

## 協調フィルタリングに基づく改良版授業推薦システムの構築と評価

蒋 再興<sup>†</sup> 檜垣 泰彦<sup>†</sup> 荒井 幸代<sup>†</sup>

†千葉大学大学院工学研究科 〒263-8522 千葉県千葉市稻毛区弥生町 1-33

E-mail: †higaki@tu.chiba-u.ac.jp

あらまし 協調フィルタリングを用いたユーザベースの授業推薦システムのプロトタイプを作成し評価を行なったところ、網羅性、精確性の改善、ユーザインターフェースの改善、時間割による制限への対応が必要であることがわかった。それに対応するため、アイテムベースの推薦を追加し類似度の計算尺度の改善を行なった。また、自分と類似した先輩の履修授業を推薦する先輩推薦と、時間割を考慮した協調推薦の機能を設けた。さらにアイテムベースの相関推薦も併用することで学生の主体性の維持を図った。推薦画面を時間割ベースに変更した。この評価の結果、この改良版プロトタイプは、網羅性と精確性を備えていること、高い主体性の維持が確認できた。

**キーワード** 協調フィルタリング、授業推薦、アイテムベース、コサイン類似度、Item-to-Item

## Construction and Evaluation of Improved Lecture Recommendation System based on Collaborative Filtering

Zaixing JIANG<sup>†</sup> Yasuhiko HIGAKI<sup>†</sup> Sachiyo ARAI<sup>†</sup>

†Graduate School of Engineering, Chiba University 1-33,Yayoi-cho,Inage-ku, Chiba, 263-8522 Japan

E-mail: †higaki@tu.chiba-u.ac.jp

**Abstract** The Lecture Recommendation System based on User-Based Collaborative Filtering had been completed. The improvement of comprehensiveness, precision and user interface would be needed. In order to improve the system, Item-based Collaborative Filtering and Cosine Similarity were introduced. The Senior Recommendation could recommend lectures the seniors who are similar to the user registered. The Collaboration Recommendation based on User-Based Collaborative Filtering and the Item Recommendation based on Item-Based Item-to-Item Collaborative Filtering had been developed to improve the initiative of students. In evaluation of the improved prototype with the form of a timetable, it is confirmed that the system performance was increased.

**Keyword** Collaborative Filtering, Lecture Recommend, Item-Based, Cosine Similarity, Item-to-Item

### 1. はじめに

大学環境において、授業の多様化が進み、学生自身の趣味趣向に合わせて、受講することが可能になってきており、授業選択の自由度が極めて高くなっている。そのため、授業選択の際には、授業シラバスを何度も参照し、自分が履修すべき授業なのか、また、自分が興味の持てる授業なのか、などの情報収集を行わなければならない。このため、授業選びの助言を必要とする学生などには、不親切となり得る。また、履修計画に頭を悩ます学生や、履修漏れをしてしまい、進級・卒業できない学生が出てきてしまうことが懸念される。

千葉大学工学部学生の履修データをもとに授業選びの効率化と履修漏れ回避、そして学生自身が存在を知らなかった「興味がある授業」との出会いの可能性の拡大を目的として、授業推薦システムのプロトタイプ（以下これを「プロトタイプ1」とする）を試作した[1]。本研究ではさらに、推薦エン

ジンとユーザインターフェース両方について改良したプロトタイプ（「プロトタイプ2」とする）を試作し、評価する。

### 2. 目標

#### (1) 網羅性と精確性を改善し、発見性を高める

情報推薦と検索の重要な要素として、網羅性と精確性が挙げられる。授業推薦においても、網羅と精確が重要である。データの質と量が不足すると、推薦の精度に影響する。推薦の精度が低いと、システムは利用される頻度が低下すると考えられる。システムの発見性は高い網羅性と精確性に依存している。また、嗜好一致の問題を改善するため、学生にとって興味ある授業を網羅することや学生の嗜好を精確に把握することを目標として、システムの性能を改善する。

#### (2) 学生に主体性を十分に發揮させる

## Item-to-Item Algorithm

```
For each item in product catalog, I1
  For each customer C who purchased I1
    For each item I2 purchased by customer C
      Record that a customer purchased I1
      and I2
  For each item I2
    Compute the similarity between I1 and I2
```

