

HMM による非タスク指向型対話エージェントの対話構造のモデル化

Dialogue Modeling of Non-task-oriented dialogue agent using HMM

磯村直樹^{1*} 鳥海不二夫¹ 石井健一郎¹
Naoki ISOMURA¹ Fujio TORIUMI¹ Kenichiro ISHII¹

¹ 名古屋大学大学院情報科学研究科

¹ Graduate School of Information Science, Nagoya University

Abstract: Recently, computerized dialogue systems have been actively investigated and used in various fields. In order to develop a practical non-task-oriented dialogue agent, the dialogue structure of the agent should be evaluated. We design a hidden Markov model (HMM) by learning a sequence of SWBD-DAMSL tags to model the dialogue structures. We analyzed the dialogue structures of the two non-task-oriented dialogue agents and human. As a result of an experiment, We successfully modeled typical dialogue structures using HMM

1 はじめに

近年、人間と対話のできるコンピュータ(対話エージェント)が社会に浸透しつつある。対話エージェントが扱う対話の範囲も、航空機のチケット予約や接客などを目的としたタスク指向型対話だけでなく、人間の話し相手になり、人間を楽しませるための非タスク指向型対話へと広がりつつある。このような状況を反映して、対話エージェントの研究も盛んになりつつあるが、その多くはタスク指向型対話に関するものであり、非タスク指向型対話に関する研究は少ない。

非タスク指向型対話エージェントの設計には、適切な対話モデルが必要となる。例えば、人間同士の対話では、質問に対してすぐに返答があるとは限らない。質問と返答の間に詳細を尋ねる別の質問が行われることや、複数の質問に対して同時に返答を返す場合、答えが複数にまたがって返される場合などがある。このような話のまとまりを談話構造と呼び、この構造を解析することが対話のモデルには必要となる。非タスク指向型対話エージェントの設計においても、対話モデルを用いることで、対話の状態がどのようになっているかを把握し、対話の流れを制御することができる。

本論文では、非タスク指向型対話エージェントの対話構造をモデル化する手法を提案する。対象とする対話は1対1のテキスト対話に限定する。人間にとって最も自然な対話は音声対話であるが、本研究では問題を言語処理に特化するため、音声認識・合成技術の問

題は扱わないことにする。

談話構造をモデル化する手法として Support Vector Machine(SVM)を用いた解析手法 [1] がある。これはテキスト自動要約タスク TSC で使用されている新聞記事データに対して、談話構造を付与したコーパスを対象としている。談話構造は木構造であると仮定し、談話構造を文と文の二項関係の集合で表現している。二項関係には因果、並列、対比といった8種類の係り受け関係を用いている。本論文では対話を対象とするため、新聞記事を用いた談話構造解析は適さない。対話構造を解析したものとしては、確率・統計的手法によるモデル化手法 [2] がある。これは対話に付与された発話行為のタグを Hidden Markov Model(HMM) や決定性の有限オートマトンを用いて対話構造をモデル化している。対話モデル作成のデータには ATR 対話コーパス中の国際会議参加登録タスクを使い、タスク指向型対話を対象としている。本論文では非タスク指向型対話を対象としているため、発話行為タグでは十分に対話を表現できないと考えられる。そこで、本論文では、HMM と SWBD-DAMSL タグを用いて非タスク指向型対話エージェントの対話構造をモデル化する手法を提案する。

2 HMM による対話のモデル化

2.1 概要

本節ではテキスト対話を取り上げ、非タスク指向型対話エージェントの対話のモデル化手法を提案する。な

*連絡先: 名古屋大学大学院情報科学研究科
〒464 - 8603 愛知県名古屋市千種区不老町
E-mail: naoki@kishii.ss.is.nagoya-u.ac.jp

お、ここでいう対話とは、一回の発言を発話と定義したときに、会話の始まりから終わりまでの発話の系列である。

系列から隠れた構造を学習によってモデル化する手法として、Hidden Markov Model(HMM) が広く知られている。そこで、HMM を用いて対話をモデル化し、対話の構造を分析する。

