

学生の修学データを用いた要注意学生の傾向分析

伊藤宏隆^{†1} 伊藤圭佑^{†1} 舟橋健司^{†1} 山本大介^{†1}
齋藤彰一^{†1} 松尾啓志^{†1} 内匠逸^{†1}

名古屋工業では、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ージメントシステムを導入した。システム導入から7年が経過し、大量のデータが蓄積されている。通常、蓄積された出欠データや学習データは成績評価に用いられるが、修学指導支援にも活用することを計画している。多くの大学では、特に留年者や退学者の存在に頭を痛めている。将来、留年、退学する可能性の高い学生を要注意学生と定義し、早期に発見し指導することで留年、退学する学生を減少することができる。本研究では、過去の学生のデータを分析することで要注意学生の修学傾向を把握し、現時点だけではなく将来の学生の傾向を予測した上での修学指導支援を行うことを目的としている。本文では、学生の修学データを用いた分析結果について論述する。

Analysis of Anxious Student using Student's Study Data.

HIROTAKE ITOH^{†1} KEISUKE ITOH^{†2} KENJI FUNAHASHI^{†1}
DAISUKE YAMAMOTO^{†1} SHOICHI SAITO^{†1} ICHI TAKUMI^{†1}
HIROSHI MATSUO^{†1}

In April 2007, the Nagoya Institute of Technology introduced an attendance system and a course management system. More than seven years have passed since these systems have been used, and the systems have accumulated large volumes of data over this period. The authors planned the use in educational guidance of the data collected in these systems. Many universities suffer from the presence of repeaters and dropouts. We define student who has high possibility to be a repeater or dropout as anxious student. If we discover the anxious students and provide them educational guidance early, we will reduce the repeaters and dropouts. The purpose of this research is supporting the educational guidance by identifying trends of the anxious students through the analysis of past students. In this paper, we describe the analysis result using student's study data.

1. はじめに

現在、学生に関する情報を電子データとして集積する事例は珍しいものではなくなっている。その背景には情報通信技術の急速な発展が大きく関係している[1][2]。

名古屋工業では、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ージメントシステム[3]を導入した。システム導入から7年が経過し、大量のデータが蓄積されている。出欠管理システムには1年で全学生約5000名分約180万件の出欠打刻データが蓄積される。通常、蓄積された出欠データや学習データは成績評価に用いられるが、修学指導支援にも活用することを計画している。修学指導の課題として、留年者や退学者の存在がある。多くの大学では、留年者や退学者に頭を痛めている。これまでの指導は、学生課や教務課等の関連部署によるもの、各学生にチューター教員を割り当て、成績配布時にアドバイスを行うものであったが、スタッフや教員一人の担当する学生の数が多く各学生にあった指導を行うことができていない。また、教員は授業や研究が優先となり、修学指導にまで手が回っていない。また、指導は学期終了後の成績配布時等、その時点までの修学状況を参考

にした指導であったため、学生が学期途中で落ちこぼれた場合には手遅れとなっている。

そこで、本研究では、これまでに蓄積された学生の出欠データや成績データを分析することにより、過去の留年者や退学者の傾向を発見し、指導に役立てることを目的としている。特に将来、留年、退学する可能性の高い学生を要注意学生と定義し、早期に発見し重点的に要注意学生を指導することで、指導コストを抑えつつ、留年、退学する学生を減らすことが可能であると考えている。

要注意学生の分析にあたっては、データマイニング[4]を用いる。データマイニングでは、ある時点までは順調な学生であっても、このパターンの場合、その後、成績が悪化する可能性がある等、それまで思いもよらなかった新しい知識の発見が期待できる。

教育現場におけるデータマイニングの活用例をいくつか挙げると、学生による授業アンケートをもとに学生個人の最終成績や学習状況の相互関係を調査したもの[5]や、授業の出席状況や課題提出状況から学生の特性を分析したもの[6]、授業への出欠状況、課題提出状況、試験採点データ及び授業アンケートの分析から授業改善策発見を目指したもの[7]などがある。これらの研究では、目的や手法に差異があるが、既知の知識やデータから未知の傾向や事象を分析・予測するという共通点が存在する。

