

# 対話行為予測に基づいた レファレンスインタビューのモデル化とその分析

河野 誠也<sup>1,a)</sup> 吉野 幸一郎<sup>1,2</sup> 鈴木 優<sup>1</sup> 中村 哲<sup>1</sup>

概要：対話システムによってユーザの情報検索を補助する場合、相手の意図や知識に応じて適切な補助を行うために、対話行為を用いた対話参加者のモデル化が有効であると考えられる。対話行為を用いたモデル化により、ユーザの各状態に応じた対話システムの応答戦略を考えることができる。本稿では、こうした応答戦略を図書館における書籍検索サービスの一部であるレファレンスインタビューにおける参加者の振る舞いからモデル化することを志向して、レファレンスインタビューにおける参加者の行動予測モデルを構築する。具体的には、発話レベルの観測系列と意図レベルの観測系列をそれぞれ素性として用いて、サービスの利用者、被利用者の対話行為予測モデルを構築し、その評価・分析を行う。また、構築した対話行為予測モデルの対話システムへの応用可能性についての検討を行う。

キーワード：統計的対話管理、対話行為予測、ユーザーシミュレーター、レファレンスインタビュー

## 1. はじめに

情報検索をタスクとする対話システムにおいて、ユーザの質問はしばしば曖昧であり、その質問に対して1ターンで明確な応答をすることが難しい場合がある。このとき、システム側から何らかの聞き返しや情報提示を行い、ユーザの検索意図の明確化を行うことが重要となる [1]。

人対人で情報検索をタスクとする対話が行われる場面として、図書館におけるレファレンスサービスがある。レファレンスサービスとは、図書館利用者が学習・研究・調査を目的として必要な情報・資料などを求めた際に、図書館員が情報そのもの、あるいはそのために必要とされる資料を検索・提供・回答することによってこれを助ける図書館サービスである [2]。レファレンスサービスにおいては、図書館利用者の曖昧な情報要求を明確化することを目的としてレファレンスインタビューという事前対応が行われる。このレファレンスインタビューを事前に行った場合、行わなかった場合と比較して簡易な事実調査に関するレファレンスサービスの回答の正確さが大幅に向上することが知られている [3]。こうした機能を対話システムに導入することができれば、ユーザが必要とする情報にたどり着く可能性

を高めることができると期待できる。

こうした機能のモデル化を行う場合、どのようなケースで聞き返しを行うかなど対話行為レベルでの抽象化が有効であると考えられる。そこで、本稿では、対話システムにおけるユーザの検索意図明確化のための応答戦略を、図書館における書籍検索サービスの一部であるレファレンスインタビューにおける参加者の振る舞いからモデル化するにあたり、まず、参加者の対話行為予測モデルを構築する。具体的には、発話レベルの観測系列と意図レベルの観測系列をそれぞれ素性として用いて、サービスの利用者（ユーザ）、被利用者（図書館員）の対話行為予測モデルを構築し、その評価・分析を行う。また、本稿で構築した対話行為予測モデルの対話システムへの応用可能性についての検討を行う。

本稿の構成は次の通りである。続く2章では、図書館におけるレファレンスインタビューとそのコーパスについて紹介する。3章では関連研究について述べる。4章では、対話行為予測モデルについての詳細を説明する。5章において実験と結果を述べ、その考察を行う。最後に6章において本論文のまとめを行う。

