グラフ表現を用いた知識獲得予測による潜在知識構造の抽出と活用

Extracting and Exploiting Latent Knowledge Structure by Graph-based Knowledge Tracing

中川 大海*1 岩澤 有祐*1 松尾 豊*1 Hiromi Nakagawa Yusuke Iwasawa Yutaka Matsuo

> *¹東京大学工学系研究科 The University of Tokyo

Recent advancements in computer-assisted learning systems have increased research in the area of *knowledge tracing*, which estimates student proficiency based on their past performance. In this context, deep learning-based methods, such as Deep Knowledge Tracing (DKT), show remarkable performance; however, existing methods do not consider latent knowledge structure. This limits not only the prediction performance but also the interpretability and validity of models' prediction, which prevents the application to real educational environments. In this paper, we propose a graph-based knowledge tracing model, Graph Knowledge Tracing (GKT). Representing the knowledge structure as a graph, we model students' time-series mastery to each skill using Graph Neural Networks. We consider two problem settings, one is to exploit the pre-defined graph structure and the other is to learn the implicit graph structure from data, and provided two models to deal with them. Using two open datasets, we empirically validated that our method shows higher prediction performance and more interpretable and valid prediction compared to the previous methods. These results show the potential of our proposed method to enhance the performance and the application to real educational environments of knowledge tracing, which could help improve the learning experience of students in more diverse environments.

1. はじめに

近年,教育と情報技術の融合が進む中で,学習体験の効率化に 関する研究が注目を集めている.中でも知識獲得予測(knowledge tracing) [Corbett 94]の研究は,過去の生徒の学習行動 を元に個々の生徒の習熟度を予測するものであり,個人に最適 化された学習内容を提供する上で核となる技術である.歴史的 に様々な分析手法が考案されてきたが,特に深層学習の活用に より,人手での複雑な特徴量設計なしに従来の手法より高い精 度での予測が可能になった [Piech 15].

一方,深層学習を用いた知識獲得予測に関する既存の研究 では,いずれも知識の構造を十分に考慮したモデルが設計さ れていない.本研究では、知識獲得予測の文脈に倣い、「知識」 をある学問を抽象的な個別の概念に分割したもの、「知識構造」 を各知識が相互の影響関係のもとに一つの体系をなしている様 態として定義する.知識構造の設計は学習や指導において重要 な役割を担い、知識獲得の効率にも影響を与えることが知られ ており,こうした知識構造を明示的に考慮しない既存研究は, 二つの課題を生んでいる.一つは,機械学習のモデル設計にお いて重要な,帰納バイアスと呼ばれる,データの性質に対する 知識に基づく最適なモデルの設計ができていないため、学習の 効率や未知データへの汎化が妨げられている可能性が高いこと である.もう一つは、知識ごとに異なる習熟の変化や、知識間 の関係性などが学習に明示的に反映されないため、モデルが知 識ごとの習熟度合いをどのように判断しているかという予測の 解釈性が低下したり、 カリキュラムを設計した側が意図しない 挙動をモデルが学習し予測の妥当性が低下するなど,実際の教 育現場における知識獲得予測の実用を妨げていることである.

本研究では、知識間の潜在構造をグラフ表現を用いて定式化 し、近年発展が進む、深層学習を用いてグラフを扱う Graph Neural Network を拡張したモデルによって、これらの問題の 解決を図る.実験では,提案手法が既存手法に比べて,高精度 かつ解釈性と妥当性の高い予測を行えることを,オープンデー タを用いて実証的に検証し,またデータから学習されたグラフ 構造を分析することで,効率的な知識構造の設計に関して考 察する.また,実験結果を踏まえ,グラフの活用を含めた今後 の知識獲得予測の研究が考慮すべき方向性を整理することで, 知識獲得予測研究の発展に向けて有益な示唆をもたらす.

2. 関連研究

2.1 Deep Knowledge Tracing

Deep Knowledge Tracing [Piech 15] (DKT) は RNN を利 用して知識獲得予測を行う手法である.モデルの入力は,ある 生徒のある時刻における問題と正誤の組み合わせを符号化し た one-hot ベクトルであり,問題の数を M とすれば,長さは 2M となる.訓練時に用いられる目的関数は,モデルの予測 の,生徒の解答行動の観測系列に対する負の対数尤度である.

