

## 言語理解・生成部を備えた協調的説得対話システムの構築と評価

## Construction and Evaluation of a Cooperative Persuasive Dialog System with Natural Language Understanding and Natural Language Generation

平岡拓也  
Takuya Hiraokaニュービッググラム  
Graham Neubigサクテイサクリアニ  
Sakriani Sakti戸田智基  
Tomoki Toda中村哲  
Satoshi Nakamura奈良先端科学技術大学院大学 情報科学研究科  
Graduate School of Information Science, Nara Institute of Science and Technology

In this paper, we construct and evaluate a fully automated text-based cooperative persuasive dialogue system, which is able to persuade the user to take a specific action while maintaining user satisfaction. In our previous works, we created a dialogue management module for cooperative persuasive dialogue, but only evaluated it in a wizard-of-Oz setting, as we did not have the capacity for natural language generation and understanding. In this work, the main technical contribution is the design of the natural language understanding (NLU) and the natural language generation (NLG) modules which allows us to remove this bottleneck and create the first fully automatic cooperative persuasive dialogue system. Based on this system, we performed an evaluation with real users. Experimental results indicate that the learned policy is able to effectively persuade the users: the reward of the proposed model is much higher than baselines, and almost the same as a dialogue manager controlled by a human. This tendency is almost the same as our previous evaluation using a wizard-of-Oz framework demonstrating that the proposed NLU and NLG modules are effective for cooperative persuasive dialogue.

## 1. はじめに

これまで、我々は協調的説得対話というパラダイムを提唱してきた [1, 2]. 協調的説得対話では、システムはユーザの目標だけでなく、システム自身の目標も達成するように振舞う。実社会における協調的説得対話の例として、商品販売における販売員と客の対話が挙げられる。一般的には、客は好みに合う商品を購入することを、また販売員は高い利益を得る商品販売することを目標とするであろう。この場合、販売員は、客の好みの商品の中で出来る限り利益を得る商品販売するように対話を進めるであろう。この販売員のように双方の目標を両立して達成するように対話を進める対話システムを協調的説得対話システムと定義する。

先行研究では、我々は協調的説得対話に強化学習を適用し、学習された協調的説得対話システムの方策を Wizard-of-Oz の枠組みに基づいて評価してきた [1]. この研究では、我々は部分観測マルコフ決定過程 (Partially Observable Markov Decision Processes; POMDP) に基づいて協調的説得対話をモデル化し、そのモデルを用いてシステムの方策を強化学習した。また、協調的説得対話をモデル化する際、システムのアクションにフレーミング [3] を導入した。フレーミングは、感情極性を持った語で意思決定候補を修飾する対話行為である。言語理解部と言語生成部を人間の Wizard が担当するシステムを用いて、学習されたシステムの方策の性能評価を行った。しかし、この評価の枠組みでは、自然言語理解部と自然言語生成部の性能による影響が無視されてしまう。そして、評価結果が人間の Wizard の能力に大きな影響を受けてしまう。

本稿では、言語理解部と自然言語生成部を備えた協調的説得対話システムを構築し、その性能を評価する。2. 節では、協調的説得対話システムを構築する際に利用する人同士の対話コーパスについて説明する。3. 節では、協調的説得対話システムが従う方策を学習する際に用いる対話モデルについて説明する。4. 節では、テキストベース協調的説得対話システムの自然言語理解部と自然言語生成部の構築について説明する。そして、

5. 節では、構築したテキストベース協調的説得対話システムの評価について説明する。

## 2. 説得対話コーパスとフレーミング

本稿では、協調的説得対話システム構築に、人同士の説得対話コーパス [4] を利用する。本コーパスは、説得対話の一例として、家電販売店でのカメラ販売における販売員 (説得者) と客 (被説得者) の対話を想定する。販売員は客に対して、複数のカメラ (意思決定候補) の中から特定のカメラ (説得目標) を購入 (意思決定) させることを目的とする。本研究では、収録された 35 対話 (340 分) の模擬対話コーパスを利用する。

