

報告番号	※甲	第	号
------	----	---	---

## 主 論 文 の 要 旨

論文題目 非タスク指向型対話エージェントの設計と評価に関する研究  
氏 名 磯村直樹

### 論 文 内 容 の 要 旨

近年、コミュニケーション不足、コミュニケーションの断絶を原因とする様々な現象が顕在化し、大きな社会問題となっている。親しみのあるコミュニケーションに対する人間の期待は大きく、そのため、ペットロボットのような、見た目やしぐさにより癒しを与えるロボットの研究が盛んになりつつある。しかし、話し相手になることを目的とした対話エージェントの研究は少なく、対話エージェント研究の発展の余地は非常に大きい。そのため、コミュニケーション能力を向上させるための話し相手として、また老人に癒しを与えるための話し相手として今後、一層、人間と対話を行うコンピュータ(対話エージェント)の需要が増すものと考えられる。

従来の対話エージェントの研究は、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エージェントが中心であり、人間の話し相手になり、対話を盛り上げる非タスク指向型対話に関する研究は少なく、未だ発展途上にある。タスク指向型対話エージェントについては、上で述べた自動応答システムなど実現されてきているが、非タスク指向型対話エージェントは、対話を破綻せずに続けることさえ困難である。そのため、質の高い質問を通じて対話を盛り上げ、対話相手を楽しませるような、人間らしい対話を可能とするレベルに達していないのが現状である。そこで、本研究では、質の高い質問を通じて対話を盛り上げ、対話相手を楽しませる非タスク指向型対話エージェントの実現を目指す。

本研究の目標とする対話エージェントKELDIC (Ken's Laboratory Dialogue Computer) は、質の高い質問を通じて対話を盛り上げ、対話相手を楽しませることを目指している。そのため、対話相手について十分調査し、綿密に練り上げられた質問を大量に作成する必要があると考えられる。これは、優れたインタビューの振舞いを模したものである。そこで、綿密に作成された質問のように、あらかじめ発話候補が大量に作成されていると仮定し、その中から発話を選択する手法の確立を目

指す。

まず、対話相手の情報、Wikipediaや単語 n-gram データベース、これまでの対話内容を基に発話集を生成する。対話相手の情報は、対話相手のブログや著書などから自動的に取得する。特に、趣味や特技などの情報を利用することで、対話を盛り上げる発話を生成可能であると期待される。さらに、Wikipediaなどの一般的知識を利用することで、発話の種類を増やし、幅広い話題に対応することが可能となる。

次に、生成した発話集の中から、最適な発話を自動選択し、KELDIC の応答とする。最適な発話を自動選択するためには、発話集に含まれる発話を比較評価する手法が必要となる。対話の内容に応じて発話に評価値を付与し、評価値の最も高い発話を最適な発話とすることで、発話集を利用した対話が可能となる。

そして、対話エージェントを評価することで、発話集の自動生成法、自動選択法を改善していく。特に、客観的・定量的な評価は必須な技術であるが、確立されていないのが現状である。

以上のプロセスのうち、本研究では最適な応答の自動選択と、非タスク指向型対話エージェントの評価を対象とする。そこで、本研究の目的を以下のように定める。

#### **【目的1】非タスク指向型対話エージェントの客観的・定量的な評価法の確立**

#### **【目的2】非タスク指向型対話における発話自動選択法の確立**

【目的1】(その1) 非タスク指向型対話エージェントを客観的・定量的に評価するため、対話エージェントの行った対話の自然さを評価した。本手法では、対話の評価するにあたり、いわゆる対話の「浅い構造」にのみ着目し、発話間の繋がりという最低限の自然さを評価した。前提として、人間同士の対話は自然で理想的な対話であると仮定し、人間同士の対話と、人間と対話エージェントの対話との類似度を計算した。

まず、対話に、談話の浅い構造を表現できる Switchboard Discourse Annotation and Markup System of Labeling(SWBD-DAMSL) タグをまとめたものである簡易 DAMSL タグの系列を付与し、対話をタグの系列で表現した。次に、類似度の計算のため、隠れマルコフモデル (Hidden Markov Model(HMM)) によって人間同士の対話タグ系列をモデル化した。評価すべき対話(たとえば人間と対話エージェントとの対話)をこの HMM に入力し、HMM の出力確率を計算することにより、対話の自然さを評価した。