2.2 Graph Neural Network

グラフは、データを物体(ノード)とそれらの関係性(エッジ) から表現するデータの表現方法であり、Graph Neural Network (GNN)は深層学習を用いてグラフを扱う手法の総称である. 画像認識において飛躍的な成果を発揮した CNN は、画像デー タに特有の局所性や位置不変性などを考慮したモデル構造に よってパラメータ数を大幅に節約しながら高い表現力を獲得し たことが知られているが、GNN は同様の機構を画像以外のよ り一般的な構造のデータに対して拡張したものと見なすことが できる [Battaglia 18].

3. 提案手法

本研究では,知識構造をグラフ表現を用いて定義し,GNNを 拡張した機構によって知識獲得予測を行う Graph Knowledge Tracing (GKT)のモデルを提案する.

連絡先:中川大海,東京大学工学系研究科技術経営戦略学専攻, nakagawa@weblab.t.u-tokyo.ac.jp

3.1 GKT の定式化

1. で述べた定義に倣うと、知識構造は、各知識をノード、知 識間の関係性をエッジとした重み付き有向グラフと見なすこと が出来る. GKT では、ノード v_i に対応する知識 i に対する 生徒固有の習熟状態を表す隠れ状態 \mathbf{h}_i を定義し、各時刻で生 徒の解答正誤が観測される際には、解答ノード v_i に加え、隣 接ノード $v_{j\in\mathcal{N}_i}$ の隠れ状態も更新する. ここで、 \mathcal{N}_i はノード i に隣接するノードの集合を表す. モデルは、DKT 同様、生 徒の時刻 T までの解答正誤系列 $\mathbf{x}^{t\leq T}$ を入力として、次の時 刻 T+1 の解答正誤予測 $y^{t=T}$ を出力し、DKT と同様に観測 系列に対する負の対数尤度を最小化するように学習する. 以下 ではモデルの処理を三つの段階に分けて説明する.

まず,生徒の時刻 t における解答 \mathbf{x}^t が観測されたときに, 隠れ状態の更新に利用する特徴の集約を行う. 解答ノード $i = \{i_0, ..., i_{a_t}\}$ と隣接ノード $j \in \mathcal{N}_i$ に対応する隠れ状態を抽出 し,該当スキルのインデックスおよび解答正誤の埋め込み表現 と連結する.

$$\begin{split} \mathbf{x}_{emb}^{t} &= \mathbf{x}^{t} \, \mathbf{E}_{\mathbf{x}} \\ \mathbf{h}_{i}^{'t} &= \left[\mathbf{h}_{i}^{t}, \mathbf{x}_{emb}^{t}\right] \\ \mathbf{h}_{j}^{'t} &= \left[\mathbf{h}_{j}^{t}, \mathbf{E}_{q}(j)\right] \end{split}$$

ここで、 a_t は、時刻 t において同時に解答したスキルの数を 表しており、 \mathbf{x}^t は、解答されたスキルのインデックスと解答 正誤を示す長さ 2M の one-hot ベクトルであり、 $\mathbf{E}_{\mathbf{q}} \in \mathbb{R}^{M \times e}$ はスキルの埋め込み行列であり、 $\mathbf{E}_{\mathbf{x}} \in \mathbb{R}^{2M \times e}$ はスキルに対 する解答正誤の埋め込み行列であり、e は埋め込みの次元数で ある.

