

## 教材コンテンツ双方向推薦システムの実装

松澤 俊典<sup>†</sup> 山口 未来<sup>‡</sup> 和田 雄次<sup>††</sup> 土肥 紳一<sup>††</sup>

東京電機大学大学院 情報環境学研究所<sup>†</sup> 〒270-1382 千葉県印西市武西学園台 2-1200

東京電機大学大学院 工学研究科<sup>‡</sup> 〒101-8457 東京都千代田区神田錦町 2-2

東京電機大学 情報環境学部<sup>††</sup> 〒270-1382 千葉県印西市武西学園台 2-1200

E-mail: DEN03302@nifty.com<sup>†</sup>

### 概要

近年、e ラーニングが普及し、企業の研修や大学の講義などで取り入れられている。この背景には SCORM や WebCT などの LMS の発展が上げられ、システムの運用だけでなくコンテンツの導入も容易となったことにある。コンテンツが増大することは学習の幅を広める一方、学習者がコンテンツ選択の際、何を学べば良いのか迷うといった懸念がある。本研究では、学習者が遷移したコンテンツの順番、つまり学習者がどのように学習を行ったかという学習履歴を基に、コンテンツの関連性を求め、推薦する双方向推薦システムの開発を行った。本稿では、著者らが開発した復習支援システム AIRS(An Individual Reviewing System)の概要及び、双方向推薦システムの詳細と、その評価と有効性について述べる。

キーワード e ラーニング, 情報推薦技術, 双方向推薦, AIRS

## Implementation of recommend system for learning texts

Syunsuke Matsuzawa<sup>†</sup> Mikuru Yamaguchi<sup>‡</sup> Yuji Wada<sup>††</sup> Shinichi Dohi<sup>††</sup>

Graduate School of Information Environment, Tokyo Denki University<sup>†</sup>

Graduate School of Engineering, Tokyo Denki University<sup>‡</sup>

School of Information Environment, Tokyo Denki University<sup>††</sup>

### Abstract

Recently, e-learning systems have begun to be used in companies and education institutions. The spread of e-learning is mainly owing to development of LMS for instance SCORM and WebCT. Thereby, the introduction of learning texts have become much simpler. The increase of the contents has realized the wider width of learning. As the result, we are afraid that learners are confused with selecting best learning texts. In our research, we are developing bidirectional recommendation system that extracts the relationship among the learning texts with historical logs and recommends the effective text for the learners. In this paper, we firstly discuss the design of the bidirectional recommendation system, secondly show its evaluation results and finally we conclude our recommendation system is useful for the learners.

**Keywords** e-learning, recommendation technology, bidirectional recommend, AIRS

### 1. はじめに

大学の講義や企業研修などにおいて、e-learning を用いた教育方法が普及し、これらの学習のサポートに限らず、単独で学習を完結するシステムも散見するようになった。しかし、現在運用されているシステムの多くは、単純に教材を閲覧するだけのものや、オンラインテストを実施するものなど、e-learning ならではの利点が見出せていない。また、学習システム・学習履歴データベース・教材コンテンツを統一

規格とした SCORM[1]の普及によって、コンテンツが増大することで、学習者がコンテンツを選択の際に迷う恐れがある。

これらを踏まえ、本研究では学習者が遷移したコンテンツの履歴情報、つまり学習者がどのように学習を行ったかという学習履歴を基に、コンテンツの関連性を求めることで、学習に最適な双方向推薦システムの開発を行った。また、復習支援システム AIRS の開発を行った[2]。本稿では AIRS の概要および、双方向推薦システムの実装と評価について述べる。

## 2. AIRS の概要

AIRS は、復習に特化した学習教材の配信を行い、学習者に効率良い復習を提供するための Web ベースシステムである。ここでいう復習とは、「一度講義などで学習した内容の再確認や、理解の強化を行うこと」と定義している[3]。

AIRS は、Java Servlet と JSP を使用し、Servlet コンテナとして Tomcat5.5、Web サーバに Apache2 を使用している。学習者の学習履歴および、推薦用データベースには MySQL を使用している。図 1 に AIRS の概要図を示す。

