

# 学修ライフログからの確率モデル構築による修学支援

近藤 伸彦† 島中 利治‡

大手前大学 CELL 教育研究所†

大阪大学大学院情報科学研究科‡

## 1 はじめに

近年、高等教育における大規模データ活用の必要性が認識され、大学運営における意思決定をエビデンスベーストで支援する機能である IR(Institutional Research)や、入学前から卒業後までの包括的修学支援である EM(Enrollment Management)など、データに基づく教育改善がさまざまな文脈で研究されている。

なかでも、データマイニングや機械学習の知見を用いた手法が近年提案されつつあり[1][2]、筆者もこれまでに、IR や EM への活用を見据えて、データマイニング手法を用いた学修成果や修学状況の予測を試みている[3]。ここでは、ニューラルネットワークと決定木を用いてモデル構築を行ったが、教学上の意思決定や教職員による学生支援といった実際の活用場面を想定すると、非線形性と不確実性を扱うことができるとともに、内部構造の可解釈性にすぐれた確率モデルの構築が望ましい。

そこで本研究では、不確実性を定量的に表現でき、構造を解釈しやすいベイジアンネットワークに注目し、学生の学修ライフログからこれを構築することを考える。本稿では、ベイジアンネットワークによる数値実験のようすと、修学支援に向けた活用についての展望を述べる。

## 2 学修ライフログからのベイジアンネットワークの構築

### 2.1 学修ライフログ

入学前から 4 年間の学士課程を通してはさまざまなタイプのデータが存在する。入学までの個人属性をはじめ、授業の出欠席状況、LMS 等のシステムログなど、多岐にわたる大規模なデータを時系列に整理し、学修ライフログとしてこれを活用することを本研究では想定する。修学支援の観点からは、学習困難に陥っている学生や、中途退学の可能性がある学生を発見し、適切な支援を行うことが大きな目的となる。

### 2.2 ベイジアンネットワーク

ベイジアンネットワークは、確率変数をノードとする非循環有向グラフネットワーク構造と条件付確率パラメータ集合による離散確率分布の近似モデルである[4]。ノード間の有向リンクは確率変数間の依存関係を示し、その依存関係は条件付確率分布により定量的に表される。

ベイジアンネットワークはその確率モデルとしての表現力の高さから、不確実性を含むモデリングの有用な手法として知られ、障害診断、パターン認識、データマイニング、ヒューマンモデリングなど、幅広く応用されている[5]。

### 2.3 ベイジアンネットワークによる学修状態のモデル化

学修ライフログには、デモグラフィック属性を含め、大規模な変数が存在する。この変数間の依存関係を適切に表現するベイジアンネットワークを構築し、あらゆる学生の実際のデータをもとに学習を行うことで、学士課程にわたる学修過程に関するヒューマンモデリングが可能になると考えられる。ネットワークの構築においては、教職員の経験による事前知識を活用することもできる。構築したベイジアンネットワークにより、学士課程の任意の時点において、それまでのデータをもとに、今後の学籍異動や出欠席、成績などの状況を確率的に推定するなど、さまざまな活用方法が考えられる。

### 3 数値実験

本研究では、X 大学（総合的教養教育型の中小規模私立大学）における 2009 年度および 2010 年度入学生の学修ライフログを用いた。学生数はそれぞれ 794, 817 である。学修ライフログから着目した変数は表 1 の通りであり、これらの変数からなる 13 次元のデータを、各学生について学修ライフログから抽出した。

グラフの構造は、イベントの順序と経験的知識から図 1 のように定めた。このグラフのもとで、2009 年度生のデータから条件付確率を求めた。GPA や出席率のような連続値の変数と単位数については、ここでは等頻度になるよう複数のクラスに離散化した。そのように作成したベイジアンネットワークに対して、2009 年度生お

Enrollment Management by a Probabilistic Model based on the Students' Lifelog

† Nobuhiko KONDO, CELL Institute for Educational Development, Otemae University

‡ Toshiharu HATANAKA, Graduate School of Information Science and Technology, Osaka University

よび 2010 年度生のデータを入力し、「X<sub>13</sub>:学籍状況」を目的変数としてパターン分類を行った。ここで、「X<sub>13</sub>:学籍状況」は、(1) 4 年間のうちに退学・除籍（以下「退学」）、(2) 4 年間で卒業（以下「卒業」）、(3) 4 年で卒業できず留年（以下「留年」）の 3 クラスである。分類結果を表 2、表 3 に示す。なお、「適合率」はモデルの予測が正解である割合、「再現率」は真のクラスが正しく分類されている割合を表わす。

