

K-015

学生の得意不得意分野の自動推定と時間割作成への応用

Automatic Estimation of Student's Expertise/Weak Field and its Application to University Course Scheduling

西森友省*
Yusei Nishimori

堀幸雄†
Yukio Hori

今井慈郎‡
Yoshiro Imai

1. はじめに

学生にとって自分の得意不得意分野を知ることが重要である。大学において学生は自分の能力に適していない授業科目を履修してしまうと、単位を取得できなかったり、授業科目の内容を十分に理解できず留年したり、進級や就職活動などに不利になると考えられる。シラバスは大学の授業科目の多様化や日々進歩する技術に対応するため年々内容が変化し、まだ専門知識がない学生や開講されている授業科目を熟知していない学生にとってはシラバスを閲覧するだけでは開講されている授業科目の内容を把握しきれず、学生の能力に適した時間割を作成することは非常に困難であると考えられる。

大学における留年者数の全学生に占める割合である留年率は、1988 年度から増加傾向を示し、2002 年度には全国 89 の国立大学において留年者率が 64.9% となっている [1]。2002 年度以降男子学生における留年者率は減少しているが、女子学生についての留年者率は増加傾向であることが分かっており、留年学生に対するサポートや学生を留年させない取り組みが必要であると考えられる。

多くの大学では専門委員が新生に対して修学について相談を受けるアドバイザー制度を設けている。この制度では学生が直接アドバイスを受けることができるが、学生と専門委員共に時間や労力がかかってしまう。

我々は、学生の得意不得意分野を考慮して学生の能力に適した時間割を作成するシステムを考案、開発した。このシステムでは学生の履修履歴から分析した学生個人の学習能力と授業科目が過去に開講されたときの不合格者数、履修者数を用いた授業科目の難易度を用いることで学生の得意不得意分野を分析し、学生の能力に適した時間割の作成を行うシステムである。これにより、学生の得意不得意分野を反映した時間割を作成し、学生に効率良く学習させることで学生の留年率の減少や学生の授業科目の内容への理解を支援することを目的とする。

本稿では本システムの提案手法、及び評価実験を行った結果について述べる。以下、2 章では本研究の問題の所在を明確化するために、先行研究について述べる。3 章では本提案手法の詳細について述べる。4 章では実際のシラバス、及び学生の時間割を用いて、本システムの有効性の評価を行う。5 章は結論であり、今後の課題をまとめる。

2. 関連研究

我々は、Active Syllabus という活性伝播モデル [2] を利用した時間割自動作成システムを開発している [3]。Active Syllabus では、時間割作成を制約充足問題に帰着し、学生の興味をシラバス閲覧履歴などからユーザプロファイルとして作成する。

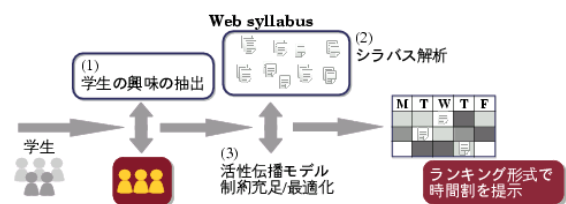


図 1: Active Syllabus の概要

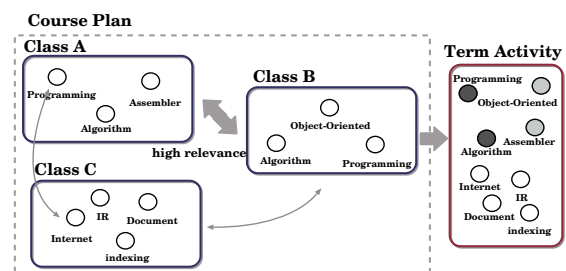


図 2: 科目間ネットワークにおける活性伝播モデル

Active Syllabus は図 2 に示すように科目間ネットワークの活性状態に着目し、科目シラバスに含まれる単語の活性状態をスケジューリング問題の目的関数に用いることで学生の興味に直接マッチしていない授業科目を時間割に含めることができる。時間割作成のスケジューリング問題を解くための探索手法として、ユーザプロファイルに適合する科目群より初期解を求め、その後反復改善法を初期解に適用するというヒューリスティック探索を用いている。Active Syllabus では学生の興味を満たすという点のみを考慮していたため、学生の得意不得意分野を考慮しておらず、提案手法は学生の履修履歴から得意不得意の推定を行なう点が異なる。

