

1. はじめに
2. 教学データ分析  
と情報推薦
3. 授業科目の成績  
予測と教材の最  
適化
4. 提案手法
5. 数値実験並びに  
考察
6. おわりに

# 教学データに基づく GPA 向上とレビュー信頼性を考慮した 情報推薦・学習支援システムの開発

Development of a Recommendation and  
Learning Support System Based on Teaching and Learning  
Data Considering Review Reliability for GPA Improvement

滝沢 光介 (Kosuke Takizawa)  
t815043@st.pu-toyama.ac.jp

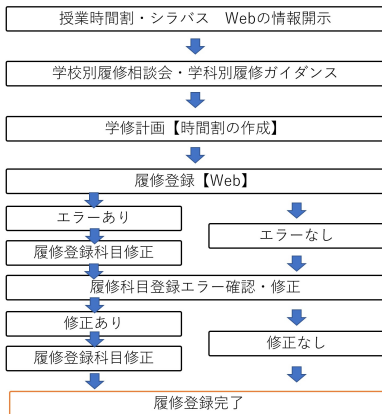
富山県立大学 工学部 電子・情報工学科 情報基盤工学講座

Teams, 14:00-14:25 Wednesday, February 16, 2022.

# 1.1 研究の背景

2/16

大学において上位の GPA を収めることで学生は就職，進学双方において有利になる。しかし，自分がよい成績を修めることができるかどうかはその科目を履修してみないとわからない。さらには，大学における取得単位の組み合わせは膨大であり，学生にとって負担となる。



図：単位履修までの流れ

表 成績からGPへ変換

| 評価 | 数字 | 点数            | 可否  |
|----|----|---------------|-----|
| S  | 4  | 90 点以上        | 合格  |
| A  | 3  | 80 点以上 90 点未満 |     |
| B  | 2  | 70 点以上 80 点未満 |     |
| C  | 1  | 60 点以上 70 点未満 |     |
| 不可 | 0  | 60 点未満        | 不合格 |
| 空白 |    | 履修なし          |     |

## GPAの計算方法

$$GPA = \frac{(S \text{ の数}) \times 4 + (A \text{ の数}) \times 3 + (B \text{ の数}) \times 2 + (C \text{ の数}) \times 1}{\text{履修した単位数}}$$

- はじめに
- 教学データ分析と情報推薦
- 授業科目の成績予測と教材の最適化
- 提案手法
- 数値実験並びに考察
- おわりに

## 1.2 研究の目的

3/16

過去の卒業生の教学データに対して分析を行い、成績評価の予測を行い、高い GPA を獲得できるように取得単位を推薦するとともに、全ての科目について良い成績がとれるように Web 上から関連情報を推薦するようなシステムの開発を行う。

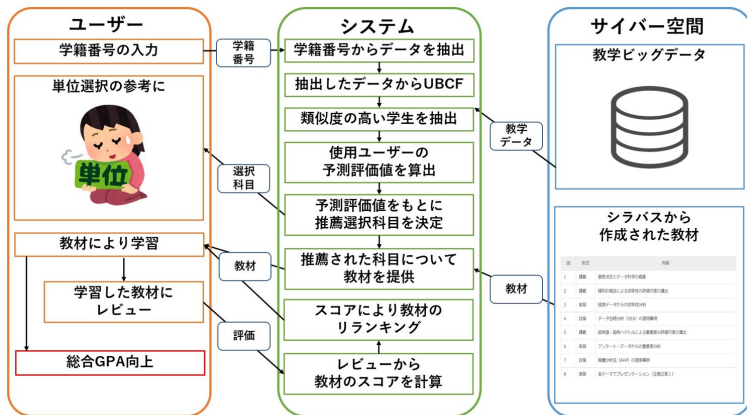


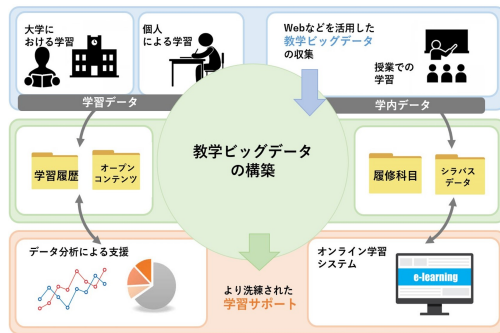
図 1.2: 提案システムの流れ

- はじめに
- 教学データ分析と情報推薦
- 授業科目の成績予測と教材の最適化
- 提案手法
- 数値実験並びに考察
- おわりに

## 2.1 教学におけるビッグデータ・アナリティクス

4/16

近年、ICT 技術の発展で多くの情報がデータとして蓄積されるようになってきている。これは教育現場においても例外ではない。教育現場におけるさまざまな情報のことを教学データと呼び、教学データに対して分析を行うことを教学ビッグデータアナリティクスと呼ぶ。教学データには 6 つの特徴があるとされている<sup>1</sup>。



