

# K-037 eラーニングにおけるレコメンデーション技術の実装 — 苦手克服に貢献する教材の抽出 —

Implementation of recommendation technology in e-learning  
- The extraction of the teaching materials contributing to weak point conquest -

浜詰 祐馬<sup>†</sup>      和田 雄次<sup>‡</sup>      土肥 紳一<sup>‡</sup>  
Yuma Hamadume      Yuji Wada      Shinichi Dohi

## 1. はじめに

現在多くの高等教育機関や企業等の教育において、eラーニングが盛んに取り入れられている。この背景には、WBT(Web Based Training)という考え方が注目を集め、多くの学習管理システムLMS(Learning Management System)が登場し世界標準規格であるSCORM(Sharable Content Object Reference Model)が提案されたことでその普及に拍車をかけたことが挙げられる[1]。しかし学習者にとってeラーニングを利用することは、紙媒体で学習を行うこと以上のメリットがあるかどうか議論が分かれている[2]。一方、AHS(Adaptive Hypermedia System)による適応型の学習支援システムは、従来のeラーニング学習に比べ学習効果などの面で優れている事が明らかにされているが、オンライン小テストの結果を基に学習コースの推薦を行う事等に留まっている[3]。又、昨本年研究室で開発された双方向推薦システム[4]は、学習履歴情報を基に関連する教材コンテンツを推薦するシステムであり、アンケート調査によって学習効果が確認出来た(2章)。しかし、双方向推薦システムは特定の学習者に適応した教材コンテンツを推薦する機能は持っていない。そこで本研究では、学習履歴情報を基に特定の学習者に適応した“苦手科目を克服する教材コンテンツ”(以下推薦教材)を推薦する協調学習推薦システムの開発を行う。本システムは各学習者との類似度及び各教材コンテンツの閲覧回数から各教材コンテンツを閲覧する回数を予測し、その値が大きい教材コンテンツを苦手克服に有効な教材コンテンツとして推薦するものである。又、学習履歴を持たない新規学習者に対しては、属性情報を利用する事で推薦を行う特徴を有する。

## 2. AIRS

AIRS(An Individual reviewing System)[5]とは、本学で開発している個別復習支援システムである(図1)。AIRSは学生の目線で教材(以下教材コンテンツ)の作成を行い、分かり易い教材コンテンツを配信する事で、復習の促進を促す目的で開発された。その中で提供される教材コンテンツは本学の

東京電機大学大学院 情報環境学専攻 情報環境工学専攻<sup>†</sup> 東京電機大学 情報環境学部<sup>‡</sup>  
<sup>†</sup> Graduate School of Information Environment, Tokyo Denki University <sup>‡</sup> Graduate School of Engineering, Tokyo Denki University



図1 AIRSトップページ

開講科目である「情報処理の基礎」と「データベースシステム」の講義を基に作成された。教材コンテンツはwebページとして表示されるページを単位として配信する。

### 2.1 教材コンテンツ

講義データベースシステムの教材コンテンツ「データベースとは」を図2に示す。このように新しく学習するキーワードの解説や、mysqlのクエリを発行する手順等を示したのになっている。特徴としては、1つの教材コンテンツにつき3種類の表現方法をもつ事である。これは、学生によって分かり易い表現は異なる為、簡潔な表現、詳細な表現、そして動きによる表現を用いてそれぞれ作成される。



図2 教材コンテンツ

### 2.2 学習履歴

図3に示すように、学習履歴は日時、学習者の識別ID、教材コンテンツの識別IDからなる。1行目のaccess\_timeは20051013134253, user\_idは0, content\_idは43となっている。これは2005年10月13日13時42分53秒にid0番の学習者がid43

番の教材コンテンツを閲覧した事を示している。2005年から2008年までの運用で、学習履歴は延べ2万7000件を超える数が収集されている。

vector_id	access_time	user_id	content_id
1	20051013134253	0	43
2	20051013134254	0	19
3	20051013134258	0	46
4	20051013134300	0	1
5	20051013134300	0	1
6	20051013134307	0	1
7	20051013134309	0	4
8	20051013134310	0	7
9	20051013134314	0	88

図3 学習履歴の一部

### 3. 双方向推薦システム

双方向推薦システムは学習履歴から教材コンテンツ間の関連度を導き出し、その値が大きい教材コンテンツを推薦するシステムである。図4ではAIRSを利用して「代入の基本」を閲覧している学習者に対して、「代入の基本」と関連の強い「変数の種類」と「while文の基本」を推薦する事を表している。

