

電気通信大学情報理工学域
I類 (情報系) 情報数理工学プログラム卒業論文

グラフニューラルネットワークを用いた ドロップアウト予測

2025年4月10日

情報数理工学プログラム

学籍番号 2110701

和田 爽花

指導教員 植野 真臣

令和6年度 情報数理工学プログラム卒業論文概要

令和2年度 入学	学籍番号 2110701
指導教員 植野 真臣	氏名 和田 爽花
題目	グラフニューラルネットワークを用いた ドロップアウト予測

概要

近年、CBTやLMSを通じて収集された教育ビッグデータの活用方法について、学生のドロップアウト予測が注目されている。ドロップアウト予測では、リスクのある生徒を判別し、その学生に適したサポートを行うため、予測の精度、特徴量の解釈性の双方が求められる。

本研究では、入学生学力調査の出題項目を入力特徴量として、階層的な構造を持つグラフニューラルネットワークを用いたドロップアウト予測手法を提案する。提案手法は、項目を頂点とする分野グラフ、分野を頂点とする教科グラフの2種類のグラフ構造を持つ。分野グラフはグラフプーリングを用いて埋め込みに変換され、教科グラフの頂点となる。このような階層構造により入力である出題項目を分類することで、分野ごとの分析、分野内の項目ごとの分析といった粒度の異なる分析が可能となる。

評価実験では、実データを用いて比較を行い、従来手法の予測精度を上回った。また、学習を通して最適化された隣接行列、グラフプーリングの重みを分析し、ドロップアウト予測に影響を及ぼした分野、分野同士の関連性を分析した。

目次

1	まえがき	4
2	グラフニューラルネットワーク	6
2.1	グラフ畳み込み	6
2.1.1	Graph convolution networks	7
2.1.2	Graph attention networks	7
2.2	グラフプーリングと Diffpool	8
3	提案手法	9
3.1	分野ネットワーク	10
3.2	教科ネットワーク	12
4	評価実験	14
4.1	データの概要	14
4.2	予測精度の評価実験	15
4.3	クラスタ割り当て行列の分析	17
4.4	隣接行列の分析	18
5	まとめ	19

1 まえがき

近年、Computer Based Testing (CBT) や Learning Management System(LMS) などで収集された教育ビッグデータの活用法が課題となっている。その中でも、ドロップアウト予測が注目されている。ドロップアウト予測とは、教育機関において、学生が単位を取得できない（または途中で離脱する）可能性をその学生のデータから予測することである。学生がドロップアウトをする要因は多岐にわたると考えられ [26, 27], その予測には多様なデータが用いられる。例えば、学生の学力、性格や環境などさまざまなデータからドロップアウト予測が行われている [2, 3, 4, 6, 12, 13, 14, 15, 16, 22].

特に、大学一年生におけるドロップアウトはその後の学修成果に強く影響する [23]. また、ドロップアウトした学生は、学生の成績のみならず精神状態、進路に影響を及ぼす [24]. そのため、ドロップアウトのリスクがある学生を早期に予測し、教育機関が適切にサポートすることが重要である [1, 25].

近年、機械分野の発展により、さまざまなドロップアウト予測手法が提案されている。[2, 3, 4, 6, 12, 13, 14, 15, 16, 22]. 特に、ドロップアウト予測の精度向上のための特徴量エンジニアリングに焦点を当てた手法が数多く研究されている [2, 3, 4, 6].

例えば、sivakumar ら (2016) は、アンケート調査から得られた学生の居住地、家庭環境、成績などの特徴量から、ドロップアウトと相関が高いが、特徴量間の相関は低い特徴量を選ぶ相関ベースの特徴量選択手法を提案した [4]. sivakumar ら (2016) の手法では選択した特徴量から決定木を用いてドロップアウト予測を行う。決定木では、経路をたどることで、ドロップアウト予測に寄与した特徴量とその分岐条件を解釈できるため、学生へのサポートへ活用できる。

さらに、ドロップアウトの予測精度を向上させるために、勾配ブースティングを用いた手法が数多く提案されている [6, 13, 14, ?]. その中で最も予測精度が高い手法として、Elbouknify ら (2024) は、XGBoost, LightGBM を用いたマルチタスクモデルでドロップアウト予測手法を提案している [14]. この手法では、学生の成功、失敗、中退の分類と脱落リスクの有無を同時に予測し、マルチタスクモデルの有効性を示している。その結果、マルチタスクモデルはシングルタスクモデルと比較して精度が向上しただけでなく、学習時間の短縮も達成した。さらに、この手法では、予測精度を高精度に維持しつつ、SHapley Additive exPlanations(SHAP)[34] 値からドロップアウト予測に対する特徴量ごとの寄与を定量化できる。

しかし、この手法では、強い相関をもつ特徴量間で SHAP 値が分散し、どの特徴量が予測に寄与したのかを解釈することが困難となる。また、SHAP 値はドロップアウト予測に対する特徴量間の関連性も解釈ができない。この問題の解決方法の一つとして、近年では、グラフニューラルネットワーク (Graph Neural Networks; GNN) が注目されている [28]. GNN は、グラフ構造を持つデータに対して適用できる深層学習手法であり、ノード、エッジ、およびグラフ全体の表現を効果的に学習することで、グラフの要素間の関係性をモデル化し、予測を行う。これにより、深層学習による高い予測精度を維持しつつ、特徴量とその関係について、グラフ構造から解釈できる。ゆえに、教育分野の Knowledge Tracing [35, 10, 11] においても、GNN を用いた手法が最も高精度の予測精度と解釈性を持つことで知られている [5]. 他にも、GNN は、分子構造 [7], ソーシャルネットワーク [8] やナレッジグラフ [9] など

さまざまな分野で活用されている。

GNNで近年注目されている手法として、グラフプーリングが知られている [32]。グラフプーリングは、類似した特徴量を持つノードを縮約したグラフ構造を学習する。グラフプーリングは特徴量の数を削減できるため、モデルの複雑さを抑えられ過学習を防ぐことができる。また、グラフのノード数とエッジが減少するため、モデルの学習時間が軽減される。これらの最先端手法として、Ying らはグラフ畳み込みで得られたグラフ構造を階層的にクラスタリングするプーリング手法として、diffpool[32] を提案した。diffpoolでは、ノードの特徴量と隣接行列をクラスタリング後の新しいサブグラフに変換する。diffpoolではこのサブグラフを動的に学習することで、予測精度が向上し、特徴量の階層構造も解釈できる。本研究では、diffpoolを用いた階層的なグラフ構造を持つドロップアウト予測手法を提案する。