図1. Item-to-Item アルゴリズム

推薦された結果だけしか見ずに授業を選択してしまうと、主体性の育成が阻まれる。ちなみに、授業推薦時に、学生自身の意思を發揮させることも重要である。プロトタイプ1の機能だけでは主体性維持には十分ではない。推薦と主体的選択のバランス点を見つけなければならない。よって、各種の方法を通じて、授業の比較と取捨は学生自身が主体的に決定できるシステム設計にすることが望ましい。

### (3) 利便性と効率を改善する

プロトタイプ1では推薦結果の計算・表示速度は測定の結果、約3秒であった。データセットの作成から推薦結果の出力まで含むこの値は十分に速い。しかし、プロトタイプ2ではさらに多くの機能を実装することになり、このレスポンスを維持するためにはさらにシステムの効率を上げる必要がある。多機能であっても高速に授業を選択して履修登録できることが望まれる。利便性とレスポンスが悪ければ、システムを実用することは困難である。そのため、システム自体の高いレスポンスの維持とユーザビリティを感じさせる安定したサービスを目指す。

## 3. システムの改良

### 3.1. 推薦エンジンの改良

#### 3.1.1. アイテムベースの導入

アイテムベースとは、ユーザの履歴を基にアイテム間の関連性を分析し、あるアイテムをユーザが選択した時に関連性の高いアイテムを表示する手法である。プロトタイプ1では、ユーザベースの手法だけを用いてシステム設計を行ってきた。しかし、授業同士、似ているものを知りたい場合には、図1に示す Item-to-Item アルゴリズムによるアイテムベース協調フィルタリングの手法が有効である。

Item-to-Item 協調フィルタリングでは、特定のアイテムを誰が好きなのかを調べ、彼らが好きな他のアイテムを探す[2]。本質的には先ほど行ったユーザ間の類似度を決めるやり方と同じで、ユーザとアイテムを入れ替えるだけで実現できる。

プロトタイプ2では、この手法で各授業の相関推薦の機能

を導入し、網羅性と発見性の高い推薦システム構築を目指す。それにより学生自身の主体性を高めることができる。

#### 3.1.2. 類似度の計算

本研究では、二値変数で構成されるベクトルを計算する場合には、ピアソン相関係数などの線形相関尺度は使用できない。プロトタイプ1では正規化されたユークリッド距離(式(1))またはコサイン類似度(式(2))を利用可能である。

$$d(f, g) = \frac{1}{1 + \sqrt{\sum_{i=1}^n (f_i - g_i)^2}} \quad (1)$$

$$\cos\_sim(x, y) = \frac{\sum_{i=1}^n x_i \cdot y_i}{\sqrt{\sum_{i=1}^n x_i^2 \cdot \sum_{i=1}^n y_i^2}} \quad (2)$$

これらを比較すると以下のことが言える。ユークリッド距離の基本的概念としては、人々が評価したアイテムを軸にとって嗜好空間上に人々を配置し、それぞれの距離がどれくらい近いかを測るものであり、コサイン類似度計算法とは、それぞれの二つずつベクトルのなす角の余弦値を測るものである。授業推薦の場合には、点の間の距離よりベクトルの余弦値が適切である。もある学生あるいは授業はゼロベクトルであれば他の学生、授業との類似度の計算結果は0になるべきである。しかし、正規化されたユークリッド距離では二つのゼロベクトルの類似度は1になり、授業推薦の場合には不適切である。更に、正規化されたユークリッド距離はベクトルの次元が増すにつれ、収束の速度も緩くなる。つまり、ベクトルの次元が高くなれば各ベクトルの類似度間の差別は小さくなる。最後の加重平均計算から精確な結果が与えられない。コサイン類似度ではこのような心配はない。

本研究では、このコサイン類似度の尺度  $\cos\_sim(x, y)$  を用いて、類似度を算出する。

#### 3.1.3. データセットの拡大

網羅性を実現するためには、全ての学生のデータを利用するのが一番良い。しかし、データが多すぎるとシステムの速度は遅くなる。その反面、データセット中のデータが不足する場合と、速度は速くても適切な推薦を行えず役に立たない。このようなバランスに配慮し、「同学科の一つ上の先輩の全履修データ」の代わりに「同学科の全ての先輩の全履修データ」を用いることで、速度と網羅性を両立できる新たなデータセットを作成する。

