

異種グラフを用いた知識モニタリング推定とメタ認知適応型学習支援

李 根[†] 陳 莉^{††} 唐 成[†] 馬 博軒[†] 出口 大輔^{†††}
山下 隆義^{††††} 島田 敬士[†]

[†]九州大学 〒819-0395 福岡県福岡市西区元岡 744 番地

^{††}大阪教育大学 〒582-8582 大阪府柏原市旭ヶ丘 4 丁目 698 番 1 号

^{†††}名古屋大学 〒464-8601 愛知県名古屋市千種区不老町

^{††††}中部大学 〒487-8501 愛知県春日井市松本町 1200 番地

E-mail: [†] {gen.li,tang,atsushi}@limu.ait.kyushu-u.ac.jp, boxuan@artsci.kyushu-u.ac.jp, ^{††} chen-l68@cc.osaka-kyoiku.ac.jp, ^{†††} ddeguchi@nagoya-u.ac.jp, ^{††††} takayoshi@isc.chubu.ac.jp

あらまし 従来の学習支援システムは主にテストログから推定される知識レベルに基づいて支援を行うが、学習者が自身の理解度を適切に判断できるかという知識モニタリング能力の観点には十分に考慮されていない。知識モニタリングとは自身の理解状態を客観的に判断する能力であり、学習の質や効率に大きく影響するため、知識レベルと併せて考慮した支援が重要である。本研究では、テストログと学習リフレクションを統合し、学習者の潜在的な知識モニタリング状態を推定する手法を提案する。学習リフレクションでは全ての知識概念が言及されるとは限らないため、学習者・知識概念・概念間関係を異種グラフとして構築し、欠損する学習者と概念関係の補完をリンク予測問題として定式化する。さらに、推定結果に基づき学習者を知識レベルと知識モニタリング能力の二軸で類型化し、Feed Up・Feed Back・Feed Forward の三段階に従ったメタ認知適応型フィードバックを提示するシステムを実装した。実データを用いた推定精度の検証およびユーザ実験によるシステムの評価を通じて、提案手法の有効性と教育実践への適用可能性を示す。

キーワード 異種グラフ, 知識モニタリング, 学習支援, LLM

1 はじめに

効果的な学習支援の実現には、学習者が自身の理解状態を正確に把握し、適切な学習方略を選択する能力が不可欠である [31]。近年、学習管理システム (LMS) や知的学習支援システム (ITS) の普及により、テストログや行動ログから学習者の知識レベルを推定し、個別化された支援を提供することが可能になっている [17], [23]。これらのシステムは主に学習者が「何を知っているか」を評価し、不足領域に対する学習教材の推薦などにより一定の成果を上げている [4]。

しかしながら、知識レベルの評価のみでは効果的な学習支援には不十分な場合がある。教育現場では、実際には理解していないにも関わらず「わかったつもり」になる過信や、十分な知識があるにも関わらず自信を持ってない過小評価が頻繁に観察される [14], [18]。このような現象は、学習者の知識モニタリング能力、すなわち自身の理解状態を正確に判断するメタ認知能力の不足に起因する [3]。したがって、学習支援システムは知識レベルだけでなく、知識モニタリング能力も考慮した支援を提供することが重要である。

従来、知識モニタリング能力の評価には、各問題に対する自信度を回答させる構造化質問紙が用いられてきた [25]。しかし、この手法は評価項目を事前に固定する必要があり、多様なカリキュラムへの柔軟な適用が困難であった。一方、自由記述式の

学習リフレクションは、自己調整学習 (SRL) における自己省察フェーズの中核的实践として多くの教育機関で実施されている [26]。学習者が能動的に自らの理解状態を言語化するこの形式は、事前定義された選択肢への受動的な回答と比較して、より真正なメタ認知プロセスを反映できる。また、評価項目が固定されないため、カリキュラム変更にも柔軟に対応可能である。しかし、学習者が評価対象となる全ての知識概念について言及するとは限らないため、言及されなかった概念に対する認知状態の推論が課題となる。

本研究では、テストログと学習リフレクションを統合し、学習者の知識レベルと知識モニタリング能力を同時に評価するフレームワークを提案する。大規模言語モデル (LLM) を用いて教材と自己報告テキストから知識エンティティを抽出し、学習者・知識概念・概念間関係からなる異種グラフを構築する。学習者が言及しなかった概念に対する認知状態の推定は、異種グラフニューラルネットワーク (HGNN) によるリンク予測問題として定式化する。さらに、推定結果に基づき学習者を知識レベルと知識モニタリング能力の二軸で5つのパターンに類型化し、Hattie & Timperley のフィードバックモデル [11] に基づく Feed Up・Feed Back・Feed Forward の三段階適応型フィードバックを提供するシステムを実装した。

本研究では、以下の研究課題に取り組む：

- **RQ1:** 異種グラフに基づくリンク予測は、学習者の潜在的な知識モニタリング状態をどの程度正確に推定できるか？

- **RQ2:** 知識モニタリング能力に基づく適応型フィードバックは、学習者と教育者にどのように評価されるか？

本論文の貢献は以下の通りである：(1) 自由記述式の学習リフレクションから知識モニタリング能力を推定する異種グラフベースの手法、(2) 知識レベルと知識モニタリング能力を統合した学習者類型化に基づく適応型フィードバックシステムの設計と実装、(3) 実データによる推定精度の検証およびユーザ実験によるシステム評価

2 関連研究

2.1 知識モニタリング評価

知識モニタリングとは、自身の知識状態を正確に判断するメタ認知能力であり、効果的な学習に不可欠な要素である。初期の研究では、学習者の自信判断と実際のパフォーマンスを比較することでメタ認知的校正を評価する構造化パラダイムが確立された [27]。これらの研究は、自己調整学習における校正の役割を強調し、メタ認知的制御やタスク解釈を含む理論的拡張を提案した [24]。