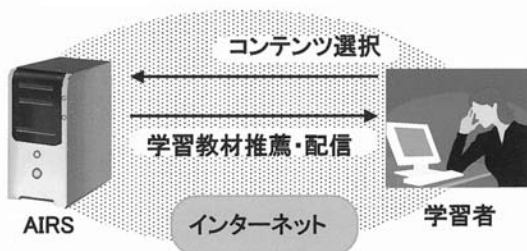


図 1. AIRS 概要図

### 2.1 教材コンテンツ

AIRS の教材コンテンツは本学で開講されている「データベースシステム」と「情報処理の基礎」の講義内容を基に、要点を押さえ、授業の流れに即した教材コンテンツの配信を行っている。各コンテンツは章と節で階層的にまとめ、更に節ごとに3つの異なる表現方法を持ったコンテンツを用意している。異なる表現方法とは、表現方法1が公式や定義など必要最低限の説明や図のみ、表現方法2は文と図などの静的な説明、表現方法3はFLASHなどを用いた動的な説明、というように同一内容でも教材の構成が異なることを言う。例えば、データベースシステムの3章8節の表現方法1では、正規形の定義のみが書かれているが、同表現方法3では動的に正規形の解説を試みている。表現方法の定義を表1に、教材コンテンツの階層構造の様子を図2に示す。

表 1. 表現方法の定義

表現方法 1	公式と定義，最低限の図のみ
表現方法 2	静的説明
表現方法 3	動的説明

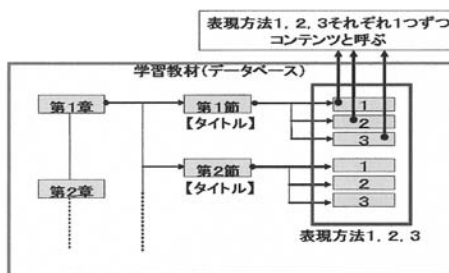


図 2. 教材コンテンツの階層構造

### 2.2 教材コンテンツの推薦

AIRS の特徴に教材コンテンツの推薦機能がある。この推薦機能について述べる。

#### 2.2.1 表現方法の推薦機能

利用している推薦技術に、協調フィルタリング[4]がある。協調フィルタリングは、履歴データからユーザーの特徴を分析し、似た特徴を持つ他のユーザーの履歴から情報を推薦する技術である。AIRS では「各表現方法の自動推薦機能」を目的に、この技術を利用している[5]。

いま学習者 A がどの節を、どの表現方法で閲覧したかという閲覧履歴情報がデータベースに保存されている。このとき、似た閲覧履歴情報をもつ学習者 B を発見し、その情報を参考に学習者 A に推薦する表現方法を決定する。協調フィルタリングによる推薦の概要を図 3 に示す。

学習者 A の閲覧履歴						
節の番号	1	2	3	4	5	6
表現方法の閲覧履歴	0	1	0	1	3	2

学習者 B の閲覧履歴						
節の番号	1	2	3	4	5	6
表現方法の閲覧履歴	0	1	2	1	3	2

↑  
学習者 A がコンテンツ 4 を閲覧するとき、表現方法 1 を推薦

図 3. 協調フィルタリングによる推薦の概要

表現方法の閲覧履歴 1~3 は各表現方法を閲覧したことを表し、0 はそのコンテンツを閲覧していないことを表す。このとき、学習者 A がコンテンツ 3 を閲覧した際に、似た特徴を持つ学習者 B のコンテンツ 3 の履歴が 2 であるため、学習者 A も 2 の表現方法を好む可能性が高いと考え、表現方法 2 を推薦する。