高橋らは、フィルタリング技術を活用したシステム Airs (An Individual Reviewing System) を構築している [4]。Airs は e-Learning 教育において、提供されるコンテンツがどの学習者に対しても同一の内容の教材コンテンツを使用するため、学習者の学習レベルによ

*香川大学大学院工学研究科

†香川大学総合情報基盤センター

‡香川大学工学部

て、学習者の感じる難易度の差異が出てきてしまう問題を解決するために構築されている。Airs では協調フィルタリングを活用することで、学習者の学習履歴データをその学習者の特徴データとしてデータベースに保存し、他の類似している特徴データを持つ学習者と比較することで、その学習者の特徴を推測している。この研究では、学習者の感じる難易度の差異を回避するため、学習者の学習履歴を用いる点では本研究と同じであるが、本研究とは学生に提供するものが時間割であることが異なり、大学が開講している授業科目に対して支援できるという特徴がある。

由らは、電子シラバスに現れる用語などを用いた教養教育のカリキュラム分析を行っている [5]。HTML 形式で記述されているシラバスを NIAD-UE が開発しているシラバス XML スキーマを用いて XML に変換し、シラバス項目で科目の情報が主に記されている項目から名詞とその複合語を抽出し、これらをキーワードと位置付けている。そして一科目を構築する文字列におけるキーワードが出現する回数などに基づいてキーワードに重みを与えている。これにより科目の内容を、キーワードの重みを成分とする一つのベクトルとして表現し、科目間の類似度を二つの科目のコサイン類似度を用いることで表現している。この研究では、コサイン類似度を用いて科目間の類似度を算出している点では本研究と同じであるが、本研究では科目間の類似度を用いて時間割を作成することで、カリキュラムを把握できていない学生にも科目間の繋がりを反映した時間割を提供できる。

時間割作成などの組み合わせにおける最適化はスケジューリング問題として良く知られている [6]。スケジューリング問題は病院におけるナースの勤務時間を作成するナーススケジューリング問題 [7][8]、鉄道会社における乗務員割り当て問題など様々な応用事例が報告されている。また、遺伝的アルゴリズムを用いた時間割作成に対する取り組みも多く見られる [9][10]。しかしこれは、授業科目の開講スケジュールであり、教員の都合に合わせて作成される場合が多く、実際に学生自身が作成する時間割とは異なるものである。

3. 提案手法

提案手法は、学生の履修履歴と科目の難易度を用いて、履修可能な科目の適正値を計算する。本節では提案手法の実現方法について説明する。

3.1. カリキュラム分析

大学で開講されている授業科目のシラバスデータを WEB 上で公開されているシラバスから取得する。WEB 上で公開されているシラバスから、対象学科、年次、区分、授業科目、科目コード、科目区分、科目区分 (種類)、教員名、学期、曜日、校時、単位数、授業の目標、授業の概要、授業の方法、成績の評価、受講上の注意、授業計画、教科書、参考書、質問の受付に於けるテキストデータを抽出する。

各シラバスに出現する単語を形態素解析を用いて抽出し、その出現頻度により、授業科目 c_i のシラバスの文書ベクトルを作成する。1 つの授業科目 c_i に出現する

単語を tf-idf により重み付けを行い、 c_i を次のように表わした。

$$c_i = (\omega_{i1}, \omega_{i2}, \dots, \omega_{in})$$

$$\omega_{i,t_j} = tf(t_j, c_i)idf(t_j)$$

$$idf(t_j) = \log(N/df(t_j))$$

ただし、 $tf(t_j, c_i)$ は授業科目シラバス c_i における単語 t_j の出現頻度であり、 N は総科目数、 $df(t_j)$ は単語 t_j が 1 回以上出現する科目数である。

今回我々の所属する科目シラバスデータの香川大学全学共通科目の出現単語数は 17168 個、信頼性情報システム工学科の出現単語数は 8878 個であった。また、下記の単語は除外した。

