

# 講義資料と QA ログを活用したプログラミング学習支援のための RAG ベースチャットシステムの試作と評価

## Prototype and Evaluation of a RAG-Based Chat System for Programming Learning Support Using Lecture Materials and QA Logs

越智 洋司<sup>\*1</sup>, 濱田結衣<sup>\*2</sup>

Youji Ochi, Yui Hamada

<sup>\*1</sup>近畿大学情報学部／近畿大学情報学研究所

Faculty of Informatics/ Cyber Informatics Research Institute, Kindai University

<sup>\*2</sup>近畿大学理工学部

Faculty of Science and Engineering, Kindai University

ochi@info.kindai.ac.jp

**あらまし**：本研究では、Retrieval-Augmented Generation (RAG) を活用し、プログラミング学習における初学者向けにヒントを提供する講義連動型チャットシステムを試作した。本システムでは、学生と TA の過去のやりとりによるチャットログ、および PDF 形式の課題資料と異なる性質を持つ 2 種類の情報源を活用した。生成された応答と過去ログの比較実験の結果、本システムによる応答も多くの場合において正確であったが、人間による応答の方が概して好まれる傾向はあるなど課題が確認された。

**キーワード**：チャットシステム、RAG、プログラミング学習、チャットログ

### 1. はじめに

近年、生成 AI (例：ChatGPT や GitHub Copilot) を活用した学習支援の導入が進み、教育分野でも大きな変革が起きつつある。特にプログラミング初学者にとって、自然言語で質問し、即座にコード例を得られることは、学習の敷居を下げる有効な手段である。一方で、こうした利便性の裏には、深い理解を伴わない「AI 依存的な学習」が常態化する懸念も存在する。初学者が課題に直面した際、本来必要とされる問題の構造化や試行錯誤の過程を経ず、即座に AI に回答を求めるような利用は、自己解決力や論理的思考力の習得を阻害する恐れがある。また、生成 AI が必ずしも正確な情報を返すとは限らず、誤情報に基づく学習の危険性もある。しかし、教員やティーチング・アシスタント (TA) がすべての質問に対応するには限界があり、学習支援の質を担保しつつ、人的負担を軽減する新たな仕組みの必要性が高まっている。本研究では生成 AI を単なる「答えを提供するツール」としてではなく、過去の学習記録や教材を活用した情報提示を通じて、学習者の自己解決を促す支援システムの構築を目指す。

### 2. 関連研究と本研究の位置付け

本研究は、Retrieval-Augmented Generation (RAG) を活用し、過去の質問応答ログや課題資料といった学習コンテキストを検索・抽出し、それをもとに動的に回答を生成するチャットシステムを設計した。

Swacha ら<sup>(1)</sup> は、教育分野における RAG ベースのチャットボットに関する 47 件の研究を包括的に調査し、応用領域、対象学習者、基盤となる大規模言語モデル (LLM)、評価手法の観点から分類した。その結果、RAG アーキテクチャが多様な教育ニーズに

柔軟に適應できることが明らかとなった。

しかしながら、既存の多くの実装は教科書や講義ノートといった単一ソースのデータに依存しており、チャットログや PDF 課題資料といった各種メディアの統合については限定的である。Khan ら<sup>(2)</sup> は、PDF 文書を主要なデータソースとして用いた RAG システムの開発に関する経験報告を行い、構造化文書の処理における課題とその解決策を強調しているが、学習者と指導者間のやりとりの機微を捉える上で重要な会話データの統合には触れていない。

本研究の提案するシステムでは、Slack 上に蓄積された過去の学生の質問と教員・TA とのやり取りを活用し、質問とそれに対応する適切なヒントの生成に参考となるように活用する。また、課題資料には講義名や講義回などのメタデータを付与することで検索精度を高める方針とした。

### 3. 研究手法

本稿では、Slack 上の複数情報源データを統合し、共通のベクトルデータベースを用いた統一検索システムを提案する。本節では、システム全体のアーキテクチャと実装について述べる。

#### 3.1 データ構造と前処理

本システムでは、以下の主な 3 種類のデータソースを使用する。

(1) 学生と教員 (TA) との Slack でのやり取りを匿名化した質問応答チャットログ。各ログには「学生」または「TA」の役割が注記されている。

(2) PDF 形式の課題資料。講義中に学生に提示している資料である。

(3) 講義で使用されたサンプルコード。プログラミング言語は Java である。

これらの資料は、講義名や講義番号に基づいて分類され、検索効率を高めるため、講義タイトルやタイムスタンプといったメタデータが付与した。

### 3.2 埋め込みとベクトルストア

テキストの埋め込みには Amazon Titan-Embed (AWS Bedrock 経由) を使用し、Chroma を用いてベクトルストアを構築した。チャットログと課題資料はそれぞれ別々のベクトルストアとして管理する。ユーザからのクエリを受け取ると、システムは意味的類似度に基づいて、各ストアから最も関連性の高い上位  $k$  件 (通常は  $k=5$ ) の文書を取得する。

