

認知タスク分析のためのインタビューエージェント に関する研究

An Interview Agent for Cognitive Task Analysis

越智 悠馬 菅野 太郎 古田 一雄

Yuma Ochi, Taro Kanno, and Kazuo Furuta

東京大学大学院工学系研究科 システム創成学専攻

Department of Systems Innovation, School of Engineering, the University of Tokyo

Abstract: This paper introduces an interview agent system that can conduct Cognitive Task Analysis for disaster nursing. This system generates queries and responses based on, 1) an interview guide based on Critical Decision Method (CDM), 2) typical patterns in Japanese interview dialogue, and 3) several practical techniques for verbal interviews. The system was tested by mock interviews and the results partly showed the effectiveness of the response generation algorithm and the potential of this system.

背景

様々な分野で、アンケート調査やグループインタビューなどの人から知識や情報を収集する手法が実践されている。小林ら[1]による調査では、対話を行うことで人からより多くの情報が引き出せることが報告されており、様々な情報収集手法の中でも対話によるインタビューは幅広い分野で活用されている。例えば、認知工学の分野では、認知タスク分析 (CTA : Cognitive Task Analysis) においてインタビューが主要な手法として用いられている。認知タスク分析は、インタビューや観察により専門家のタスク遂行や問題解決、意思決定の際の思考プロセスを分析するための方法論である。

しかし、一般的にインタビューは、そこで得られる情報の質と量がインタビュアの技量に大きく依存し、また、その技能習得の効率的な方法は確立されていない。さらにも、インタビューそのものに必要な時間と労力も非常に大きく、これを多人数に対して行うことが難しい等、その適用には限界がある。

このような背景のもと本研究では、人に代わりエージェントがインタビュー、特に認知タスク分析を行うインタビューエージェントシステムを提案し、効率的な情報収集を支援する手法の開発を行う。開発にあたって、このシステムの適用領域を災害現場の看護活動とし、インタビューの対象を看護師、インタビューによって抽出する情報を、「震災時の看護活動における、看護師が行動に至るまでの思考プロセス」とした。

関連研究とその課題

対話エージェントに関連する研究は従来から数多く行われている。そのほとんどは、何らかの目的のための対話を志向するものではなく、純粋に対話を楽しむためのチャットエージェントに関する研究である。一方、特定のタスクを志向した対話エージェントに関する研究としては、例えば渡辺ら[2]の提案する航空券の予約サービスなどがある。しかしながら、これらのタスク指向型の対話エージェントは領域依存性が高く、構築の際にそれぞれのタスクに応じたシナリオやルール、辞書を作り込むのに大きな労力を要するという課題がある。

また、本研究と同様に対話エージェントによる知識抽出を目的とした研究も行われている。例えば、鳥澤[3]の研究では、対象領域を限定しないインタビューを行う対話エージェントの構築を実現している。しかし、エージェントの応答生成アルゴリズムやインタビューのモデルが単純なため、対話の効用やインタビューという文脈による効用を十分に活用できてはいない。

一般的に対話エージェントは、発言中の特定のキーワードに反応して予め辞書に登録された応答を返す辞書型の人工無能と自然言語処理技術を基礎技術として利用している。そのため対話エージェントを用いて目的的に情報収集を行う場合、応答生成の自由度が高すぎることや、文脈を考慮した発話をする事ができないことが問題となる。また、辞書の内容を学習することができても、それを用いて応答を

生成するルール自体を学習することができない点も、人工無能型の対話エージェントの限界であると言える。

以上で述べた問題点を表 1 にまとめた。

本研究では、インタビューに関する様々な分野の知見および技法を統合し、対話エージェントシステムに導入することで、表 1 に示した従来の対話エージェントの研究課題を解決することを目指す。

次節では本研究で用いるインタビューに関する知見および技法について説明する。

表 1.従来の対話エージェントの問題点

項目	内容
自由度	応答生成の自由度が高すぎる場合、特定のタスクを行うことが困難となる
文脈の考慮	文脈を考慮できない場合、場あたりのな応答しか生成できない
対話の効用	応答生成アルゴリズムが単純すぎる場合、対話の効用を活用できない
学習	応答生成ルールそのものを学習することができない