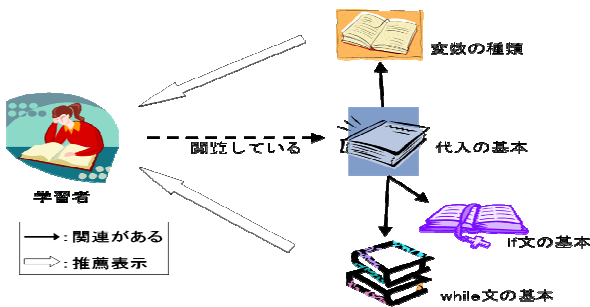


図4 双方向推薦システム

又、昨年AIRSと連携して実施したアンケート調査では、復習が「時間短縮出来た」という意見が寄せられ学習効果を示す事が出来た(表1)。逆に「講義資料を見た方が良い」という意見も見受けられたが、これは教材コンテンツに対する意見だと考えられる。教材コンテンツは比較的簡潔に表現されている為、講義について詳細に表現している講義資料を参照する事は自然である。

表1 “双方向推薦システム” アンケート調査

「推薦結果は適切」or 「まあ適切」と答えた人	
22名 = 61%	
「適切」or「まあ適切」 と答えた人の意見	「適切でなかった」 と答えた人の意見
次にどこを見るべきか分 かり易く時間短縮出来た	講義資料や授業プリ ントを見た方が良い

### 4. 協調学習推薦システム

本システムは、次の仮定に基づいている。すなわち閲覧回数の多い教材は苦手克服に有効であるという事である。なぜならば、何度も閲覧しなければ理解出来ない程難しく、又学習効果が期待出来る可能性が高いと考えたからである。そこで閲覧回数の測定と、閲覧回数を予測する方法を述べる。

#### 4.1 閲覧回数の測定と予測

閲覧回数の測定は、学習履歴を用いる。図3の学習履歴におけるuser\_idとcontent\_idを用い、誰がどの教材コンテンツを何度閲覧したかという表を作成する。その結果が図5に示す表である。1行目のuser\_idは0, con1は53, con2は0, con3は0, con4は49・・・となっている。これはid0番の学習者がid1の教材コンテンツを53回, id2の教材コンテンツを0回, id3の教材コンテンツを0回, id4の教材コンテンツを49回閲覧した事を示している。カラムの数は教材コンテンツの総数、行の数は学習者の総数である。

user_id	con1	con2	con3	con4	con5	con6	con7
0	53	0	0	49	0	0	40
1	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0
3	2	0	0	1	0	0	2
4	9	0	0	7	0	0	5
5	5	0	11	1	0	0	0
6	4	0	0	0	0	0	1
7	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0
9	9	0	0	12	0	0	1
10	0	0	0	0	0	0	0

図5 閲覧回数を示す表

次に、閲覧回数の予測を行う。閲覧回数の予測には、他の学習者の学習履歴を用いて行う。そこで協調フィルタリングの1手法である相関係数法[6]を利用する。協調フィルタリングとは、例えばデータベースを苦手としている学習者Aがいたとする。Aさんは期末試験に向け、データベースの苦手を克服する為にAIRSを利用して復習を行っている。他の学習者BもAIRSを使って復習を行っており、AさんとBさんは学習のパターンが類似しているとする。この時、Aさんが復習に行き詰まったら、Bさんから苦手克服に有効な教材を教わることが出来る。相関係数法では学習パターンの類似している学習者を発見する為、類似度の計算を行う。

#### 4.2 類似度の抽出

本研究における類似度とは、各教材コンテンツに対する閲覧回数が似ている度合を表す。例えば図6に類似度を求める例を示す。学習者とBさんCさんは学習履歴を持っている

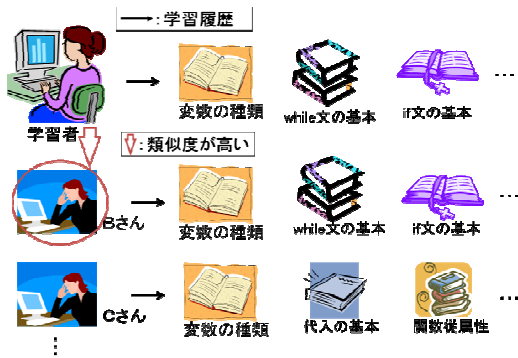


図6 類似ユーザ抽出の例

とする。この時、学習者は変数の種類、while文の基本、if文の基本を閲覧。Bさんも同様に変数の種類、while文の基本、if文の基本を閲覧。Cさんは変数の種類、代入の基本、関数従属性を閲覧しているとする。この場合、学習者とBさんが閲覧した教材は類似している為、学習者とBさんは類似度が高く、学習者とCさんの類似度は低いといえる。

