

令和元年5月16日現在

機関番号：22604

研究種目：若手研究(B)

研究期間：2016～2018

課題番号：16K16331

研究課題名（和文）機械学習による学生の修学状態モデルの構築と学習支援への展開

研究課題名（英文）Model Construction of Students' Learning States Using Machine Learning and its Application to Learning Support

研究代表者

近藤 伸彦（Kondo, Nobuhiko）

首都大学東京・大学教育センター・准教授

研究者番号：10534612

交付決定額（研究期間全体）：（直接経費） 3,000,000円

研究成果の概要（和文）：本研究では、大規模な教育データを活用して一人ひとりに応じた適切な学習支援を行うことをめざし、機械学習手法による学生の修学状態のモデル化と、これを用いた学習支援について検討した。大学の保有する学生に関するデータをもとに、いくつかのモデル化手法を用いて数値実験を行いその有用性を検討した。ベイジアンネットワーク、ランダムフォレスト等のよく知られたアルゴリズムを用いて、修学状態推移のモデル化や学習成果の予測モデル構築等を行い、一定の精度をもつモデル化が可能であることを示した。また、大学における教育データという特殊性をふまえたモデル化のフレームワークを提案し、汎用性のある知見として研究成果をまとめた。

研究成果の学術的意義や社会的意義

本研究は、汎用的な確率モデルとして知られるベイジアンネットワークを用いた大学生の修学状態の推移のモデル化のフレームワークの提案、機械学習による学習成果の予測モデルを構築することについての数値的な検証、の2つに大きくまとめられる。これらに共通するのは、「どのような大学にも対応できるモデル」ではなくそれぞれの大学の状況に応じた「モデル構築の方法論」について提案したことと、学期ごとの成績のような粒度の粗いデータと学習管理システムのログのような粒度の細かいデータをあわせて用いることの可能性について検討を行ったことであり、実際の大学の現場での活用に向けた知見をまとめたことに意義があると考えられる。

研究成果の概要（英文）：In this research, with the aim of providing appropriate learning support according to each student by using educational big data, modeling of students' learning states by machine learning method and learning support using such models were examined. Based on the students' data held by a university, usefulness of some modeling methods were examined by numerical experiments. By using well-known algorithms such as Bayesian networks and random forests, the transition of learning states were modeled and the prediction models of learning outcomes were constructed, and modeling with a certain accuracy was shown to be possible. In addition, a modeling framework based on the particularity of educational data in universities was proposed, and the research results as general findings were summarized.

研究分野：ラーニングアナリティクス、IR、教育工学

キーワード：機械学習 教学IR 予測モデル 確率モデル 修学状態 学習支援 ラーニングアナリティクス

## 様式 C-19, F-19-1, Z-19, CK-19 (共通)

### 1. 研究開始当初の背景

研究を開始した時期には、大規模なデータに基づく教育改善の必要性が広く認識されてきていた。なかでも、大規模な教育データをデータマイニング等の手法を用いて分析・可視化し、学習プロセスの改善をめざす Learning Analytics (LA) や Educational Data Mining (EDM) が新しい研究領域として注目されていた。

一方、大学全入時代に入って久しい日本の高等教育では、学力や学習意欲などあらゆる面で学生は多様化しており、かつこれまで以上の高等教育の質保証が求められてきていた。ある層には中途退学を未然に防ぐ施策が必要であり、また別の層にはより学習成果を上げるための施策が必要であるといった、多様な層に対する多様な施策が求められていた。ここで急速に注目を集めていたのが教学 IR (Institutional Research) であり、学生の多様さへの対応と学習成果の保証を同時に解決するためには、教育データを有効に活用し、個に応じたきめ細かい学習支援を行うことが必要であるという課題があった。そのためには、教学 IR に LA や EDM の観点を導入し、教育ビッグデータ分析の知見や手法を融合させた研究・開発が今後ますます必要になるという認識も生まれつつあった。

