大規模言語モデルを用いた授業対話の学びに関するラベル分類

Classification of Learning Label in Class Dialogues Using LLM

大西 朔永[*1]，椎名 広光[*2]，保森 智彦[*3]
Sakuei ONISHI[*1], Hiromitsu SHIINA[*2], Tomohiko YASUMORI[*3]
[*1] 岡山理科大学大学院　総合情報研究科　数理・環境システム専攻
[*1]Graduate School of Informatics, University of Educational Systems
[*2] 岡山理科大学　情報理工学部
[*3] 岡山理科大学　教育学部
Email: i22ed08bf@ous.jp

あらまし：小学校では授業改善を目的に教員の省察活動が行われているが，省察の負担を軽減する支援が必要である．そこで，本研究では，大規模言語モデルを用いることで，授業中の発話に対して，「主体的・対話的で深い学び」の観点における学びのラベルを推定している．ラベル推定を行うことで，「主体的・対話的で深い学び」に関する発話の頻度などを分析でき，授業の客観的な評価に繋がると考えられる．
キーワード：授業分析，対話分析，対話モデル，大規模言語モデル

1. はじめに

文部科学省の小学校指導要領において，「主体的・対話的で深い学び」を実現する授業が目指されており，教員の省察活動においても取り組まれている．授業の分析研究では，文字起こしした授業発話を分析するシステム(1)が提案されているが，教員と児童の対話としては扱われていない．また，近年では Large Language Model（LLM）が発展し，LLM を用いた ChatGPT などの対話システムが活用されている．汎用的な LLM を小規模な計算資源で特定用途にファインチューニングする手法として，Low Rank Adaptation（LoRA）や QLoRA が提案されている．

本研究では，大規模言語モデル（LLM）を用い，小学校の授業発話に対して，「主体的・対話的で深い学び」のラベルを推定している．最初に，小学校の授業において教員と児童の発話を文字起こしし，「主体的・対話的で深い学び」に関する発話へラベルを付与した対話データを構築している．次に，QLoRA を用いて LLM に小学校の授業を学習させることで，小学校の授業用途に LLM をファインチューニングしている．最後に，ファインチューニングした LLM を用いて，授業対話のラベルを推定している．

2. 対話データとラベル

授業の対話データとして，小学校における算数の授業を録画した上で，教員と児童の発話に対して，文字起こしをしている．そして，「主体的・対話的で深い学び」に関連する発話に対して，ラベルを人手で付与している．ラベルは，「主体的な学び」，「対話的な学び」，「深い学び」，「非主体的な学び」の 4 種類である．

対話データの概要を表 1 に示す．対話は，授業 1 から授業 4 の 4 授業から作成しており，授業 3 と授業 4 は同じ教員と児童の授業である．4 授業を合わせた発話数は 869 で，その内 106 発話に対してラベルが付与されている．ただし，1 発話に複数の学びが関連する場合には，同じ発話に複数のラベルを付与している．

表 1 授業の対話データ

データ	教員	授業概要
授業 1	教員 A	階段の周りの長さを求める
授業 2	教員 B	体積の関係を求める
授業 3	教員 C	余りのある小数の割り算
授業 4	教員 C	正方形の棒の数を求める

3. LLM を用いたラベル推定

3.1 ラベルの推定手法

本研究では，小学校の授業に特化させた LLM を用いて，授業対話のラベルを推定している．推定手法の手順を次に示す．最初に，LLM のモデルを用いて，推定対象対話d_iとラベル例対話d_jをそれぞれベクトルv_iとv_jに埋め込む．次に，推定対象対話のベクトルv_i^{target}とラベル例対話のベクトル$v_j^{supervised}$のコサイン類似度$s_{i,j}$を求める．

$$s_{i,j} = Cos\left(v_i^{target}, v_j^{supervised}\right)$$

そして，閾値σを決定し，閾値以上の類似度$s_{i,j}$となるラベル例対話d_jを抽出した対話集合D_iを作成する．

$$D_i = \left\{d_j \middle| s_{i,j} \geq \sigma\right\}$$

最後に，対話集合D_iの各ラベル例対話d_jに付与されているラベルl_jを推定対象対話d_iのラベルL_iであると推定する．

$$L_i = \left\{l_j \middle| d_j \in D_i\right\}$$

3.2 ラベル推定の実験設定

モデルの比較として，拡張 GVTSC(2)というパラメータ数が LLM に対して約 1/300 のモデルを用いて，LLM の汎用性とパラメータ数による影響を確認している．LLM のベースモデルとして，日本語の対話が可能な Youri 7B Chat(3)を用いており，授業の対話

