

一対一英語個別指導における教師の戦略予測モデル構築

Pureumae Lee^{1,2} 三浦直己^{1,2} Steven Coyne^{1,2} 松林優一郎^{1,2}¹ 東北大学 ² 理化学研究所{lee.pureumae.s6, miura.naoki.p6, coyne.steven.charles.q2}@dc.tohoku.ac.jp
y.m@tohoku.ac.jp

概要

一対一個別指導において、教師は対話履歴に基づいて次の発話で使うべき戦略を決める。この戦略を適切に決めることで学習者のニーズに合わせた個別教育を提供し、学習効果を上げることができる。本研究では、チャット形式の英語個別指導において、対話履歴から次の発話に必要な教師の戦略を予測するモデルを構築する。特に、発話と戦略が相互作用し合う関係に着目し、戦略予測と発話生成を同時に学習させるマルチタスク学習手法を提案する。実験の結果、既存手法より提案手法が戦略の予測性能を向上したことが確認された。

1 はじめに

一対一個別指導は一般的な授業環境より教師が学習者に合わせた学習を提供することができる。そのような個別指導を受けた学習者集団は、通常の授業を受けた学習者集団に比べて学習成果が高いことが示されている [1]。しかし、個別指導には拡張性の問題がある。個別指導は一対一ないし少数の学習者を対象に教育を提供するが、教師の数が学習者より少ないため大規模な展開が難しい。

そのため、近年は大規模言語モデル (LLMs) を AI 教師として構築する研究が進められている。これは LLM の発話を実際の教師発話に近づけるように学習させることで実現している [2] [3]。しかし、現在の LLM は教育的に適切な発話を生成できていないことが報告されている [4] [5]。この課題に対し、教師の戦略 (Strategy) を活用することが有効であると示されている。戦略とは発話の性格であり、その発話が何をしているのかを表すものである (例: 質問、修正、知識伝達など)。教師は現在の対話履歴に基づき、次の発話に必要な戦略を選択することで学習を円滑にし、学習効率を向上させる。このような戦略を LLM に組み込むことで、AI 教師の発話を

表 1 TSCC2 の戦略: 一対一英語個別指導で、学習を促進するために用いている戦略。

教師の戦略	説明
scaffolding (足場かけ)	学習者の理解をサポートする行為
eliciting (引き起こし)	学習者から回答を引き出す行為
repair (修正)	教師が発話を修正する行為
presentation (提示)	ある知識や能力を示す行為
enquiry (質問)	学習者に質問をする行為
clarification (明確化)	不明確な発話をクリアにする行為
reference (参照)	外部資料を参照する行為
recap (復習)	学習内容をサマリーする行為
revision (見直し)	前の学習セッションを見直す行為

人間教師の発話に近づけたり [6]、より教師の発話として好まれる [7] ことが報告されている。

本研究では、一対一チャット形式および自由対話型の英語個別指導において、与えられた対話履歴から次の教師戦略を予測するタスクに着目する。タスク指向型学習では学習目標が明示的に決まっており、それに沿って対話が進む。一方、自由対話型学習では、教師と学習者が日常会話を挟んだり、学習内容が柔軟に変わったりする。そのため、自由対話型では、各発話が学習目的か否かを判断するという、タスク指向型にはない追加的な課題が生じる。

具体的な手法は、現在までの対話履歴を LLM に入力し、その次の教師発話に必要な戦略を予測するタスクを実験する。特に、発話の生成過程が戦略の理解を深め、生成した発話が適切な戦略の予測を支援するという相互補完関係を仮定し、戦略予測と発話生成を同時に学習させるマルチタスク学習手法を提案する。実験結果、提案手法によってモデルの一般的な予測性能が向上された。しかし、戦略分布の不均衡による予測性能の低さが課題として残っている。

本研究の貢献は以下の通りである:

- 英語学習および自由対話型の個別指導における LLM の戦略予測の性能を検証する。

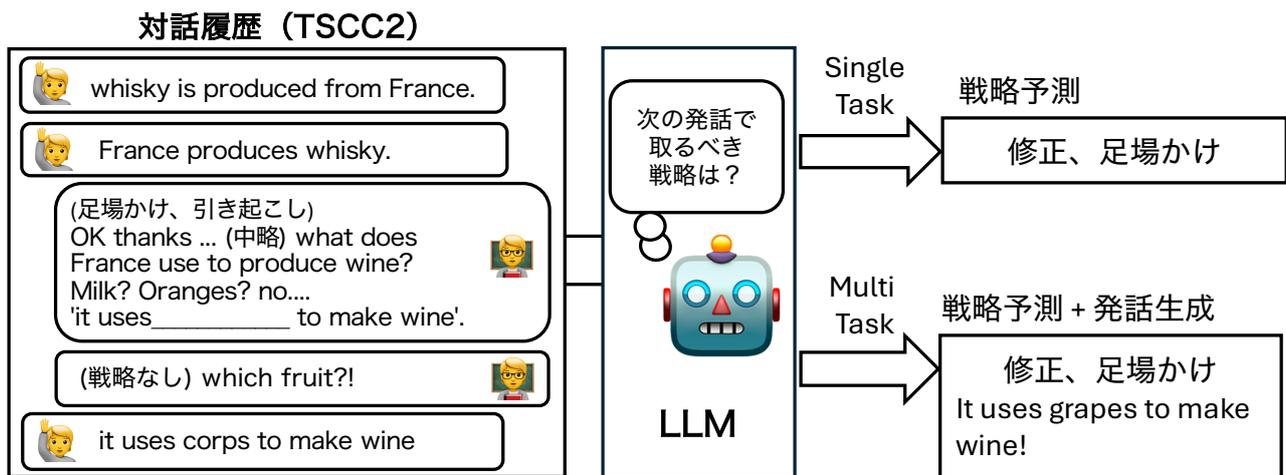


図1 一対一英語個別指導における教師戦略の予測タスク。本研究では戦略予測の性能を引き上げるため、戦略予測と発話生成を同時に学習する。

- 戦略予測の性能を向上させるため、戦略予測とともに発話を同時に生成するマルチタスク手法を提案する。

2 関連研究

教師戦略予測タスクとは、与えられた対話履歴に基づいて、教師が次の発話で使うべき戦略を予測するタスクである。従来は時系列モデル (GRU [8]) やエンコーダーデコーダーモデル (BART [9]) などのモデルを用いて対話履歴を埋め込み、次の発話で必要な教師戦略を予測している [10] [6]。Ikram ら [11] が数学学習ドメインにおいて LLM を用いた戦略予測を行い、従来手法より高い性能を達成している。

しかし、先行研究には二つの限界がある。第一に、LLM を用いた教師戦略予測の研究は数学・タスク指向ドメインに限定されており、英語・自由対話型ドメインへの適用可能性が十分に検証されていない。第二に、先行研究は戦略予測のみを単独タスクとして扱っている。関連研究 [12, 13] ではトピック予測と発話生成の同時学習により発話品質が向上したが、予測タスク自体の精度向上は評価されていない。本研究では、戦略予測と発話生成のマルチタスク学習が戦略予測性能に与える影響を検証する。

3 データセット

本研究では The Teacher-Student Chatroom Corpus version 2 (TSCC2) [14] をデータセットとして用いる。TSCC2 は一対一のチャット形式英語個別指導において、2 人の経験豊富な英語教師と 13 人の学

習者が実際に行った対話 260 セッションを収集したコーパスである。学習者の年齢は 12 歳から 40 歳まで、英語の実力はヨーロッパ言語共通参照枠の CEFR レベルで B1 から C2、国籍は 7 ヶ国にわたる。また、既存の多くの個別指導データセット (例: CIMA [2]、MathDial [15]) がタスク指向型であるのに対し、TSCC2 は特定の課題ではなく、学習者の興味や理解度に応じて教師が柔軟に対話を展開する自由対話型のデータセットになっている。