提案手法の入力は、学生が受験した入学生学力調査における出題項目の正誤データである。入学生学力調査とは、電気通信大学において行われる、大学に入学した直後に学生が受験する学力調査である。数学、物理の試験が行われる。このような入学生学力調査のデータを用いてドロップアウト予測を行うことで、大学の授業が始まる前に、受講予定の科目のうちドロップアウトのリスクのある科目を知ることができる。これにより、ドロップアウト予測で重視される早期の予測を可能にする。提案手法は分野グラフ、教科グラフという2種類のグラフ構造を持つ。分野グラフは項目に対応するノードを持ち、教科グラフは分野に対応するノードを持つ。このとき、分野グラフをdiffpoolを用いて縮約し、埋め込みとしたものを教科グラフのノードとする。これにより、分野グラフと教科グラフの階層構造を構成する。分野は、入力である入学生学力調査の全ての出題項目に予め定められており、全ての項目が特定の分野に属する。分野を集めたものが教科である。例えば、分野「数と式」、「図形と計量」は教科「数学」に属する。分野グラフは、項目同士の関係を表現したグラフである。同様に、教科グラフは分野同士の関係を表現したグラフである。これらのグラフを分野ネットワーク、教科ネットワークにより処理し、予測を行う。分野ネットワークは分野グラフをdiffpoolを用いて縮約し、分野の埋め込みベクトルを出力する。教科ネットワークは、教科ネットワークは全ての分野ネットワークから分野埋め込みを受け取り、それらを特徴ベクトルとして教科グラフを生成する。さらに、教科グラフをdiffpoolを用いて縮約し、ドロップアウト予測スコアを出力する。また、分野グラフと教科グラフの隣接行列は学習とともに最適化される。

提案手法では、前述のGNNを使用したモデルにより、特徴量間の相互関係を効果的に活用した従来手法より精度の高い予測を行うことが可能となる。さらに、教科グラフ、分野グラフの隣接行列から、分野間の関連性、特徴量間の関連性を分析できる。また、diffpoolにて使用されるクラスタ割り当て行列の重みに注目することで、予測結果に寄与した分野を明示化できる。分野グラフ、教科グラフの2種類の粒度の異なるグラフを使用し、階層的な構造を構成することで、分野間の包括的な分析から詳細な特徴量同士の分析まで幅広い解釈が可能となる。このように、提案手法では、予測精度に加え、既存手法の課題であった予測結果に影響を及ぼす特徴量の分析、相互に関連する特徴量の可視化などドロップアウトの要因の分析を可能にする。

評価実験では、既存手法 (LightGBM, XGBoost, DNN) と提案手法を用いてドロップアウト予測を行い、既存手法と比較して予測精度が向上したことを示す。さらに、学習を通して得られたグラフ構造を隣接行列によって分析し、分野同士の関連性を考察する。加えて、diffpool 層のクラスタ割り当て行列の可視化を行い、ドロップアウトに影響を及ぼした分野を考察する。

2 グラフニューラルネットワーク

本章では、本研究の基礎となるグラフニューラルネットワークを紹介する。グラフ $G = (V, E)$ はノードの集合 V 、エッジの集合 E を構成要素にもつデータ構造である。エッジはノード間の接続を表す。グラフはソーシャルネットワーク、分子構造の分析など実社会における様々な構造の表現が可能であり、幅広い分野で使用されている。グラフニューラルネットワーク (Graph Neural Networks; GNN)[28] はグラフを扱うために考案されたニューラルネットワークである。GNN は高精度な予測能力を備えるだけでなく、グラフの構造的特性を活用した多角的な分析を可能にする点で注目されており、幅広い研究において高い評価を得ている [5, 7, 8, 9]。GNN の入力であるグラフは、初期状態でノードごとに独自の情報である特徴ベクトルを持っている。GNN を通して、全てのノードの特徴ベクトルを表す特徴行列と、エッジの接続状態を表す隣接行列を更新する。代表的な GNN が行う処理として、グラフ畳み込み、グラフプーリングが挙げられる。本章では、これらの主要な GNN アーキテクチャについて、その基本的な理論と特徴を概説する。

2.1 グラフ畳み込み

深層学習で画像認識などに使用される畳み込みは、隣接する画素に対して計算処理を適用し、空間的な特徴を抽出する手法である [19]。この手法は主に画像認識や物体検出といったタスクにおいて、画像データの細部の特徴を段階的に抽出し、データ全体の複雑なパターンを効果的に捉えることを可能にしている。一方、グラフ畳み込みは、隣接するノードに対して計算処理を適用し、グラフの構造的な特徴を抽出することを目的とする。GNN の入力であるグラフは、初期状態でノードごとに独自の情報である特徴ベクトルを持っている。特徴ベクトルを、メッセージ伝達を通じて更新し、隣接ノードの特徴を反映した埋め込みに変換する。ノード v が第 l 層でのグラフ畳み込みにて行う処理は以下のように定式化される。

$$\begin{aligned} \mathbf{h}_v^{(0)} &= \mathbf{X}_v \quad (\forall v \in V) \\ \mathbf{h}_v^{(l+1)} &= f_{\theta, l+1}(\mathbf{h}_v^{(l)}, \{\mathbf{h}_u^{(l)} \mid u \in \mathcal{V}(v)\}) \\ \mathbf{z}_v &= \mathbf{h}_v^{(L)} \end{aligned}$$

ここで、 V は入力グラフのノードの集合、 $\mathbf{X}_v$ はノード v の特徴ベクトル、 $\mathbf{h}_v^{(l)}$ は第 l 層でのノード v の埋め込み、 $f_{\theta, l+1}$ は集約関数、 $\mathcal{V}(v)$ はノード v に隣接するノードの集合、 $\mathbf{z}_v$ は最終的なノードの埋め込みベクトルである。全てのノードにおいて、

複数回にわたってグラフ畳み込みを行うことで、グラフ全体の構造を考慮した埋め込み表現を得る。代表的なグラフ畳み込みに、グラフ畳み込みネットワーク (Graph convolution networks; GCN)[33]、グラフアテンションネットワーク (Graph attention networks; GAT)[31] がある。集約関数 $f_{\theta,l+1}$ の設計は、このようなグラフ畳み込みの種類により異なる。

2.1.1 Graph convolution networks

GCN は、グラフ畳み込みを行う際、自身と隣接する全てのノードを対称に扱う。GCN において、ノード v が適用する集約関数は以下のように定式化される。

$$\begin{aligned} f_{\theta,l+1}(\mathbf{h}_v^{(l)}, \{\mathbf{h}_u^{(l)} \mid u \in \mathcal{V}(v)\}) \\ = \sigma \left(\sum_{u \in \mathcal{V}(v)} \frac{1}{\sqrt{|\mathcal{V}(v)|} \sqrt{|\mathcal{V}(u)|}} W^{(l+1)} \mathbf{h}_u^{(l)} \right) \end{aligned}$$

ここで、 $\sigma(\cdot)$ は活性化関数、 $\mathcal{V}(v)$ はノード v に隣接するノードの集合、 $|\mathcal{V}(v)|$ はノード v の次数、 $W^{(l+1)} \in \mathbb{R}^{d_{l+1} \times d_l}$ は $l+1$ 層目の重み行列、 d_l は l 層目のノード埋め込みの次元を指す。式の右辺において、隣接ノード $u \in \mathcal{V}(v)$ から得られる情報は、各ノードの次数（隣接ノードの数） $|\mathcal{V}(v)|$ および $|\mathcal{V}(u)|$ に基づく正規化係数

$$\frac{1}{\sqrt{|\mathcal{V}(v)|} \sqrt{|\mathcal{V}(u)|}}$$

を通じて重み付けされる。この正規化はノードの次数に依存する影響を抑え、安定した学習を可能にするために行う。この式に基づいて、各ノードのノード埋め込みはその周囲のノードからの情報を集約し、更新される。重み行列 $W^{(l+1)}$ は学習可能なパラメータであり、ネットワーク全体の学習を通じて最適化される。