この HMM の学習には、出力シンボルの選択が重要である。本手法では発話の種類を出力シンボルとする。発話の種類を記述するためのタグとして、SWBD-DAMSL(Switchboard Discourse Annotation and Markup System of Labeling) タグがある [3]。この SWBD-DAMSL タグには、Yes/No 質問を表す qy や、5W1H の質問を表す qw といったタグが存在する。これらのタグの組み合わせを付与することによって発話の意味を表すことができる。そこで、HMM の出力シンボルとしては SWBD-DAMSL タグを用いる。モデル化したい対話をタグの系列に置き換え、HMM にタグの系列を学習させることで、対話の構造をモデル化する。

2.2 SWBD-DAMSL タグ

SWBD-DAMSL タグは発話の種類を記述するタグであり、48 種類のタグで発話の種類を記述する。本研究において重要なタグを表 1 に示す。発話毎に一つまたは複数のタグが付与される。複数のタグが付与される例としては、「私は猫を飼っていますが、あなたは何か飼っていますか？」のように WH 質問と客観的発言を同時に行うような発話「qw + sd」などがある。手動でタグを付与した例を表 2 に示す。

本研究で用いた対話は表 3 の通りであり、合計 72 対話である。

対話は人間同士の対話 24 対話 (表 3 の (1))、人間と対話エージェントとの対話 48 対話 (表 3 の (2)(3)) より構成されており、以下の制約を課した。

- 30 分のテキスト対話とする
- 1 対 1 で交互に発話する
- 話題は制限しない
- 顔文字や方言は使用しない

評価対象となる対話の発話数が一定ではないことを考慮し、発話数ではなく 30 分を対話の単位とした。

人間同士の対話は、互いに面識のない被験者 10 人により行われ、全て異なるペアによる対話とした。また、人間と対話エージェントの対話は、それぞれ 24 人の異なる被験者により行われた。

表 1: SWBD-DAMSL タグの一部

	タグの種類	内容
1	qy	Yes/No 質問 「ペットを飼っていますか？」
2	qw	WH 質問 「どのチームが好きですか？」
3	qo	自由回答式の質問 「どう思う？」
4	qy ^ g	付加疑問文/確認 「東京にお住まいですね」
5	co	提案 「お入用ならレシピありますよ」
6	cc	遂行 「私がやりましょう」
7	ny	答え「はい」 「うん」「はい」
8	nm	答え「いいえ」 「いや」「ううん」
9	na	「はい」を使わない肯定 (「猫を飼っていますか?」) 「3匹おります」
10	ng	「いいえ」を使わない否定 (「あのドラマを見ましたか?」) 「うちはテレビがないんですよ」
11	no	その他の答え 「わかりません」
12	aa	同意 「ええ」「私もそう思います」
13	ar	拒否 「いいえ」
14	b	相槌 「ふんふん」「なるほど」
15	fe	感嘆 「ええっ」「へえー」
16	fp	開始の言葉 「こんにちは」
17	fc	閉めの言葉 「失礼します」
18	sv	主観的文 「これはおいしいね」
19	sd	客観的文 「これはリンゴです」
20	%	不明語、曖昧な発話

表 2: ある対話に対する手動タグ付与の例

発話者	発話 [タグ] (タグの意味)
C(人)	駅からちょっと行ったところに、有名な焼き屋さんがありますよ。 [sd] (客観的文)
D(人)	たこ焼き屋さんですか！これまた魅力的ですね！ [fe] (感嘆)
C	駅からの南北はちょっとわかりませんが、大通り沿いにある大きな服屋さんの裏にあります。 [sd] (客観的文)
D	インターネットで探してみます！お店の名前ご存じですか？ [cc,qy] (遂行, Yes/No 質問)

表 3: 実験に用いた対話

対話の種類	対話数	対話の自然さ
(1) 人間同士の対話	24	4.12
(2) 人間対 ELIZA 型 KELDIC の対話	24	1.67
(3) 人間対インタビュー型 KELDIC の対話	24	2.29

対話エージェントには、我々が考案した日本語対話エージェントである“KELDIC(Ken’s Laboratory Dialogue Computer)”のうち、ELIZA [4] を模して作られた ELIZA 型 KELDIC と、プロのインタビューを模したインタビュー型 KELDIC の 2 種類を用いた。

