

大規模言語モデルによる授業生成システムの開発

臼井 悠人^{*1} 呉 陽登^{*1} 守屋 燎吾^{*1} 中島 陸貴^{*1} 日室 佑輔^{*1} 島田 憲人^{*1}

指導教員：和田 貢^{*2}

Email: mwada@g.nagano-c.ed.jp

*1: 長野県諏訪清陵高等学校 2年

*2: 長野県諏訪清陵高等学校

◎Key Words

LLM, GPT

1. はじめに

近年、学校教育における人材不足は深刻化しており、教員は多忙な業務や指導要領の変更に対応するために大きな負担を抱えている。特に、学習指導要領改定により追加された新科目である情報 I などに対応できる指導教員の不足は目下の課題である。また、現在 ChatGPT, Stable Diffusion などの Generative AI の進歩が顕著である。そのため、私達はこの学校教育の問題を Generative AI によって緩和できるのではないかと考え、学校教員の負担を軽減し、効率的な授業資料の作成を支援することを目的としてシステムを開発した。

2. 使用するソフトウェア・ライブラリなど

2.1 Markdown

Markdown とは、文書を記述するための軽量マークアップ言語である。プレーンテキスト形式で入力でき、ChatGPT のレスポンスとして適しているため、この言語を採用した。

2.2 ChatGPT

OpenAI が 2022 年 11 月に公開した GPT モデルを用いた人工知能チャットボットである。今回のシステムでは、この ChatGPT の API を利用することで資料を作成した。今回使うモデルは GPT-3.5 である。

2.3 Embedding

Embedding とは、文章をベクトルに変換する事である。ベクトルにすることによって、文字列同士の関連性を測定できる。またこれにより、単語などの検索や分類ができるようになる。今回は Embedding のモデルとして OpenAI によって開発された text-embedding-ada-002 を使用する。

2.4 Chroma

Chroma は、ベクトルデータベースの一つである。ベクトルデータベースとは、ベクトル化した文章を保存し、他の LLM への情報提供や、検索ツールとしての使用ができるものである。

2.5 LangChain

ChatGPT の機能拡張ライブラリ。このような機能拡張ライブラリとして、Semantic Kernel, guidance などがあるが、最も有名なものが LangChain であり機能も豊富なため LangChain を選択した。具体的には、Chroma

などのベクトルデータベースや、各 LLM (大規模言語モデル, Large Language Model) の API をラッピングして使うことができる。また、情報量も他のライブラリに比べて多く、開発期間が短かったことも考慮して、このライブラリを使用した。

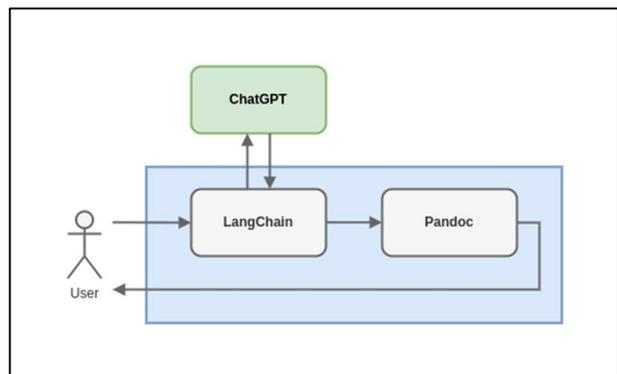
2.6 Pandoc

Pandoc は、Haskell 製のドキュメント変換ツールである。Pandoc により、Markdown, reStructuredText, HTML, LaTeX, PDF などの多様なフォーマットを相互に変換できる。

3. システムの設計

本システムの設計は以下の図 1 のようになっている。

図 1 システムの設計



1. ユーザーからのリクエストをシステムが受け取る。
2. システムがユーザーから受け取ったリクエストを元にして、プロンプトを生成する。
3. 生成したプロンプト LangChain を通して ChatGPT API に送信する。この際、資料をベクトル化して読み込み、それをもとにした Markdown を生成する。今回は、資料として学習指導要領を読み込ませた。
4. この Markdown を PDF などの任意の形式に変換する。