本研究で用いるインタビューに関する知見および技法

認知タスク分析

認知タスク分析とは問題解決行動や意思決定行動の際の思考過程を分析するための方法論である。この技法をインタビューエージェントに導入し、応答生成に一定の規則を持たせることで、表 1 で示した「自由度」に関する問題の克服を試みる。

本研究では Klein ら [4] が提案した、Critical Decision Method (CDM) と呼ばれる専門家の知識抽出を目的としたインタビュー手法を用いる。この手法では、まずシナリオを選択しインタビューイが経験したインシデント(出来事)や観察されたインシデントを詳細に分析する。その後シナリオをフェーズに細かく分割し、インシデントをタイムラインで整理する。そして整理されたインシデントについて、CDM Probe と呼ばれる質問項目を用いた半構造化インタビューを行う。

情報収集の話段展開の分析と機能文型

話段とは、談話の内容と目的の相対的なまとまりから区分される談話の成分である。発話機能とは

「個々の文が発話中に用いられて果たす機能」で、ザトラウスキー[5]はこれに着目し発話表現を 12 類 22 種に分類している。

小林[6]は、この発話機能分類を用いてインタビュー内容を分析し、インタビューにおける話段構造と発話機能の出現パターンを明らかにした。この分析結果を対話エージェントに導入し、人間のインタビューが行うインタビューに近い文脈制御を実現することで、表 1 で示した「文脈の考慮」に関する問題の克服を試みる。また小林は発話機能ごとの表現の類型である機能文型も示しており、これを応答生成ルールに組み込むことでより多様で自然な応答の出力が期待できる。

インタビューの技法

人間が行ったインタビュー内容を分析することで、その特徴を洗い出した。例えば、インタビューイの回答をオウム返しすることでさらに詳細な情報を引き出す技法や、回答を予想して行う質問を効果的に使う技法など、対話の効用を十分に発揮するための技法が数多く発見された。これらに対話システムに実装し、より効果的なインタビューを行うことを目指す。

応答生成ルールの学習

インタビューに関する複数の知見及び技術を用いた応答生成を行うため、応答生成には多数のルールが関係する。インタビューを行う際、どのようなルールで応答生成を行うか、またはこれらのルールをどのように組み合わせて用いるかを学習することで、表 1 に示した「学習」に関する問題の克服を試みる。

次節ではインタビューエージェントシステムの概要と各モジュールについて説明する。

提案システム

提案するシステムの構成を図 1 に示す。実装には Ruby を用いた。「知識ベース管理」モジュールは知識ベースからの情報の入出力を行う。「文脈制御」モジュールが文脈の制御を行い、「応答生成」モジュールが文章を生成し、出力する。「記録・学習」モジュールは知識ベースの更新と学習を行う。

なお、本研究ではインタビューに関連する言語的要素の研究に特化するため、音声認識や音声合成技術については扱わず、テキストによる対話のみを扱う。

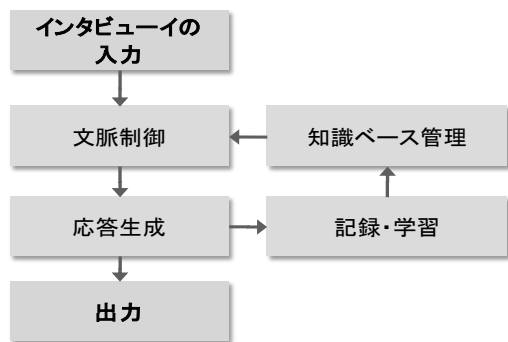


図 1. システム構成

知識ベース

知識ベースには領域オントロジーとインタビューガイドが含まれる。

インタビューガイドは図 2 に示すようにマトリックスで表現され、列の項目に「因子」、行の項目に「概念」という要素をもつ。「概念」の項目は、認知タスク分析の CDM Probe（質問項目）と対応している。このインタビューガイドに基づき、それぞれの「因子」について「概念」の項目である情報、判断、行動がどのように行われたかを質問する。

具体的には、インタビューイの入力をインタビューガイドの因子の項目と関連付け、その項目について、インタビューイが情報を得てから判断を下し行動に移るまでのプロセスがどのようであったかを質問する。つまり、インタビューイの入力と、出力する質問をインタビューガイドのマトリックスにマッピングすることで、発話意図解釈や次の質問生成の際の話題の決定を行っていると言える。