全ての対話は、対話を行った被験者によって対話の自然さをアンケートにより 1(極めて不自然) から 5(極めて自然) の 5 段階で主観評価されており、表 3 に示されている。この結果から、インタビュー型 KELDIC は ELIZA 型 KELDIC より自然な対話が可能であるといえる。また、当然ながら人間同士の対話は最も自然な対話と評価されている。

3 非タスク指向型対話エージェント

3.1 ELIZA 型 KELDIC

対話エージェント「KELDIC」(Ken’s Laboratory Dialogue Computer) は、我々が研究している非タスク指向型対話エージェントの総称である。本研究では、非タスク指向型対話エージェントとして、ELIZA 型 KELDIC とインタビュー型 KELDIC の 2 種類を使用

した。ELIZA 型 KELDIC の基本的な仕組みは ELIZA [4] と同じであり、対話には予め用意された発話生成規則を使用する。ただし、英語ではなく日本語を用いる点が異なっている。ELIZA 型 KELDIC は、人間からの質問に対して直接答えず、「私にはよくわかりません」「あなたはどうですか？」といったようにはぐらかし、相手の意見を尋ねるといった対話戦略をとっている。

3.1.1 ELIZA 型 KELDIC による発話生成規則

ELIZA 型 KELDIC は対話相手の発話ごとに、対応する発話生成規則を探して返答を生成する。以下に生成規則の例を挙げる。

key: 人

decomp: * (彼|彼女) * 有名 だ 人 * ?

reasmb: そうかもしれませんね。

reasmb: そうらしいですね。

decomp: * 誰 * 有名 だ 人 * ?

reasmb: あなたは誰か有名人に会ったことがありますか？

reasmb: あなたは有名人ですか？

まず、ユーザの発話を形態素解析し、キーワード (key) を探す。キーワードが見つかった場合、そのキーワードの持つ decomp と比較する。形態素解析の解析結果と decomp が一致した場合、reasmb の中から文をランダムに一つ選び、ユーザの発話に対する返答とする。

3.1.2 ELIZA 型 KELDIC の問題点

どのような発話に対しても適切な返答を返すためには、広範なキーワードと、キーワードごとに十分な量の decomp、さらに decomp それぞれに適切な reasmb が必要となる。そのため、多くの対話に対応するためには膨大な量の生成規則を記述する必要がある。生成規則の全てを人が考え、記述するのは困難であるため、生成規則を自動生成すること求められるが、生成規則は複雑であり、自動生成は難しい。

3.2 インタビュー型 KELDIC

インタビュー型 KELDIC は ELIZA 型 KELDIC を改良した対話エージェントであり、バラエティ番組の司会など、対話のプロであるインタビューを模した対話エージェントである。インタビュー型 KELDIC は、対話の前に対話相手の情報をもとに、大量の発話集 (スクリプト集) を用意して対話を行う対話エージェントであり、情報を聞き出すための対話ではなく、対話相手を楽しませ、話を盛り上げる対話を目指している。

3.2.1 インタビュー形式の対話

人間を楽しませる対話に着目した研究の一つとして、我々が行った調査 [5] がある。この調査より、人間同士の対話では、何の制約もない通常の対話よりも、片方がインタビューとなるインタビュー形式の対話のほうが、インタビューイ（インタビューを受ける発話者）が楽しく感じる事が明らかになった。インタビュー形式の対話には、以下の特徴がある。

1. YES/NO 型の質問文 (closed question) ではなく 5W1H 型の質問文 (open question) を使う
2. 対話の前にインタビューイについて調査する
3. インタビューイの意見に理解を示す
4. インタビューイの好む話題を選択する
5. 安易に話題を転換しない

インタビュー型 KELDIC は、インタビュー形式の対話を行う非タスク指向型対話エージェントである。

3.2.2 インタビュー型 KELDIC による発話の生成

インタビュー型 KELDIC は質問、相槌、話題転換、相手の質問への回答のいずれかを返答として行うことで人間と対話を行う。インタビュー型 KELDIC は以下の手順でユーザの発話から返答を出力する。

1. ユーザが発話をテキストで入力
2. 入力発話に文法的な情報を付与
3. 返答として回答、話題転換、質問、相槌のいずれかを生成
4. 生成された返答を過去に発言していないかチェックを行う。過去に発言していたら 3 に戻って、別の返答を生成
5. 返答を出力