知識モニタリング能力の定量化には、信号検出理論に基づく指標が広く用いられている [2], [20]。学習者の認識（理解している/していない）と実際のパフォーマンス（正答/誤答）の対応関係から、誤警報率、見逃し率、および弁別力 d' などの指標が算出される。これらの指標により、単なる正答率では捉えられない学習者の自己認識の正確さを評価できる。しかし、これらの手法は事前定義された閉鎖的な回答形式に依存しており、学習者の表現の柔軟性が制限されるという課題がある。

2.2 教育分野における LLM 活用

教育分野において、LLM の重要な応用の一つは、教育コンテンツからの知識構造抽出である。LLM は教材から知識概念を抽出し [22]、知識グラフと組み合わせることで非構造化コンテンツを構造化表現に変換できる [5], [6], [15], [21]。このグラフベースの文脈モデリングは、教育応用における概念間の依存関係を捉える機会を提供する。また、プログラミングバグへの対応 [9]、自動評価 [12] などのタスクにも活用されている。さらに、学習者へのフィードバック生成にも LLM が活用されており [13], [28]、適応型学習支援の実現に貢献している。

2.3 グラフベースの学習モデリング

グラフベースのモデリングは、教育コンテンツと学習者行動の関係性を表現する有効な手法である。教育知識グラフは前提関係や概念階層を符号化し [19]、その構築手法はニューラルシーケンスラベリングから確率的アソシエーションマイニングまで多岐にわたる [7]。近年では、強化学習を用いた解釈可能な推薦パスの生成 [10] や、RDF ベースの表現 [1] も提案されている。

グラフニューラルネットワーク (GNN) は、演習問題と概念間の潜在的依存関係を捉えることができる [16], [30]。GKT-CD [29] のようなアプローチは、学習者-概念インタラクションを通じて習熟度を推論する。また、教育理論に基づくグラフベースフレームワークにより認知プロセスのモデル化も試みら

れている [8]。しかし、既存のグラフベースモデルは知識モニタリング評価を十分に扱っていない。

3 提案手法

本章では、テストログと学習リフレクションを統合し、学習者の知識モニタリング状態を推定するフレームワークを提案する。図 1 に提案フレームワークの概要を示す。

3.1 問題設定

学習者集合 $S = \{s_1, \dots, s_n\}$ がコースに参加している教育環境を考える。 \mathcal{K} をコースで扱われる知識概念（概念、理論、原理など）の集合とする。学習者は以下の 2 つの活動に従事する：(1) 自由記述式の学習リフレクションにおいて、自身の理解状態を振り返り自由に記述する中で、理解していると認識する知識概念 (\mathcal{K}_i^+) と理解していないと認識する知識概念 (\mathcal{K}_i^-) に言及する；(2) 特定の知識概念を評価するための評価項目に回答する。各評価項目 $q_j \in \mathcal{Q}$ は主に知識概念 $k_j \in \mathcal{K}$ を評価し、学習者のパフォーマンスは正誤の二値として記録される。

しかしながら、課題となるのは評価される知識概念の集合 $\mathcal{K}_Q = \{k_j | q_j \in \mathcal{Q}\}$ が、学習者が明示的に言及した知識概念 $\mathcal{K}_i^+ \cup \mathcal{K}_i^-$ に完全には含まれない場合がある。すなわち、一部の学習者において $\mathcal{K}_Q \not\subseteq (\mathcal{K}_i^+ \cup \mathcal{K}_i^-)$ となる。このような「言及されなかった概念」に対する認識状態を潜在的認識状態 (Latent Perceived State) と呼ぶ。潜在的認識状態が存在する場合、テスト結果との対応付けが不完全となり、知識モニタリング能力の正確な評価が困難となる。本研究の目標は：(1) 潜在的認識状態を推定し、各学習者の知識モニタリング状態を補完すること；(2) 推定結果に基づく学習者類型化と適応型フィードバックの有効性を検証すること。

3.2 知識グラフの構築

3.2.1 知識概念の抽出と関係同定

LLM を用いて、コース教材から知識概念とその間の階層的・意味的關係を抽出し、知識グラフ $G = (\mathcal{K}, \mathcal{E}_{\mathcal{K}-\mathcal{K}})$ を構築する。ここで \mathcal{K} は知識概念の集合、 $\mathcal{E}_{\mathcal{K}-\mathcal{K}}$ は概念間の関係（前提関係、包含関係など）を表すエッジ集合である。

3.2.2 評価項目と知識概念のマッピング

各評価項目 q_j とそれが主に評価する知識概念 k_j を関連付けるため、LLM を用いてマッピング $M_Q : \mathcal{Q} \rightarrow \mathcal{K}$ を構築する。これにより、各学習者について正答した概念集合と誤答した概念集合を特定できる。

3.2.3 明示的認識概念の抽出

各学習者 s_i の学習リフレクションに対して LLM を適用し、言及された知識概念を抽出する。抽出された概念は、記述の文脈に基づき \mathcal{K}_i^+ （「理解している」と認識）または \mathcal{K}_i^- （「理解していない」と認識）に分類される。