この際領域オントロジーを利用し語彙間の関係性に関する情報を利用することでより精度の高いマッピングを行う。また、知識ベースとオントロジーは XML で記述し外部ファイルから読み込む方式とした。これにより、知識ベースの編集を行うことで領域に依存しない認知タスク分析用ツールとしての活用が可能になる。

		参加者	タスク	環境			
		看護師	患者	権限 ... 施設 ...			
情報	徴候	「看護師」の行動について教えてください			
	タイミング						
判断	優先度				その時の目標は何でしたか？
	必要性						
行動	目標			
	手順						

図 2. インタビューガイド

表 2. 発話の種類

1	新しい話題について質問を開始する発話
2	説明要求の発話
3	追求質問でより深く尋ねる発話
4	判定要求および確認要求の発話
5	促しによってさらなる情報を引き出す発話
6	質問の意図が理解されなかった時などの異常状態修復のための発話
7	質問終了の発話

文脈制御

本研究ではインタビューの談話における発話を 7 種類に分類した。発話の種類の一覧を表 2 に示す。これらを文脈における個別の状態と捉え、文脈制御を状態遷移モデルとして実装した。情報収集の話段展開の分析結果に従い、人間のインタビューにおける発話の種類の出現パターンに近い文脈の推移になるよう状態遷移確率を与えた。

応答生成

文脈制御モジュールで発話の種類が決定された後、応答生成モジュールで出力される文章が生成される。この際、発話機能ごとの表現の類型である機能文型を用いることで、自然で多様な応答や質問の生成を行う。また追求質問でより深く尋ねる発話を行う際に、Critical Decision Method の質問項目である CDM Probe を用いることで、インタビューイがタスクの遂行や意思決定を行う際の思考過程を明らかにする。

記録・学習

記録・学習モジュールでは、インタビュー結果の記録と、文脈制御と応答生成ルールの学習を行う。インタビュー結果の記録は、インタビューエージェントの質問とインタビューイの回答、さらにインタビューガイドのマトリックスのマッピング情報を 1 セットとし、CSV 形式で保存する。このような形式で記録することにより、インタビューによって得られた情報を分析する際や、別のインタビューに対してインタビューを行う際に、記録された情報を効率的に再利用することが可能となる。

文脈制御、応答生成ルールの学習は、インタビューイからのフィードバックを用いて行う。インタビューイは、インタビュー中に質問や話題を変更する権限を持っており、システムの入力に対し、インタ

ビューイが不自然さ等を感じた場合、これを矯正することができる。この際システムは応答生成のルールを学習する。

インタフェース

インタフェースを図3に示す。画面には、親しみやすさを考慮しキャラクターを表示した。また、対話の履歴だけでなく、現在の話題や関連語句を表示し、インタビューイに刺激を与えることでより多くの情報を引き出すことを狙う。

さらにインタビューイがインタビューガイドを随時拡張するためのインタフェースも備えており、柔軟に新しい話題に対応することが可能であるため、半構造化インタビューの利点を活かすことができると考えられる。また、図4に示すように、インタビュー中にインタビューガイドのどの部分の話題に触れたかを、マトリックスのセルの色の濃淡によって可視化する。これによりインタビューガイドの網羅度を視覚的に把握することが可能になる。



図3. 画面イメージ

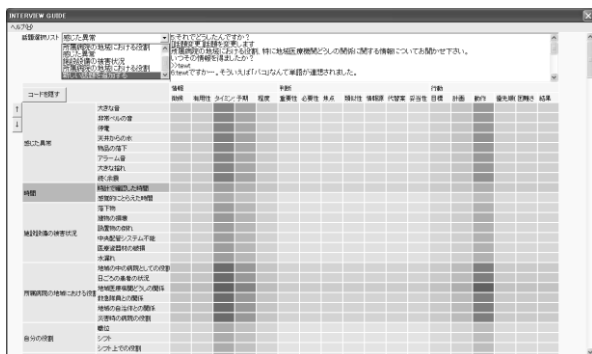


図4. インタビューガイドの可視化

模擬インタビュー

実装したシステムの性能を確認するため、模擬インタビューを行った。模擬インタビューは、「震災時の看護活動における、看護師が行動に至るまでの思考プロセス」を明らかにすることを目標とし、大学院生2名に対して行った。この際、インタビューイは震災の体験者の手記等が掲載された資料[7]を参考にして質問に答えることで、知識の不足を補った。

