

# 教員・学生特性の抽出に基づく 多様な授業シミュレーションの構築

大西朔永<sup>1</sup> 椎名広光<sup>1</sup> 保森智彦<sup>2</sup>

<sup>1</sup>岡山理科大学 情報理工学部 <sup>2</sup>岡山理科大学 教育学部

{s-onishi,shiina,yasumori}@ous.ac.jp

## 概要

教員支援システムの開発において、多様な授業データが必要不可欠であるが、実データの収集は個人情報保護の観点から困難である。本研究では、メタプロンプト最適化を通じて、既存の少数授業記録から、授業環境（教員ペルソナ、学生ペルソナ、授業コンテキスト）を抽出し、それに基づいた場面ごとの一括生成により一貫性のある授業対話をシミュレーション・再現する。さらに、抽出した授業環境を組み替えることで、多様な組み合わせの授業対話を生成し、その再現性と制御性を検証した。

## 1 はじめに

近年、GIGA スクール構想をはじめとする教育 DX の推進に伴い、教育データの利活用による個別最適化された学びや指導の実現が期待されている [1]。教育現場においては、学習者と教員を支援する機械的な方法への需要が高まっている。そして、文部科学省中央教育審議会 [2] は継続的に学ぶ教員を求め、その一環として教員が自身の教育実践を振り返り、改善策を検討する省察活動 [3] が注目されている。しかし、学校現場における教員の多忙化は深刻であり、教員が自身の授業を振り返り、指導技術を向上させるための省察活動に十分な時間を割くことが困難な状況にある [4]。

このような背景から、大規模言語モデル (LLM) を活用し、授業対話音声から自動的に分析や改善アドバイスを生成することで、教員の省察を効率的に支援するシステムの開発を行っている [5]。教員支援の研究としては、「主体的・対話的で深い学び」の観点から、省察活動のデジタル化に着目した研究 [6]、コンピュータシステムを用いた発話分析 [7] が行われている。しかし、学習者支援の研究に比べ、教員を支援する研究は少ない。教員支援システムの開

発、特に LLM を用いたアドバイス生成や授業分析の高度化には、多様な授業データが必要不可欠である。しかし、実際の授業データは児童・生徒のプライバシーに関わる機微な情報であり、収集・共有が困難であるという「データ不足」の問題がある。

この課題に対し、近年、LLM のエージェントとしての能力を活用し、教育環境を仮想的に再現する試みが活発化している。LLM によるデジタルツインモデリングのトレンドを調査し、物理的なシミュレーションを超えた、意味的・文脈的な再現が可能になりつつあることが示されている [8]。そして、LLM エージェントを用いて教室内の多様な役割（教師、生徒）を再現し、教育メカニズムの解明や教授法の検証を行うフレームワーク [9] が提案されている。しかし、これらの研究は一般的な教室環境の再現に主眼を置いており、授業対話全体を再現する点については十分な検討がなされていない。

また、教育分野におけるデータ不足の解決策として、合成データの生成が注目されており、プライバシー保護と公平性を考慮した教育データの合成手法 [10, 11] が提案されている。しかし、これらの手法は主に表形式の学習者データの生成に焦点を当てており、教員の指導改善に資する多様な授業対話データの生成ではない。加えて、授業対話の生成においては、適切なプロンプト設計が不可欠であるが、LLM 自身にプロンプトを生成・評価・改善させる「メタプロンプト最適化」の手法 [12] は、複雑なタスクにおいて人手による設計を上回る性能を示している。

教員支援システムの評価や高度化においては、多様な熟練度（ベテラン・新人）や指導スタイル、また主体的・受動的といった多様な学生を想定した検証が必要不可欠である。しかし、実地調査において多様な属性を網羅するデータを収集することは、協力校の偏りや人的コストの観点から現実的に難しい。そこで、本研究では、既存の少数授業記録から、

