

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の評価と対話エージェントの設計に関する研究

氏名 稲葉 通将

## 論文内容の要旨

近年、人間と対話を行うコンピュータ(対話エージェント)に関する研究が活発化している。しかし、その多くはチケット予約や道案内など、特定のタスク達成を目的としたタスク指向型の対話エージェントに関する研究である。

しかし、我々の身の回りのコミュニケーションの大半は雑談のような非タスク指向型のコミュニケーションであり、それは人間関係の形成に非常に大きな役割を果たしている。今後、擬人化エージェントやロボットが人間社会へますます進出していくことが期待されている。それらが社会に受け込み、人間と良好な関係を形成するためには、非タスク指向型対話エージェントの研究は不可欠である。

非タスク指向型対話エージェント研究の必要性は日増しに高まっているものの、現状では非タスク指向型対話エージェント研究を行なっている研究者・研究機関は非常に少なく、人間らしい応答が行えるエージェントは皆無である。

そこで、本研究では人間らしい知的な応答が可能な非タスク指向型対話エージェント実現を最終目標と設定する。そのために、本研究では「①研究基盤技術の開発」と「②非タスク指向型対話エージェントの設計」を行う。

まず「①研究基盤技術の開発」であるが、前述したように、現在非タスク指向型対話エージェント研究は活発に行われているとは言い難い状況にある。研究が活発でない原因はいくつか考えられるが、非タスク指向型対話エージェント研究における研究基盤が未整備であることが、その大きな原因の1つであると考えられる。

必要な研究基盤はいくつか考えられるが、本研究では「非タスク指向型対話の客観的・定量的評価法」を対象とし、研究を行う。

対話エージェントを設計するためには対話エージェントの性能の客観的・定量的な評価が不可欠である。客観的な評価法を用いることで、新たに設計した対話エージェントの性能を既存の対話エージェントと比較することが可能となり、研究として体系化することが容易となる。また、定量的な評価法を用いることで、対話エージェントに施した修正点が性能向上にどの程度有効であったかを分析できる。これにより、修

正と評価を繰り返すことで対話エージェントの性能を効率的に向上させることが可能になる。

次に「②非タスク指向型対話エージェントの設計」であるが、本研究では非タスク指向型対話エージェント KELDIC(Ken's Laboratory Dialogue Computer)の実現を目指している。KELDIC は良い聞き手となることを目指した非タスク指向型対話エージェントである。そのために、KELDIC はトークショーにおけるインタビューを模し、エージェントが相手に応じた適切な話題を提供し、相手の話を引き出し、適切な反応を返すことによって対話を盛り上げることを目指す。良い聞き手であるインタビューは、どのような話題で話を進めるか、どのような発話を行うか、何を聞き出すかといったことをあらかじめ考え、周到に準備を重ねた上で対話に臨む。そこで、KELDIC でもその手法を模して対話を行う。つまり、KELDIC はあらかじめ発話候補を大量に用意しておき、その時々状況に応じて適切な発話を発話候補集から選択することで対話を進める対話エージェントである。

KELDIC の実現を目指し、本研究では統計的応答手法の提案も行う。近年、コンピュータの性能向上によって、大規模データを用いた統計的手法が自然言語処理や画像処理など様々な研究分野で主流となり成功をおさめてきた。本研究では、この統計的手法を非タスク指向型対話エージェントの応答に適用する。現在、非タスク指向型対話エージェントの応答手法は、人手で記述された応答ルールを用いるルールベース応答手法が主流である。ルールベース応答手法は、応答ルールの作成は非常に高コストであるという問題がある。コストの問題を解消し、また、非タスク指向型対話エージェント研究の今後の方向性を示すためにも統計的応答手法の提案は重要な意味を持つ。

なお、本研究で扱う対話はテキスト対話とする。これは、音声認識・音声合成のエラーの影響を無視し、研究対象を言語処理に特化するためである。また、対話は1対1の対話とする。すなわち、本研究では人間同士の対話、もしくは人間と対話エージェントの対話を研究対象とする。