模擬インタビューは30分間行った。その対話ログの一部を表3に示す。ここで、Userはインタビューイ、Agentはインタビューエージェントを表している。

3行目と5行目においてインタビューエージェントはオウム返しにより、インタビューイのさらなる発言を引き出している。このようにして特定の話題をについて掘り下げた後、6行目のインタビューイの発言を受けて、インタビューエージェントはインタビューガイドの「判断」にマッピングを行い、6行目においてその判断の重要性についてCDM Probeを用いた詳細な質問を行っている。

表3. 模擬インタビューのログ

1	Agent:	そうなんですか。続けてください。
2	User:	で、とりあえず、受入先に搬送しないといけないので、その振り分けをしました。
3	Agent:	非常に興味深いです。振り分けですか。例などを交えてより具体的に教えてもらえますか。
4	User:	やはり患者さんの容体によって緊急度が異なりますから。医者の指示のもと、緊急性の高い患者さんは他の病院に移ってもらいました。
5	Agent:	なるほど。容体ですか。それはどうしてですか？
6	User:	緊急性の判断が必要でしたから。人工呼吸の患者さんは特に
7	Agent:	そうですね。その時の判断に関する質問です。その判断は重要なものでしたか？
8	User:	人命にかかわるので非常に重要と思います

表 4. 評価項目

対象	評価項目
対話エージェント	インタビューイがインタビュー中に感じた「自然さ」や「人らしさ」
	対話中の出現単語数や、対話のターン数
インタビュー	質問の順序
	質疑応答の適切さ
	倫理的規範
インタビューエージェント	人が行ったインタビューとの比較
	インタビューガイドのマトリックスの網羅度
	学習結果と、インタビューイが行った質問や話題の変更の回数

評価

評価項目の設定

対話エージェントの特徴と、インタビューの特徴にそれぞれ着目して評価項目を設定した。これに加え、インタビューエージェント独自の評価項目を設定した。これらを表 4 に示す。

多層評価モデル

対話エージェントやインタビューの評価手法に関して個別の評価指標は存在するが、総合的な評価を行うための指標はない。そこで本研究ではインタビューエージェントを総合的に評価するための、多層評価モデルを構築した。このモデルは、Reaction, Quantity and Quality, Usefulness, Results の 4 つの階層から成り、それぞれ「インタビューを行った際の表面的なインタビューイの反応」、「インタビューによって得られた情報の質と量」、「得られた情報の有用性」、「得られた情報が組織や社会に与えた影響」を評価する。このモデルではインタビューエージェントによるインタビューそのものの評価は Reaction と Quantity and Quality の階層で行われ、得られたデータの評価は Quantity and Quality, Usefulness の階層で行われる。これに従い、上述した評価項目を各階層に対応させ評価することで、インタビューエージェントの評価を多面的に行うことが可能となる。これを表 5 に示す。

表 5. インタビューエージェントの評価モデル

階層名	説明	評価項目
Reaction	インタビューイの反応	インタビューイへのアンケート調査
		対話ログの観察
Quantity and Quality	得られた情報の質と量	出現単語数
		対話ターン数
Usefulness	得られた情報の有用性	インタビューガイドの網羅度
		人が行ったインタビューとの比較
Results	組織や社会への影響	CDM Probe によって抽出した知識や経験
		現場での適用結果の観察

模擬インタビューの評価

多層評価モデルを用いて模擬インタビューの評価を行った。人間が行ったインタビューとの比較にあたっては小林の日本語のインタビューの分析結果を用いた。結果を以下に示す。

1) Reaction

インタビューイへのアンケートと、対話ログの観察から以下の事柄が明らかとなった。

質問の意図は分かりやすく、言い回しも豊富なためインタビューイの思考に沿った質問を行うことが出来た部分については「最初に大まかな質問をなげかけた後、特定の話題について掘り下げ、CDM Probe を用いて質問をする」という一連の流れを自然な文脈の中で行うことが出来ている。しかし、それ以外の部分では話題の変更の仕方に唐突感が強いなど、文脈の制御が不自然な点が目立った。そのため、双方向的なコミュニケーションの意義が薄れている。