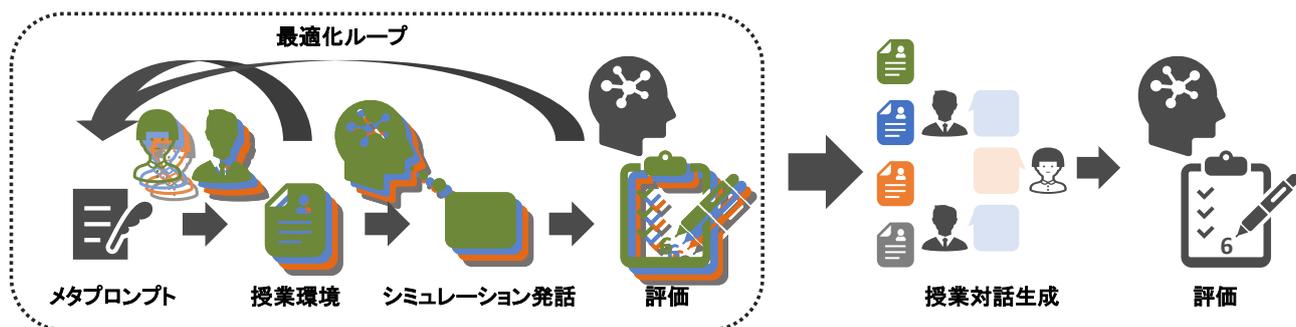


図1 授業シミュレーションの概要

授業環境（教員ペルソナ，学生ペルソナ，授業コンテキスト）を抽出し，それらを組み合わせることで多様な授業対話データを生成する手法を提案する．具体的には，(1) 既存の少数授業記録から授業環境を抽出するためのメタプロンプトを，シミュレーション評価を通じて反復的に最適化する．そして，(2) 抽出した授業環境の要素を組み替えることで，多様かつ一貫性のある授業対話を一括生成する．

## 2 提案手法

本研究では，教員の指導スタイルや学生の特性を反映した授業シミュレーションを構築し，多様な対話を生成するために，以下の2段階のアプローチを提案する．本研究の概要を図1に示す．

### 2.1 授業環境の抽出

まず，既存の授業対話データから，以下の授業環境である3要素をLLMによって抽出する．

- (1) **教員ペルソナ**: 役割，対話スタイル（口調，態度），知識の範囲，フィードバック（肯定・否定，問い返し等），発話の長さやリズム等を記述したプロンプト．
- (2) **学生ペルソナ**: 役割，対話スタイル（口調，態度），知識の範囲，クラス全体の雰囲気，意欲や発話の積極性，教員への反応の傾向，発話の長さやリズム等を記述したプロンプト．発話学生群の集合的なペルソナを表す．
- (3) **授業コンテキスト**: 科目，単元，授業概要，学習目標，授業の場面構成（導入・自力解決・集団解決・まとめ），使用した問題や教材の内容．

特に，これらの授業環境抽出においては，単なる抽出ではなく，抽出用プロンプト（メタプロンプト）の最適化を行っている．本研究では，教員ペルソナ・学生ペルソナと授業コンテキストには関連が

あると考え，これらをメタプロンプトにより一括して抽出し，授業シミュレーションにおいて組み合わせる際の素材として扱う．

メタプロンプトの最適化プロセスは以下の反復的なループによって行われる．

- (1) **抽出**: メタプロンプトを用いて，授業対話から授業環境（教員ペルソナ・学生ペルソナ・授業コンテキスト）を一括抽出する．
- (2) **シミュレーション**: 抽出された授業環境を用いて，実際の授業対話の一部を入力とし，次発話を予測生成する．
- (3) **評価**: 生成された発話と実際の発話を比較し，3.1の2つの評価指標により評価する．G-Evalの評価基準は，自然さ，ペルソナらしさ，一貫性である．
- (4) **最適化**: 評価値と理由をフィードバックとして，メタプロンプト自体をLLMにより改善する．

このループを繰り返すことで，教員の指導スタイルだけでなく，それに対する学生の反応や授業の文脈も含めた，高精度なシミュレーション環境を構築するための再現プロンプトを抽出する．

また，従来の手法[13]と比較して，本研究では抽出された授業コンテキストもシミュレーション，評価，最適化で利用することで，再現精度の向上を図っている．最適化においては，高評価を得たメタプロンプトを最大N件保持し，最適化の際に参照することとしている．