2.1.2 Graph attention networks

GAT はアテンション機構を取り入れたグラフ畳み込みの手法である。アテンション機構は、入力データ内の重要な部分に対して動的に重みを付与する手法であり、情報の選択的な集約を可能にする。画像認識 [20]、自然言語処理 [21] をはじめとするあらゆる分野で活用されている。前述の GCN が次数を用いて隣接するノードを対称に重み付けし畳み込みを行うのに対し、GAT は隣接するノードごとに計算されるアテンション係数を使用して重み付けをし、畳み込みを行う。この際、学習されたアテンション係数はどのノードがタスクに重要であるかを定量的に示す。GAT でノード v に適用される集約関数は以下のように表す。

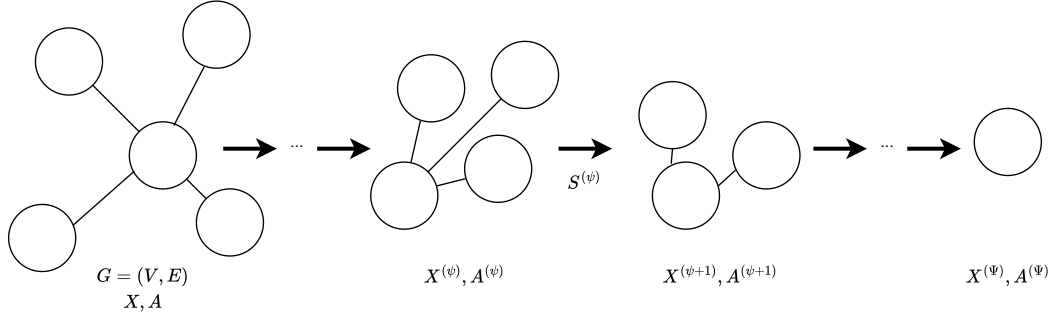


図 1

$$f_{\theta, l+1}(\mathbf{h}_v^{(l)} \{ \mathbf{h}_u^{(l)} \mid u \in \mathcal{V}(v) \}) = \sigma \left(\sum_{u \in \mathcal{V}(v)} \alpha_{vu}^{(l+1)} W^{(l+1)} \mathbf{h}_u^{(l)} \right),$$

$$\tilde{\mathbf{h}}_v^{(l)} = W_a^{(l+1)} \mathbf{h}_v^{(l)},$$

$$\alpha_{vu}^{(l+1)} = \frac{\exp \left(\text{LeakyReLU} \left(\mathbf{a}^{(l+1)\top} \cdot \text{concat} \left(\tilde{\mathbf{h}}_v^{(l)}, \tilde{\mathbf{h}}_u^{(l)} \right) \right) \right)}{\sum_{w \in \mathcal{V}(v)} \exp \left(\text{LeakyReLU} \left(\mathbf{a}^{(l+1)\top} \cdot \text{concat} \left(\tilde{\mathbf{h}}_v^{(l)}, \tilde{\mathbf{h}}_w^{(l)} \right) \right) \right)}.$$

ここで、 $\alpha_{vu}^{(l)}$ はノード v がノード u を畳み込む際のアテンション係数、 concat はベクトル同士の結合を表す。また、 $W^{(l+1)} \in \mathbb{R}^{d_{l+1} \times d_l}$ 、 $W_a^{(l+1)} \in \mathbb{R}^{f_{l+1} \times d_l}$ 、 $\mathbf{a} \in \mathbb{R}^{2f_{l+1}}$ は学習可能なパラメータである。パラメータ $W_a^{(l+1)}$ により、隣接ノードを畳み込む際の動的な重み付けが実現される。

2.2 グラフプーリングと Diffpool

GNN を使ったタスクには、グラフの分類など、グラフ全体の特徴量を考慮して処理を行うものが存在する。このようなタスクを行う目的で、グラフプーリング (Graph Pooling) が用いられる。グラフプーリングとは、GNN においてグラフのノードが持つ埋め込みをまとめ、1つの抽象的な埋め込みを得る処理の総称である。例えば、全てのノード埋め込みの平均値を取る平均値プーリングや、最大値を取る最大値プーリングなど、様々なグラフプーリング手法が考案されている。その中で、diffpool[32] は近年注目されているグラフプーリング手法の1つである。diffpool[32] は、縮約の過程を複数の段階に分割し、段階的にグラフプーリングを行う。diffpool は、ソフトクラスタリングを用いて、データセットごとに異なる動的なクラスタリングを行う点で他のグラフプーリング手法と異なる。縮約する際、diffpool はグラフのノードをクラスタに割り当てるクラスタ割り当て行列を用いる。クラスタ割り当て行列 S の

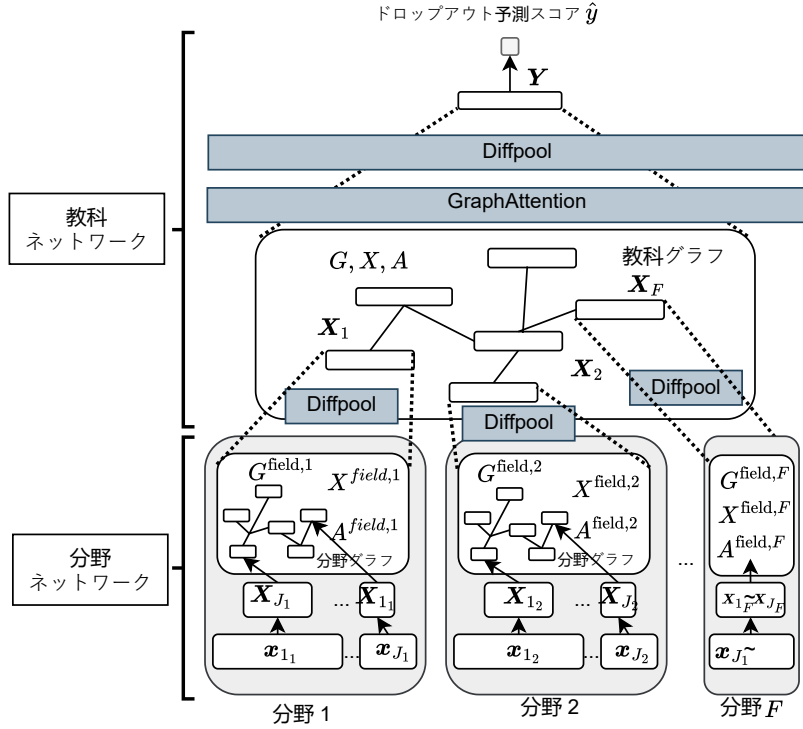


図 2: 提案手法のモデル図

構造は以下のようになる.

行列 $S \in \mathbb{R}^{\mathcal{N} \times \mathcal{N}'}$

$\mathcal{N}$: 現段階のグラフのノード数

$\mathcal{N}'$: クラスタ数 (次段階のグラフのノード数)

行列 S の各要素 S_{mn} は、ノード m がクラスタ n に属する重みを表す. これにより, diffPool はソフトクラスタリングを実現する. また, 行列 S の各行はソフトマックス正規化され, 行の合計が 1 となるように調整される. よって, 各ノードがどのクラスタに属するかが確率的に解釈できる. 以上のクラスタ割り当て行列の構造を踏まえて, diffpool が段階 ψ でのグラフ $G^{(\psi)}$ を段階 $\psi + 1$ へ縮約する際に行う手順を図 1 に示す. 段階 ψ のクラスタ割り当て行列 $S^{(\psi)}$ は $\text{GNN}_{\psi, \text{pool}}$ により計算される.

