

学生の修学データを用いた修学指導支援システムの検討

伊藤 宏隆† 舟橋 健司† 山本 大介† 内匠 逸† 松尾 啓志†

名古屋工業大学 情報基盤センター†

1. はじめに

名古屋工業大学では、双方向型修学指導システムの構築を目的として、2007年4月にICカード出欠管理システムとコースマネージメントシステムを導入した[1, 2]。ICカード出欠管理システムでは、ICカード化された学生証により学生の出欠状況を把握する。学生には、すべての授業における出欠打刻を義務づけている。コースマネージメントシステム(CMS)は、e-Learningを支援するシステムである。

システム導入から2年半が経過し、出欠管理システムには、約370万件もの大量の出欠データが蓄積されている。また、CMSにも大量の学習データが保存されている。著者らは蓄積されているデータを有効に活用し、修学指導支援に役立てることを計画している。例えば、データマイニングにより過去の学生のデータの傾向を分析することで、学生の修学傾向を把握ができる。また、過去のデータをもとに学生の将来の成長可能性を予測することで、落第候補者を早期に発見することが可能となる。データマイニングや予測により、早期にきめ細かな修学指導を行うことが可能となる。教員側は学生からデータの提供を受け、データ分析の結果を学生側に修学指導という形で還元する。これが双方向型修学指導支援システムである。

著者らは、これまでにある授業の出欠データとCMSでの課題提出の有無のデータ、課題提出時間のデータと成績データを用いてデータマイニングを行い、有用な結果を得ている[3]。また、過去の学生のデータを用いて、成績予測実験も行っている[4]。

今回は、某学科の学生の2年生の前期における取得単位数予測実験を行った。蓄積されている出欠データと単位修得状況をもとに、将来の単位取得状況を予測することで、早期に修学指

導を行う支援システムを構築するための実験である。本論文では、取得単位数予測実験について論述する。

2. 取得単位数予測実験

2. 1. 実験概要

某学科の2年生の2009年前期における取得単位数を予測する。予測に用いるサンプルデータは現3年生の1年前期における出席率と取得単位数、1年後期における出席率と取得単位数、2年前期における出席率と取得単位数である。そして、現2年生の1年前期における出席率と取得単位数、1年後期における出席率と取得単位数、2年前期における出席率から2年前期の取得単位数を予測し、実際の取得単位数と比較することで実験結果の検証を行う。なお、出席率は受講登録をしている全授業の出席率である。3年生、2年生の学生の人数はそれぞれ172名と168名である。

予測に用いたソフトウェアはMicrosoft SQL Server 2005とMicrosoft Office Excel 2007+SQL Server 2005 Data Mining Add-ins for Office system[5, 6]であり、Microsoftのデータマイニングアドインを使用した。

成績予測は、マイニングツールの「例の全体適用」機能を用いて行った。図1のように、予

| | 1前 出席 | 1前 単位 | 1後 出席 | 1後 単位 | 2前 出席 | 2前 単位 |
|-------------|----------|----------|----------|----------|----------|----------|
| 3 年 生 | 0.86 | 22 | 0.77 | 25 | 0.92 | 26 |
| | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| | 0.88 | 25 | 0.84 | 27 | 0.86 | 25 |
| 2 年 生 | 0.88 | 23 | 0.89 | 24 | 0.88 | |
| | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | |
| | 0.91 | 24 | 0.90 | 25 | 0.89 | |

図1. 例の全体適用

Examination of education support system using student's schooling data.