## 2. 図書館におけるレファレンスインタビュー

レファレンスサービスにおける図書館員は、レファレンスインタビューを通じて、利用者の情報要求を明確化したうえで回答を提供する。その際、インタビューによって質

<sup>1</sup> 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

<sup>2</sup> 科学技術振興機構  
Japan Science and Technology Agency

a) kawano.seiya.kj0@is.naist.jp

表 1 チャットレファレンスインタビューの例

No.	Speaker	Utterance
1	Patron	here is a current in a metal wire due to the motion of electrons. sketch a possible path for the motion of a single electron in this wire, the direction of the electric field vector, and the direction of conventional current.
2	Patron	you can just describe what they would look like
3	Librarian	Just a moment, please....
4	Patron	Thanks
5	Librarian	Is this for a school assignment and if so what is your grade level?
6	Patron	I 'm a junior in high school... it 's for a physics class... i have a test tomorrow and this stuff and I 'm still shaky on it
7	Librarian	What part of your physics books this question comes from: electricity?
8	Patron	ya
9	Librarian	Let me check
10	Librarian	Hold on please
11	Patron	ok
12	Librarian	I am still checking
13	Librarian	Hold on please
14	Librarian	<a href="http://www.swansontec.com/set.htm">http://www.swansontec.com/set.htm</a>
15	Librarian	The source that I just sent has good graphics that shows the electric currents
16	Librarian	And the graphic is animated so you can see the movement
17	Librarian	Can you see the page?
18	Patron	yes
19	Librarian	Let me check for more hold on please...

問の主題を確定するとともに、要求が生じた状況や背景、情報利用の目的や動機などを、利用者への質問などを通じて明らかにする。また、利用者が対話以前から明確に意識していることだけでなく、図書館員のインタビューによってはじめて意識し思い起こすようなことや、図書館員に問われその時点で考えたことも、適切な文献を推薦する上で重要な手がかりとなる。こうした情報を総合して、ユーザが真に必要とする文献を検索・提供する [4]。

Radford らは、OCLC(Online Computer Libaray Center, Inc.) が提供するチャットレファレンスサービス QuestionPoint に記録されたレファレンスインタビューのチャットログをランダムにサンプリングし、サービスの利用者と被利用者(図書館員)を対象とした分析を行った。その結果、チャットレファレンスでは相手との共感を生み出すことが重要であることを指摘しており、そのためには図書館員が自分自身についての情報を提供し、安心を与え、ユーモアのある形式張らない言葉遣いをすることが勧められている。また、レファレンスサービスにおいては正しい回答を示すだけでなく、図書館が対話をする能力を発揮することで、利用者のニーズを理解しようとし、信頼関係を築き、利用者に安心感を与えることが重要であり、それが「名前も顔もない白い検索ボックス」と図書館員を分けるものであると述べている [5]。

## 2.1 レファレンスインタビューコーパス

本研究では、Radford らが用いた QuestionPoint にお

けるチャットデータをコーパスとして用いる。本データは、700 対話セッション、約 12,634 発話から構成される。また、収録されているすべての発話について、参与者、日時(分単位)に関する情報が付与されており、個人情報に関わる部分(利用者の名前やメールアドレスなど)については、匿名化処理が施されている。

レファレンスインタビューにおける実際の対話例を表 1 に示す。表 1 の例では、利用者の質問に対して、単に回答を提示するのではなく、インタビューを通じて次の日に学校で試験があることなどの背景情報や、利用者の学年、検索履歴などの事項を明らかにし、これに応じた情報を提供していることがわかる。

## 2.2 レファレンスインタビューにおける対話行為

人間同士が行っているような自然な対話を行う対話システムを構築する上で、システムの全ての振る舞いをルールで網羅的に記述する方法は非現実的である。そこで、強化学習などの枠組みを用いてシステムの振る舞いを対話コーパスから統計的に学習する試みがなされている [6], [7]。しかしながら、このようなデータ駆動型のアプローチを用いてモデル化を行う場合、学習に用いることができる対話データの量は限られるため、入力シンボルとなる発話のある程度汎化した状態で、記述し取り扱うことが必要となる [1], [8]。そこで、本稿では発話を汎化し記述するために対話行為タグに着目する。対話行為とは、対話において参与者がどのような意図を持って発話を行ったかに応じて発

話を分類するタグとして定義される [9] .

レファレンスインタビューに特化した対話行為タグとして、Inoue によるタグセットがある [10] . Inoue は、レファレンスインタビューにおける参加者の行動は、円滑なコミュニケーションの促進と維持のために、1) 情報の伝達 (情報提供, 情報要求), 2) 情報探索活動におけるタスク管理, 3) 挨拶や感謝などによる社会的関係の維持, 4) 通信の物理的側面の管理といった 4 つの機能のいずれかを持つとしている . 具体的には、発話を機能 (Dialogue Act Function; DAF) とドメイン (Dialogue Act Domain; DAD) の 2 つの意味的レベルで取り扱い、対話行為タグの定義を行っている (表 2, 3) .