コーパスには、ネガティブ/ポジティブフレーミング [3] が付与されている。これらのフレーミングでは、感情極性を持った語で意思決定候補を修飾する。具体的には、ネガティブフレーミングはネガティブな感情極性を、ポジティブフレーミングはポジティブな感情極性を持つ語で意思決定候補を修飾する。本コーパスでは、フレーミングは 3 つ組  $\langle a, p, r \rangle$  で表される。  $a$  は論証の対象である意思決定候補を表す。  $p$  はフレーミングがネガティブの場合 NEG, ポジティブの場合 POS の値をとる。  $r$  は論証中に被説得者の嗜好に合致した決定要因 (例: カメラの性能や値段) への言及が存在するかを表す。  $r$  は言及が存在する場合 TRUE, 存在しない場合 FALSE の値をとる。被説得者の嗜好に合致する決定要因はアンケート結果に基づいて決定する。例えば、コーパス中の「(カメラ A は) ポケットに入る大きさで一眼並みの性能で撮っていただけるといいよ」という販売員の発話には、 $\langle a = A, p = POS, r = FALSE \rangle$  が付与されている。なお、この例では、アンケート結果に基づいて、客の嗜好に合致する決定要因は「カメラの値段」であると想定されている。

また、一般的な対話行為 (例: 質問や情報提示) として、一般目的機能 (GPF) タグ [5] も付与されている。

## 3. 協調的説得対話のモデル化

本節では POMDP に基づいた協調的説得対話のモデル化について説明する。モデルはユーザとシステムのモデルで構成

されており、ユーザ (2. 節の被説得者) のシミュレーションモデルを 3.1 節で、またシステム (2. 節の説得者) のモデルを 3.2 節でそれぞれ説明する。

### 3.1 ユーザシミュレータ

強化学習時の報酬計算のため、ユーザ (2. 節の被説得者) の以下の振る舞いのシミュレータを構築する：

1. ユーザの一般的な対話行為。
2. ユーザへの嗜好の通知。

ユーザの一般的な対話行為は GPF を用いて表わされる。また、ユーザへの嗜好の通知とは、説得者が代替案のフレーミングに引用した決定要因が被説得者の嗜好に合致することである。例えば、表??では、販売員のカメラ A のポジティブフレーミングに“性能”が引用されている。もし、“性能”が被説得者の好みに合致 (i.e.  $\text{pref}=\text{YES}$ ) する場合は嗜好の通知がされたとする。

ターン  $T_{t+1}$  における、ユーザの GPF  $G_{user}^{t+1}$  と嗜好の通知  $C_{alt}^{t+1}$  はそれぞれ以下の確率に基づいて計算される。

$$P(G_{user}^{t+1} | G_{user}^t, F_{sys}^t, G_{sys}^t, S_{alt}^t) \quad (1)$$

$$P(C_{alt}^{t+1} | C_{alt}^t, F_{sys}^t, G_{sys}^t, S_{alt}^t) \quad (2)$$

$G_{sys}^t$  はターン  $T_t$  におけるシステムの GPF,  $F_{sys}^t$  はターン  $T_t$  におけるシステムのフレーミングを表す。これらはいずれもシステムのアクションであり、3.2 節で説明する。 $G_{user}^t$  はターン  $T_t$  におけるユーザの GPF,  $C_{alt}^t$  はターン  $T_t$  における嗜好の通知状態を表す。 $S_{alt}$  は代替案の初期選択である。本研究では、ユーザが最初に嗜好に合致するとして選んだカメラである。

### 3.2 協調的な説得対話方策の学習

本節では、システム (2. 節の説得者) に関するモデルについて述べる。特に、強化学習を行う上で必要な情報である報酬や、システムの行動と信念状態について説明する。