次に,集約された特徴を元に,生徒の隠れ状態を更新する.

$$\mathbf{m}_{k}^{t+1} = \begin{cases} f_{self}(\mathbf{h}_{i}^{'t}) & (k=i) \\ f_{neighbor}(\mathbf{h}_{i}^{'t}, \mathbf{h}_{j}^{'t}, \theta) & (k\neq i) \end{cases}$$
(1)
$$\tilde{\mathbf{h}}^{t+1} = \tanh\left(\frac{1}{a_{t}}\sum_{i_{a_{t}}\in i}\mathcal{G}_{ea}(\mathbf{m}_{i_{a_{t}}}^{t+1})\right)$$

$$\mathbf{h}^{t+1} = \mathcal{G}_{gru}(\tilde{\mathbf{h}}^{t+1}, \mathbf{h}^{t})$$

ここで、 $f_{self}, f_{neighbor}$ は任意の関数であり、本研究では f_{self} には二層の多層パーセプトロンを用いる。 θ は任意のパラメータであり、 \mathcal{G}_{ea} は [Zhang 16] で用いられた Erase-Add ゲートからバイアス項を除いたものであり、 \mathcal{G}_{gru} は GRU [Cho 14] からバイアス項を除いたものである。

最後に,次の時刻における解答正誤予測 y^tを出力する.

$$\mathbf{y}^t = \mathbf{W}_{\mathbf{02}} \mathbf{W}_{\mathbf{01}} \mathbf{h}^{\mathbf{t+1}} + \mathbf{b}_{\mathbf{0}}^T$$

ここで、 W_{o1}, W_{o2} はノード共通で用いられる重み行列であり、 b_o はノードごとに独立した値を持つバイアス項である.

3.2 GKT の性質

GKT の性質を既存手法との比較から考察する.まず,DKT [Piech 15] は、単純な RNN を元にしたモデルであり、隠れ状 態をスキルごとに区別して定義できないため、スキル間の関 係性を学習に明示的に反映したり、長期系列において過去の回 答の習熟情報を維持することが困難であった.また、この課題 を解決するためにメモリ行列を導入した DKVMN [Zhang 16] も、メモリの各スロットは入力空間のスキルとは明示的に対応 しない低次元表現であり、スキルとの関係も単純な内積でしか 表現できなかった.一方,GKTは、知識構造をグラフとして 明示的に学習に反映することが可能な上、スキルごとの隠れ状 態を区別して定義した上で、スキル間の関係性を複数のエッジ や多層のニューラルネットワークによる非線形変換を用いて複 雑に定義可能である.このような性質から、従来の手法に比べ てより高い精度での予測が可能になり、また予測の解釈性や妥 当性が向上することが期待される.

3.3 GKT におけるグラフの定義

GKT は知識構造をグラフとして定式化して知識獲得予測を 行うモデルだが、モデルの学習時には、既知のグラフを明示的 (*explicit*)に仮定することも、グラフを明示的に仮定せず(*implicit*)にモデルの学習の過程で同時に学習することも可能で ある.本研究では、前者を Explicit Graph Knowledge Tracing (EGKT),後者を Implicit Graph Knowledge Tracing (IGKT)として区別し、以下でそれぞれの実装を説明する.

3.3.1 Explicit Graph Knowledge Tracing (EGKT)

EGKT では、隣接ノードの隠れ状態の更新の際に、集約した特徴をエッジの向きごとに別々の関数を用いて変換した後、仮定した既知の隣接行列 **A** の値を元に加重和を計算する.このとき、式1における $\mathbf{m}_{k\neq i}^{t+1}$ は以下の式で定義される.

$$\mathbf{m}_{k\neq i}^{t+1} \quad = \quad \mathbf{A}_{i,k}^T \odot f_o([\mathbf{h'}_i^t, \mathbf{h'}_k^t]) + \mathbf{A}_{k,i} \odot f_i([\mathbf{h'}_i^t, \mathbf{h'}_k^t])$$

ここで、 f_i, f_o は任意の関数であり、本研究では二層の多層パー セプトロンを用いる。EGKT は、任意の方法で定義した隣接 行列を利用出来るが、本研究では学習済みの DKT のモデルか ら [Piech 15] の手法に基づいて抽出した「DKT グラフ」、遷 移確率を元に定義した「遷移グラフ」、自己以外の全ノードと 等しい重みで接続する「全結合グラフ」を定義して検証を行う。 **3.3.2 Implicit Graph Knowledge Tracing (IGKT)**

