

平成 20 年度

卒業研究論文

データマイニングによる学生の修学傾向分析と
その修学指導への適応有効性の検証

指導教員

舟橋 健司 准教授

伊藤 宏隆 助教

名古屋工業大学 工学部 電気情報工学科

平成 16 年度入学 16213029 番

堀江 匠

目次

1	はじめに	1
2	データマイニングの基本知識	4
2.1	データマイニングによる知識獲得工程	4
2.2	データマイニングに使用する道具	7
2.3	データマイニングの使用方法	8
3	データマイニングによる修学傾向分析	9
3.1	修学傾向分析方法	9
3.2	データマイニングに使用するデータ	10
3.3	データマイニング環境	11
3.4	分析結果	11
4	修学指導への適応有効性の検証	25
4.1	有効性の検証方法	25
4.2	検証に使用するデータ	27
4.3	検証結果および考察	28
4.3.1	授業成績判定の有効性の検証	28
4.3.2	単位習得不可者に対する修学指導への適応有効性の検証	31
4.4	授業成績判定傾向と考察	34
5	むすび	38
	謝辞	40
	参考文献	41

第1章

はじめに

近年、大量の情報が増えつづけているため重要なデータがどれなのか分からず、その中に埋もれていることが多々ある。そこで、増え続けている情報を有効活用するため、データマイニングという技術が注目されている。

データマイニング(Data Mining)とは、日々増え続けている大量の情報に対して、人が対応することができず、その大量の情報の中に埋もれてしまった有用な知識(情報やルール)を発見する技術である[1]。

データマイニングに使われている解析技術の基礎的理論は、20世紀半ばまでには出来ている。しかし、従来の統計解析やデータ解析とは違い、データマイニングでは多数のデータベースを統合して高度なシステム化されたデータウェアハウス上で行われ、大量のデータが扱えるようになり、数値だけではなくテキストデータ、画像データといった様々な種類のデータの蓄積、処理を行うことができる。また、データウェアハウスはネットワークで結ばれているので、ただのデータベースの役割しかないものではなく、データベースを管理、分析を行うコンピュータとも結ばれているシステムである。そのため、従来の統計解析やデータ解析だけでなく、ニューラルネットワークやパターン認識などのあらゆる解析方法が可能である。

ところが、解析技術そのものは20世紀半ばまでにあったものの、その頃はコンピュータ技術やデータベース技術といった計算や記録管理の技術がまだ発展しておらず、データマイニングを実用化できるところまでの域に達していなかった。しかし近年、大容量で安価な記憶装置や高性能な演算処理装置が開発され、ネットワーク環境が整備されたため、データウェアハウスのような大規模なシステムの構築が可能になったことが、データマイニングが注目されている要因である。

また、現在も解析技術は研究されている。GNP(遺伝的ネットワークプログラミング)を用いたニューラルネットワーク[2]といった解析方法もあり、データマイニングは様々な視点から発展している。

社会的な背景として、ビジネスや企業経営関連では、データマイニングが登場する1990年代以前にも、経営管理システム、意思決定支援システム、戦略的情報システム、定形業務の自動化、オフィスオートメーションなどのコンピュータシステムの導入に熱心であった。インターネットの普及によりビジネス形態も変化していき、より時代に対応させる必要が出てきた。そのため、

大量のデータの中から有用な情報を探し出す必要性が高まったことも、データマイニングが注目されている要因である[3].

本来、データマイニングは蓄積された莫大なデータをもとにビジネスに有効なデータを特徴や傾向、区分、予測したりするもので、ビジネス用として生まれたのがはじまりである。しかし、現在は大量のデータを扱えたり処理できたりするので、ビジネスの分野だけでなく、様々な分野でも応用されている。

ビジネス分野として、流通、サービス業では顧客の購入パターンのルールを発見して戦略を立てたり、地域や天候で仕入れる品物の量を調整してコスト削減を行うといった使い方をされる。また、通信の分野と連携してインターネットショッピングやインターネットオークションといったインターネットを用いられているビジネスに、会員登録や購入履歴、メールマガジン登録から様々な情報が得ることができると、顧客の性格や特徴からニーズに合った商品の提供やおすすりめなサービスの表示といった活用方法もある。他にも製造業では、不良品になりやすい物の原因や工程を発見することで、製品の品質の向上やコスト削減に活用され、金融、保険業ではそれぞれの顧客にあった貸付や保険の設定などに使われている。

また、ビジネス以外の分野として医療分野では、病気の原因を調べる、薬の作用や影響を改善するといった使用方法がある。また、株価予想といった経済関連、天気予想や地震予想といった環境、災害関連の大量の観測データをもとに予測したりするときにも、データマイニングは活用されている。その他、競馬予想や心理学によるカウンセリングなどにも用いられている。

データマイニングは、大量にデータとそれを蓄積、管理するためのデータベース(データウェアハウス)、処理する演算装置やアプリケーションといったデータマイニングが使用可能である環境ならば、どの分野でも用いることは可能である。

本研究では、大量の情報の中から有用な情報を発見するというデータマイニングの特徴を生かし、現在問題となっている学力低下の解決策として、データマイニングを使い学生の修学傾向を分析し、その結果から修学指導への適応が有効であるかの検証を行う。具体的には、過去の授業データを用いてデータマイニングを行い、学生の修学傾向を分析する。その結果から、現在まで行われている授業データを使い、その授業の各受講生の成績を予想する。ただし本研究の目的は、各受講生の授業成績の予想を的確に的中させることではなく、成績を予想した結果、特に単位習得不可の成績になる可能性がある者に対し、危機意識を持たせることである。そうすることで、学習意欲が向上し結果的に授業成績が改善されると考えられる。また、その修学傾向分析結果が

危機意識を持たせるのに有効であるかを判断するために、予想した授業成績の正答率を調べる検証も行い修学指導への適応有効性を判定する。

なお、本研究における全ての授業データは、各受講生の氏名などのプライバシー情報を特定できないものを用いる。

以下、データマイニングの詳しい内容を第2章で述べ、データマイニングによる修学傾向分析方法、結果について第3章で示す。また、分析結果から修学指導へ適用したときの有効性を調べる検証と考察については第4章で述べる。第5章で本研究のまとめについて述べる。

第2章

データマイニングの基本知識

本章では、データマイニングの基本知識であるデータマイニングによる知識獲得工程や使用する道具、データマイニングの使用方法について述べる。

2.1 データマイニングによる知識獲得工程

実際にデータマイニングを使用して知識(法則やルール)を発見するためには、図2.1に示される工程を行う。図2.1に示す知識獲得とは、今までは発見できなかったが、データマイニングを行ったことで有益であったり改善できるようなルールや法則を発見できたことを指す。また、知識の表現方法は文字や文章だけであるとは限らず、図2.1の(5)視覚化、解釈、意思決定のように図や表など視覚的にわかりやすく表現することで、発見した知識に対して理解されやすくなり、より良い知識となる。

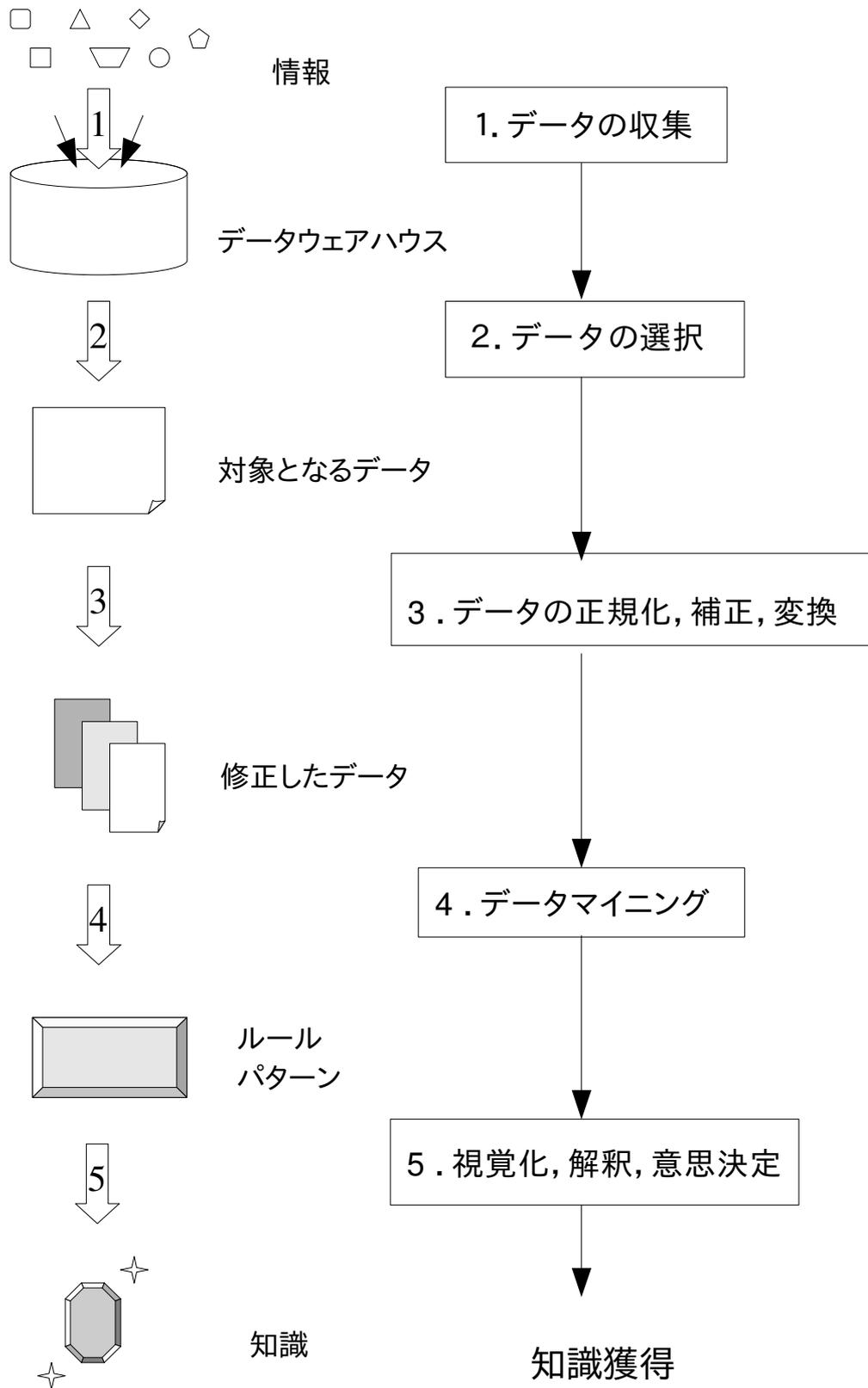


図 2.1 データマイニングによる知識獲得工程

データマイニングによる知識獲得は以下の工程で行われる。

(1) データの収集

大量の情報をネットワークなどを通じ、データベースを統合したデータウェアハウス上に保管管理する。

(2) データの選択

データウェアハウス上に収集されたデータからデータ解析に使うためのデータを選択し取り出す。

(3) データの正規化, 補正, 変換

取り出してきたデータを比較解析などを行えるように正規化(データを管理しやすいように整理する), 補正, 修正などを行い, データマイニング可能な状態にするため平均値や標準偏差などに変換する。