ユーザの満足度 (被説得者の目標の達成度合い)、システムの説得成功 (説得者の目標の達成度合い) と自然性を用いて報酬を設計する。1. 節で述べたような協調的説得対話システムは、ユーザとシステム両方の目的を達成するよう対話を進めなければならない。このような対話の達成度合いを表す各ターンにおける報酬を以下のとおり定義する。

$$r_t = \frac{Sat_{user}^t - \overline{Sat_{user}^t}}{\text{Stddev}(Sat_{user}^t)} + \frac{PS_{sys}^t - \overline{PS_{sys}^t}}{\text{Stddev}(PS_{sys}^t)} + \frac{N^t - \bar{N}}{\text{Stddev}(N)}, \quad (3)$$

$Sat_{user}^t$  は、[0,1] に規格化された、ターン  $t$  における 5 段階のユーザの満足度の主観評価値の (1: Not satisfied, 3: Neutral, 5: Satisfied) を表す。 $PS_{sys}^t$  はターン  $t$  における説得の成功 (1: Success, 0: Failure) の期待値である。 $N_t$  はターン  $t$  におけるシステムとユーザの対話の bi-gram 尤度である。なお、 $Sat_{user}^t$  と  $PS_{sys}^t$  は、先行研究 [4] で得られた予測モデルに基づき、対話状態 (表 1) を利用して計算される。上線付きの変数は上線なしの変数の平均を表す。また、Stddev 関数は引数の標準偏差を表す。これらの統計量は、本節で提案する対話モデルに基づく約 6000 対話 (計 60000 ターン) のシミュレーションを通して計算される。

システムのアクションはフィルタリングされたフレーミングと GPF の組  $\langle F_{sys}, G_{sys} \rangle$  である。これらは 2. 節で述べた販売員 (説得者) の対話行為を表す。フィルタリングのために、実対話コーパスから販売員のユニグラム  $P(G_{sales}, F_{sales})$  を構築する。本研究では、 $P(G_{sales}, F_{sales})$  が 0.005 以下の  $\langle G_{sys}, F_{sys} \rangle$  を削除し、残った 13 個組をアクションとして利用する。

表 1: 報酬計算のための特徴量

$Sat_{user}$	システムの <i>commisive</i> (GPF の一種) の頻度 システムの <i>question</i> (GPF の一種) の頻度
$PS_{sys}$	経過時間 ユーザへの嗜好の通知 $C_{alt}$ ユーザの代替案の初期選択 $S_{alt}$
$N$	システムとユーザの現在のターンの GPF システムとユーザの直前のターンの GPF システムのフレーミング

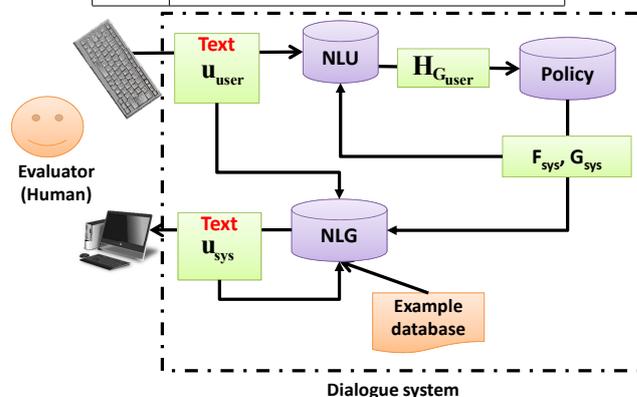


図 1: 我々の協調的説得対話システムの構成。円柱はシステムのモジュールを表しており、四角はそのモジュール間でやり取りされる情報を表している。NLU は言語理解部、NLG は言語生成部、Policy は方策部を表している。

