

LLM を用いた日本語学習者支援システムの実装及び評価

亀田 隆雅¹ 馬 青¹

¹龍谷大学大学院先端理工学研究科

y24m002@mail.ryukoku.ac.jp qma@math.ryukoku.ac.jp

概要

生成 AI 技術の著しい進歩により、これを活用した製品やサービスが数多く登場している。我々はこれまで、LLM を用いて架空の日本での生活をシミュレーションする日本語学習者支援ゲームを提案し、複数のエージェントが連携することで、学習者にパーソナライズされた会話練習を提供できる可能性を示してきた。しかし、これまでの研究では、実用的なシステムとしての完成度や、評価手法の十分な検証という点で課題が残されていた。

本研究では、これらの課題に対応するため、提案手法を基に Streamlit を用いた Web アプリケーションとしてシステムを再構築し、誰でも利用可能な形で実装を行った。さらに、疑似学習者を用いた評価実験を系統的に行い、システムの有効性を多角的に検証した。

1 はじめに

外国人労働者の受け入れ拡大や日本文化への関心の高まりを背景に、日本語学習者の数は国内外で増加している[1]。日本語学習において、日常会話スキルの向上は重要である。しかし、従来の教室やテキストを中心とした学習方法では、実際の会話を再現し、文脈に応じた適切な表現を習得することが困難であるという課題が指摘されている。また、日本語教師の適性についても議論があり、学習者の目的やレベルに応じた柔軟な対応が求められている[2]。

一方、E ラーニングを活用した学習が近年行われるようになった。これは、時間や場所に縛られない学習を可能にする一方で、学習意欲の持続が難しいこと、即時の質疑応答が困難なこと、教師が学習者の進捗を把握しづらいといった課題がある[2]。

また、学習者の主体性やモチベーションを重視するアプローチも提案されている。アドバイザーの助言を受けつつ、学習者が主体的に学習ツールを選択することの重要性が強調されている[3]。しかし、この方法には、アドバイザーを確保するハードルや、

学習ツールが静的データに基づいているため柔軟性に限界があるといった課題がある。

このような背景のもと、我々は先行研究として、LLM を活用した日本語学習者支援ゲームを提案してきた[4]。先行研究では、会話エージェント、評価エージェント、要約エージェント、動的プロンプト生成エージェントの4種類のエージェントが連携するアーキテクチャを設計し、学習者の会話履歴に基づいてパーソナライズされた会話練習を提供できることを、疑似学習者実験を通じて示した。LLM を活用した教育やゲームに関する事例はいくつか存在するが、本研究のように日本語学習に特化し、学習者の会話スキル向上を目的とした適応的なゲームの開発は他に例がなく、本研究は新規性を持つものであった。

一方で、先行研究は主にゲームの概念的な提案と基本的な動作確認に焦点を当てたものであり、実用的なゲームとしての完成度や、評価の点では十分とは言えなかった。特に、一般の利用者が実際に使用可能な形の構築は今後の課題として残されていた。

そこで本研究では、先行研究で提案したアーキテクチャを基盤として、ゲームの実用化および評価の系統化を目的とした。具体的には、オープンソースの Streamlit を用いて Web アプリケーションとして再構築することで、誰でも利用可能な日本語学習者支援ゲームを実装した。また、疑似学習者を用いた評価実験を拡張し、提案システムの有効性を系統的に検証した。

2 ゲームの概要

本ゲームは、日本での生活を想定したシミュレーション型の学習ゲームであり、実際の日本語使用場面を舞台とした会話練習を提供する。

本ゲームは、以下の4種類のエージェントによって構成される。

1. 会話エージェント

ゲーム内の登場人物をシミュレートし、学習者と会話を行う。

2. 評価エージェント

会話終了後に学習者の発話を分析し、誤りの指摘や改善案を提示する。

3. 要約エージェント

会話履歴から学習者の発話傾向や誤用の特徴を抽出し、行動履歴として蓄積する。

4. 動的プロンプト生成エージェント

蓄積された情報を基に、次回以降の会話内容を学習者の特性に応じて調整する。

図1に、本ゲームの進行過程を示す。学習者はシチュエーションを選択し、会話エージェントとの会話を通じて課題に取り組む。会話終了後には評価エージェントによるフィードバックが提示され、会話履歴は要約および次回以降のプロンプト生成に利用される。この循環構造により、学習者個人に最適化した会話練習が継続的に提供される。

