

## 協調フィルタリングを利用した英語教材推薦アルゴリズムの研究

岩下文香 津田塾大学大学院理学系研究科

来住伸子 津田塾大学学芸学部情報科学科

協調フィルタリングとは、映画やCDなどに対する評価情報を集め、あるユーザと似た評価をした他のユーザの情報を利用して、あるユーザがまだ知らない作品の評価を推定する手法のことである。この報告では、協調フィルタリングを利用した英語学習者に適したポッドキャストを推薦するアルゴリズムの提案と実験結果について報告する。まず、予備実験で、ポッドキャストの評価の項目として 1. 内容の面白さ、2. 難易度、3. 教材としての有用度の3点を選んだ。本実験では、GroupLens [1]などで使われたユーザの類似度計算方法を利用して、3項目について予測評価値を算出した。実際の評価値と比較した結果、平均を使った予測値より、協調フィルタリングを使った予測のほうが予測誤差が小さいことが観察できた。また、予測値を利用した推薦結果の一致率と再現率についての考察も行った。

### **Study on Recommender System of English Listening Material for EFL Students using Collaborative Filtering**

Fumika Iwashita, Graduate School, Tsuda College

Nobuko Kishi, Department of Computer Science, Tsuda College

Collaborative filtering is a method of making prediction about the interest of a user on items such as music and movies, based on the information from other users. We applied collaborative filtering to recommend podcasts as listening material for ESL students. This report describes the results from two experiments to evaluate our recommendation algorithm. We asked ESL students to rate podcasts in three respects: interest, difficulty and perceived effectiveness. We made rating predictions based on the similarity metric used in GroupLens (Resnik et al. 1994). We found that the collaborative filtering-based method predicted better than the rating average method. We also evaluated the validity of recommendation results derived from the collaborative filtering-based prediction using precision and recall.

## 1. 背景と目的

インターネット上には英語で提供される様々なコンテンツがあり、英語学習者はそれらを学習教材として利用できる。特に音声データについては、ポッドキャストの普及と共に利用可能な学習教材数が増加してきた。しかし、利用できる数があまりに膨大であるため、その中から各自に適したポッドキャストを発見するのは困難な作業である。そこで、教材として学習者に適したポッドキャストを推薦するシステムの開発を考える事にした。

このようなシステムでは、適切な推薦手法が重要である。現在、推薦手法としては内容に基づくフィルタリングや協調フィルタリングを用いたものがよく知られている。内容に基づくフィルタリングでは、ユーザが過去に行った評価や行動を基に、ユーザプロフィールを作成し、それを利用してユーザに推薦するコンテンツを決定する。一方、協調フィルタリングでは、ユーザ同士の評価や行動の類似性を利用して、推薦コンテンツを決定する。

本研究では、NetNews 推薦システムである GroupLens[1]で初めて使用された協調フィルタリングの手法を英語学習教材向けに拡張した手法を提案し、その評価実験を行った。従来、コンテンツに対する評価は、「内容の面白さ」のような1つの観点からのみ行われていた。本研究では、推薦すべきコンテンツが学習教材である点を考慮して、1.内容の面白さ、2.難易度、3.学習素材として利用したいか、の3つの観点の評価を同時に扱うことにした。

本報告では、評価実験のために作成した

ポッドキャストディレクトリについて説明した後、3観点の独立性についての予備実験、協調フィルタリングを用いた評価予測値に関する実験、評価予測値を元にした推薦結果について考察を行う。

## 2. 実験システム

ポッドキャストは RSS(RDF Site Summary, Rich Site Summary など)を利用してWeb上で提供された音声データのことである。音声ブログのように定期的に配信されるものも多い。異なる Web サイトから提供されるポッドキャストの RSS をまとめて表示し、さまざまなポッドキャストをダウンロードしやすくしたアプリケーションをポッドキャストディレクトリ、またはアグリゲータなどと呼ぶ。よく知られたポッドキャストディレクトリとしては、iTunes Music Store や Yahoo ポッドキャストなどがある。これらのシステムでは、ポッドキャストのお勧め度を5段階の星印で評価する機能を提供している。今回の推薦手法では、3個の観点から評価する機能が必要なので、評価実験のためのポッドキャストディレクトリを作成した。WhitePlum ポッドキャストディレクトリ(図1)は、RSS フィードを利用して、ポッドキャストの登録、評価、お気に入り登録、ポッドキャストの視聴という機能を提供している。ポッドキャストの評価は、1.内容の面白さ、2.難易度、3.学習素材として利用したいか、の3項目について5段階評価が行える。