システムの信念状態は、報酬計算に用いた特徴量 (表 1) と報酬で表わされる。ただし、本研究では、システムは  $C_{alt}$  を観測できないと仮定し、式 (2) を用いて計算された推定値を利用する。また、システムはユーザの対話行為 (i.e. GPF) も観測できないと仮定し、以下の式に基づいた推定値を利用する。

$$P(C_{user}^{t+1} | H_{G_{user}}^t) = \frac{\sum_{G_{user}^t} P(H_{G_{user}^{t+1}} | G_{user}^{t+1}) P(G_{user}^{t+1} | G_{user}^t) P(G_{user}^t)}{\sum_{G_{user}^t, G_{user}^{t+1}} P(H_{G_{user}^{t+1}} | G_{user}^{t+1}) P(G_{user}^{t+1} | G_{user}^t) P(G_{user}^t)}, \quad (4)$$

$H_{G_{user}}^t$  は  $t$  におけるシステムの言語理解結果 (4.1 節) を表している。他の変数は式 (1) と式 (2) 中のものと同じである。 $P(H_{G_{user}^{t+1}} | G_{user}^{t+1})$  は真のユーザの GPF と言語理解結果の混同行列を表している。なお、混同行列は、5.1 節での言語理解部の評価の際に得られたものを利用する。また、 $P(G_{user}^{t+1} | G_{user}^t)$  は 2. 節で述べられた販売員と客の対話コーパスを利用して構築される。

## 4. テキストベース協調的説得対話システム

3. 節の対話モデルを用いて学習したシステムの方策を評価するために、テキストベースの協調的説得対話システムを構築する。構築するシステムは言語理解部、方策部と言語生成部によって構成される (図 1)。方策部は 3. 節で構築した対話モデルを用いて強化学習された方策に従う。本節では、システムの言語理解部 (4.1 節) と自然言語生成部 (4.2 節) について説明する。

### 4.1 言語理解部

言語理解部では、統計的分類器を用いて、ユーザの入力テキスト  $u_{user}$  中の GPF を検出する。本稿では、統計的分類器として決定木を弱分類器とする Bagging[6] を利用する。

分類器への入力  $u_{user}$  とシステムの出力  $\langle u_{sys}, \langle G_{sys}, F_{sys} \rangle \rangle$  の履歴から計算される特徴量である。特徴量は主に以下の 4 種類に分類される：

**Uni:**  $u_{user}$  中のユニグラム単語頻度。

表 2: 学習データにおける GPF ラベルの分布.

Other	Question	SetQuestion	PropQ
46	4	12	156
Inform	Answer	Directive	Commissive
260	117	36	63

**Bi:**  $u_{user}$  中のバイグラム単語頻度.

**DAcl:** 直前のシステムのアクション ( $\langle G'_{sys}, F'_{sys} \rangle$ ).

**Unicl:** システムの直前の発話  $u'_{sys}$  中のユニグラム単語頻度.

言語理解結果  $H_{G_{user}}$  は, 8 種類の GPF の所属確率値である.

我々は説得対話コーパス (2. 節) 中の客の 694 発話を分類器の学習データとして利用する. この学習データでは表 2 に記載されている通りに GPF ラベルが分布している.

## 4.2 自然言語生成部

言語生成部は,  $u_{user}$ ,  $u'_{sys}$  とシステムのアクション  $\langle G_{sys}, F_{sys} \rangle$  を入力とし, システム発話  $u_{sys}$  を出力する. 本稿では, 容易にシステム発話が修正・管理出来るように, 用例ベース対話管理 [7] の枠組みに基づいた言語理解部を構築する.

我々は, 2. 節の説得対話コーパスを修正することで, 用例データベース  $D = \{d_1, d_2, \dots, d_M\}$  を構築する. この用例データベースは,  $i$  番目のデータ  $d_i = \langle s, u, g, f, p \rangle$  は話者  $s$ , 発話  $u$ , GPF  $g$ , フレーミング  $f$  と直前のデータ  $p$  によって構成される. なお,  $M$  はデータ総数を表す. 用例データベースとして, 以下の修正が施された 2. 節の説得対話コーパスを用いる:

- 冗長な語 (e.g. 言い淀みや言いなおし) の削除.
- 省略された語 (e.g. 主語や述語) と文の補完.