本研究においては、まず、GPA (Grade Point Average) を

^{†1} 名古屋工業大学
Nagoya Institute of Technology

用いて、学生の一般的な修学傾向を分析し、1年次のGPAが1.0以下の学生は、ほぼ留年・退学していることを突き止めた。これらの学生は指導が必要なことが明白であるため、データマイニングの対象から除き、1年次順調だった学生のうち、どのような学生がその後、留年・退学するかをデータマイニングにより、明らかにする。科目的GPA、獲得成績数、打刻回数の3種類のデータを用いて、主成分分析とクラスタリングによる傾向分析を行い、さまざまな知識を得ることができた。要注意学生への修学指導は早ければ早いほど、留年・退学者の減少に効果があると考え、今回1年次のデータを用いている。

本論文では、まず、対象とする要注意学生を定義するため、GPAによる分析を行った。そして、主成分分析、クラスタリングといったデータマイニング手法を用いて行った科目別GPA、獲得成績数、打刻回数と修学状況との関連性の分析結果について述べる。

2. 要注意学生の定義

2.1 学生の修学状況調査

要注意学生を定義するため、まず、学生の修学状況の調査を行った。用いたデータはA年度入学171名とB年度入学167名の2年度分338名の成績データと出欠打刻データと修学状況データである。成績データは約1万件、出欠打刻データは約24万件存在する。この二つの年度入学生は同一学科の学生で順調に行けば、4年ですでに卒業しているが、一部の学生はまだ在学中である。本学科では、4年次開始時に卒業研究着手となり研究室に配属されるが、一定の単位取得条件を満たしていない場合、卒業研究着手は許可されず留年となる。表1に二つの各年度の学生の卒業研究着手までに要した年数を示す。表1に示すとおり338名中283名が順調に卒業研究に着手できているが、約15%にあたる55名は4年次開始時に卒業研究に着手できていない。

表2に各年度における卒業に要した年数を示す。表2から約20%にあたる70名が4年で順調に卒業できていないことが分かった。

次に留年・退学者70名の3年次終了までの各学期において履修登録の未登録者数を表3に示す。学期の数字は年次、前は前期、後は後期を示す。学期が進むにつれて、その人数が増加している。表3では、複数の学期で未登録の場合があるので、履修登録をしていない学期が存在する学生数を調査したところ、18名であった。換言すると、70名中75%の52名は単位取得の意志があつたにも関わらず留年・退学してしまったことになる。

次に、対象の70名の学期別GPAを調査したところ、1年次から成績が低迷している学生もいれば、3年次まで平均以上の成績を収めている者もいることが確認できた。また、学期間でのGPAの差が1.0以上の開きがある学生が70名中36名であったところから、入学以降に順調に修学しな

表1 卒業研究着手までの年数

Table 1 Years until starting graduation research.

年度	3年	4年以上	未着手	退学	合計
A年度	145	15	5	6	171
B年度	138	15	6	8	167
合計	283	30	11	14	338

表2 卒業までの年数

Table 2 Years until graduation.

年度	4年	5年以上	在学中	退学	合計
A年度	134	22	8	7	171
B年度	134	12	10	11	167
合計	268	31	18	18	338

表3 各学期における履修登録の未登録者数

Table 3 Number of course non-registration students at each term.

学期	1前	1後	2前	2後	3前	3後
A年度	0	1	2	2	4	4
B年度	0	1	3	8	5	6
合計	0	2	5	10	9	10

がらもある時期を境に成績が急低下する学生が半数以上いることになる。

1年次に成績値とその後の修学傾向をさらに調査するため、全学生を対象に、1年次前期と後期のGPAの値域別人数を調査した。表4に結果を示す。

表4より1年次のGPAが高い学生ほど留年・退学する割合が少なかった。1年後期のGPAが1.0を下回る学生に注目すると、全員がその後、留年・退学していた。1年前期の場合、GPAが1.0を下回る学生13名のうち11名が留年・退学している。すなわち、1年次のGPAが1.0を下回る学生は、その後、留年もしくは退学する可能性が高いといえる。これらの学生については、指導対象とすべきことは明白である。そこで、本研究では、これらの学生を除き、1年次前期後期のGPAが1.0以上の学生を調査対象とし、その中からその後、留年・退学する学生の傾向を発見することとする。そこから、指導対象とすべき要注意学生の指摘を試みる。なお、1年次前期後期のGPAがともに1.0以上の学生は307名である。

3. データマイニングによる分析手法

3.1 手法の概要

前節で調査対象とした1年次前期後期のGPAがともに1.0以上の学生307名についてデータマイニングにより、修学傾向を分析した。

表 4 GPA 値域別人数

Table 4 Number of Students at each GPA scope.

