

協調フィルタリングを用いた初等教育の授業向け Web 検索手法の提案

松原 有希[†] 長田 智和[†] 玉城 史朗[†]

Web 上の適切なサイトから目的の情報を素早く的確に探し出す能力は、高度情報化社会を担う子どもたちに身に付けさせたい能力のひとつである。しかし、現在 Web 上の情報を検索する最も一般的な手法である検索エンジンにおいては、ユーザが検索を行うための検索キーワードを自分自身で入力することが前提とされており、情報を得るためにはある程度の知識や経験が必要とされ、多くの初心者にとって使いこなすのは容易ではない。一方、協調フィルタリングとは、大量の情報の中からユーザに対して有用であると推定される未知のアイテムを推薦する、近年注目されている推薦手法のひとつである。そこで本研究では、既存の Web 検索エンジンを初等教育で活用する場合の問題点を明らかにし、協調フィルタリングを用いて、初等教育において積極的に活用される Web 検索システムを提案する。

Proposal of the Web Search Technique for Lessons of the Elementary School Education using Collaborative Filtering

YUKI MATSUBARA,[†] TOMOKAZU NAGATA[†] and SHIRO TAMAKI[†]

To search the information exactly and quickly from the Internet is important capability, we want children to learn it. However, the search engine is general method to search informations in the Internet, it is based on the premise that users input keywords oneself, and the user needs amount of knowledge and experience to get informations. Consequently, to search information is not easy for beginners. On the other hand, collaborative filtering attracts attention in recent years. It is the method of making automatic predictions about the interests of a user by collecting information from many users. In this research, we clear up the questionable points to use the Internet search engine by elementary school education, and propose a Web filtering system which make active use in elementary school education by using the collaborative filtering.

1. はじめに

インターネットは現代社会の中に急速に浸透しており、そのサービスの中でも特に WWW(World Wide Web)はその情報量を増大させている。Web 上の適切なサイトから目的の情報を素早く的確に探し出す能力は、高度情報化社会を生きていく上で重要な能力のひとつである。しかし、現在最も一般的な Web 上の情報検索手法である検索エンジンにおいては、ユーザが検索を行うための検索キーワードや検索式などのクエリを自分自身で入力することが前提とされており、情報を得るためにはある程度の知識や経験が必要とされ、多くの初心者にとって使いこなすのは容易ではない。

新しい学習指導要領で述べられている「生きる力」を養う上で、初等中等教育における情報教育では、「情

報活用能力」の育成が目標とされている。とくに初等教育段階における情報教育では、各教科間の関連を図った取り組みが行われやすい。この特色を生かし、各教科などの具体的、体験的活動の中で「情報活用の実践力」の育成を図ることを基本とし、子どもたちが情報手段に慣れ親しみ、適切に活動する学習活動を充実することとしている。情報活用の実践力とは「課題や目的に応じて情報手段を適切に活用することを含めて、必要な情報を主体的に収集・判断・処理・創造し、受け手の状況などを踏まえて発信・伝達できる能力」と定義されている¹⁾。これを達成するために初等教育の授業では、各教科で調べ学習教材のひとつとして Web 検索が活用されている。しかし、情報収集能力の低い児童らでは、大量の情報の中から自分が必要とする情報を得ることができない場合がある。

このような大量の情報の中から有用な情報を取得する有効な手法のひとつとして、推薦システムが近年注目されている²⁾。推薦システムとは、ユーザの過去の嗜好、購入履歴、デモグラフィック情報（年齢や性別、地域などの人口統計的な情報）に基づいてアイテムを

[†] 琉球大学大学院理工学研究科情報工学専攻
Engineering and Science,
Information Engineering Course,
Graduate School of the Ryukyus