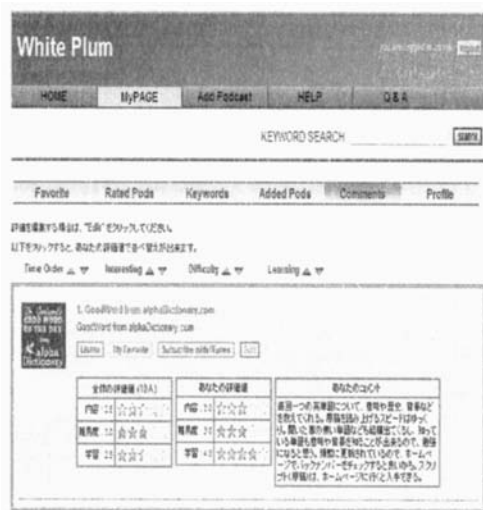


図 1 : WhitePlum ポッドキャストディレクタリの登録コメントと評価値の表示画面

### 3. 予備実験

どのような観点から行った評価が英語リスニング教材の推薦に必要なかを調査するために予備実験を行った。予備実験は、次のような条件で行った。

- ・ 被験者：情報科学科 1 年生 57 名
- ・ 実験時間：30 分
- ・ 音声データ：15 個
  - 1 個の長さは 1 - 2 分程度で、内容は童話、ニュース、情報科学などさまざまな話題のもの。

被験者は上記の音声データを聞いて、下記 3 項目について、5 段階で評価した。

1. どの程度聞き取れたか (難易度)
2. 続きを聞きたいか (興味)
3. 学習素材として利用したいか

3 項目の評価値間の相関関係を調べたところ、次のような相関係数になった。

- [1. 難易度-2. 興味] 間 : 0.27

- [1. 難易度-3. 学習] 間 : 0.136

- [2. 興味 - 3. 学習] 間: 0.583

これらの相関係数の値から、3 項目の間には強い相関関係はないと考え、3 項目ともユーザ評価の対象とすることにした。

### 4. 協調フィルタリングによる評価値予測

本研究では、GroupLens[1][2]で提案された計算方法で、ユーザ間の類似度とユーザがポッドキャストに与える評価の予測値を算出した。具体的には、ユーザ同士が共通に評価したポッドキャストの評価値で、ピアソン積率相関係数(Pearson correlation coefficient)を算出し、それをユーザ間の類似度とする。得られた類似度は、各ユーザ評価値に対する重みとして予測評価値の算出に利用する。

以下に、GroupLens で使われた予測評価値の算出方法について説明する。

$i$ : データ  $I$  (記事やポッドキャスト)

$a$ : ユーザ  $a$  (予測対象ユーザ)

$k$ : ユーザ  $k$  (ユーザ  $a$  の予測値算出に利用するユーザ)

$I_k$ : ユーザ  $k$  が評価したデータ集合

$n_i$ : データ  $i$  を評価し、かつユーザ  $a$  との類似度が 0 でないユーザ  $k$  の総数

$r_{k,i}$ : ユーザ  $k$  のデータ  $i$  に対する評価値

ユーザ  $k$  の評価値の平均:

$$\bar{r}_k = \frac{1}{|I_k|} \sum_{i \in I_i} r_{k,i} \quad (\text{式 1})$$

ユーザ  $k$  とユーザ  $a$  間の類似度：

$$w(a, k) = \frac{\sum_{i \in I_a \cap I_k} (r_{ai} - \bar{r}_a)(r_{ki} - \bar{r}_k)}{\sqrt{\sum_{i \in I_a \cap I_k} (r_{ai} - \bar{r}_a)^2} \sqrt{\sum_{i \in I_a \cap I_k} (r_{ki} - \bar{r}_k)^2}}$$

(式 2)

(式 2) で計算される類似度が 1 に近ければ正の相関でユーザ同士は類似していると言える。-1 に近ければ負の相関で互いに逆の好みを持つと言える。0 であれば相関はない。

ユーザ  $a$  のデータ  $i$  に対する評価の予測値

$$P_{a,i} = \bar{r}_a + \frac{\sum_{k=1}^n w(a, k)(r_{ki} - \bar{r}_k)}{\sum_{k=1}^n |w(a, k)|}$$