表 2 Dialogue Act Function(5 クラス)

No	Dialogue Act Function	Description
1	Information Provision	To provide information
2	Information Request	To request information
3	Task Management	To assign or commit to tasks
4	Social Relationship Management	To manage socio-emotional aspects of communication
5	Communication Management	To manage physical aspects of communication

表 3 Dialogue Act Domain(19 クラス)

No.	Function	Domain
1	Information Transfer	Information Problem
2	- Information Provision	Search Process
3	- Information Request	Information Object
4		Feedback
5		Other
6	Task Management	Librarians' Task
7		Users' Task
8		Other
9	Social Relationship Management	Greeting
10		Valediction
11		Exclamation
12		Apology
13		Gratitude
14		Downplay
15		Closing Ritual
16		Rapport Building
17	Communication Management	Channel Checking
18		Pausing
19		Feedback

### 3. 関連研究

対話システムにおける重要な研究課題の一つに対話管理がある . 対話管理とは、あるユーザ状態とその認識結果に対応する適切なシステム行動を決定する問題である . しか

しながら、2.2 節で述べたように、人手で定義されたルールを用いて対話管理を実現しようとする場合、ルールが増大すると、ルール間の一貫性を保つことが難しくなる . また、記述コストそのもの問題もある . この問題を解決するため、近年、マルコフ決定過程 (MDP) や部分観測マルコフ決定過程 (POMDP) などの枠組みを用いた強化学習による対話管理が研究されている [6], [7], [8], [11] . 具体的には、ユーザの対話行為やタスク達成に必要な情報を定義した対話状態などから環境の状態を、システムの対話行為をそれに対応する行動として定義し、決定過程として問題を定義する . つまり、強化学習を用いた対話システムにおける対話管理では、ユーザの状態および対話履歴を入力としたシステムの対話行為を予測する問題として考えることができる .

MDP, POMDP を用いた対話管理では、抽象化された意図レベルの入力 (対話行為や対話状態) からシステムの行動を決定する . 一方で、ユーザの発話と対話の履歴から直接システム応答を予測する研究もあり、Bordes らは、レストラン情報案内ドメインにおいて、End-to-End Memory Network を用いた応答選択モデルを提案している [12] . また、Min らはゲームベースの学習環境に対話システムを組み込むことを目的として、WoZ 実験により収集されたシステムとユーザとの対話ログから対話行為を予測するモデルを構築している [13] .

### 4. レファレンスインタビューにおける対話行為予測モデル

本研究では、今までの研究と同様に、利用者の発話に対する被利用者の応答対話行為の予測を行う . これに加えて、被利用者に対する利用者の応答予測も行うことで、対話のモデル化への知見を得る . 具体的には、対話行為系列を利用した予測と単語系列を利用した予測をそれぞれ試行し、対話の文脈というマクロな情報と発話内の単語というミクロな情報が対話行為の決定にそれぞれどの程度寄与するかを確認した .

本章では、具体的にどのようにしてレファレンスインタビューにおける対話行為予測のモデル化を行ったかについて述べる . まずデータの前処理について説明し、次に対話行為予測モデルの構築方法について述べる .

#### 4.1 データの前処理

本研究では、2.1 節で述べた Radford らによるチャットコーパスを用いる . 学習用対話データの作成にあたっては、次の規則によってデータを整形する .