表 1 アンケート回収率

小学校	回答者	対象者	回答率
A	18	26	69.2%
B	18	28	64.3%
C	20	34	58.8%
計	56	88	63.6%

推薦するシステムである。推薦システムには、ユーザが情報に対して評価を行い、その結果を利用してユーザの評価と類似した評価を行った他のユーザを検索し、検索された類似ユーザの評価をもとに必要な情報をフィルタリングする協調フィルタリングという手法がある。この手法は、アイテムに対する多くのユーザの評価情報を収集する必要があるが、近年においては、アイテムへのアクセス履歴から自動的に習得し、その履歴をもとに推薦を行う手法が提案されている。

そこで、本研究では、現場の教員にアンケート調査等を行い、既存の Web 検索エンジンを授業で利用する場合の問題点を明らかにし、既存の検索エンジンに協調フィルタリング手法を導入することで、児童が短時間で目的の情報を取得することができる授業支援ツールとしての Web 検索システムを提案する。

本稿の構成は以下の通りである。第 2 章では、教員を対象とした調査と既存の Web 検索エンジンにおける問題点を示す。第 3 章では、協調フィルタリングの概要を述べ、第 4 章では、提案システムの構成と推薦手順を述べる。第 5 章では、提案システムで使用する協調フィルタリングのアルゴリズムを示す。第 6 章では、評価実験について述べ、その結果を示し、考察する。第 7 章においてまとめる。

2. 調 査

初等教育における情報検索教育について現状を把握するために、アンケート調査を行った。調査は、県内 3 校の教職員を対象に、アンケート用紙 (A4 両面 1 枚) を配布した。このアンケート結果をもとに各学校別に、IT 担当の教員など情報教育に関心を持っている教員と意見交換を行った。

2.1 アンケート調査結果

アンケートの依頼は県内の小学校 3 校の教職員 88 名に対して行い、回収された 56 件 (回収率 63.6%) について分析を行った。表 1 は回答者の人数の内訳を示す。

図 1 は、授業で Web 検索を利用した調べ学習を実施したことがあるかという問いに対する回答である。各学校とも約 80% の教員が、授業で Web 検索を利用したことがあると答えたことから、Web 検索が初等教育において情報収集手段のひとつであることが明らかとなった。図 2 は、インターネット検索を利用したことがある科目を示し、図 3 はその内容を示している。表 2 に、図 3 のデータを学校別に集計した結果を

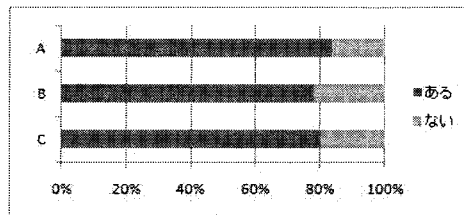


図 1 Web 検索を授業で利用した経験

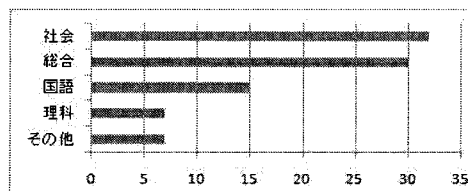


図 2 Web 検索を利用したことがある科目

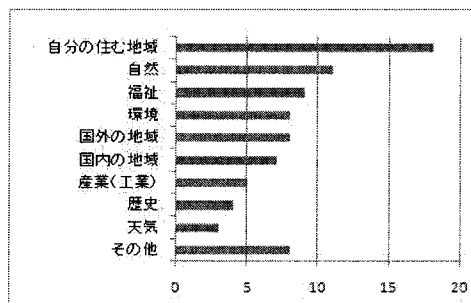


図 3 Web 検索の内容

示す。どの学校でも、各教科の授業で Web 検索を利用する場合、書籍には掲載されていない最新の情報や、自分の住んでいる地域の歴史や特産物といったローカルな情報、近年の社会問題である環境・福祉についての情報を検索されることが多い。これより、初等教育における Web 検索を利用した授業では、異なる学校・クラスにおいても、ある程度テーマが重複することが確認された。