以上を踏まえ、本研究では以下に示す2つの対話評価法と応答手法を提案する。

#### 【評価法1 (①-1)】発話の意味的な正しさの評価法

#### 【評価法2 (①-2)】対話の盛り上がりの評価法

#### 【応答手法 (②)】統計的応答手法

【評価法1 (①-1)】対話エージェントの発話の意味的な正しさという観点から対話エージェントの性能評価を行う手法である。発話の意味的な正しさとは、「ある発話が前後の文脈からみて論理的に矛盾がなく、意味的に自然であること」と定義した。本手法では、対話の中で語彙的結束性を有する部分は意味的に正しい対話が行われていると仮定した。語彙的結束性は結束性を持つ語、すなわち意味的なつながりのある語の出現により、テキストの意味的なつながりを明示する表層的な情報である。意味的なつながりのある語を獲得するため、大規模コーパスから共起語対を獲得した。次に共起語対が発話間に含まれるかどうかによって語彙的結束性の判定を行い、対話エージェントの生成する発話の評価を行った。

複数の対話エージェントを用いた評価実験の結果、人手によって性能が良いと評価された対話エージェントほど評価値が高くなることを確認した。また、人手による評価と提案

手法による評価値との間に相関があることを確認した。相関係数は 0.81 であり、人間同士の対話を除いた場合でも相関係数は 0.50 であった。

【評価法 2 (①-2)】既存の非タスク指向型対話エージェントの性能は、意味的に正しい発話を十分に行うことができる段階には達していないことから、本論文ではまず、発話の意味的な正しさの評価法を提案した。ただし、将来的には非タスク指向型対話エージェントは意味的に正しい対話が可能なだけでは十分とは言えず、話を盛り上げ、相手を楽しませることにより、人間が対話を続けたいと思えるような対話を展開することが求められる。

しかし、どのようにすれば盛り上がる対話が展開できるのかは自明ではない。それを明らかにするためには、人間同士の対話を分析し、対話の盛り上がりに関する何らかの知見を得ることが必要である。その知見は対話エージェントの設計に役立つだけでなく、人間同士のコミュニケーションの支援にも有効であると思われる。しかし、対話の盛り上がりを手で判定するのは極めてコストが大きい。また、盛り上がりの判定基準は個人差が大きいと思われるため、首尾一貫した判定を行うためにも、自動判定法が必要である。

提案手法は、人間同士の対話における盛り上がり度を CRF(Conditional Random Field) を用いて自動判定する手法である。本手法では、盛り上がり度の判定を「盛り上がり度：高」と「その他」という 2 種類のラベルを用いて、それらを発話系列に付与する系列ラベリング問題として解くことにより行った。なお、盛り上がりの度合いは「話者がある発話を行った時点で、対話を継続することに対してどの程度積極的であるか」と定義した。

本手法では、発話間の語彙的結束性を素性として用いることにより、対話中の各部分発話系列において、どの程度の密なやりとりがなされたのかや、対話中の話題ごとのまとまりの分布を把握することを目指した。

評価実験の結果、提案手法は人手による評価と同等の性能を示すことを確認した。また、語彙的結束性の素性を用いた場合と用いない場合の実験を行い、語彙的結束性の素性の有効性を確認した。

【応答手法(②)】非タスク指向型対話エージェントの応答手法の新たな試みとしてランキング学習を用いた統計的応答手法を提案した。提案手法では、文脈とそれに対する応答候補から素性の抽出を行う。そして、その素性を用いて、主に情報検索の分野で研究が進められているランキング学習によって適切な応答を実現する。

ランキング学習とは、複数のデータの中から目的のデータを求める際に、目的のデータが上位に順位付けされるよう学習を行う教師あり学習法である。例えば文書検索では、ある検索語に対して適切な順に文書を順位付けするためにランキング学習を用いることができる。提案手法ではこの例における検索語を文脈、文書を応答発話候補としてランキング学習を適用する。すなわち、本研究で構築を目指す対話エージェントは、あらかじめ大量の発話候補を用意しておき、ランキング学習を用いて、それまでの対話の流れに応じて発話候補を順位付けし、最も上位に順位付けされた発話を選択することで対話を進める。

なお、ランキング学習には、様々なデータセットを用いた実験で他のランキング学習手法と比べて高い性能が確認されている ListNet を用いた。評価実験の結果、70.1%の文脈に対して、自然な発話候補を 1 位に順位付けできることを確認した。