4	23	43.4
3	3	5.7
2	8	15.1
1	1	1.9

表 2: クラウドワーカのコメント例

評価	コメント
5	かなり嫌がっているのに押し通すのが忍びないと思うほど、自然なやり取りが出来て感動しました。
4	受け答えがしっかりとできていました。話題が終わってしまいそうになっても、向こうから質問を投げかけようとしていたので、リアリティがありました。
2	話があまりかみ合いませんでした。
1	会話が成り立たないことがしばしばあった。

課題

・訓練資源の追加・ユーザシュミレータの作成

・人らしさの自動評価

・言いずらさ等を考慮したタイミング管理