† Hirotaka Itoh, Kenji Funahashi, Daisuke Yamamoto, Ichi Takumi and Hiroshi Matsuo

† Information Technology Center, Nagoya Institute of Technology

測のサンプルデータとして用いる 2 年前期の取得単位数も入った 3 年生のデータと 2 年前期の取得単位数を予測のために空白にした 2 年生のデータをつなげて同じテーブルとする。例の全体適用を 2 年前期の取得単位数に対して行くと、2 年前期の取得単位数以外の列の各項目が 2 年前期の取得単位数にどのように影響を与えているかを分析し、2 年前期の取得単位数が空欄の場合、分析を基に推論し結果を埋めてくれる。

2. 2. 実験結果

実験結果として、2 年生の 2 年前期の取得単位数の予測値と実値との誤差の絶対値の分布を表 1 に示す。誤差の平均値は 2.86 であり、約 90% が誤差 5 以内である。

通常、半期には 25 前後の単位数が取得でき、誤差の少ない場合はコンスタントに単位が取得できているパターンが多い。

誤差 10 以上の学生の出席率と取得単位数のパターンに注目してみると、2 つのパターンに分かれていることがわかった。一つ目のパターンは、1 年前期、後期に取得した単位数は多いが、2 年前期に取得した単位数が少なかった場合で、もう一つのパターンは 1 年前期、後期の取得単位数は少ないが、2 年前期の取得単位数が多い場合である。このようなパターンを分析すると、取得単位数が少ないときには、出席率が低くなっている傾向が見られるので、予測は可能と思われるが、これらは例外的なパターンで、サンプルデータの中にもこのようなパターンは少ない

表 1. 取得単位数予測値と実値の誤差分布

| 誤差 | 件数 | 比率 |
|----|-----|-------|
| 0 | 25 | 14.9% |
| 1 | 34 | 20.2% |
| 2 | 28 | 16.7% |
| 3 | 33 | 19.6% |
| 4 | 18 | 10.7% |
| 5 | 13 | 17.7% |
| 6 | 3 | 1.8% |
| 7 | 4 | 2.4% |
| 8 | 1 | 0.6% |
| 9 | 3 | 1.8% |
| 10 | 1 | 0.6% |
| 11 | 1 | 0.6% |
| 13 | 3 | 1.8% |
| 14 | 1 | 0.6% |
| 計 | 168 | 100% |

ため、予測が困難になっているものと思われる。例外的なパターンについては、それだけを抽出して別に予測を行う必要があるかもしれない。

なお、1 年前期からずっと半期の取得単位数が少ないパターンもあるが、これは予測誤差が少なかった。パターンとしての件数は少ないが、コンスタントに多くの単位が取得できているパターンと傾向が似ているため、予測ができているものと思われる。

3. まとめ

IC カード出欠データ、過去の単位取得状況を用いて将来の取得単位数の予測実験を行った。出席率と取得単位数のパターンにより、予測値と実値の誤差が異なることが判明した。誤差が大きい場合のパターンは、例外的なパターンでサンプルも少ないため予測が難しい。逆に例外的なパターンとして、予測前にクラスタ分析などで抽出することも考えられる。他の学科のデータも使うなどサンプルデータを多く用いることで精度の向上が期待できる。また、ニューラルネットワークなどによる予測も検討している。

今後は、就学指導支援システムの実用化に向けて、他の学科の学生データの整理と更なる実験を進める予定である。

参考文献

- [1] 伊藤宏隆, 舟橋健司, 中野智文, 内匠逸, 松尾啓志, 大貫徹, “名古屋工業大学における Moodle の構築と運用”, メディア教育研究, Vol. 4, No. 2, pp. 15-21 (2008)
- [2] 伊藤宏隆, 舟橋健司, 中野智文, 松尾啓志, 内匠逸, 大貫徹, “コースマネジメントシステム Moodle と IC カード出欠システムとの連携”, 平成 19 年度電気系学会東海支部連合大会講演論文集, 0-376 (2007)
- [3] 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”, 日本 e-learning 学会誌, Vol. 9, pp. 95-108 (2009)
- [4] 伊藤宏隆, 舟橋健司, 山本大介, 内匠逸, 松尾啓志, “出欠データと学習データを用いた学生の成績予測”, 日本 e-Learning 学会 2009 年度秋季学術講演会論文集, pp. 52-57 (2009)
- [5] <http://www.microsoft.com/japan/sql/dmaddin/default.aspx>
- [6] 平井明夫, 松井浩輔, “ひと目でわかる Microsoft Office Excel 2007 データマイニングアドインを使用したビジネスデータ簡単分析術”, 日系 BP ソフトプレス (2007)