### 2.2 授業シミュレーションの構築

本研究では，抽出された授業環境に基づき，一貫性のある対話生成を行う「場面分割生成」と，授業環境を操作して多様な授業環境を作り出す「環境制御」を統合することで，高度なシミュレーション環境を構築する．

### 2.2.1 場面分割による一括生成

従来のチャットボットのようなターン制の対話シミュレーションでは、45分間という長い授業全体の一貫性を保つことが難しく、また生成コストも増大する傾向がある。そこで、本研究では、抽出した授業コンテキストに基づき、**授業全体（約45分間）の対話ログを一貫性を保ちつつ生成する手法**を提案する。具体的には、授業を問題解決型学習のプロセスに従い、「導入」「自力解決」「集団解決」「まとめ」の4場面に分割し、場面単位で対話を生成した後、それらを時系列順に結合することで1つの授業対話データを構築する。

各場面の生成では、LLMに以下を入力する。

- **授業環境**: 教員ペルソナ、学生ペルソナ、授業コンテキスト
- **対象場面**: 生成対象とする場面の名称と内容（授業環境から抜粋）
- **コンテキスト**: 直前の場面までの授業対話

これにより、前の場面での議論を受けて次の場面が展開するという、授業特有の連鎖的な文脈を保持した生成が可能となる。

### 2.2.2 環境制御による多様化

抽出された授業環境は、授業を構成する3要素（教員ペルソナ、学生ペルソナ、授業コンテキスト）のペアであり、それを組み替えて生成を行うことで、多様な授業データを生成する。組み合わせの変更においては、抽出された授業環境の構成要素は独立した要素として扱う。

本研究では、**授業コンテキストを固定し**、「ベテラン教員」のペルソナと「活発な学生」のペルソナを組み合わせるなど、異なる授業から抽出したペルソナ同士を掛け合わせることで、未知の授業環境を合成する。これにより、同じ授業内容（教材・単元）に対して、異なる指導アプローチや学生のリアクションが生じた場合の対話データを生成できる。

## 3 実験設定

### 3.1 評価手法

本研究では、生成されたデータの品質を多角的に検証するため、LLM評価フレームワークであるDeepEval [14]を基盤とし、以下の2つの評価指標を用いる。これにより、評価プロンプトの最適化およ

び評価バイアスの抑制を図り、評価の客観性を担保する。

- (1)**埋め込みベクトルによる類似度評価**: 埋め込みモデルにはGoogleのtext-embedding-005を用いて、テキストをベクトルに埋め込み、そのコサイン類似度を算出することで、語彙の表層的な一致に依存しない意味的類似性を定量的に評価する。
- (2)**LLM-as-a-judgeによる評価**: LLMに対して2個のテキストを提示し、その類似性や一貫性、自然さなどを多角的な観点から評価させる。DeepEvalに実装されているG-Eval[15]を用いる。これは、指定した評価基準（類似性、一貫性、自然さ等）に基づき、LLMがChain-of-Thought（思考過程）を生成した上でスコアリングを行う手法である。人間による評価と高い相関を持つことが知られており、多角的な観点から生成データの品質を検証する。

また、Round-trip Correctness (RTC) [16]を用いて、コードから自然言語への変換に対し、その自然言語からコードへ逆変換した際に、一致するかを検証する手法が提案されている。本研究では、コードと自然言語の対応関係に対し、授業対話 $D$ と授業環境 $E$ の対応で一致するかを検証することで評価を行う。

### 3.2 評価実験

提案手法の有効性を検証するため、再現性と制御性の観点で評価を行う。

**再現性の評価** 抽出された授業環境 $E_{in}$ を用いて**授業全体の対話を生成し**、それが実在する元の授業をどの程度忠実に再現できているかを評価する。

- **対話レベルの評価**: 生成された対話テキスト $D_{out}$ と元の対話テキスト $D_{in}$ の類似度を上記の2手法で評価する。G-Evalの評価基準は、自然さ、ペルソナらしさ、一貫性である。
- **特徴レベルの評価**: 生成された対話 $D_{out}$ から再度授業環境 $E_{out}$ を抽出し、元の授業環境 $E_{in}$ との類似度を上記の2手法で評価する。G-Evalの評価基準は、ペルソナの一貫度、授業コンテキストの一貫度である。