以上が、先行研究で提案した日本語学習者支援ゲームの基本構成である。本研究では、この構成を基盤として、実際の利用を可能とするため、Webアプリケーションとして再実装を行い、その評価を行った。

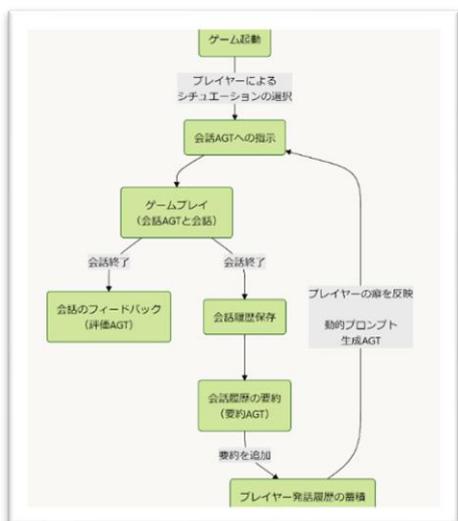


図1 ゲームのプロセスフロー

3 Webアプリケーションの実装

Webアプリケーションの実装に Streamlit を用いた。Streamlit は Python のみで Web アプリケーションを構築でき、Web フロントエンドに関する専門的知識を必要としない。実装したアプリケーションは、Web ブラウザ上で動作し、利用者は特別な環境構築を行うことなく、日本語学習者支援ゲームを利用できる。

3.1 本アプリケーションの操作フロー

基本的な操作フローを以下に示す。

1. ログインとシチュエーション選択

アプリケーション起動後にログインまたは新規登録を行う。ホーム画面のサイドバーから学習したい会話シチュエーションを選択すると、チャット画面へ遷移する。この際、画面上部には当該シチュエーションにおける課題が提示され、学習者はその達成を目標として会話エージェントとの対話を開始する。

2. 対話の実行とサポート機能

画面下部の入力ボックスに発話内容を入力し、送信することで会話を進める。対話中、必要に応じて「ヒント」機能を利用し、「単語の意味の確認」や「次の行動に関する助言」を得ることが可能である。

3. 会話終了

シチュエーションの課題を達成する、もしくは課題の達成が困難であると会話エージェントに判断された場合は、評価画面へ自動で遷移する。評価画面の下部には、同一シチュエーションを再度プレイするためのボタンが用意されており、学習者は評価結果を踏まえて再挑戦することが可能である。

4. 履歴の閲覧

サイドバーのメニューを介して、過去の「会話履歴」や「フィードバック結果」をいつでも閲覧できる。

3.2 ゲーム終了後のフィードバック

会話終了後、本ゲームでは学習者の発話履歴に基づき、評価エージェントによるフィードバックを提示する。

Web アプリケーション化に伴い、各発話に対する in-line 形式の添削を新たに導入した。in-line 添削では、誤用箇所を指摘および改善案を該当する発話と対応付けて表示することで、学習者が自身の発話を文単位で振り返ることを可能としている。

図2に、in-line 形式による添削結果の表示例を示す。図に示すように、誤用箇所と修正例が発話と併せて提示されるため、学習者は修正内容を文脈と対応付けながら確認することができる。

なお、会話全体に対する評価として、「文法と語彙の正確さ」、「表現の自然さと適切さ」、「会話

の論理性と流暢さ」の3つの観点からの評価は、先行研究と同様に評価エージェントによって出力される。これらの評価は、in-line 添削と併せて提示され、学習者が自身の会話を総合的に振り返るための情報として利用される。

フィードバック内容



表 1 評価エージェントの出力分類

	同じ指摘 適切な訂正	同じ指摘 不適切な訂正	同じ指摘 より適切な訂正	新規指摘 適切な訂正	新規指摘 指摘内容が誤り	新規指摘 指摘は不要	指摘 漏れ
初級中 A	28	1	0	10	2	9	3
初級中 B	44	9	0	12	3	2	10
初級上 A	30	2	2	8	0	2	14
初級上 B	44	2	0	10	2	6	8
初級上 C	26	1	0	6	3	0	7