## 3.2. ユーザインターフェースの改良

### 3.2.1. 時間割表現

プロトタイプ2では、時間割表現を用いて各时限の推薦授業を推薦値の順に並べて提示する。また、登録するすべ

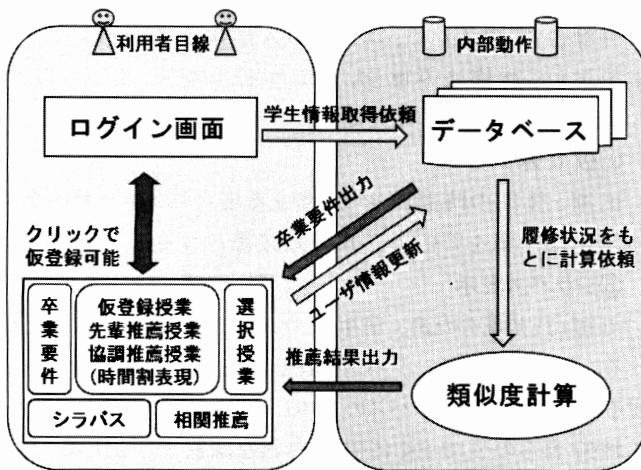


図2. システムの処理の流れ

ての授業も時間割形式で表示する。時間割表現の採用で1ページずつの推薦リストより見やすくなり、授業を探すにも便利になる。ユーザが「興味のある授業」を仮登録すると、データベースのユーザ情報を更新して類似度の再計算を行い、新たな推薦授業を表示する。

### 3.2.2. 多様な推薦方式の導入

本システムでは、先輩推薦、協調推薦と相関推薦の3種類の推薦方式を用いた。先輩推薦では、「システムを利用する学生の履修データ」と「同学科の全ての先輩の全履修データ」を用いて類似度計算を行い、ユーザと似ている先輩の該当する时限の履修履歴を順に表示する。協調推薦では、協調フィルタリングの技術に基づいて、同じデータセットで評価値を計算して、時間割中に順に表示する。相関推薦では、アイテムベースでアイテム間の類似度を使用して、推薦する機能を導入する。多様な推薦方式を用いて、学生の主体意思を十分に發揮させる。授業網羅の範囲を拡大して、先輩と授業という二つの視点から推薦すれば、興味のある授業を発見する機会が多くなる。

## 4. システムの設計

### 4.1. 全体の構成

本システムは MySQL, PHP と jQuery を用いて、Web アプリケーションとして作成した。

図2に示したログイン画面で学生証番号を入力し、学生の情報をデータベースから取得する。ここで取得した情報はまず始めに卒業要件の表示で用いる。卒業要件の表示は履修漏れの確認を数値で判断できる材料である。同時に卒業要件と推薦結果とを学生自身が照らし合わせることで、主体的に授業の取捨選択を行うことが可能になる。

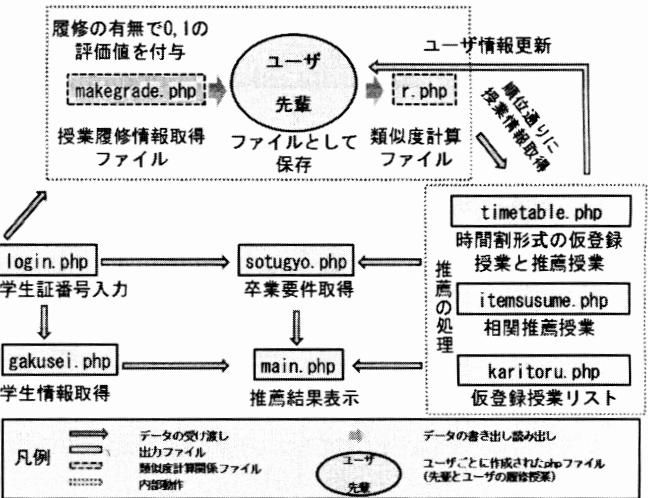


図3. データの流れ

また、推薦結果を表示するにあたっては、上記で述べた学生の情報と先輩の履修情報も取得することになる。その際、推薦結果を瞬時に計算・表示できるようにするために、ログイン学生と同じ学科の全ての先輩の履修データを利用することにした。これにより、結果表示までの時間を少なくすることができる。また、もし授業をまったく履修していない場合でも先輩の情報を元に推薦が可能である。同時に、情報を絞ることで授業の変更にあまり左右されない仕様にもなっている。推薦結果の中で授業名称をクリックすると、この授業のシラバスと相関授業リストの閲覧が可能になる。また、もし興味がある授業を発見した場合には、仮登録の機能を使用できる。これにより、仮登録した授業は時間割中と Web ページの右下に「選択授業」として表示され、任意に増加や、取り消しを行うことができる。同時に、仮登録分の単位をもとに Web ページ左下の「卒業要件」を再計算する。