授業でインターネット検索を利用することに対しての意見を自由記述式で求めたところ、「必要な情報まで辿りつけない」「情報の取捨選択ができない」といった意見が複数見受けられた。

2.2 意見交換

アンケート結果をもとに、各学校別に情報教育に関心を持っている教員 (計 7 名) と意見交換を行った。この意見交換から、既存の Web 検索エンジンを初等教育で利用する場合には、以下の問題があることが把握できた。

- 適切な検索キーワードを選択することができない

表 2 学校別 Web 検索内容

小学校	自分の住む 地域	Web 検索内容					産業 (工業)	歴史	天気	その他
		自然	福祉	環境	国外の 地域	国内の 地域				
A	4	4	3	3	2	5	3	1	1	5
B	5	2	3	2	3	2	1	2	1	3
C	9	5	3	3	3	0	0	1	1	0

- 情報が多すぎて、インターネット上から必要な情報を取捨選択できない
- インターネット上の情報は一般向けが多いため、児童は内容を理解しきれない

Web 検索において自分が必要とする情報を検索するためには、検索キーワードが必須となる。語彙や経験が少ない児童にとって、適切な検索キーワードを設定することは容易ではない。また、適切な検索キーワードを設定しても、膨大な情報の中から自分が理解できる内容が掲載されている Web ページを見つけ出すことが困難である。

そこで、これらの問題を解決するために、本研究では、協調フィルタリングを用いて検索キーワードを推薦する Web 検索システムを提案する。

3. 協調フィルタリング

協調フィルタリングは、大量の情報の中から有用な情報を選び出す情報フィルタリングの一手法である。ここで、情報とは、記事、ウェブページ、書籍、楽曲、映画作品など、その有用度（関心度）がユーザによって異なるもの（アイテム）である。また、「有用な情報を選び出す」とは、ユーザによって未知のアイテムの有用度を推定し、有用度の高いアイテムをそのユーザに提示（推薦）することである³⁾。ユーザの嗜好を過去の行動という形で記録し、そのユーザと似たような行動を取っているユーザの嗜好情報をもとに、ユーザの嗜好を推測する、リコメンデーションサービスを提供する際に使用される代表的な手法である。

協調フィルタリングの特徴は、未知アイテムの有用度推定において、他ユーザの意見（既知アイテムに対する有用度評価値）を利用する点にある⁴⁾。すなわち、協調フィルタリングへの入力は、各ユーザの各アイテムに対する評価値を要素とする行列（評価値行列）であり、その出力は、あるユーザにとって未知のアイテムの評価値（評価値行列の特定要素）の推定値である（図 4）。

協調フィルタリングの手法が初めて導入された Netnews 記事推薦システム Tapestry⁵⁾では、記事の推薦を受けようとするユーザは、記事に対する関心が自分と似ている他ユーザ（類似ユーザ）を指定する必要があった。推薦アルゴリズムは単純で、ユーザの未知記事のうち、類似ユーザが高い vote 値を与えている記事が推薦される。Web ページ推薦システム PHOAKS⁶⁾では、類似ユーザを指定する必要はないが、推薦アル

入力(評価値行列)

	i_1	i_2	...	i_j	...	i_n
u_1	$v_{1,1}$	$v_{1,2}$...	$v_{1,j}$...	$v_{1,n}$
u_2	$v_{2,1}$	$v_{2,2}$...	$v_{2,j}$...	$v_{2,n}$
...
u_a	$v_{a,1}$	$v_{a,2}$...	$v_{a,j}$...	$v_{a,n}$
...
u_m	$v_{m,1}$	$v_{m,2}$...	$v_{m,j}$...	$v_{m,n}$



$$u_a \quad \boxed{P_{a,j}}$$

出力(未知アイテムの推定評価値)