### 4.3 閲覧回数の予測

4.2節で求めた類似度を使って、予測閲覧回数を抽出する方法を図7にて説明する。図7では学習者aが提案システムを利用して推薦を受ける対象者とし、学習者bの閲覧回数は教材Aを3回、教材Bを9回、教材Cを5回、教材Dを2回閲覧しているとする。この時、学習者bと学習者aの類似度が高いならば、学習者aが教材Bを閲覧する回数は9回に近く、Cを閲覧する回数は5回に近いと予測できる。ここで求めた予測閲覧回数の高い教材を推薦教材とする。

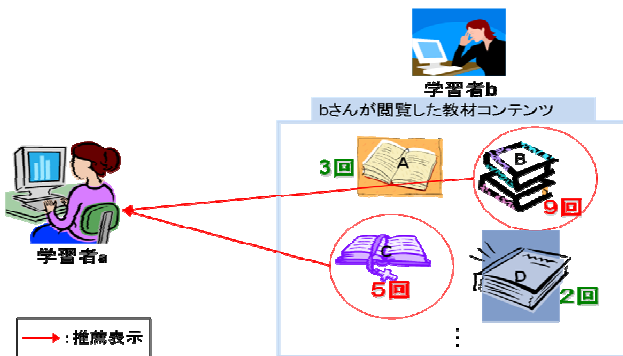


図7 推薦教材抽出の例

### 4.4 学習履歴を用いた推薦教材の抽出

2.2節の学習履歴を用いて、実際に類似度の計算及び予測閲覧回数の測定を行った。それぞれの計算には、ResnickらのGroupLensで使用されたCorrelation coefficientアルゴリズム[7]を用いて行う。

#### 4.4.1 学習履歴の分析

まず、閲覧回数による学習者の分布を以下のグラフ図8に示す。縦軸は学習者の数(人)、横軸は閲覧回数(回)を示している。ここに示した学習者は最低でも1回教材コンテンツを閲覧しており、その数は368人である。この368人に対して、推薦教材の抽出を行った。

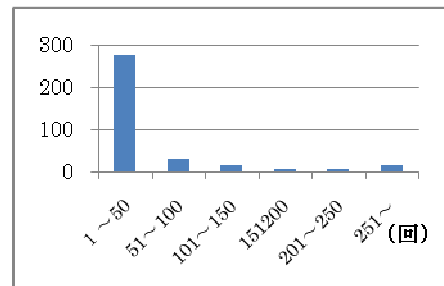


図8 閲覧回数による学習者の分布

#### 4.4.2 各学習者における推薦教材の比較

各学習者における推薦教材を比較する為、図8の368人の学習履歴から無作為に4人のデータを選択した。その結果、1~50回閲覧している学習者から2人、101~150回閲覧している学習者から1人、201~250回閲覧している学習者から1人が抽出された。表2は4人の推薦教材を表しており、推薦教材1は予測閲覧回数が最も大きかった教材コンテンツ、推薦教材2以下はその次に予測閲覧回数が大きかった教材コンテンツを示している。ここで推薦教材1、3、4に関しては共通する推薦教材が無い事がわかり、学習者に適した教材コンテンツを推薦している可能性が高いと考えられる。一方、推薦教材2に関してはid10とid427が同じ「プログラム内蔵方式」となり、id336の推薦教材3でも同じ教材コンテンツが推薦教材となった。この要因としては「プログラム内蔵方式」が「情報処理の基礎」の1章第1節に配置されており、多くの学習者が閲覧する為推薦され易くなったと考えられる。同様にid427の推薦教材3と4は「データ

表2 推薦教材の比較

user_id	10	200	336	427
推薦教材1	線形探索	ハッシュ法の衝突	変数	配列
推薦教材2	プログラム内蔵方式	チェイン法	バブルソートの複雑性	プログラム内蔵方式
推薦教材3	制御部とデータパス部	オープンアドレス	プログラム内蔵方式の特徴	データベース
推薦教材4	バイト	論理データモデル	挿入ソートの複雑性	3層スキーマ