我々の用例データベースは, 2022 発話 (695 のシステム発話と 1327 のユーザの発話例) から構成される.

言語生成部は,  $u_{user}$ ,  $u'_{sys}$  と  $\langle G_{sys}, F_{sys} \rangle$  から, システムの応答  $u_{sys}$  を決定する. より具体的には, 言語生成部は以下の手順に基づいてシステムの応答を決定する.

1. システムの応答候補  $R$  がユーザ入力の有無 ( $u_{user} \neq \phi$ ,  $u_{user} = \phi$ ) により決定される. もし  $u_{user} \neq \phi$  ならば,  $R$  は直前の発話がユーザ発話 ( $r.p.s = User^{*1}$ ) である発話  $r$  の集合と定義される. そして, もし  $u_{user} = \phi$  ならば,  $R$  は直前の発話がシステム発話 ( $r.p.s = Sys$ ) である発話  $r$  の集合と定義される.
2. システムの応答候補  $R$  は以下の類似度を用いて評価される.

$$\cos(r.p.u, u_{input}) = \frac{\text{words}(r.p.u) \cdot \text{words}(u_{input})}{|\text{words}(r.p.u)| \cdot |\text{words}(u_{input})|} \quad (5)$$

$$u_{input} = \begin{cases} u'_{sys} & (u_{user} = \phi) \\ u_{user} & (u_{user} \neq \phi). \end{cases}$$

システムの応答候補中の直前の発話  $r.p.u (r \in R)$  と入力された発話  $u_{input}$  間のコサイン類似度が評価に利用される.  $u_{input}$  は,  $u_{user}$  に応じて,  $u'_{sys}$  か  $u_{user}$  の値が設定される. words 関数は tf-idf により重み付けられた語の頻度を返す.

3. 最も高い評価をされた  $r^*.u$  が言語生成部の出力  $u_{sys}$  として選択される.

$$r^* = \arg \max_{r \in R} \cos(r.p.u, u_{input}) \quad (6)$$

$$u_{sys} = r^*.u. \quad (7)$$

\*1本稿では, “.” を変数間の所属関係を表す記号として利用する. 例えば,  $Var1.Var2$  は「 $Var2$  は  $Var1$  のメンバー変数である」ことを意味する.

## 5. 実験結果

本節では 2 種類の実験的評価を行う. まず最初に, 予備実験として, 4.1 節で構築した言語理解部の評価を行う. 次に, 4. 節で構築したテキストベース協調的説得対話システムの評価を行う.

### 5.1 言語理解部の評価

4.1 節で提案した特徴量を組み合わせて言語理解部の性能評価を行う. 本稿では, 4 通りの特徴量の組み合わせ (Uni, Uni+DAcl, Uni+CAcl+Unicl and Uni+CAcl+Bi) を用いる言語理解部がそれぞれ構築される. そして, 構築された言語理解部は, 客の発話中の GPF ラベルの正答率に基づいて, 評価される. この評価は, 4.1 節で述べた客の 694 発話を用いた 15 分割交差検定に基づいて行われる.

実験結果 (図 2) から, Uni+CAcl+Bi を用いた言語理解部が最も高い正答率を達成したことがわかる. そのため, 次節の評価では, この Uni+CAcl+Bi を用いた言語理解部を利用する. 詳細な分類誤りに注目するため, 表 3 に Uni+CAcl+Bi を用いた言語理解部の分類結果に対する混同行列を示す. この行列から, Answer が Inform に, SetQ と Question が PropositionalQ にそれぞれ誤って分類されていることがわかる. この結果は, 構築した言語理解部では, 上位・下位関係にある対話行為を弁別することが困難であることを示唆する.

### 5.2 協調的説得対話システムの全体評価

本節では, 4. 節で実装した対話システムを用いて, 3. 節の対話モデルに基づいて学習されたシステムの方策の評価を行う. 評価のために, 以下の 4 種類の方策を用意する.

