

ローカル LLM を用いて保育記録と資質・能力の三つの柱との関連性を読み取る手法の検討

Methods for Analyzing the Relationship Between Childcare Records and Children's Qualities and Abilities Using Local Large Language Models

湯地 調*¹ 加藤 直樹*¹ 樋山 淳雄*¹
Shirabe YUJI*¹ Naoki KATO*¹ Atsuo HAZEYAMA*¹

東京学芸大学*¹
Tokyo Gakugei University*¹

<あらまし> 平成 29 (2017) 年告示の幼稚園教育要領等では、「幼児期において育みたい資質・能力」が明示されている。本稿では、保育記録に記された幼児の姿と資質・能力の三つの柱との関連性を読み取る手法を、ローカル LLM で実行可能な形で設計し、有識者との一致率、相関から有用性を検証した。結果として、ローカル LLM の種類とプロンプト設計の組み合わせによっては、有識者の評価傾向と一定の一致が得られることが示唆された。

<キーワード> 資質・能力の三要素, 保育記録, ローカル LLM

1. はじめに

平成 29 (2017) 年に告示された幼稚園教育要領 (文部科学省 2017) 等では、「幼児期において育みたい資質・能力 (知識及び技能の基礎, 思考力・判断力・表現力等の基礎, 学びに向かう力, 人間性等)」(以降, 三要素と記す) を一体的に育むように努めることが示されている。

筆者らはこれまで, 生成 AI を用いて保育記録に記された幼児の姿と三要素がどの程度関連しているかの度合い (以降, 関連度と記す) を読み取る手法の開発を行ってきた (Yuji et al 2025)。しかし, 生成 AI の外部 API の利用には, 個人情報や運用コスト等の制約がある。

そこで, 本研究では, ローカル環境で使用できる大規模言語モデル (ローカル LLM) を用いた手法を設計し, 有用性を検証する。

2. 資質・能力を読み取る手法の設計

2.1. 基本コンセプト

本手法は, 保育記録の幼児の姿を記した文章から, 有識者 (幼児教育専門の大学教員 2 名) の意見を基に作成したループリック (表 1) に基づき, 幼児の姿と三要素それぞれとの関連度を一文単位, 5 段階で算出する。

生成 AI には GPT-4o を用いる。開発主体による違いや, 日本語のデータの追加学習の有無による違いを検討するため, ローカル LLM に

は, gpt-oss-20b, gemma-3-12b, Meta-Llama-3-8B-Instruct に日本語の追加学習を行った Llama-3-ELYZA-JP-8B を選定した。

2.2. 詳細設計

関連度を算出するための GPT-4o 及び各ローカル LLM (以降, 対象モデルと記す) への指示文 (プロンプト) について, 手法の基盤となる基本プロンプトと, 追加の有無を制御する制御プロンプトを設計した。

2.2.1. 基本プロンプト

前節で述べたループリックを含め, 入力した文における関連度を 1~5 の整数で評価するよう指示する。役割を明確にするため, 「あなたは幼児教育の専門家です」といった内容も含める。

2.2.2. 制御プロンプト

(1) 三要素の説明の提示

幼稚園教育要領から, 三要素の内容を説明した部分を抜粋し, 制御プロンプトとする。

表 1 関連度のループリック

関連度	評価の基準
1	発揮している姿が明示されており, 発揮していることが明確に推察できる
2	発揮している姿が明示されていないが, 発揮していることが明確に推察できる
3	発揮している姿が明示されていないが, 発揮している可能性があることを推察できる
4	発揮している姿が明示されておらず, 発揮していないことが推察できるが, 断定できない
5	発揮している姿が明示されておらず, あらわれていないと断定できる

(2) 三要素の具体例の提示

幼稚園教育要領にある「幼児期の終わりまでに育ってほしい姿」の説明文から、三要素に関連する部分を抜粋し、関連度が5になる行動の具体例として制御プロンプトとする。

3. モデル間の比較検証

3.1. 検証方法

検証に用いた文は、「幼保連携型認定こども園における園児が心を寄せる環境の構成」(文部科学省 2012) の実践事例より、無作為に選択した6事例各3文ずつの計18文である。