また、これを実際の学習支援において活用するためには、学習支援者、あるいは学生自身が理解しやすく、使いやすいシステムを構築しなければならない。その意味で、可解釈性に優れた分析・予測モデルと、その支援システム化が必要であった。

### 2. 研究の目的

本研究は、多様化する学生への対応と学習成果の保証という 2 つの課題を同時に解決するために、LA や EDM と教学 IR を融合し、大規模な教育データを活用して一人ひとりに応じた適切な学習支援を実現するために、機械学習手法による大学学士課程全体にわたる学生の修学状態のモデル化と、これを用いた学習支援について検討するものとし、数値実験を通じた包括的な検討により、有用なモデル構築と実際の学習支援への活用についての指針を得ることが本研究の当初の目的であった。

### 3. 研究の方法

上述の目的を達成するにあたって、次のような課題を設定した。

- 《課題 1》予測モデルにおけるモデル選択：将来の修学状態の予測を可能とするような、学士課程にわたる修学状態モデルを構築する際には、どのような種類のモデルを用いるのが適当か。汎化能力の高いモデルの獲得に適したモデル構造やパラメータ設定はどのようなものか。
- 《課題 2》有用な修学状態推移モデルの構築：確率モデルによる修学状態推移モデルを構築するうえで、汎化能力の高いモデルを得るためには、その構造をどのように決定すればよいか。
- 《課題 3》実問題への適用：構築したモデルをどのように実際の学習支援へ結びつけるか。実際に活用することが容易であり、かつ成果を生むようなシステムをいかに構築するか。

これらの課題について、具体的に以下の点を明らかにすることをめざした。

《課題 1》では、さまざまな機械学習手法について大規模な数値実験を行い、この課題における有用なモデルの選択と学習手法についての指針を得る。

《課題 2》では、ベイジアンネットワークに代表される確率モデルの可能性を探るため、適切な構造選択や、特徴量の離散化方法などについて実験的に検討し、有用な指針を得る。また、《課題 1》《課題 2》に共通する課題として、修学状態のモデル化において有用な特徴量についての知見を得る。LMS のログなど、扱うデータの種類を拡張することも含めて考える。

《課題 3》に関しては、実問題適用のさまざまな可能性のひとつとして、実際の学習支援者の支援活動をサポートするシステムの実証実験を行い、実運用上有用な知見を得る。本研究により得られた修学状態モデルを用いて修学予測システムを構築し、学習支援者の支援活動に活用してもらい、その成果を検討する。

これらに基づきつつ軌道修正しながら、より汎用的な知見が得られることを念頭に研究を進めた、次節にてその成果を示す。

### 4. 研究成果

主な研究成果を以下の 2 つの観点からまとめる。

#### (1) ベイジアンネットワークによる修学状態推移モデルの構築

3. の《課題 2》《課題 3》に対応する成果である。

5. の〔雑誌論文〕の①②では、大学学士課程における学生の修学状態の推移過程をベイジアンネットワークによりモデル化するフレームワークを提案した。ベイジアンネットワークとは、確率変数をノードとする非循環有向グラフネットワーク構造と条件付き確率パラメータ集合によって離散確率分布の同時分布を近似するモデルであり、変数間の依存関係がグラフィカルに表現される可解釈性、ひとつのモデルにおいて説明変数・目的変数を任意に指定できる柔軟性など多くの利点を持ち、幅広い応用事例があるモデルである。本研究では、大学の保有できる教学データから、学生の修学状態の推移を確率的に表現するベイジアンネットワークを構築する手順と、用いる変数の分類を行い、実際の教学 IR や学習支援に用いることを想定した

