

履修授業レコメンドシステムの設計と実装

鴻野弘明† 山本知典‡ 武田圭史‡ 村井純‡

†慶應義塾大学総合政策学部 ‡慶應義塾大学環境情報学部

1 はじめに

近年の情報爆発に伴い、多様な情報から適切な情報を解析抽出することにより新しい情報の価値を見出す研究が行われている。しかし大学などの教育機関においては、学生の目的に応じて効率的な授業の履修選択ができていないなど、未だ十分に活用できていないのが現状である。本研究では、複数のフィルタリング手法を組み合わせることで、学生に応じた効率的な授業履修を実現する手法およびシステムを提案する。

本システムは大学のシステムと連携することによって、履修履歴データを取得、解析し、科目の分類、難易度により授業のレコメンドをする。本手法を実際に学生に提供することによって効率的に授業をレコメンドできるか検証する。

2 背景と目的

大学で授業の履修申告を行う際、学生は自らさまざまな授業について調べ、候補の授業を絞り込み、その中からどの授業を実際履修するかを検討する必要がある。そのため、授業の選択に多くの時間を消費してしまい、適切な授業選択ができていない場合もある。また、学生が閲覧できる情報には限りがあり、最適な判断をすることが困難である。

本研究では、授業の性質と学生の趣向や嗜好を考慮したレコメンドシステムを設計、実装し、効率的な履修授業の選択を可能とすることを目的とする。

3 先行研究

レコメンドに用いられるフィルタリング手法についての研究は多く存在する。本章では協調フィルタリング、内容ベースフィルタリングという二つの代表的な考え方について述べる。

3.1 協調フィルタリング

協調フィルタリングとはユーザと他ユーザの相関やユーザのアイテム評価値を基に嗜好の予

測を行う手法であり、アイテムベースとユーザベースの2種類に分けられる。アイテムベースとはアイテム同士の相関を求め、アイテムに類似した評価のアイテムBをレコメンド対象とする手法である。ユーザベースとはユーザ同士の相関を求め、対象ユーザに類似したユーザBが実際に利用したアイテムをレコメンド対象とする手法である。協調フィルタリングはユーザの評価値や利用履歴を用いてフィルタリングを行うため、コールドスタート型である。

3.2 内容ベースフィルタリング

内容ベースフィルタリングとは各アイテムについての属性を予め決めておき、ユーザに利用されたアイテムと似た属性を持つアイテムをレコメンド対象とする手法であり、ウォームスタート型である。

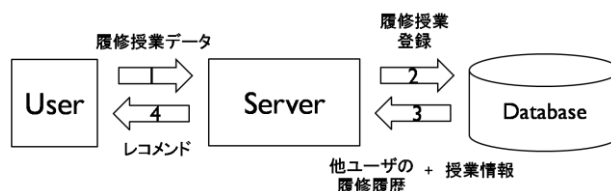
4 アプローチ

本研究では協調フィルタリング手法と内容ベースフィルタリング手法を用いて、筆者の所属する大学の学生を対象としたレコメンドシステムを作成する。本章ではレコメンドに用いる手法の具体的な実装方法及び用いるアルゴリズムについて述べる。

4.1 実装手法

本研究ではPHP言語とMySQLデータベースを用いてシステムの実装を行い、Webをユーザのインターフェースとして用いる。以下にシステムの簡単な概要図を示す。

図1：システムの概要



ユーザはブラウザからアクセスし、用意されたフォームから大学システムで取得した履修授業の履歴ファイルを送信し、ログインすると、PHPのプログラムは履歴データファイルを解析し、データベースに格納する。既にデータベースに登録されているユーザの履修授業の履歴情報などを基に4.2節で述べるアルゴリズムを用いてレコメンドする科目を算出し、ユーザのブラウザ

Design and Implementation of the course recommendation system

Hiroaki Kono†, Tomonori Yamamoto‡, Keiji Takeda‡, Jun Murai‡

†Faculty of Policy Management, ‡Faculty of Environment and Information Studies

に表示する。表示方法については一見して理解のしやすいタグクラウドを用いる。

4.2 フィルタリング手法

本研究において用いるフィルタリングの手法は協調フィルタリングと内容ベースフィルタリングを合わせた、ハイブリッド型のフィルタリング手法である。本節ではそれぞれのフィルタリング手法についての具体的な実装アルゴリズムと、その組み合わせの方法について述べる。

