

1 はじめに

近年、様々な教育機関で、学生に関する情報が電子的に扱われるようになってきている [1]. 蓄積された修学データは、教員が学生の成績を決定する指標などに用いられていた。しかし、蓄積された大量の修学データをより有効に活用したいという考えから、データ提供元である学生側に、データから予測された学生の修学傾向に応じた学習指導を与えるという構想が多方面で提唱されている。そこで本研究では、予測手法にベイジアンネットワークの採用を提案し、ベイジアンネットワークによる予測モデルを用いた学習指導を検討する。

2 ベイジアンネットワーク

ベイジアンネットワーク [2] は、未知事象の予測や推論などに用いられる手法で、確率変数、有向グラフ構造、変数間の条件付き確率の 3 つの要素で定義される。ベイジアンネットワークに最適な予測モデルを構築するには、3 要素の最適な決定、採択をしなければならない。しかしこの処理は、人手によるものでは大変手間であり困難でもある。そこで、3 要素の決定にデータマイニングの考えを取り入れることで問題を解消している。

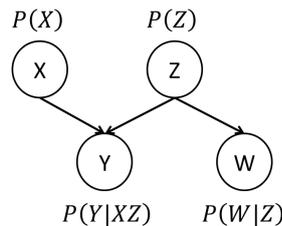


図 1: ベイジアンネットワークの例

3 最適な予測モデルの構築

3.1 用いるデータとその拡張

本研究では、講義成績のデータを 2 年分用いている。このデータから様々な変数を定義した。例えば、講義の成績から各学生の GPA 値を算出している。変数定義されたデータをもとに最適な予測モデルを構築している。

3.2 変数の取捨と離散化

定義された変数は多数あり、数値の形式であるものが多い。変数の取捨と離散化に、CFS[3] の概念と決定木の生成を適用した。CFS はエントロピーと情報利得でその値を計算できる属性選択指標である。決定木は 2 分木と各ノードに設置されたルール記述によって定義され、事象の分類などに用いられている。これらデータマイニングの手法で、最適モデルの構築を支援した。

4 構築されたモデルの検証

学生の修学傾向を予測するためのモデルを構築した後、leave one out 法による的中率の評価により、モデルの検証を行った。本研究において構築されたモデルの 1 つに、2 年後期の GPA 値が偏差値 40 を下回るかどうかを予測するものがある。このモデルにより、成績が低迷してしまいそうな学生を救済することを試みる。その精度評価を以下の表 1 に示す。全体として高い予測精度を示しており、また、「YES」に該当する学生、つまり学習指導を与えられる対象である学生が 25 人いるが、その内の 18 人に正しく指導が与えられることが結果からも分かる。また、ベイジアンネットワークの出力は表 2 のような形式であるため、表 2(B) のような確率値出力であれば、予測分類の結果では「NO」でも、「YES」である可能性、つまり、未来の成績低迷を示唆するような表現が可能となる。このような表現の柔軟性により、多くの学習指導対象者を拾い上げることができる。このことから、ベイジアンネットワークの予測手法は、本研究の学習指導案に有効であると言える。

表 1: 評価結果

		予測出力	
		YES	NO
実 際	YES	18	7
	NO	4	142
的中率		93.56%	

表 2: 2 つの出力例

(A)	YES	NO
		1.8%
(B)	YES	NO
		46.7%

5 むすび

ベイジアンネットワークが、学習指導のための予測に有効であることを示した。また、ベイジアンネットワークのモデル構築の際、データマイニングの手法を取り入れることで、より精度の高い予測モデルを構築することができた。今後は、予測精度の更なる向上と、それぞれの予測モデルは別年度のデータを考慮していないため、本研究の検証を含めた考慮をしていきたい。

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, 4 巻, 2 号, 15-21 (2008)
- [2] 本村陽一, “ベイジアンネットワーク:入門からヒューマンモデリングへの応用まで”, 行動計量学会セミナー資料 (2004)
- [3] Duan, S. and Babu, S, “Processing Forecasting Queries”, Proc. 2007 Intl. Conf. on Very Large Data Bases, pp.711-722 (2007)