### 4.2. データの処理

図3で示すように、login.php で学生証番号を入力すると、gakusei.php と sotugyo.php でユーザの学生情報と卒業要件を取得し、main.php で表示する。同時に、makegrade.php で先輩評価値とユーザ評価値の行列を生成することになり、そのデータは r.php でユーザの履修データと、すべての先輩の履修データとの類似度計算ならびに評価値の加重平均の算出に使用される。協調推薦のため、算出された推薦値は、順位付けされた授業コードとして、timetable.php へと値を渡し、時間割形式で推薦結果を出力する。先輩推薦のため、ユーザと類似度の高さの順に先輩の履修データを順位付けされた先輩たちの学生番号として、timetable.php へ渡して結果を時間割形式で表示する。そして、シラバス中の授業の相関推薦処理は

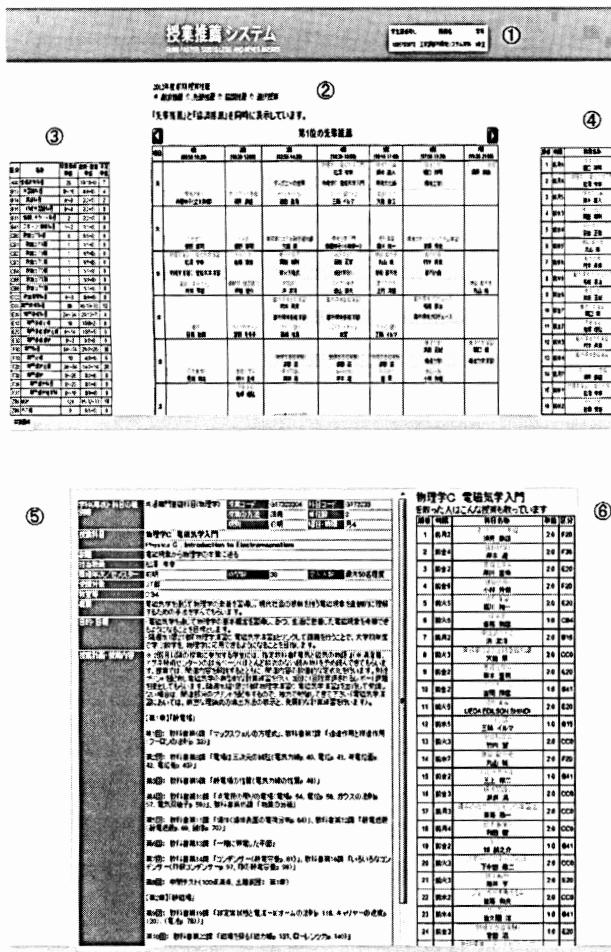


図4. メイン画面

`itemsusume.php` の中で行われて、授業リストで結果を出力する。仮登録機能に関しては、`timetable.php`, `itemsusume.php`と`karitoru.php`など推薦処理の中で行われる。ユーザ評価値の行列を更新して、もう一度前述の計算を行う。これが新たな推薦結果として出力される。それと一緒に推薦処理のファイルと`sotugyo.php`にも値が渡り、仮登録分の取得単位計算が行われる。取り消す処理も同様の手順を踏む。

#### 4.3. 実際の動作

プロトタイプ2では学生がログイン後、図4のような表示を展開する。

##### ①学生情報

ユーザの学年・所属名・学生証番号を表示。

##### ②仮登録と各推薦の時間割表示

ユーザの仮登録リスト、先輩の履修授業リストと協調推薦授業リストを時間割形式で表示する。授業の科目名称と教員名の二つを時間割から評価値の順に並べる。また、「取る」と「消す」を選択すると再計算(再推薦)が行われる。