- 参加者によるチャット入力 1 行分を 1 発話単位として取り扱う
- 対話における参加者は 1 ターンにつき 1 発話を行う : ある参加者が 1 ターンに複数の発話を行う場合、発話

間に他方の参加者による沈黙が行われているものとみなし、ラベル SILINCE を挿入する

- 対話は利用者によって開始される：実際のデータがこれに従わない場合、ラベル START\_STATE を挿入する
- 対話は利用者によって終了される：実際のデータがこれに従わない場合、ラベル END\_STATE を挿入する

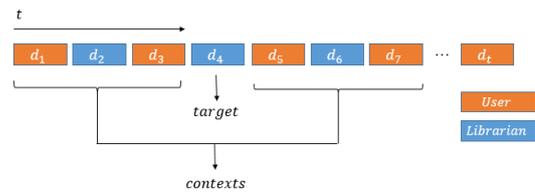


図 1 文脈の対話行為とターゲットの対話行為の決定 (窓幅=3)

#### 4.2 意図レベル観測系列を用いた対話行為予測

この節では、レファレンスインタビューにおける対話行為予測モデルについて述べる。具体的には、時刻  $t$  における参加者の発話の対話行為を  $d_t$  としたとき、その発話に対する他方の参加者の応答発話の対話行為  $d_{t+1}$  を予測するモデルについて説明する。これにより、意図レベルに抽象化されたマクロな情報が次の対話行為へ与える影響を確認する。

予測モデルには、RNN(Recurrent Neural Network) の一種である LSTM(Long Short-Term Memory)[14] を用いる。RNN は系列データを扱うためのモデルであり、前時刻の中間層の状態を現時刻の入力としても用いることで、内部状態を保持しながら学習を行う。しかし、通常の RNN は逆誤差伝播による学習を行う際、勾配が消失するという問題が起こりうる。LSTM はこの勾配消失の問題を解決するために改良されたネットワーク構造を持ち、RNN と同様にネットワーク構造に前時刻の状態を保持する隠れ層 (メモリブロック) と、隠れ層の各ノードの入力、出力、忘却を制御する 3 種類のゲート (入力ゲート、出力ゲート、忘却ゲート) から構成される。

構築したモデルの概要を図 1 に示す。時間  $t$  の時点で観測された対話行為の系列  $\{d_1, \dots, d_{t-1}, d_t\}$  を各々 one-hot ベクトルで表したものをモデルの入力とする。各入力ベクトルは Embedding 層において 128 次元の密ベクトルにしたうえで LSTM へ入力される。Embedding 層の重み初期値については、対話において似た文脈で出現する対話行為はそれぞれ似た意味を持つという仮定に基づいて、CBow(Continuous Bag-of-Words model)[15] を用いて事前学習する。CBow は生テキスト (単語の系列) から単語埋め込みを学習するための効率的な予測モデルである。本論文では、図 2 のように対話における対話行為の系列を用いて、文脈の対話行為から、ターゲットの対話行為を予測する CBow モデルを構築した。これによって得られた重みベクトルを、Embedding 層の重み初期値として用いる。

#### 4.3 発話レベル観測系列を用いた対話行為予測

この節では、レファレンスインタビューにおける対話において、時刻  $t$  における参加者の発話を  $u_t$  としたとき、その発話に対する他方の参加者の応答発話の対話行為  $d_{t+1}$  を予測するモデルについて説明する。これにより、発話内

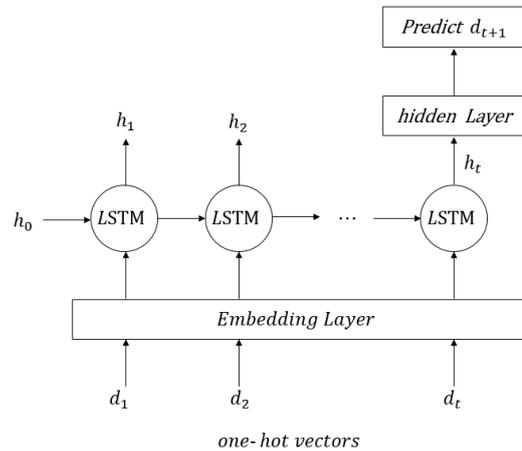


図 2 LSTM による対話行為予測モデル

容そのものが次の対話行為に与える影響を確認する。

予測モデルには、4.2 節で述べた LSTM によるモデルと同様のものを用いる。しかし、予測モデルへの入力、時間  $t$  の時点で観測された発話の系列  $\{u_1, \dots, u_{t-1}, u_t\}$  を単語分割して連結した長さ  $N$  の単語列  $\{w_1, \dots, w_{N-1}, w_N\}$  とする。各単語は、one-hot ベクトルで表され Embedding 層において 128 次元の密ベクトルに変換される。Embedding 層の重み初期値はランダムに設定する。