**Random:** システムのアクションをランダムに選択するベースライン.

**NoFraming:** GPF のみをシステムのアクションに用いて学習されたベースライン. この方策では, システムのアクションは, 3.2 節のシステムのアクション中の, フレーミングが None のアクションのみから選択される.

**Framing:** 3.2 節のモデルを用いて学習された提案手法. NoFraming とは異なり, この方策では, システムのアクションは, 3.2 節で定義された全てのシステムのアクションから選択される.

**Human:** 人間により適切なアクションが選択されるオラクル. この研究では, 一番目の著者 (1 年程度のカメラ販売対話の分析経験あり) がアクションの選択を行う.

方策 (i.e. NoFraming and Framing) を強化学習するため, 我々は Neural fitted Q Iteration (NFQ) [8] を用いる.

各方策は実ユーザとの対話における報酬と応答の正答率に基づいて評価される. 報酬は 3.2 節中の報酬を指す. また, 応答の正答率は, 全システムの応答の中でシステムが正しく応答できた割合である. この評価では, システムは販売員約を演じ, ユーザは客役を演じて対話を行う. そして, 対話の最後には, システムに与える報酬を計算するためユーザは以下の質問に回答する.

**満足度:** システムとの対話に関する 5 段階の主観的な満足度 (1: 満足しなかった, 3: どちらともいえない, 5: 満足した).

**カメラの最終選択:** 最終的にユーザが購入したいカメラ.

また, システムの応答の正答率を計算するため, ユーザは各システムの応答にたして, 正しく返答できているかを注釈する. 評価のために, 13 人のユーザそれぞれが各方策に従うシステムと 1 回ずつ対話を行う.

報酬に関する実験結果は図 3 に記載される. この結果から, Framing の報酬は, NoFraming と Random の報酬よりも高く,

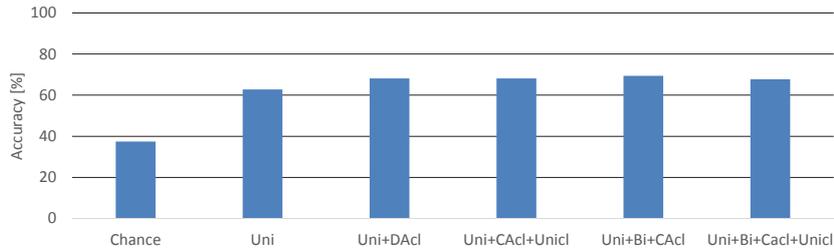


図 2: 言語理解部の正答率. 縦軸は正答率を表し, 横軸は言語理解部が用いた特徴量セットを表している. なお, チャンスレートとして, 常に Inform を出力する言語理解部の正答率も合わせて記載している.

表 3: 混同行列. 各行は真の GPF ラベルの分布を表している. また, 各列は言語理解部の分類結果を表している.

	Other	Commissive	PropQ	Directive	Answer	Inform	SetQ	Question	
Other	43	0	0	0	0	3	0	0	Other
Commissive	6	31	2	4	0	20	0	0	Commissive
PropQ	0	1	112	3	0	40	0	0	PropQ
Directive	2	2	6	13	0	13	0	0	Directive
Answer	0	3	5	0	53	56	0	0	Answer
Inform	1	12	4	4	9	230	0	0	Inform
SetQ	0	0	10	0	0	2	0	0	SetQ
Question	0	0	3	0	0	1	0	0	Question

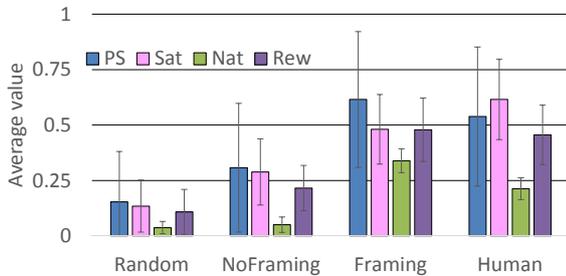


図 3: 実ユーザにおける評価結果. エラーバーは 95%信頼区間を表している. Rew は報酬, Sat はユーザの満足度, PS はシステムの説得成功, Nat は自然性をそれぞれ示している.