3.3 HGNN に基づくリンク予測

3.3.1 異種学習者グラフの構築

構築したコンテキスト知識グラフから、異種学習者グラフ

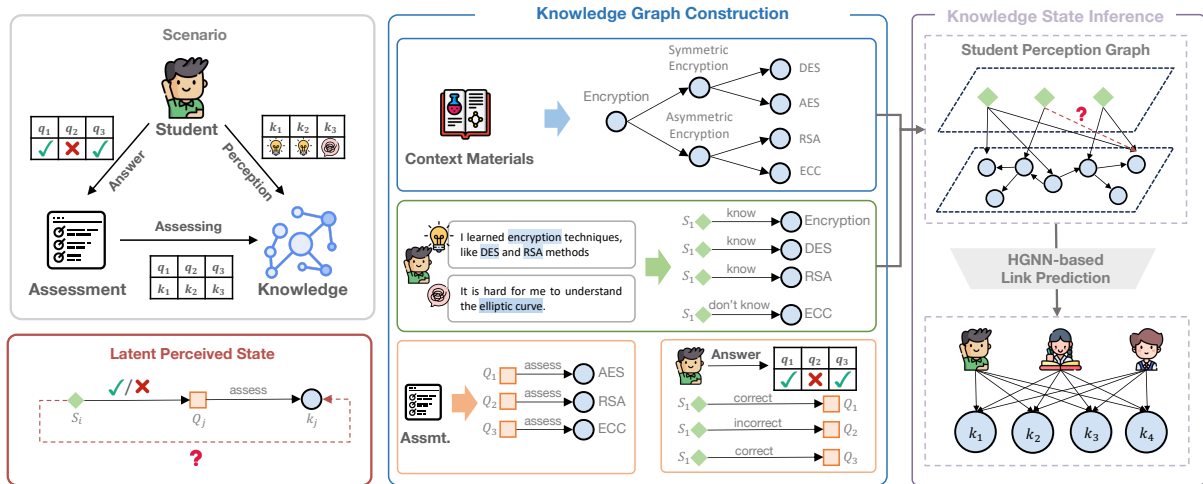


図 1: 提案フレームワークの概要. 左部のシナリオに示すように, 学習者はテストへの回答と学習リフレクションの記述という 2 つの活動に従事する. 中央部の知識グラフ構築では, LLM を用いて教材から知識概念の階層構造を抽出し, 学習リフレクションから学習者の認識状態(「理解している」/「理解していない」)を抽出する. 右部の知識状態推論では, 学習者が明示的に言及しなかった概念(図中の?マーク)に対する認識状態を, HGNN によるリンク予測として推定する.

$G_H = (V, E)$ を構築する. このグラフは学習者ノード $V_S = S$ と知識概念ノード $V_K = K$ を含む. エッジ集合 E は以下の 3 種類から構成される:

- **学習者-概念エッジ \mathcal{E}_{S-K}** : 学習者と, その学習者が「理解している」と認識した知識概念 (K_i^+) を接続する.
- **概念-概念エッジ \mathcal{E}_{K-K}** : 知識概念間の意味的・階層的关系を表現する.
- **概念-学習者エッジ \mathcal{E}_{K-S}** : 双方向メッセージパッシングのための逆エッジ.

重要な点として, 「理解していない」と認識した概念 (K_i^-) へのエッジはこの段階ではグラフに含めない. この設計により, 潜在的認識状態の推定は「学習者-概念間のエッジが存在するか否か」を予測するリンク予測問題として定式化される. すなわち, エッジの存在は「理解している」という認識を, エッジの不在は「理解していない」という認識を表す.

3.3.2 異種グラフニューラルネットワークによるリンク予測

潜在的認識状態を推論するため, HGNN を用いたリンク予測を行う. HGNN は異種情報をタイプ別メッセージパッシング機構により処理する:

$$h_v^{(l+1)} = \text{Aggregate}_{\tau(v)}^{(l)} \left(h_v^{(l)}, \{h_u^{(l)}, r : u \in N_r(v), \forall r \in R\} \right) \quad (1)$$

ここで, $\tau(v)$ はノード v の種類, $N_r(v)$ は関係タイプ r で接続された v の近傍ノード, $R = \{\mathcal{E}_{S-K}, \mathcal{E}_{K-K}, \mathcal{E}_{K-S}\}$ は全関係タイプ集合を表す.

ノード埋め込み h_s (学習者) と h_k (知識概念) を計算後, 学習者 s が知識概念 k を「理解している」と認識している確率を $\hat{y}_{s,k} = \sigma(f(h_s, h_k))$ として予測する. モデルは二値交差エントロピー損失を最小化するように学習される.

3.3.3 明示的信息に基づく負例サンプリング (EINS)

効果的なリンク予測モデルの学習のため, 明示的信息に基づ

く負例サンプリング (EINS: Explicit-Informed Negative Sampling) 方式を設計する. 前述の通り, K_i^- (「理解していない」と認識した概念) はグラフのエッジとして含まれていない. EINS はこの情報を負例として積極的に活用する:

1. **明示的負例**: K_i^- から直接サンプリング. 学習者が明示的に「理解していない」と報告した概念である.
2. **暗黙的負例**: 潜在的認識状態 (言及されなかった概念) からランダムサンプリング (比率パラメータ ρ で制御).

学習者の明示的な「理解していない」という報告を負例として優先的に活用することで, 学習者の自己認識の境界をより正確に学習できる.

3.3.4 潜在的認識状態の推論

学習済みモデルを用いて, 明示的に言及されなかった知識概念 $k_j \notin (K_i^+ \cup K_i^-)$ に対する学習者 s_i の潜在的認識状態を, 予測リンク確率に基づき推論する:

$$k_j \in \begin{cases} \hat{K}_i^+ & \text{if } \hat{y}_{s_i, k_j} \geq \theta \\ \hat{K}_i^- & \text{if } \hat{y}_{s_i, k_j} < \theta \end{cases} \quad (2)$$

ここで θ は閾値パラメータである. 推論の結果, 学習者の認識状態は $K_i^+ \cup \hat{K}_i^+$ (理解認識) と $K_i^- \cup \hat{K}_i^-$ (非理解認識) として補充される.