表 2 授業 3 に対する全授業のラベル例を用いたラベル推定の例

推定対象対話	正解ラベル	類似対話	推定ラベル
児童: いい。あってる。 **教員:** 合ってそう?ほんでな、 **児童:** はい。 **教員:** ○○さんが言ったことをもう一回言える?大丈夫?先生、○○さんが言ったことをちゃんとね、図に書き込んでほしいなと思うんよ。書いた?素晴らしいね。はい。○○さん、**ここからここまでがどれくらいだった?**	深い学び	**児童:** わかる。 **教員:** 問題ないよね?じゃあ全部書こうね、この図を。 **児童:** え? **教員:** ちょっとめんどくさいよね、それを書くのは。**この筆算の段階で何か気づいたことない?**これ、3 じゃいけませんよってもう、ほぼほぼ言ってたけど。	**深い学び**
児童: もとのままで計算すると良い。 **教員:** で、もう、今、**○○君が言ってくれたのでほぼいいよね?**それで。もう 1 回言って、○○君、はい。だから!	主体的な学び	**児童:** えっこれ長くならん？ **教員:** うん。わかりやすく書いてね。いいよ。はいじゃあ途中でもいいです。よしじゃあ。自分が書いた**まとめ**をね**隣の人に伝えてみましょう**どうぞ。	**主体的な学び** 対話的な学び

表 3 ラベル推定の評価結果

ラベル例	Model	Precision	Recall	F1
対象授業以外	拡張 GVTSC	0.021	0.042	0.014
対象授業以外	LLM	0.045	0.254	0.055
全授業	拡張 GVTSC	**0.302**	0.189	**0.158**
全授業	LLM	0.130	**0.327**	0.097

データで QLoRA を用いてファインチューニングしている．拡張 GVTSC は，ベースモデルを使用せず小学校の授業から作成した対話データのみで学習しており，汎用的な知識などは含まれていない．

類似度の閾値σには，0.90，0.95，0.97，0.98，0.99 を用いている．推定に使用するラベル例の対話データとして，全授業と，推定対象の授業以外を使用して比較を行っている．発話単体の 0 に加えて，対話として扱うためにコンテキストを 1 から 5 と変化させた対話を使用してラベルを推定している．
評価指標には，Micro-average の Precision，Recall，F1 を用い，正解ラベルと推定ラベルで評価している．

3.3 ラベル推定の評価

ラベル推定の例と評価結果を表 2 と表 3 に示す．

表 3 ではラベル例とモデルごとの平均評価を示している．ラベル例については，全授業の場合，ラベル例と同じ対話の推定は自明であるため，その例は除外しているが，どちらのモデルも全授業のラベル例を使用した方が高い評価となっている．つまり，推定対象授業における話し方や話題の共通点が影響したと考えられる．

LLM の評価としては，対象授業のラベル例を用いない場合，拡張 GVTSC に対して LLM の評価が勝っているため，汎用性が高いと考えられる．

表 2 は，授業 3 に対する全授業のラベル例を用いたラベル推定の例である．授業 3 は授業 4 とともに，授業別の評価において，同授業のラベル例も用いた推定で特に高い評価となった．要因として，授業 3 と 4 が同じ教員と児童による授業であったことが影響したと考えられる．

1 行目は，F1 が拡張 GVTSC 中で最高の 0.308 であった閾値σを 0.90，コンテキスト数を 3 とした場合の推定例である．この類似発話は，同授業の対象発話から 5 分後の発話であり，コサイン類似度は 0.936 で同じラベルを推定できている．2 行目は，LLM で F1 が最高の 0.229 であった閾値σを 0.98，コンテキスト数を 1 とした場合の推定例である．この類似発話は，授業 2 のまとめに関する発話であり，コサイン類似度は 0.980 と高いが，複数ラベルが付けられており，対象発話のラベルとしては適切でない対話的な学びも推定している．

4. おわりに

本研究では，小学校の授業でファインチューニングした LLM を用いて，授業対話に対するラベルを推定した. LLM は未知の授業に対する汎用性が確認されたため，データ数の少ない教員や授業内容に対するラベル推定や，ラベルを付与する際の補助としての活用が考えられる．

今後の課題としては，閾値などを変えた複数のラベル推定を統合した推定や，ラベル例対話の収集がある．

参考文献

(1) Y. 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ョン，pp. 753-757 (2021)
(2) 大西朔永，椎名広光，保森智彦: “対話モデルを用いた授業の発話分析とシーンの可視化”，教育システム情報学会 2022 年度特集論文研究会, Vol. 37, No. 7, pp. 152–159 (2023)
(3) Zhao, T. and Sawada, K.: “rinna/youri-7b-chat”, https://huggingface.co/rinna/youri-7b-chat (参照 2024.3.5)