(式 3)

この式では、類似度を重みとして、ユーザ  $k$  のデータ  $i$  に対する評価から、ユーザ  $k$  の全ての評価値の平均で減じたものの加重平均を取っている。

本研究では、1.内容の面白さ、2.難易度、3.学習素材として利用したいかの 3 項目についてユーザに評価をさせている。そこで、(式 3) を上記 3 項目それぞれに用いて、ユーザの類似度と予測評価値の計算をすることにした。3 項目の 1 つを  $x$  とすると、本研究で使う予測評価値は(式 4)により求められる。

$$P_{a,i,x} = \bar{r}_{a,x} + \frac{\sum_{k=1}^n w(a_x, k_x)(r_{k,i,x} - \bar{r}_{k,x})}{\sum_{k=1}^n |w(a_x, k_x)|}$$

(式 4)

## 5. 評価値の予測実験

評価値の予測実験では、ユーザがつけた実際の評価値を集め、それらの評価値を利用して予測評価値を算出し、実際の評価値と予測評価値とを比較検討した。

実験では、情報科学科 1 年生 26 名が、ポッドキャストを聞き、1.内容の面白さ、2.難易度、3.学習素材として利用したいか、について 5 段階で評価をつけた。聞いたポッドキャストは *WhitePlum* ポッドキャストディレクトリにあらかじめ登録済みの 26 個の中から自由に選択する。選択したポッドキャストを聞いた直後に、上記 3 項目について評価する。時間は約 1 時間で、評価ポッドキャストの個数の上下限は設定しなかった。その結果、1 人の学生が平均 7.42 個のポッドキャストを評価した。

次に、評価値データを使って 4 章で述べた計算方法で、学生間の類似度を計算し予測評価値を算出した。予測評価値は次の (I)-(V) の 5 パターンで計算した。以下の記述では、 $S$  を予測算出対象の学生とする。

- (I) 項目別に、 $S$  以外の評価値の平均を予測値とする。
- (II) 項目別に、 $S$  以外の学生の評価値データから協調フィルタリングを使って予測する。
- (III) 項目別に、 $S$  以外の学生のうち、ポッドキャストを 4 個以上で評価した学

生の評価値データから協調フィルタリングを使って予測する。

(IV) 項目別に、(II)の半数の学生の評価値データから協調フィルタリングを使って予測する。

(V) 3項目の類似度を全て使って、S以外の学生の評価データから協調フィルタリングを使って予測する。

既存のポッドキャストディレクトリは、(I)のS以外の平均値を予測平均として使うことが多い。そこで、パターン(I)と、協調フィルタリングの代表パターンとして(II)について主に比較する。

### 5.1 内容の面白さの予測

図2はパターン(I)を使ったときの、内容の面白さの予測評価値の散布図である。横軸は、各被験者を実際の評価値の高い順に並べている。図3はパターン(II)の協調フィルタリングを使ったときの散布図である。