3 提案手法

本章では, グラフプーリングの手法である diffpool[32] を利用し, 2 種類のグラフを階層的な構造にしたドロップアウト予測モデルを提案する. 提案手法の詳細を図 2 に示す.

本研究では、ドロップアウト予測のデータセットに学生の入学生学力調査を用いる。そのため、提案モデルの入力特徴量は出題項目の正誤である。入学生学力調査には教科ごとに分野が存在し、それぞれの分野から項目が出題される。例えば、入学前試験の教科「数学」には、分野として「数と式」、「図形と計量」などが存在し、これらの分野から一様に項目が出題される。つまり、項目は分野に属し、分野は教科に属するという包含関係がある。このような3段階の階層関係に対し、「分野グラフ」と「教科グラフ」の2種類のグラフを定義する。分野グラフは、その分野に属する項目同士の関係を表現するグラフである。入力特徴量である項目と1対1に対応する項目埋め込みをノードに持ち、関連が強い項目の埋め込み間にエッジが存在する。教科グラフは、その教科に属する分野同士の関係を表現するグラフである。分野と1対1に対応する分野埋め込みをノードに持ち、分野グラフと同様に関連が強い分野の埋め込み間にエッジが存在する。これらのグラフに対し、グラフプーリングを用いて分野グラフを埋め込みに縮約し、教科グラフの分野埋め込みとして階層構造を構成する。さらに、教科グラフをグラフプーリングを用いて埋め込みに縮約することで最終的なドロップアウト予測スコアを得る。このような階層構造を構成する際、グラフプーリング手法に diffpool を用いて動的なグラフプーリングを行う。diffpool のクラスタ割り当て行列はグラフプーリングに使用する重みを表しており、分野グラフ、教科グラフを埋め込みに縮約するとき、それぞれのノードに割り当てられた重みが可視化できる。これにより、教科グラフを縮約する際のクラスタ割り当て行列を観察することで、ドロップアウト予測に影響を及ぼした分野、その程度を分析できる。さらに、分野グラフを縮約する際のクラスタ割り当て行列を可視化することで、それぞれの分野で学習された項目の重みを分析できる。

提案モデルでは、次の手順により入学生学力調査の結果を持つ学習者 i のある科目におけるドロップアウト予測スコアを求める。

3.1 分野ネットワーク

分野ネットワークは、入力である入学時事前調査項目の正誤から分野グラフを構成する。そして、分野グラフをグラフプーリングを用いて分野埋め込みに縮約し出力する GNN である。分野 $f(= 1, 2, \dots, F)$ のグラフ $G^{\text{field},f}$ を、分野 f に属する事前調査の項目の集合を表現するノードの集合 $V^{\text{field},f}$ 、エッジの集合 $E^{\text{field},f}$ から $G^{\text{field},f} = (V^{\text{field},f}, E^{\text{field},f})$ と定義する。分野 f の分野グラフ $G^{\text{field},f}$ は特徴行列 $X^{\text{field},f}$ 、隣接行列 $A^{\text{field},f}$ を持つ。

分野グラフ $G^{\text{field},f}$ の特徴行列 $X^{\text{field},f}$ は、学生 i が受験した事前調査の、分野 f に属する項目の正誤を入力として、分野ネットワークにより以下の手順で計算される。学生 i が分野 $f(= 1, 2, \dots, F)$ に属する項目 $j_f(= 1, 2, \dots, J_f)$ に正答したか否かを表す x_{j_f} を次のように定める。

$$x_{j_f} = \begin{cases} 1 & \text{正答} \\ 0 & \text{誤答} \end{cases}$$

なお、可読性のため以降では学生 i に限定し、添字の表記を省略する． $\mathbf{j}_f^{\text{one-hot}} \in \mathbb{R}^{J_f}$ を項目 j_f の j_f 番目の要素のみ 1、他の要素が 0 である one-hot ベクトルとすると、図 2 に示す分野 ネットワークの入力 $\mathbf{x}_{j_f} \in \mathbb{R}^{j_f+1}$ は以下ようになる．

$$\mathbf{x}_{j_f} = \text{concat}(\mathbf{j}_f^{\text{one-hot}}, x_{j_f})$$

ここで、concat はベクトルの結合を表す． 入力 $\mathbf{x}_{j_f}$ は多層パーセプトロン層を通り、項目の埋め込み $\mathbf{X}_{ij_f} \in \mathbb{R}^d$ に変換される． d は項目埋め込みの次元である．

$$\begin{aligned} \mathbf{X}_{j_f}^{(1)} &= \mathbf{x}_{j_f} \\ \mathbf{X}_{j_f}^{(k)} &= \text{relu}(\mathbf{X}_{j_f}^{(k-1)} W_{j_f}^{(k)} + \mathbf{b}_{j_f}^{(k)}) \quad (1 \leq k \leq K) \end{aligned}$$

$W_{j_f}^{(k)} \in \mathbb{R}^{d \times d}$ 、 $\mathbf{b}_{j_f}^{(k)} \in \mathbb{R}^d$ はそれぞれ分野 f に属する j_f 番目の項目を処理する多層パーセプトロン層の第 k 層目の重み行列、バイアスベクトルである． 多層パーセプトロン層の出力である項目埋め込み $\mathbf{X}_{j_f}^{(K)}$ から、分野 f のグラフの特徴行列 $X^{\text{field},f} \in \mathbb{R}^{J_f \times d}$ を次のように定義する．

$$X^{\text{field},f} = \begin{bmatrix} \mathbf{X}_1^{(K)} \\ \mathbf{X}_2^{(K)} \\ \vdots \\ \mathbf{X}_{j_f}^{(K)} \\ \vdots \\ \mathbf{X}_{J_f}^{(K)} \end{bmatrix}$$

また、分野グラフ $G^{\text{field},f}$ の隣接行列 $A^{\text{field},f} \in \mathbb{R}^{J_f \times J_f}$ を行列 $\Omega^{\text{field},f} \in \mathbb{R}^{J_f \times J_f}$ を用いて次のように定める． $A^{\text{field},f}$ の m 行 n 列の要素 $a_{mn}^{\text{field},f}$ は、1 であれば項目ノード m 、項目ノード n 間にエッジが存在し、0 であればエッジが存在しないことを表す． $A_{mn}^{\text{field},f}$ は行列 $\Omega^{\text{field},f}$ の m 行 n 列の要素 $\omega_{mn}^{\text{field},f}$ 、閾値を表すハイパーパラメータ λ^{field} から次のように定められる．

$$A_{mn}^{\text{field},f} = \begin{cases} 1 & \omega_{mn}^{\text{field},f} > \lambda^{\text{field}} \\ 0 & \text{それ以外} \end{cases}$$