また、インタビューイが質問や話題の変更を行ったのは全体の約 36% の出力に対してであった。このようにインタビューイからの能動的な話題の変更を行った場合、インタビューイからの回答が多く得られていることがわかったため、今後応答生成ルールの学習が進むことで、文脈制御の改善が期待される。

2) Quantity and Quality

対話ターン数は 30 分で平均 16.5 ターン、発話数は 832 文字であった。これは、人が行ったインタビューと比較して少ない結果である。

インタビューイの発話の総数はインタビューアの 1.3 倍であった。小林の調査によれば、日本語のイン

インタビューにおけるインタビューイの発話数はインタビューアの1倍から2倍であるため、本システムを用いた場合の発話総数の割合は、人が行うインタビューと比較して大きな差は見られないと言える。

インタビュー中にインタビューエージェントが質問で触れた話題の数は、30分間で平均6.5個であったが、インタビューイの回答は一回で複数の話題に触れている場合が多いため、これらの話題についてそれぞれ適切に掘り下げて質問を行うことができれば、より円滑な対話が可能になると考えられる。

3) Usefulness

インタビューによって得られた情報を用いて、本研究の目的である認知タスク分析を行うためには、CDM Probeによって知識および経験を抽出することが不可欠である。しかし、今回の模擬インタビューでは、インタビューイの意図に沿って対話を掘り下げることができない場合が多く、CDM Probeを用いて質問を行い得た情報はわずかであった。

4) Results

この項目については、今後の課題とする。今後、看護師にインタビューを行う際、併せてインタビューエージェントシステムの現場での適用の可能性を調査する予定である。

以上の評価をまとめたものを表6に示す。

表6. インタビューエージェントの評価結果

階層名	評価項目	結果(30分間平均)
Reaction	対話ログの観察	36%の質問に対して変更
	インタビューイへのアンケート	個々の出力は自然 文脈制御が不自然
Quantity and Quality	対話ターン数	16.5 ターン
	発話総数	832 文字
	インタビューガイドの網羅度	6.5個の話題について質問した インタビューイの回答にはより多くの話題が含まれていた
Usefulness	CDM Probeにより抽出した情報	有効に活用できず
Results	現場での適用結果の観察	今後の課題

まとめ

インタビューエージェントを構築し模擬インタビューを行いその評価を行った。従来研究の課題に対し、インタビューに関する様々な分野の知見および技法を用いて対処した。評価の際に、インタビューエージェントのための評価モデルを構築した。そして模擬インタビューの評価結果から、これらの技術と知見を用いて妥当な質問応答生成が個々の出力については出来ていることを確認した。しかし、インタビューイの思考に沿った文脈制御の部分で課題が残ることがわかった。

今後は、インタビューイの思考に沿った文脈制御により効率的に情報を収集することを目標にシステムの改良を進めるとともに、模擬インタビューによる応答生成ルールの学習を行う。また実際に人によるインタビューを受けた看護師を被験者に、対話データを収集し、再度システムの評価を行う予定である。

またさらなる取り組みとして、対話データとシステムの学習結果の分析により効果的なインタビューについて考察することを検討している。

参考文献

- [1] 小林宏教, 川崎善司朗: 知識獲得のための対話方式に関する一考察, 電子情報通信学会論文誌, pp901-905, (2000)
- [2] 渡辺卓夫, 大澤一郎, 米澤明憲: 対話領域の独立性を指向した日本語対話理解システム, 人工知能学会誌, Vol.2, No.2, pp.206-213, (1987)
- [3] 鳥澤健太郎: 一般ユーザーにインタビューする対話エージェント, 情処研報, Vol.2007, No.76, pp.25-30, (2007)
- [4] Klein, Calderwood R., Macgregor D.: Critical Decision Method For Eliciting Knowledge, IEEE Transactions on systems man and cybernetics, Vol.19 ,No.3, pp.462-472, (1989)
- [5] ポリー・ザトラウスキー: 談話の分析と教授法 (I), 日本語学, (1986)
- [6] 小林英美: 日本語の談話のインタビューにおける情報収集の談話展開の分析, 早稲田大学大学院日本語教育研究科修士論文, (2007)
- [7] 財団法人小千谷総合病院看護部: 小千谷総合病院看護部活動記録, 小千谷総合病院, (2007)