- 名詞の数、接続、記号
- ひらがな、カタカナ、アルファベット 1 文字の単語、これらは形態素解析に失敗している可能性が高いため除外した。

提案手法では科目間の内容の類似度を用いて学生の得意不得意分野を推定する。科目間の内容の類似度を算出するためにベクトル空間におけるコサイン類似度を用いた。コサイン類似度を用いて科目間の類似度は式 (1) より算出される。この科目間の類似度を用いて、当該学期に履修可能な授業科目と学生が過去に履修した授業科目との関係性を推定し、過去に内容が類似している科目の単位を多く取得できているならば得意分野、そうでなければ不得意分野と推定する。

$$L_{ij} = \frac{c_i c_j}{\|c_i\| \|c_j\|} = \frac{\sum_{k=1}^N \omega_{ik} \omega_{jk}}{\sqrt{\sum_{k=1}^N \omega_{ik}^2} \sqrt{\sum_{k=1}^N \omega_{jk}^2}} \quad (1)$$

図 3 に全学共通科目の科目間の類似度の度数分布表を、図 4 に信頼性情報システム工学科の科目間の類似度の度数分布表を示す。

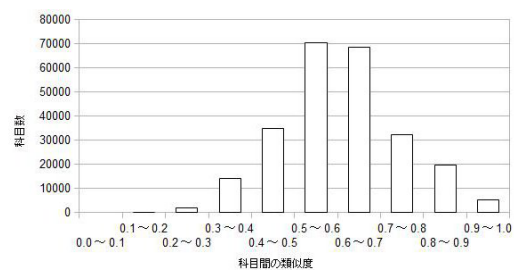


図 3: 全学共通科目における科目間の類似度度数分布表

3.2. 授業科目の難易度と適性値

過去に授業科目が開講されたときの不合格者数 $failed$ 、履修者数 $total$ から授業科目の難易度 $difficult$ を推定する。授業科目の難易度は式 (2) により定義される。

$$difficult = failed/total \quad (2)$$

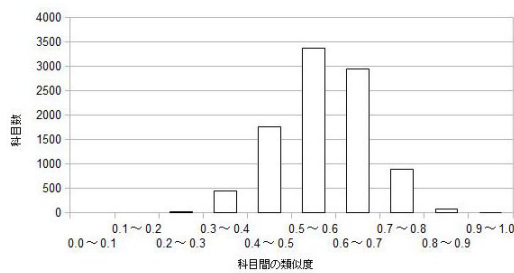


図 4: 信頼性情報システム工学科における科目間の類似度度数分布表

ただし、不合格者数には授業科目を放棄した学生の人数は含まれていない。

図 5 に全学共通科目の難易度の度数分布表を、図 6 に信頼性情報システム工学科の難易度の度数分布表を示す。

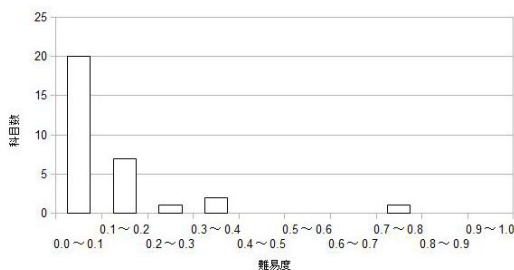


図 5: 全学共通科目における難易度度数分布表

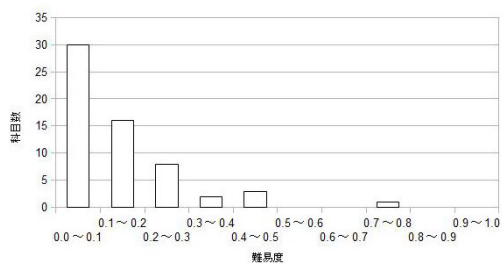


図 6: 信頼性情報システム工学科における難易度度数分布表

授業科目の難易度、科目間の類似度を用いて、授業科目が学生に対してどれだけ向いているかを表す適性値 apt を定義する。ただし、適性値の計算は選択科目にのみ行う。授業科目 a の適性値は式 (3)、(4) により算出される。

$$apt_a = \frac{\sum_{i=1}^n L_{ai_s} \times difficult_{is}}{n} - \alpha_a \quad (3)$$

$$\alpha_a = \frac{\sum_{i=1}^m L_{ai_f} \times difficult_{if}}{m} + difficult_a \quad (4)$$