(4) データマイニング

主成分分析やクラスター解析, 決定木などの解析技術を用いて変換したデータからルールやパターンなどを発見する。

(5) 視覚化, 解釈, 意志決定

データマイニングによって発見されたルールやパターンなどを視覚化したり, わかりやすく解釈する。

これらの工程を経てデータマイニングを行ったとしても, 必ずしも良い知識が発見できるとは限らない。いくつかの工程を繰り返したり, やりなおすことも必要となる。例えば, (3)データの変換や(4)データマイニングの工程でうまくいかなかった場合, (1)データの収集や(2)データの選択の工程からやりなおすことも必要である。

2.2 データマイニングに使用する道具

大量のデータやそれを処理するだけのコンピュータ環境やネットワーク環境があってもデータマイニングを行うことはできない。図2.2のように各工程ごとに必要に応じたソフトウェアを使いわけの必要がある。ただし、図2.2にあるように、データマイニングを行うには特定のソフトウェアが必要であるというわけではなく、規模や使いやすさ、何をしたいかということに応じて道具を選択することも可能である。例えば、小規模なものであれば、グラフ作成機能、データベース機能、計算機能などをもっている表計算ソフトでデータマイニングを行うほうが、操作が容易であり、ある程度コンピュータを使う知識さえあれば使うことが可能である。しかし、ビジネスなどの大規模なデータマイニングとなれば、表計算ソフトでは扱いきれないほどのデータを使うので、データベース管理システムでデータを扱ったり、それぞれの用途にあったデータマイニング専用のソフトウェアを使う必要がある。そのため、大規模なものになってくるとそれなりの知識と技術を持っていないとデータマイニングを行うことができない。また、現在では簡単にデータマイニングでき、一般的な表計算ソフトである excel を用いてデータマイニングの練習をするという使い方もされている[4]。

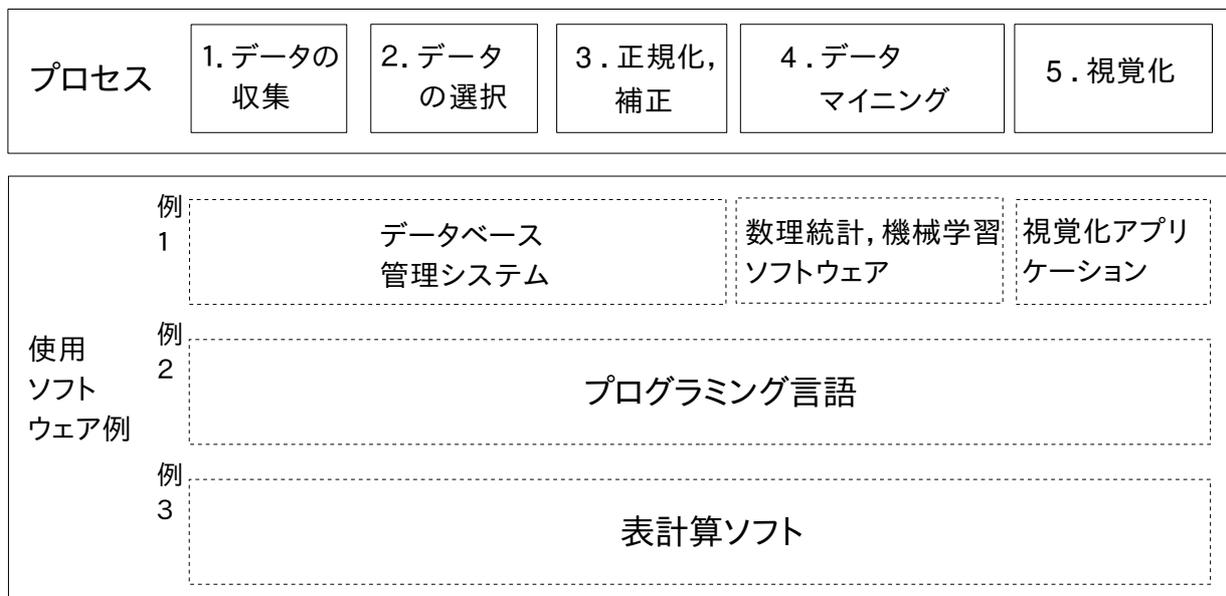


図 2.2 データマイニングに使用されるソフトウェア

2.3 データマイニングの使用法

データマイニングでは、主に以下の使い方をされる。

(1) 分類

新たに得たデータの特徴を調べ、それらの特徴からすでに定義されているグループや数値などに分類する。分類の例として、新しい薬の副作用を高リスク、中リスク、低リスクに分ける、ある材料に対して不良品になりやすいのかどうかの判定、といったものがある。

(2) 予測

過去の得られたデータからどのようなパターンがあるかを見つけ、そのパターンより未来の値や行動を予測する。分類との違いは時間の変数があることである。予想の例として、過去の株価から将来の株価を予想する株価予想、過去の気圧や雲の動きなどのデータから今後の動きを予想する天気予報などがある。

(3) クラスタリング(グループ化)

1つの多様なデータ群を、いくつか同様の特徴や性質を持ったクラスタ(グループ)に分割する。分類とは異なり、クラスタリングは事前に定義されているものに分類するのではなく、得たデータ群を似た特徴のクラスタに分けるだけである。そのため、顧客グループをクラスタリングしてから複数のクラスタに分け、各クラスタごとに販売戦略を立てる、といった探索的な使い方をする。

(4) 関連付け、アソシエーションルール

Xが起これば、それによりYが起きるといった関連性のあるルールを見つける。例えば、ある商品Aを買えば、ある商品Bを買う、というものが関連付けにあたる。しかし逆に、ある商品Bを買えば、ある商品Aを買うというルールがあるとは限らない。

これらは、それぞれの用途にあった解析方法やソフトウェアを用いて使い分けられる。

第3章

データマイニングによる学生の修学傾向分析

本研究では、大量のデータの中から知識を獲得するというデータマイニングの特性を生かすため、過去の授業データからデータマイニングを行い、学生の修学傾向を分析する。その分析結果より現在進行している授業の成績を予想し、特に成績不可や失格のような単位の習得できない者に対して単位習得の危機意識を持たせ、結果的に授業成績が改善されることを目的としている。

本章では、データマイニングによる学生の修学傾向分析方法や結果について述べる。

3.1 修学傾向分析方法

修学傾向分析方法として、データマイニングに用いる授業データを1つの授業データしか使わない場合だと、データマイニングによって分析した結果が他の授業でも同じであるとはいえない。そのため、複数の授業データからデータマイニングを行い分析する。

本研究では、4つの授業データ、授業1(07')、授業2(07')、授業3(07')、授業4(07')を使用しデータマイニングを行った(ただし、「(07')」は2007年度の授業であることを意味する)。

手順として、最初にデータマイニングを各授業ごとで行い4授業分のデータマイニングの知識を発見する。各授業1(07')、2(07')、3(07')、4(07')を1つの授業データとはしない。次に、得られた4つの授業のデータマイニングの知識より共通、類似した法則を発見してまとめる。それを本研究における学生の修学傾向の分析結果とする。

3.2 データマイニングに使用するデータ

以下のデータを用いてデータマイニングを行い分析した。

授業 1(07') : 人数 61 人, (成績 S 12 人, A 18 人, B 16 人, C 12 人, D 2 人, X 1 人)
講義回 8 回, 課題提出 7 回, (第 1~7 回講義回後の課題)

授業 2(07') : 人数 110 人, (成績 S 13 人, A 30 人, B 38 人, C 9 人, D 9 人, X 11 人)
講義回 15 回, 課題提出 3 回, (第 3, 5, 8 回講義回後の課題)

授業 3(07') : 人数 106 人, (成績 S 10 人, A 26 人, B 43 人, C 10 人, D 12 人, X 5 人)
講義回 15 回, 課題提出 14 回, (第 1~14 回講義回後の課題)

授業 4(07') : 人数 94 人, (成績 S 9 人, A 17 人, B 30 人, C 18 人, D 14 人, X 6 人)
講義回 14 回, 課題提出 14 回, (第 1~14 回講義回後の課題)

各授業の 出欠データ(受講生ごとの各講義回の打刻時間と出席, 欠席, 遅刻, 早退状況)
課題提出データ(受講生ごとの各課題提出回の課題提出時間)
課題提出期間(各課題提出回の課題提出期間)
成績データ(受講生ごとの成績)

(授業 1(07'), 2(07')については 2007 年前期の授業, 授業 3(07'), 4(07')は 2007 年後期の授業)

成績は S,A,B,C,D の順番で 5 段階あり, S が最高成績で A,B,C と続き D が最低成績である。また, S~C は成績可, D は成績不可を表しており, 履修した授業において成績が D であった場合, 単位は与えられない。また, 出席日数が足りないなど受験資格がない場合は, X により表す。

3.3 データマイニング環境

本研究のシステムではデータマイニングが容易にできる点、扱うデータの量がそれほど大規模なものではない点から Microsoft Office excel 2007 のデータマイニングアドインを用いてデータマイニングを行う[5].

この研究でデータマイニングを行ったコンピュータの環境は以下の通りである.

OS: Microsoft Windows Server 2003 R2 Standard x64 Edition Service Pack2

CPU: Intel(R) Xeon(R) E5310 1.60GHz

メモリ: 8.00 GB

アプリケーション: Microsoft Office excel 2007

ソフトウェア: Microsoft SQL Server 2005 Data Mining Add-ins for Office system

3.4 分析結果

各授業 1(07'), 2(07'), 3(07'), 4(07')について以下のことを行った.

- ・ 出欠、課題提出が成績とどのような関連性があるのかを調べるため、各授業の出欠データと課題提出データに対して、クラスタおよび関連付けの機能を使い、上位成績や下位成績といったクラスタごとの特徴分析、出欠や課題提出の関連パターンを判定.
- ・ 出欠、課題提出が成績にどのような影響を与えているのかを調べるため、各授業の出欠データと課題提出データに対し主要な影響元の分析の機能を使い、影響力があるものを判定.
- ・ 課題提出データの課題提出の有無だけでなく、提出期間内における課題提出の早さも成績との影響や関連性があるのかを調べるため、主要な影響元の分析、クラスタ、関連付けの機能を使い分析.