まず、文法的な情報として表 4 の時制、態、体、法を付与する。例えば、文節に助動詞「た」を含むとき「過去形」、文節に助動詞の「たい」を含むとき「希望」といったように、複数の規則を設けた。

返答の生成では、入力発話が質問であるかどうかを調べる。入力が質問であれば、KELDIC のプロフィールまたは過去の発話を使い回答する。また、入力発話が否定であればプロフィールに関する質問を生成し、話題転換を行う。入力発話が質問でも否定でもない場合、格フレームを用いて入力発話から質問を生成する [6]。質問を生成できなかった場合、相槌を行う。

表 4: インタビュー型 KELDIC の文法情報

用語	意味
時制	過去・過去進行・現在・現在進行
態	能動態・受動態
体	常体（だ、である調）・敬体（です、ます調）
法	否定・疑問・指示・依頼・意思・忠告・命令・禁止・許可・義務・希望・依頼・推量・確認など

4 対話のモデル化実験

4.1 実験の概要

非タスク指向型対話エージェントの対話の構造を分析するため、表 3 で示した各対話を、HMM を用いてモデル化する。対話によって HMM の適切な状態数は異なると考えられるため、HMM の状態数を 1 から 20 まで変化させ、最もエントロピーが低くなるモデルを採用する。HMM のパラメータを λ とすると、エントロピー $H(\lambda)$ は以下の式で求められる。

$$H(K|i) = - \sum_{k=1}^K p(v_k|i) \log_2 p(v_k|i) \quad (1)$$

$$H(\lambda) = \sum_{i=1}^N \omega_i H(K|i) \quad (2)$$

ただし、

$$N : \text{状態数} \quad (3)$$

$$K : \text{シンボル数} \quad (4)$$

$$p(v_k|i) : \text{状態 } i \text{ でシンボル } v_k \text{ を出力する確率} \quad (5)$$

である。 ω_i は定常状態確率を表し、十分な遷移の後で状態 i にいる確率を表す。シンボルには SWBD-DAMSL タグを用いる。

4.2 実験結果

HMM の状態数とエントロピーの関係を図 1 に示す。全体の傾向として、HMM の状態数が多くなるに従いエントロピーが減少している。エントロピーが最小となった状態数は、人間同士の対話では状態数 19、人間対 ELIZA 型 KELDIC の対話では状態数 8、人間対インタビュー型 KELDIC の対話では状態数 11 のときであった。このときのそれぞれの HMM の構造を図 2、図 3、図 4 に示す。人間同士の対話は状態数が多いため、一部 (14 個の状態) のみを示す。図中の楕円は各状態を表している。楕円内に書かれている文字は出力シンボル

である SWBD-DAMSL タグを表す。DAMSL タグの右隣の数字は、そのタグを出力する確率を表す。DAMSL タグの種類が多いため、出力確率の大きいもののみを記述している。矢印の太さは遷移確率の大きさを示している。遷移確率も出力確率と同様に、確率の大きいもののみを記述している。

4.2.1 人間同士の対話

まず、人間同士の対話について分析する。図 2 より、[qy] (Yes/No 質問) を高確率で出力する状態からは、[na][ng] といった質問に対する返答を表すシンボルを高確率で出力する状態に遷移しやすいことがわかる。また、[qw] (WH 質問) を高確率で出力する状態からは、[sv][sd] といった意見を表すシンボルを高確率で出力する状態に遷移しやすいことがわかる。これらは、質問-返答といった対話の流れを示していると考えられる。さらに、[sv][sd] を高確率で出力する状態は多くの状態から高確率で遷移があり、対話における中心を担っていると考えられる。

4.2.2 人間対 ELIZA 型 KELDIC の対話

次に、人間対 ELIZA 型 KELDIC の対話について分析する。図 3 を見ると、[qy][qw][qo] といった質問の状態から質問または [no] (その他の答え) に遷移する可能性が高い。ELIZA 型 KELDIC では、人間からの質問に対して「その質問には答えられません」「あなたは どう思いますか?」のように、答えをはぐらかす対話戦略をとっているためこのような状態遷移が表れたと考えられる。また、[ar] (拒否)、[b] (相槌) といったタグが頻出しているが、キーワードが見つからない場合に「そう」「ええ」「はあ」といった相槌や「私はそうは思いません」といった返答を行うためと考えられる。