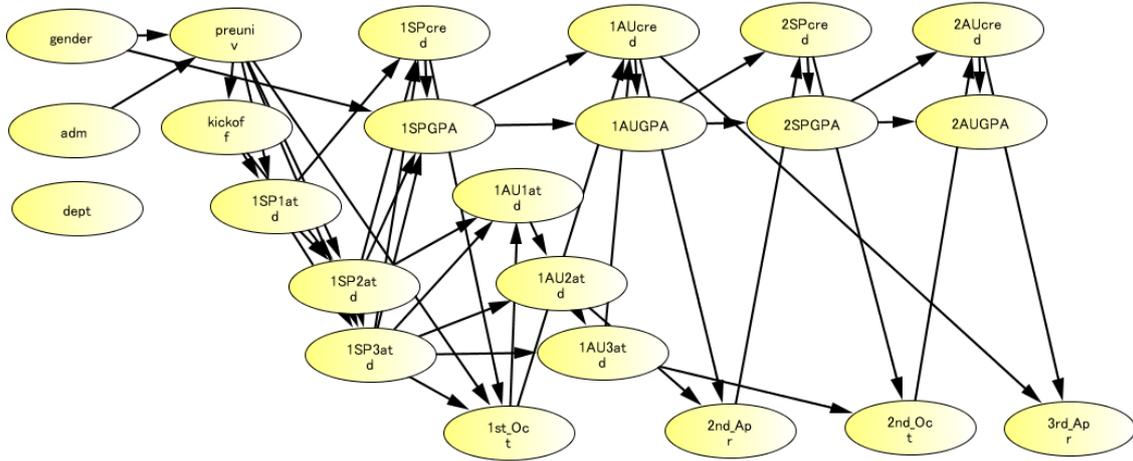


図1 修学状態の推移を示すベイジアンネットワークの例 (図は5. [雑誌論文] ②より)

汎用的なフレームワークを提案した。

また、考えられるいくつかの活用例について、実際の教学データを用いた数値実験により挙動を確認した。その結果は以下の通りである。

- ①少なくとも実験で用いたデータについて、適切なパラメータを設定すれば、1年次前期の時点で2年後の退学者の60%程度が予測できるなどの結果が得られたことから、本フレームワークにもとづいて有用な予測精度をもつモデルが構築できる可能性がある。
- ②学士課程の進行にしたがって変化する動的な「予測される修学状態」を、ベイジアンネットワークを用いることによって、ひとつのモデルで統一的に表現できる。
- ③学生が能動的にコントロール可能な変数に着目して確率推論を行うことにより、未来に取り得る行動による修学状態の変化を予測することが可能であり、修学支援に援用するなどさまざまな応用可能である。

以上から、学士課程における修学状態の推移をベイジアンネットワークにより表現することで有用なモデルの運用ができる可能性があることがわかり、本フレームワークの機関レベルでの運用についての適用可能性が示唆された。

また、ここで提案した変数の分類方法 (Type 1: 性別、入試種別など変更不可能な個人属性に関する変数, Type 2: 成績や就職などの学習成果に関する変数, Type 3: 出席率や課題提出率、サークル所属の有無、在籍状態など、学生の行動の結果や状態を示す変数, Type 4: アカデミックアドバイザーや学習支援員からの連絡など、大学からの修学上の介入の有無を示す変数) は、本フレームワークに限らず、大学の保有する教学データを機械学習や統計モデルで扱う際に一般的に適用可能なものである。

[学会発表] の①②⑥では、このフレームワークを用いて、上述の実験では用いていなかった LMS (学習管理システム) のログから得られる特徴量を変数としたベイジアンネットワークの構築の方法と数値実験の結果を示した。これも、さまざまな大学がそれぞれの保有するデータに応じたモデル化を行うことの一つの例として、本フレームワークの可能性を示すものである。

## (2) 機械学習手法による教学 IR で用いる予測モデルの構築

3. の《課題1》《課題3》に対応する成果である。

5. の [学会発表] の⑦⑩⑪では、LMS (学習管理システム) ログを用いた成績予測を、いくつかの機械学習手法を用いて行った結果を発表した。図2に示すように、LMS のログから計算した特徴量 (および実際の授業の出席率) を説明変数とし、学期末の GPA (成績平均値) の高低を予測するという問題設定で数値実験を行った。図3は、1年次の前期において、授業15週の各時点での低成績に対する予測性能を示したものである (予測性能の指標である Precision (モデルの予測が正解である割合), Recall (実際に判別したいものが判別できている割合), F値 (Precision と Recall の調和平均) を用いている)。図3の結果はランダムフォレストを用いたときのものであるが、週ごとに性能が向上することや、3週目の時点で低成績の学生の40%程度が予測できることなどがわかる。また、ランダムフォレストのように説明変数の重要度が計算できるアルゴリズムを用いれば、重要度の比率の変化をみることで、教学施策の分析などにつながる可能性も示した。