授業 1(07)について、以下のような分析結果となった。

ただし、出欠データの各図では出欠を 1、欠席を 2、遅刻を 3、早退を 4 とし、課題提出データの各図では提出を 1、未提出を 0 として表している。

図 3.1 に授業 1(07)の出欠データをクラスタ化した結果を示す。図 3.1 のクラスタ 1, 2, 3 から、成績 S,A,B は全体的に出欠率が高い。逆に、クラスタ 4 より成績 X は全体的に出欠率が低い。また、クラスタ 5 より成績 C については徐々に出席率が低くなっている。さらに、最終講義は全体的に出席率が高い。

変数	状態	母集団 (すべて)	クラスタ 1	クラスタ 2	クラスタ 3	クラスタ 4	クラスタ 5
サイズ		61	27	8	15	5	6
20070604	1	54	99 %	90 %	100 %	0 %	99 %
20070604	2	7	1 %	10 %	0 %	100 %	1 %
20070611	1	54	100 %	89 %	100 %	0 %	98 %
20070611	2	7	0 %	11 %	0 %	100 %	2 %
20070618	1	53	100 %	98 %	100 %	0 %	42 %
20070618	2	8	0 %	2 %	0 %	100 %	58 %
20070625	1	52	91 %	87 %	100 %	0 %	97 %
20070625	2	8	4 %	12 %	0 %	100 %	3 %
20070625	3	1	5 %	1 %	0 %	0 %	1 %
20070702	1	53	100 %	91 %	100 %	0 %	69 %
20070702	2	8	0 %	9 %	0 %	100 %	31 %
20070709	1	53	100 %	93 %	90 %	0 %	98 %
20070709	2	8	0 %	7 %	10 %	100 %	2 %
20070717	1	50	99 %	82 %	100 %	0 %	40 %
20070717	2	11	1 %	18 %	0 %	100 %	60 %
20070723	1	57	100 %	100 %	87 %	60 %	100 %
20070723	2	3	0 %	0 %	13 %	20 %	0 %
20070723	4	1	0 %	0 %	0 %	20 %	0 %
成績	S	12	8 %	0 %	67 %	0 %	0 %
成績	A	18	41 %	45 %	3 %	40 %	4 %
成績	B	16	31 %	35 %	3 %	40 %	39 %
成績	C	12	20 %	20 %	15 %	0 %	57 %
成績	D	2	0 %	0 %	12 %	0 %	0 %
成績	X	1	0 %	0 %	0 %	20 %	0 %

図 3.1 授業 1(07)の出欠データのクラスタ化

授業 1(07')の出欠データに関連付けを行った結果の成績 S, 成績 C と出欠の関係をそれぞれ図 3.2, 3.3 に示す. 図 3.2 で成績 S は授業 01 から 07 まで連続で出席している影響が強く, 図 3.3 で成績 C と授業に対する影響がまばらであることより, まばらで出席するよりも連続で出席する方が, 成績は良くなる傾向があるといえる.

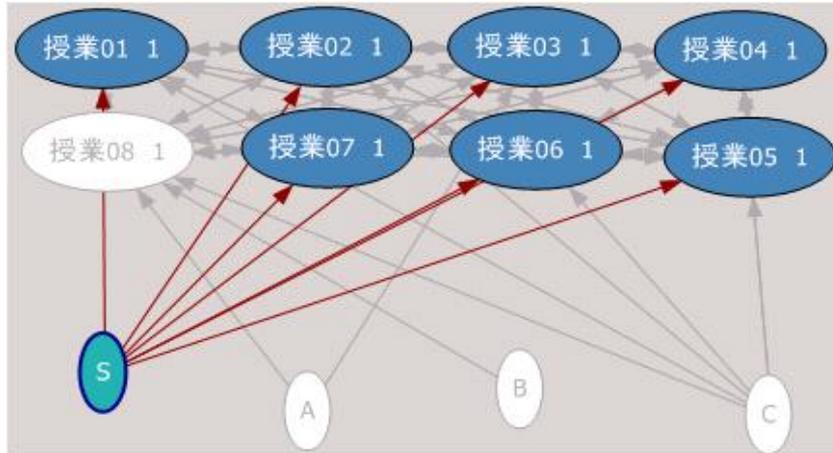


図 3.2 授業 1(07')の成績 S と出欠の関連性

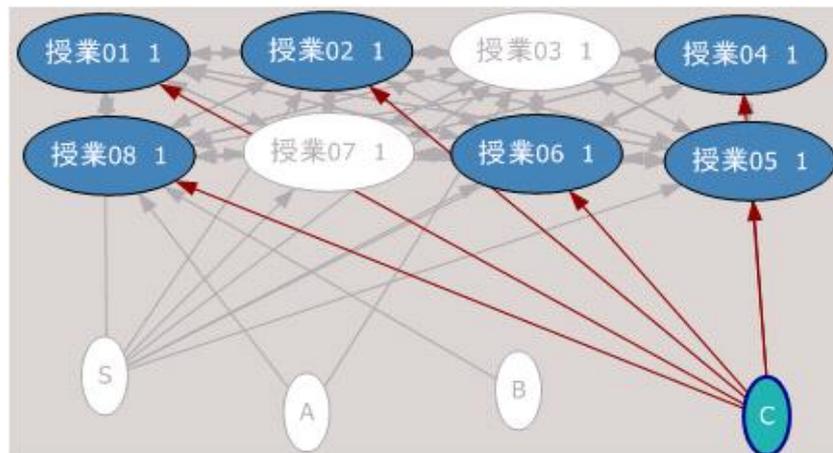


図 3.3 授業 1(07')の成績 C と出欠の関連性

図 3.4 に授業 1(07)の出欠データの主要な影響元の分析結果を示す。図 3.4 から遅刻は打刻が遅いものだけであるが、早退は本当に早退しているのか打刻を忘れているのかは分からない。

列	値	優先	相対的影響	影響値
開始前打刻時間	8:32:45 - 8:40:33	1		6
開始前打刻時間	< 8:32:45	1		6
終了後打刻時間	10:03:26 - 10:17:14	1		6
終了後打刻時間	10:00:15 - 10:03:26	1		5
終了後打刻時間	< 10:00:15	1		5
開始前打刻時間	8:44:30 - 8:49:12	1		4
終了後打刻時間	10:17:14 - 10:23:35	1		4
開始前打刻時間	>= 8:49:12	1		3
開始前打刻時間	8:40:33 - 8:44:30	1		3
終了後打刻時間	>= 10:23:35	1		3
開始前打刻時間	なし	2		100
終了後打刻時間	なし	2		75
終了後打刻時間	10:17:14 - 10:23:35	3		100
開始前打刻時間	>= 8:49:12	3		89
終了後打刻時間	なし	4		100
開始前打刻時間	8:44:30 - 8:49:12	4		92

図 3.4 授業 1(08)の出欠データの主要な影響元の分析

図 3.5 に授業 1(07)の出欠データと全ての成績を含めた関連付けの分析結果を示す。図 3.5 より出席を連続で行っているといたり、出席以外のルールがないことから、出欠や欠席は連続で起こりやすい、授業終盤の出欠の影響が成績への強いなどの法則は見つからなかった。

確率	重要度	ルール
92 %	0.66	授業03 1 -> 授業07 1
98 %	0.64	授業01 1 -> 授業02 1
98 %	0.64	授業02 1 -> 授業01 1
94 %	0.49	授業02 1, 授業01 1 -> 授業04 1
91 %	0.47	授業05 1 -> 授業07 1
94 %	0.44	授業01 1 -> 授業03 1
96 %	0.35	授業06 1, 授業01 1 -> 授業05 1
96 %	0.35	授業05 1, 授業08 1 -> 授業06 1
96 %	0.35	授業05 1, 授業01 1 -> 授業06 1
94 %	0.27	授業06 1, 授業01 1 -> 授業03 1
94 %	0.27	授業04 1, 授業02 1 -> 授業05 1
94 %	0.27	授業03 1, 授業01 1 -> 授業06 1

図 3.5 授業 1(07)の出欠データの関連付け

図 3.6 に授業 1(07')の課題提出データをクラスタ化した結果を示す。図 3.6 から、課題を提出しなければ成績 X になりやすい。また、X 以外の成績の者であっても授業中盤以降に課題提出していない場合がある。

変数	状態	母集団 (すべて)	クラスタ 1	クラスタ 2	クラスタ 3
サイズ		61	30	30	1
課題1	1	60	100 %	100 %	13 %
課題1	0	1	0 %	0 %	87 %
課題2	1	60	100 %	100 %	13 %
課題2	0	1	0 %	0 %	87 %
課題3	1	60	100 %	100 %	13 %
課題3	0	1	0 %	0 %	87 %
課題4	1	59	100 %	97 %	13 %
課題4	0	2	0 %	3 %	87 %
課題5	1	58	100 %	93 %	12 %
課題5	0	3	0 %	7 %	89 %
課題6	1	59	98 %	99 %	8 %
課題6	0	2	2 %	1 %	92 %
課題7	1	57	94 %	96 %	10 %
課題7	0	4	6 %	4 %	90 %
評価	S	12	16 %	24 %	1 %
評価	A	18	13 %	48 %	5 %
評価	B	16	36 %	17 %	6 %
評価	C	12	30 %	9 %	2 %
評価	D	2	5 %	2 %	0 %
評価	X	1	0 %	0 %	87 %

図 3.6 授業 1(07')の課題提出データのクラスタ化

授業 1(07')の課題提出データに関連付けを行った結果の成績 S、成績 A、成績 B、成績 C と課題提出との関係をそれぞれ図 3.7~3.10 に示す。図 3.7~3.10 よりどの成績でも課題 03 以前の影響が強くなく課題 04 以降のいずれかの影響が大きいため、課題提出は授業中盤から終盤にかけての影響が大きいといえる。