4 タスク設定

図 1 に本研究のタスク概要を示す。今回用いるデータセットの TSCC2 では、一つの教師発話に複数の戦略が同時に使われることがあるため、本タスクはマルチラベル分類となる。学習セッション S は以下のように定義する：

$$S = \{(u_i, r_i, m_i)\}_{i=1}^N$$

一つのセッション S はそれぞれ N 個の発話単位 (u_i, r_i, m_i) で構成され、それぞれの発話単位は話者 (r_i) と発話内容 (u_i) 、そして発話戦略 (m_i) を含めている。

Single Task Learning (既存手法) 既存手法 [11] のタスク設定は、予測対象となる教師発話の直前の対話履歴をモデルに入力する。既存手法では、対話履歴 C_i から次の教師発話で必要な戦略を予測する：

$$C_i = \{(u_j, r_j, m_j)\}_{j=i-4}^i$$

$$\hat{m}_{i+1} = f_{\theta}(C_i)$$

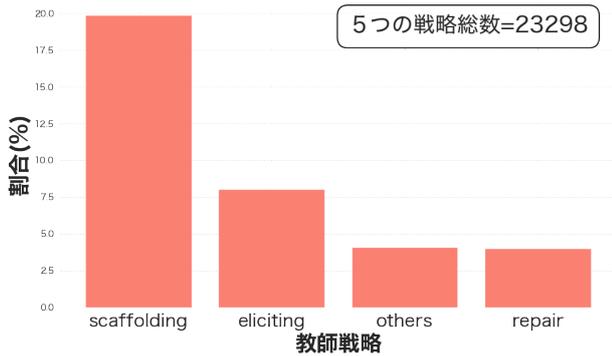


図2 TSSC2の戦略分布

5つの発話単位で構成された対話履歴 C_i を戦略予測モデル f_θ に入力、次の教師発話に必要な戦略 \hat{m}_{i+1} を出力する。

Multi Task Learning (提案手法) ベースラインからさらなる性能向上を目指すため、本研究では戦略予測と発話生成を同時に行うマルチタスク学習手法を提案する：

$$\hat{u}_{i+1}, \hat{m}_{i+1} = f_\theta(C_i)$$

本手法は戦略と発話の相互依存関係を活用する。適切な戦略予測には「どのような発話で実現すべきか」の考慮が必要であり、逆に適切な発話生成には「どのような教育的意図（戦略）を持つか」の理解が必要である。発話生成を補助タスクとして導入することで、モデルは戦略と言語パターンの対応関係を学習でき、戦略予測の精度向上が期待される。

タスクは人間の思考過程（戦略決定後、発話を生成）に従い、戦略を先に予測してから発話を生成する。ただし、今回の発話生成は戦略予測の補助として取り入れたタスクであり、評価対象ではない。

4.1 教師戦略

TSSC2には教師の発話ごとに学習の大きな転換を表すメインシークエンスと小さな転換を表すサブシークエンスがアノテーションされている。その中、9つあるサブシークエンスを教師の発話戦略として扱う。本研究で予測する教師戦略は5種類である：scaffolding、eliciting、repair、others（そのほか）、none（戦略なし）。scaffoldingはVygotskyのZone of Proximal Development (ZPD)理論[16]に基づき、サポートを通じた学習者の現在のレベルと到達可能なレベルの間隔を縮める戦略である。elicitingは学習者が答えにたどり着けるように思考を促進し、学習者の積極的な参加と理解を深める[17]。repairは学

習者の誤りを訂正することによってテスト成績の向上に寄与することが報告されている[18]。othersは三つの戦略を除外した教師戦略をまとめたラベルであり、説明や復習などの戦略を含む（表1参照）。最後に、これら4つの戦略に属しない発話はnoneとして分類する。

5 実験設定

5.1 データ分割

TSSC2には260個の学習セッションがあり、総教師発話数は22,012個である。これをセッション基準で6:2:2の比率で訓練・検証・テストデータで分割し、先行する対話履歴が存在しない各セッションの最初の教師発話は除外している。最終的に各データセットに含まれる教師発話の数は13,256、4,557、4,198個となる。検証データはハイパーパラメータとエポック数の調整に使用している。