無効化した場合と有効にした場合の2つの条件で実験を行った。

評価には、性格および日本語能力の異なる10種類のペルソナを持つ疑似学習者エージェントを用い、それぞれを仮想プレイヤーとして使用した。各仮想プレイヤーに対し、シチュエーション 1:空港での手続き、シチュエーション 2:スーパーでの買い物、シチュエーション 3:友人との会話の3つのシチュエーションを対象とし、各シチュエーションを5回ずつ反復してプレイさせた。同一シチュエーション内では、各プレイ終了後に評価エージェントが出力するフィードバックおよび評価スコアを学習済み情報として次回以降の発話生成に反映させた。

一方、シチュエーションが切り替わる際、それまでの学習内容を引き継がない設定とした。これにより、各シチュエーション内における反復プレイによる学習効果を独立に評価できるようにしている。

評価エージェントは会話終了後、「文法と語彙の正確さ」「表現の自然さと適切さ」「会話の論理性と流暢さ」の3つの観点からスコアを出力する。

図3に、パーソナライズ機能を有効にした場合と無効にした場合それぞれについて、反復プレイ回数に対する平均スコアの推移を示す。横軸は反復のプレイ回数で、縦軸は10ペルソナの、3シチュエーションでのプレイの平均スコアである。

図3より、パーソナライズ機能の有無を問わず、反復プレイ回数の増加に伴い評価スコアが段階的に上昇する傾向が確認された。これは、評価エージェントによるフィードバックを受けながら反復プレイを行うことで、疑似学習者の発話内容が改善されていることを示している。以上より、本システムは反復的な対話とフィードバックを通じて一定の学習効果を提供できることが示唆された。

また、パーソナライズ機能を有効化した場合の方が、無効化した場合に比べて一貫して高い値を示している。この結果は、評価エージェントによるフィードバックに加えて、学習者の特性を考慮した対話生成が行われることで学習の効率が上がった可能性を示唆している。

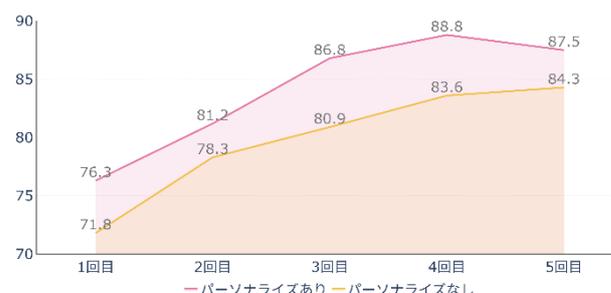


図3 反復プレイ回数に対する平均スコアの推移

5 おわりに

本研究では、LLMを用いた日本語学習者支援ゲームについて、先行研究で提案したシステムを基盤としてWebアプリケーション化および系統的な評価を行った。システムを日本語学習者が容易に利用可能なWebアプリケーションとして実装することにより、会話に対するin-line形式のフィードバックを実現した。評価では、評価エージェントの正確性、疑似学習者による学習効果、およびパーソナライズ機能の効果の三つの観点から検証を行い、提案システムが日本語学習支援において一定の有効性を有することが示唆された。

謝辞

本研究を進めるにあたり、京都外国語大学の麻先生より、研究内容に関する貴重な助言ならびにご協力を賜りました。ここに記して深く感謝申し上げます。

参考文献

1. 文化庁. 日本語教育の展開に関する調査研究事業（報告書）
https://www.bunka.go.jp/tokei_hakusho_shuppan/tokeic_hosa/nihongokyoiku_suishin/nihongokyoiku_tenkai/hokokusho/1/, 2024-12 閲覧.
2. 迫田久美子. 外国語としての日本語教育の現状と課題. 日本外国語教育推進機構会誌, 2020.
3. 国際交流基金. 学習者オートノミー、自己主導型学習、日本語ポートフォリオ、アドバイジング、セルフ・アクセス
<https://www.jpf.go.jp/j/project/japanese/teach/tsushin/research/201003.html>, 2024-12 閲覧.
4. 亀田隆雅, 馬青. LLM を用いた日本語学習者支援. 言語処理年次大会, 2025
5. 国立国語研究所. 日本語学習者会話データベース.
<https://mmsrv.ninjal.ac.jp/kaiwa/index.html>, 2025-12 閲覧.