幼児教育の有識者6名(幼児教育専門の大学教授3名、幼稚園等の教師3名)は、事例全文を読んだ上で、ループリックを基に各文の関連度を決定した。提案手法には、対象の18文のみを入力し、関連度算出を行った。

前章で設計した基本プロンプト及び二つの制御プロンプトの有無を組み合わせた計4つの手法(表2)を対象モデルに適用して関連度算出を行い、比較を行った。

3.2. 結果

3.2.1. 一致率

過半数の有識者の関連度評価が一致した文において、有識者が評価した関連度と、各手法を適用した対象モデルそれぞれが算出した関連度が一致した割合を一致率とした。対象モデル及び手法ごとの一致率を表3に示す。

結果として、GPT-4o が最も高い一致率を示す場合が多かったが、「知識及び技能の基礎」において、gemma-3-12b+手法cが95%の一致率を示すなど、手法によってローカルLLMが高い一致率を示す場合があった。「学びに向かう力、人間性等」の一致率は全体として低かった。

3.2.2. 相関分析

検証18文それぞれに対して、各手法を適用した対象モデルが算出した関連度(10回試行の平均値)と、有識者6名が評価した関連度の平均値を用いて、三要素ごとに相関分析(有意水準

5%)を行った。

結果として、gpt-oss-20b+手法c、gemma-3-12b+手法dが三要素すべてで有意な相関を示すなど、ローカルLLMと有識者の傾向との一定の一致が確認された。

4. おわりに

本稿では、保育記録の幼児の姿と三要素との関連度を読み取る手法を、ローカルLLMで実行可能な形で設計し、有識者との一致率、相関から有用性を検証した。結果として、ローカルLLMを用いた手法は、GPT-4oを用いた手法よりも全体的に一致率が低かった。この要因として、事前学習したデータ量の差による影響が考えられる。一方、ローカルLLMの種類とプロンプト設計の組み合わせによっては、有識者の評価傾向と一定の一致が得られ、ローカル環境においても、教師の幼児理解を支援する関連度算出を実現できる可能性があることが示唆された。

今後の課題として、多様な事例での検証、データの前処理等による一致率の向上、本手法を実装したツールの現実的評価が挙げられる。

参考文献

- 文部科学省(2017) 幼稚園教育要領, フレーベル館
- Yuji et al. (2025) Proposal and Development of Tools for Visualizing the Relationship Between Childcare Records and Children's Qualities and Abilities Using Generative AI. *Procedia Computer Science*, 270:1727-1736.
- 文部科学省(2012) 幼保連携型認定こども園における園児が心を寄せる環境の構成, フレーベル館

表3 対象モデル及び手法ごとの一致率

対象モデル	手法	知	思	学
GPT-4o	a	53.0%	82.0%	47.1%
	b	82.0%	66.7%	30.0%
	c	87.0%	53.3%	42.9%
	d	70.0%	77.8%	38.6%
gpt-oss-20b	a	33.0%	60.0%	31.4%
	b	58.0%	53.3%	4.3%
	c	61.0%	60.0%	7.1%
	d	44.0%	58.9%	31.4%
gemma-3-12b	a	62.0%	41.1%	28.6%
	b	77.0%	64.4%	41.4%
	c	95.0%	67.8%	10.0%
	d	34.0%	56.7%	41.4%
Llama-3-ELYZA-JP-8B	a	46.0%	24.4%	28.6%
	b	78.0%	55.6%	21.4%
	c	87.0%	38.9%	11.4%
	d	63.0%	58.9%	30.0%

表2 手法の一覧

手法	制御プロンプト	
	三要素の説明	具体例
a	○	×
b	×	○
c	○	○
d	×	×