3.4 学習者類型化

3.4.1 知識モニタリング評価のためのカテゴリ分類

推論により補充された認識状態と実際のテスト結果を比較し, 知識モニタリング能力を評価する. 各知識概念を, 認識状態(「理解している」/「理解していない」)と実際の正誤の組み合わせにより, 以下の 4 カテゴリに分類する (表 1):

各カテゴリの解釈は以下の通りである:

- **ヒット (A)**: 理解していると認識し, 実際に正答
- **フォルスアラーム (B)**: 理解していると認識するが, 実際は誤答

表 1: 知識モニタリング評価のためのカテゴリ分類

	正答	誤答
「理解している」と認識	ヒット (A)	フォルスアラーム (B)
「理解していない」と認識	ミス (C)	正棄却 (D)

- ミス (C) : 理解していないと認識するが、実際は正答
- 正棄却 (D) : 理解していないと認識し、実際に誤答

3.4.2 知識モニタリング能力の定量化

信号検出理論に基づき、各学習者の知識モニタリング能力を定量化する。表 1 の 4 カテゴリから、以下の指標を導出する。

a) 弁別力 (d')

学習者が「理解している」状態と「理解していない」状態を正確に区別できる能力を表す総合指標である：

$$d'_i = z\left(\frac{|A_i|}{|A_i| + |C_i|}\right) - z\left(\frac{|B_i|}{|B_i| + |D_i|}\right) \quad (3)$$

ここで $z(\cdot)$ は標準正規分布の逆累積分布関数である。 d' が高いほど、学習者は自身の理解状態を正確に弁別できている。

b) 誤警報率 (False Alarm Rate)

誤答した概念のうち、「理解している」と誤認した割合を表す：

$$FAR_i = \frac{|B_i|}{|B_i| + |D_i|} \quad (4)$$

この値が高い学習者は、実際には理解していない概念を「理解している」と認識する過信傾向を示す。

c) 見逃し率 (Miss Rate)

正答した概念のうち、「理解していない」と誤認した割合を表す：

$$Miss_i = \frac{|C_i|}{|A_i| + |C_i|} \quad (5)$$

この値が高い学習者は、実際には理解している概念を「理解していない」と認識する過小評価傾向を示す。

これらの指標により、知識モニタリング能力の全体的な弁別力 (d') だけでなく、その内訳として過信傾向 (FAR) と過小評価傾向 (Miss) を把握できる。

3.4.3 学習者の 5 パターン類型化

学習者を知識レベル (テスト正答率) と知識モニタリング能力に基づき、以下の 5 パターンに分類する (表 2)：

表 2: 学習者の 5 パターン類型化

パターン	知識レベル	弁別力	追加条件	特徴
校正良好	高	高	—	理想的な状態
限界認識	低	高	—	不足を正確に認識
自信不足	高	低	見逃し率 高	能力を過小評価
自信過剰	低	低	誤警報率 高	理解したつもり
イエス傾向	高	低	見逃し率 低	常に『理解している』と回答

3.5 適応型フィードバック

学習者類型化の結果に基づき、個別最適化されたフィードバックを提供するための方略を設計する。

3.5.1 3段階フィードバックモデル

本研究では、Hattie & Timperley [11] が提唱した 3 段階フィードバックモデルを採用する。このモデルは、効果的なフィードバックが以下の 3 つの問いに答えるべきことを示している：

- **Feed Up** : 目標は何か (Where am I going?)
- **Feed Back** : 現状はどうか (How am I going?)
- **Feed Forward** : 次に何をすべきか (Where to next?)

本研究では、このモデルを知識レベルと知識モニタリング能力の二軸フレームワークに適用し、各学習者類型に応じた方略を定義する。

3.5.2 類型別フィードバック方針

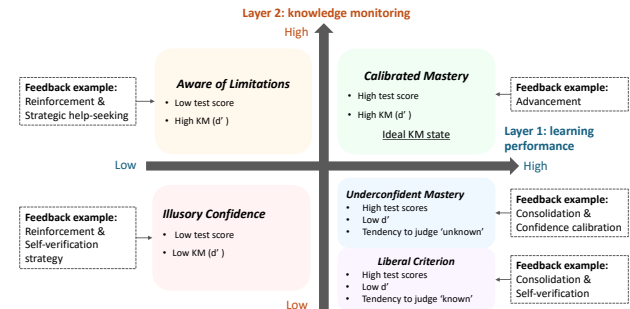


図 2: 学習者類型とフィードバック方略の対応

図 2 に、各類型に対応するフィードバック方略の概要を示す。知識レベルの改善に対しては、低成績の学習者には基礎概念の復習と定着を促す**強化学習**を、高成績の学習者には習得済み知識の発展・応用を促す**定着学習**を提供する。知識モニタリングの改善に対しては、類型別の方略を提供する。自信過剰の学習者には、理解を「既知」と判断する前に自己説明や回答確認などの**自己検証方略**を促す。自信不足の学習者には、正答した概念を振り返り成功体験を認識させる**自信校正**を促す。限界認識の学習者には、不明な概念に対する質問や教材参照などの**戦略的援助要請**を促す。