図 4 協調フィルタリング

ゴリズムは比較的単純で、ユーザの未知ページのうち、評価値を与えているユーザが多いページが推薦される。Usenet 記事推薦システム GroupLens⁷⁾では、まず、vote 値行列に基づいて、ユーザと他ユーザとの類似度（記事に対する関心の類似度）の推定が行われる。そして、推定した類似度で重みづけられた評価値行列を用いて推薦記事が決定される。類似度推定の精度が高くなれば、システムユーザが類似ユーザを指定した場合と同様の推薦が可能となる。

本提案では、検索キーワードに対する明示的评价値は用いず、かわりに、ユーザによる検索履歴を用いて、検索キーワードの推薦を行う。

4. 提案システムの構成と推薦手順

4.1 構成

提案システムは、図 5 に示すように、ユーザのブラウザと Web 検索エンジンの間にあるフィルタサーバと考える。提案システムはユーザの計算機とネットワークで接続された環境に 1 台設置され、ユーザはブラウザから提案システムへアクセスする。提案システムは各ユーザが使用した検索キーワードを蓄積する「検索履歴データベース」を具備する。Web 検索エンジンには、Yahoo!JAPAN が無料で提供している Web サービス Yahoo!検索 API⁸⁾を利用する。

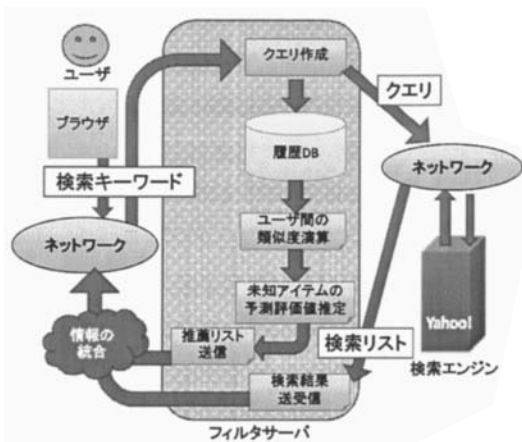


図 5 提案システム



図 6 システム実行画面

4.2 実装環境

システムは FedoraCore6 上にて実装する。Web サーバに Apache2.2.x を用いる。ユーザからの検索キーワードの送受信, Web 検索エンジンへのクエリ送受信およびデータベースとの通信部分に PHP5.1.6 を用い, 検索履歴を蓄積するデータベースには MySQL5.0.27 を用いた。実行画面を図 6 に示す。

4.3 システムの動作

提案手法の手順は次のとおりである。

- (1) 一般の Web 検索エンジンと同様に, ユーザは検索キーワードを入力する。
- (2) システムはユーザから受け取った検索キーワードを履歴データベースに蓄積する。
- (3) (2) と同時に, Web 検索エンジンに検索キーワードを送信し, その結果リストを受信する。
- (4) 履歴データベースのデータをもとに, 全ユーザ

との類似度を算出する。

- (5) 未使用の検索キーワードの評価値を類似度を重みとして予測し, 予測評価値による順位付けを行う。
- (6) Web 検索エンジンからの結果リストとともに, 予測評価値の高い未使用の検索キーワードをユーザに送信する。

5. 提案システムにおける推薦アルゴリズム

協調フィルタリングの手法は, 2 つのステップで行われる。1 つ目のステップは, アイテムを推薦する対象ユーザと他ユーザの類似度を求め, それをもとに類似ユーザを見つけ出すことである。2 つ目のステップは, 類似ユーザの評価に基づいて, 推薦対象ユーザが評価していないアイテムの推定評価値を求め, ユーザに推薦するアイテムを決定することである。