### 5. 評価実験

Radford らによるチャットレファレンスインタビューコーパスのうち、Inoue によりアノテーションされた 200 対話セッション (5,327 発話) を学習データとして用いて 10 分割交差検証により評価した。乱数シード値は固定で実験を行った。評価指標としては、対話行為の予測精度を確認するために、Accuracy, F-score を用いる。しかしながら、人対人における対話においては、必ずしも適切な応答が一意に定まらない場合がある。そこで、5-best Accuracy, Multi-class Log Loss を用いた評価も行う。単語分割は、spaCy<sup>\*1</sup>で行った。予測に用いる対話行為タグとしては、Inoue により定義された 19 クラスの Dialogue Act Doamin に加えて、4.1 節で述べた SILENCE 及び END\_STATE を用いた 21 クラスのタグを用いる。

#### 5.1 ベースライン手法

この節では、比較に用いた対話行為予測モデルについて

\*1 <https://github.com/explosion/spaCy>

て述べる．予測モデルには，Random Forests(以下，RF)，Multinomial Naive Bayes(以下，MNB)を用いる．対話行為系列を入力とした予測モデルには，以下の素性を用いた．

- 現在の時間  $t$
- 参加者情報
- 時間  $t$  における発話の対話行為
- 時間  $t-1$  における発話の対話行為
- 時間  $t-2$  までに出現した対話行為の出現回数
- 時間  $t-2$  までに出現した対話行為の隣接ペアの出現回数

単語系列を入力とした予測モデルには，以下の素性を用いた．

- 現在の時間  $t$
- 参加者情報
- 時間  $t$  における発話の bag-of-words, bag-of-bigrams
- 時間  $t-1$  における発話の bag-of-words, bag-of-bigrams
- 時間  $t-2$  までに出現した発話の bag-of-words, bag-of-bigrams

## 5.2 実験結果・考察

4.2 節，4.3 節で述べた対話行為予測モデルの評価結果を表 4，各対話参加者毎の応答対話行為予測結果を表 5 に示す．対話行為系列を利用した対話行為予測では，ナイーブベイズ，Random Forest による予測と比較して，対話行為系列を入力とした LSTM の性能が全ての評価指標で優る結果となった．1-best Accuracy では 0.51，5-best の Accuracy では 0.86 で参加者の応答を予測が可能であることを確認した．一方で，単語系列を利用した対話行為予測では，ナイーブベイズによる予測器の性能が最も性能が低く，Random Forest，LSTM による予測器が優れるという結果となった．ナイーブベイズによる予測の LogLoss は 22.32 であり，全く予測できていないことが確認できる．また，提案手法である LSTM は Random Forest に対し，F-score で劣る一方で，5-best Accuracy 及び LogLoss で優れていることが確認できた．これは，LSTM による予測モデルにおける正解の対話行為の出力確率がより高く，上位に含まれていることを示唆している．

各対話参加者毎の応答対話行為予測結果としては，対話行為系列を用いた予測モデルと単語系列を用いた予測モデルの両方において，利用者の応答予測が図書館員の応答予測に比べて容易であるということが確認できた．具体的には，対話行為系列を用いたモデルでは，1-best Accuracy で 0.62，5-best Accuracy で 0.91 の高精度で利用者の応答予測が可能であった．一方で，図書館員の応答予測は 1-best Accuracy で 0.41，5-best Accuracy で 0.82 と性能差があり，単語系列を用いたモデルでも同様の傾向となった．2.1 節で述べたように，レファレンスインタビューにおいて図書館員は能動的な聞き返しや情報提供などにより対話を主