IGKT は、知識獲得の予測と同時にグラフ構造をデータか ら学習するモデルであり、本研究では、グラフ構造を学習する 方法として三つのモデルを提案する.一つ目は, EGKT と同 様のモデル構造において隣接行列 A をパラメータと見なして 最適化する, Parametric Adjacency Matrix (PAM) である. 二つ目は、複数ヘッドの注意機構 [Vaswani 17] を用いて、各 ノードの特徴を元に, K 個のヘッドを用いて 2 ノード間の K 通りのエッジの重みを計算することでグラフ構造を推定する, Multi-Head Attention (MHA) である.本研究では生徒共通 の静的なグラフ構造を学習するため、注意係数の計算にはスキ ルのインデックスや解答正誤の埋め込みを入力として用いる. 三つ目は、各エッジの種類を定義する離散のカテゴリを表す潜 在変数 z の存在を定義し, VAE を用いて z を推定することで グラフ構造を推定する, Variational AutoEncoder (VAE) で ある. [Kipf 18] に着想を得たモデルであり,予測の負の対数 尤度に加え,潜在変数 \mathbf{z} の予測分布 $q(\mathbf{z}|\mathbf{x})$ と事前分布 $p(\mathbf{z})$ の カルバック・ライブラー情報量を最小化するように学習させる ことで、グラフ構造に対するスパース性などを仮定することが できる.

4. 実験

4.1 実験設定

実験には、オンライン教育サービスの ASSISTments^{*1} に おける生徒の数学の問題解答ログの「skill_builder」^{*2} から

^{*1} https://www.assIstments.org/

^{*2} https://sites.google.com/site/
assistmentsdata/home/assistment-2009-2010-data/
skill-builder-data-2009-2010

表 1: 各データセットの統計量

データセット	生徒数	スキルタグ数	ログ数
ASSISTments	1,000	101	62,955
KDDCup	1,000	211	98,200

表 2: 各手法のモデルの予測精度の比較

手法		AUC		
		ASSISTments	KDDCup	
ベースライン	DKT	0.709	0.751	
	DKVMN	0.710	0.753	
EGKT	DKT グラフ	0.723	0.764	
	遷移グラフ	0.721	0.769	
	全結合グラフ	0.722	0.762	
IGKT	PAM	$\bar{0}.\bar{7}1\bar{9}$	0.762	
	MHA	0.723	0.766	
	VAE	0.722	0.769	

なるデータセット(以下, ASSISTments)と, KDDCupの Educational Data Mining Challenge で使用された「Bridge to Algebra 2006-2007」[Stamper 10]のデータセット(以下, KDDCup)を用いる. 各データセットでは各問題に人間の専 門家が定義した「スキルタグ」が紐付けられており,本研究で はこの各スキルタグを「知識」の定義とする. 各データセット より,同時解答を意味する重複ログを一つにまとめ,名前が割 り当てられておりダミーでないスキルの解答ログを抽出し,最 低 10回以上解答されているスキルの解答ログを抽出した後, 各ユーザ 100回目までのログを抽出し,ランダムに 1000 ユー ザを抽出した. 以上の条件から抽出されたデータセットの統計 量を表 1 に示す.

4.2 実験1:モデルの予測の精度の比較

まず,提案手法が既存手法より高い精度で生徒の知識獲得を 予測できるかを検証するために,DKT [Piech 15] とDKVMN [Zhang 16] をベースラインとして予測のAUCを比較し,表 2 に示した.各データセットで最高精度のものを太字で記載した.

表より,いずれのデータセットにおいても,提案手法が最も 高い精度を示した.このことから,知識の潜在的な構造をグラ フを用いて定式化し,GNNを拡張してグラフ表現上で生徒の 知識獲得をモデリングをする提案手法が,知識獲得の予測にお いて有効であることが示された.

4.3 実験2:モデルの予測の解釈性と妥当性の比較

次に,提案手法が既存の深層学習を用いた手法と比較して, 解釈性と妥当性の高い予測が行えるかを検証した.まず,ある 生徒の時刻 T までの解答ログを抽出して学習済みモデルに入 力し,出力からバイアス項を除いて正規化した値を蓄積する ことで,学習済みのモデルによる,生徒の各スキルに対する時 系列の習熟度予測を分析可能な行列を作成した.この行列を 基に,学習済みモデルの予測について,各時刻において関連す る知識の習熟度のみ更新しているかによって解釈性を評価し, 仮定したグラフ構造と連動して生徒の習熟を予測しているかに よって妥当性を評価した.