5.1 ユーザ間の類似度推定

GoupLens で使用された, ピアソンの積率相関係数を用いたアルゴリズムを, 提案システムに応用する。相関係数は, 対象を一定の基準で数値化する 2 種類の尺度があり, 複数のアイテム (要素) について評価した場合に, アイテム間の基準の相関を求める手法である。システムの全ユーザを表す集合を $U = \{u_1, u_2, u_3, \dots, u_m\}$ とし, ユーザ u_a が使用した全検索キーワードを表す集合を $W_a = \{w_1, w_2, w_3, \dots, w_K\}$ とする。

u_a が w_j を使用した場合, 評価値 $v_{a,j} = 1$ とし, w_j を使用しない場合は評価値 $v_{a,j} = 0$ とする。これにより, 評価値ベクトル v_a は $\{0, 1\}$ の要素からなる, u_i のすべての評価の平均値を \bar{v}_i と表す。推薦対象ユーザ u_a とその他の各ユーザ u_i について, ユーザ間の類似度 $sim(u_a, u_i)$ は u_a と u_i の共分散を, u_a の標準偏差と u_i の標準偏差の積で除算した値であり, 次式で求める。

$$sim(u_a, u_i) = \frac{\sum_{w_k \in W_a \cap W_i} (v_{a,k} - \bar{v}_a)(v_{i,k} - \bar{v}_i)}{\sqrt{\sum_{w_k \in W_a \cap W_i} (v_{a,k} - \bar{v}_a)^2} \sqrt{\sum_{w_k \in W_a \cap W_i} (v_{i,k} - \bar{v}_i)^2}} \quad (1)$$

$sim(u_a, u_i)$ は $[-1, 1]$ の範囲の値をとり, 1 に近づくほど両者の類似度が高く, -1 に近づくほど両者の類似度は低くなる。

5.2 評価値推定

u_a の類似ユーザ u_i が使用しており, かつ u_a が未使用である検索キーワード w_j すべてに対して, u_i の評価値に基づいて算出し, w_j に対する u_a の推定評価値を $P_{a,j}$ とする。

$$P_{a,j} = \bar{v}_a + K \sum_{u_i \in U} sim(a, i)(v_{i,j} - \bar{v}_i) \quad (2)$$

ここで K は, $sim(a, i)$ の総和が 1 になるよう設定さ

れた正規化のための係数である。

最終的に、推定評価値 $P_{a,j}$ に基づいて上位 N 個の推薦リスト $L_a : \{q_1, q_2, q_3, \dots, q_N\}$ が求められる。 L_a は最も高い推定値をもつアイテムを 1 位とした降順の推薦ランキングを示す。

6. 評価実験

提案システムの精度を実験的に評価するため、本実験においては、あるテーマに従って被験者が Web 検索を行った。その検索履歴を用いて第 4 章、第 5 章で提案した協調フィルタリングの性能を検証する。

6.1 実験データ

実験において設定した検索テーマは、「作った人の名前になんで名付けられた楽器で、大きさによって種類に分けられ、その中でも日本に数台しかない珍しい楽器の写真」とした。被験者は学生 15 人、総検索クエリ実行回数は 109 回、使用されたキーワードは 59 件であった。

6.2 精度評価基準

情報検索における検索性能を定量化する指標として再現率 (*Recall*) と適合率 (*Precision*) が用いられており、協調フィルタリングによる推薦リストの正確さの指標としても用いられている²⁾。一般に、再現率と適合率はトレードオフ関係にあり、一方が高い値をとると、他方の値は低くなる。再現率はユーザが選択したアイテムの中に、システムが選択したアイテムが含まれる割合を示す。一方適合率は、システムが提案する推薦アイテムの中に、ユーザが選択したアイテムが含まれる割合を示す。全被験者を表す集合を $R = \{r_1, r_2, r_3, \dots, r_m\}$ 、被験者 r_i が使用したすべての検索キーワードの集合を $W_i = \{w_1, w_2, w_3, \dots, w_K\}$ とし、ランダムに選ばれた $K - n$ 個の要素がトレーニングセットを形成する。残りの n 要素は予測のためには使われず、テストセット T_i とする。 T_i は、 r_i 以外の被験者にも評価されているアイテムとする。システムが推薦するアイテムの集合を L としたとき、再現率は式 (3)、適合率は式 (4) と定義される。

$$Recall = \frac{|T_i \cap L|}{|L|} \quad (3)$$

$$Precision = \frac{|T_i \cap L|}{|L|} \quad (4)$$

再現率、適合率ともに $[0, 1]$ の範囲の値をとり、1 に近づくほど推薦の精度が高いことを示す。

6.3 実験結果

被験者 14 人を履歴データをもとに、被験者 r_i のデータから $n = 3$ とし、ランダムに T_i を 5 セット選択した。 $|L| = 10$ のときのそれぞれの再現率および適合率を求めた結果を表 3 に示す。

