

生成 AI を用いた数学教材の類似度判定

Determining The Similarity of Mathematics Teaching Materials
using Generative AI荒関 虹希^{*1}, 上野 春毅^{*2}, 小松川 浩^{*1}Koki ARASEKI^{*1}, Haruki UENO^{*2}, Hiroshi KOMATSUGAWA^{*1}^{*1} 公立千歳科学技術大学大学院 理工学研究科^{*1} Graduate School of Science and Technology, Chitose Institute of Science and Technology^{*2} 公立千歳科学技術大学 理工学部^{*2} Faculty of Science and Technology, Chitose Institute of Science and Technology

Email: m2230020@photon.chitose.ac.jp

あらまし：本研究では、eラーニング教材に含まれる文脈・画像情報や知識といった様々な観点から問題の類似度を判定するために、分散表現を用いた手法と生成 AI を用いた手法を試みた。各手法を用いて、実運用されている問題の類似度を算出し、実際の問題文とプロンプトに対する出力、ヒートマップから読み取れる類似度の分布などを比較しながら求められた類似度の特徴を分析した。

キーワード：適応的支援, 類似度, 生成 AI, LLM

1. はじめに

近年, Giga スクール構想の推進により初等中等教育における eラーニングシステムの活用が期待されている。例として, 文科省が提供する CBT システム (MEXCBT⁽¹⁾) や商用の AI ドリルなどが挙げられる。しかし, MEXCBT は適応的に問題を出題する CAT 形式にはなっておらず, 個別最適な学習支援の観点で課題が残る。また, 商用の AI ドリルは企業固有の教材となっており, 各教育現場の地域特色や教育事情の観点から課題が残る。そこで本研究チームの先行研究では, 各教育現場レベルの教材提供者が柔軟に新規教材を提供できることを目標にし, CAT における適切な難易度を担保するために, テキストから成る教材の難易度更新アルゴリズムを構築した。構築したアルゴリズムは, 教育ポリシーや解答履歴, 文脈情報などをもとに難易度更新を行い, 公開性が担保された難易度と比較して高精度での難易度更新を実現した⁽²⁾。しかし, 受験者の能力に応じて出題を制御する CAT 形式の特性を考慮すると, 難易度の分類だけでなく各教材の類似度による分類も必要となる。ただし, 教材における類似度は, 文脈だけでなく, 教材に含まれる知識や画像情報, 出題方式など様々な観点からの判断が必要とされる。そのため, 先行研究で扱った教育ポリシーや文脈情報のみから類似度を判定するのは現実的ではない。

一方, 近年, 大規模言語モデル (LLM) を用いた生成 AI の登場により, 教材の文脈情報だけでなく, 教材に含まれる知識情報や解答までのプロセス, 詳細な画像情報といった様々な情報を取得できる可能性が出てきた。

そこで本研究では, 教材が持つテキストや画像情報から, 文脈や出題方式, 知識面の構成要素などを抽出し, 多視点から類似度判定を行う手法を構築する。また, 演習やテストといったユースケースによ

って扱う手法を変化させることで, 適切な出題ロジックの確立を目指す。本稿では, 3 つの手法から算出された類似度の比較を行い, 各手法における類似度の特徴について述べる。

2. 本研究で扱う教材

本研究では, 類似度を判定する教材として, 本大学が運用している eラーニングシステム (CIST-Solomon) が保有する演習問題を扱う。CIST-Solomon では, 演習問題が 7 段階のレベルに分けられており, レベル 1~2 は「知識の定着」, レベル 3~5 は「知識の活用」, レベル 6~7 は「知識の応用的な活用」と, 求められる知識によって 3 段階のループリックに分類された構造になっている。また, 本研究が想定している教材と同じく, テキストだけでなく画像情報を保持しており, 出題方式も様々である。そのため, 多視点からの教材の類似性を算出しやすく, アルゴリズムの評価は CIST-Solomon が持つループリックの特性と照らし合わせて行う。

3. 類似度判定の手法

本研究では, 類似度の算出方法として, 問題文の分散表現を用いた手法と, 生成 AI を用いて類似度を算出させる手法を試した。このとき, 類似度を算出する問題については, 中学数学の一次関数のグラフを対象とした。

3.1 分散表現をもとにした類似度の算出

OpenAI が提供するモデルの中から「Embeddings V3 small」を用いて問題文の分散表現を取得し, cos 類似度を算出した。このとき, 表 1 に示した問題文のみを入力とした場合 (手法 1) と, 問題文を含めたプロンプトを入力とした場合 (手法 2) の 2 つのパターンで類似度算出を行った。表 2 に簡略化したプロンプトを示す。