ASSISTments における DKT と GKT の出力から作成した 行列から,一部のスキル集合に関する値のみを抽出したものを 図1に示した.各時刻において正答・誤答したスキルに対応す る要素に,それぞれ「〇」と「×」を記載した.各項目におけ るヒートマップは、各行が各スキルを、各列が左から右に進む 時刻を表しており、各要素の色は習熟度の変化を可視化したも ので、赤色に近づくほど習熟度合いが減少し、緑色に近づくほ ど習熟度合いが増加したと見なすことが出来る.

まず,生徒が頻繁に解答したスキル集合に関する値のみ抽出 して可視化したものを図 1a に示した.図より,DKT では解 答の観測と予測される習熟度の間の関係性が小さく,全ての値 が常に値が小さく変動している一方,GKT では各スキルに対 する習熟度がそのスキルの解答が観測された時刻で明確に更新 されていることがわかる.この結果より,提案手法が既存手法 に比べ解釈性の高い予測を行えていることが検証された.

次に、グラフ上で隣接するスキル集合に関する値のみ抽出 して可視化したものを図 1b に示した.図の右側は、ネット ワーク図における対応する部分を抽出して可視化したものであ る.図より、GKT において.スキル 29(Effect of Changing Dimensions of a Shape Proportionally)の解答が観測されて いないものの習熟度予測が大きく変化した 2 箇所において、グ ラフ上で接続するスキル 4(Addition and Subtraction Positive Decimals)の正答が観測されており、モデルがグラフ上の接続 関係に基づいた予測を行っている様子が確認された.一方,該 当のグラフを抽出した元のDKTの出力では、そのような性質 は確認できなかった.この結果より、提案手法が既存手法に比 ベ妥当性の高い予測を行えていることが検証された.

4.4 実験3:学習された知識構造の分析

最後に、学習後の IGKT のモデルから学習されたグラフ構造 を抽出してネットワーク図として可視化し、分析した. IGKT は自身が学習するグラフ構造を用いて生徒の知識獲得を予測 するため、高い精度で予測が可能な IGKT のモデルから抽出 したグラフからは、効率的な知識構造の設計について有益な 知見を得られる可能性がある. 各手法のモデルから作成した ネットワーク図を図2に示す. 各スキルが解答ログ内で解答 される平均順序を計算し、値が大きくなるに連れて青色から 黄色,赤色と変化するように対応するノードを着色した.比 較のために DKT から抽出したグラフ [Piech 15] は, 色の近 いノード間に多くエッジが張られており、クラスタを形成して いる様子が確認できた. DKT は全てのスキルに関する習熟状 態を同一の隠れ状態で扱うなど、長期の系列のモデリングを 行いにくいモデル機構であるため、解答順序の近いノード間 の依存関係を学習しやすかった可能性が考えられる. PAM か ら作成したグラフは、他の MHA や VAE と比較すると比較的 DKT に近い, 色の近いノードを主に結びながら, クラスタを 形成している様子が確認できた.局所的な接続に注目した図 3a からは、12(Area Rectangle) や 20(Complementary and Supplementary Angles) などの図形に関するスキル同士の接 続が確認された. MHA から作成したグラフは,特定のノード から多くのエッジが流出している様子が見て取れる.他の手法 が学習しなかった特殊な依存関係を学習している可能性が高い が, 偏った予測をする可能性があり, スキルごとの予測精度の 影響を検証する必要がある. VAE から作成したグラフは,他の グラフに比べ,特定のノードにエッジが集中せずに多くのノー ドが繋がりあった密なグラフとなった.人間には直感的には解 釈しにくいものの,一部,図 3b に示したように,46(Mode), 38(Histogram as Table or Graph), 78(Scatter Plot) といっ た統計に関連するスキルが直列的に接続している部分などもあ り、従来の手法では発見しにくいが、意味的な関係性が深いと 考えられるエッジも確認された.