図：教学データの例

| 取得時期   | 教学データ                    | 内容  |
|--------|--------------------------|---|
| 入学前    | 出身高校<br>入試情報<br>入学前学習    | 課程差別、判定値, etc.<br>入試区分、成績<br>取組状況、提出物                       |
| 入学時    | 導入教育                     | オリエンテーション、テスト結果, etc.                                       |
| 各セメスター | 履修登録<br>授業<br>学生生活<br>成績 | 履修科目<br>出欠状況、課題提出, etc.<br>部活、アルバイト, etc.<br>科目成績、GPA, etc. |
| 4年次    | 就職活動                     | 活動履歴、内定状況   |
| 卒業後    | 卒業後                      | 満足度、アンケート   |

- ・データ量はそれほど大きくない
- ・対象人数は少ないがデータの種類の急増している
- ・匿名性が低い
- ・データの意味解釈が容易
- ・因果関係は複雑
- ・多様化・細分化が進行している

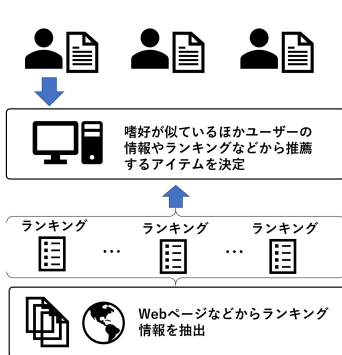
図：教学ビッグデータアナリティクス

<sup>1</sup>松田岳士, 渡辺雄貴, “教学 IR, ラーニング・アナリティクス, 教育工学”, 日本教育工学会論文誌, Vol. 41, No. 3, pp. 199-208, 2017.

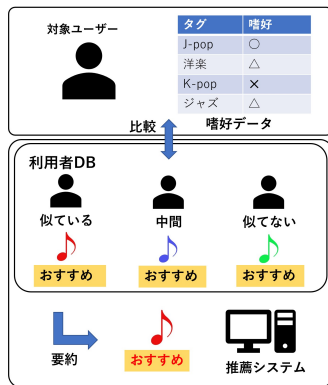
## 2.2 情報推薦と協調フィルタリング

5/16

情報推薦のアルゴリズムとして協調フィルタリングがある。協調フィルタリングは与えられたデータから規則性を見つけ出し，利用者がまだ知らない情報を予測するものである<sup>2</sup>。本研究では，教学データに対してユーザーベース協調フィルタリング(UBCF)を適用する。



図：情報推薦



図：協調フィルタリング

<sup>2</sup>神嶋敏弘, “推薦システムのアルゴリズム”,

## 2.3 レビューの信頼性の判断支援

6/16

1. はじめに
2. 教学データ分析  
と情報推薦
3. 授業科目の成績  
予測と教材の最  
適化
4. 提案手法
5. 数値実験並びに  
考察
6. おわりに

Amazon などの EC サイトにおいてユーザーから商品へのレビューは重要な役割を担っている。レビューの価値を使用し偽のスパムレビューを投稿し商品の評価を不当に操作する行為が問題となっている。その解決のためにレビューが信頼できるものなのかを判断するための研究が行われている。



図：ECサイトにおけるレビュー



図：日本の主なECサイト

サクラグループ



図：サクラグループによるスパムレビューの投稿

## 3.1 協調フィルタリングからの科目の推薦

7/16

UBCF における利用者=学生, アイテム=科目, 評価値=成績と置き換えて学生に対して成績を予測し, 科目の推薦を行う。

### 嗜好データにおけるUBCF

表: 成績値行列

|       | 1: そば | 2: ラーメン | 3: うどん | 4: パスタ |
|-------|-------|---------|--------|--------|
| 1: 高橋 |       | 2       | 3      | 2      |
| 2: 山本 | 1     | 2       | 3      |        |
| 3: 伊藤 | 2     |         | 2      | 1      |
| 4: 吉田 | 1     | 3       | 3      | 2      |