同一テーマに基づいた検索履歴なので、各被験者間の類似度が高く、被験者全体が利用した検索キーワー

表 3 再現率および適合率

	$ T_i \cap L $	再現率	適合率
T_{i1}	2	0.67	0.2
T_{i2}	2	0.67	0.2
T_{i3}	1	0.33	0.1
T_{i4}	2	0.67	0.2
T_{i5}	2	0.67	0.2
平均	1.8	0.60	0.18

ド数も少ないため、高い再現率および適合率を得ることができた。評価実験におけるデータ数が少なかつたため、今回の実験結果は、信頼性という点では不十分である。

今後は、より多くの実験を行ってデータ数を増やし、推薦精度を高めていく。また、教員の評価を考慮した協調フィルタリングのアルゴリズムについて検討する。実際の授業で本提案システムを使用した検証実験を実施し、初等教育向けの Web 検索システムとして、従来の検索エンジンと比較して、本提案手法の有効性を検証する必要がある。

7. まとめ

本論文では、初等教育での Web 検索を利用した授業における問題点を明らかにし、これらの問題点を解決するために、協調フィルタリングを用いた Web 検索システムを提案した。本提案システムは、ユーザからの検索履歴をデータベースに蓄積し、協調フィルタリングアルゴリズムを適用することで、各ユーザに対して未使用の検索キーワードを推薦する。推薦アルゴリズムは、ユーザ間の類似度推定し、求めた類似度をに基づいてユーザ評価値を予測する。推薦アルゴリズムを再現率と適合率によって評価したが、データ数が少なかつたため、十分な結果を得ることができなかった。初等教育の授業における Web 検索エンジンの問題点を解決するために、今後は、より多くの実験を行って推薦制度を高めていく。

また、実際の授業で検証実験を行い、従来の検索エンジンと比較して、本提案システムの有効性を検証していく。

参考文献

- 1) 経済産業省商務情報政策局情報処理振興課：”eラーニング白書” 2006/2007 年版, pp267.
- 2) 清水拓也, 土方嘉徳, 西田正吾：発見性を考慮した協調フィルタリングアルゴリズムに関する基礎検討, 情報処理学会研究報告, Vol.2006, No.59, FI-83-7, pp53-60, 2006
- 3) 大杉直樹, 門田暁人, 森崎修司, 松本健一：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, Vol.45, No.1, pp.267-278, (2004) .
- 4) 森田昌宏, 速水治夫：情報フィルタリングシステ

- △, 情報処理, Vol.37, No8, pp751-758(1996).
- 5) Goldberg, D., Nichols, D., Oki, B.M. and Terry, D.: Using Collaborative Filtering to Weave an Information Tapestry, Comm.ACM, Vol.35, No12, pp.61-70 (1992) .
 - 6) Terveen, L., Hill, W., McDonald, D., and Creter, J. : PHIAKS: A System for Sharing Recommendations, Comm.ACM, Vol.40, No.3, pp.59-62 (1997) .
 - 7) Resnick, P., Iacovou, N., Suchak, M., Bergstorm, P. and Rjedl, J. : GoupLens : An Open Architecture for Collaborative Filtering of Netnews, Proc. ACM 1994 Conference on Computer Supported Work (CSCW'94) , pp.175-186 (1994)
 - 8) <http://developer.yahoo.co.jp>