##### ③卒業要件

目標(2)実現のため、ユーザの入学年度や学科に応じた卒業単位、履修済み単位、不足単位を表示。また、「取る」と「消す」の選択を行うことで、取得単位を再計算する。

##### ④仮登録リスト

前期と後期の時間割を切り替える場合に、ユーザの全ての仮登録授業をリストの形式で直接表示する。

##### ⑤シラバス表示

目標(2)実現のため、学生自身が推薦結果と内容を見比べができるようシラバスを表示する。科目名称をクリックするとその授業のシラバスの URL にジャンプするようになっており、自らの目ですぐに授業内容を確認できる仕様となっている。

##### ⑥相關授業の推薦

シラバスの隣に相關授業(「この授業をとった人はこんな授業も取っています」)を表示する。

## 5. 実験

プロトタイプ2について客観と主観両方から性能を確認する。

### 5.1. 客観的実験

推薦授業と学生の履修データを利用して、先輩推薦と協調推薦の評価を行う。実験対象としたのは、特にシステムを利用するとと思われる1年生と2年生であり、それぞれ実験データとして、今までの履修データと、被験者と同学科の全ての先輩の履修データを使用する。実験方法としては、1年生と2年生が2011年度後期の授業は登録されないと仮定し、先輩の履修データのみを使って、授業推薦表示を行う。

時間割中の推薦授業は集合であり、要素には順番がない。それを協調フィルタリングの評価値を用いて順位付けすることにより、適合率と再現率によって本システム性能の優劣を比較可能となる。なお、1年生と2年生の実際の履修データはともにテストコレクションとし、評価用のデータとして使用する。実験の評価指標には、適合率と再現率、そして F 尺度の 3 つを用いる[3][4]。先輩推薦は推薦する先輩の順に、一人ずつ先輩の履修した授業を推薦値の順に並べ、重複の授業を削除して、推薦結果を生成する。協調推薦は時間割のコマ毎に順位順に、第 1 行目から重複の授業を消して、推薦結果を生成する。その際、推薦結果数を 10 ずつ増やしていく、それぞれの数値を記録する。

推薦結果内の履修授業数を R、推薦結果の授業数を N、実際の履修授業数 C とすると、適合率 precision、再現率 recall と F 尺度 F-measure は、次のように表せる。これを推薦結果数ごとのシステム性能評価値とする。