$\Omega^{\text{field},f}$ はモデルが持つ学習可能なパラメータであり、誤差逆伝播により学習される． 以上のように定義された分野グラフ $G^{\text{field},f}$ を入力として、隣接する項目ノード間で畳み込みを行う．最後に、Diffpool を使用して Ψ 回の段階に分けて分野 f のグラフ $G^{\text{field},f}$ を縮約し、分野 f を表す中間表現 $H^f \in \mathbb{R}^{\mathcal{N}^{(\Psi),\text{field}} \times d}$ を出力する． $S^{(\Psi),\text{field},f} \in \mathbb{R}^{\mathcal{N}^{(\Psi),\text{field}} \times \mathcal{N}^{(\Psi-1),\text{field}}}$ は段階 $\Psi - 1$ から段階 Ψ に縮約する際の Diffpool のクラスタ割り当て行列である． $\mathcal{N}^{(\Psi),\text{field}}$ は段階 Ψ での分野グラフのノード数を表すハイパーパラメータである．

$$H^f = S^{(\Psi),\text{field},fT} X^{\text{field},f} \quad (1)$$

最終的な分野 ネットワークの出力である分野埋め込み $\mathbf{X}_f \in \mathbb{R}^d$ は、縮約後の分野グラフに残存するノードの埋め込みの平均値である。

$$\mathbf{X}_f = \frac{1}{\mathcal{N}^{(\Psi), \text{field}}} \sum_{n=1}^{\mathcal{N}^{(\Psi), \text{field}}} \mathbf{H}_n^f \quad (2)$$

ここで、 $\mathbf{H}_n^f$ は H^f の n 行目の行ベクトルである。

このように、全ての分野ネットワークにおいて項目埋め込みから分野グラフを構築し、出力である分野埋め込み $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_F$ を得る。これらの F 個の分野埋め込みは独立した分野ネットワークから出力されるため、それぞれの分野の埋め込みとなっている。

3.2 教科ネットワーク

教科ネットワークは分野ネットワークの出力である分野埋め込みを用いて教科グラフを構成し、教科グラフをグラフプリーングによって縮約してドロップアウト予測スコアを出力する。教科グラフ G は、分野ノードの集合を V 、エッジの集合を E とすると $G = (V, E)$ と定義する。このとき、教科グラフ G の特徴行列 X と隣接行列 A は以下のように定められる。

$$X = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_f \\ \vdots \\ \mathbf{X}_F \end{bmatrix} \quad (3)$$

教科グラフのノード m ，ノード n 間を接続するエッジの有無を表す隣接行列 A の m 行 n 列の要素 A_{mn} は、提案モデルの持つ学習可能なパラメータ $\Omega \in \mathbb{R}^{F \times F}$ の m 行 n 列の要素 ω_{mn} ，閾値を表すハイパーパラメータ λ から次のように定められる。

$$A_{mn} = \begin{cases} 1 & \omega_{mn} > \lambda \\ 0 & \text{それ以外} \end{cases} \quad (4)$$

このとき、 Ω はモデルが持つパラメータであり、誤差逆伝播により学習される。構築された教科グラフをグラフアテンションネットワーク (GAT; Graph Attention Networks)[31] に入力する。通常のグラフ畳み込み層が隣接ノードを対称に畳み込んでノード埋め込みを更新するのに対し、GAT は隣接ノード全てにアテンション係数を計算し、異なる重みをつけて畳み込む。その際、GAT はアテンション係数を算出する重み行列をパラメータに持つため、各ノードの重要度を動的に考慮し学習を行いながら処理を行う。独立した分野ノードを複数回畳み込むことで、グラフ全体に

わたってメッセージ伝達を行い、ノード間の関係や局所的な特徴を学習する。特徴行列 X が GAT によって更新される様子は以下である。

$$Z = \alpha F, \quad F = XW, \quad (5)$$

$$\alpha_{mn} = \text{softmax} \left(\text{LeakyReLU} \left(\mathbf{a}^\top \left[\mathbf{F}_m \parallel \mathbf{F}_n \right] \right) \right). \quad (6)$$

ここで、 $Z \in \mathbb{R}^{F \times d}$ は GAT によって畳み込みが行われた後の特徴行列の中間表現、 $\alpha \in \mathbb{R}^{F \times F}$ はアテンション行列、 $W \in \mathbb{R}^{d \times d}$ は重み行列、 α_{mn} は GAT によって計算された頂点 m, n 間のアテンション係数、 $\mathbf{a} \in \mathbb{R}^{2d}$ は学習可能なパラメータ、 $\mathbf{F}_m \in \mathbb{R}^d$ は行列 H の m 行目の行ベクトルである。これらを diffpool 層を通して縮約を行う。

$$H = S^{(\Psi)\top} Z \quad (7)$$

縮約後のノード数はハイパーパラメータである。その数を $\mathcal{N}^{(\Psi)}$ とすると、最終的なドロップアウト予測スコア $\hat{y} \in \mathbb{R}^1$ は $\mathcal{N}^{(\Psi)}$ 個のノードの平均から求められる。

$$\mathbf{Y} = \frac{1}{\mathcal{N}^{(\Psi)}} \sum_{n=1}^{\mathcal{N}^{(\Psi)}} \mathbf{H}_n, \quad (8)$$

$$\hat{y} = \sigma(\mathbf{Y}W_{out} + \mathbf{b}_{out}). \quad (9)$$

ここで、 $\mathbf{Y} \in \mathbb{R}^d$ は縮約後の教科グラフの持つノード埋め込みの平均、 $\mathbf{H}_n \in \mathbb{R}^d$ は H の n 行目の行ベクトル、 $W_{out} \in \mathbb{R}^{d \times 1}$ は重み行列、 $\mathbf{b}_{out} \in \mathbb{R}^{1 \times 1}$ はバイアスベクトルである。

最終的な出力のドロップアウト予測スコアは、閾値を使用して学生がドロップアウトをするか否か予測を行うために用いる。閾値を 0.5 とし、閾値を上回れば学生がドロップアウトをせず、下回ればドロップアウトをすると定義する。

また、ドロップアウト予測では、ほとんどのデータセットがドロップアウトをする学生が少数クラス (サンプル数が少ないクラス) である不均衡なデータである。不均衡なデータセットにおいて、機械学習モデルは損失関数を最小化することを目的として学習を行う。そのため、サンプル数が多い多数クラスに対して正確に予測することで、全体の損失を効率的に削減しようとする傾向が生じる。この結果、少数クラスのデータはモデルに十分考慮されず、誤って多数クラスに分類されやすくなる。しかし、ドロップアウト予測では、ドロップアウトをしない学生をドロップアウトすると誤って予測することより、ドロップアウトをする学生をドロップアウトしないと誤った予測をすることの方がその後の影響において重大である。ドロップアウト予測の先行研究でも、不均衡なデータセットの中で少数クラスである、ドロップアウト予測をする学生を誤って予測しないことを重視した対策が行われてきた [12, 15]。そこで、本研究では、損失関数に focalloss [36] を用いる。focalloss [36] はドロップアウト予測に限らず、不均衡なデータへ有効な手法として知られている [17, 18]。focalloss はバイナリクロスエントロピー (Binary Cross entropy; BCE) を用いて計算する。BCE は以下のような数式で表される。

$$\text{BCE}(\hat{y}, y) = -y \log(\hat{y}) - (1 - y) \log(1 - \hat{y}), \quad (10)$$