図2のような予測モデルの枠組みは汎用的なものであり、説明変数、目的変数およびアルゴリズムに何をを用いるかによってさまざまな課題に対応できる。このような観点から、5. の [学会発表] の⑨では、教学 IR における予測モデル活用の一般的な枠組みを提案した。教学 IR において予測モデルを活用する場合に、各大学の保有するデータの違い、直面する課題 (退学予測、ハイパフォーマー予測等) の違いなどをふまえて、各大学の状況に応じた予測モデルを自

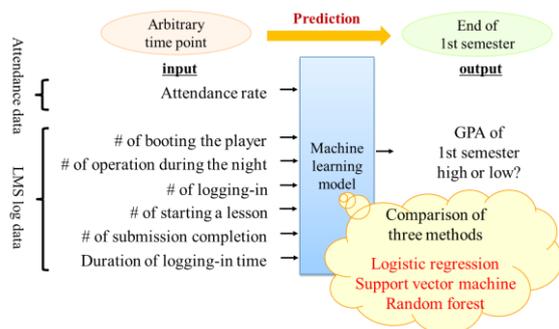


図2 LMS ログからの成績予測モデル

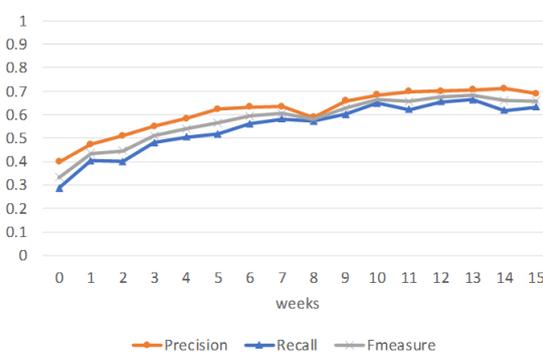


図3 成績予測の週ごとの性能  
(図は5. [学会発表] ⑦より)

ら構築できるようにするための枠組みをまとめたものである。ここでは、(1)で述べた教学データを用いる場合の変数の分類法も含まれる。当初想定していた先述の《課題1》～《課題3》においては、どのような場合にも共通して用いることが可能な技術的な知見を得ることを想定していたが、各大学のおかれた状況や背景、課題などが大きく異なることから、「モデルを構築するための枠組みそのもの」をまとめるほうが現実的で有用であるという考えにつながり、このような枠組みの提案に至った。

この汎用的な予測モデルの枠組みを用いて、5.の[学会発表]の⑤では、自己主導型学習レディネスに基づく成績予測を、5.の[学会発表]の③では、大規模な教学IRデータを用いた学習成果予測を行った結果を報告した。前者は、SDLRSとよばれる自己主導型学習レディネス尺度の質問紙調査の回答値を説明変数とし、個別授業の成績の高低を予測できるかを検討したものである。授業の特性、機械学習のアルゴリズムや分類の閾値の設定によって予測結果が異なり、より詳細な検討が必要ではあるものの、適切な設定を用いれば一定の性能をもって予測できることが示唆された。後者は、ある大学で入手可能な多数の教学IRデータをもとに170を超える説明変数を用意し、卒業時の学習成果を予測することについて数値的に検討した。またここでは、予測性能を向上させる説明変数の組合せを進化計算により自動的に選択する仕組みを提案した。この仕組みは予測性能を向上させることが示され、かつ大規模な説明変数から重要なものが絞られることから、可解釈性も向上することが期待できる。