図 1: 一部のスキル集合に関する生徒の習熟予測の変遷の可視化



図 2: 学習済みモデルから作成したネットワーク図



図 3: IGKT から作成したネットワーク図の局所的な可視化

5. 結論

本研究では、既存の深層学習を用いた知識獲得予測の研究 が、知識構造を明示的に考慮したモデル構造を用いていない ことにより生じている、二つの課題に着目し、知識構造をグ ラフ表現を用いて定式化し、GNN を拡張した機構で知識獲得 予測を行う GKT のモデルを提案した.実験結果から、提案 手法は既存手法に比べて高精度かつ解釈性・妥当性の高い予測 が行えることが示され、知識獲得予測の教育現場での活用を 促進する上で有益なものであることが確認された.提案手法 は特殊なデータを必要としない汎用性の高い拡張であるため, 多様なデータセットへの適用が期待でき、さらに知識獲得予測 や GNN の近年の研究における多様な拡張と組み合わせること で、より予測の精度や解釈性などを高めることが期待できる. また、提案手法によってデータから学習した知識構造は、従来 の人間の設計とは異なる構造を有することから、良い知識構造 の定義は何かという根本的な問題を問いかけるものでもあり, タグ付け自体をデータから学習する [Nakagawa 18] の研究と 組み合わせることで、より有益な示唆を得られる可能性があ る.本研究では数学に関する二つのデータセットを用いたが. より網羅的なデータセットを用いて検証することで,手法の汎 用性の検証や,効率的な知識構造に関する新たな知見の獲得が 期待でき,今後の研究課題である.本研究が,今後の知識獲得 予測研究が考慮すべき重要な方向性を示し,知識獲得予測の研 究や実用を加速させ,多様な環境における生徒の学習効率の向 上に寄与するものとなると信じている.

参考文献

- [Battaglia 18] Battaglia Peter W, Hamrick Jessica B, Bapst Victor, Sanchez-Gonzalez Alvaro, Zambaldi Vinicius, Malinowski Mateusz, Tacchetti Andrea, Raposo David, Santoro Adam, Faulkner Ryan and others. Relational inductive biases, deep learning, and graph networks. arXiv preprint arXiv:1806.01261 (2018).
- [Cho 14] Kyunghyun Cho, Bart Van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078 (2014).
- [Corbett 94] Albert T Corbett and John R Anderson: Knowledge tracing: Modeling the acquisition of procedural knowledge, User modeling and user-adapted interaction 4, 4 (1994), 253278.
- [Hochreiter 97] Sepp Hochreiter and Jurgen Schmidhuber: Long short-term memory, Neural computation 9, 8 (1997), 17351780.
- [Kipf 18] Thomas Kipf, Ethan Fetaya, Kuan-Chieh Wang, Max Welling and Richard Zemel. Neural relational inference for interacting systems. arXiv preprint arXiv:1802.04687 (2018).
- [Nakagawa 18] Hiromi Nakagawa, Yusuke Iwasawa and Yutaka Matsuo. End-to-End Deep Knowledge Tracing by Learning Binary Question Embedding, In 2018 IEEE International Conference on Data Mining Workshops (ICDMW). (2018).
- [Piech 15] Chris Piech, Jonathan Bassen, Jonathan Huang, Surya Ganguli, Mehran Sahami, Leonidas J Guibas, and Jascha Sohl-Dickstein: Deep Knowledge Tracing, In Advances in Neural Information Processing Systems. 505513 (2015).
- [Stamper 10] J. Stamper, A. Niculescu-Mizil, S. Ritter, G.J. Gordon, and K.R Koedinger. Bridge to Algebra 2006-2007, Development data set from KDD Cup 2010 Educational Data Mining Challenge, http://pslcdatashop.web.cmu.edu/KDDCup/downloads.jsp. (2010).
- [Vaswani 17] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser and Illia Polosukhin. Attention is all you need, In Advances in Neural Information Processing Systems. 5998– 6008 (2017).
- [Zhang 16] Jiani Zhang, Xingjian Shi, Irwin King and Dit-Yan Yeung. Dynamic Key-Value Memory Network for Knowledge Tracing, arXiv preprint arXiv:1611.08108 (2016).