対話の自然さの評価実験では、人間による主観評価と HMM による評価を比較した。その結果、人間に近い自然な対話を行うと主観評価された対話エージェントほど、HMM の出力確率が高くなることを確認した。

さらに、複数の対話エージェントを評価することにより、人間同士の対話は人間と対話エージェントとの対話より HMM の出力確率が高くなり、その中でも自然な対話ほど出力確率が高くなることを確認した。これによって、HMM の出力確率の大小によって対話の自然さを評価できることが明らかになった。

本手法は、非タスク指向型対話エージェントを評価の対象としているが、タスク指向型対話エージェントを評価することも可能である。これまでのタスク指向型対話エージェントの評価法では、タスクの達成度や達成時間という側面を評価してきたが、本手法を用いることにより、対話の自然さという側面を評価することができる。

ただし、本実験では HMM による評価法の有効性のみを確認するため、タグ付与を全て手動で行い、タグ付与による誤りはないものとしている。しかし、実際に本手法を適用する場合には、全ての処理を自動化する必要があるため、タグ付与は自動化する必要がある。

【目的 1】(その 2) そこで、非タスク指向型対話エージェントの評価法を発展させ有用性を高めるため、タグ付与の自動化法を提案した。まず、一定量の対話に手動タグ付与を行い、標準データとした。標準データを用いてテキスト対話に対して自動タグ付与を行った。自動タグ付与手法として、DICE 係数を用いた手法、情報量を用いた手法、Naive Bayes を用いた手法、SVM を用いた手法、CRF を用いた手法の 5 種類を提案し、比較実験を行った。実験の結果、CRF を用いた手法が最も精度が高く、正解率は 75.77%、意味的に近いものを許容した場合で 83.20% であった。また、 $\kappa$  統計量は 0.6371 であった。以上のことから、CRF を用いた自動タグ付与の有効性が確認できた。

そこで、CRF を自動タグ付与手法として使い、HMM による対話の評価を行った。まず、手動でタグを付与した標準データを用いて、CRF のパラメータを学習した。CRF を用いて各発話に自動タグ付与を施し、全ての対話をタグ系列として表現した。

そして、人間同士の対話のタグ系列を学習した HMM を作成した。評価すべき対話をこの HMM に入力し出力確率を計算することにより、対話の自然さを評価した。実験の結果、より人間に近い自然な対話を行う対話エージェントほど HMM の出力確率が高くなることを確認した。さらに、人間同士の対話は人間と対話エージェントとの対話より HMM の出力確率が高くなり、その中でも自然であると主観評価された対話ほど出力確率が高くなることを確認した。これは自動タグ付与を行わず、手動タグ付与のみで対話の評価した際の結果と同等の結果であり、本手法の有用性を確認した。

【目的 2】続いて、非タスク指向型対話の発話を自動選択するため、機械学習アルゴリズムを用いた、対話エージェントのための統計的な発話選択法を提案した。

発話選択を、入力に対して大量に用意された発話候補集合の中から、応答として適切な発話を選択する問題と定義した。発話候補を自動評価するために、入力と発話候補の対に対して、「適切」または「不適切」を手動で評価し、教師データとした。教師データから発話候補の評価値の相対的な大小関係を学習し、発話選択を行った。本手法の有効性を評価するために、発話候補に対する手動評価の結果と自動評価の結果を比較した。実験の結果、入力発話に対して適切な応答が1位となる割合は47.5%、発話候補中上位10位以内に適切な応答が含まれる割合は78.0%となることを確認した。ただし、対話エージェントに本手法を実装する場合は、適切な応答が1位となることが重要であり、本手法にはまだ改善の余地がみられる。

一方、人間の応答を補助するエージェントに本手法を実装する場合は、エージェントの提案する候補に適切な応答が含まれることが重要である。そのため、上位10位以内に適切な応答が78.0%と多く含まれている本手法を用いることにより、このようなエージェントを実現できる可能性が確認できた。