4 学習支援ダッシュボードの実装

4.1 概要

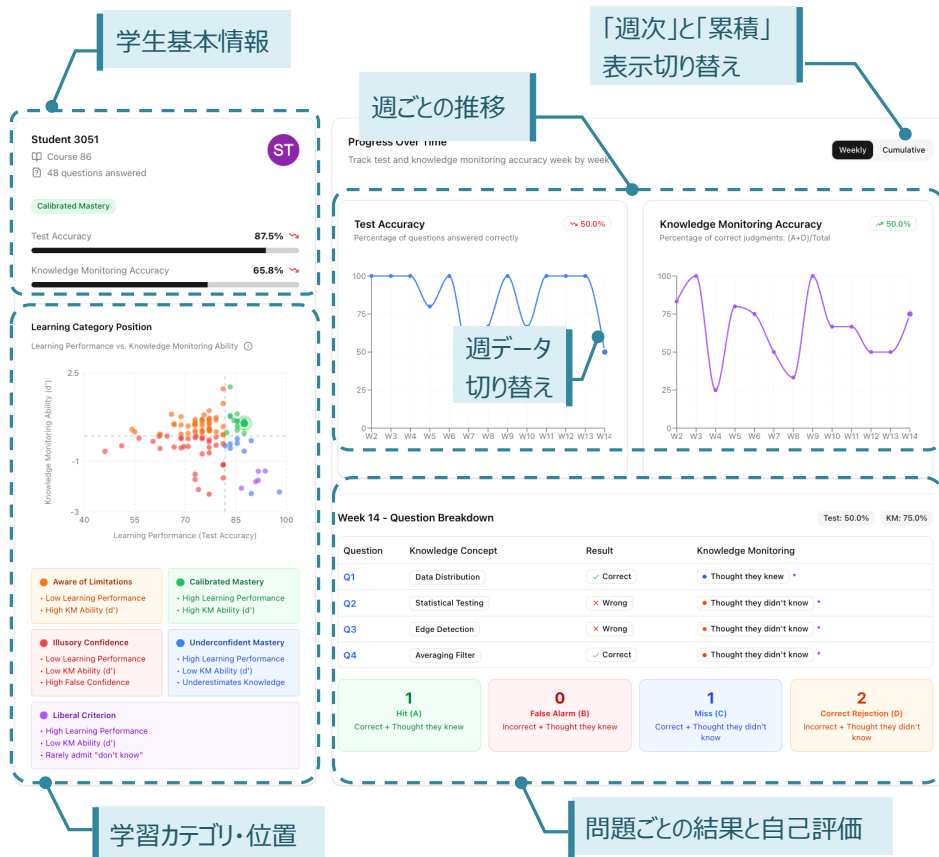
提案ダッシュボードは、学習者の知識レベルと知識モニタリング能力を可視化し、第 3 章で設計した適応型フィードバック方略に基づく支援を提供する Web アプリケーションである。図 3 に全体構成を示す。ダッシュボードは上部の**結果可視化部** (図 3a) と下部の**フィードバック部** (図 3b) から構成される。

4.2 結果可視化部

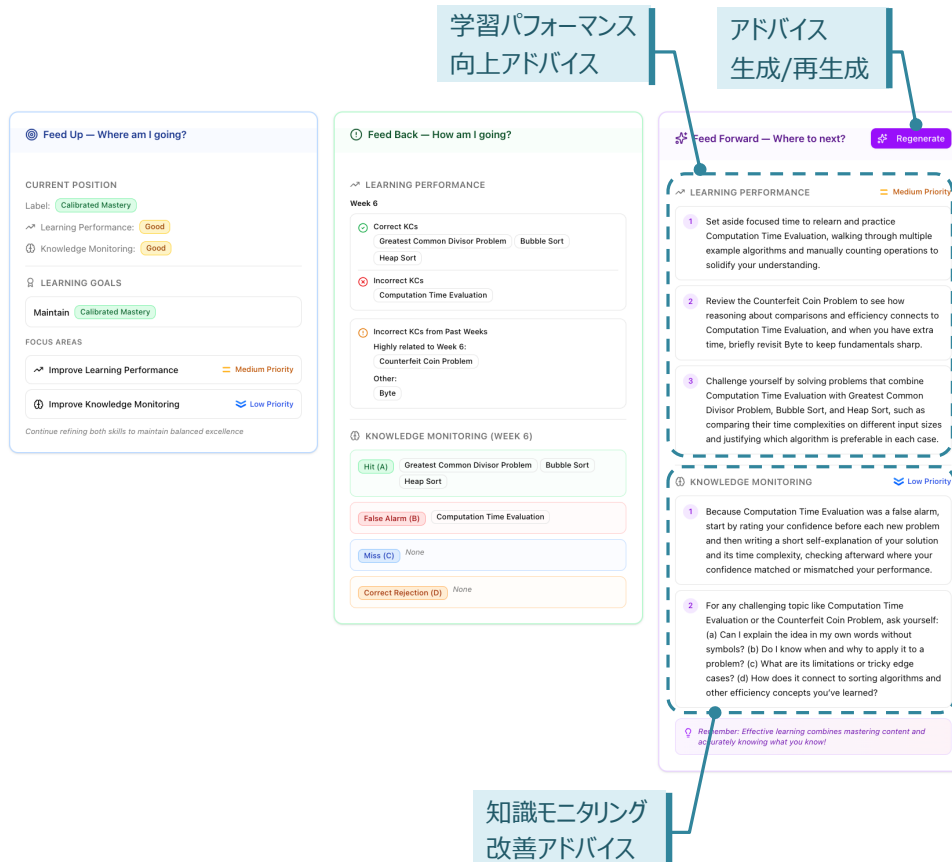
結果可視化部では、学習者が自身の学習状態を客観的に把握できるように、4 つのコンポーネントを配置する (図 3a)。

4.2.1 学習者基本情報

学習者 ID、累積解答数、現在の学習者類型 (5 パターンのいずれか)、テスト正答率、および知識モニタリング能力 (d') を表示する。



(a) 結果可視化部



(b) フィードバック部

図 3: 学習支援ダッシュボードの全体構成

4.2.2 学習者ポジションマップ

知識レベル（横軸）と知識モニタリング能力（縦軸）の二次元空間上に、学習者の現在位置を表示する散布図である。5つの学習者類型に対応する領域を色分けして示すことで、学習者が自身の位置づけと目指すべき方向を直感的に把握できるようにする。