### 2.2.2 節の推薦機能

本研究で提案する、「節同士の関連を算出し、関連の高い節の推薦表示機能」である。これを「双方向推薦システム」と名づけ、アルゴリズムの提案および、開発を行った。

## 3. 双方向推薦システム

AIRS は、効率の良い復習を提供するシステムであるため、単一で学習を完了するものではない。分からない箇所や確認したい箇所をピンポイントで選択し、教材を閲覧することで、短時間で要点を抑え、効率の良い復習を提供するコンセプトの基に開発を行っている。しかし、学習者は授業の流れに沿って最初から順番に閲覧をしてしまう傾向が強い。節の数が 210 個と教材の選択肢が多いことも、学習者が順を追ってしまいう要因と考えられる。そこで、現在閲覧している節と関連の強い節を推薦することができれば、復習している箇所と関連の高い箇所を提示し、効率よく復習ができると考えた。

### 3.1 推薦の種類

参考文献[6]では、推薦方式の代表的なものとして以下の二つが上げられている。

- ① ユーザの関連による推薦
- ② アイテムの関連による推薦

ユーザの関連による推薦とは、各ユーザの行動履歴を分析し、ユーザの関連性から推薦する技術である。これをユーザ相関と言い、前述の協調フィルタリングはユーザ相関を用いた推薦手法となる。

アイテムの関連による推薦とは、ユーザ相関同様、アイテムの関連性から推薦する技術である。これをアイテム相関と言い、分析の手法として以下の二つがあげられる。

- ① ユーザの行動履歴を基にした相関
- ② アイテム自身の情報を基にした相関

双方向推薦では、前者のユーザの行動履歴を基にした相関を用いて、節の関連を分析する。

### 3.2 双方向の必要性

現在、協調フィルタリングなど実用化されているレコメンドエンジンは、商品などの推薦が主な目的のため、購入したものを推薦する必要がない。しかし、学習においては現在閲覧している教材から発展した内容に進むだけでなく、

前の内容に戻ることは日常的に行う。そこで、新たな内容の推薦のみ行う単一方向の推薦ではなく、以前閲覧した内容も推薦を行う双方向の推薦が必要であると考えられる。

### 3.3 双方向推薦の概要

いま、学習者が「情報処理の基礎」の教材コンテンツ「代入の基本」を閲覧している。このとき、次のステップである「while 文の基本」や「if 文の基本」に遷移することは自然である。しかし、基本的な内容である「変数の種類」を再び閲覧することも、学習する上で不自然な行動ではない。つまり「代入の基本」から多くの遷移があったコンテンツのみを推薦するのではなく、「代入の基本」に多く遷移したコンテンツを同時に推薦することで、学習効率の良い推薦が期待できる。双方向推薦の概要を図 4 に示す。

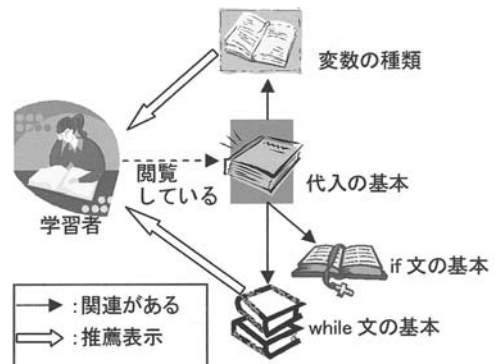


図 4. 双方向推薦の概要図

### 3.4 双方向推薦アルゴリズム

双方向推薦の実現方法として、各節を root としたバイナリツリーを構成する。この子ノードはその節と関連の高い節が繋がり、高さ 2 までの規模の小さなツリーとなる。このツリーは節を root とし、すべての節の数と同じ数のツリーが構成される。また、このツリーは前述の教材コンテンツの階層構造とは別のものであり、節の関連性を示したものである。このツリーを参照することで、前後に関連の高い節を発見し、推薦を行う。

