

ベイジアンネットワークを用いた学生の成績予測

伊藤 宏隆^I 伊藤 圭佑^{II} 舟橋 健司^{III}
 名古屋工業大学^I 名古屋工業大学^{II} 名古屋工業大学^{III}
 情報基盤センター 工学部情報工学科 情報基盤センター

1. はじめに

名古屋工業大学では、2007年4月に教育支援システムとして、ICカード出欠管理システムとコースマネジメントシステムを導入した。ICカード出欠管理システムは、ICカード化された学生証により学生の出欠を管理する。コースマネジメントシステム（CMS）は、情報技術やインターネットを使ったe-Learningを支援するシステムであり、教材の作成支援、課題の提出管理、小テストの実施、学生の受講管理を行う機能を持っている。導入から約5年が経過し、出欠管理システムには学生の授業の出欠情報が蓄積され、コースマネジメントシステムには、学生の提出した課題ファイルや小テストの回答などの学習データが蓄積されている。通常、これらの出欠データや学習データは成績評価の際のデータとして用いられる。著者らは、これらのデータを成績評価のデータとしてだけではなく、早期の学習指導用のデータとして活用することを計画した。学生の理解度や学習意欲を適切に把握することは容易ではなく、理解度や意欲の高い学生に高度な学習機会を与えることは困難であった。また、理解度の低い学生が発見されるのは学期末の成績算出時で手遅れとなっていた。そこで、著者らは学期途中において過去の受講生の学習データと成績データを活用し、現在の受講生の学習データと比較することで現在の受講生の成績が予測可能となり、学生の理解度が把握できれば、理解度に合わせた学習指導が可能となると考えた。中間テストによる理解度把握も考えられるが、過去のデータを基にした予測では、学生の成長可能性や落第可能性までも含めた予測ができる。

著者らはこれまでに、ある授業の出欠データや学習データを用いてデータマイニングを行い、出欠データ、学習データと成績の関係を分析し、有意義な結果を得ている[1]。また、落第候補者の早期発見を目的として、学期途中までの出欠や学習データから成績を予測する実験を行っている[2,3]。さらには、著者らはこれらの実験の発展として、予測した成績を基に教育指導の実践を行いその効果を検証し、有用な効果を得ることができた[4]。

本研究では、学生の学習状況の各要素間の関係を条件付き確率で表すことができる確率推論モデルであるベイジアンネットワークを用いて成績予測を行うことを試みた。本論文では、ベイジアンネットワークの構成と成績予測実験の結果について述べる。

2. ベイジアンネットワーク

ベイジアンネットワークは、確率変数間の定量的な関係を条件付き確率で表す。ベイジアンネットワークは図1のようにグラフ構造で表され、確率変数をノード（○）、確率変数間の関係をリンク（→）で表す。ベイジアンネットワークの特徴の一つが確率推論である。ある確率変数にあらたなデータが与えられたとき、図2のようにネットワーク上の確率変数の確率の伝搬が行われ、目的変数の事後確率が求められる。

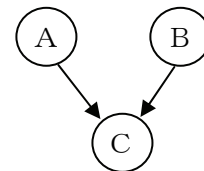


図1. ベイジアンネットワーク

3. ベイジアンネットワークを用いた成績予測

ベイジアンネットワークを用いてある授業の成績の予測実験を行った。今回、実験に用いたデータは1年生の情報基礎科目の授業のデータである。最終的な成績評価は、出欠、課題の提出と授業の

Forecasts of Result using Bayesian Network

I Hirotaka Itoh, Information Technology Center, Nagoya Institute of Technology

II Keisuke Itoh, Department of Computer Science, Faculty of Engineering, Nagoya Institute of Technology

III Kenji Funahashi, Information Technology Center, Nagoya Institute of Technology

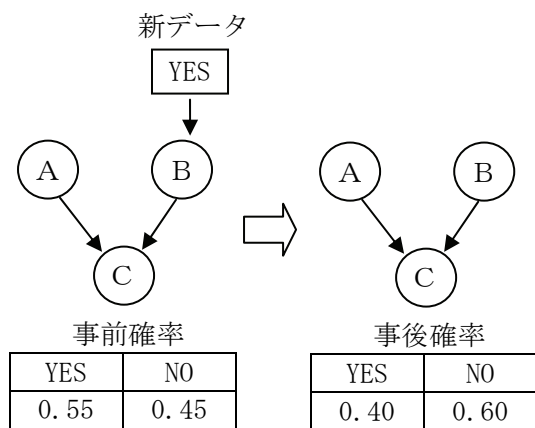


図2. ベイジアンネットワークの確率推論

終盤で課される最終課題 2 題の採点結果で判定している。使用したデータは出席回数，課題提出回数，課題の提出値の合計である。課題の提出値とは課題提出の早さを表し，課題提出受付開始時を 1 とし，課題提出締め切り時を 0 としている。課題提出が早いほど提出値は高くなる。確率変数として以下の項目を用いた。

- ・成績面
 - ①成績が S もしくは A (成績優秀者)
 - ②成績が B もしくは C (合格)
 - ③成績が D もしくは X (不合格)
- ・出席面
 - ①1 回目～8 回目の出席が 7 回以上
 - ②9 回目～12 回目の出席が 3 回以上
 - ③13 回目～15 回目ですべて出席
 - ④13 回以上出席
 - ⑤連続して 2 回欠席したことがある
- ・課題提出
 - ①すべての課題を提出
 - ②2 回連続課題未提出となったことがある
 - ③課題提出値の平均が 0.8 以上
 - ④課題提出値の平均が 0.3 以下
 - ⑤ (前半平均提出値) - (後半平均提出値) が 0.1 以上
 - ⑥提出値の分散が 0.1 以上

決定木を用いて変数間の関係进行分析し，ベイジアンネットワークを構成した。学期途中での予測を行うものとして第 8 回終了時までのデータを用いて予測を行った。50 人分の学生のデータのうち，1～40 番の学生のデータを用いてベイジアンネットワークを学習し，41 番～50 番の学生のデータを用いて成績を予測した。第 8 回までにに関するノードにのみ情報を入力した。結果を表 1 に示す。

表 1. 成績予測結果

	予測値	実成績
4 1	DX	DX
4 2	DX	DX
4 3	BC	BC
4 4	BC	SA
4 5	SA	BC
4 6	BC	SA
4 7	BC	SA
4 8	SA	BC
4 9	—	SA
5 0	SA	SA

正答率は 40% で，予測不可が 1 つ発生した。不合格者 (DX) の判定には有効ではないかと考えられる。成績優秀者 (SA) と合格者 (BC) の実成績との逆判定が多くみられる。データを精査したところ，課題提出時間の傾向が起因していた。課題の提出が遅くなると予測成績は良くなるが，実際の成績はその反対であった。

4. まとめ

ベイジアンネットワークを用いて学生の成績予測を行った。不合格者の判定には有効と考えられる。成績優秀者と合格者の判定にはまだまだ熟考が必要である。今後の課題は，ネットワークの構成方法についての検討である。

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”，日本 e-learning 学会誌，Vol. 9, pp95-108 (2009)
- [2] 伊藤宏隆，堀江匠，舟橋健司，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修学傾向分析，情報処理学会第 71 回，全国大会講演論文集第 4 分冊，pp. 357-358 (2009)
- [3] 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)
- [4] 伊藤宏隆，舟橋健司，山本大介，内匠逸，松尾啓志，齋藤 彰一，“過去のデータに基づく成績予測による教育指導の実践”，ICT 活用教育方法研究，Vol. 13 No.1 pp. 41-45 (2010)