5.2 モデル

今回用いるモデルは二つである。BERT[19]：今回LLMの性能を比較するため、Transformer系で分類タスクに高い性能を出すbert-based-uncasedを用いる。モデルの出力に多ラベル分類ヘッドを追加し、入力文脈から戦略ごとのロジットを通じて確率を出力する。Llama-3.1[20]：LLMはオープンソースモデルの中でLlama-3.1-8B-Instructを用いる。学習にはSupervised Fine-Tuning (SFT)のprompt-completion形式を用い、出力として戦略と発話のみを学習するように設定した。

5.3 評価指標

データセット内の戦略分布の不均衡を考慮し（図2参照）、三つの F_1 スコアを評価指標として報告する。Micro F_1 ：マルチラベル分類において、全予測の正確度 (Accuracy) と同値である。Macro F_1 ：各戦略の F_1 スコアの単純平均であり、各クラスを平等に評価する指標である。Weighted F_1 ：各戦略の F_1 スコアを出現頻度で重み付けして平均したもので、全体的な予測性能と少数事例の戦略予測の性能を両方反映する。また、最低限の予測性能を測るため、ランダム (Random) に予測した時の結果と一緒に報告する。これは、テストデータセットにおける各戦略の出現頻度に従ってランダムに予測した時の性能である。なお、noneは四つの戦略のいずれも該

表 2 教師戦略の予測性能 (%)

モデル	Micro F_1	Macro F_1	Weighted F_1
Random	11.43	7.49	11.46
BERT	35.89	21.15	30.76
Single Task	<u>42.01</u>	<u>31.30</u>	<u>38.99</u>
Multi Task	42.20	32.47	39.67

表 3 教師戦略ごとの予測性能 (F_1 , %)

戦略	Single Task	Multi Task
scaffolding (n=4,627 / 19.86%)	50.06	50.41
eliciting (n=1,875 / 8.05%)	24.95	24.79
repair (n=957 / 4.11%)	2.99	9.14
others (n=931 / 4.00%)	47.22	45.55
Micro F_1	42.01	42.20
Macro F_1	31.30	32.47
Weighted F_1	38.99	39.67

当しない負例クラスとして扱い、 F_1 スコアの算出対象には含めない。

6 実験結果

6.1 戦略の予測性能

モデルの戦略予測結果を表 2 に示す。各評価指標で最も高い性能を太字で、その次を下線で示す。実験の結果、マルチタスク学習が全ての評価指標で性能向上の傾向を示した (Weighted F_1 : +0.68%, Macro F_1 : +1.17%, Micro F_1 : +0.19%)。特に、Macro F_1 の向上幅が大きいことから、マルチタスク学習による低頻度戦略予測への寄与が示唆された。

6.2 戦略ごとの予測性能

具体的な戦略ごとの予測結果を表 3 に示す。表 3 から、提案手法によって repair の F_1 スコアが 6.15% 向上されていることがわかる。これにより、マルチタスク学習が repair 戦略に対して有効である可能性が示唆された。表 4 に repair の例を示す。

表 4 を見ると、学習者の間違いに対して教師がエラーを指摘し、その後の発話で正解を教えている (修正)。シングルタスクモデルは前の発話で行った修正で十分と判断し、戦略なしを予測したと考えられる。一方、マルチタスクモデルは、まだ修正が必要と予測し、学習者のエラーを適切に直している。修正戦略を取るためには、前の教師発話では学習者のエラーに十分に対応していないと判断する必要がある。従って、マルチタスク手法が、現段階で修正が十分かに対する正確な判断に寄与していると考えられる。

