認知的情報処理を活用した看護診断支援の学習法の提案

-学習データ生成とアセスメントパターン分類器の構築-

久保　貴弘　武蔵野大学

Ⅰ　背景

看護記録は看護診断過程を反映する価値の高いデータであるが、個人情報保護法・医療法の保護対象であり、家族等の第三者情報も含むため実データの二次利用は容易ではない。代替策である書籍からの模擬事例抽出も、事例数・多様性の両面で機械学習モデルの学習規模には届きにくい。近年の AI・大規模言語モデル（LLM）の進展は看護分野に新たな可能性を開くが、看護師の認知的情報処理過程を AI の学習データとして体系化する研究は限られている。看護診断支援 AI の構築には、S・O・アセスメント・看護診断を体系的に対応づけた学習データを、実データに依存せず確保する方法の確立が課題である。

Ⅱ　目的

本研究は、生成 AI を用いた患者ペルソナ×臨床シナリオの二段階生成により、看護診断支援 AI の学習データを創造する方法を提案し、生成データの構造的整合性を検証すること、さらに生成した合成データを用いて看護診断の前段階となるアセスメントパターン（Gordon's FHP【1】）分類器を構築し、その有効性を検証することを目的とする。

Ⅲ　方法

研究デザインは、看護診断支援システムの基盤となる学習データを構築する方法論的研究である。大規模言語モデル（OpenAI GPT-4o、temperature=0.85）による二段階生成アーキテクチャを構築し、第 1 段階で属性を辞書型データから組み合わせ多様な患者ペルソナを生成、第 2 段階で各ペルソナに基づき S・O・アセスメント・看護診断を生成した（図 1）。

プロンプトには熟練看護師としての役割定義、Gordon's FHP 全パターンの網羅指定、NANDA-I【2】指標参照による論理的整合性の担保等を組み込んだ。JSON 構文検証と自己修復による品質保証機構を通し（最終成功率 99.8%）、所定の基準でデータクリーニングを行った。

生成データの構造的整合性については、無作為抽出した 30 事例に対し、看護資格を有する単一の評価者が、①データの自然さ、②S・O データおよび Gordon's FHP 間の内的整合性、③専門用語等の記述適切性の 3 観点から 5 件法 Likert 尺度で質的評価を行った。単一評価者による評価は本研究の限界である。

さらに、構築した合成データセットを用いて看護診断の前段階となるアセスメントパターン（Gordon's FHP）分類器を構築した。日本語事前学習済み BERT（cl-tohoku/bert-base-japanese-v3）をファインチューニングし、Gordon's FHP11 パターンへの分類タスクとして 5 分割のグループ層化交差検証により学習・評価した。

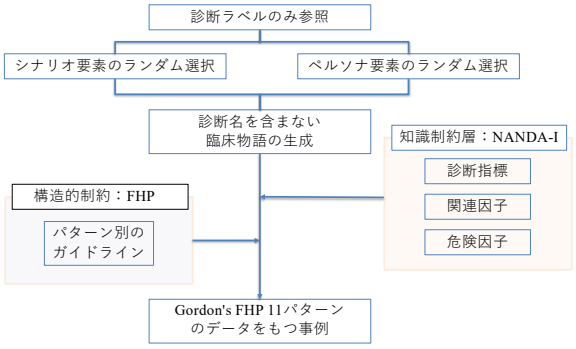

図 1　ペルソナ×シナリオ二段階生成アーキテクチャ

Ⅳ　結果

二段階生成により当初 558 事例を生成し、データクリーニングを経て 519 事例・97,307 件（約 10 万件）のデータセットを構築した。NANDA-I 看護診断 56 種類を網羅し、診断あたりの生成事例数は 6〜10 事例（目標 10 事例に対し達成率約 9 割）であった。この未達は、LLM 特有の input-conflicting

hallucination【3】によるものと考えられる。質的評価（30 事例）では総合スコア 3.5 未満の事例が 12 件（40%）あった。

表 1　質的評価における問題パターンの分類（低評価 12 事例の内訳、複数該当あり）

問題カテゴリ	該当数	主な内容
臨床的不自然さ	5 件	訴えの強さとバイタルサインの矛盾　等
論理的矛盾	11 件	基本属性・社会的背景に関する明らかな矛盾　等
記述の不備	11 件	病態を裏付ける客観的データの不足　等

詳細分析の結果、低評価事例の多くは年齢・職業等の基本属性の設定誤りに起因し、看護診断に直結する S・O データの整合性は保たれていた。これを踏まえ、生成事例全体の約 70%が学習データとして使用可能な水準にあると判断した。統計的評価では、S・O データ比率 48.5：51.5（理論値との差は統計的に有意だが実務上無視できる水準）、Gordon's FHP11 パターン間のデータ量の変動係数 4.1%と高いバランスを確認した。

FHP 分類器は 11 クラス分類で正解率 42.85%を達成し、ベースライン 8.6%から約 5 倍の性能向上を示した（Macro F1=0.408、Weighted F1=0.423）。一方、S・O まで区別する 22 クラス分類は、データ量不足により正解率 0.68%にとどまり、分類粒度とデータ量のバランスが性能に影響することが示唆された。

Ⅴ　考察

生成 AI による二段階生成は、実データの二次利用が制約される看護分野で学習データを効率的に確保する方法として有効性が示唆された。診断ラベルの事前指定による生成制御は実データ収集にはない利点である。一方、LLM 特有の input-conflicting hallucination や約 3 割の事例に見られた整合性・根拠づけの不備は、生成 AI が指示や妥当性を常に完全には担保できない限界を示す。単一評価者による質的評価は評価者間信頼性の観点から限界があり、複数評価者による検証が今後の課題である。また、合成データで学習した FHP 分類器がベースラインを大幅に上回ったことは、合成データが看護診断に必要な意味構造を含む客観的証拠であるが、この正解率は看護診断そのものではなくその前段階であるアセスメントパターン分類の精度である点には留意が必要である。本研究の方法論は、看護分野における合成データ活用の新たな知見を提供すると考える。今後は実データへの転移可能性の検証と生成プロセスの改善を進める必要がある。

倫理的配慮

本研究は書籍由来の模擬事例および大規模言語モデルによる合成データのみを使用し、実在する患者データは一切使用していない。そのため、個人情報の匿名化処理および倫理審査委員会（IRB）の承認は不要であった。今後、実データを用いた検証の際は、倫理審査の承認を得た上で適切な手順を遵守する。

引用文献

【1】M. Gordon, Manual of Nursing Diagnosis, 13th ed. Burlington, MA, USA: Jones & Bartlett Learning, 2014.

【2】T. H. Herdman, S. Kamitsuru, and C. T. Lopes, NANDA International Nursing Diagnoses, 12th ed. Thieme Medical Publishers, 2021.

【3】Z. Ji et al., "Survey of Hallucination in Natural Language Generation," ACM Comput. Surv., Nov. 2022.