#### 3.4.1 ツリーの構成方法

この双方向推薦に用いられるツリーを構成するために、各学習者がどのようにコンテンツを遷移したのかを記録した履歴データを用い

る。これを学習者全体で総合したものを、コンテンツ遷移履歴と言う。この様子を図5に示す。

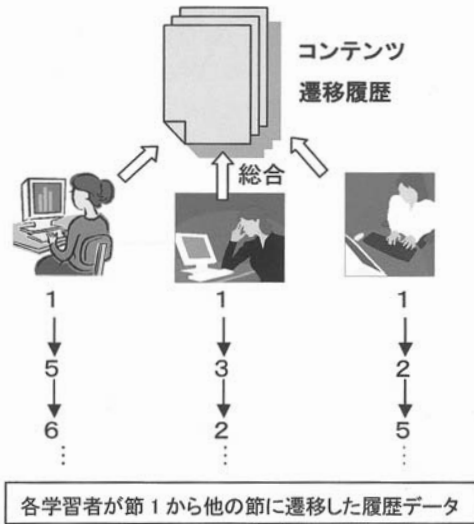


図5. コンテンツ遷移履歴の作成

このコンテンツ遷移履歴から、節1と関連の高い節を算出するために、Terveen[7]らの web 推薦システム PHOAKS で使用された、User count アルゴリズムを用いる。これは、多くのユーザに使用されているアイテムを算出するアルゴリズムである。多くの学習者がその節から遷移している節は、関連が強いと考えられる。コンテンツ遷移履歴から教材コンテンツの関連の度合いを求める式を以下に示す。

$$p_{a,j} = \frac{\sum_{u \in U} \text{binaryOf}(q_{i,j})}{|U|} \quad (1)$$

ただし、

$$\text{binaryOf}(q_{i,j}) = \begin{cases} 0 & (q_{i,j} = 0.0) \\ 1 & (q_{i,j} > 0.0) \end{cases} \quad (2)$$

このとき、教材コンテンツ  $U_i$  から  $U_j$  に遷移した頻度を  $q_{ij}$  と表すと、 $U_a$  に対する  $U_j$  の関連度  $p_{a,j}$  が求められる。式(1)の右辺は教材コンテンツ  $U_a$  に遷移した学習者の割合に等しくなる。節1から遷移した割合の多い節上位2つをツリーの階層1に、それらから遷移した割合の高い節上位2つずつを階層2として、バイナリツリーを構成する。これを全ての節に行うことで、

関連を示すツリーが節の数だけ構成される。節「関数」の関連を示したツリーを図6に示す。

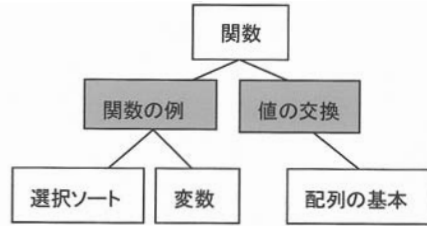


図6. 関連を示すツリーの例

### 3.4.2 ツリーを用いた推薦方法

前節で述べたツリーを用いて双方向推薦を行う。学習者が節「関数」を閲覧しているとき、システムは図6の関連を示すツリーを参照する。このときシステムは、このツリーから節「関数の例」と「値の交換」を推薦項目に決定する。次に、全てのツリーから、節「関数」が含まれるツリーを検索する。ここでは図7の例に示す節「if文」が見つかったとする。このツリーから、節「関数」は節「while文の基本」と関連が高いことが分かるので、これを推薦項目に決定する。この推薦の様子および、節「if文」のツリーの例を図7に示す。

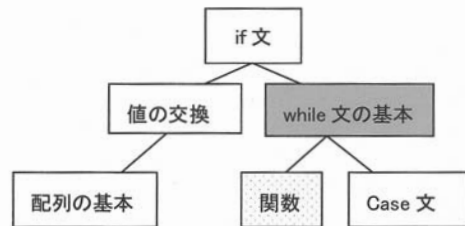


図7. 節「関数」を含むツリーから推薦の様子

更に節「関数」を含むツリーを発見したとき、同様の処理を行い、推薦項目に決定する。システムはこれらの処理によって決定した推薦項目を、学習者に提示する。これによって現在学習している節から、前後に遷移される割合の高い節の推薦が可能となる。つまり、次に発展する内容と、復習すべき内容を双方向で推薦することが可能になる。この前後に関連のあるアイテムを推薦する推薦方式を「双方向推薦システム」と名づけた。図8に双方向推薦のフローチャートを示す。

