

非タスク指向型対話システムの比較評価 Comparative Evaluation of Non-task-oriented Dialogue Systems

竹村 圭祐[†] 大村 英史[‡] 太原 育夫[‡]
Keisuke Takemura Hidefumi Omura Ikuo Tahara

1. はじめに

タスク指向型対話システムに比べて、対話自体を目的とした非タスク指向型の対話システムに関する研究はそれほど多くない。その理由の一つに客観的な評価が難しいということが考えられるが、近年その評価方法についていくつかの提案がなされている。磯村らはHMMを用いた評価方法を提案している[1]。本稿では、その評価方法を用いて、ログ型、辞書型、マルコフ型それぞれの非タスク遂行型対話システムと人間同士の実会話（名大会話コーパス）との比較検証を行う。また、与えるタグ数を限定することが結果に及ぼす影響についても検討する。

2. 非タスク指向型対話システム

雑談のような達成すべき目標を持たず、単に楽しむためのシステムを非タスク指向型対話システムと呼ぶ。既存の非タスク対話システムは数多く存在するが、以下の3種に大別することができる。

● ログ型

チャットログなどの人同士の対話ログをあらかじめ用意しておき、ユーザの発話とログの文字列を比較し、類似度の高い文章の次の文を応答とするものがログ型対話システムである。応答文の単語を適切な単語に置き換えて応答を生成することもある。

● 辞書型

発話内にキーとなる単語又は単語列が出現したときに、あらかじめ設定された文を応答として返すパターンマッチング型の対話システムである。対話の流れや意味を崩さずに応答を行うことができる反面、大量のパターン辞書を用意しなければならないという欠点もある。応答に用いるパターン辞書の例を表1に示す。

表 1: 応答のパターン辞書の例

キーワード	返答
こんにちは	こんにちは。いいお天気ですね
ありがとう	どういたしまして
疲れた	ゆっくり休んでください
コーヒー	ミルクを入れてください

● マルコフ型

発話の中にある単語が出現した際、次にどのような単語が出現するかは、発話中に出現した単語によってある程度絞ることができる。このような、ある状態から別の状態へ遷移する時に、その遷移の確率が

直前の状態に依存するモデルのことをマルコフモデルと呼ぶ。このモデルを用い、発話を形成するのが、マルコフ型対話システムである。文を形態素解析して単語単位に分割し、後続の単語とセットで保存し、出現頻度を求める n -gram を用いて発話辞書を生成する。ユーザの発話にある単語を注目語とし、その単語を含む n -gram の組み合わせが辞書に保存されていた場合、マルコフ連鎖で辞書内の文章を組み合わせ、応答を生成する。以下に、文章を分割した例を示す。例えば「東京へ行きました」という文は2単語を接頭辞として表2のようにモデル化される。

表 2: 2-gram によるモデル化

単語 1	単語 2	→	単語
東京	へ	→	行き
へ	行き	→	まし
行き	まし	→	た
まし	た	→	END

3. 対話システムの客観的評価 [1]

3.1 対話の構造

対話や文章で内容やその流れを理解するためには、文章構造を理解することが重要な役割を果たす。対話における発話情報を表す指標として DAMSL がある。DAMSL は、対話における各発話に対してその種類を記述するためのタグセットである。例えば、「ありがとう」という発話に関しては「Thanking」のタグが付与され、「車に詳しくありませんか？」という発話には「Question」というタグが付与される。本稿でも、タグの種類を 14 種にまとめた表3に示す簡易 DAMSL[1] を使用する。

3.2 HMM による評価

HMM (Hidden Markov Model) は、不確定な時系列のデータをモデル化するための有効な統計的手法である。ここではタグの出現時系列のモデル化に利用する。HMM は以下の 5 項で表現できる。

- $\Sigma = \{o_1, \dots, o_M\}$ タグの有限集合
- $Z = \{z_1, \dots, z_N\}$ 状態の有限集合 (潜在変数)
- $A = \{a_{i,j}\} (\sum_j a_{i,j} = 1)$ 遷移確率の有限集合
- $\mu = \{\mu_{N,M}\}$ タグの出現確率分布
- $\pi = \{\pi_i\}$ 初期状態確率分布

ここで、 M は付加するタグ数、 N は状態数、 i, j は任意の状態である。これらの集合を用い、EM アルゴリズムによる最尤推定で出力確率と遷移確率のパラメータを更新する。