GPA 値域	1 年前期			1 年後期		
	全学生	留年・退学者	割合	全学生	留年・退学者	割合
0.0 以上 0.5 未満	5	5	100%	11	11	100%
0.5 以上 1.0 未満	8	6	75%	12	12	100%
1.0 以上 1.5 未満	11	7	64%	31	14	45%
1.5 以上 2.0 未満	48	23	48%	67	15	22%
2.0 以上 2.5 未満	105	16	15%	96	10	10%
2.5 以上 3.0 未満	100	10	10%	70	2	3%
3.0 以上 3.5 未満	53	3	6%	42	4	10%
3.5 以上 4.0 未満	8	0	0%	7	0	0%
合計人数	338	70		338	70	

分析にはデータマイニング手法の一つであるクラスタリング[8]を用いている。これにより、同じ傾向・特性を持つ学生を1つのグループに集団化し、その集団の傾向を分析する。本研究では、扱う属性数が多くクラスタリングの結果に解釈を与えるのが難しい。そこで、クラスタリングを行う前に、主成分分析[9]により情報を縮約し、説明を与えやすい変数を作り出す。その後、クラスタリングを行うことで、結果に対する解釈を平易化した。分析の手順は以下ようになる。

1. 属性変数の選択
2. 主成分分析による情報の縮約
3. 主成分の意味付け
4. 主成分スコア値を用いたクラスタリング
5. クラスタ毎の傾向調査

本研究においては、分析に用いるソフトウェアとして、Microsoft Office Excel 2010 と、フリーのデータマイニングツールである Weka[10] を採用した。Weka はニュージーランドのワイトカ大学で開発されたソフトウェアで、JAVA で実装されている。主にデータ解析、予測モデリングや視覚化ツールとして多くの機能が搭載されている。

3.2 主成分分析

主成分分析 (Principal Component Analysis : PCA とも表記される) は、多変数で表されるデータから、各変数によって新しい変数を作り出す手法である[9]。主成分分析を用いる目的は、情報の縮約であり、データを主要な変数に要約することで特徴の分析・把握を容易にする働きがある。

「主要な変数」は主成分と呼称され、生成された各主成分に対し解釈を与えることで新たな意味を有した変数が生成される。主成分は元の変数の数 p 個まで作られる。しかし、大抵の場合は全ての主成分は用いられず、十分に全データを説明しうる分の主成分だけ採択され、不必要な主成分は無視される。これにより、主成分分析の主目的である情報の縮約が完了される。主成分分析において、用いる主成分

の個数を決定する指標として、寄与率と累積寄与率が挙げられる。寄与率を直感的に説明すれば「ある主成分が全体のデータの何%を説明しているかを表すもの」だと言える。主成分を採択する基準として、「寄与率が何%以上であるものを採択する」「寄与率が大きい順に加算していった、累積寄与率が何%を上回る主成分の個数だけ採択する」といった基準が存在する。一般的には累積寄与率が 60%~80%を上回った分だけを採択する。本研究では 70%を基準として、主成分の採択を行っている。

3.3 クラスタリング

クラスタリングは、対象間の類似度に基づき対象をグループ分けすることによって、グループの共通の性質が何であるかを検討しやすくするために使われる手法である[8]。データマイニングに内包されている概念であり、元のデータ集合を特徴のある集団に分割するため、分析手法として用いられている。クラス (ラベル) が不要であるため、教師なし学習に分類される。

クラスタリングは主に階層的、非階層的に大別できる。階層的クラスタリングは似たもの同士を併合していくつかのグループにまとめていく手法であり、非階層は似た固体が同じグループになるように集合を分割していく手法である。非階層的の例として K-means 法を解説する。

K-means 法は、元のデータを似た事例同士が同じクラスに属するように集合を分割する手法である。階層的クラスタリングと異なる点は、クラスタ数を予め決めておかなければならない点が挙げられる。

クラスタ数を k 、データ (事例) の個数を n としたとき、K-means 法は次の手順で行われる。

1. k 個のシード $p_i (i=1,2,\dots,k)$ を生成する。
2. 各データ $x_j (j=1,2,\dots,n)$ に対し、最も近いシード p_i を求め、データ x_j をクラスタ C_i に入れる。
3. 各クラスタの重心を求め、その重心を新たなシード p_i とする。