### 3.3 応答生成プロンプト

LangChain を用いて、取得した文書から応答プロンプトを構築し、自然言語による応答を動的に生成する。これにより、単純な文書検索ではなく、文脈に応じた一貫性のある回答が可能となる。生成される応答は、実際の過去のチャットコミュニケーションの語調や文体に合わせて調整されており、ユーザにとっての親しみやすさや心理的なハードルを下げることを目指した。図1に応答生成のプロンプト例を示す。

```
以下は公式な PDF 資料と過去のチャットログから得られた情報です。PDF 資料をもとに質問に答えてください。ただし、回答の形式や語調は過去のチャットログを参考にしてください。
回答は自然な文章で簡潔に説明してください。
【PDF 資料の情報】 :
{pdf_context}
【過去のチャットログの情報】 :
{chat_context}
【質問】 :
{question}
【回答】 :
""",
input_variables=["pdf_context", "chat_context", "question"]
```

図1 応答生成プロンプト例

## 4. 評価実験

### 4.1 実験方法

本実験の目的は、提案するシステムが実際の学生からの質問に対して、有用かつ適切なヒントを提供できるかを検証することである。特に、システムによる回答の関連性、明確さ、教育的有用性について、人手による回答 (教員や TA によるもの) と比較しながら評価を行うことを目指した。実験では、過去の Slack ログから代表的な 10 問の質問を選定し、それぞれに対して以下の 2 種類の回答を用意した。回答 X は TA によって実際に記述されたオリジナルの回答であり、回答 Y は本システムによって生成され

た回答である。これらについて、情報系を専攻する 11 名の参加者に対して実験を実施し、各設問ごとに「どちらの回答 (X または Y) がより適切か」を選択してもらった。

### 4.2 結果

全体として、回答 X (人手による回答) に対しては肯定的な評価が多く寄せられた。「より適切」と評価した参加者は約 44%、さらに「どちらかという X」と答えた参加者を含めると、全体の約 70% が人手による回答を支持する結果となった。一方で、回答 Y (システムによる回答) については、「より適切」が 15%、「どちらかという Y」が 16% であり、肯定的な評価の合計は約 30%にとどまった。

回答 X (人手) は、文脈に即した具体的な説明が多く、「適切」と評価された割合が 69%に達した。一方で、一部の回答には抽象的すぎる、あるいは情報が不足していると判断されたケースも見られた。

回答 Y (システム) は、誤答率は人手の回答と同程度で平均 9%にとどまったものの、「説明が過剰である (教えすぎ)」とされた割合が 55%にのぼった。

### 4.3 考察

この結果は、現時点においては人手による回答の方が学習支援として効果的であると認識されていることを示唆している。ただし、すべての設問において人手の回答が優れているわけではなく、特定の質問においてはシステムによる回答が高い評価を得る場面も見られた。

また、特定の設問では、人手による回答が抽象的すぎると評価された一方で、システムの回答が明確かつ有用であると判断された事例も確認された。

## 5. 今後の課題

現在のプロンプト設計では、生成される回答が冗長になる傾向があり、学習者の「自分で考える余地」が損なわれる懸念がある。今後は、学習者にとって適切な粒度のヒントを提供するため、プロンプト設計と応答生成の制御方法の最適化が重要な課題となる。また、本システムは現在、問い合わせ内容のみをトリガーとしているが、学習文脈を補完する枠組みの追加が求められる。これにより、より個別化された学習支援が可能となり、システムの教育的価値をさらに高めることが期待される。

### 謝辞

本研究は JSPS 科研費 25K15376 の助成を受けた。

### 参考文献

- (1) Swacha, J., & Gracel, M. (2025). Retrieval-Augmented Generation(RAG) Chatbots for Education: A Survey of Applications. *Applied Sciences*, 15(8), 4234
- (2) Khan, A., Saeed, A., Ahmad, W., & Asgher, U. (2024). PDF RAG: An experience report of building retrieval-augmented generation systems using PDFs. *arXiv*. <https://arxiv.org/abs/2410.15944>