4.2.3 人間対インタビュー型 KELDIC の対話

同様に、人間対インタビュー型 KELDIC の対話について分析する。図 4 より、3 個の状態 [qo] が最大の出力確率となっていることがわかる。これは、インタビュー型 KELDIC が「[固有名詞] はどうでしたか?」という質問文を生成しているためであると考えられる。これらの状態は、特定の状態から遷移することがないため、話題転換を示していると考えられる。また、7 個の状態から高い確率で遷移してくる状態では、[sv][sd] といった意見を述べる一般の発話が多くなっており、これは人間同士の対話の傾向に近い。

実際に、人間同士の対話との類似度を定量的に評価した結果が図 5 である。これは、人間同士の対話を学

習した HMM に、それぞれの対話の SWBD-DAMSL タグ系列を入力したときの出力確率の平均を求めたものである。人間同士の対話は学習用対話であるため、出力確率は高い値になる。一方 2 種類の対話エージェントの SWBD-DAMSL タグ系列を入力した結果を比較すると、インタビュー型 KELDIC の方が ELIZA 型 KELDIC より明らかに出力確率が高くなっており、インタビュー型 KELDIC の方が ELIZA 型 KELDIC より人間同士の対話に近い対話構造を持っていると言える。

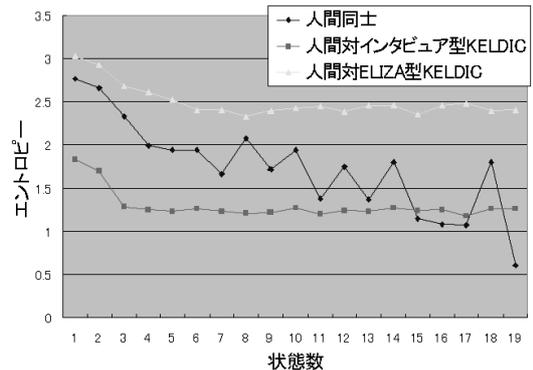


図 1: HMM の状態数とエントロピーの関係

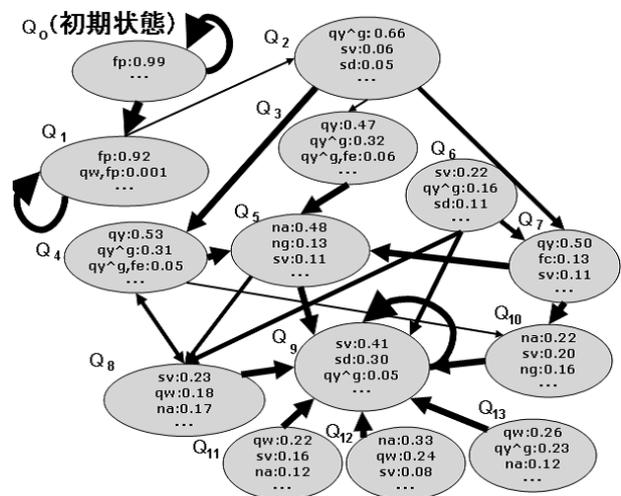


図 2: 人間の対話構造 (一部)