4. 上記手順においてシード p_i が移動していた場合、手順 2 へ戻る。移動していなかった場合、クラスタリングを終了する

手順 2 において、ある事例と重心の距離を算出する。その距離の指標としてユークリッド距離が最も有名であり、本研究でも採用している。K-means 法の利点は、実装が容易かつ実行が早い点である。そのため、新規のサンプルを入力してもすぐに再計算できる。ゆえに多くの場面で活用されている。しかし、階層性はないため、クラスタリングの結果がクラスタ数や初期のシードに大きく影響を受けてしまうのが難点である。

4. データマイニングを用いた要注意学生の分析

4.1 科目別 GPA に関する分析

2 節で述べた 307 名を対象として、科目別 GPA と未来の修学状況に関する分析を行った。属性変数は 1 年次の科目別 GPA データを用いた。属性変数として 1 年次の全授業科目を、外国語、人文、数学、体育、理科、専門、その他に分類し科目別 GPA を求め、前期、後期毎で全 14 個の変数を用いている。まずは、主成分分析によりデータの縮約を試みた。主成分分析の結果を表 5 に示す。第 6 主成分において累積寄与率が 70% を超えたため、採用する主成分数を 6 とした。次に各主成分の固有ベクトルを表 6 に示す。各主成分の固有ベクトルを解釈することで、主成分に意味付けを行った。その結果を表 7 に示す。

第 1 主成分では、すべての固有値が負の値となっていた。このことから、第 1 主成分を「全科目成績低迷度」と意味付けした。第 2 主成分では、特に外国語科目の値が大きくなっていた。このことから第 2 主成分を「外国語の得意度」と意味付けした。第 3 主成分では、数学と理科の科目の値が低くなっていたため、「理系科目の不得意度」とした。

表 5 科目別 GPA を用いた主成分分析結果

Table 5 Result of PCA using each subjects GPA.

	固有値	寄与率	累積寄与率
第 1 主成分	2.258	0.364	0.364
第 2 主成分	1.198	0.102	0.467
第 3 主成分	1.117	0.089	0.556
第 4 主成分	1.012	0.073	0.629
第 5 主成分	0.992	0.070	0.699
第 6 主成分	0.888	0.056	0.756
⋮	⋮	⋮	⋮
第 13 主成分	0.504	0.018	0.988
第 14 主成分	0.410	0.012	1.000

表 7 表 6 の各主成分の意味付け

Table 7 Meaning of each principal component in Table 6.

	解釈
第 1 主成分	全科目成績低迷度
第 2 主成分	外国語の得意度
第 3 主成分	理系科目が不得意度
第 4 主成分	やる気増加度
第 5 主成分	専門科目の得意度
第 6 主成分	人文科目の得意度

第 4 主成分では、前期よりも後期の値が大きかった。よって「やる気増加度」とした。第 5 主成分では、他の分野と比較して専門科目の値が高かったため、「専門科目の得意度」と意味付けした。第 6 主成分では人文科目において高い値が確認できたため、「人文科目の得意度」と意味付けした。

この主成分スコアを属性変数として、K-means 法による

表 6 表 5 の主成分分析の固有ベクトル

Table 6 Eigenvector of PCA on Table 5.

	学期	外国語	人文	数学	体育	理科	専門	その他
第 1 主成分	前	-0.168	-0.226	-0.325	-0.12	-0.313	-0.342	-0.232
	後	-0.188	-0.206	-0.34	-0.183	-0.306	-0.359	-0.288
第 2 主成分	前	0.579	-0.15	-0.006	-0.106	0.046	-0.19	-0.26
	後	0.616	0.092	0.087	-0.034	0.114	-0.105	-0.322
第 3 主成分	前	0.334	0.221	-0.244	0.533	-0.306	0.204	0.282
	後	0.21	-0.056	-0.355	0.129	-0.237	-0.096	0.161
第 4 主成分	前	-0.271	0.099	-0.11	0.391	-0.126	-0.269	-0.568
	後	0.047	0.333	-0.066	0.358	0.261	0.103	0.13
第 5 主成分	前	0.004	-0.141	-0.323	-0.246	-0.256	0.046	-0.027
	後	0.149	0.356	-0.081	-0.529	0.027	0.35	0.434
第 6 主成分	前	0.078	0.625	0.133	0	0.101	-0.11	-0.001
	後	-0.184	0.538	-0.033	-0.329	-0.18	-0.195	-0.245

表 8 主成分スコアを用いたクラスタリング結果
 Table 8 Result of clustering using principal component score.