表 4 対話履歴と戦略の例

対話履歴	
Teacher:	(戦略なし) <i>Thanks could you choose two of those and give me an example?</i>
Student:	<i>The film gave me impression.</i>
Teacher:	(修正) <i>Thanks <STUDENT>do you mean : I was impressed by the film?</i>
Student:	<i>Yes. Is that correct?</i>
Teacher:	(修正) <i>Sorry I mean <STUDENT>unfortunate error!</i>
教師発話 (ゴールドデータ)	
Teacher:	(修正) <i>You need to say: the film made an impression on me</i>
各モデルの出力	
Single Task:	(戦略なし)
Multi Task:	(修正) <i>You mean: "The film gave me a good impression."</i>

7 おわりに

本研究では、1対1英語学習ドメインに対する LLM の教師戦略予測の性能を確かめた。実験の結果、発話を同時に生成する提案手法によって性能は引き上げられた。また、提案手法が修正戦略が必要かどうかに対する判断に寄与していることが示唆された。しかし、戦略予測タスクの最大予測性能が Micro F_1 で 42.20% であり、性能の低さが見られている。その原因として、戦略頻度の不均衡が挙げられる。表 3 では、others を除いた三つの戦略に対して、戦略数が多い戦略ほど予測性能が高い傾向が観察されている。今後の方針は二つ考えられる。一つ目は、上記で述べた戦略頻度の不均衡の解消である。これは、頻度に比例したサンプリングや損失重み付与などの学習手法が考えられる。二つ目は、教師の戦略を予測するだけでなく、その理由を分析することである。現在のモデルは次の教師発話に必要な戦略だけを予測し、「なぜその戦略が必要なのか」に対して根拠を提示していない。したがって、モデルが対話履歴のどのようなところに基づいて戦略を予測したのかを説明可能な AI 手法 (例: LIME [21]、SHAP [22] など) を用いて分析する。モデルの戦略予測根拠を提示することで、AI の透明性を提供し、信頼性の確保に寄与できると期待される。