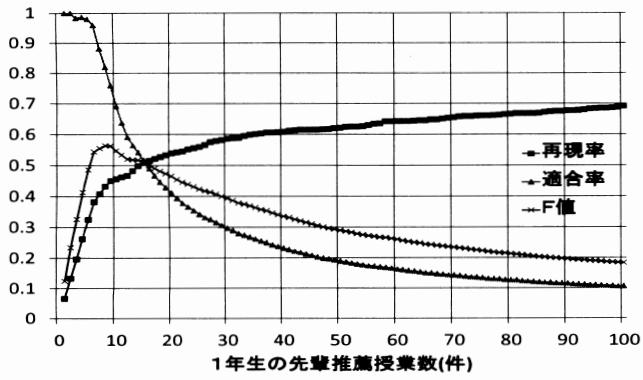


図5. 先輩推薦評価グラフ

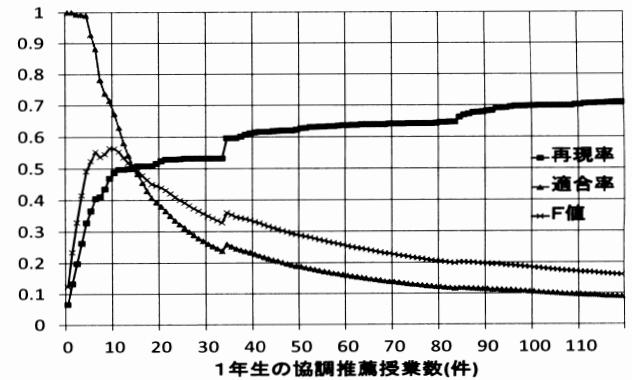


図6. 協調推薦評価グラフ

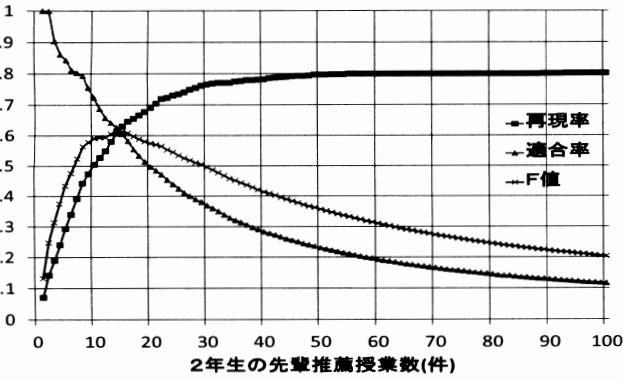


図5. 先輩推薦評価グラフ

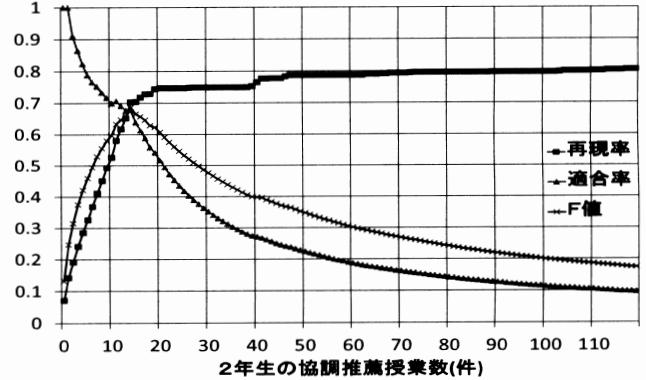


図6. 協調推薦評価グラフ

$$precision = \frac{R}{N} \quad recall = \frac{R}{C}$$

$$F\text{-measure} = \frac{2 \cdot precision \cdot recall}{precision + recall}$$

## 5.2. 主観的実験

本実験では、都市環境システム学科の1年生を対象として、プロトタイプ2について、実際に授業推薦を体験してもらった。また、四つの視点からアンケートを実施した。

### (1) 嗜好一致についての質問

本システムの推薦結果が、どの程度ユーザの嗜好に一致しているかを探るために、「取りたいと思える授業」があったかを問う。また、5段階のリッカート尺度を使用し、学生がどの程度合意できるかを確認する。

### (2) 発見性についての質問

本システムの推薦結果内に、「今まで知らなかつたけど、取つてみたい授業」はあつたかということを問う。また、この選択には発見される授業の数と履修したい授業の数はいくつかということを収集する。

### (3) 推薦方式についての質問

協調推薦と先輩推薦の精確性と有効性について質問す

る。そして、相関推薦（「あの授業を取つた人はこんな授業も取つています」）の機能は便利かどうかを確認する。

### (4) 便利度についての質問

ユーザが授業推薦システムを欲しているかを問うために、本システムは便利かを問う。

## 6. 評価

### 6.1. 客観的実験の評価

客観的実験では、先輩推薦と協調推薦の推薦結果を用いてプロトタイプ2に対する評価を行つた。客観的実験の被験者人数は、1年生が59人、2年生が54人である。

先輩推薦の評価は先輩順に推薦授業の数を増やしていく、先輩推薦授業の和集合とテストコレクションの集合を用いて適合率、再現率とF値を計算した。協調推薦の評価は時間割の各コマについて各同一推薦順位毎の和集合とテストコレクションの集合で計算した。先輩推薦は図5、協調推薦は図6に示すような数値になった。

### (1) 1年生と2年生の評価

図5と図6から、1年生の場合には、先輩推薦でF値が8件、協調推薦で10件のときに最高値となつた。2年生の場合には、先輩推薦でF値が14件、協調推薦で15件のときに最高値となつた。なお、1年生の履修授業の平均件数は

14件、2年生は15件である。全部の数値が、2年生は1年生より高い。それに、再現率と適合率の収束速度も速く、システムのF尺度の最高値も平均件数の位置に近い。以上より2年生に対する推薦精度は1年生より高い。

## (2) 先輩推薦と協調推薦の評価

協調推薦の最大再現率は先輩推薦より高く、先輩推薦のF値は協調推薦より早く最大値になる。先輩推薦では一人につき約14件表示され、協調推薦では最初に曜日毎に約30件表示される。そのため先輩推薦は精度が高く、協調推薦は全ての履修可能な授業を見る能够ため網羅性がよいといえる。この二つ方式を併用すれば、プロトタイプ1より、推薦授業の精度と範囲を向上させることができる。