クラスタ	主成分						人数	留年・ 退学	割合
	1	2	3	4	5	6			
1	1.32	-1.67	0.45	-0.07	0.15	-0.29	42	12	29%
2	-1.09	-0.23	-0.51	0.4	-0.74	0.57	44	6	14%
3	1.44	1.14	-0.27	-1.01	-0.62	-0.05	37	6	16%
4	-0.42	-0.11	1.11	0.44	0.8	-0.04	36	1	3%
5	3.37	0.55	0.6	-2.89	2.16	-0.22	4	1	25%
6	-2.45	-0.78	-0.55	-0.39	0	-0.49	36	1	3%
7	-2.41	0.98	-0.18	0.1	0.6	0.25	45	2	4%
8	0.51	0.39	0.07	-0.03	-0.71	-1.44	26	4	15%
9	2.28	-0.02	0.29	0.17	-0.13	1.15	35	10	29%
10	4.47	1.47	-1.34	2.28	0.79	-0.51	10	4	40%

クラスタリングを行った。クラスタ数を 10 に設定した際のクラスタリング結果を表 8 に示す。各クラスタの人数、その中で留年者及び退学者の人数を記載する。

留年・退学者の割合が最も高かったクラスタはクラスタ 10 であった。主成分 1 のスコアが高いので、このクラスタは全科目の成績がかなり悪いが、主成分 4 も高いので、学前期の成績で危機感を抱き、学習意欲は改善したのかもしれない。ところが、最終的には、留年・退学しているののでその後、悪化したのかもしれない。次に留年・退学者の人数が多かったクラスタはクラスタ 1 と 9 であった。2 つのクラスタに共通した傾向として、主成分 1 と 3 が高いので、全科目成績が悪く、特に理系科目が不得意である。またクラスタ 1 では、主成分 2 が低いので、外国語が苦手という傾向が確認できた。また、クラスタ 6 とクラスタ 7 はほとんど留年・退学者が含まれていなかった。この 2 つの共通点は、主成分 1 が低いことから、全科目成績が良好であることであり、「1 年次の成績が良い学生は留年・退学する確率が低くなる」という一般的に述べることのできる傾向が見出された。

4.2 獲得成績数に関する分析

獲得成績数とは、講義の最終評価である「秀」や「不可」などの成績の獲得個数を意味する。獲得成績数データを用いて、成績評価の分布の差異における修学状況の傾向分析を試みた。

属性変数は 1 年前期と後期の各成績の獲得数を示す 12 変数とした。各成績とは、秀、優、良、可、不可、失格である。まず、主成分分析により情報の縮約を行った。各主成分の寄与率及び累積寄与率を表 9、各主成分の固有ベクトルを図示化したものを図 1 に示す。主成分分析の結果、第 5 主成分にて累積寄与率が 0.7 を超えたため、主成分数を 5 とした。また、固有ベクトルを検討することで各主成分の意味付けを行った。その結果を表 10 に示す。

表 9 獲得成績数を用いた主成分分析結果

Table 9 Result of PCA using number of academic score.

	固有値	寄与率	累積寄与率
第 1 主成分	1.903	0.302	0.302
第 2 主成分	1.323	0.146	0.448
第 3 主成分	1.072	0.096	0.543
第 4 主成分	1.042	0.090	0.634
第 5 主成分	1.003	0.084	0.718
⋮	⋮	⋮	⋮
第 11 主成分	0.091	0.001	0.999
第 12 主成分	0.032	0.000	1.000

図 1 から、第 1 主成分では、前後期ともに秀と優の値が大きくなっていったため「優秀さ」と意味付けした。第 2 主成分では、前後期共に最高評価である秀と、単位取得が認められない不可と失格の値が大きくなっていった。ゆえに、中間の成績が少なく良い成績と悪い成績が多い「成績の極端さ」と意味付けした。第 3 主成分では秀と良の値が大きく、優と可が小さい。また、不可と失格の値が増加していった。1 年次はチュートリアルも兼ねた講義も多いため秀が取りやすい背景がある。これらを踏まえ「優可の少なさ」と意味付けした。第 4 主成分では、前期と後期で値の分布に変化が見られた。前期では優と不可の値が大きく、後期では良と失格の値が大きくなっていった。優から良、不可から失格に、全体的に低下しているの、「成績低下度」と意味付けした。第 5 主成分でも前期と後期において大きな変化が見られた。前期では可の値が大きくなっていったが、後期では良と可が多く、平滑化されていた。このことから「学習レベル改善度」と意味付けした。