謝辞

本研究は JSPS 科研費 JP25K00470 の助成を受けたものです。

参考文献

- [1] Benjamin S Bloom. The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational researcher*, Vol. 13, No. 6, pp. 4–16, 1984.
- [2] Katherine Stasaski, Kimberly Kao, and Marti A. Hearst. CIMA: A large open access dialogue dataset for tutoring. In Jill Burstein, Ekaterina Kochmar, Claudia Leacock, Nitin Madnani, Ildikó Pilán, Helen Yannakoudakis, and Torsten Zesch, editors, **Proceedings of the Fifteenth Workshop on Innovative Use of NLP for Building Educational Applications**, pp. 52–64, Seattle, WA, USA → Online, July 2020.
- [3] Anaïs Tack, Ekaterina Kochmar, Zheng Yuan, Serge Bibauw, and Chris Piech. The BEA 2023 shared task on generating AI teacher responses in educational dialogues. In Ekaterina Kochmar, Jill Burstein, Andrea Horbach, Ronja Laarmann-Quante, Nitin Madnani, Anaïs Tack, Victoria Yaneva, Zheng Yuan, and Torsten Zesch, editors, **Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023)**, pp. 785–795, Toronto, Canada, July 2023.
- [4] Rose Wang and Dorottya Demszky. Is ChatGPT a good teacher coach? measuring zero-shot performance for scoring and providing actionable insights on classroom instruction. In Ekaterina Kochmar, Jill Burstein, Andrea Horbach, Ronja Laarmann-Quante, Nitin Madnani, Anaïs Tack, Victoria Yaneva, Zheng Yuan, and Torsten Zesch, editors, **Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023)**, pp. 626–667, Toronto, Canada, July 2023.
- [5] Shashank Sonkar, Kangqi Ni, Sapana Chaudhary, and Richard Baraniuk. Pedagogical alignment of large language models. In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen, editors, **Findings of the Association for Computational Linguistics: EMNLP 2024**, Miami, Florida, USA, November 2024.
- [6] Lingzhi Wang, Mrinmaya Sachan, Xingshan Zeng, and Kam-Fai Wong. Strategize before teaching: A conversational tutoring system with pedagogy self-distillation. In Andreas Vlachos and Isabelle Augenstein, editors, **Findings of the Association for Computational Linguistics: EACL 2023**, pp. 2268–2274, Dubrovnik, Croatia, May 2023.
- [7] Rose Wang, Qingyang Zhang, Carly Robinson, Susanna Loeb, and Dorottya Demszky. Bridging the novice-expert gap via models of decision-making: A case study on remediating math mistakes. In Kevin Duh, Helena Gomez, and Steven Bethard, editors, **Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)**, pp. 2174–2199, Mexico City, Mexico, June 2024.
- [8] Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation, 2014.
- [9] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault, editors, **Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics**, pp. 7871–7880, Online, July 2020.
- [10] Ananya Ganesh, Martha Palmer, and Katharina Kann. What would a teacher do? Predicting future talk moves. In Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli, editors, **Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021**, pp. 4739–4751, Online, August 2021.
- [11] Fareya Ikram, Alexander Scarlato, and Andrew Lan. Exploring LLMs for predicting tutor strategy and student outcomes in dialogues. In Ekaterina Kochmar, Bashar Alhafni, Marie Bexte, Jill Burstein, Andrea Horbach, Ronja Laarmann-Quante, Anaïs Tack, Victoria Yaneva, and Zheng Yuan, editors, **Proceedings of the 20th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2025)**, pp. 765–779, Vienna, Austria, July 2025.
- [12] Aseem Srivastava, Ishan Pandey, Md. Shad Akhtar, and Tanmoy Chakraborty. Response-act guided reinforced dialogue generation for mental health counseling, 2023.
- [13] Hongru Wang, Mingyu Cui, Zimo Zhou, Gabriel Pui Cheong Fung, and Kam-Fai Wong. Topicrefine: Joint topic prediction and dialogue response generation for multi-turn end-to-end dialogue system, 2021.
- [14] Andrew Caines, Helen Yannakoudakis, Helen Allen, Pascual Pérez-Paredes, Bill Byrne, and Paula Buttery. The teacher-student chatroom corpus version 2: more lessons, new annotation, automatic detection of sequence shifts. In David Alfter, Elena Volodina, Thomas François, Piet Desmet, Frederik Cornillie, Arne Jönsson, and Evelina Rennes, editors, **Proceedings of the 11th Workshop on NLP for Computer Assisted Language Learning**, pp. 23–35, Louvain-la-Neuve, Belgium, December 2022. LiU Electronic Press.
- [15] Jakub Macina, Nico Daheim, Sankalan Chowdhury, Tanmay Sinha, Manu Kapur, Iryna Gurevych, and Mrinmaya Sachan. Math-Dial: A dialogue tutoring dataset with rich pedagogical properties grounded in math reasoning problems. In Houda Bouamor, Juan Pino, and Kalika Bali, editors, **Findings of the Association for Computational Linguistics: EMNLP 2023**, pp. 5602–5621, Singapore, December 2023.
- [16] Lev S Vygotsky. **Mind in society: The development of higher psychological processes**, Vol. 86. Harvard university press, 1978.
- [17] Shanjida Halim and Tanzina Halim. Elicitation: A powerful diagnostic tool for actively involving learners in the learning process. **Arab World English Journal (AWEJ) Special Issue: The Dynamics of EFL in Saudi Arabia**, 2019.
- [18] Shaofeng Li. The effectiveness of corrective feedback in sla: A meta-analysis. **Language learning**, Vol. 60, No. 2, pp. 309–365, 2010.
- [19] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In Jill Burstein, Christy Doran, and Thamar Solorio, editors, **Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)**, pp. 4171–4186, Minneapolis, Minnesota, June 2019.
- [20] Llama Team. The llama 3 herd of models, 2024.
- [21] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "why should i trust you?": Explaining the predictions of any classifier. In **Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**, KDD '16, p. 1135–1144, New York, NY, USA, 2016. Association for Computing Machinery.
- [22] Scott M. Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In **Proceedings of the 31st International Conference on Neural Information Processing Systems**, NIPS'17, p. 4768–4777, Red Hook, NY, USA, 2017. Curran Associates Inc.