4. 実装と詳細

4.1 ユーザーから LangChain へのリクエスト

質問の形でどのような資料を作ってほしいかをリクエストする。このリクエストは、REST API に準拠している。

4.2 生成に使うプロンプト

生成に使うプロンプトとして、次のようなプロンプ

トを使用した。

『化学の演習問題（授業資料）を作成してください。また、回答は Markdown 記法でしてください。PDF を必ず使う必要はありません。』

まず初めに何を作成して欲しいのか要求したうえで、その後、出力などに対する条件を加えた文章にした。1 つ目の条件は出力形式である。今回の場合は Markdown を指定した。二つ目はベクトル化した学習指導要領に、生成した資料の内容が偏ることを防ぐために加えた。

4.3 プロンプトから Markdown を生成

プロンプトを LangChain を通して ChatGPT API に送信し、Markdown を生成する。この際、ベクトル化した学習指導要領が使用される。

5. 結果

5.1 演習問題

生成した演習問題は、以下の通りである。

図 2 生成した演習問題

化学の演習問題

- 次の化学式をバランスさせてください。
 $Mg + HCl \rightarrow MgCl_2 + H_2$
回答：
 $Mg + 2HCl \rightarrow MgCl_2 + H_2$
- 次の化学反応式において、酸化剤と還元剤をそれぞれ答えてください。
 $2K + Cl_2 \rightarrow 2KCl$
回答：
酸化剤：Cl₂
還元剤：K
- 次の化合物の組成式を答えてください。
亜硝酸カリウム
回答：
KNO₂
- 次の化学式において、水素イオン濃度 (pH) を求めてください。
 $HNO_3 + KOH \rightarrow KNO_3 + H_2O$
HNO₃ の濃度：0.1 mol/L
KOH の濃度：0.05 mol/L
回答：
HNO₃ と KOH は 1:1 の反応をするため、反応後の HNO₃ の濃度は 0.05 mol/L となる。
HNO₃ の水溶液は強酸であるため、pH は $-\log(0.05) \approx 1.3$ となる。

1

問題と解答が出力できているが、いくつか問題点が見られた。

一つ目は問題文に関してである。化学式や化学反応式など、名称が統一されていない箇所がある。また、一つ目の問題は文章が間違っており、最後の問題では、何の pH を求めれば良いのかが分からない。

二つ目は表現に関してである。文頭の見出しの付け方が誤っている。

5.2 授業資料

生成した授業資料は、以下の通りである。

図 3 生成した授業資料

化学の授業資料

化学とは

化学とは、物質の性質や構造、反応などを研究する科学分野です。化学は、私たちの身の回りにある物質や、自然界で起こる現象を理解するために欠かせない学問です。

化学の役割

化学は、私たちの生活に欠かせない役割を果たしています。以下に、化学が果たす役割の一例を挙げてみます。

- 医薬品の開発や製造
- 食品の生産や加工
- 燃料の開発や利用
- 環境問題の解決

化学の学習方法

化学を学ぶには、以下のような方法があります。

- 教科書や参考書を読む
- 実験を行う
- 課題やレポートを作成する
- 話し合いやディスカッションをする

化学の応用例

化学は、様々な分野で応用されています。以下に、化学の応用例をいくつか挙げます。

医療分野

- 薬物の開発や製造
- 検査や診断に用いる試薬の開発

1

食品分野

- 食品の生産や加工における品質管理
- 食品添加物の研究や開発

環境分野

- 大気汚染や水質汚染の原因究明
- 環境汚染物質の分解や除去技術の開発

化学の授業資料

目的

- 化学が果たす役割を理解する
- 物質とその変化に対する興味・関心を高める

内容

- 化学とは
 - 物質の性質や構造、反応などを研究する科学分野
 - 私たちの身の回りにある物質や、自然界で起こる現象を理解するために欠かせない学問
- 化学の役割
 - 医薬品の開発や製造
 - 食品の生産や加工
 - 燃料の開発や利用
 - 環境問題の解決
- 化学の応用例
 - 医療分野
 - 薬物の開発や製造
 - 検査や診断に用いる試薬の開発
 - 食品分野
 - 食品の生産や加工における品質管理
 - 食品添加物の研究や開発
 - 環境分野

