

## 1 はじめに

近年、大量の情報が増え続けているために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、埋もれてしまった情報を発見するためにデータマイニング (Data Mining) という技術が目ざされている。そこで、本研究では大量の情報の中から有用な知識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、その結果を修学指導へ適応したとき有効であるかの検証を行う。

## 2 データマイニング

データマイニングは大量のデータとそれを処理できるコンピュータ、ネットワーク環境があり、規模に応じてソフトウェアを使い分けることで使用可能である。また、データマイニングには以下のような使用方法がある。

**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。

**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。

**クラスタリング** 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。

**関連付け** 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。

## 3 データマイニングによる修学傾向分析

本研究では、より正確に学生の修学傾向を分析するために、4つの授業のデータを用いてデータマイニングを行う。各授業でデータマイニングを行い、それぞれの分析結果から共通、類似したものをまとめる。それを本研究の修学傾向分析結果とする。分析結果は以下ようになった。

- 成績 S,A,B は出席率、課題提出率ともに高い。しかし、出欠でいずれかを予想するのは難しい。
- 成績 C,D,X は徐々に出席率、課題提出率が下がる。また、成績 C,D は授業後半からは徐々に増えるのに対し、成績 X は下がったままである。
- 成績への影響は出欠は終盤、課題は中盤が強い。
- 出席や欠席は連続性が強い。
- 課題を早く提出するほど成績は良くなる。
- 出欠よりも課題提出の方が成績への影響が大きい。

## 4 有効性の検証と結果

検証では授業成績を予測し、予測結果から単位修得不可 (成績 D,X) になることに対して注意喚起するものとする。授業成績がわかっている授業を使い予測の正答率を調べ、有効性があるか判定する。判定基準を次のように定める。

判定 1 上位 (S,A,B)、中位 (A,B,C)、下位 (C,D,X) に分けて成績を予測

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

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つを予測

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測

なお、判定 1, 2 は授業データにクラスタリングを行い、各クラスタの特徴から成績を予測する。

### 4.1 授業成績判定の有効性の検証

修学傾向分析結果から予測した授業成績判定が有効であるかを検証するために判定 1~3 を行った。結果を図 1 に示す。検証した結果、範囲的な成績の予測は有効であり、1つの成績を予測する場合もある程度授業が進めば多少有効であった。

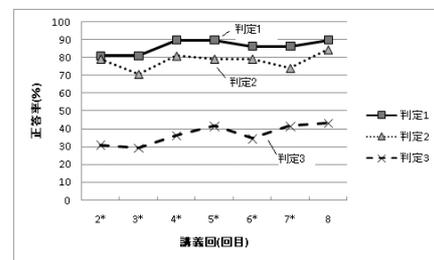


図 1: 授業成績判定結果

### 4.2 修学指導への適応有効性の検証

単位修得不可者 (成績 D,X) に対して修学指導したとき有効性があるのかを検証するため成績 D と X に対して判定 4 を行った。検証した結果、授業中盤以降であれば注意喚起は有効であり、修学指導への適応有効性があることがわかった。

## 5 まとめ

本研究では、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適応することが有効であることを検証した。今後の課題として、学生に修学指導を効果的に行う方法を検討する必要がある。

### 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