

## 1 はじめに

名古屋工業大学では、IC カード出欠管理システムにより学生の打刻情報 (IDm, 打刻時間, 打刻 IC カードリーダ番号) がサーバにリアルタイムで蓄積されている。しかし学生はデータを提供するだけで何かしらのフィードバックを得られておらず大量のデータを有効活用しているとは言えない。そこで従来の研究 [1] では学習レベルの予測をフィードバックすることを目的とし、そのためのデータに関して打刻情報の部分は出欠データ (出席, 欠席, 遅刻, 早退) しか使われていなかった。本研究では打刻情報より得られる打刻データに着目して学習レベル予測を行い、また打刻における特徴の分析を行った。

## 2 本研究で用いる手法

本研究では、学習レベルを予測する手法として計算過程がモデル構造より確認することができるベイジアンネットワークと決定木を採用している。またデータ量が多い場合、予測精度が下がる可能性があるためデータの削減方法として情報利得と主成分分析を採用している。これにより最も高い予測的中率を示す手法によるモデルから特徴を分析する。

## 3 ベイジアンネットワーク, 決定木による学習レベル予測

本研究では 2007, 2008 年度における学生 171 人の打刻データと修学データより学習レベルの予測を行っている。打刻データはそのままでは情報が少ないためデータの拡張を行い、各月の打刻回数・欠席回数、各曜日の分散・平均、祝日前後の欠席回数を求めた。なお前期は 4 月～7 月、後期は 10 月～1 月としている。また 1, 2 年複合の GPA を等距離で 5 段階 (S, A, B, C, D) に離散化したものを目的変数としている。

1, 2 年複合を予測するのに 2 年前期までのデータから予測するのと 2 年後期半学期までのデータから予測するのは後者のほうが少し予測精度が高かったが、本研究では早い段階で予測を行いたいので前者を用いて予測を行った。2 年前期までのデータから 1, 2 年後期を予測した結果、ベイジアンネットワークでは予測的中率 75.4386 %、決定木では予測的中率 70.1754 %を示した。さらに予測的中率を上げるため情報利得によるデータの削減および主成分分析による次元削減をした結果、ベイジアンネットワークでは情報利得によるデータの削減したとき 80.117 %、決定木では主成分分析による次元削減をしたとき 74.8538 %を示し予測的中率を上げることができた。これにより成績上位、中間、下位程度ならば予測可能であることが分かった。

## 4 特徴分析

ベイジアンネットワークおよび決定木により構築されたモデルを用いて特徴を分析した。1, 2 年複合の予測において情報利得を用いてベイジアンネットワークより構築されたモデルでは 1 年前期の GPA が 1.98 以下だと 1, 2 年複合の成績は中間より下位になる傾向が見られた。また 1 年前期の GPA が 3.13 より上だと 1, 2 年複合の成績は中間より上位になる傾向が見られた。全体的に 2 年前期金曜における打刻回数の平均が 2.31 より上だと成績が良くなる傾向があることがわかった。

主成分分析を用いて決定木により構築されたモデルは主成分データは使われず成績データのみで構築されていることがわかった。これは細かい部分の分類について打刻データのとある変数が使われている可能性があるが、主成分分析により複数の変数が統合されてしまったため細かい分類ができなくなったことが原因として考えられる。

ここである曜日における打刻回数の平均について考えてみると平均が低いといっても講義をあまり取っていない、あるいは講義を休んでいる、あるいは遅刻が多いなどさまざまなパターンが考えられる。よって平均のかわりにある曜日における打刻回数の最頻値とある曜日における欠席回数を用いることである程度パターンを絞ることにした。結果、1, 2 年複合におけるベイジアンネットワークによる予測的中率が 80.117 % に対し、平均を置き換えて予測した結果 80.7018 % と予測的中率を上げることができた。これより打刻データが成績と関係があるということが言えた。

## 5 むすび

打刻データを用いてベイジアンネットワークと決定木により予測した結果、成績上位、中間、下位程度の予測ならば可能であることを示した。また打刻データを情報利得、主成分分析を用いて削減することにより予測的中率を上げることができた。構築されたモデルより打刻データが学習レベル予測に関わっていることを示し、説明変数をさらに詳しく調べる置き換えることにより予測的中率を上げることができた。今後は他大学やカリキュラムの違う学年に対応できるよう大学の人数による相対値、必須科目や選択科目の単位数を説明変数として取り入れることにより打刻が成績とどう関係しているか検証していきたい。

### 参考文献

- [1] 伊藤宏隆, 舟橋健司, 内匠逸, 松尾啓志, “IC カード出欠データと CMS 学習データを用いたデータマイニング”, メディア教育研究, 4, 2, pp.15-21, 2008