4.2.3 週次推移グラフ

テスト正答率と知識モニタリング正確度の時系列変化を折れ線グラフで表示する。「週次」モードでは各週の値を、「累積」モードでは学期開始からの累積値を表示する。グラフ上の任意の週を選択することで、当該週の詳細データを表示できる。表示モードは「週次」（各週の値）と「累積」（学期開始からの累積値）を切り替え可能である。

4.2.4 問題別詳細テーブル

選択された週の各評価問題について、対応する知識概念、正誤結果、および学習者の自己評価（理解認識/非理解認識）を一覧表示する。これにより、知識モニタリングの4カテゴリ（ヒット、フォルスアラーム、ミス、正棄却）が問題単位で確認できる。

4.3 フィードバック部

フィードバック部では、第3章で設計した3段階フィードバックモデルに基づき、Feed Up・Feed Back・Feed Forwardの3セクションを配置する（図3b）。

4.3.1 Feed Up：目標提示

学習者の現在の類型、目指すべき目標状態、および改善の優先順位を提示する。図2に基づき、知識レベルとモニタリング能力のどちらを優先的に改善すべきかを明示する。

4.3.2 Feed Back：現状提示

学習者の現状を知識レベルとモニタリング能力の2観点から詳細に提示する。

知識レベルの観点では、当該週で正答した知識概念と誤答した知識概念をリスト表示する。さらに、知識グラフの関係性に基づき、過去に誤答した概念のうち当該週の内容と意味的に関連の高いものを抽出し、「関連する過去の誤答概念」として提示する。モニタリングの観点では、4カテゴリに分類された知識概念をそれぞれ提示する。

4.3.3 Feed Forward：アドバイス生成

学習者類型と現状に応じた個別最適化されたアドバイスを提示する。アドバイスは知識レベル向上とモニタリング改善の2種類から構成される。アドバイス生成にはLLMを使用する。学習者の類型、正答・誤答履歴、および類型別フィードバック方針に基づき、学習者の具体的な状況に応じたアドバイスを生成する。

4.4 実装詳細

ダッシュボードはWebアプリケーションとして実装した。フロントエンドはReact.jsを用いて実装し、バックエンドはPython (FastAPI)を用いた。第3章で述べたHGNNによる潜在的認識状態の推論機能、およびLLMベースのアドバイス

生成機能を統合した。学習者類型化における知識レベルおよび知識モニタリング能力の高低判定には、各クラスの中央値を閾値として使用した。アドバイス生成にはOpenAI GPT-5.2 APIを使用し、学習者の状況に応じたプロンプトエンジニアリングにより個別最適化を実現した。

5 評価実験

5.1 潜在的認識状態の推定精度 (RQ1)

5.1.1 実験設定

情報科学コースの5クラス、計684名の学生を対象としたデータセットを使用した。コースは14週間にわたり、情報理論、暗号、アルゴリズム、データ構造、機械学習などの基礎概念を扱う。各学生は各講義後に(1)自由記述式の学習リフレクション、(2)評価テストに従事した。表3にデータセットの統計情報を示す。#LPSは潜在的認識状態の数。

表3: データセットの統計情報

	FC _A	FC _B	FC _C	FC _D	FC _E
学生数	109	181	101	148	145
評価項目数	48	38	43	42	50
# \mathcal{E}_{S-K}	6,349	10,128	6,908	9,396	6,918
# LPS	2,187	3,002	1,727	2,634	3,807

全クラス共通：知識概念数 211, # $\mathcal{E}_{K-K} = 226$

比較手法として、Random Guesser (RG), GCN, GAT, Label Propagation (LP) を用いた。RGはリンクをランダムに予測するベースライン、GCNとGATは同種グラフ (\mathcal{E}_{S-K} のみ)上のグラフニューラルネットワーク、LPは異種グラフ (\mathcal{E}_{S-K} と \mathcal{E}_{K-K})上のラベル伝播法である。提案手法はHANベースの異種グラフニューラルネットワークにEINSを組み合わせたものである。評価指標にはROC-AUCを使用し、各クラスについて80%を訓練データ、20%をテストデータとして分割し、30回の試行の平均値を報告する。

5.1.2 結果

表4に各手法の性能比較を示す。提案手法は全クラスにおいて最高のAUC値を達成し、平均85.21%のAUCで潜在的認識状態を推定できた。

表4: 推定精度 (AUC%) の比較

手法	FC _A	FC _B	FC _C	FC _D	FC _E	平均
RG	50.00	50.00	50.00	50.00	50.00	50.00
GCN	77.91	76.97	79.18	79.53	76.45	78.01
GAT	75.44	74.71	75.81	77.18	73.89	75.41
LP	81.30	81.85	80.68	82.85	81.47	81.63
提案手法	86.52	85.17	84.81	87.23	82.33	85.21
w/o EINS	76.52	76.49	78.47	79.49	77.58	77.71

提案手法は同種グラフを用いるGCN (78.01%) およびGAT

(75.41%)を大幅に上回った。この結果は、知識概念間の意味的・階層的関係(\mathcal{E}_{K-K})を活用することの有効性を示している。また、異種グラフを用いる LP (81.63%)と比較しても 3.58 ポイント高い精度を達成した。EINS の導入により 7.50 ポイントの向上が確認され、学習者が明示的に報告した非理解認識を負例として活用することの重要性が示された。

5.1.3 考察

提案手法は 85.21%の AUC を達成し、ベースライン手法を上回った。特に EINS による 7.50 ポイントの向上は、学習者の明示的な自己報告を負例として活用することの有効性を示している。これは、「理解していない」という学習者の認識が、モデルの学習において重要な境界情報となることを示唆している。

また、同種グラフ手法 (GCN, GAT) と比較して異種グラフ手法 (LP, 提案手法) が高い性能を示したことから、知識概念間の意味的・階層的関係が潜在的認識状態の推定に有効であることが確認された。