[†] 東京理科大学大学院理工学研究科情報科学専攻

[‡] 東京理科大学理工学部情報科学科

表 3: 簡易 DAMSL タグ

	タグの種類
1	Uninterpretable
2	Self-talk
3	3rd-party-talk
4	Statement
5	Question
6	Directive
7	Influencing-addressee-future-action
8	Committing-speaker-future-action
9	Other-forward-function
10	Thanking
11	Apology
12	Agreement
13	Understanding
14	Other

4. 非タスク指向型対話システムの比較評価実験

4.1 実験項目とデータ

実験では以下の項目について検証する。

- 各対話の出力確率の比較
- タグの出現頻度と出力確率の関係

比較に利用した非タスク指向型対話システムは、ログ型としてロイディ[2]、辞書型としてハル[3]、マルコフ型としてししゃも[4]という既存のものを使用した。また、元となる人同士の対話データは、名大会話コーパス[5]を利用した。その中から、自由な対話が行われると期待できる大学生同士の対話を 29 対話、5005 の発話を経験に基づき手動で簡易 DAMSL のタグ付与を行った。タグ数の増加を防ぐために、各発話に対して付与するタグ数は 2 つまでとした。

4.2 各対話の出力確率の比較

本実験によって生成された非タスク遂行型対話システムの HMM による出力確率の結果を示す。

表 4: 出力確率の比較

状態数	人	ログ型	辞書型	マルコフ型
5	78.5%	66.3%	61.1%	62.8%
10	78.2%	66.0%	60.8%	62.5%
20	61.4%	55.9%	53.3%	53.5%

表 4 より、状態数の変化にかかわらず、人対人の対話が最も出力確率が高いことから、HMM における非タスク遂行型対話システムの評価は可能であることが確認できた。ログ型、辞書型、マルコフ型のそれぞれの出現確率について、表 4 よりログ型の出現確率が最も人対人の対話に近いことがわかるが、実際、入力系列に用いた対話を主観的に見たときもログ型が最も人対人の対話に近

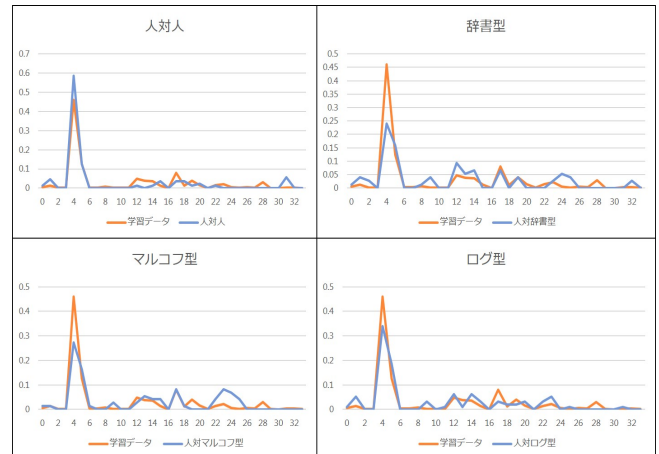


図 1. 各対話におけるタグの出現頻度

いと感じられた。また、人対人の対話とその他の対話の出力確率の差異は 10% 前後である。差異が小さい原因としては、人対人の対話においても、タグの並びが HMM の学習データと乖離すると、その時点で出力される確率が極端に小さくなってしまふことが考えられる。

4.3 タグの出現頻度と出力確率の関係

図 1 に、各対話におけるタグの出現頻度をグラフで示す。横軸は出現するタグを表し、縦軸は相対頻度を表す。人対人の対話は、出現頻度のグラフが学習データに最も近いことがわかる。しかし、図 1 からわかるように、出現タグ 4,5 にあたる「Statement」「Question」タグが付与されたタグの半数近く占めている。精度の高い結果を得るためには、この 2 種のタグをさらに細分化するか、さらに対話の深い構造に着目することが必要である。

5. おわりに

HMM により、非タスク遂行型対話システムを定量的に評価することは可能である。本稿では、この考えに基づき、発話の種類をタグ付けしたものを学習データとして用い、各種の非タスク遂行型対話システムを定量的に比較評価できることを示した。

参考文献

- [1] 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”, 電子情報通信学会論文誌, Vol.J92-D, No.4, pp.542-551, 2009
- [2] ロイディ <http://magazine.rubyist.net/?0035-Reudy>
- [3] ハル <http://www.okiniiri.com/chat/>
- [4] ししゃも <http://ashitani.jp/sixamobbs/>
- [5] 名大会話コーパス <https://dbms.ninjal.ac.jp/nknet/ndata/nuc/>