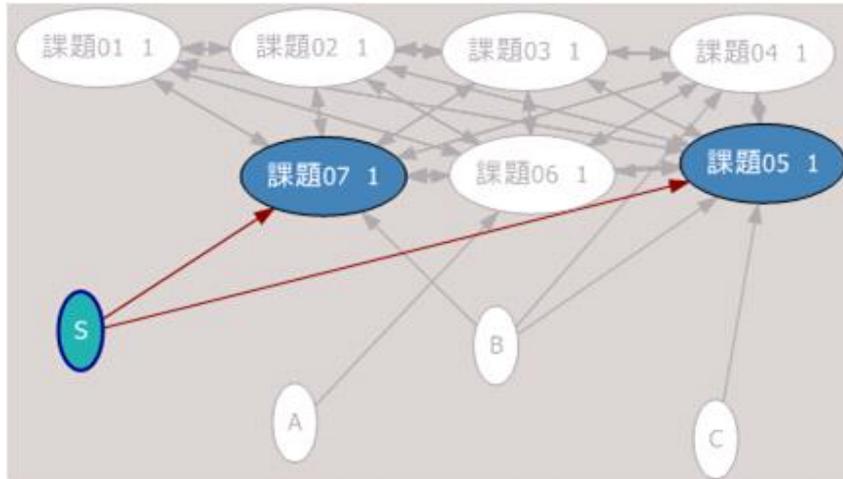


図 3.7 授業 1(07)の成績 S と課題提出の関連性

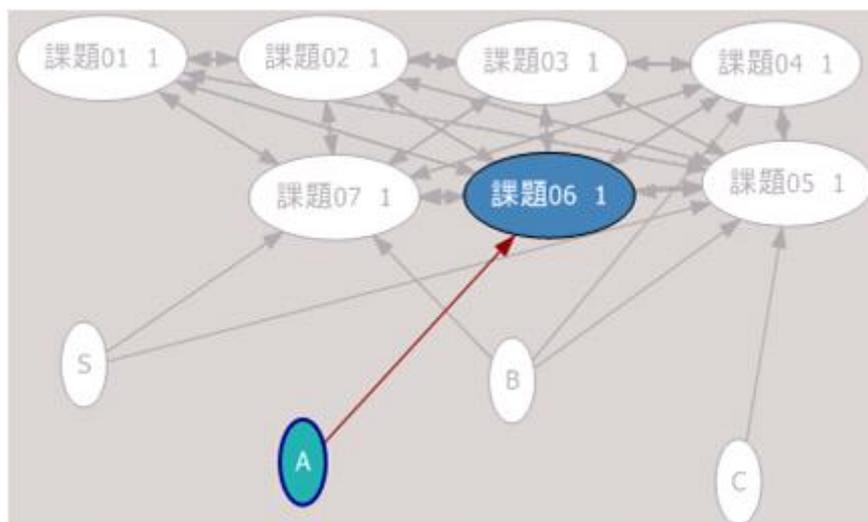


図 3.8 授業 1(07)の成績 A と課題提出の関連性

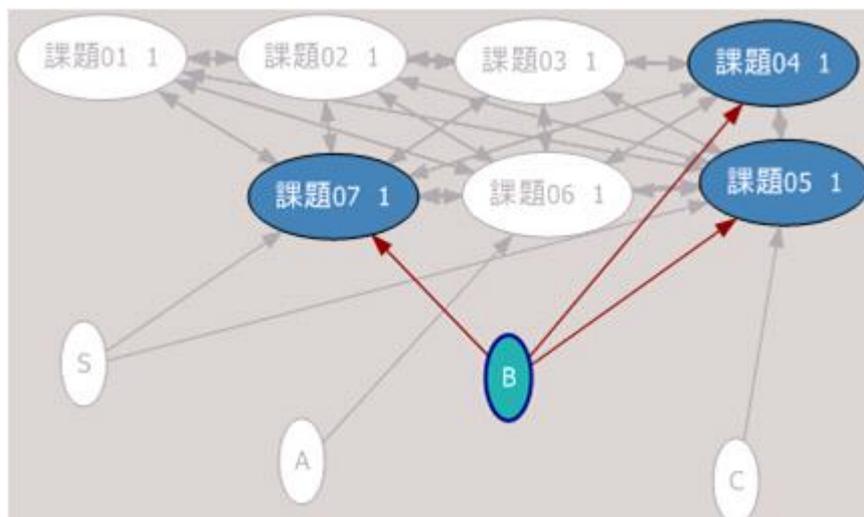


図 3.9 授業 1(07)の成績 B と課題提出の関連性

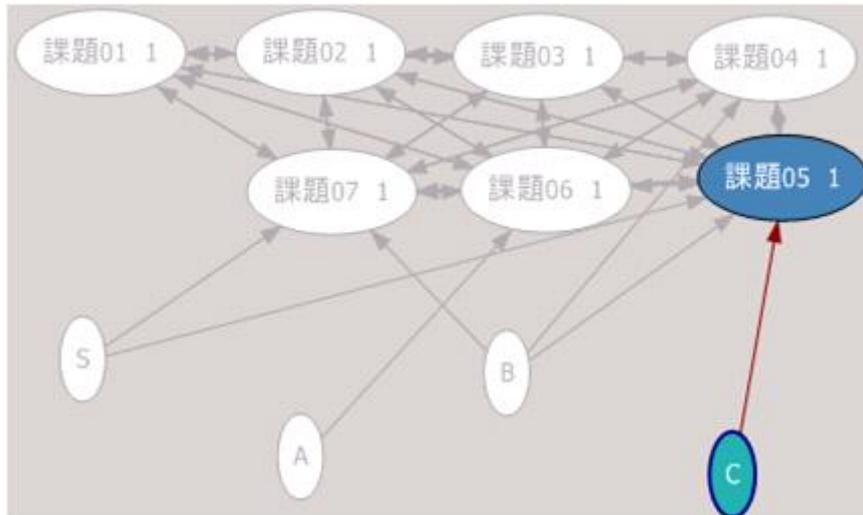


図 3.10 授業 1(07')の成績 C と課題提出の関連性

図 3.11 に授業 1(07')の課題提出データの主要な影響元の分析結果を示す。図 3.11 より成績 D は課題提出，成績 X は課題未提出の影響が大きい。

列	値	優先	相対的影響	影響値
課題5	0	A	100	100
課題7	0	C	100	100
課題3	1	D	100	100
課題1	1	D	100	100
課題2	1	D	100	100
課題6	1	D	47	47
課題4	1	D	47	47
課題5	1	D	12	12
課題2	0	X	100	100
課題3	0	X	100	100
課題1	0	X	100	100
課題6	0	X	79	79
課題4	0	X	79	79
課題5	0	X	66	66
課題7	0	X	56	56

図 3.11 授業 1(07')の課題提出データの主要な影響元の分析

図 3.12 に授業 1(07')の課題提出データと全ての成績を含めた関連付けの分析結果を示す。図 3.12 から特に授業序盤は課題を提出しやすく，課題提出すれば次回も提出しやすいと分析できる。

確率	重要度	ルール
100 %	0.47	課題02 1 → 課題01 1
100 %	0.47	課題01 1 → 課題02 1
100 %	0.47	課題03 1, 課題02 1 → 課題01 1
100 %	0.47	課題03 1, 課題01 1 → 課題02 1
100 %	0.29	課題06 1, 課題03 1 → 課題02 1
100 %	0.29	課題04 1, 課題02 1 → 課題01 1
100 %	0.29	課題04 1, 課題01 1 → 課題02 1
100 %	0.29	課題04 1, 課題03 1 → 課題01 1
100 %	0.21	課題05 1 → 課題02 1
100 %	0.21	課題05 1 → 課題03 1
100 %	0.21	課題05 1, 課題02 1 → 課題03 1
100 %	0.21	課題06 1, 課題04 1 → 課題01 1
100 %	0.14	課題07 1, 課題04 1 → 課題03 1
100 %	0.14	課題07 1, 課題06 1 → 課題01 1

図 3.12 授業 1(07)の課題提出データの関連付け

図 3.13 に授業 1(07)の提出値(課題提出が早ければ値が大きく、遅くなるほど小さくなるように課題提出の早さを数値化したもの)の主要な影響元の分析結果を示す。図 3.13 の提出値と成績の関係から、課題を早く提出するほど成績が良くなる傾向があるといえる。

列	値	優先	相対的影響	優先値
提出値合計	>= 2.9041	S		100
課題04提出値	>= 0.5658	S		32
課題07提出値	>= 0.5890	S		32
課題07提出値	0.0144 - 0.1054	A		100
提出値合計	1.7855 - 2.2658	A		46
課題04提出値	< 0.0500	A		46
提出値合計	1.0490 - 1.7855	B		100
課題04提出値	0.0500 - 0.0762	B		100
課題07提出値	< 0.0144	C		100
課題02提出値	< 0.1874	C		36
課題04提出値	0.0762 - 0.1089	D		100
提出値合計	1.7855 - 2.2658	D		100
課題02提出値	0.1874 - 0.2758	D		100
課題01提出値	>= 0.5025	D		90
課題07提出値	>= 0.5890	D		90
課題07提出値	< 0.0144	X		100
提出値合計	< 1.0490	X		100
課題04提出値	< 0.0500	X		100
課題01提出値	< 0.4168	X		100
課題02提出値	< 0.1874	X		100

図 3.13 授業 1(07)の課題提出の早さによる成績への影響分析

以上の結果から、授業 1(07')の修学傾向分析結果をまとめると次のようになった。

出欠に関する結果

- ・成績 S,A,B は全体的に出欠率が高い。成績 C については徐々に出席率が低くなっている。成績 X は全体的に出欠率が低い。
- ・最終講義は出席率が高い。
- ・まばらで出席するよりも連続で出席する方が、成績は良くなる傾向がある。
- ・遅刻は打刻が遅いものだけであるが、早退は本当に早退しているのか打刻を忘れているのかは分からない。
- ・出欠や欠席は連続で起こりやすい、授業終盤の出欠の影響が成績への強いなどの法則は見つからなかった。

課題提出に関する結果

- ・課題を提出しなければ成績 X になりやすい。また、X 以外の成績の者であっても授業中盤以降に課題提出していない場合がある。
- ・成績は授業中盤から終盤の課題提出の影響が大きい。
- ・特に授業序盤は課題を提出しやすく、課題提出すれば次回も提出しやすい。
- ・課題を早く提出するほど成績が良くなる傾向がある。

同様の方法を授業 2(07'), 3(07'), 4(07')でも分析を行った。以下、各授業の修学傾向分析結果のまとめを示す。

授業 2(07')の修学傾向分析結果のまとめ

出欠に関する結果

- ・成績 S,A,B は全体的に出席率が高く、特に成績 S はその中でも少し高い。
- ・成績 D,X は徐々に出席率が低くなる。また、成績 X は成績 D に比べ出席率が低く、授業中盤以降はほとんど欠席している。
- ・一部の成績 C,D は、授業中盤まではほとんど出席せず、終盤から出席する。
- ・成績は授業中盤から終盤にかけての出欠の影響が大きい。特に中盤の方が影響力がある。
- ・成績が良いと出席の影響が大きく、悪いと欠席及び遅刻の影響が大きい。
- ・終盤の授業を欠席すると、かなり成績 X になりやすい。
- ・遅刻は打刻が遅いものだけであるが、早退は本当に早退しているのか打刻を忘れているのかは分からない。
- ・出席が続くと欠席しにくく、欠席が続くと出席しにくい。また、後者の方が影響が大きい。

課題提出に関する結果

- ・課題を提出していなければ成績 D,X になりやすい。特に X は終盤の提出率が悪い。
- ・課題をすべて提出していれば、成績 S,A,B になりやすい。
- ・課題を早く提出するほど成績が良くなる傾向がある。

授業 3(07)の修学傾向分析結果のまとめ

出欠に関する結果

- ・成績 A,B は全体的に出席率が高い。
- ・成績 C,D,X は出席率が低く、遅刻より欠席の方が多い。また、全体的に出席率が低いものではなく、徐々に出席率が悪くなる。ただし、成績 C,D は終盤にかけて徐々に出席率が上がるのに対し、成績 X は終盤でも出席率は低い。
- ・成績が良いと出席の影響が大きく、悪いと欠席及び遅刻の影響が大きい。ただし、成績 S は全員が必ず出席率が高いというわけではない。
- ・遅刻の約 7 割は授業前の打刻時間が早すぎるため遅刻となっている。また、早退については本当に早退したのか打刻を忘れたのかは分からない。その中でも、成績 S,A,B は打刻時間が早い遅刻、成績 C,D,X は普通の遅刻であることが多い。
- ・出席が続くと欠席しにくくなり、逆に欠席が続くと出席しにくい。特に後者の方が影響が大きい。また、授業終盤になるほどその影響が大きくなる。

課題提出に関する結果

- ・課題を提出していなければ成績 C,D,X になりやすい。逆に課題を提出していても必ず成績 S,A,B になるということではないが、提出回数が多いほど成績はよくなりやすい。
- ・成績 A は授業中盤の課題提出の影響が大きく、成績 B は授業終盤の課題提出の影響が大きい。
- ・成績 C,D,X の中でも、授業中盤に課題未提出の方が終盤に課題未提出の場合より成績が悪くなる傾向がある。
- ・ある回の課題を提出すると、この回の課題を提出するといった法則はみつからなかった。また、課題を提出すれば次回も提出するという連続性もなかった。
- ・課題を早く提出するほど成績が良くなる傾向がある。