導的に進める一方で，利用者は受け身になることが多い．そのため，利用者の応答は一意に決まりやすい場面が多く予測が容易であったことが推察される．一方で，図書館員の応答については，個々人の対話戦略の異なりにより自由度が高いため応答が一意に決定されにくく，文脈によって複数の正解候補が存在したのではないかと考えられる．ここから，文脈に応じて複数の許容可能な応答候補を考慮することが有効であることが示唆される．また，疑似ラベル SILENCE を発話間に挿入した影響もあると考えられる．

表 4 応答対話行為の予測結果

手法	1-best	5-best	F-score	LogLoss
意図レベル予測				
MNB	0.4114	0.7951	0.3771	4.2712
RF	0.4517	0.8079	0.3834	2.0496
LSTM	<b>0.5154</b>	<b>0.8667</b>	<b>0.4470</b>	<b>1.6111</b>
発話レベル予測				
MNB	0.1461	0.5596	0.1550	22.3260
RF	0.4190	0.7899	<b>0.3582</b>	2.0509
LSTM	<b>0.4376</b>	<b>0.8085</b>	0.3458	<b>1.8958</b>

表 5 各参加者毎の応答対話行為の予測結果

手法	1-best	5-best	F-score	LogLoss
意図レベル予測				
MNB(図書館員応答)	0.3009	0.7284	0.2756	5.3030
RF(図書館員応答)	0.3442	0.7535	0.3054	2.3536
LSTM(図書館員応答)	<b>0.4151</b>	<b>0.8218</b>	<b>0.3618</b>	<b>1.8778</b>
発話レベル予測				
MNB(図書館員応答)	0.1470	0.4859	0.1509	23.0166
RF(図書館員応答)	0.3116	0.7412	0.2826	2.3491
LSTM(図書館員応答)	<b>0.3520</b>	<b>0.7575</b>	<b>0.2912</b>	<b>2.1251</b>
利用者レベル予測				
MNB(利用者応答)	0.1452	0.6371	0.1738	21.5984
RF(利用者応答)	<b>0.5321</b>	0.8411	<b>0.4582</b>	1.7369
LSTM(利用者応答)	0.5278	<b>0.8622</b>	0.4287	<b>1.6542</b>

今回の実験では，対話行為系列を用いたモデルと単語系列を用いたモデルでは，対話行為系列を用いたモデルの予測精度がやや優る結果となった．これは，入力として用いた対話行為系列にアノテーションされた正解を用いたことや，学習データ量，素性設計による問題でもあったと考えられる．しかしながら，対話において出現した単語系列の情報は，抽象化された対話行為の情報を潜在的に含むため，両者を組み合わせることで性能向上は期待できると考えられる．今回構築した LSTM によるモデルは，対話において観測された発話の単語列を順に LSTM への入力とするシンプルなモデルであった．そのため，対話のターン経過に応じた入力系列長の増加に由来する予測性能の低下の問題がある．したがって，対話における履歴の効率的な利用を

行うための注意機構の導入 [12] や、発話単位や話者交代をモデルとして導入 [16] するなどによるモデルの改善を検討する必要があると考えられる。

## 6. まとめ

本研究では、レファレンスインタビューにおける参加者の応答の対話行為を予測するために、リカレントニューラルネットワークを用いたモデルを構築しその性能評価を行った。評価実験の結果、対話行為系列を用いた予測モデルと単語語系列を用いた予測モデルの 5-best Accuracy はそれぞれ、0.86, 0.80 の精度で予測が可能であることを確認した。また、レファレンスインタビューにおいては、図書館利用者の応答は図書館員の応答予測と比較して高精度で予測が可能であることを示した。対話参加者の応答モデル化を行うことは、対話システムにおける対話管理や強化学習におけるユーザシミュレーション等々への応用が考えられるが、予測性能向上に向けたモデル改良や、対話システム構築へ向けた対話行為タグ設計などの課題が残る。また、本研究では予測モデルへの入力として対話行為系列と発話における単語系列のみを考えたが、レファレンスインタビューにおける本質は利用者の情報要求の認識とその充足にある。そのため、利用者の情報要求の状態や検索意図などを対話状態として定義し、対話参加者の行動を追跡できるような枠組みやその評価方法も、検討していく必要があると考えられる。