ここで、 y は実際のドロップアウトをしたかを表すラベル、 $\hat{y}$ はドロップアウト予測スコアである．次のように p_t を定めると、BCE は以下のように表すことができる．

$$p_t = \begin{cases} \hat{y}, & \text{if } y = 1, \\ 1 - \hat{y}, & \text{if } y = 0, \end{cases} \quad (11)$$

$$\text{BCE}(p_t) = -\log(p_t). \quad (12)$$

これらを用いて、focalloss は次のように定められる．

$$\text{FocalLoss}(p_t) = -(1 - p_t)^\gamma \log(p_t) \quad (13)$$

$(1 - p_t)^\gamma$ は、少数クラスの予測時に大きく、多数クラスの予測時に小さくなる． γ はその変動の程度を定めるハイパーパラメータである．これにより、少数クラスの予測時に損失を大きく算出し、少数クラスを重視した予測を行うことが可能となる．

4 評価実験

本章では、これまでドロップアウト予測に用いられてきた代表的な機械学習手法 (LightGBM, XGBoost)、深層学習手法 (DNN) と提案手法を用いてドロップアウト予測を行う．

4.1 データの概要

表 1: 入学生学力調査データ (数学)

No.Students	No.Fields	No.items
666	18	141

提案手法の入力データは電気通信大学 2023 年度新入生の入学生学力調査のデータ、ラベルデータは同新入生が受講した大学初年度の成績データである．入力データの概要を表 1、表 2、表??に示す．入学生学力調査とは、電気通信大学の新入生が入学直後に受験する、入学時点での学生の数学の学力、物理の学力を測定する試験である．本研究では、この試験のうち数学の正誤データを入力として使用する．表 1 が示す数学の事前調査は学生 666 人が受験し、141 項目のアイテムバンクの中から項目が出題された．No.Fields は数学の分野の数であり、具体的には「数と式」、「図形と計量」など 18 個の分野が該当する．全ての数学の事前調査の項目は 1 つの数学の分野と紐づけられており、項目ごとに問われる能力が異なる．144 項目が属する数学の分野の一覧、その分野に属する項目の数を表 2 に示す．数学の事前調査では、学生ごとに出席項目が異なり、学生ごとに出席されていない項目が存在する．このような欠損項目の正誤は、事前に求めたパラメータを用いた項目反応理論 (Item

表 2: 入学生学力調査のデータ (数学)
の分野一覧

Field name	No.item
数と式	14
図形と計量	9
二次関数	6
データの分析	2
いろいろな式	11
図形と方程式	5
指数関数・対数関数	9
三角関数	8
微分・積分の考え	10
極限	7
微分法	7
積分法	8
図形の性質	6
場合の数と確率	13
数学と人間の活動	4
数列	8
ベクトル	6
平面上の曲線と複素数平面	9

Response Theory; IRT) の正答確率の推定値で補完した。ドロップアウト予測を行う対象の科目は電気通信大学で一年次に開講する 7 つの数学科目である。これらの科目は秀, 優, 良, 可, 不可からなる 5 段階の評定を持つ。そのうち, 不可を取得した人数が少数であること, 可を取得した学生も成績不良の危険性があることを考慮し, 可, 不可を取得した学生をドロップアウトと定義する。科目ごとの受講人数, その中でドロップアウトした人数, 受講生に占める割合を表 3 に示す。このような実験設定のもと, 提案手法のハイパーパラメータのグリッドサーチを行い, 最も AUC が高い値を採用した。具体的には, 分野ネットワークの縮約数を 1, 教科ネットワークの縮約数を 4, 隠れ層の次元を 16, focal loss の γ は 2.0 とした。さらに, 全ての diffpool 層の段階数を $\Psi = 2$ とした。比較手法の DNN は表 4 が示すと通りの構成とし, 層の数などハイパーパラメータは最も AUC が高いものを採用した。

4.2 予測精度の評価実験

本節では, データセットを 5 分割交差検証を用いて訓練データ, 検証データ, 評価データに分割し, 訓練データ, 検証データから得たパラメータを利用して評価データのドロップアウト予測を行う。本研究で利用したデータセットは, 7 科目を平均したドロップアウトする学生の割合が 0.35 であり, 不均衡なデータセットである。

表 3: 大学初年度の成績データ

Subject name	No.Students	No.Dropout	Rate Dropout
微分積分学第一	697	233	0.33
線形代数学第一	698	248	0.36
数学演習第一	698	247	0.35
微分積分学第二	687	223	0.32
線形代数学第二	688	227	0.33
解析学	687	294	0.43
数学演習第二	687	236	0.34

表 4: DNN の構成

層番号	層の種類	次元 (入力 → 出力)
1	全結合層 (Linear)	150 → 16
2	活性化関数 (ReLU)	—
3	全結合層 (Linear)	16 → 16
4	活性化関数 (ReLU)	—
5	全結合層 (Linear)	16 → 16
6	活性化関数 (ReLU)	—
7	全結合層 (Linear)	16 → 16
8	活性化関数 (ReLU)	—
9	全結合層 (Linear)	16 → 16
10	活性化関数 (ReLU)	—
11	全結合層 (Linear)	16 → 1
12	活性化関数 (Sigmoid)	—

表 5: 各手法の AUC の比較

科目	提案手法	DNN	XGB	LGBM
微分積分学第一	0.723 ± 0.0713	0.684 ± 0.0707	0.677 ± 0.035	0.677 ± 0.032
線形代数学第一	0.679 ± 0.0362	0.644 ± 0.0476	0.664 ± 0.066	0.667 ± 0.075
数学演習第一	0.709 ± 0.0411	0.712 ± 0.0316	0.644 ± 0.037	0.653 ± 0.035
微分積分学第二	0.675 ± 0.0298	0.650 ± 0.0280	0.592 ± 0.034	0.598 ± 0.023
線形代数学第二	0.637 ± 0.0371	0.611 ± 0.0260	0.610 ± 0.033	0.601 ± 0.018
解析学	0.693 ± 0.0524	0.689 ± 0.0354	0.635 ± 0.062	0.637 ± 0.060
数学演習第二	0.729 ± 0.0647	0.717 ± 0.0673	0.644 ± 0.059	0.643 ± 0.047
全科目平均	0.692 ± 0.019	0.672 ± 0.018	0.638 ± 0.018	0.639 ± 0.017