授業 4(07')の修学傾向分析結果のまとめ

出欠に関する結果

- ・成績 S,A,B は全体的に出席率が高い。
- ・成績 A は授業終盤の出席，成績 B は授業中盤の出席の影響が大きい。
- ・成績 C,D と一部の成績 B は徐々に出席率が悪くなるが最後の授業には出席している。また，欠席率と遅刻率のどちらが高いかは分からなかった。
- ・成績 X と一部の成績 C,D は徐々に出席率が悪くなり授業終盤でも出席率は増えない。また，欠席率の方が遅刻率よりも高い。
- ・遅刻は打刻が遅いものだけであるが，早退は本当に早退しているのか打刻を忘れているのかは分からない。
- ・成績が良いと出席の影響が大きく，悪いと欠席及び遅刻の影響が大きい。
- ・出席が続くと欠席しにくく，逆に欠席が続くと出席しにくい。特に後者の方が影響が大きく，授業終盤になるほどその影響が大きくなる。

課題提出に関する結果

- ・課題を提出していなければ成績 D,X になりやすい。
- ・成績 A は授業中盤から終盤にかけて，成績 B は授業終盤，成績 C は授業中盤の課題提出の影響が大きい。
- ・成績 C,D,X は徐々に課題の提出率が悪くなる。また，成績が悪くなるほど全体の提出率も悪い。
- ・ある回の課題を提出すると，この回の課題も提出するといった法則はみつからなかったが，後半の方が課題を提出したとき次の課題も提出しやすい傾向がある。

以上の分析結果より、各授業 1(07'), 2(07'), 3(07'), 4(07')の学生の修学傾向の共通、類似したものをまとめると以下の通りとなった。

出欠に関する結果

- ・上位成績者(S,A,B)は全体的に出席率が高い。逆に、Sならば必ずすべての授業に出席しなければならない、というわけではないため、出欠データのみで成績S,A,Bのいずれかを判断するのは困難である。
- ・全員ではないが大半の下位成績者(C,D,X)は、徐々に出席率が悪くなる傾向がある。その傾向は成績が悪くなるほど多くなる。そのなかでも出席していない場合は、成績Xのほとんどが欠席であるのに対して、成績C,Dは遅刻と欠席が同じくらいである。つまり成績C,Dよりは成績Xの方が欠席率が高い。また、最終的に成績C,Dは終盤には出席率がまた徐々に良くなる傾向があるが、成績Xはそのまま下がりつづけ最終的には、ほとんど出席していない。詳しくは成績C,D,Xの中で成績が良いほど前者、悪いほど後者になりやすい傾向がある。
- ・最後の講義は全体的に出席率が高い。
- ・終盤の出欠になるほど、成績への影響が大きくなる。出席なら良い、欠席なら悪い成績になる影響が大きい。
- ・出席が続くと欠席しにくく、欠席が続くと出席しにくい。また、後者の方が影響が大きい。出席または欠席は連続で起きやすい。

課題提出に関する結果

- ・課題未提出ならば、下位成績者(C,D,X)になりやすい。また、すべて提出したからといって必ず上位成績者(S,A,B)になるわけではないが、提出率が高いほど成績はよくなる傾向がある。
- ・成績Xについては早い段階から未提出の傾向があり、終盤ではほぼ未提出である。
- ・中盤から終盤の課題提出の有無が成績への影響力が強い。特に、終盤よりも中盤の方が影響力がある。
- ・課題を早く提出するほど成績は良くなる傾向がある。

以上、4 授業分の修学傾向分析の結果をまとめたものを、本研究における学生の修学傾向分析結果とする。

また、出欠が成績に与える影響と課題提出が成績に与える影響を比較すると、各授業の出欠で出席しても成績 C,D である可能性より、課題提出で提出しても成績 C,D である可能性が高いため、課題を未提出したときの影響が大きいと考えられる。そのため、課題提出の方が出欠に比べ成績に与える影響が少し大きいといえる。

第4章

修学指導への適応有効性の検証

データマイニングでは、知識を得ただけではあまり意味が無い。もし、得た知識が有効性の低いものである場合、すでに知られているルールであった場合は良い知識とはいえず、その知識を実用的に使用することは困難である。知識を検証し、有効性を確認することは実用化する上で重要なことである。

本章では、第3章で分析した修学傾向を修学指導へ適応したとき有効性があるかを調べるための検証方法、結果と考察について述べる。

4.1 有効性の検証方法

第3章のデータマイニングによって修学傾向分析した結果を修学指導への適応したとき有効性があるか調べるため、データマイニングで使用した授業とは別の授業のデータを用いて検証を行う。また、有効性をより高めるため、複数の授業を用いる。

本研究では、授業1(08'), 授業2(08'), 授業3(08'), 授業4(08')を使用し有効性の検証を行った。授業1(08'), 2(08'), 3(08'), 4(08')はデータマイニングに用いた授業1(07'), 2(07'), 3(07'), 4(07')の次年度の授業であるが、これらはデータ収集が容易であったために用いているだけで、これらを用いている特別な意味は無い。また、検証を行ったコンピュータの環境は第3章と同様である。

具体的な検証方法はまず、授業1(08'), 2(08')において、全講義回の出欠、それに対応する課題の提出状況がわかった状態で成績を予想し正誤を調べる。

予想については、修学傾向の分析結果より出欠と課題提出の2種類あるので、出欠データと課題提出データそれぞれで予想し、それらを合わせたものを今回の検証に用いる成績データの予想結果とする。予想の合わせ方について、出欠と課題提出の予想に相違がある場合は、分析結果から、課題の予想を多少優先するものとする。また、本研究目的である、分析結果から成績を判定して特に単位習得できない者になる危機感を持たせるという点、より詳細に成績を判定できるかを調べるという点から、検証に使う判定基準を次のように定める。

授業成績判定の有効性を検証

判定1：成績の上位(S,A,B)，中位(A,B,C)，下位(C,D,X)に分けていずれか予想

判定2：判定1を詳細にし，2～3つの成績(A,B,CやD,Xのように)候補に分けて予想

判定3：判定2の候補のうち出欠や課題提出の状況よりどの成績になるか1つを予想

修学指導への適応有効性を検証

判定4：判定1，3の状況から成績不可(D)および失格(X)になる危険性の高さを予想

判定1，2についてはクラスタ化の機能を使い，同じような出欠，課題提出の特徴や性質を持った者を機械的にクラスタ分けする．その後，クラスタリングしたクラスタを修学傾向分析結果から各クラスタごとの判定1，判定2の予想を行う．判定3については，判定1，2と出欠，課題提出状況から一番なる可能性高いものを手動で判断する．

判定1で大まかな範囲の成績を予想，判定2で判定1を絞れるものは絞った範囲の成績を予想，判定3で一番なりやすい成績を予想とする．

判定4については，1=判定1が上，2=判定1が中かつ判定3がC以外，3=判定3がC，4=判定3がDまたはXの4段階とし，4に近いほど成績Dや成績Xになりやすいように判断し注意喚起する．

また，判定結果は

判定1が 上 中 下 の3段階

判定2が S,A S,A,B A,B A,B,C B,C C,D C,D,X D,X の8段階

判定3が S A B C D X の6段階

判定4が 1 2 3 4 の4段階

とする．ここで，判定2にB,C,Dがない理由は，B,C,Dと判定された場合成績不可になる可能性があるが，それと同時に判定3でBと判定されたとき，成績Dより2段階も成績が上なので単位

習得の危機感を与えるという効果が薄れる可能性があるためである。

次に、最新回分の講義の出欠データを1回分消去する。また、消去した期間内に対応している課題があればその課題も消去する。その状態で成績を予想し正誤を調べ、成績予想ができなくなるまで繰り返す。ただし、課題提出データが無くなった場合、出欠データでデータマイニングが可能であれば、出欠データのみの成績予想をその講義回の成績予想とする。

正誤の判定は実際の成績が成績データを用いて予想した範囲内に含まれていれば正答とする。

例：判定1=下，判定2=D,X，判定3=X，（判定4=4）と予想し実際の成績がDだったとすると判定1，2が正解で判定3が間違いと判断する。

すべて判定した結果，データマイニングによって修学傾向分析して修学指導へ適応したとき有効であるか判断する。また，有効であれば，どの段階の講義の進み具合から予想は有効であるかも判断する。

また，授業3(08')，4(08')については現時点で進行中の授業であり，成績データがないため，予想結果の正誤が分からないので検証を行うことができない。そのため，これらの授業は成績予想のみ(判定1～3)を行い，講義回が進むにつれて予想にどのような変化があるかを調べ，その結果について考察を行う。

4.2 検証に使用するデータ

以下のデータを用いて検証を行った。

授業 1(08') : 人数 58 人, (成績 S 12 人, A 18 人, B 16 人, C 12 人, D 2 人, X 1 人)
講義回数 8 回, 課題回数 7 回(第 1~7 回講義後の課題)

授業 2(08') : 人数 102 人, (成績 S 12 人, A 18 人, B 16 人, C 12 人, D 2 人, X 1 人)
講義回数 15 回, 課題回数 2 回(第 3, 9 回講義後の課題)

授業 3(08') : 人数 99 人, (成績未定)
講義回数 14 回, 課題回数 12 回(第 5, 14 回を除く講義後の課題)

授業 4(08') : 人数 109 人, (成績未定)
講義回数 14 回, 課題回数 13 回(第 1~13 回講義後の課題)

各授業の 出欠データ(受講生ごとの各講義回の打刻時間と出席, 欠席, 遅刻, 早退状況)
課題提出データ(受講生ごとの各課題提出回の課題提出時間)
課題提出期間(各課題提出回の課題提出期間)
成績データ(受講生ごとの成績)

(授業 1(08'), 2(08')については 2008 年前期の授業, 授業 3(08'), 4(08')は 2008 年後期の授業)

ただし, 成績データについては予想の正誤判定のみに使用する。また, 授業 3(08'), 4(08')については現時点で進行中の授業のため, 現在分かっている範囲の授業データを使用する。

4.3 検証結果及び考察

検証は、授業成績判定の有効性(判定1~3)と単位習得不可者(成績D, X)に対する修学指導への適応有効性(判定4)についての2種類を行った。

それぞれを検証可能な授業データ、授業1(08'), 2(08')で検証を行った結果、以下の通りになった。

4.3.1 授業成績判定の有効性の検証

データマイニングを用いて分析した修学傾向の結果から予想した授業成績判定が有効であるかを検証するため、授業1(08'), 2(08')の全体の判定1~3の検証結果を図4.1, 4.2に示す。

