

成績予測のための過去の受講生のデータのクラスター分析

伊藤 宏隆^I 伊藤 暁人^{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、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ージメントシステムを活用している。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、コースマネー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。

これまでに、著者らは成績を予測し、予測成績を基に高レベル、中レベル、低レベルの3つのグループに分け、レベル別の課題を実施した[1]。この方法では、予測成績は各学生に対し一つであり、教員のみが予測成績を把握し、学生には予測成績は伝えていない。また、予測にはオフラインの予測ツールを用いている。将来的には、学生に直接的に予測を伝えることを計画

している。そのためには、オフラインのツールではなく、オンラインの Web アプリケーションで用いることができる手法が必要であり、学生に伝える予測成績も一つではなく、“現在の学習状況では、Sの確率60%、Aの確率30%、Bの確率10%です”のように複数の成績可能性を示す形式がよいのではないかと思いついた。

本論文では、Web アプリケーションで利用可能で、複数の可能性を提示する方法を提案する。提案方法は、過去の受講生を学習状況を基にクラスター分析し、いくつかのクラスターに分割する。クラスター分析には、Web アプリケーションで利用可能な k -平均法を用いる。各クラスター毎に成績分布を作成し、その成績分布を成績確率とし、現在の受講生の学習データがどのクラスターに入るかで成績確率を求める。次節で、成績予測のための過去の受講生データのクラスター分析について論述する。

2. 過去の受講生データのクラスター分析

今回、クラスター分析を行った授業は1年生の情報基礎科目である。最終的な成績評価は、出欠、課題の提出と授業の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。過去の受講生102名の第8回終了時までデータを用いて分析を行った。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。第8回終了時において、出席は8回、提出すべき課題は5個であった。

クラスター分析の結果を表1に示す。クラスター数は7個とした。表1より、クラスター1では、成績確率が、S=17%、B=33%、C=50%となる。表2にクラスター1に分類された受講生の学習データを示す。表2よりクラスター1の受講生は、課題はすべて早期に提出しているが、欠席が数回ある。他の例として、表3にクラスター7の受講生のデータを示す。表3よりクラスター7の受講生は、ほぼ出席し、課題もすべて提出してい

Cluster Analysis of Past Student's Data for Forecasts of Result

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

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

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

表 1. クラスタ分析結果

クラスタ	成績						総計
	S	A	B	C	D	X	
1	1 17%	0 0%	2 33%	3 50%	0 0%	0 0%	6
2	0 0%	3 43%	1 14%	2 29%	0 0%	1 14%	7
3	2 18%	3 27%	3 27%	3 27%	0 0%	0 0%	11
4	16 59%	7 26%	2 7%	2 7%	0 0%	0 0%	27
5	15 42%	10 28%	7 19%	4 11%	0 0%	0 0%	36
6	0 0%	0 0%	0 0%	0 0%	0 0%	10 100%	10
7	3 60%	2 40%	0 0%	0 0%	0 0%	0 0%	5

表 2. クラスタ 1 の受講生の学習データ

番号	出席回数	課題提出	提出値計	成績
37	6	5	4.275	B
75	5	5	4.793	B
67	4	5	4.852	C
68	4	5	4.653	C
74	3	5	4.652	C
86	6	5	4.852	S

表 3. クラスタ 7 の受講生の学習データ

番号	出席回数	課題提出	提出値計	成績
48	8	5	1.150	S
51	8	5	2.443	S
95	7	5	1.520	S
71	7	5	1.521	A
72	7	5	2.233	A

るが、提出値計が 1~2.5 で課題提出が遅いという特徴がある。各クラスターの特徴をまとめたものを表 4 に示す。特徴ごとにクラスターにまとまっているが、成績はその後の 9 回以降の出欠と課題提出、特に最終課題 2 題の評価によって変わってくる。例えば、クラスター 7 の受講生の 9 回以降の学習データを示す。参考までに成績に大きく影響する最終課題の結果も付ける。表 5 より、9 回以降の出席回数と最終課題の評価によって成績が S と A にはっきり分かっている

表 4. 各クラスターの特徴

クラスタ	特徴
1	課題はすべて早期に提出しているが、欠席が数回ある
2	数回欠席があるか、あるいは数回課題提出をしていない
3	欠席が多い
4	ほぼ出席、課題もすべて早期に提出している
5	ほぼ出席、課題もすべて提出しているが、提出が少し遅い
6	欠席が多く、課題も全く提出していない
7	ほぼ出席、課題もすべて提出しているが、提出が遅い

表 5. クラスタ 7 の受講生の 9 回以降のデータ

番号	出席回数	課題提出	提出値計	最終 1	最終 2	成績
48	7	5	3.000	◎	◎	S
51	7	5	3.665	◎	◎	S
95	7	5	1.255	◎	◎	S
71	3	5	0.524	○	◎	A
72	5	5	2.863	○	◎	A

ことがわかる。したがって、クラスター 7 の場合、成績確率と共に“今後欠席せず、最終課題を頑張れば、成績が S になる可能性が高い”とアドバイスを加えれば、受講生の学習意欲の向上が期待できる。

3. まとめ

現在の受講生の成績を予測し、学習指導に活用するため、過去の受講生データのクラスター分析について論述した。成績確率だけではなく、過去の受講生のその後の学習経過をもとに学習アドバイスを付けることで学習意欲の向上を図ることも検討している。今後の課題は本分析を用いた成績予測による学習指導の実践である。

参考文献

- [1] 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)