さらに主成分スコアを変数として K-means 法によるクラスタリングを行った。クラスタ数を 10 に設定した時の

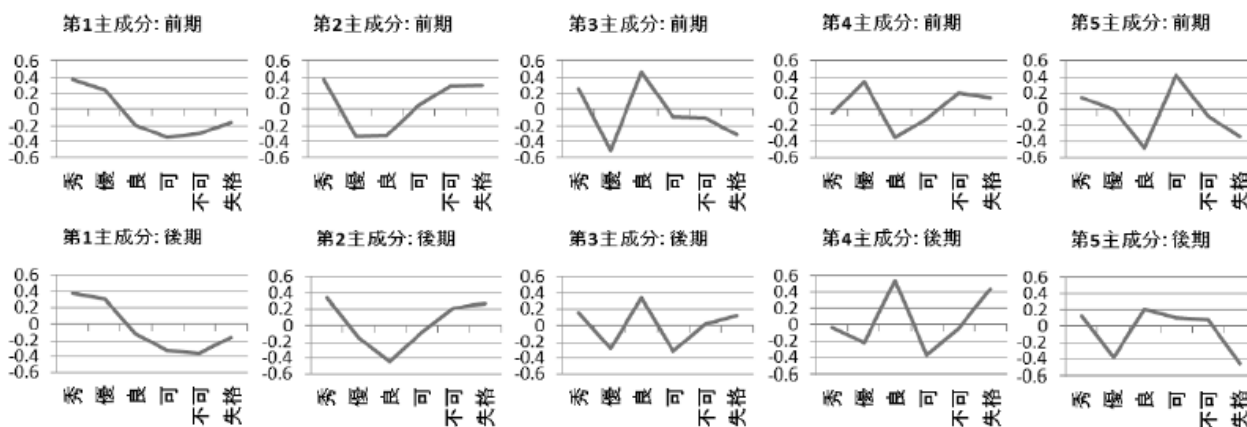


図1 表9の主成分分析の固有ベクトル

Figure 1 Eigenvector of PCA on Table 9.

表10 図1の各主成分の意味付け

Table 10 Meaning of each principal component in Figure 1.

	解釈
第1主成分	優秀さ
第2主成分	成績の極端さ
第3主成分	優可の少なさ
第4主成分	やる気喪失度
第5主成分	学習レベル改善度

結果を表11に示す。

表11より、優秀さの主成分1がマイナス、すなわち優秀でないクラスタほど、留年・退学者の割合が高くなっている傾向が確認できた。また、主成分2より成績が極端なほどその割合が高く、安定するほど割合が低くなることも確認できた。クラスタ7は該当する3名中3名が留年していることが分かった。このクラスタの傾向として、優秀ではなく、成績が極端で、成績が低下、学習レベルが追いつい

ていない傾向が確認できた。クラスタ8も割合が高かった。クラスタ7とクラスタ8の共通点として、優秀ではなく、成績が極端で、成績低下が認められる点を確認できた。クラスタ6も割合が比較的高く、優秀ではなく、成績が極端である傾向があった。同じ成績が極端なクラスタ3は、優秀さがあるため留年・退学者は少なかった。このことから、「成績が優秀ではなく獲得する成績が極端な学生は要注意学生になりやすい」といえる。

対して、割合が低いクラスタは、クラスタ3、クラスタ4、クラスタ5であった。クラスタ3とクラスタ5は優秀であるが、クラスタ4はそこまで優秀でなかった。しかし、成績が極端ではないという傾向を示していた。つまり、「優秀でなくても成績が極端でない学生は要注意学生になりやすい」といえる。

4.3 打刻回数に関する分析

打刻データを用いて、出席率と打刻回数と要注意学生に関する調査を行った。打刻は学生が教室での入退室の際に学生証を端末に認識させることを意味する。打刻回数が出

表11 主成分スコアを用いたクラスタリング結果

Table 11 Result of clustering using principal component score.