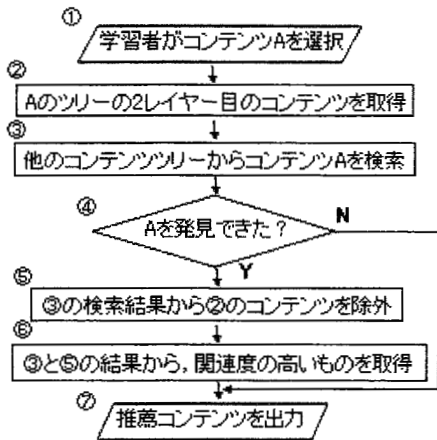


図 8. 双方向推薦システムのフローチャート

以上のフローにより、節「関数」を閲覧している学習者に対して、関連ページとして提示する節を決定する。この提示する数は最大4つまでとしている。図9に関連ページを提示する様子を示す。

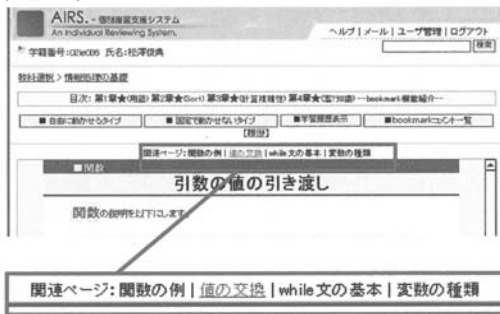


図 9. 関連ページを提示する様子

#### 4. 評価

本章では、授業の流れに沿って並べられたコンテンツを利用したログデータから得られる推薦結果と、ランダムに並べられたコンテンツを利用したログデータから得られる推薦結果を比較する。これによって、双方向推薦システムで提示される関連ページの妥当性について検証した。この結果を、以下に述べる。

##### 4.1 検証用の履歴データ

本学部で2007年9月から開講された2年次生向けの講義である、「データベースシステム」の受講者が、AIRSを利用した。受講者数は125名であった。2007年9月19日からコンテンツ遷移履歴の取得を開始し、授業終了後の2007

年12月27日までの履歴データを検証に利用した。履歴データの総数は7355件である。また、実際に運用されているシステムであるので、コンテンツのリンクは講義の流れに沿って並んでいる。この履歴データから構成された関連を示すツリー群を実験データ1とする。

比較のためのコンテンツ遷移履歴を得るため、大学院生1名、研究生1名、学部生22名(4年次生13名、3年次生9名)、合計24名がAIRSを利用した。利用期間は2008年1月8日から2008年1月18日まで、1日10分程度、AIRSを利用して学習を行った。ここでは比較のため、コンテンツのリンクをランダムに配置し、授業の流れとは全く関係なく並んでいる。履歴データの総数は768件となった。この履歴データから構成されたツリー群を実験データ2と呼ぶ。表2に実験データのまとめを示す。

表 2. 実験データのまとめ

	実験データ1	実験データ2
利用期間	2007年9月~12月	2008年1月
利用者	受講者125名	被験者24名
履歴総数	7355件	768件
コンテンツの並び	講義に沿って整列	ランダム

表3はツリーをデータベースに格納した一部を示したものである。content\_idが節を表し、layer1\_1とlayer1\_2がツリーの1層、layer2\_1からlayer2\_4が2層目を表している。

表 3. データベースに格納されたツリー

content_id	layer1_1	layer1_2	layer2_1	...
14	20	23	11	...
15	173	10	4	...
16	19	12	1	...