3.2 生成 AI を活用した類似度の算出

3.1 とは異なり、モデル「GPT-4o」を用いて対話形式で類似度を算出した（手法3）。具体的には、2つの問題文と付随する画像をプロンプト（表3）とともに送信し、返却された Json データを解析することで、類似度を取得した。

表1 入力として扱う問題文の例

	問題文
例1	$y=ax+b$ の a を $y(1)$ と $y(2)$ という。
例2	下のグラフは直線 $y=2x+1$ である。切片 b は $y(1)$ である。（+画像情報）

表2 手法2で入力としたプロンプトの一部

プロンプト文
役割 ### ・数学領域の教師
指示 ### ・知識やプロセス

表3 手法3で入力としたプロンプトの一部

プロンプト文
役割 ### ・数学領域の教師
指示 ### ・知識やプロセス ・ループリック ・0~1の類似度
出力フォーマット #### ・JSON形式"""

4. 検証・評価

レベル1~7が各10問ずつ存在する70問の演習問題の類似度を算出し、実際の問題文を見ながら各手法の特徴を比較した。手法1, 手法3によって求められた70x70個の類似度をヒートマップとして図1, 図2に示す。

手法1では、類似度が0.8以上の問題が約6%であり、全体的に類似性が低いと判定される傾向にあった。しかし、図1からレベルが高くなるにつれて、類似度が高くなるのが読み取れる。これは低レベルの問題が基礎知識を問う問題であるため、出題フォーマットが様々であるのに対し、高レベルの問題は数値が異なるだけで出題フォーマットが比較的共通化していることが影響していると考えられる。

手法2では、全問題の類似度が0.8を超える結果となり、プロンプトとして入力された文章がそのまま分散表現として変換されたと考えられる。そのため、分布の特徴としては、手法1と同じ傾向を示し、プロンプトに記述した問題に含まれる知識といった観点からは、類似度を求める上で影響が小さかったと考えられる。

手法3では、全手法の中で類似度の分布が最も散らばっており、類似度が0.8以上の問題は約50%であった。また、図2から他の手法と同じく高レベルの問題は高い類似度を示していることが読み取れる。

一方で、低レベルの問題でも約50%が類似していると出力された。低レベルにおける類似度の算出過程を確認すると、問題に含まれる知識レベルが大きく影響していることが分かった。また、中レベルでは、含まれるグラフや図が様々であり、これらの読み取りに用いる知識が異なることから低い類似度を示していた。

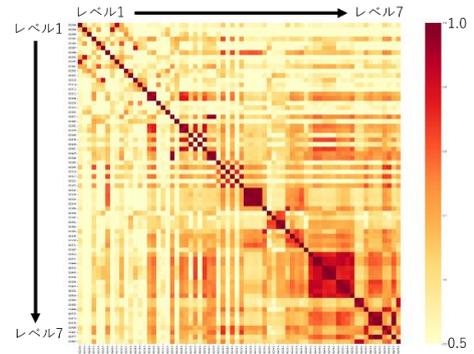


図1 手法1によって求められた類似度 (70問x70問)

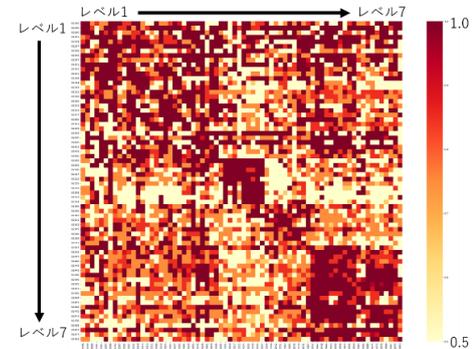


図2 手法3によって求められた類似度 (70問x70問)

5. おわりに

本研究では、教材が持つテキスト・画像データから、教材に含まれる知識などを考慮して類似度判定を行う手法の検証を行った。その結果、「GPT-4o」を利用した手法が最も知識や画像内容を考慮し、類似度を算出できることが確認された。また、プロンプトに与える条件を重視するため、類似度の定義をユースケースによって変化させることも可能であると考えられる。一方で、「GPT-4o」を用いた手法では、今回の検証の場合、約6時間の実行時間がかかり実システムでの運用上課題が残る。

参考文献

- (1) 総合教育政策局教育DX推進室: “文部科学省 CBT システム (MEXCBT: メクビット) について”, https://www.mext.go.jp/a_menu/shotou/zyouhou/mext_00001.html (参照日 2024.6.2)
- (2) Araseki, K., Ueno, H., and Komatsugawa, H.: “A Machine Learning Algorithm for Classification of Difficulty Levels of CATs Considering Education Institution Circumstances”, Proceedings of Chitose International Forum on Science & Technology 2023 (CIF 23), pp.17-21(2023)