このように本研究では、教学IRにおける予測モデルを具体的に活用するための枠組みと、これを実際にさまざまな課題へ適用した場合の数値的検討を行った。今後はこれらの結果をもとに本枠組みをより具体化し、広く教学IRにて活用されうるガイドラインとして発展させていきたいと考えている。

## 5. 主な発表論文等

[雑誌論文] (計2件)

- ①Nobuhiko KONDO and Toshiharu HATANAKA, “Modeling of Learning Process based on Bayesian Networks”, Educational Technology Research, Vol.41, pp.57-67, 2018. (査読あり)  
DOI: <https://doi.org/10.15077/etr.42136>
- ②近藤伸彦, 畠中利治, “ベイジアンネットワークによる修学状態推移モデルの構築”, 日本教育工学会論文誌, 41巻, 3号, pp.271-281, 2017. (査読あり)  
DOI: <https://doi.org/10.15077/jjet.41031>

[学会発表] (計13件)

- ①Nobuhiko Kondo and Toshiharu Hatanaka, “Estimation of Learning Process based on Bayesian Network using LMS Log data”, SICE Annual Conference 2018, pp.1049-1050, 2018. (査読あり)
- ②Nobuhiko Kondo and Toshiharu Hatanaka, “Bayesian Network about Students’ Learning Process using Bayesian Network with LMS Log Data”, 7th IIAI International Congress on Advanced Applied Informatics, pp.474-477, 2018. (査読あり)
- ③近藤伸彦, 松田岳士, 林祐司, 渡辺雄貴, 松河秀哉, 立石慎治, 椿本弥生, 山下英明, “自動選定した教学IRデータに基づくアカデミック・サクセスの予測”, 第27回教育学習支援情報システム(CLE)研究発表会, 2019.
- ④近藤伸彦, 畠中利治, “計算知能による教育/学習システムの構築に関する考察”, 第28回インテリジェント・システム・シンポジウム, pp.29-32, 2018.
- ⑤近藤伸彦, 松田岳士, 渡辺雄貴, 重田勝介, 加藤浩, “科目選択支援のためのSDLRSに基づく

- く成績予測に関する考察”，教育システム情報学会第43回全国大会，pp.341-342，2018
- ⑥近藤伸彦，畠中利治，“ベイジアンネットワークによるLMSログからの修学状態推定”，第62回システム制御情報学会研究発表講演会，2018.
  - ⑦Nobuhiko Kondo, Midori Okubo, and Toshiharu Hatanaka, “Early Detection of At-Risk Students Using Machine Learning Based on LMS Log Data”, 6th IIAI International Congress on Advanced Applied Informatics, pp.198-201, 2017. (査読あり)
  - ⑧近藤伸彦，畠中利治，“教学IRにおけるLMSログデータ活用の試み”，教育システム情報学会第42回全国大会講演論文集，pp.119-120，2017.
  - ⑨近藤伸彦，松田岳士，“教学IRにおける予測モデル活用の枠組み”，第6回大学情報・機関調査研究会MJIR，pp.42-47，2017.
  - ⑩近藤伸彦，大久保緑，畠中利治，“機械学習を用いた高リスク学生の早期発見”，計測自動制御学会第11回コンピューターショナル・インテリジェンス研究会講演論文集，pp.82-85，2017.
  - ⑪大久保緑，畠中利治，近藤伸彦，“ランダムフォレストによる課題非従事学生の検出”，計測自動制御学会第56回システム工学部会研究会，2017.
  - ⑫近藤伸彦，畠中利治，“隠れマルコフモデルによる修学状態のモデリングと学習支援”，計測自動制御学会システム・情報部門学術講演会2016講演論文集，pp.749，2016.
  - ⑬近藤伸彦，畠中利治，“LMSログデータからの修学モデル構築”，教育システム情報学会第41回全国大会講演論文集，pp.401-402，2016.

〔図書〕(計0件)

〔産業財産権〕

○出願状況(計0件)

○取得状況(計0件)

〔その他〕

なし

## 6. 研究組織

(1) 研究分担者 なし

(2) 研究協力者 なし

※科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。