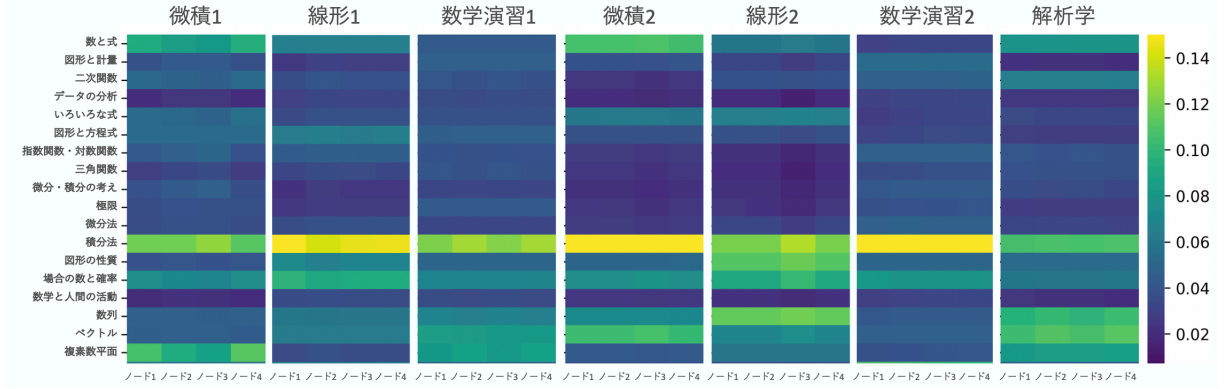


図 3: S^{all} のヒートマップ

そのため、予測精度の評価指標として、正例と負例の数に依存せずモデルが正例と負例をどの程度正確に区別できるかを示す AUC を用いる。AUC は、ROC 曲線の下側面積を数値化したものである。ROC 曲線は、モデルの予測結果に基づいて異なる閾値における True Positive Rate と False Positive Rate の関係をプロットしたものであり、下側面積を求めることでモデルのクラス分離能力を測定する。AUC の値は 0.0 から 1.0 の範囲をとり、1.0 に近い値ほどモデルの予測性能が高い。具体的には、AUC が 1.0 である場合、モデルは正例と負例を完全に区別できることを意味し、0.5 の場合はランダムな予測と同等の性能を示す。一方、AUC が 0.5 未満である場合、モデルの予測が逆効果であることを示唆する。全ての手法でドロップアウト予測を行った結果を表 5 に示す。数学演習第一では DNN が、それ以外の科目では提案手法が最も高い AUC であった。全科目平均では、提案手法が従来手法を大きく上回った。このことから、提案手法が従来手法より優れた予測性能を持っていることが確認できる。また、ほとんどの科目において、XGBoost, LightGBM の勾配ブースティング系のアルゴリズムより DNN, 提案手法の深層学習手法が予測性能が高いことから、ドロップアウト予測での深層学習手法の有効性が明らかになった。

4.3 クラスタ割り当て行列の分析

本節では、前述の実験設定で行った学習後の提案手法のモデルにおいて、教科ネットワークのクラスタ割り当て行列 $S^{(1)} \in \mathbb{R}^{19 \times 10}$, $S^{(2)} \in \mathbb{R}^{10 \times 4}$ の詳細な分析を行う。クラスタ割り当て行列は、DiffPool 層で学習可能なパラメータとして定義され、教科グラフを縮約する過程で用いられる。これを解析することにより、教科ネットワークを構成する分野ノードのうち、予測に対して影響を及ぼすノードおよびその寄与度を特定することが可能となる。

行列 $S^{all} \in \mathbb{R}^{19 \times 4}$ を $S^{(1)}, S^{(2)}$ の積として次のように定義する。

$$S^{all} = S^{(1)}S^{(2)}.$$

このとき、行列 S^{all} はそれぞれの分野ノードを縮約後のノードに割り当てる程度を表す。例えば、 S^{all} の m 行 n 列の要素 s_{mn}^{all} は、 m 番目の分野が縮約後の n 個目の

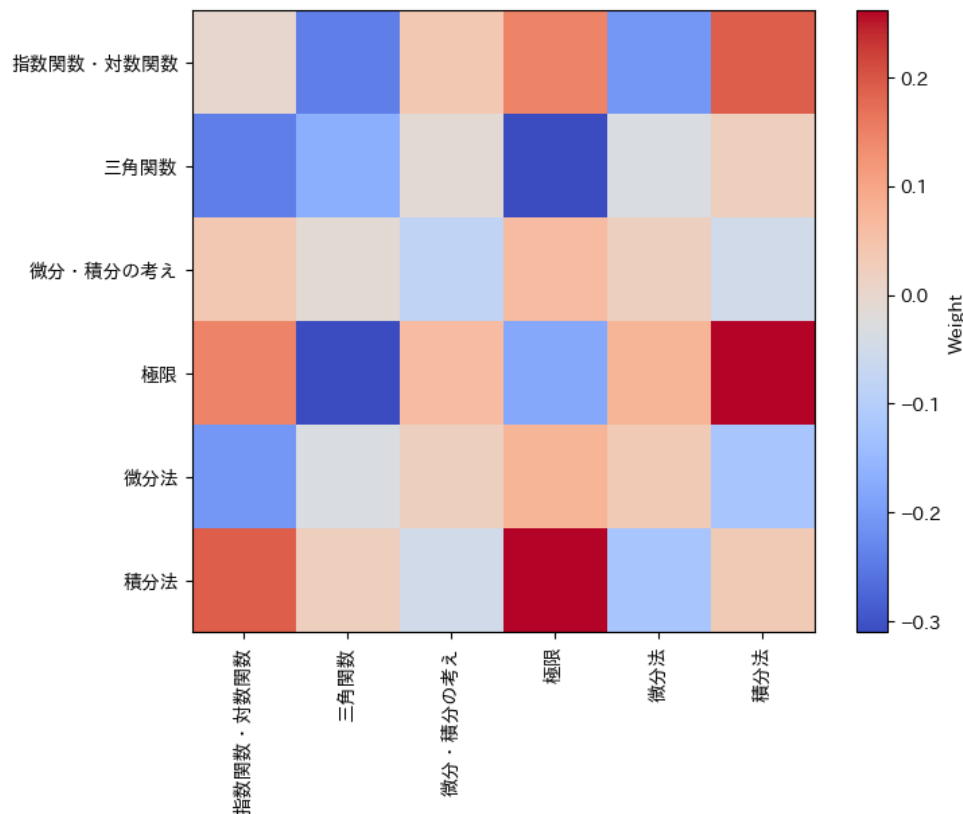


図 4: 微分積分学第一の教科グラフにおける隣接行列の一部

ノードに寄与した割合を示す。前述の数学7科目についてモデルの学習を行い、各科目において学習を通して得られた行列 S^{all} のヒートマップを図3に示す。縦軸は分野の名前、横軸は縮約後の教科グラフのノードに対応する。ヒートマップの色はクラスタ割り当て行列の数値によって異なり、黄色に近いほど予測結果に寄与している。ヒートマップから、全ての科目において、「積分法」が最も大きな影響を及ぼしていることが読み取れる。その他の分野は教科によって寄与しているものが異なり、それぞれの教科が違う分野と関連していると解釈できる。また、微分積分学第一と微分積分学第二は類似した科目であるが、両者とも「数と式」が他の教科と比較して影響しており、反対に「データの分析」は予測に関与していないことが観察される。

4.4 隣接行列の分析

提案手法の持つグラフは学習可能なパラメータとして隣接行列を持っている。そこで、本節では予測を通して学習された教科グラフの隣接行列の分析を行う。教科グラフは無向グラフなので、隣接行列は対称行列である。教科グラフのノードは分野と1対1に対応しており、隣接行列を分析することで分野間の関係を可視化することができる。図4は微分積分学第一の教科グラフにおける隣接行列の一部をプロットしたヒートマップである。色が赤に近いほど隣接行列の数値が大きく学習されていることを示す。この図から、「積分法」と「極限」、「積分法」と「指数関数・対数関

数」の間に強いエッジが学習されていることが観察される．一方で、「極限」と「三角関数」の間には比較的低い重みが付与されている．これらは学習内容間の関連性を反映していると考えられる．「積分法」と「極限」、「積分法」と「指数関数・対数関数」は、数学的概念やその応用の観点で密接に関連していることから、これらの結果は直感的にも妥当である．「極限」と「三角関数」に関しても同様である．