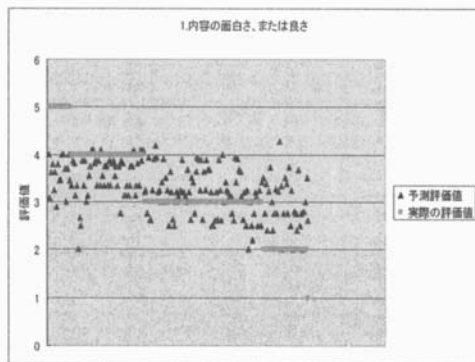


図2: (I) S 以外の評価値の平均

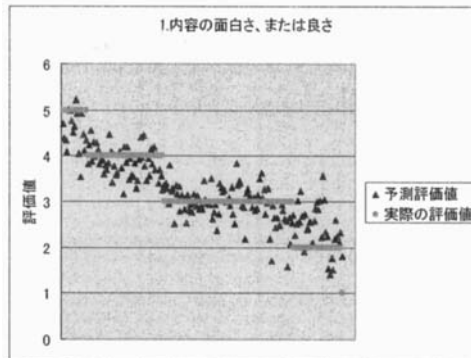


図3: (II)協調フィルタリング

図2と図3を比べると、図3のほうが、実際の評価値の近くに、予測評価値が散布していることが分かる。

### 5.2 難易度の予測

図4は難易度の平均による予測評価値の散布、図5は難易度の強調フィルタリング(II)による予測評価値の散布を示している。

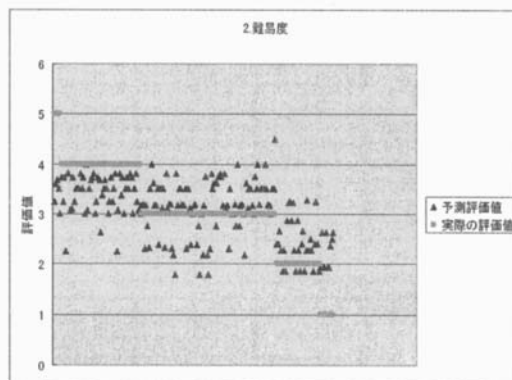


図4: (I) S 以外の評価値の平均

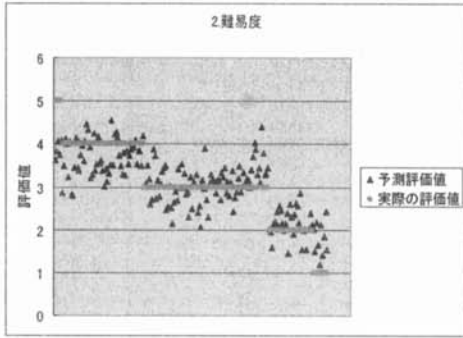


図 5: (II)協調フィルタリング

### 5.3 学習素材としての有効度の予測

図 6 は学習素材としての有効度の平均による予測評価値、図 7 は学習素材としての有効度の強調フィルタリング (II) による予測評価値の散布を示している。

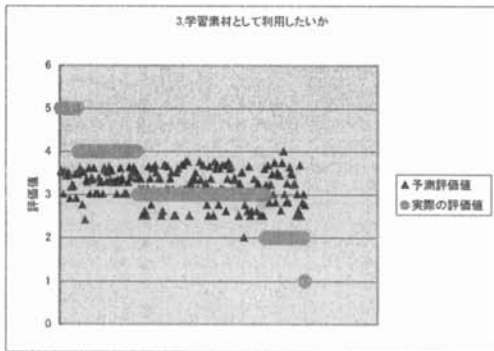


図 6: (I) S 以外の評価値の平均

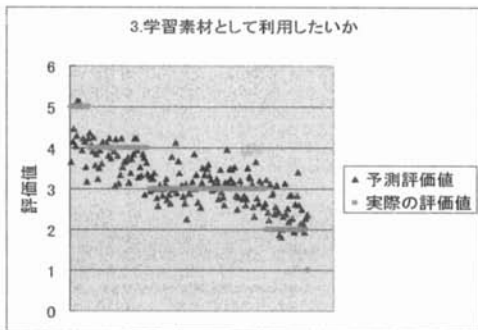


図 7: (II)協調フィルタリング

### 5.4 予測誤差の評価

予測誤差を、予測評価値と実評価値の差の平均偏差と標準偏差によって評価することにした。平均偏差は(式 5)、標準偏差は(式 6)により算出した。

$$\overline{E}_x = \frac{\sum_{i=1}^n |P_{a,i,x} - r_{a,i,x}|}{N} \quad (式 5)$$

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^n (E_x - \overline{E}_x)^2}{N}} \quad (式 6)$$

パターン	1. 内容の面白さ		2. 難易度		3. 学習素材としての有効度	
	E	$\sigma$	E	$\sigma$	E	$\sigma$
	(I)	0.635	0.807	0.57	0.717	0.688
(II)	0.366	0.473	0.40	0.518	0.389	0.519

表 1: 予測誤差の平均偏差と標準偏差

表 1 から、3 項目のどれも、パターン(I)の平均による予測より、パターン(II)の協調フィルタリングによる予測のほうが誤差が小さいことがわかる。また、協調フィルタリングによる予測誤差は、0.366 から 0.519 の値に収まっている。

表 2 は、パターン (II) 以外の協調フィルタリングを使った予測についての予測誤差を示している。パターン (II) は、どの項目



においても1番小さいか、2番目に小さい予測誤差を示している。

パターン	1. 内容の面白		2. 難易度		3. 学習素材と	
	E	$\sigma$	E	$\sigma$	E	$\sigma$
(I)	0.635	0.807	0.578	0.717	0.688	0.846
(II)	0.366	0.473	0.404	0.518	0.388	0.519
(III)	0.345	0.439	0.473	0.592	0.392	0.524
(IV)	0.430	0.542	0.485	0.617	0.397	0.563
(V)	0.397	0.522	0.466	0.613	0.540	0.663
(VI)	0.442	0.554	0.438	0.569	0.445	0.591