## 謝辞

本研究は JST さきがけ (JPMJPR165B) および JST CREST (JPMJCR1513) の支援を受けた。

## 参考文献

- [1] Yoshino, K. and Kawahara, T.: Conversational system for information navigation based on POMDP with user focus tracking, *Computer Speech & Language*, Vol. 34, No. 1, pp. 275–291 (2015).
- [2] 日本図書館学会. 用語辞典編集委員会: 図書館情報学用語辞典, 丸善株式会社 (1997).
- [3] Radford, M. L., Connaway, L. S., Confer, P. A., Sabolcsi-Boros, S. and Kwon, H.: "Are we getting warmer?" Query clarification in live chat virtual reference, *Reference & User Services Quarterly*, pp. 259–279 (2011).
- [4] 泰則 斎藤: レファレンス・インタビューにおける利用者モデル, *Library and information science*, No. 27, pp. p69–85 (オンライン), 入手先 (<http://ci.nii.ac.jp/naid/40000019897/en/>) (1989).
- [5] Radford, M. L. and Connaway, L. S.: Seeking synchronicity: Evaluating virtual reference services from user, non-user, and librarian perspectives, *Proposal for a research project, submitted February*, Vol. 1, p. 2005 (2005).
- [6] Williams, J. D.: Applying POMDPs to dialog systems in the troubleshooting domain, *Proceedings of the Workshop on Bridging the Gap: Academic and Industrial*

- Research in Dialog Technologies*, Association for Computational Linguistics, pp. 1–8 (2007).
- [7] 目黒豊美, 東中竜一郎, 南泰浩, 堂坂浩二: POMDP を用いた聞き役対話システムの対話制御, 言語処理学会第 17 回年次大会, pp. 912–915 (2011).
  - [8] 翠輝久, 大竹清敬, 堀智織, 柏岡秀紀, 中村哲ほか: 京都観光案内対話コーパスにおける対話行為タグの設計と分析, 情報処理学会研究報告音声言語情報処理 (SLP), Vol. 2009, No. 10 (2009-SLP-075), pp. 39–44 (2009).
  - [9] Stolcke, A., Ries, K., Coccaro, N., Shriberg, E., Bates, R., Jurafsky, D., Taylor, P., Martin, R., Van Ess-Dykema, C. and Meteer, M.: Dialogue act modeling for automatic tagging and recognition of conversational speech, *Computational linguistics*, Vol. 26, No. 3, pp. 339–373 (2000).
  - [10] Inoue, K.: An Investigation of Digital Reference Interviews: A Dialogue Act Approach, PhD Thesis, Syracuse University (2013).
  - [11] Hiraoka, T., Neubig, G., Sakti, S., Toda, T. and Nakamura, S.: Reinforcement Learning of Cooperative Persuasive Dialogue Policies using Framing., *COLING*, pp. 1706–1717 (2014).
  - [12] Bordes, A. and Weston, J.: Learning end-to-end goal-oriented dialog, *arXiv preprint arXiv:1605.07683* (2016).
  - [13] Min, W., Wiggins, J. B., Pezzullo, L., Vail, A. K., Boyer, K. E., Mott, B. W., Frankosky, M., Wiebe, E. N. and Lester, J. C.: Predicting Dialogue Acts for Intelligent Virtual Agents with Multimodal Student Interaction Data., *EDM*, pp. 454–459 (2016).
  - [14] Hochreiter, S. and Schmidhuber, J.: Long short-term memory, *Neural computation*, Vol. 9, No. 8, pp. 1735–1780 (1997).
  - [15] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. and Dean, J.: Distributed representations of words and phrases and their compositionality, *Advances in neural information processing systems*, pp. 3111–3119 (2013).
  - [16] Serban, I. V., Sordani, A., Bengio, Y., Courville, A. C. and Pineau, J.: Building End-To-End Dialogue Systems Using Generative Hierarchical Neural Network Models., *AAAI*, pp. 3776–3784 (2016).