**制御性の評価** 授業環境の構成要素であるペルソナ（教員・学生）の組み合わせ変更が生成結果に与える影響を確認するため、RTCを実施する。ここでは、授業コンテキストを固定し、教員および学生ペ

ルソナのみを変更した新たな授業環境  $E_{target}$  から対話  $D_{out}$  を生成する。生成された対話  $D_{out}$  から再度授業環境  $E'_{target}$  を抽出し、 $E_{target}$  と  $E'_{target}$  (特に変更したペルソナ部分) の類似度を上記の2手法で評価する。

### 3.3 使用データ

本研究では、2021年度から2024年度にかけて小学校で実施している教員の省察活動支援の中で収集した、複数の教員による算数科目の3授業分の対話データを用いる。これらのデータは、授業音声の文字起こしから構成されており、各発話には話者IDを付与している。

### 3.4 実験条件

LLMには、すべてのプロセス(抽出、最適化、生成、評価)において、Googleのgemini-2.5-flashを使用した。

## 4 実験結果

### 4.1 メタプロンプトの最適化結果

メタプロンプトのバージョンごとの平均評価について述べる。初期プロンプトにおいて、抽出した授業環境における評価は0.53であった。v1においては、0.50、v2が0.51と改善は見られなかった。

授業ごとの評価では、授業1が0.49、授業2が0.51、授業3が0.53と差がある結果になった。

### 4.2 再現性の評価結果

### 4.3 環境制御の評価結果

## 5 おわりに

本研究では、教員支援に向けたデータ不足の課題を解決するため、教員・学生特性の抽出に基づく多様な授業シミュレーションの構築手法を提案した。教員ペルソナ・学生ペルソナ・授業コンテキストからなる授業環境を抽出し、それらの組み合わせを変更することで、個別の授業環境の再現だけでなく、多様な教育場面の生成が可能であることを示した。今後は、生成した授業対話データの妥当性や信頼性のさらなる検証や、教育現場での実証実験に取り組む。マルチモーダル情報(音声など)を含めたより高度なシミュレーションも課題である。

また、本研究の制限として、授業対話データでは学生を区別して扱っておらず、音声をベースとしているため、抽出された学生ペルソナは実際に発話を行った学生群の特徴となっている。それを踏まえて、沈黙している学生を含むクラス全体の環境再現することは今後の課題である。