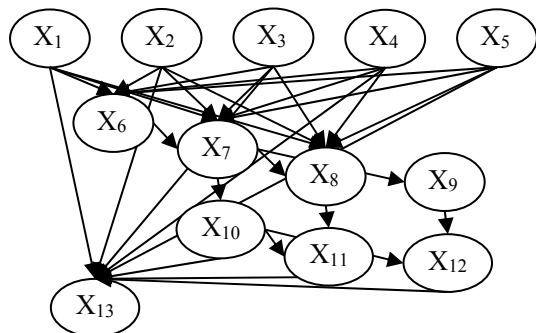


図 1 ベイジアンネットワークの構造

表 1 使用した変数

変数	値
X <sub>1</sub> : 性別	名義尺度, 2 クラス
X <sub>2</sub> : 学部	名義尺度, 3 クラス
X <sub>3</sub> : 入試種別	名義尺度, 7 クラス
X <sub>4</sub> : 出身校課程	名義尺度, 5 クラス
X <sub>5</sub> : 出身校評定値	連続値, [0, 5]
X <sub>6</sub> : 導入教育 (入学前学習, 新入生オリエンテーション) 参加率	連続値, [0, 1]
X <sub>7</sub> : 1 年次必修科目出席率(春学期)	連続値, [0, 1]
X <sub>8</sub> : 1 年次春学期末通算 GPA	連続値, [0, 4]
X <sub>9</sub> : 1 年次春学期末修得単位数	離散値, [0, 20]
X <sub>10</sub> : 1 年次必修科目出席率(秋学期)	連続値, [0, 1]
X <sub>11</sub> : 1 年次秋学期末通算 GPA	連続値, [0, 4]
X <sub>12</sub> : 1 年次秋学期末修得単位数	離散値, [0, 48]
X <sub>13</sub> : 学籍状況	名義尺度, 3 クラス

表 2 2009 年度生のデータに対する結果

	予測	実際のクラス			適合率
		退学	卒業	留年	
退学		128	36	11	0.731
卒業		45	533	24	0.885
留年		0	0	17	1.000
再現率		0.740	0.937	0.327	

表 3 2010 年度生のデータに対する結果

	予測	実際のクラス			適合率
		退学	卒業	留年	
退学		118	87	46	0.470
卒業		46	479	41	0.846
留年		0	0	0	0.000
再現率		0.720	0.846	0.000	

表 2 が示す 2009 年度生についての結果は、トレーニングデータに対する分類精度にあたる。表 3 が示す結果は、2010 年度生のデータによりモデルの汎化性を検証したものに当たる。「留年」は再現率・適合率ともに 0 で、「退学」も適合率は 2009 年度生より大幅に低い。「卒業」の再現率・適合率は、ともに 2009 年度生よりやや低いが、85%ほどの水準は保っている。4 年間で卒業するという「順調な」ケースは、退学や留年などの「順調でない」ケースよりもモデル化しやすいことが示唆され、入学年度に左右されにくい情報を獲得できる可能性がある。

#### 4 修学支援への活用

本実験では、予測対象として学籍状況に着目したが、GPA や単位数など、学修成果に着目することも同一のモデル上で可能である。さらには、変数を 4 年間にわたって追加することで、1 年次終了時、2 年次終了時など、より細かく適時的な退学リスクや学修状況の推定も可能になると考えられる。

ベイジアンネットワークの特性を生かし、確率伝搬によるシミュレーションを行うことも可能である。例えば、ある時点までのデータが得られたとき、その後の状況がどのような確率で生起するかをシミュレーションにより予測することも考えられる。または、春学期の時点で出席状況や成績がふるわなかった場合に、仮に秋学期の出席率が良くなったときに退学する確率がどう変化するか、といったことも推定できる。

今後はこうした観点から、実際の修学支援に生かせるモデル構築について検討していきたい。

#### 参考文献

- [1] 雨森聡, 松田岳士, 森朋子, “教学 IR の一方略: 島根大学の事例を用いて,” 京都大学高等教育研究, 第 18 号, pp.1-10, 2012.
- [2] S. K. Yadav, B. Bharadwaj, S. Pal, “Data Mining Applications: A Comparative Study for Predicting Student’s Performance,” Int. J. of Innovative Tech. & Creative Eng., Vol.1, No.12, pp.13-19, 2012.
- [3] 近藤伸彦, 畠中利治, “大規模学修データからのライフログ抽出とその活用,” 教育システム情報学会第 39 回全国大会, pp.305-306, 2014.
- [4] 植野真臣, ベイジアンネットワーク, コロナ社, 2013.
- [5] 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,” 日本行動計量学会第 7 回春のセミナー, 2004.