

1 はじめに

近年の情報通信技術の発達により、大学では大量のデータが電子化され保持されている。本研究ではこのデータを活用するために、ロジスティック回帰分析を用いた修学指導支援を提案する。将来、留年・退学する可能性の高い学生を「要注意学生」と定義し、過去の留年・退学者の履修した講義の成績データや出席時の打刻データを用いて要注意学生となる学生の傾向の発見・予測を目的としている。従来研究ではページアンネットワークと交差検証 (leave-one-out 法)(以下、従来手法) による要注意学生の推定 [1] が行われてきたが、要注意学生の傾向の発見は行われていない。そこで本研究では、推定モデルが回帰式で簡単に表すことができるロジスティック回帰分析を用いて要注意学生の推定と傾向の発見を目指した。

2 データの概要

2.1 学生データの変換

本研究では、ある 2 つの年度に名古屋工業大学に入学した学生 338 名に関するデータを用いる。保存されているデータは大きく分けて 3 種類あり、講義の成績データ、講義室への入退出時の打刻データ、卒業研究着手および卒業に要した年数や退学した学生の退学理由といった修学データである。しかしこのままでは分析に用いるには不適なデータなため、従来より採用されていた以下の変数に変換を行った。

- 講義区分別 GPA
- 獲得成績数
- 必修科目不合格数
- 偏差値別講義数
- 出席低偏差講義数

2.2 要注意学生の定義

従来研究では将来、留年・退学してしまう学生のうち、本当に指導が必要な学生に十分な指導が行きわたるように次のように要注意学生を定義している。『1 年次前期・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・退学する可能性の高い学生。ただし、3 年で卒業研究に着手した学生は対象外、在学年数が 3 年に満たない退学者と、経済的理由による退学者のデータは除外する。』これにより予測対象者は 302 名、要注意学生は 25 名となった。

3 ロジスティック回帰分析による分析

3.1 変数選択

変数選択の方法として次の 3 種類の手法を採用した。手法 1 は変換データをそのまま説明変数とした。手法 2 は変換データを主成分分析し、各主成分を説明

変数とした。手法 3 は変換データを種類別に分け、それぞれを主成分分析し、各主成分を説明変数とした。いずれも説明変数を決定した後、ロジスティック回帰分析を行い、予測モデルの生成・評価を行った。

3.2 評価方法

分析結果の検証のため、再現率と適合率とその調和平均である F 値を比較指標とした。再現率は要注意学生を予測できた割合であり、これが大きいほど漏れなく要注意学生を予測できたことを意味する。適合率は指導対象と判定されたうちの要注意学生の割合であり、これが大きいほど指導コストの削減ができたことを意味する。

4 検証実験

手法 1・手法 2・手法 3 で実験を行った結果、手法 3 が一番優れていた。そこで、従来手法による分析結果と、本研究で用いた手法 3 による分析結果をそれぞれ表 1、表 2 に示す。すべての閾値で予測精度の向上が確認できた。さらに、用いた説明変数の回帰係数を求めることができたことにより、要注意学生の傾向を発見することができた。例えば、説明変数の意味が「成績の良さ」であり、回帰係数がマイナスであれば成績が良くなるほど要注意学生の確率が下がることを意味している。このように、回帰係数の符号と大きさで傾向を知ることができた。

表 1: 従来手法 (2 年次後期終了時点まで)

閾値	再現率	適合率	F 値
50%	80%	56%	0.656
30%	80%	48%	0.597
8.3%	80%	41%	0.541

表 2: 本研究手法 3 (2 年次後期終了時点まで)

閾値	再現率	適合率	F 値
50%	80%	87%	0.833
30%	96%	77.4%	0.857
8.3%	96%	55.8%	0.706

5 むすび

本研究では、ロジスティック回帰分析を用いることで要注意学生の推定を行い精度向上が確認できた。回帰係数から要注意学生の傾向を発見することができた。今後の課題として、予測モデルの汎用性の評価が挙げられる。よりデータを増やした分析の必要があるだろう。

参考文献

- [1] 西脇雅弥：“教育支援を目的とした要注意学生の推定精度改善法”，平成 27 年度名古屋工業大学修士論文，2015。