また、各表内の講義回の数値横に"*"がある場合はその講義回後に課題が課されたことを示し、判定可能な講義回から予想を行った。

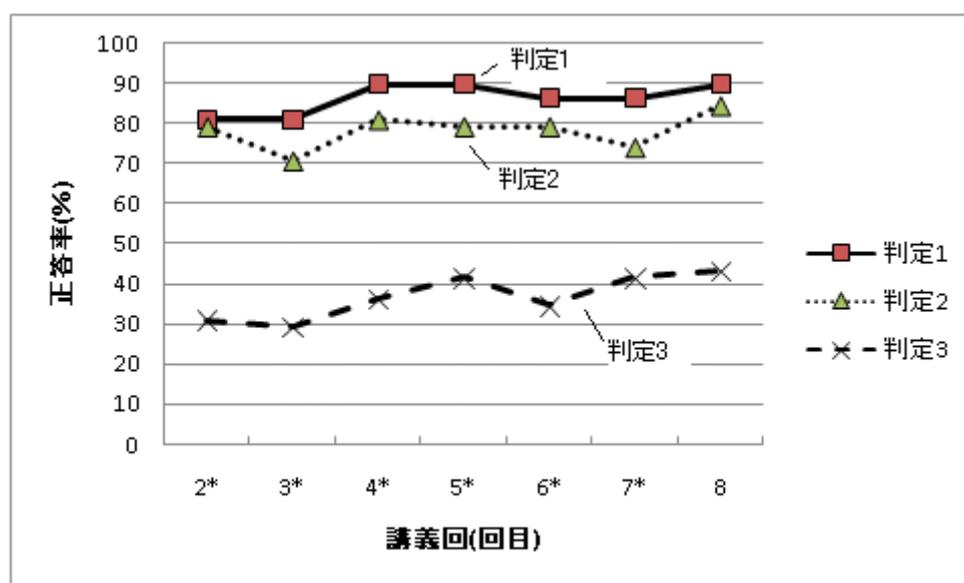


図 4.1 授業1(08')の判定1~3の全体の検証結果

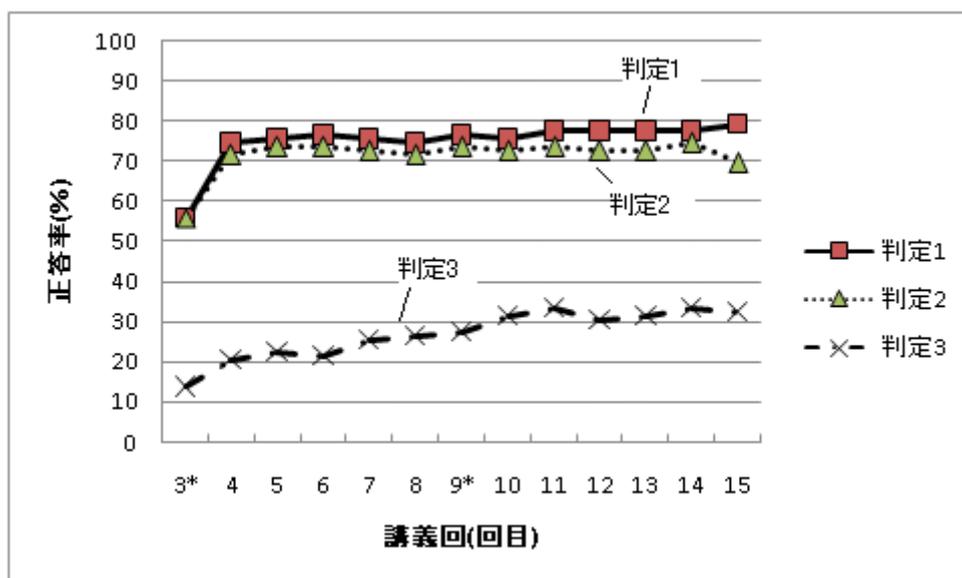


図4.2 授業2(08')の判定1～3の全体の検証結果

授業1(08')について、判定1はある程度課題があれば比較的高い正答率を出しているため有効であるといえる。判定2についても授業がある程度進んでいて、課題もある程度課されているという条件なら十分有効であるといえる。判定3については正答率が低く有効であるとはいえない。

授業2(08')について、この授業は課題自体少ないため、それぞれ判定1～3の正答率が授業1(08')よりも低くなった。ただし、正答率が低くなったとはいえ判定1, 2については、多少の成績予想の有効性はあるといえる。判定3については正答率が低いため有効であるとはいえない。

このような検証結果になった原因は、分析の結果から出欠のデータでは大体の範囲は予想できるが、上位成績のいずれか1つに決める判定にはあまり向かないためだと考えられる。そのため、判定1, 2では高い正答率を出していても判定3では正答率が低くなったと思われる。

次に、判定3が主に低くなった原因を調べるために、判定3を各成績別で見ると次の図4.3, 4.4となった。

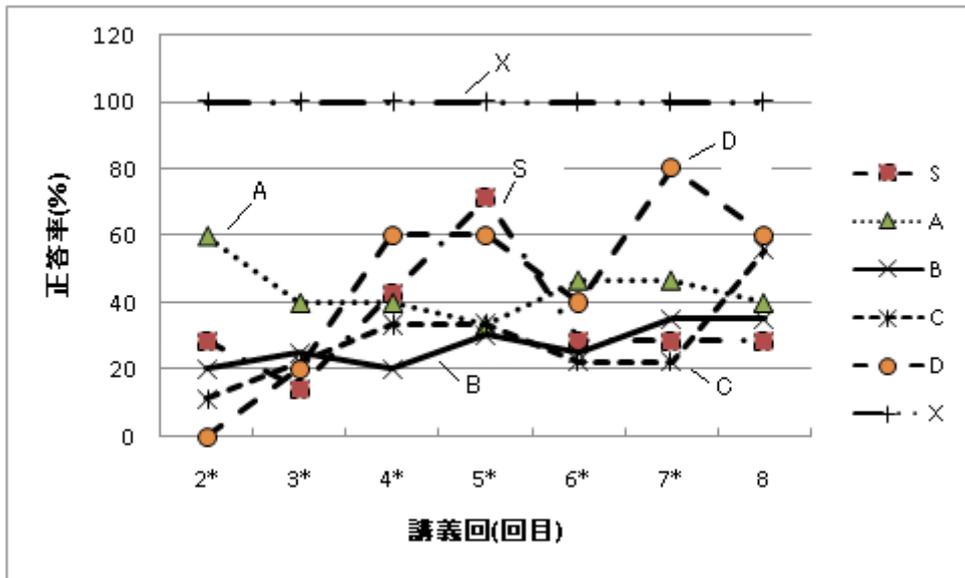


図 4.3 授業 1(08')の判定 3 の各成績別の推移

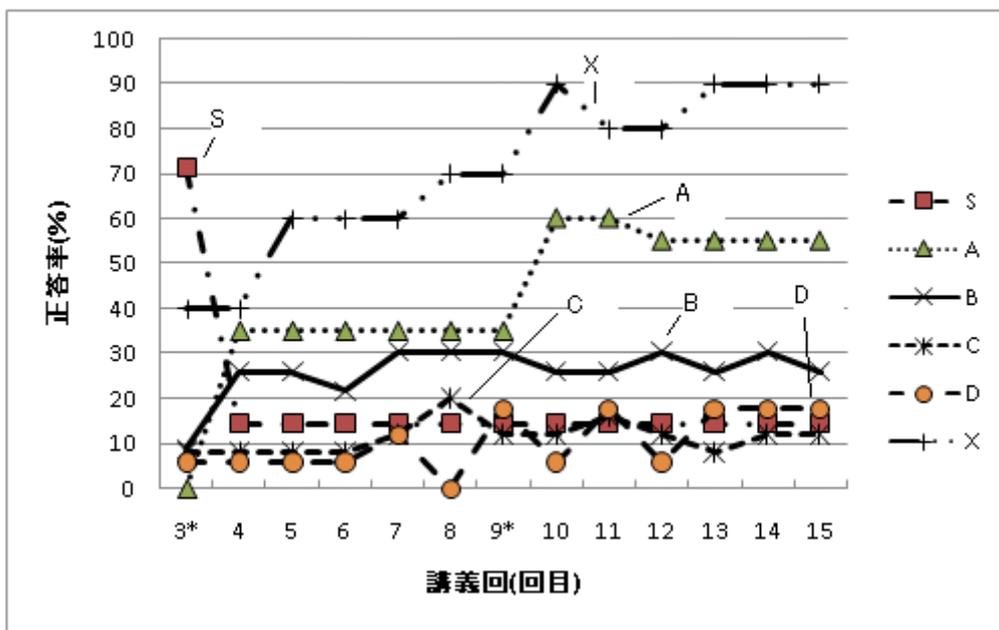


図 4.4 授業 2(08')の判定 3 の各成績別の推移

図 4.3, 4.4 より分かることは、どの成績も同じような値ではなく、成績ごとに正答率の値が全く違うことがわかる。授業 1(08')については、講義が進むごとに高い成績よりも低い成績の方が正答率が上がっている。授業 2(08')についても、授業 1(08')ほどではないが低い成績の方が正答率が上がっている。

このことより、成績別で見るとある程度講義が進み、課題もある程度課されている状態なら、判定 3 の低いと判断された成績予想は多少有効であると考えられる。

以上の結果から、成績予想の有効性として1つの成績を予想するのは難しいが、範囲的な予想に関しては有効である、課題が多い授業の方が正答率が高いため成績予想の有効性が高い、という結果となった。

分析結果も、出欠データでは上位成績のいずれかを定めることは困難とされる点や出欠よりも課題提出の方が成績への影響が強い点などに関しては、正しく検証されたものと考えられる。

4.3.2 単位習得不可者に対するの修学指導への適応有効性の検証

成績不可者(D)と失格者(X)に対して修学傾向分析結果が修学指導へ適応したとき有効性があるのかを検証するため、授業1(08'), 2(08')の成績Dと成績Xの判定4の検証結果を図4.5~4.8に示す。