ここで、 $difficult_a$ とは授業科目 a の難易度、 i_s は単位を取得した授業科目であり、 n はその総数、 i_f は単位を落とした授業科目であり、 m はその総数である。

3.3. 時間割作成

式 (3)、(4) により算出された授業科目の適性値が高い授業科目から履修科目として時間割を作成する。必修科目は必ず履修するものとし、演習や実験など数時間に及ぶ科目は一科目として数える。表 1 に該当学期に履修可能な選択科目一覧を、表 2 に学生により作成された時間割例を、表 3 に提案手法により作成された時間割例を示す。

表 1: 履修可能な選択科目一覧

科目名	適性値 apt	区分
上級英語 I (3)	0.093	全学共通科目
上級英語 I (2)	0.092	全学共通科目
光デバイス	0.055	学部開設科目
データベース	0.050	学部開設科目
数値シミュレーション	0.050	学部開設科目
デジタル信号処理	0.049	学部開設科目
信頼性設計	0.047	学部開設科目
電気電子計測	0.045	学部開設科目
メディア情報処理	0.043	学部開設科目
技術英語 (信頼)	0.041	学部開設科目
通信工学	0.041	学部開設科目
計算機アーキテクチャ	0.036	学部開設科目

表 3 において科目名の横に * のある科目は変更のあった科目である。「ソフトウェア工学 I」、「信頼性情報システム工学実験 I」、「計算機ネットワーク I」、「情報システム演習」、「国際コミュニケーション II」は必修科目であるため時間割に入った授業科目であり適性値は存在しない。表 3 の時間割では「光デバイス」が最も適性値が高く、「通信工学」が最も適性値が低い。今回の時間割の場合、「計算機アーキテクチャ」が学生が作成した時間割に含まれているが、それよりも適性値が高い「信頼性設計」と入れ替わっていることが分かる。

今回は、学生が単位を取得し易いであろう得意分野と推定された授業科目を優先して履修科目としたが、逆に学生が苦手とする授業科目を時間割に入れることで学生の苦手分野の克服を支援することもできる。

4. 評価と考察

4.1. 実験内容

提案手法 apt の有効性を調べるため学生が実際に履修した時間割を用いて、落とした科目に関する精度

表 2: 学生により作成された時間割例

曜日	時限	科目名
mon	2	電気電子計測
tue	2	ソフトウェア工学 I
	3	信頼性情報システム工学実験 I
wed	3	通信工学
	4	計算機アーキテクチャ*
	5	メディア情報処理
thu	1	データベース
	2	数理シミュレーション
	3	デジタル信号処理
fri	1	技術英語（信頼）
	2	計算機ネットワーク I
	3	光デバイス
	4	情報システム演習
集中科目		国際コミュニケーション II

表 3: 提案手法により作成された時間割例

曜日	時限	科目名
mon	2	電気電子計測
tue	2	ソフトウェア工学 I
	3	信頼性情報システム工学実験 I
wed	1	信頼性設計*
	3	通信工学
	5	メディア情報処理
thu	1	データベース
	2	数理シミュレーション
	3	デジタル信号処理
fri	1	技術英語（信頼）
	2	計算機ネットワーク I
	3	光デバイス
	4	情報システム演習
集中科目		国際コミュニケーション II

precision, 誤認率 *misrecognition* を算出した。
精度, 誤認率は式 (5), (6) により算出される。

$$precision = \frac{\text{不可科目回避数}}{\text{不可の科目総数}} \quad (5)$$

$$misrecognition = \frac{\text{取得単位を外した数}}{\text{取得単位総数}} \quad (6)$$

本研究と本研究以外に作成した 5 つの時間割作成手法を精度, 誤認率により比較する。比較対象に用いた時間割作成手法は kw, as, ds, apt+kw, apt+as である。kw は時間割における選択科目を学生の興味のあるキーワードに関連の強い授業科目を優先して時間割を作成する手法である。as とは Active Syllabus を用いて時間割を作成する手法である。Active Syllabus は, 時間割における履修科目の活性値が最大値となる時間割を選択する。活性値は式 (7) により, 算出する。

$$\text{活性値} = \frac{\sum_{i=1}^n \sum_{j=1}^n (\omega_{im} + \omega_{jm}) L_{ij}}{2} \quad (7)$$

