

帰納論理プログラミングを用いた授業教室割り当てルールの抽出

大谷紀子研究室

0932070 加藤 由人

1. 研究の背景・目的

学校の授業の時間割編成を最適化するスケジューリング問題については、さまざまな研究がなされており、すでに実用的なソフトウェアも発売されている。しかし、編成された時間割を運用した結果、使用する教室と受講人数が見合わず、開講後に教室を変更せざるを得ない状況が発生することがある。本学部では、每期 20 件～30 件の教室変更が発生しており、教室の手配や情報の掲示などの学生支援センター職員の業務が増加する要因となっている。また、教室を変更した授業についての掲示情報の増加により、掲示板の利便性が低下している。本学部での時間割編成作業や先行関連研究[1]では、時間割編成の際に使用教室については考慮されず、基本的に前年度と同じ教室を割り当てている。しかし、開講される曜日や時限、同時開講科目の有無などの配置条件により受講人数は毎年大きく変動している。本研究では、教室変更の件数を減少させることを目的とし、授業の配置条件と受講人数との関連性を示すルールを帰納論理プログラミング (Inductive Logic Programming; ILP) により抽出する手法を提案する。

2. 帰納論理プログラミング

ILP とは、述語論理上で帰納推論を行う機械学習の手法であり、主に分類問題の解決に適用されている。ILP は、与えられた多くの事例から目標とする概念を抽出する。目標概念として正しい事例を正例、正しくない事例を負例と呼ぶ。ILP では知識表現言語として述語論理を使用する。述語論理は命題論理では記述できない対象間の関係を記述することができるため、複雑な問題構造の記述に有用である。述語論理の表現力により、ILP は事例に関する背景知識を与えられ、複雑な規則の学習を可能としている。本研究では、代表的な ILP システムの一つである Progol を用いる。

3. 適用手法

Progol で学習を行う際、問題に関する記述とバイアスと呼ばれる目標概念や述語の定義に関する記述を入力する。ILP は、一般的に連続値の扱いが苦手であるとされる。本研究で扱う受講人数は連続値で表されるが、閾値を用いて複数の領域に分割し、記号に置き換えることにより、ILP への適用を可能にする。表 1 に受講人数を離散化したクラスと閾値を示す。

表 1：受講人数の離散化

抽選やクラス指定により受講人数の操作が行われる授業については、受講人数を推定する必要がないため、除外する。目標概念を事例の受講人数の規模として設定し、事例ごとの離散化した受講人数の規模を正例として与える。背景知識として、開講曜日・時限、同時開講科目、担当教員、配当学年、必修選択などの科目に付属する情報を与える。さらに科目固有の情報を背景知識として与える。本研究では、複数年のデータを取り扱

クラス	閾値(単位:人)
Huge	300 以上
Large	200 以上 300 未満
Medium	100 以上 200 未満
Small	50 以上 100 未満
Few	50 未満

うため、異なる年度に開講した同一内容の科目を結びつける記述をする。曜日や時限といった配置条件ではなく、科目自体の要因によって生じるルールを抽出することに役立つと考える。表2は目標概念やモード宣言とその定義の一部を示したものである。

表2：モード宣言と定義（一部）

宣言	定義
<code>:- modeh(1, class(+subject, #class))?</code>	subject は class に分類される
<code>:- modeb(*, teach(+subject, #teacher))?</code>	subject の担当教員は teacher である
<code>:- modeb(*, belong(+subject, #basicsubject))?</code>	subject は科目固有情報 basicsubject を持つ
<code>:- modeb(*, required(+subject))?</code>	subject は必修科目である

4. 評価実験

提案手法の有用性を検証するために、評価実験を行った。平成19～23年度の5カ年分の時間割データのうち、4年分の事例データを訓練事例、残りの1年分をテスト事例とするクロスバリデーションにより予測正解率を評価する。予測正解率とは、あるルールが未知事例を分類する際の正しさの程度を示すものである。抽出したルールの本体部に該当する事例数を a、頭部に該当する事例を b としたときの b/a の百分率で求める。平成19～22年度を訓練事例、23年度をテスト事例とした場合の一例を示す。抽出したルールのう

ち高い予測正解率を得ることができたルールと予測正解率を表3に示し、表4には、クラスごとの該当数と正解数、予測正解率を示した。

表3：抽出したルール（一部）と予測正解率

ルール	予測正解率
<code>class(A, large) :- required(A).</code>	100%
<code>class(A, medium) :- with(A, 某特定講義B), g2(A).</code>	66.6%
<code>class(A, few) :- fifth(A).</code>	66.6%

表4：クラスごとの予測正解率

クラス	Huge	Large	Medium	Small	Few	総合
該当数	23	64	46	32	29	234
正解数	1	40	15	15	16	99
正解率	4.3%	62.5%	32.6%	46.9%	55.2%	42.3%

5. 考察

評価実験においては、高い予測正解率は得られなかったが、おおよそ半数の事例において正しく分類できたことから、受講人数規模の予測手法としてはある程度の有用性が確認できた。受講人数規模のクラス分けにより実験結果は大きく変わる。クラス分けを大まかにするほど正解率は上がるが、時間割運用上の実用性は低くなる。今回は、各領域のデータが取りうる範囲を、単純に同程度の大きさの領域に分割して学習に用いた。しかし、受講人数規模のクラス分けをする際に用いる離散化の手法を改善することにより、正解率と実用性の両方の向上が期待されると考える。

参考文献

[1]佐野 裕也, “授業時間表自動作成システム”, 東京都市大学環境情報学部情報メディア学科卒業論文, 2011.