ベースシステム」の1章第1節と第2節に配置されている事が要因だと考えられる。又、id427の推薦教材3,4以外の推薦教材は、学習者が一度も閲覧していないにもかかわらず、推薦教材として抽出されていた。この事から学習者にとって閲覧した事がない未知の教材コンテンツが推薦されたといえる。これは368人のデータからもいえる。なぜならば368人中約67%の246人に対する推薦教材は、全て一度も閲覧していない教材コンテンツが抽出されており、368人に対する全推薦教材1476の約88%の1302が一度も閲覧していない教材コンテンツである事がわかった。

#### 4.5 属性データによる類似度の計算

eラーニングを管理するLMSとして有名なMoodle, 本研究室で開発された双方向推薦システム, そしてAHSとの機能比較を行った。表3では、例えばMoodleは新規学習者でもオンライン小テストを受ける事でその結果を基に適応した学習コースを提供するシステムである事を示している。本研究では、新規学習者に対する類似度計算において属性データを用いる手法を提案する。

属性データは、新規学習者でも持つ情報であり、学習者の特性を表す情報として類似度の計算に有効だと考えられる。なぜならば、属性の共通点が多い学習者が高い類似度をもつと考えたからである。例えば、属性データとして個人の年齢、性別、趣味趣向、得意科目、苦手科目、平均学習時間、科目の嗜好、AIRSの利用時期等を取得する。図9に属性データによる類似度計算の例を示す。ここで新規学習者の苦手科目は情報、得意科目は英語、趣味は水泳だったとする。Dさんの苦手科目は情報、得意科目は英語、趣味は野球だったとする。Eさんの苦手科目は英語、得意科目は情報、趣味は天体観測だったとする。よって、新規学習者はDさんと高い類似度、Eさんとは低い類似度となる。

表3 既存システムとの比較

	Moodle	双方向	協調学習	AHS
推薦される教材	コース	関連教材	苦手科目 克服教材	コース
推薦に使うデータ	テスト	関連強さ	閲覧回数	テスト
新規ユーザ	△	○	○	△
適応性	△	×	○	○

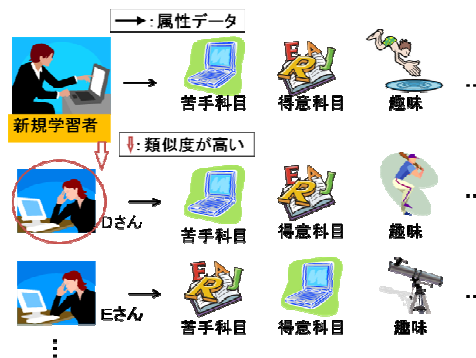


図9 属性データを用いた類似度計算の例

#### 5 まとめ

本研究では学習者の閲覧回数を基に苦手克服に貢献する教材を推薦するシステムを開発中であるが、その中で今回は実際にAIRSを利用した学習者の学習履歴を利用し、各学習者に対して推薦教材の抽出を行った。今後は推薦教材が苦手克服に貢献する教材であるかどうかについて分析を行い、属性データによる類似度計算機能の実装を行う予定である。

#### 参考文献

[1] eラーニング白書2007/2008年版  
 [2] gooリサーチ, “第3回ビジネスにおけるeラーニングの利用に関する調査結果”  
<https://research.goo.ne.jp/database/data/000055>  
 [3] Evangelos Triantafyllou, Andreas Pomportsis, Stavros Demetriadis and Elissavet Georgiadou, “The value of adaptivity based on cognitive style: an empirical study”, British Journal of educational technology, Vol.35, pp95-106(2004)  
 [4] 松澤俊典, 山口未来, 和田雄次, 土肥紳一: 教材コンテンツ双方向推薦システムの実装, 情報処理学会研究報告, 2008-CE-93(18), (2008/2/17)  
 [5] 高橋泰樹, 松澤俊典, 山口未来, 和田雄次, 土肥紳一: 履歴データを活用した教育支援システムAirs, PCカンファレンス2006  
 [6] 管坂玉美, 横尾真, 寺野隆雄, 山口高平: ”第3章. 情報推薦システム”, eビジネスの理論と応用, pp69-74(2003)  
 [7] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. and Riedl, J.: GroupLens: An Open Architecture for Collaborative Filtering of Netnews, Proc. ACM 1994 Conference on Computer Supported Cooperative Work (CSCW'94), pp.175-186 (1994).