表 2: 全パターンの予測誤差の平均偏差と標準偏差

- ・ 予測評価値 3.0 以上 4.0 以下と実評価値 3.0 以上 4.0 以下
- ・ 予測評価値 3.0 以上 3.5 以下と実評価値 3.0 以上 3.5 以下
- ・ 予測評価値 2.5 以上 3.0 以下と実評価値 2.5 以上 3.0 以下

	5.0 以上	4.5以上	4.0以上	4.0~3.0	3.5~3.0	3.0~2.5
1.内容	100%	100%	100%	-	-	-
2.難易度	-	-	-	97%	72%	75%
3.学習	100%	100%	96%	-	-	-

表 3: 推薦結果の適合率

## 6. 予測評価値を利用した推薦

強調フィルタリングのパターン(II)で得られた予測評価値を使ったポッドキャストを推薦するには、いくつかの方法が考えられる。「内容の面白さ」や「学習素材として有用度」を利用する場合は、予測評価値が最も高いものを推薦することが考えられる。また、「難易度」は、本人指定の難易度に近い予測評価値のものを推薦することが考えられる。

そこで、「内容の面白さ」と「学習教材としての有効度」については、次の集合の組み合わせの適合率 (precision) と再現率 (recall) をそれぞれ調べた。

- ・ 予測評価値 5 以上と実評価値 5 以上
- ・ 予測評価値 4.5 以上と実評価値 4.5 以上
- ・ 予測評価値 4 以上と実評価値 4 以上

「難易度」については、以下の集合の適合率と再現率をそれぞれ調べた。

	5.0以 上	4.5以上	4.0以上	4.0~3.0	3.5~3.0	3.0~2.5
1.内容	12%	19%	46%	-	-	-
2.難易度	-	-	-	63%	51%	28%
3.学習	16%	16%	37%	-	-	-

表 4: 推薦結果の再現率

表 3 と表 4 は推薦結果の適合率と再現率をそれぞれ示す。これらの表から、予測評価値を使った推薦は、適合率が高いが再現率は高くないことが分かる。

## 7. まとめと今後の課題

本研究では、英語リスニング教材の推薦のために、3項目の評価値を協調フィルタリングを使って予測する実験を行った。実験は、ポッドキャストディレクトリを利用して、26人の評価データを集め、25人のデ

一人から残りの1人の評価値を予測した。ユーザ評価の平均を使った予測より、協調フィルタリングを使った予測のほうが予測誤差が小さいことが観察できた。協調フィルタリングによる予測では5段階評価を予測誤差0.4程度で予測できた。

予測評価値を利用した推薦結果の適合率は、「内容の面白さ」や「学習教材としての有効度」では高いが、「難易度」についてはやや低かった。

データ数が少ない割に予測誤差が0.4程度と小さかったのは、被験者の考え方や好みがよく似ているためではないかと考えられる。また、難易度については推薦結果の適合率がやや低いのは、好みが似た被験者の集合とはいえ、英語のレベルに差があるので、難易度の評価が分かれるのではないかと考える。

協調フィルタリングは、データ数が多くなるほど予測精度が高くなると通常考えられているので、今回の実験結果は、より大きなデータ数を扱う場合も、協調フィルタリングを使った推薦が有望であることを示唆している。

今後は、より多くのユーザ評価を集めて実験し、ポッドキャストの推薦方法を確立したい。また、現在のポッドキャストディレクトリに推薦機能以外の機能を追加し、英語学習者向けポッドキャストディレクトリの実用化を目指したい。

## 9. 参考文献

[1] Paul Resnick , Neophytos Iacovou , Mitesh Suchak , Peter Bergstrom , John Riedl, *GroupLens: an open architecture*

*for collaborative filtering of netnews*, Proceedings of the 1994 ACM conference on Computer supported cooperative work, p.175-186, October 22-26, 1994, Chapel Hill, North Carolina, United States

[2] Joseph A. Konstan , Bradley N. Miller , David Maltz , Jonathan L. Herlocker , Lee R. Gordon , John Riedl, *GroupLens: applying collaborative filtering to Usenet news*, Communications of the ACM, v.40 n.3, p.77-87, March 1997