5.2 ユーザ実験 (RQ2)

5.2.1 実験設定

システムの受容性を評価するため、大学生 36 名および教員 11 名、計 47 名を対象としたユーザ実験を実施した。

実験手順は以下の通りである：(1) システム使用 (ダッシュボードの操作と学習状況の確認)、(2) 評価アンケート (フィードバック品質 16 項目, SUS 10 項目, 自由記述 2 項目)。

5.2.2 結果

a) フィードバック品質

三段階フィードバックの評価結果を表 5 に示す。学生群 ($M=4.19$)、教員群 ($M=4.52$) ともに高い評価を得た。特に Feed Back (現状提示) の評価が高く、誤答した知識概念、関連する過去に誤答した知識概念の提示が有用と評価された。

表 5: フィードバック品質評価 (5 段階)

セクション	学生		教員	
	M	SD	M	SD
Feed Up	3.95	1.01	4.41	0.70
Feed Back	4.39	0.81	4.59	0.61
Feed Forward	4.06	0.93	4.51	0.61
全体	4.19	0.30	4.52	0.15

b) システムユーザビリティ

System Usability Scale (SUS) による評価結果を表 6 に示す。教員群は 70.68 点を達成した一方、学生群は 63.26 点と平均をやや下回った。

表 6: SUS スコア

	平均	SD
学生 ($n=36$)	63.26	14.37
教員 ($n=11$)	70.68	17.45

c) 定性評価

自由記述の分析から、最も多く言及されたのは「知識モニタリングの 4 カテゴリ可視化」(15 件)であり、「単なる正解・不正解だけでなく、自己評価との組み合わせで復習の優先順位をつけやすい」という意見が得られた。次いで、「過去に誤答した知識概念との関連性提示」(12 件)、「学習者類型とポジションマップ」(7 件)が有用と評価された。

改善要望としては、「UI/UX」に関する指摘 (10 件) が最も多く、文字サイズや週の切り替え操作に関する意見があった。また、「ヒット」「フォルスアラーム」等の専門用語の分かりにくさ (5 件) が指摘された。

5.2.3 考察

フィードバック品質は学生・教員ともに高い評価を得た ($M>4.0$)。特に知識モニタリングの 4 カテゴリ可視化と過去の誤答との関連性提示が有用と評価され、提案システムの核心機能が受容されたことを示している。

一方、学生群の SUS スコアは平均を下回った。教員群のスコアが高かった要因として、教員は学習分析ダッシュボードの使用経験があり、類似システムへの親和性が高かったことが考えられる。今後の改善点として、(1) UI の操作性改善、(2) 専門用語のより直感的な表現への変更が挙げられる。

6 まとめ

本研究では、学習者の知識モニタリング能力を評価するための異種グラフベースのフレームワークと、その結果に基づく適応型フィードバックシステムを提案した。LLM による知識概念抽出と異種グラフ構築により、学習者の潜在的認識状態をリンク予測問題として推定する手法を開発した。また、知識レベルと知識モニタリング能力の二軸による学習者類型化と、Feed Up・Feed Back・Feed Forward の三段階フィードバックを提供するダッシュボードを実装した。評価実験により提案手法の有効性を確認し、ユーザ実験により知識モニタリングの可視化と過去の誤答との関連性提示が有用であることを示した。今後は、他分野への適用、UI の改善、長期的な学習効果の検証に取り組む。

謝辞

本研究は、JST CREST (JPMJCR22D1)、JST AIP チャレンジ、ならびに JSPS 科研費 (JP22H00551, JP25K21360, JP24K16759) の支援を受けて実施された。

文献

- [1] Ismail Aliyu, A Kana, Salisu Aliyu, et al. Development of knowledge graph for university courses management. *International Journal of Education and Management Engineering*, Vol. 10, No. 2, p. 1, 2020.
- [2] Adam B Barrett, Zoltan Dienes, and Anil K Seth. Measures of metacognition on signal-detection theoretic models. *Psychological methods*, Vol. 18, No. 4, p. 535, 2013.
- [3] Tony Blackwood. Business undergraduates' knowledge monitoring accuracy: How much do they know about how