協調フィルタリング手法において用いるアルゴリズムを以下に示す。

$$C_{ab} = \frac{\sum (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum (a_i - \bar{a})^2} \sqrt{\sum (b_i - \bar{b})^2}} \quad \dots(1)$$

$${}_a D_i = \bar{a} + \frac{\sum (h_i - \bar{h})C_{ah}}{\sum |C_{ah}|} \quad \dots(2)$$

a_i はユーザ a が科目 i に付けた評価、 \bar{a} はユーザ a が付けた評価の平均、 C_{ab} はユーザ a とユーザ b の相関係数を表し、 ${}_a D_i$ はユーザ a に対する科目 i についての推薦度合いを表している。

図2: 協調フィルタリングの例

	A	B	C
1	3	2	3
2	5	4	1
3	1	3	2
4		5	2

例として、図2に示すようにユーザ A, B, C が科目 $1, 2, 3, 4$ について図のように評価し、ユーザ A の科目 4 に対する評価を予測するには、式(1)から $C_{AB} = 0.5$ 、 $C_{AC} = -0.5$ 、また式(2)から

$$\begin{aligned} {}_A D_4 &= \bar{a} + \frac{\sum (h_4 - \bar{h})C_{Ah}}{\sum |C_{Ah}|} \\ &= 3 + \frac{3.5C_{AB} + 2C_{AC}}{|C_{AB}| + |C_{AC}|} \quad \dots(3) \\ &= 4.75 \end{aligned}$$

式(3)のようになり、相関の強いユーザとより近く、相関の弱いユーザとより遠い予測値となる。

内容ベースフィルタリング手法において用いるアルゴリズムは、著者の意図的な操作をできる限り介さないレコメンドを行うため、予め決められている属性のみを適用することにした。つまり、必修である科目についてレコメンド内容に介入を行うなどに留める。

協調フィルタリングの結果について内容ベースフィルタリングの結果を上書いた結果が最終的にユーザにレコメンドされる値となる。

5 実験

本研究では実際に作成したレコメンドシステムを学生に利用してもらい、本章で述べる手法を用いてレコメンドシステムの有用性について測定する。

レコメンドシステムの評価に用いられる手法として、レコメンドされるアイテムとユーザの嗜好との合致度を測定する、というものが挙げられる。しかしながらレコメンドシステムの有用性には、ユーザの嗜好との合致度の他にユーザの予想と反してはいるが、ユーザにとって良いアイテムという要素も大きく影響すると考えられる。

以上を踏まえて本研究ではユーザにレコメンドされた科目それぞれについてのユーザの評価に加え、本レコメンドシステムを用いなかった場合の履修授業予定と比較し、どのように変化したかも評価の対象とする。

レコメンドされた授業の評価予測値配列とユーザの実際の評価値配列に相関が有意が見られると考えられる。

6 参考文献

- [1] 小出幸典 2010 「リコメンデーションシステムの構築と評価方法に関する研究 - 履修科目選択支援システムの例 -」 慶應義塾大学大学院 政策・メディア研究科 修士論文
- [2] 神宮敏弘 2006 「推薦システム-情報過多時代をのりきる」 『情報の科学と技術』 56 巻 10 号 pages.452-457
- [3] 神宮敏弘 2007 「推薦システムのアルゴリズム (1)」 『人工知能学会誌』 22 巻 6 号 pages.826-837
- [4] 増井俊之 2005 「UNIX MAGAZINE 2005. 12」 インターフェイスの街角 (93) - 本棚演算
- [5] Barry Schwartz, Andrew Ward, John Monterosso, Sonja Lyubomirsky, Katherine White, Darrin R. Lehman "Maximizing Versus Satisficing: Happiness Is a Matter of Choice" Journal of Personality and Social Psychology 2002, Vol. 83, No. 5, 1178 -1197