

平成 27 年度 修士論文概要

主査	舟橋 健司	副査	北村 正	研究室	舟橋研究室
入学年度	平成 23 年度	学籍番号	23417596	氏名	西脇 雅弥
論文題目	教育支援を目的とした要注意学生の推定精度改善法 Estimation Accuracy Improvement Methods of Students Requiring Guidance for Education Support				

1 はじめに

情報通信技術が発展し、教育現場でも電子的にデータを扱うようになったことで、蓄積したデータを教育支援に有効活用する手法が数多く報告されている [1]。また、多くの大学において、留年・退学者の存在が大きな問題となっている。これらを背景に、本研究室では学生が留年・退学する前に指導を与えることを目的としたベイジアンネットワーク (Bayesian Network, BN) による要注意学生推定モデルを提案している [2]。要注意学生とは、1 年次前期と 1 年次後期における GPA がともに 1.0 以上であるにも関わらず、将来的に留年・退学する学生である (ただし、学業不振によらない退学者はこれに含めない)。このように、今後指導を与えるべき学生を要注意学生と定義することで、指導にかかる時間的なコストを削減している。従来より BN の構築には過去のある年度 A, B に本学のある学科に入学した 338 名の学生に関する 2 年分の成績データと打刻・出欠データから生成された変数を用いているが、本研究では推定精度改善のため「科目属性 (必修・選択) を考慮した変数」、「クラス内の偏差値を考慮した変数」、「成績と出席を関連付けた変数」を新たに導入した。また、従来研究では交差検証 (Leave-one-out) 法しか用いておらず、入学年度の違いに起因する学生の傾向変化に対する推定モデルの汎化能力の検証は行ってこなかった。そこで、Leave-one-out 法だけでなくホールドアウト検証も採用することで、推定モデルの汎化能力を確かめた。

2 データ概要

学生データはレコード形式で与えられており、この形式のままでは推定に不適なため、変換を施す必要がある。本研究では、従来より採用している各学生の講義区分別 GPA と、各成績評価の獲得数に加えて、

- 必修科目で不合格判定を受けた講義の数が多いと留年しやすいという予想から生成した、必修科目不合格数
- 講義の難易度を反映するため、クラス内の偏差値 s_l の値域によって集合 $S_{low} = \{s_l \mid s_l < 40\}$, $S_{mid} =$

$\{s_l \mid 40 \leq s_l < 60\}$, $S_{high} = \{s_l \mid 60 \leq s_l\}$ を定義して生成した、 $\#(S_{low})$, $\#(S_{mid})$, $\#(S_{high})$ 。

- 成績と出席の関連付けにより計算したクラス別出席数の偏差値 $s_{l,att}$ が 40 未満、標準偏差が 1 以上、かつ成績評価が失格でない講義の数 $\#(S_{att})$

の 3 つの変数を新たに導入した。すべての変数は開講期ごとに求めるため、2 年分で 68 個の変数が生成される。

3 検証実験

生成した変数を入力とする BN を用いて要注意学生の推定を行う。BN は 1 ノードあたりの状態数が増加すると計算効率が悪化するため、ウォード法により離散化を行うことで、各変数の状態数を 4 に抑えている。また、変数の次元が大きすぎると有用でない特徴量がノイズを引き起こし、結果に悪影響を与える場合があるため、CFS によって次元を削減している。

要注意学生かどうかの判定は BN の出力 (事後確率値) によって行う。すなわち、この事後確率値が閾値以上となる学生が要注意学生であると推定する。閾値は、より柔軟な推定を行うため、50%, 30%, 事前確率 (全学生に対する留年・退学者の割合) の 3 パターン用意し、正例クラスを要注意学生とみなした 2 値分類問題を解くことで再現率、適合率、F 値を計算する。本研究では、より多くの要注意学生を発見する能力 (再現率) と指導対象者を絞り込む能力 (適合率) の両方を評価するため、再現率と適合率の調和平均である F 値が高いモデルを良い推定モデルと位置付けている。また、推定を行う時期による推定精度の違いを確認するため「1 年次前期」、「1 年次後期」、「2 年次前期」、「2 年次後期」の 4 つの時期が終わるまでに得られる変数を用いた推定を行う。