ただし, m はユーザプロフィールに含まれるキーワードを示す。活性値が最大となる時間割は以下の処理を行うことで見つける。

1. ユーザプロフィールに適合する授業科目を選択する。これは各科目の文書ベクトルを学生の興味空間に射影したときの絶対値が大きいものから順に選ぶことによって行う。その絶対値を選択した科目の初期活性値とする。それらの科目が初期時間割を構成する。
2. 現在の時間割のうち活性値が最も高い値をもつ科目を 1 つ取り出し, 類似度の高いほかの科目候補を入れ換える。

3. 活性伝播を行い, 活性値がより高い値に更新されたならば, 現在の時間割を 2. で入れ替えた科目のものに更新する。

4. 全科目を探索終了, 活性値が更新されなくなるまで 2. 以下を繰り返す。

apt+kw とは本研究と kw を合わせた手法であり, apt における適性値と kw における学生の興味のあるキーワードに対する関連度の総和が高い授業科目を優先して時間割を作成する。apt+as とは本研究と as を合わせた手法であり, as における初期値を apt にして時間割を作成する。

表 4 は 2009 年度シラバスにおける香川大学工学部の区分とその卒業要件単位数である。

表 5 に実験に使用したデータを示す。

表 5: 実験データ

学科	香川大学工学部信頼性情報システム工学科
学年	1 年後期 ~ 3 年後期
人数	10 名
シラバス	2009 年度

また, 学生の興味を示すキーワードには, 学生が過去に履修した科目内で出現する単語の出現回数が多いものの上位 20 件を学生のキーワードとした。表 6 に学生のキーワード例を示す。

履修履歴のない 1 年前期, 及びほとんど講義を履修しないであろう 4 年次は実験の対象から除外する。

表 4: 実験対象カリキュラム

区分			卒業要件単位数	
全学共通科目	主題科目		8 単位以上	24 単位以上
	教養ゼミナール（選択）		（2 単位）	
	共通科目		8 単位以上	
	健康・スポーツ科目（選択）		（2 単位）	
	高学年向け教養科目（選択）		（4 単位）	
	外国語科目	初修外国語	（選択 1 種類）（4 単位）	
		既修外国語	6 単位以上	
小計		30 単位以上		
学部開設科目	工学教養科目	多角的思考能力	8 単位以上	
		コミュニケーション能力	6 単位以上	
		数理の基礎能力	10 単位以上	
	専門科目	専門基礎	30 単位以上	
		専門専攻	32 単位以上	
		卒業研究	6 単位以上	
	自由科目		6 単位	
	小計		98 単位以上	
合計		128 単位以上		

表 7: 精度

	1 年後期	2 年前期	2 年後期	3 年前期	3 年後期	全体平均
kw	0.25	0	1	0	0.33	0.20
as	0.25	0	1	0.13	0.10	0.18
ds	0.50	0	1	0.13	0.39	0.32
apt	0.50	0	1	0.25	0.26	0.33
apt+kw	0.25	0	1	0.13	0.33	0.25
apt+as	0.25	0	1	0.13	0.10	0.18

表 6: キーワード例

回路	測定	データベース	信頼性	実験
英語	力	交流	積分	演習
法	物理	電気	関数	確率
分布	行列	電気回路	構造	交流回路

4.2. 実験結果と考察

精度、誤認率の実験結果を表 7、表 8 に示す。表中の値は有効数字 2 桁で表したものである。

表 7 から、精度では apt が最も値が良く、as と apt+as が最も値が悪い結果となった。また、次点が ds であることから授業科目の難易度を用いることによって時間

割の精度を良くすることができることが考えられる。また表 8 から、誤認率では as と apt+as が最も値が良く、apt が最も値が悪い結果となった。このことから学生の興味を時間割に考慮すれば時間割の誤認率を良くすることができると考えられる。また、apt の結果が最も値が悪いことから現在の学生は授業科目が自分に適しているかどうかを考慮せずに時間割を作成していると考えられる。

また、2 年後期では精度、誤認率共に全て同じ値を取っていることが分かる。これは 2 年後期に開講される授業科目に必修科目が非常に多く存在するため、ほとんどの学生が同じ時間割になってしまうことが原因である。