##### 4.2 推薦の適合率

前節で述べた二つのツリー群を用いて、それぞれ双方向推薦を行い、推薦結果についての比較を行った。ここで対象にしたコンテンツは「データベースシステム」の第1章と第3章の58個の節である。実験データ1を用いた推薦では、152個の関連ページが提示され、実験データ2を用いた推薦では、212個が提示された。実験データ2のほうが多く提示されたのは、コンテンツがランダムに配置されているため、多

様な遷移履歴を得られたためと考えられる。

実験データ1と実験データ2の推薦結果を比較し、実験データ1と推薦項目が同じものがヒットした個数は32個、ヒットした割合は15.1%である。1つでもヒットしたものを含んだ節の個数は25個、平均で43.1%であった。これを完全ヒットという。続いて、全く同じではないが、ヒットと考えてよい推薦を加えた結果を比較した。ヒットと考えてよい推薦とは、「B+木へのkeyの挿入」と「B+木のkeyの削除」や、「和演算」と「差演算」などの、推薦結果として同一であると考えてよいものである。この条件でヒットした個数は74個、割合は34.9%である。1つでもヒットしたものを含んだ個数は45個となり、平均すると77.6%になった。これを近似ヒットという。ヒットした個数と、ヒットと考えてよい個数の差の平均から求めた標準偏差は0.80であった。

表4. ヒットした個数と割合

	完全ヒット	近似ヒット
個数(コンテンツ)	32	74
割合(コンテンツ)	15.1%	34.9%
個数(節)	25	45
割合(節)	43.1%	77.6%

・実験データ2の推薦結果総数:212個

・節の総数 :58個

### 4.3 考察

今回は、実験データの件数や期間が異なっているため、正確な比較が行えたとは言いがたい。しかし、この比較でほぼ同一と考えてよい推薦を行えている節の割合が、77.6%と高いことを確認できた。

協調フィルタリングなどの個人の履歴データを用いるレコメンドエンジンでは、初めてそのシステムを使用するユーザに最適な推薦を行うことができない。しかし、ユーザ全体の履歴データを使ってアイテムの関連を算出し、推薦をするレコメンドエンジンは、初めて使用するユーザにも同一の結果を提示することが可能である。この方法を取る双方向推薦システムは、コンテンツの遷移履歴さえあれば、全ての学習者に対して適切に教材コンテンツを推薦することが可能といえる。

## 5. まとめ

今回の実験によって、双方向推薦システムの推薦結果はどのようにコンテンツを配置しても、適切に推薦できることを示すことができた。また、個人の履歴データを必要としない双方向推薦システムが教材コンテンツの推薦に有効であることも示せた。今後は、双方向推薦システムを使った節の推薦と、協調フィルタリングを使った表現方法の推薦を合わせることにより、学習者が学習したいコンテンツを適切に推薦できているかを検証することが今後の課題である。

### 謝辞

本研究の実験にご協力いただいた被験者、AIRSをご利用いただいた多くの学生の皆様に感謝いたします。

本研究は、科学研究費補助金(基盤研究(C)課題番号18599731「E ラーニングにおける学習教材レコメンドーション技術に関する研究」)の支援による。

### 参考文献

- [1]日本eラーニングコンソシアム:eラーニング情報ポータルサイト <http://www.elc.or.jp>
- [2]山口未来, 高橋泰樹, 松澤俊典, 高橋時市郎, 土肥紳一: 複数のコンテンツを利用可能なeラーニングシステムの開発と評価 - 学習コンテンツのAIRSへの実装 -, 電子情報通信学会技術研究報告, ET2006-118 (2007,3)
- [3]土肥紳一, 高橋泰樹, 松澤俊典, 山口未来, 田窪昭夫, 和田雄次: 学習者の履歴データを活用した個別復習システム, 平成18年度情報教育研究集会 講演論文集(2006,11)
- [4]大杉直樹, 門田暁人: 協調フィルタリングに基づくソフトウェア機能推薦システム, 情報処理学会論文誌, Vol.45, No.1(2004)
- [5]高橋泰樹, 松澤俊典, 山口未来, 和田雄次, 土肥紳一: 学習者に適した学習教材の推薦と配信, 情報処理学会研究報告, Vol.2007, No.12 (2007,2)
- [6]山崎徳之: レコメンドエンジン開発室 第1回, Software Design, 通巻268号(2007,8)
- [7]Terveen, L., Hill, W., McDonald, D. and Creter, J.: PHOAKS: A System for Sharing Recommendations, *Comm. ACM*, Vol.40, No3 (1997)