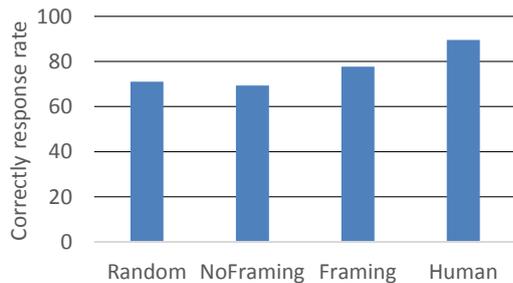


図 4: システムの発話の正答率.

Human の報酬とほぼ同じであることがわかる. これはフレーミングを用いて学習した方策は, 構築したシステムにおいて, 有効であることを示唆する. また, 本実験結果における各方策の報酬の傾向は, Wizard-of-Oz に基づいた実験結果 [1] と類似している.

システムの応答の正答率 (図 4) は, 我々の協調的説得対話システムはユーザの入力におおよそ正しく答えていることを示唆している. Framing に従うシステムは 77% の正答率を達成している. また, Random に従うシステムでさえ約 70% の正答率を達成している. このことから, システムが利用している言語生成部 (4.2 節) では, 文脈に沿った返答を生成できていることを示唆している.

## 6. 結論

本稿では, テキストベース協調的説得対話システムの構築と評価について説明した. 特に, システムの方策が従う対話モデル, システムの言語生成部と言語理解部の構築に焦点を当てて説明した. そして, 構築したシステムを用いて, システムの方策部の評価を行った. 実験結果から, 学習した方策部は本稿で構築したシステムでは, 人間が方策部を操作した場合と, ほぼ同等の性能を達成した. また, 異なる方策部に従うシステムの報酬の傾向は, 我々の先行研究での報酬の傾向と類似していた [1].

今後は, 協調的説得対話システムを, ロールプレイ対話ではなく, より現実的な状況で評価を行う予定である. また, 非言語情報を考慮して, システムの説得成功やユーザ満足度の予測モデルの性能を改善する予定である.

## 参考文献

- [1] Takuya Hiraoka, Graham Neubig, Sakriani Sakti, Tomoki Toda, and Satoshi Nakamura, "Reinforcement learning of cooperative persuasive dialogue policies using framing," *Proceedings COLING*, 2014.
- [2] Takuya Hiraoka, Yuki Yamauchi, Graham Neubig, Sakriani Sakti, Tomoki Toda, and Satoshi Nakamura, "Dialogue management for leading the conversation in persuasive dialogue systems," *Proceedings of ASRU*, 2013.
- [3] Levin Irwin, Sandra L. Schneider, and Gary J. Gaeth, "All frames are not created equal: A typology and critical analysis of framing effects," *Organizational Behavior and Human Decision Processes* 76.2, 2013.
- [4] Takuya Hiraoka, Graham Neubig, Sakriani Sakti, Tomoki Toda, and Satoshi Nakamura, "Construction and analysis of a persuasive dialogue corpus," *Proceedings of IWSDS*, 2014.
- [5] ISO24617-2, *Language resource management-Semantic annotation frame work (SemAF), Part2: Dialogue acts*. ISO, 2010.
- [6] Leo Breiman, "Bagging predictors," *Machine Learning*, 1996.
- [7] Cheongjae Lee, Sangkeun Jung, Seokhwan Kim, and Gary Geunbae Lee, "Example-based dialog modeling for practical multi-domain dialog system.," *Speech Communication*, 2009.
- [8] Martin Riedmiller, "Neural fitted Q iteration - first experiences with a data efficient neural reinforcement learning method," *Machine Learning: ECML*, 2005.