客観的実験のグラフを見てみると、本システムの全体的な性能も安定したものになっている。

## 6.2. 主観的実験の評価

主観的実験では、本システムの有効性および、授業推薦の必要性をアンケート結果から確かめることを行った。アンケートの被験者は、都市環境システムの1年生に呼びかけた結果、40人中23人の学生から回答を得た。アンケート回収率は約56%となった。

### (1) 本システムの嗜好一致

本システムの嗜好一致に関しては、平均評価値は4.0であり、ほぼ一致している。

### (2) 本システムの発見性

本システムの発見性に関しては、「知らなかったけど、受けてみたい授業」があったと答えた人は、23人中19人となり、高い発見性を示している。

### (3) 本システムの推薦方式

先輩推薦は協調推薦より精確性と有効性が高い評価になった。一つずつの授業推薦より、学生は時間割のような履修経験が欲しいためである。また、学生は自身の意思で調査・比較する意欲が強く、相関推薦を用いて自身の主体性をもっと良く發揮させることできる。

### (4) 本システムの便利度

本システムの便利度に関しては、一番高い評価値4.2が得られており、高い便利性を示した。

### (5) 本システムへの意見

本システムのアンケート実施の際に、意見・感想記入欄も設けており、合計11人の意見が寄せられた。それらは肯定的意見と是正要求の意見であり、システムや授業推薦に対しての反対意見などはなかった。

## 6.3. 利便性と効率の評価

本システムでは、様々な新機能を導入した。そのためユーザベース用のデータセットだけではなく、アイテムベース用のデータセットも必要になっている。高い効率を維持するた

め、データセットは事前に用意した。データセットの作成時間が必要ないので、推薦計算・表示速度は測定の結果、1秒ほどとなっており、プロトタイプ1よりも早い。アンケートの便利度の結果から、目標(3)を満たすことに成功した。

## 7. 考察

客観と主観の実験評価により、プロトタイプ2の全般的な性能と効果はプロトタイプ1より高くなった。しかし、アンケートにおいて、専門科目と普遍教育科目、必修選択の別に分けて処理し、有効な単位の授業だけを推薦するべきであることが指摘された。卒業要件上無効な単位を取ったために必修授業の履修漏れになってしまふ場合があり、改良する必要がある。

また、評価実験において、推薦された授業は実際履修した授業を網羅できておらず、学生の嗜好に沿った高レベルな授業推薦が十分にできていないと判断できる。この問題を解決するため、データセットをさらに拡大し、最大再現率を1に近づけるべきである。二値変数ではなく、五段の評価をつける形式に変えれば、さらに嗜好に合った推薦を行うことが可能となってくる。

さらに、新規の学生・履修・授業の3つのデータの取得・管理の問題や、協調性の高いシステムにしていかなければならぬという問題などが残されたままである。

## 8. おわりに

授業推薦プロトタイプ1の網羅性と精確性を改良する目標を達成できたと考えている。また、プロトタイプ1既存の機能に加え相関推薦を利用することで、時間割から学生が主体的に授業を選択できるように改良できた。さらに1秒ほどのレスポンスで計算・出力が可能な使いやすい利便性と効率を改良する目標も同時に達成できたと考えている。

しかし、授業推薦の網羅性と精確性には、さらに改良の余地がある。また、学生状況を詳しく分析し、有効な推薦を行うことについて改良の必要もある。授業推薦の実用化と運用のためには、これらの問題をクリアし、学生の期待に答えられるようなシステムとすることが必要である。

## 文 献

- [1] 蒋再興, 江村裕介, 檜垣泰彦, LOIS2010-94“協調フィルタリングに基づく授業推薦システム”, 電子情報通信学会技術研究報告, vol.110, no.450, pp.173-178, Mar.2011.
- [2] Greg Linden, Brent Smith and Jeremy York, Amazon.com Recommendations Item-to-Item Collaborative Filtering Internet Computing, IEEE, pp.76-80, Feb.2003.
- [3] 北研二, 津田和彦, 獅々堀正幹, “情報検索アルゴリズム”, 共立出版, 東京, 2002.
- [4] 土方嘉徳, “嗜好抽出と情報推薦技術”, 情報処理学会論文誌, vol.48, no.9, pp.957-965, Sep.2007.