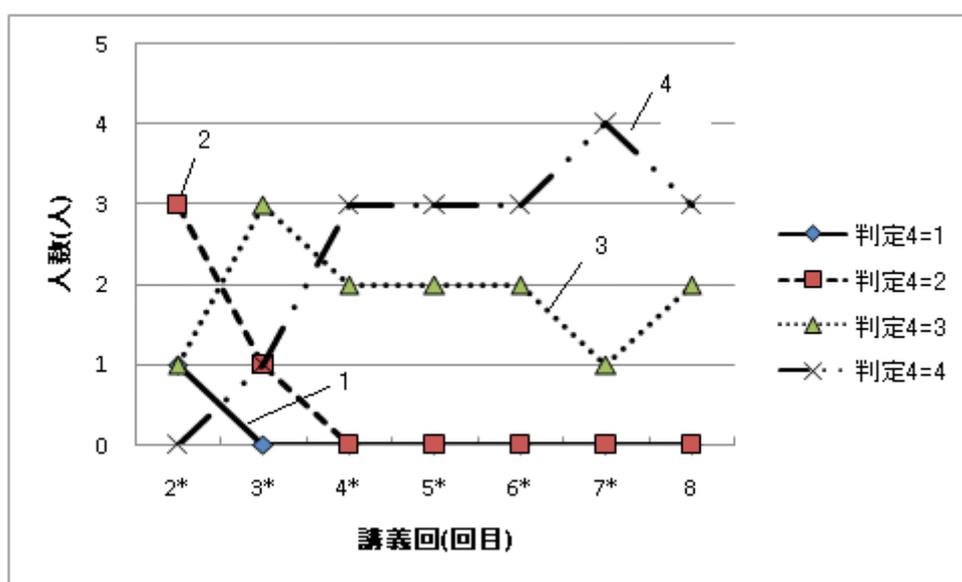


図4.5 授業1(08')の成績Dに対する単位習得不可の危険判定(判定4)の検証結果

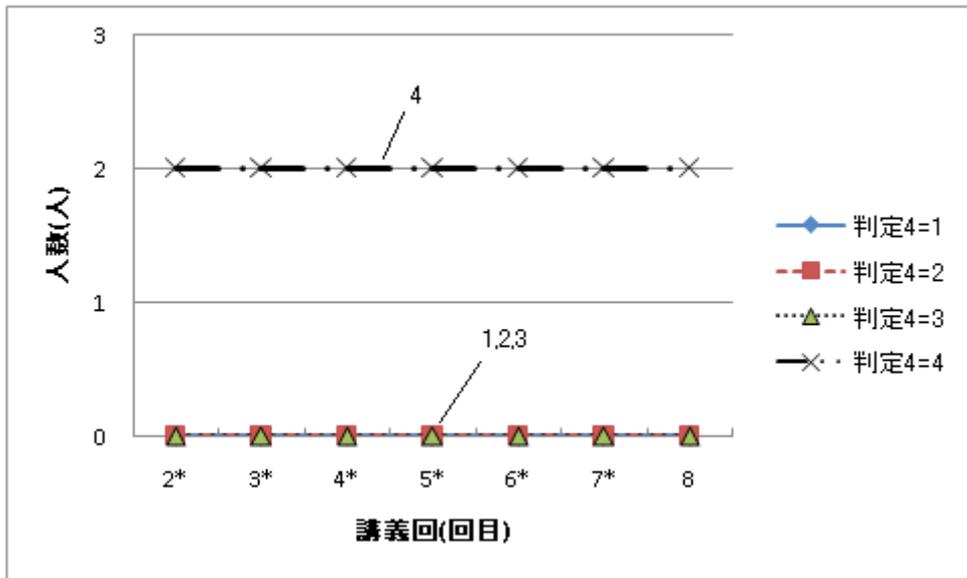


図 4.6 授業 1(08')の成績 X に対する単位習得不可の危険判定(判定 4)の検証結果

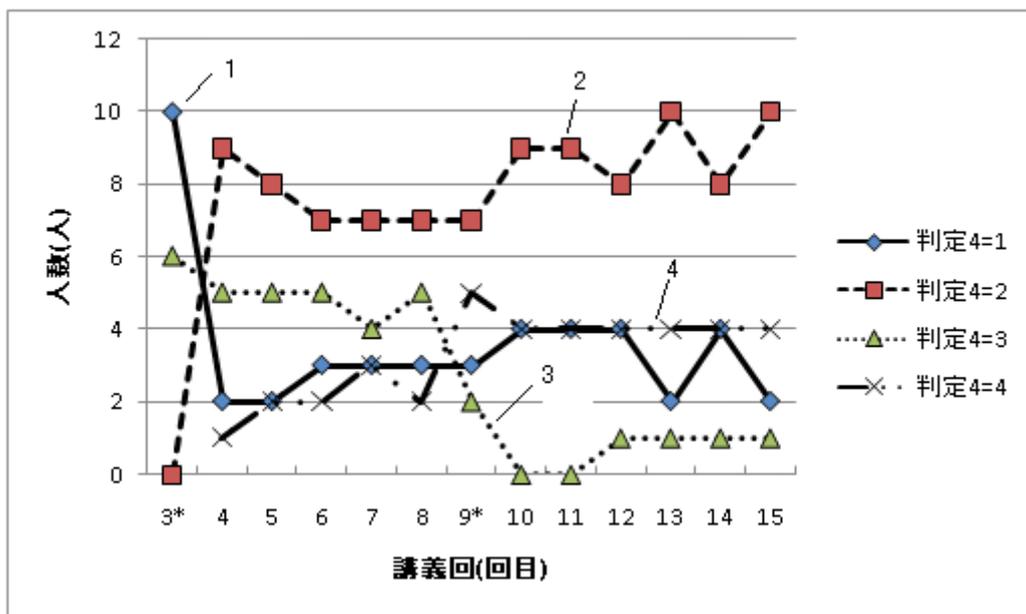


図 4.7 授業 2(08')の成績 D に対する単位習得不可の危険判定(判定 4)の検証結果

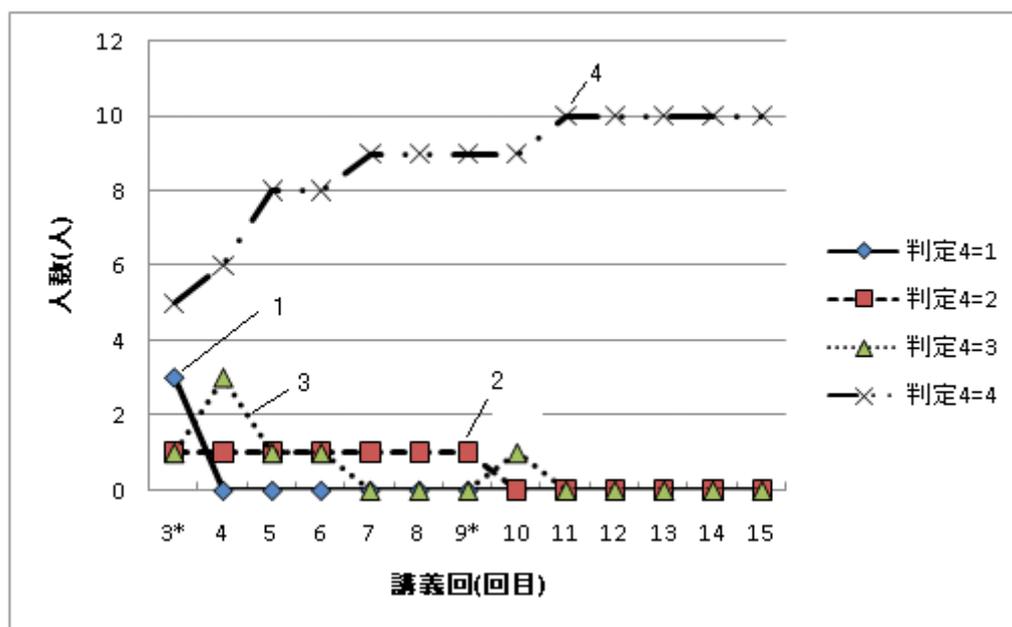


図 4.8 授業 2(08)の成績 X に対する単位習得不可の危険判定(判定 4)の検証結果

授業 1(08)について、成績 D はある程度授業が進み、ある程度課題も課されていれば、判定 4 は十分有効であるといえる。また、成績 X の検証結果については、十分有効であるが判定可能な第 2 回目講義からすべて同じ結果と判定されたため例外として考えた。

授業 2(08)について、成績 D は授業が進むにつれて多少有効であるように判定されているが、危険判定に使えるとはいえない結果である。また、成績 X に関しては、ある程度授業が進んでいるなら十分有効であるといえる。

以上の結果より、成績 D, X に対する単位習得不可の危険判定の有効性は、ある程度授業が進めば成績 X に関してはあるといえる。また、成績 D に関しては、出欠データに加えて、ある程度課題提出データがあれば判定は有効であるといえる。よって、特に課題の多い授業ならば、授業中盤以降、実際に修学指導として危険判定を用いることも可能であると考えられる。

なお、今回の検証では成績 D や成績 X に対しての危険判定が有効であるかを検証の対象としているので、判定 4 で成績 D と成績 X の者以外がどの判定であっても良いものとした。その理由として、成績 S が判定 4 で 4 と判定されたとしても、単位を習得できない者になることに対して危機感が増えるだけでデメリットは無いと考えられるからである。

4.4 授業成績判定傾向と考察

授業が進むことで授業成績判定がどのように変化している傾向があるのかを調べるため、現在進行中の授業3(08'), 4(08')について判定1~3を行った。判定結果を以下の図4.9~4.12および表4.1, 4.2に示す。なお、判定2については判定結果の項目が多いため図での表示を省略した。

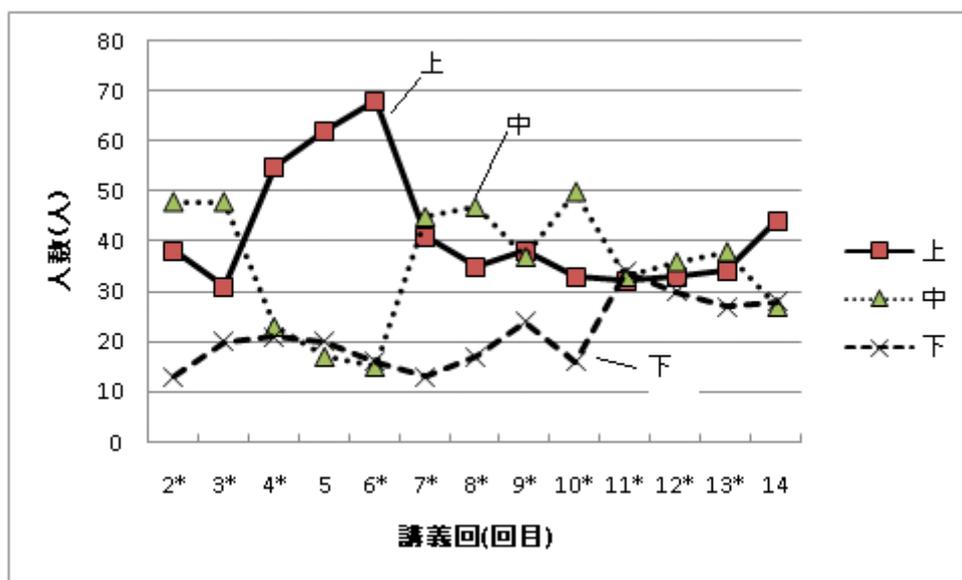


表 4.9 授業 3(08') の判定 1 の判定結果

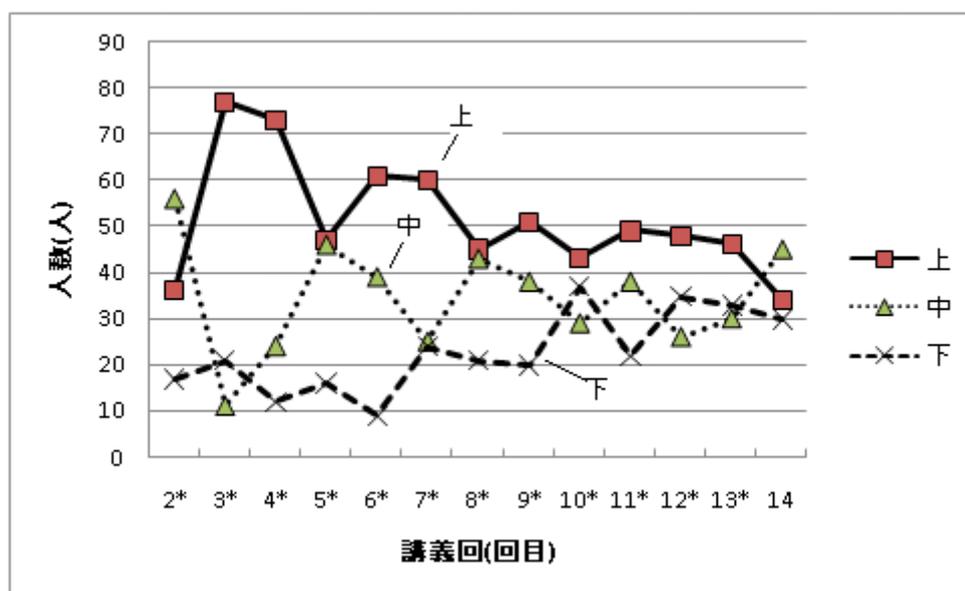


図 4.10 授業 4(08') の判定 1 の判定結果

表 4.1 授業 3(08')の判定 2 の判定結果[人]