## 謝辞

本研究は JSPS 科研費 JP23K11378 の助成を受けたものです。

## 参考文献

- [1] 文部科学省. 教育振興基本計画. [https://www.mext.go.jp/a\\_menu/keikaku/](https://www.mext.go.jp/a_menu/keikaku/), 2023. 参照 2025-02-07.
- [2] 中央教育審議会. 教職生活の全体を通じた教員の資質能力の総合的な向上方策について (答申), 2012.
- [3] 秋田喜代美. 変貌する教育学, 教師教育から教師の学習過程研究への転回—マイクロ教育実践研究への変貌—, pp. 45–75. 世織書房, 2009.
- [4] 国立教育政策研究所. 教員環境の国際比較 OECD 国際教員指導環境調査 (TALIS)2018 調査報告書. ぎょうせい, 2018.
- [5] 大西朔永, 椎名広光, 保森智彦. マルチモーダル llm によるインタラクティブな教員アドバイス生成システム. 第 50 回教育システム情報学会全国大会, 2025.
- [6] 保森智彦. 省察方法のデジタル化に関する一考察: 「主体的・対話的で深い学び」の観点から, pp. 3–11. 学習開発学研究, No. 14. 広島大学大学院人間社会科学研究所学習開発学領域, 2022.
- [7] yuchen Wang, 大井翔, 松村耕平, 野間春生. 新任教員の授業力向上のための授業振り返りシステムに関する研究. 情報処理学会インタラクシオン, pp. 753–757, 2021.
- [8] Linyao Yang, Shi Luo, Xi Cheng, and Lei Yu. Leveraging large language models for enhanced digital twin modeling: Trends, methods, and challenges, 2025.
- [9] Zheyuan Zhang, Daniel Zhang-Li, Jifan Yu, Linlu Gong, Jinchang Zhou, Zhanxin Hao, Jianxiao Jiang, Jie Cao, Huiqin Liu, Zhiyuan Liu, Lei Hou, and Juanzi Li. Simulating classroom education with LLM-empowered agents. In Luis Chiruzzo, Alan Ritter, and Lu Wang, editors, **Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)**, pp. 10364–10379, Albuquerque, New Mexico, April 2025. Association for Computational Linguistics.
- [10] Mohammad Khalil, Farhad Vadiiee, Ronas Shakya, and Qinyi Liu. Creating artificial students that never existed: Leveraging large language models and ctgans for synthetic data generation. In **Proceedings of the 15th International Learning Analytics and Knowledge Conference**, LAK '25, p. 439–450, New York, NY, USA, 2025. Association for Computing Machinery.
- [11] Kadir Kesgin. Fairsyn-edu a diffusion-based model for fair and private educational data synthesis. **Discover Education**, Vol. 4, No. 1, p. 336, sep 2025.
- [12] Hayoung Park, Dugki Min, Tuan Anh Nguyen, and Eunmi Choi. Adaptive prompting in the metaverse: An iterative prompt optimization framework for enhancing llm performance across diverse tasks. In **2025 International Conference on Metaverse Computing, Networking and Applications (MetaCom)**, pp. 13–18, 2025.
- [13] 大西朔永, 椎名広光, 保森智彦. アドバイス生成における教員対話の再現を通じたメタプロンプト最適化. 電子情報通信学会教育工学研究会 (ET), Vol. 125, No. 171, pp. 40–47, 2025.
- [14] Confident AI. Deepeval: The open-source llm evaluation framework. <https://github.com/confident-ai/deepeval>,.
- [15] Yang Liu, Dan Iter, Yichong Xu, Shuohang Wang, Ruochen Xu, and Chenguang Zhu. G-eval: NLG evaluation using gpt-4 with better human alignment. In Houda Bouamor, Juan Pino, and Kalika Bali, editors, **Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing**, pp. 2511–2522, Singapore, December 2023. Association for Computational Linguistics.
- [16] Miltiadis Allamanis, Sheena Panthaplackel, and Pengcheng Yin. Unsupervised evaluation of code llms with round-trip correctness. In **Proceedings of the 41st International Conference on Machine Learning**, ICML'24. JMLR.org, 2024.

表1 授業データ概要

授業 ID	小学校	学年	教員	科目	単元
授業 1	小学校 A	6 年生	教員 A	算数	体積の関係を求める単元
授業 2	小学校 B	6 年生	教員 B	算数	三角形の数を調べる単元
授業 3	小学校 B	5 年生	教員 C	算数	同じ四角形を見つける方法を探す単元

表2 メタプロンプトの最適化における G-Eval

評価基準	評価基準の詳細
自然さ	再現された発話が、元の{role_name_jp}の発話と比較して、人間にとって自然に感じられるか
ペルソナらしさ	再現された発話が、元の{role_name_jp}の発話と比較して、{role_name_jp}の話し方や思考パターンをどの程度反映しているか
一貫性	提供された対話コンテキストと、元の{role_name_jp}の発話、再現された発話全体を通して、{role_name_jp}のキャラクターや対話スタイルに一貫性があるか

表3 対話レベルの評価における G-Eval

評価基準	評価基準の詳細
自然さ	生成された授業対話が、日本の小学校の授業として自然で、違和感がないか。
ペルソナらしさ	教員と学生の話し方、役割、相互作用が、元の授業の特徴（ペルソナ）を反映しているか。
一貫性	授業の流れ（導入からまとめまで）や文脈が一貫しており、矛盾がないか。

表4 特徴レベルの評価における G-Eval

評価基準	評価基準の詳細
ペルソナの一致度	再抽出された教員・生徒のペルソナ（役割、スタイル、知識など）が、元のペルソナとどの程度一致しているか。
授業コンテキストの一致度	再抽出された授業コンテキスト（単元、目標、フェーズ構成）が、元のコンテキストとどの程度一致しているか。