3.1 従来手法との精度比較

従来手法を用いた場合の推定精度を表 1 に、新しい変数を加えて推定モデルを構築した場合の推定精度を表 2 に示す。モデル構築に用いたのは全体 338 名のうち GPA と退学理由のフィルタで残った 302 名分のデータであり、うち 25 名が要注意学生である。2 年次前期終了時の推定を除いて精度の改善がみられるため、新

しい変数は要注意学生の推定において有用な変数であるといえる。

表 1: 従来手法の推定精度

時期	閾値	再現率	適合率	F 値
1 年次前期	50%	32%	38%	0.348
1 年次後期	50%	52%	32%	0.394
2 年次前期	50%	72%	45%	0.554
2 年次後期	50%	72%	51%	0.600

表 2: 提案手法の推定精度

時期	閾値	再現率	適合率	F 値
1 年次前期	50%	40%	31%	0.351
1 年次後期	30%	64%	35%	0.451
2 年次前期	30%	64%	40%	0.492
2 年次後期	50%	80%	56%	0.656

3.2 入学年度別の精度検証

年度が変わると授業のカリキュラムや担当教員が変更されることがある。これにより、入学年度により将来的に要注意学生となる学生の傾向が変わることが考えられる。推定モデルがこの変化にどの程度対応できるかを調べるため、学生データを入学年度別に分けてホールドアウト検証を行った。まず、A 年度入学の学生データに対して Leave-one-out 法を適用したときの検証結果 (表 3) と A 年度入学の学生データで学習した推定モデルで B 年度入学の学生群における要注意学生をどの程度推定できるかを調査した結果 (表 4) を示す。

表 3: Leave-one-out (学習, 評価: A 年度)

時期	閾値	再現率	適合率	F 値
1 年次前期	8.4%	62%	30%	0.400
1 年次後期	8.4%	69%	47%	0.563
2 年次前期	30%	69%	47%	0.563
2 年次後期	50%	77%	56%	0.645

表 4: ホールドアウト (学習: A 年度, 評価: B 年度)

時期	閾値	再現率	適合率	F 値
1 年次前期	8.1%	58%	23%	0.326
1 年次後期	50%	58%	32%	0.412
2 年次前期	50%	42%	42%	0.417
2 年次後期	50%	75%	35%	0.474

次に、B 年度入学の学生データに対して Leave-one-out 法を適用したときの検証結果 (表 5) と B 年度入学の学生データで学習した推定モデルで A 年度入学の学

表 5: Leave-one-out (学習, 評価: B 年度)

時期	閾値	再現率	適合率	F 値
1 年次前期	30%	33%	27%	0.296
1 年次後期	50%	50%	50%	0.500
2 年次前期	8.1%	67%	36%	0.471
2 年次後期	30%	58%	39%	0.467

表 6: ホールドアウト (学習: B 年度, 評価: A 年度)

時期	閾値	再現率	適合率	F 値
1 年次前期	30%	23%	33%	0.273
1 年次後期	30%	62%	30%	0.400
2 年次前期	8.4%	85%	44%	0.579
2 年次後期	30%	77%	59%	0.667

生群における要注意学生をどの程度推定できるかを調査した結果 (表 6) を示す。

一般に、推定精度はホールドアウト法よりも交差検証法が上回ることが多いため、B 年度入学生の 2 年次における検証で逆の結果を得たことになる。このことから、入学年度の違いに対する推定モデルの汎化能力は希望の持てる結果になったといえる。

4 むすび

本研究では、要注意学生の推定に用いる変数を見直すことで、推定精度の向上を確認し、ホールドアウト検証を用いることで入学年度の違いに対する推定モデルの汎化能力を検証した。CFS の結果からは、クラス内偏差値が低い学生や相対的な出席数が少ない学生、1 年次において必修科目が不合格となる科目が多い学生は、将来的に要注意学生になりやすい傾向があることが確認できた。今後の課題は実用化に向けた更なる検証である。例えば、扱うデータを拡張して検証実験を行うこと、相関ルール分析を利用し要注意学生となる学生の傾向を新たに見出すことなどが挙げられる。

参考文献

- [1] 佐藤和彦: “修学指導支援のための学生の質的傾向を可視化する手法の検討”, 電子情報通信学会技術研究報告, Vol.110, No.334, pp.25-28, 2010.
- [2] M. Nishiwaki, H. Itoh, K. Funahashi: “A Method of Identifying Students Who Require Guidance Using Bayesian Network”, *Proc. IEEE GCCE*, pp.283-287, 2015.