講義回	S, A	S, A, B	A, B	A, B, C	B, C	C, D	C, D, X	D, X
1*	判定不能							
2*	0	11	27	3	45	0	10	3
3*	7	9	15	27	21	2	17	1
4*	10	13	32	6	17	0	17	4
5	10	24	28	7	10	9	6	5
6*	12	25	31	7	8	3	9	4
7*	10	24	7	27	18	5	4	4
8*	8	20	7	21	26	6	7	4
9*	8	15	16	6	30	0	17	7
10*	7	17	9	10	40	7	4	5
11*	8	6	19	1	31	15	14	5
12*	8	3	22	2	34	6	15	9
13*	10	4	23	2	33	5	13	9
14	9	15	20	12	15	9	13	6

表 4.2 授業 4(08')の判定 2 の判定結果[人]

講義回	S, A	S, A, B	A, B	A, B, C	B, C	C, D	C, D, X	D, X
1*	判定不能							
2*	0	12	24	2	54	7	7	3
3*	11	19	47	0	11	8	9	4
4*	10	25	38	9	15	5	5	2
5*	13	13	21	27	19	6	9	1
6*	13	25	23	19	20	0	5	4
7*	13	17	30	14	11	11	9	4
8*	10	15	20	27	16	4	8	9
9*	12	14	25	20	18	0	17	3
10*	13	15	15	5	24	0	30	7
11*	19	8	22	22	16	7	8	7
12*	11	18	21	7	17	9	19	7
13*	11	20	17	11	17	16	9	8
14	6	13	15	18	27	10	13	7

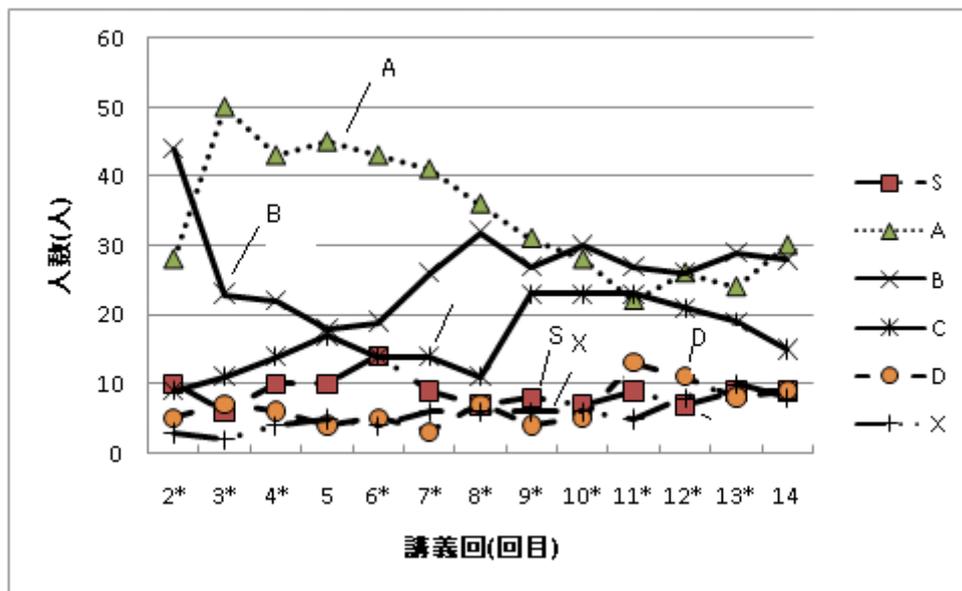


表 4.11 授業 3(08')の判定 3 の判定結果

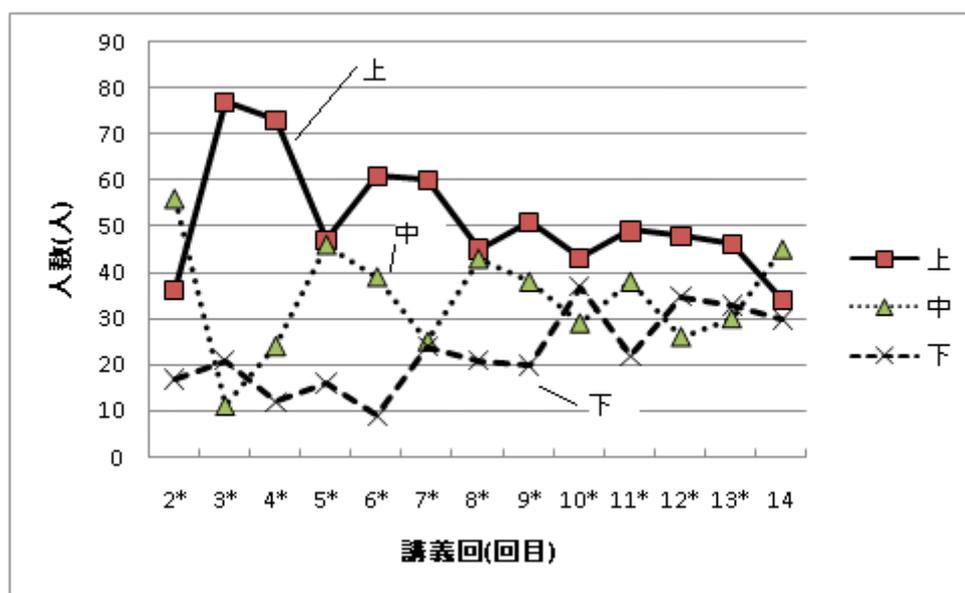


図 4.12 授業 4(08')の判定 3 の判定結果

判定 1 について、授業 3(08'), 4(08')ともに授業序盤から中盤では上の判定が多く、下の判定が少ない。しかし、中盤以降では上や下の判定が、ある特定の値に収束される傾向がある。また中の判定人数については、序盤は上の増減と反対の動きをし中盤以降では下の増減と反対の動きをしている傾向がある。

判定 2 について、判定 1 と同様に序盤では高めの判定が多く、授業が進むごとに低めの判定も増

えている傾向がある。

判定3についても、判定1, 2と同様の結果である。判定1で序盤に上が多かったのは判定3でAであった者が多く、中盤で下がかったのは判定3でCであった者が多い。また、中盤以降成績予想は比較的安定している。

以上の結果より判定1, 2, 3が、このような結果になったことを考察すると、中盤以降が判定人数が安定していることから、正しく判定されたのは中盤以降であると考えられる。また、序盤で成績が高く判定されたのは分析結果にある、下位成績は徐々に出席率が悪くなる傾向があるということから、逆にいえば序盤はまだ出席率が良い方と考えられるので、下位成績であるにもかかわらず序盤では上位や中位成績と判定されたものであると思われる。

また、授業1(08'), 2(08')の検証結果でも、低い成績の判定3の正答率が徐々に高くなったのも、そのためだと説明できる。ただし、判定1, 2の正答率が授業序盤から高めだったことより、序盤の成績判定で上位や中位成績と判定されても、それほど成績予想のずれはないもだと考えられる。

第5章

むすび

本研究では、データマイニングを用いて過去の授業データから学生の修学傾向を分析し、その結果を検証して修学傾向への適応有効性があることがわかった。修学傾向の分析結果から授業成績判定したとき、学生の成績を確実に予想しの中させるものではないが、ある程度授業が進んでおり、課題もある程度進んでいる状態の授業中盤頃以降なら、おおよそなるであろう範囲の成績の予想は十分可能である。さらに、本研究の目的である、特に単位習得不可の成績になる可能性がある者への危機意識を持たせるという点でも、危険判定を行い有効であることを検証した。

しかし、本研究では修学傾向を分析し結果が修学指導への適応有効性があるか調べることを最優先としたので、授業を受けた学生に対して判定した成績をわかりやすく正確に伝えるという点を考慮していなかった。なので、実際に学生に対して判定した結果を伝えるときにも効果があるかが疑問に残る。また、授業内容についても年々変化があると思うので、今回データマイニングによって得られた知識が今後の授業に通用するかを定期的に検討、改善することも考えられる。

今後の課題として、先に述べたように成績判定した結果を正確に伝達、単位習得不可の可能性のある者への危険判定が正しく伝わっているのかが疑問であるのでしっかり伝わるようにしたい。そのため、成績判定や危険判定をよりわかりやすくする表現する必要がある。本研究は、成績S、成績Cや安全、危険といった簡単な文字での表現になっているため、予想した結果を伝えたい相手に伝えても印象に残りづらく効果が薄いと思われる。この問題を解決するために、予想判定を視覚的またはその他の感覚的に表現された理解しやすいモデルを使って表現するということが考えられる。成績判定の見方を3次元的に見る、単位習得不可に対する危機感を表現するように予想された成績に比例して重さを体感するモデルで表現する、といった表現の仕方を変えるだけでも印象が変わったり、興味を引きつけることができるのでより伝達の効率が良くなる。今後、まずこれらのように表現可能で効率的な伝達ができるようなモデルを検討したい。

また、今後授業内容の変化によって学生の修学傾向が変わる可能性があるので、常に授業の変化に対応できるように検討したい。今回の研究では、どの授業でも修学指導が対応できるようにデータマイニングに使った授業データの数は少なくないが、決して多いともいえない。今後、全授業に対応可能であると確認できて、今後の授業内容の変化に対応できるような授業成績の改善の知識が得られるような方法があるかも検討していきたい。

今後上記のように検討，改善をしていき，授業成績の改善のシステムとして実用化できるようにしたいと思う。

謝辞

本研究を進めるにあたり，日頃から多大な御尽力を頂き，御指導を賜りました名古屋工業大学 舟橋健司 准教授，伊藤宏隆 助教に心から感謝致します。

本研究を進めるにあたり，データの提供元である出欠システムコースマネジメントシステムの開発に尽力されました，名古屋工業大学情報基盤センター長 松尾啓志 教授，内匠逸 教授，情報基盤センター教職員の皆様に感謝致します。

本研究は文部科学省特別研究経費『充実した「学びの場」の構築-教員の教育力向上および双方向型教育支援システムの整備』の支援により行われた。ここに感謝の意を表します。

また，研究を進めていく上で御協力を頂きました名古屋工業大学 山本大介 助教，中部大学 岩堀裕之 教授，名古屋工業大学 中村剛士 准教授，愛知大学 福井真二 講師に深く感謝致します。

最後に，本研究を進めるにあたり多大な協力を頂いた舟橋研究室諸氏に深く感謝致します。

参考文献

- [1] マイケルJ.A.ベリー，ゴードンS.リノフ：“データマイニングの手法 営業，マーケティング，CRMのための顧客分析(第2訂版)”，海文堂 (2008-2)

- [2] 嶋田香，平澤宏太郎，古月敬之：“獲得した情報を用いる遺伝的ネットワークプログラミングによるデータマイニング”，情報処理学会論文誌，Vol.46，No.10，pp2576-2586 (2005-10)

- [3] 石井一夫：“図解よくわかるデータマイニング”，日刊工業新聞社 (2004-12)

- [4] 近藤宏，末吉正成：“excel できるかんたんデータマイニング入門”，同友館 (2008-5)

- [5] 平井明夫，松井浩輔：“ひと目でわかる Microsoft Office Excel 2007 データマイニングアドインを使用したビジネスデータ簡単分析術”，日経BP ソフトプレス (2007-4)