クラスタ	主成分					人数	留年・退学	割合
	1	2	3	4	5			
1	-0.59	-0.48	0.11	0.6	1.09	37	4	10.80%
2	-1.84	-0.3	0.52	-1.18	0.06	39	5	12.80%
3	2.44	1.58	1.01	0.01	0.47	29	1	3.40%
4	0.32	-1.16	-0.9	0.08	-0.16	53	2	3.80%
5	2.26	0.36	-0.84	0.53	0.11	32	1	3.10%
6	-2.29	1.29	-1.59	-0.16	0.08	24	6	25.00%
7	-4.1	3.82	0.8	3.81	-4.13	3	3	100.00%
8	-2.5	1.01	0.51	0.79	-0.14	23	10	43.50%
9	-0.14	-1.46	1.46	0.34	-0.44	27	3	11.10%
10	1.57	0.19	0.14	-0.81	-0.63	40	6	15.00%

表 12 打刻回数を用いた主成分分析結果

Table 12 Result of PCA using number of time recording.

	固有値	寄与率	累積寄与率
第 1 主成分	2.269	0.515	0.515
第 2 主成分	1.329	0.177	0.691
第 3 主成分	0.877	0.077	0.768
⋮	⋮	⋮	⋮
第 9 主成分	0.401	0.016	0.988
第 10 主成分	0.351	0.012	1.000

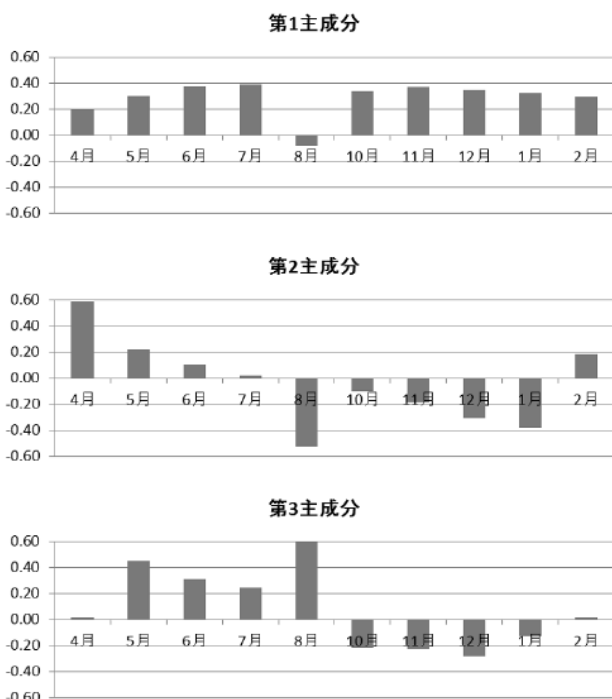


図 2 表 12 の主成分分析の固有ベクトル
 Figure 2 Eigenvector of PCA on Table 12.

席状況の一端を表していると言える。

まずは、属性変数を月別打刻数（授業のある 4 月から 8 月分と 10 月から 2 月分）の 10 変数として、調査を行った。主成分分析の結果を表 12 に示す。第 3 主成分にて累積寄与率が 0.7 を超えたため、成分数を 3 とした。さらに固有ベクトルを図 2 に示す。固有ベクトルの検討による各主成分の意味付けを行った。その結果を表 13 に示す。

第 1 主成分では全体的各月の値が大きくなっていった。よって「勤勉に出席したかどうか」と意味付けした。第 2 主成分では 4 月の値が大きく、8 月の値が小さくなっていった。さらに 4 月から 8 月、10 月から 1 月にかけて値が減少していることも確認できた。これより、第 2 主成分を「徐々に欠席率上昇」と意味付けした。第 3 主成分では前期（4 月から 8 月）の値が大きく、後期（10 月から 2 月）の値が小さくなっていった。そこで、「後期に出席状況悪化」と意味付けした。

表 13 図 2 の各主成分の意味付け

Table 13 Meaning of each principal component in Figure 2.

	解釈
第 1 主成分	勤勉に出席したかどうか
第 2 主成分	徐々に欠席率上昇
第 3 主成分	後期に出席率低下

表 14 主成分スコアを用いたクラスタリング結果

Table 14 Result of clustering using principal component score.

