

## 履修履歴を用いた科目推薦システム

西森友省 †

堀幸雄 †

今井慈郎 ‡

† 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

## 1 はじめに

大学において学生は大学から配布されるシラバスを用いて時間割を作成する。学生は大学が指定する卒業要件単位数を満たす時間割を作成しなければならない。しかし、シラバスは大学の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する。故に、まだ専門知識がない学生や開講されている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科目の内容を理解できず、時間割を作成することは容易ではない。また、卒業要件単位数を満たすだけでなく、自分の得意分野を伸ばしたり、不得意分野を克服するような学習が望ましいと考えられる。

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。

我々は学生の履修履歴から得意不得意分野を推定し、履修可能な科目群から得意な科目、不得意な科目を推薦するシステムを提案、開発した。これにより、学生が自らの得意不得意分野を理解し、体系的な学習や科目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章では本システム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章では本システムの詳細について述べる。4章では実際のシラバス、学生の履修履歴を用いて、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

我々はこれまでに学生の興味に基づいて時間割を自動的に作成するシステム Active Syllabus を開発した [1]。これまでの Active Syllabus では科目同士の内容のみを考慮していたが、本システムでは学生の履修履歴から得意不得意を推定するものである。

高橋らは、フィルタリング技術を活用したシステム Airs を構築している [2]。Airs では学習履歴に協調フィルタリングを活用することで、その学習者の特徴を推測

している。学習者の履修履歴を用いている点では同じであるが、本システムでは実際の大学のシラバスを用いており、時間割作成を支援することができる。

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 [3]。科目間の類似度をコサイン類似度を用いることで表現している。科目間の類似度を用いている点では同じであるが、本システムでは履修履歴から時間割作成を支援できる。

## 3 提案方式

## 3.1 カリキュラム分析

大学で開講されている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、質問の受付に関す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、科目  $c_i$  のシラバスの文書ベクトルを作成する。1つの科目  $c_i$  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、 $tf(t_j, c_j)$  は科目シラバス  $c_i$  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。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。

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

Course Recommendation System Based on Learning History of Students

†Yusei Nishimori †Yukio Hori ‡Yoshiro Imai

3.2 科目の成績推定

科目間の類似度, 学生の履修履歴を用いて, 当該科目を学生が履修した場合の成績を推定する.

香川大学の成績評価は表 1 のような 5 段階評価となっている. 成績推定には GPA 算出時の重みを用いた.

表 1: 成績評価

評語	評点	GPA の重み
秀	90 点以上	4
優	80 ~ 90 未満	3
良	70 ~ 80 未満	2
可	60 ~ 70 未満	1
不可	60 未満	0

科目 A の成績の推定値  $Value(A)$  は式 (3) により算出される.

$$Prob(A, a) = \frac{L_{Aa}}{\sum_{i=1}^N L_{Ai}} \quad (2)$$

$$Value(A) = \sum_{i=1}^N Grade(i) Prob(A, i) \quad (3)$$

ここで,  $N$  は履修した科目数,  $a$  は過去に履修した科目である.  $Grade(i)$  は科目  $i$  の GPA の重みの値である.

4 評価と考察

本システムの有効性を調べるため実際の学生の履修履歴を用いて, 推定した成績と実際の成績との絶対平均誤差を算出した. 評価実験に用いたデータを表 2 に示す.

表 2: 評価実験データ

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	8 名
シラバス	2009 年度
対象科目数	595 科目
推定科目数	340 科目

無作為に推定した場合の評価実験の結果を表 3 に, 本システムで推定した場合の評価実験の結果を表 4 に示す. 表 3, 表 4 から無作為に成績を推定した場合より, 本システムを利用して推定した場合の方が成績の絶対平均誤差が低いことが分かった.

5 まとめ

本稿では, 学生の履修履歴からの成績推定について述べた. 評価実験から, 本システムは得意不得意分野を考慮して時間割作成を行いたい学生にとって有効である

表 3: 無作為に推定した場合の評価実験結果

	1 年後期	2 年前期	2 年後期	3 年前期	3 年後期
学生 1	1.63	1.20	1.60	2.00	1.73
学生 2	1.53	1.40	1.40	1.78	1.56
学生 3	1.13	2.20	1.10	0.92	2.44
学生 4	1.33	2.00	1.38	1.60	1.40
学生 5	1.43	1.00	1.20	1.70	1.80
学生 6	1.38	1.60	1.40	1.31	2.11
学生 7	2.14	1.00	1.70	1.63	1.22
学生 8	1.50	2.00	1.57	2.22	1.57

表 4: 本システムで推定した場合の評価実験結果

	1 年後期	2 年前期	2 年後期	3 年前期	3 年後期
学生 1	0.38	0.80	1.30	0.89	1.00
学生 2	0.63	1.00	1.20	1.11	0.78
学生 3	1.00	0.80	1.00	1.17	0.67
学生 4	0.83	0.50	0.63	1.20	0.80
学生 5	0.71	0.60	0.80	1.20	1.30
学生 6	0.50	0.20	0.70	0.69	0.33
学生 7	1.14	0.40	1.10	0.63	0.89
学生 8	1.13	1.00	1.70	1.11	0.57

と考えられる. 今後の課題として, 様々な成績推定法と比較し, 最も学生の得意不得意分野を推定できる推定法を調査すること. 及び, 実験人数を増やすことでより正確な評価実験を行う予定である.

謝辞

本研究の一部は IPA 情報処理推進機構未踏 IT 人材発掘・育成事業の支援を受けて行われています.

参考文献

- [1] 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.
- [2] 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.
- [3] 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