2

- 大気汚染や水質汚染の原因究明
 - 環境汚染物質の分解や除去技術の開発
4. 化学の学習方法
 - 教科書や参考書を読む
 - 実験を行う
 - 課題やレポートを作成する
 - 話し合いやディスカッションをする
 5. 化学の授業で行う実験
 - 物質の変化を観察する実験
 - 化学反応の速度を測定する実験
 - 化学反応の熱変化を測定する実験
 6. 化学の授業で行う課題
 - 化学反応の式のバランスを取る課題
 - 化学反応の速度式を書く課題
 - 化学反応の熱変化を計算する課題

参考文献

- 化学の基礎（岩波書店）
- 化学の実験（化学同人）
- 化学の世界（化学同人）

3

こちらでも、改善点が見られた。

まず内容について、「化学の授業資料」という点が2か所あるなど、内容の区切り方が不適切である。また、文章が重複している箇所もあり、「化学の学習方法」の項目で起こっている。

次に、表記について。前半と後半で項目の分け方が異なっており統一されていない。

6. 考察

6.1 演習問題

ベクトル化したデータが学習指導要領のみだったため、問題文が不正確になる、Markdown が適切に使用されないなどの結果となってしまった。したがって、実際の問題集や教科書などをベクトル化し、読み込ませることで問題や解説の内容を向上させる事ができると考えられる。ただし、これらのデータを使用する場合、著作権に関連して起こる問題や、使用するデータが増える事による計算時間の増加について考慮しなければならない。

また、Markdown の記法についての資料なども読み込ませることで、生成された Markdown 記法の正確さも向上すると考えられる。

6.2 授業資料

内容に関する改善点が多く見られたため、実際に教員が作成した資料などを読み込ませることで、品質が向上すると考えられる。また、具体的な単元などをプロンプトで指定することで、資料の内容をより具体的にすることができると思われる。

7. 今後の展望

7.1 GPT-4の使用

今回のシステムでは、GPT-4 の Limited Beta の申請が通らなかったため GPT-3.5 を使用した。そのため、この申請が通った際には、GPT-4 を使用して性能を向上させたいと考えている。また、GPT-4 はマルチモーダルなので、テキスト以外を使用した資料も生成できるようにしたいと考えている。

7.2 授業ライブ配信システム

今回開発したシステムで生成した資料を応用し、AI による授業動画を生成するシステムを開発したいと考えている。

8. おわりに

8.1 プログラムの応用範囲

今回開発した大規模言語モデルを活用し開発した授業資料生成システムには、教育現場におけるさまざまな応用範囲と展望がある。

このシステムは教員が必要とする教材や資料を自動的に生成するため、教員の時間と労力を大幅に節約できる。例えば、演習問題や解説資料など、さまざまな形式の資料の高品質かつ迅速な作成が可能である。

また、この授業資料生成システムは、学生の学習支援として、学生向けの教材や学習資源を生成し、個別の学習ニーズや進度に合わせた教材を提供することができる。加えて、質問応答システムや解説支援ツールとしても利用可能である。

9. 参考文献

- (1) Embeddings - Open AI, URL: <https://platform.openai.com/docs/guides/embeddings/what-are-embeddings>, (06/29/2023)
- (2) Getting Started | Chroma, URL: <https://docs.trychroma.com/getting-started>, (06/29/2023)
- (3) 平成 29・30・31 年改訂学習指導要領（本文、解説）：文部科学省, URL: https://www.mext.go.jp/a_menu/shotou/new-cs/1384661.htm, (06/29/2023)