Given a tutoring dialogue context, predict what teaching strategy(ies) the teacher should use in their NEXT utterance.
(Given a tutoring dialogue context, first identify what teaching strategy(ies) the teacher should use, then predict what the teacher should say in their NEXT utterance.)

CORE TEACHING STRATEGIES

1. scaffolding

Definition: Providing support by giving knowledge, explanations, examples, or step-by-step guidance

When to use:

- When student needs help understanding or completing a task
- When student seems to understand when proper explanations are given

Examples:

- Teacher gives examples of the word since student seems to misunderstand it
- Teacher shows how to use specific grammar since student shows uncertainty

2. eliciting

Definition: Asking questions to check understanding or draw out student's existing knowledge

When to use:

- To assess what student knows about specific words or grammar
- To make student think about specific concepts

Examples:

- Teacher asks student if they know similar words when they're not sure about it
- Teacher gives student a question to answer to assess if they know it or not

3. repair

Definition: Correcting errors in student's response or teacher's own previous statement

When to use:

- When student makes a mistake
- When teacher needs to fix their own error (typo, incorrect information)

Examples:

- Fixing teacher's previous statement since it includes a typo
- Teacher fixes 'afraid' since it seems student wanted to say 'afford'

KEY DISTINCTIONS

- scaffolding = GIVING information/help/explanation
- eliciting = ASKING questions to retrieve/check information
- repair = CORRECTING errors (student's or teacher's)

OTHER STRATEGIES

Beyond the core three strategies, teachers may use: introducing new knowledge (presentation), revisiting previous content (revision), seeking clarity (clarification), asking non-knowledge question (enquiry), summarizing (recap), or referencing materials (reference).

****When to predict "others":****

- Teacher introduces completely new knowledge for the first time
- Teacher summarizes the lesson or revisits past material
- Teacher asks questions without checking/assessing knowledge
- Teacher makes unclear points clearer

図3 LLMの共通プロンプト。下線はマルチタスク時に変わる。

PREDICTION GUIDELINES

1. Analyze the dialogue flow:

- What just happened in the conversation?
- What does the student need next?
- What would be the logical next teaching move?

2. Multiple labels:

- Use multiple labels when teacher needs to combine strategies
- Example: ["scaffolding", "eliciting"], ["repair", "scaffolding"], ["eliciting", "others"]

3. ["NA"] usage:

- Only for non-teaching utterances
- Example: greetings, small talk, lesson closing

OUTPUT FORMAT

Return a JSON array of strings only, with NO explanation or additional text.

Valid outputs examples:

- ["scaffolding"]
- ["eliciting"]
- ["others"]
- ["NA"]
- ["scaffolding", "eliciting"]
- ["repair", "scaffolding", "others"]
- ["eliciting", "others"]

図4 図3に続くシングルタスクプロンプト。

PREDICTION GUIDELINES

1. Analyze the dialogue flow:

- What just happened in the conversation?
- What does the student need next?
- What would be the logical next teaching move?

2. Predict strategies (multiple labels allowed):

- Use multiple labels when teacher needs to combine strategies
- Example: ["scaffolding", "eliciting"], ["repair", "scaffolding"], ["eliciting", "others"]

3. Generate the teacher's next utterance:

- Make it natural and contextually appropriate
- Keep it concise and focused
- Align with the predicted strategy(ies)

4. ["NA"] usage:

- Only for non-teaching utterances
- Example: greetings, small talk, lesson closing

OUTPUT FORMAT

Return a JSON object with two fields only, no explanation.

Valid outputs examples:

- {"strategies": ["scaffolding"], "utterance": "yes right idea for the meaning = good BUT you need a noun i.e. the thing that led him..."}
- {"strategies": ["eliciting", "repair"], "utterance": "Food can be out of date / past it's best, or it can be - (the opposite)?"}
- {"strategies": ["NA"], "utterance": "yes it's the one thing really you can do"}

図5 図3に続くマルチタスクプロンプト。