- much they know? *Teaching in Higher Education*, Vol. 18, No. 1, pp. 65–77, 2013.
- [4] Dejun Cai, Yuan Zhang, and Binta Dai. Learning path recommendation based on knowledge tracing model and reinforcement learning. In *2019 IEEE 5th international conference on computer and communications (ICCC)*, pp. 1881–1885. IEEE, 2019.
 - [5] Li Chen, Gen Li, Boxuan Ma, Cheng Tang, and Masanori Yamada. A three-step knowledge graph approach using llms in collaborative problem solving-based stem education. *International Association for Development of the Information Society*, 2024.
 - [6] Li Chen, Gen Li, Buxuan Ma, Cheng Tang, Masanori Yamada, and Atsushi Shimada. Classifying knowledge nodes and analyzing activation features: An integrated knowledge graph approach for collaborative problem-solving. In *2025 IEEE International Conference on Advanced Learning Technologies (ICALT)*, pp. 107–111, 2025.
 - [7] Penghe Chen, Yu Lu, Vincent W Zheng, Xiyang Chen, and Boda Yang. Knowedu: A system to construct knowledge graph for education. *Ieee Access*, Vol. 6, pp. 31553–31563, 2018.
 - [8] Jiajun Cui, Hong Qian, Bo Jiang, and Wei Zhang. Leveraging pedagogical theories to understand student learning process with graph-based reasonable knowledge tracing. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 502–513, 2024.
 - [9] Tung Do Viet and Konstantin Markov. Using large language models for bug localization and fixing. In *2023 12th International Conference on Awareness Science and Technology (iCAST)*, pp. 192–197. IEEE, 2023.
 - [10] Jibril Frej, Neel Shah, Marta Knezevic, Tanya Nazaretsky, and Tanja Käser. Finding paths for explainable mooc recommendation: A learner perspective. In *Proceedings of the 14th Learning Analytics and Knowledge Conference, LAK '24*, p. 426–437, New York, NY, USA, 2024. Association for Computing Machinery.
 - [11] John Hattie and Helen Timperley. The power of feedback. *Review of educational research*, Vol. 77, No. 1, pp. 81–112, 2007.
 - [12] Owen Henkel, Libby Hills, Adam Boxer, Bill Roberts, and Zach Levonian. Can large language models make the grade? an empirical study evaluating llms ability to mark short answer questions in k-12 education. In *Proceedings of the Eleventh ACM Conference on Learning@ Scale*, pp. 300–304, 2024.
 - [13] Stephen Hutt, Allison DePiro, Joann Wang, Sam Rhodes, Ryan S Baker, Grayson Hieb, Sheela Sethuraman, Jaclyn Ocumpaugh, and Caitlin Mills. Feedback on feedback: Comparing classic natural language processing and generative ai to evaluate peer feedback. In *Proceedings of the 14th Learning Analytics and Knowledge Conference*, pp. 55–65, 2024.
 - [14] Priscilla Lee and Soohyun Nam Liao. Targeting metacognition by incorporating student-reported confidence estimates on self-assessment quizzes. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education, SIGCSE '21*, p. 431–437, New York, NY, USA, 2021. Association for Computing Machinery.
 - [15] Gen Li, Li Chen, Cheng Tang, Daisuke Deguchi, Takayoshi Yamashita, and Atsushi Shimada. From reflections to motifs: A graph-based analysis of learners' knowledge construction. In *International Conference on Artificial Intelligence in Education*, pp. 299–307. Springer, 2025.
 - [16] Gen Li, Cheng Tang, Li Chen, Daisuke Deguchi, Takayoshi Yamashita, and Atsushi Shimada. Llm-driven ontology learning to augment student performance analysis in higher education. In *International Conference on Knowledge Science, Engineering and Management*, pp. 57–68. Springer, 2024.
 - [17] Chien-Chang Lin, Anna YQ Huang, and Owen HT Lu. Artificial intelligence in intelligent tutoring systems toward sustainable education: a systematic review. *Smart learning environments*, Vol. 10, No. 1, p. 41, 2023.
 - [18] Klaus Lingel, Jan Lenhart, and Wolfgang Schneider. Metacognition in mathematics: Do different metacognitive monitoring measures make a difference? *ZDM*, Vol. 51, No. 4, pp. 587–600, 2019.
 - [19] Hanxiao Liu, Wanli Ma, Yiming Yang, and Jaime Carbonell. Learning concept graphs from online educational data. *Journal of Artificial Intelligence Research*, Vol. 55, pp. 1059–1090, 2016.
 - [20] Brian Maniscalco and Hakwan Lau. A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and cognition*, Vol. 21, No. 1, pp. 422–430, 2012.
 - [21] Lars-Peter Meyer, Claus Stadler, Johannes Frey, Norman Radtke, Kurt Junghanns, Roy Meissner, Gordian Dziwis, Kirill Bulert, and Michael Martin. Llm-assisted knowledge graph engineering: Experiments with chatgpt. In *Working conference on Artificial Intelligence Development for a Resilient and Sustainable Tomorrow*, pp. 103–115. Springer Fachmedien Wiesbaden Wiesbaden, 2023.
 - [22] Lin Ni, Sijie Wang, Zeyu Zhang, Xiaoxuan Li, Xianda Zheng, Paul Denny, and Jiamou Liu. Enhancing student performance prediction on learnersourced questions with sgnn-llm synergy. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38, pp. 23232–23240, 2024.
 - [23] Zilong Pan, Lauren Biegley, Allen Taylor, and Hua Zheng. A systematic review of learning analytics: Incorporated instructional interventions on learning management systems. *Journal of Learning Analytics*, Vol. 11, No. 2, pp. 52–72, 2024.
 - [24] Stephanie Pieschl. Metacognitive calibration—an extended conceptualization and potential applications. *Metacognition and Learning*, Vol. 4, pp. 3–31, 2009.
 - [25] Tricia R Prokop. Calibration and academic performance in students of health sciences. *Health Professions Education*, Vol. 6, No. 4, pp. 564–573, 2020.
 - [26] T. Ullmann. Automated analysis of reflection in writing: Validating machine learning approaches. *International Journal of Artificial Intelligence in Education*, Vol. 29, pp. 217–257, 2019.
 - [27] Christopher A Was. Discrimination in measures of knowledge monitoring accuracy. *Advances in Cognitive Psychology*, Vol. 10, No. 3, p. 104, 2014.
 - [28] Kevin P Yancey, Geoffrey Laflair, Anthony Verardi, and Jill Burstein. Rating short l2 essays on the cefr scale with gpt-4. In *Proceedings of the 18th workshop on innovative use of NLP for building educational applications (BEA 2023)*, pp. 576–584, 2023.
 - [29] Junrui Zhang, Yun Mo, Changzhi Chen, and Xiaofeng He. Gkt-cd: Make cognitive diagnosis model enhanced by graph-based knowledge tracing. In *2021 International joint conference on neural networks (IJCNN)*, pp. 1–8. IEEE, 2021.
 - [30] Yifen Zhou and Xian Yu. Multi-graph spatial-temporal synchronous network for student performance prediction. *IEEE Access*, 2024.
 - [31] Barry J Zimmerman. Becoming a self-regulated learner: An overview. *Theory into practice*, Vol. 41, No. 2, pp. 64–70, 2002.