それぞれの手法の学生の興味や時間割の活性値を算出し、順位付けを行った。その結果を表 9 に示す。キーワードの値高いことは、学生がこれまでに履修した科目との関連度が高いことを示し、活性値が高いことは

表 8: 誤認率

	1 年後期	2 年前期	2 年後期	3 年前期	3 年後期	全体平均
kw	0.57	0.87	0.59	0.23	0.37	0.52
as	0.57	0.77	0.59	0.10	0.08	0.42
ds	0.61	0.79	0.59	0.25	0.45	0.54
apt	0.64	0.79	0.59	0.27	0.53	0.56
apt+kw	0.57	0.86	0.59	0.25	0.37	0.53
apt+as	0.57	0.77	0.59	0.10	0.08	0.42

時間割内に関連の高い科目が多く含まれていることを意味する。

表 9: キーワード、及び活性値の順位

順位	キーワード	活性値
1	kw	as
2	apt+kw	kw
3	as	apt+kw
4	ds	apt
5	apt+as	apt+as
6	apt	ds

キーワードでは kw が、活性値では as が最も高いことは自明であるが、apt+kw がキーワードでは次点に、活性値では 3 番目に良いことが分かった。このことから、学生の興味や先行研究である Active Syllabus などの特性を活かしながら本研究を有効に用いるには、apt+kw を用いれば良いと考えられる。

5. おわりに

本稿では、学生の履修履歴から学生の得意、不得意分野を推定し、学生に適した時間割を提供する時間割自動作成手法について述べた。本システムの目的では学生の得意不得意を分析することが重要であり、評価実験では誤認率より精度の方が良い手法の方が好ましい。よって、本提案手法が学生の得意、不得意分野を推定する上で有効であることが考えられる。また、過去の学生の不合格率を考慮することで学生が単位を取得できるであろう科目を推定できることが分かった。しかし、本システムでは入学したばかりの学生に対しては履修履歴が存在しないため本システムを提供できない。この場合は Active Syllabus を用いることで対処する方法が考えられる。

今後の課題として、学生の得意、不得意分野をさらに正確に推定すること、そして工学部カリキュラムだけでなく他学部他学科の学生を対象に評価実験を行い、他学部他学科の本システムの有効性を確認することである。

参考文献

- [1] 内田千代子: 大学における休・退学、留年学生に関する調査 第 27 報, 全国大学メンタルヘルス研究会報告書 第 28 回, 13-25, 2007.
- [2] J.R. Anderson: A spreading activation theory of memory, Journal of Verbal Learning and Verbal Behavior, 22, pp.261-295, 1983.
- [3] 堀 幸雄, 中山 堯, 今井 慈郎: 科目ネットワーク上の活性伝播を用いた時間割の自動生成システム, 情報処理学会論文誌, Vol.52, No.7, pp.2332-2342, 2011.
- [4] 高橋泰樹, 松澤俊典, 山口未来, 土肥紳一, 和田雄次: 履歴データを活用した教育支援システム Airs, 情報科学技術フォーラム一般講演論文集 5(4), 367-368, 2006-08-21.
- [5] 由谷真之, 森幹彦, 喜多一: 電子シラバスを用いた大学教養教育のカリキュラム分析, 情報科学技術フォーラム一般講演論文集 4(4), 315-316, 2005-08-22
- [6] A. Caprara, M. Fischetti, P. Toth: A heuristic method for the set covering problem, Operations Research, Vol.47, pp.730-743 (1999).
- [7] 池上 敦子, 丹羽 明, 大倉 元宏: 我が国におけるナース・スケジューリング問題, オペレーションズ・リサーチ, Vol.41, No.8, pp. 436-442 (1996).
- [8] Aickelin, U. and Dowsland, K. A.: An indirect genetic algorithm for a nursescheduling problem. Comput. Oper. Res. 31, 5, 761-778 (2004) .
- [9] 鄭 信雨, 平井 敏之, 村尾 了, 新妻 真輔, 品野 勇治: 分散遺伝的アルゴリズムとローカルサーチを併用した大学の時間割作成システム, 情報処理学会研究報告. MPS, 数理モデル化と問題解決研究報告 2009(19), 173-176, 2009-02-26.
- [10] 難波 道弘: 分散遺伝的アルゴリズムを用いた時間割作成, 山梨英和大学紀要 8, 31-38, 2009.