以上の結果より、HMM を用いることで、非タスク指向型対話エージェントの対話構造をモデル化できているといえる。

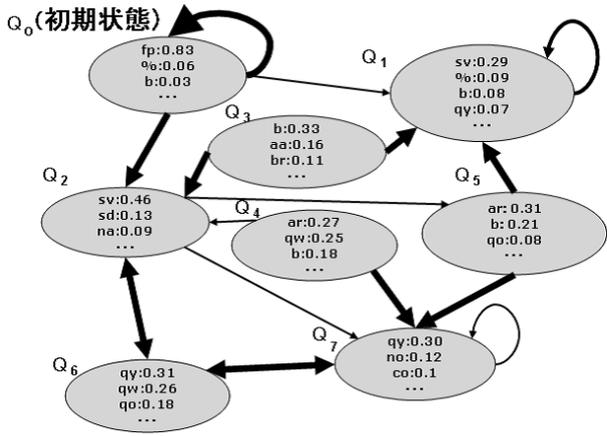


図 3: ELIZA 型 KELDIC の対話構造

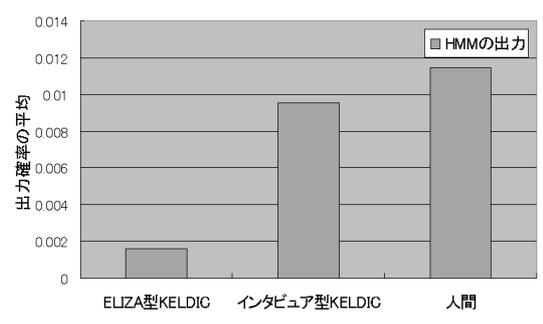


図 5: 人間同士の対話のモデルとの類似度

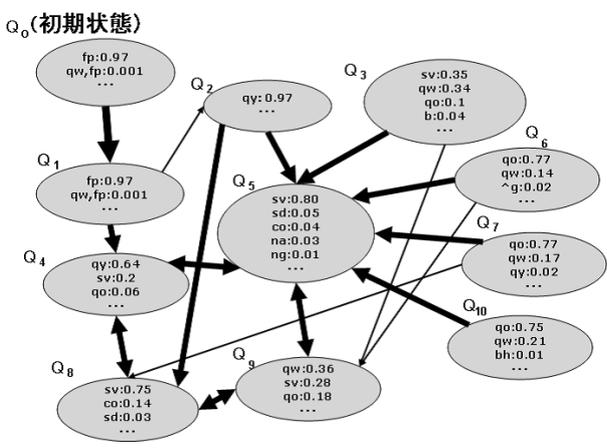


図 4: インタビュー型 KELDIC の対話構造

5 むすび

5.1 まとめ

非タスク指向型対話エージェントの対話構造を HMM を用いてモデル化した。本手法では、人間と対話エージェントの対話を SWBD-DAMSL タグで表現し、タグ系列を学習した HMM を作成した。対話エージェントのタグ系列を学習した HMM では、対話エージェントの特徴的な機能を抽出することができ、人間同士の対話を学習した HMM では、質問に対する返答といった人間同士の対話に表れる構造を抽出することができた。また、人間同士の対話を学習した HMM と比較することで、人間同士の対話の構造は人間対インタビュー型 KELDIC の対話の構造と似ていることが示唆された。HMM の出力確率を比較した結果、人間対インタビュー型 KELDIC の対話は人間対 ELIZA 型 KELDIC の対話よりも類似度が高いことが明らかとなった。

5.2 今後の課題

今後の課題としては、非タスク指向型対話エージェントの改良と、SWBD-DAMSL タグ以外の指標による対話のモデル化がある。人間の対話のモデルと非タスク指向型対話エージェントの対話モデルを比較することで、より人間らしい非タスク指向型対話エージェントを設計することが可能になると考えられる。

参考文献

- [1] 横山憲司, 難波英嗣, 奥村学. Support Vector Machine を用いた談話構造解析. 情報処理学会研究報告. 自然言語処理研究会報告, Vol. 2003, No. 23, pp. 193-200, 2003.
- [2] 福井義和, 北研二, 永田昌明, 森元逞. 確率・統計的手法による対話構造のモデル化. 情報処理学会研究報告. 自然言語処理研究会報告, Vol. 96, No. 27, pp. 111-118, 1996.
- [3] D. Jurafsky, E. Shriberg, and D. Biasca. Switchboard SWBD-DAMSL Shallow-Discourse-Function Annotation Coders Manual. *Institute of Cognitive Science Technical Report*, pp. 97-02, 1997.
- [4] J. Weizenbaum. ELIZA-a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, Vol. 9, No. 1, pp. 36-45, 1966.
- [5] 磯村直樹, 鳥海不二夫, 石井健一郎. HMM による非タスク指向型対話システムの評価. 電子情報通信学会技術研究報告, Vol. 106, No. 296, pp. 39-44, 2006.
- [6] 岡田譲二, 鳥海不二夫, 石井健一郎. インタビューを模した対話エージェントのための質問文自動生成. *Proceedings of JAWS 2007*, 2007.