5 まとめ

近年、学校教育の現場では、学生のドロップアウトを防ぐためにより精度が高く、且つドロップアウトに寄与した要因の把握が可能な手法が求められている．本論文では、従来手法と比較し精度の高い予測を行いつつ、予測に影響を及ぼした特徴量、特徴量同士の関連を分析することを目的とした．具体的には、diffpool[32]を使用した階層的な GNN を持つドロップアウト予測手法を提案した．評価実験では、提案手法は従来手法と比較し高い精度を示した．教科ネットワークの隣接行列の分析から、学習を通して得られた分野同士の関連を分析した．また、学習可能な隣接行列と動的なクラスタリングから特徴量を考察し、使用したデータセットにおいて予測結果に大きな影響を及ぼす分野を特定した．提案モデルでは、数学に焦点を当ててドロップアウト予測を行った．今後の課題として、物理や化学など複数種類の教科の予測への拡張が考えられる．これにより、教科間を縦断した分野の関連性の分析、他教科のドロップアウトに対する分野の影響などを考察することが可能となる．

参考文献

- [1] Namoun, A., & Alshantiri, A. (2020). Predicting student performance using data mining and learning analytics techniques: A systematic literature review. *Applied Sciences*, 11(1), 237.
- [2] Akçapınar, G., Altun, A., & Aşkar, P. (2019). Using learning analytics to develop early-warning system for at-risk students. *International Journal of Educational Technology in Higher Education*, 16(1), 1-20.
- [3] Aulck, L., Nambi, D., Velagapudi, N., Blumenstock, J., & West, J. (2019). Mining University Registrar Records to Predict First-Year Undergraduate Attrition. *International Educational Data Mining Society*.
- [4] Sivakumar, S., Venkataraman, S., & Selvaraj, R. (2016). Predictive modeling of student dropout indicators in educational data mining using improved decision tree. *Indian Journal of Science and Technology*, 9(4), 1-5.
- [5] Zhang, K., Ji, T., & Zhang, H. (2024). Knowledge tracing via multiple-state diffusion representation. *Expert Systems with Applications*, 124797.

- [6] Mason, C., Twomey, J., Wright, D., & Whitman, L. (2018). Predicting engineering student attrition risk using a probabilistic neural Networks and comparing results with a backpropagation neural Networks and logistic regression. *Research in Higher Education*, 59, 382-400.
- [7] Xu, K., Hu, W., Leskovec, J., & Jegelka, S. (2018). How powerful are graph neural networks?. *arXiv preprint arXiv:1810.00826*.
- [8] Perozzi, B., Al-Rfou, R., & Skiena, S. (2014, August). Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 701-710).
- [9] Wang, Z., Zhang, J., Feng, J., & Chen, Z. (2014, June). Knowledge graph embedding by translating on hyperplanes. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 28, No. 1).
- [10] Tsutsumi, E., Kinoshita, R., & Ueno, M. (2021). Deep-IRT with Independent Student and Item Networks. *International Educational Data Mining Society*.
- [11] Tsutsumi, E., Guo, Y., Kinoshita, R., & Ueno, M. (2023). Deep knowledge tracing incorporating a hypernetwork with independent student and item networks. *IEEE Transactions on Learning Technologies*.
- [12] Palacios, C. A., Reyes-Suárez, J. A., Bearzotti, L. A., Leiva, V., & Marchant, C. (2021). Knowledge discovery for higher education student retention based on data mining: Machine learning algorithms and case study in Chile. *Entropy*, 23(4), 485.
- [13] Ng, K., & Lei, P. (2022, July). A Lightweight Method using LightGBM Model with Optuna in MOOCs Dropout Prediction. In *Proceedings of the 6th International Conference on Education and Multimedia Technology* (pp. 53-59).
- [14] Elbouknify, I., Berrada, I., Mekouar, L., Iraqi, Y., Bergou, E. H., Belhabib, H., ... & Wardi, S. (2024, July). Student At-Risk Identification and Classification Through Multitask Learning: A Case Study on the Moroccan Education System. In *International Conference on Artificial Intelligence in Education* (pp. 372-380). Cham: Springer Nature Switzerland.
- [15] Nabil, A., Seyam, M., & Abou-Elfetouh, A. (2021). Prediction of students ' academic performance based on courses ' grades using deep neural networks. *IEEE Access*, 9, 140731-140746.
- [16] Yang, D., Sinha, T., Adamson, D., & Rosé, C. P. (2013, December). Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses. In *Proceedings of the 2013 NIPS Data-driven education workshop* (Vol. 11, p. 14).

- [17] Mulyanto, M., Faisal, M., Prakosa, S. W., & Leu, J. S. (2020). Effectiveness of focal loss for minority classification in network intrusion detection systems. *Symmetry*, 13(1), 4.
- [18] Pasupa, K., Vatathanavaro, S., & Tungjitnob, S. (2023). Convolutional neural networks based focal loss for class imbalance problem: a case study of canine red blood cells morphology classification. *Journal of Ambient Intelligence and Humanized Computing*, 14(11), 15259-15275.
- [19] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.
- [20] Dosovitskiy, A. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.
- [21] Vaswani, A. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*.
- [22] Tan, M., & Shao, P. (2015). Prediction of student dropout in e-Learning program through the use of machine learning method. *International journal of emerging technologies in learning*, 10(1).
- [23] Wagner, K., Merceron, A., & Sauer, P. (2020). Accuracy of a cross-program model for dropout prediction in higher education. In *Companion Proceedings of the 10th International Learning Analytics & Knowledge Conference (LAK 2020)* (pp. 744-749).
- [24] Ajjawi, R., M. Dracup, N. Zacharias, S. Bennett, and D. Boud. 2019. "Persisting Students' Explanations of and Emotional Responses to Academic Failure." *Higher Education Research & Development* 39(open in a new window) (2(open in a new window)): 185–99.
- [25] Tinto, V. (2017). Reflections on student persistence. *Student Success*, 8(2), 1-8.
- [26] Jimerson, S., Egeland, B., Sroufe, L. A., & Carlson, B. (2000). A prospective longitudinal study of high school dropouts examining multiple predictors across development. *Journal of school psychology*, 38(6), 525-549.
- [27] Spady, W. G. (1971). Dropouts from higher education: Toward an empirical model. *Interchange*, 2(3), 38-62.
- [28] Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., & Monfardini, G. (2008). The graph neural network model. *IEEE transactions on neural networks*, 20(1), 61-80.

- [29] Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).
- [30] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30.
- [31] Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y. (2017). Graph attention networks. *arXiv preprint arXiv:1710.10903*.
- [32] Ying, Z., You, J., Morris, C., Ren, X., Hamilton, W., & Leskovec, J. (2018). Hierarchical graph representation learning with differentiable pooling. *Advances in neural information processing systems*, 31.
- [33] Kipf, T. N., & Welling, M. (2016). Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.
- [34] Lundberg, S. (2017). A unified approach to interpreting model predictions. *arXiv preprint arXiv:1705.07874*.
- [35] Zhang, J., Shi, X., King, I., & Yeung, D. Y. (2017, April). Dynamic key-value memory networks for knowledge tracing. In *Proceedings of the 26th international conference on World Wide Web* (pp. 765-774).
- [36] Lin, T. (2017). Focal Loss for Dense Object Detection. *arXiv preprint arXiv:1708.02002*.