クラスタ	主成分			人数	留年・退学	割合
	1	2	3			
1	1.03	-1.2	0.03	66	2	3.00%
2	-2.15	-0.02	-0.04	44	12	27.30%
3	-2.59	-2.07	1.62	10	1	10.00%
4	1.77	1.82	0.77	47	4	8.50%
5	-1.01	-3.17	2.42	7	1	14.30%
6	0.87	0.73	-0.35	69	6	8.70%
7	-5.05	1.47	0.15	16	8	50.00%
8	-0.47	-1.93	3.44	1	1	100.00%
9	3.35	-1.77	-0.96	11	1	9.10%
10	-1.09	-0.43	-1.15	36	5	13.90%

さらに 3 つの主成分スコアを変数として、K-means 法によるクラスタリングを行った。表 14 にその結果を示す。クラスタ数は 10 とした。

クラスタ 8 の該当者は 1 名で、その 1 名は留年していた。このクラスタが示す特徴として、勤勉に出席しておらず、後期に出席しなくなっていることが確認できた。次に留年・退学の割合が高いクラスタはクラスタ 7 であった。このクラスタの傾向として、出席率がかなり悪いこと、出席率が学期開始時から低下していることが確認できた。対して、割合が低いクラスタ 1、クラスタ 4、クラスタ 6、クラスタ 9 に着目すると、総じて出席率が高いという傾向が確認できた。

5. まとめ

学生 338 名を対象に、各種データを用いた修学状況分析を行った。修学指導の対象者となるべき学生の指摘を試みた。その結果、調査対象の 338 名の内、70 名が 1 年次以降に退学もしくは留年していることが分かった。さらに、1 年次の前期と後期の GPA について調査すると、GPA が 1.0 を下回る学生のほぼ全員が 1 年次以降に退学もしくは留年していることが分かった。この結果を受け、調査対象を 1 年前期と 1 年後期の GPA が共に 1.0 以上の学生に限定し、更なる調査を進めるためデータマイニングの手法を適用した。データマイニングの手法として、主成分分析と

K-means 法におけるクラスタリングを採用した。主成分分析により多くの変数が持つ情報を縮約し、クラスタリングによって学生集団をグループ分けし、各クラスタの傾向や退学・留年者の割合を調査した。この調査により、各観点における要注意学生とすべき学生の傾向を得ることができ、単一の観点ではなく複数の観点から学生を評価することの有用性を示すことができた。

要注意学生とすべき学生の傾向をまとめる。

- ・全科目成績が低迷しているが、特に理系科目と外国語科目が不得意。
- ・1年前期から1年後期にかけて成績が多少改善しても全体の成績が悪い場合は要注意。
- ・獲得成績が中間の成績が少なく、良いか悪いか極端。
- ・出席率が低く、特に、時間がたつにつれ、出席率が低下している。

今回の打刻回数の分析では、月別の打刻回数を用いたが、週間別の打刻回数や曜日別の打刻回数もしくはそれらの分散を計算し新たなデータを獲得することで、学生の新たな傾向を得られる可能性がある。

参考文献

- 1) 岡田正, 高橋参吉, 藤原正敏, ICT 基礎教育研究会: ネットワーク社会における情報の活用と技術, 実教出版 (2010)
- 2) 山本洋雄, 中山実, 清水康敬: ICT 活用での形成的評価による学習成績・意欲に関する一考察, 電子情報通信学会技術研究報告. ET, 教育工学 Vol.108, No.247, pp.39-44 (2008)
- 3) 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)
- 4) 元田浩, 山口高平, 津本周作, 沼尾正行: データマイニングの基礎, オーム社(2006)
- 5) 原圭司, 高橋健一, 上田祐彰: ペイジアンネットワークを用いた授業アンケートからの学生行動モデルの構築と考察, 情報処理学会論文誌, Vol.51, No.4, pp.1215-1226 (2010)
- 6) 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)
- 7) 南俊朗, 大浦洋子: 授業データ解析による授業改善策発見を目指して -努力・成果・評価の関連性からのアプローチ-, 九州情報大学研究論集, Vol.15, pp.1-16 (2013)
- 8) A.K.Jain, M.N.Nurty, P.J.Flynn,: Data Clustering: A Review, ACM Computing Surveys, Vol.31, No.3 (1999)
- 9) 河口至商: 多変量解析入門 1, 森北出版 (1973)
- 10) Weka 3: Data Mining Software in JAVA
<http://www.cs.waikato.ac.nz/ml/weka/>