吉田はうどんに評価3を与えているのでうどんが好きだといえる

$$\rho_{ax} = \frac{\sum_{y \in \mathcal{Y}_{ax}} (r_{ay} - \bar{r}_a')(r_{xy} - \bar{r}_x')}{\sqrt{\sum_{y \in \mathcal{Y}_{ax}} (r_{ay} - \bar{r}_a')^2} \sqrt{\sum_{y \in \mathcal{Y}_{ax}} (r_{xy} - \bar{r}_x')^2}}$$

Pearson相関でユーザー同士の嗜好の類似度を求める

$$\hat{r}_{ay} = \bar{r}_a + \frac{\sum_{x \in \mathcal{X}_y} \rho_{ax} (r_{xy} - \bar{r}_x')}{\sum_{x \in \mathcal{X}_y} |\rho_{ax}|}$$

評価していない食べ物を予測

予測評価が高い食べ物を優先的に推薦

### 学術データにおけるUBCF

表: 成績値行列

|       | 1: 国語 | 2: 数学 | 3: 化学 | 4: 物理 |
|-------|-------|-------|-------|-------|
| 1: 高橋 | 1     | 3     |       | 3     |
| 2: 山本 |       | 1     | 3     |       |
| 3: 伊藤 | 2     | 1     | 3     | 1     |
| 4: 吉田 | 1     | 3     | 2     |       |

高橋は数学で評価3を取得しているので数学が得意といえる

$$\rho_{ax} = \frac{\sum_{y \in \mathcal{Y}_{ax}} (r_{ay} - \bar{r}_a')(r_{xy} - \bar{r}_x')}{\sqrt{\sum_{y \in \mathcal{Y}_{ax}} (r_{ay} - \bar{r}_a')^2} \sqrt{\sum_{y \in \mathcal{Y}_{ax}} (r_{xy} - \bar{r}_x')^2}}$$

Pearson相関で学生同士の成績の類似度を求める

$$\hat{r}_{ay} = \bar{r}_a + \frac{\sum_{x \in \mathcal{X}_y} \rho_{ax} (r_{xy} - \bar{r}_x')}{\sum_{x \in \mathcal{X}_y} |\rho_{ax}|}$$

取得していない科目を予測

予測成績が高い科目を優先的に推薦

- はじめに
- 学術データ分析と情報推薦
- 授業科目の成績予測と教材の最適化
- 提案手法
- 数値実験並びに考察
- おわりに

## 3.2 シラバスからの教材作成

8/16

学校で配布されているシラバスを活用し、Web 上の授業内容に関連した情報をホームページと YouTube から取得し、学生に提示する。このようにして作成された情報を教材と呼ぶ。

- はじめに
- 教学データ分析  
と情報推薦
- 授業科目の成績  
予測と教材の最  
適化
- 提案手法
- 数値実験並びに  
考察
- おわりに

|            |                      |  |  |
|------------|----------------------|--|--|
| 授業科目名      | IoTシステムデザイン          |  |  |
| 授業科目名 (英語) | Design of IoT System |  |  |
| 科目区分       | 専門                   |  |  |
| 配当学年       | 工学部 3年               |  |  |

|      |         |           |  |
|------|---------|-----------|--|
| 担当教員 |         |           |  |
| 姓    | 氏名      | 所属        |  |
| 主教授  | ◎ 沼本 健朗 | 情報システム工学科 |  |

|                  |  |
|------------------|--|
| 開講学期             | 後期   |
| 単位数              | 2  |
| 単位区分             | 選択   |
| 関連する学習・教育目標      | 電子・情報工学科：(C)-3   |
| 授業の目標・授業概要       | IoTは、モノや建物がインターネットに接続され、その状態が把握できるシステムであり、これからのAI/データ社会の基盤となる技術である。<br>一方で、物に目的、意思、自律性を与える技術(技術)だけを知覚しても技術者としてIoTを社会課題の解決にあずかることはできない。そのため本授業では、社会課題の発見から解決手法の提案、そしてシステムの実装まで、社会課題の発見・解決としてのIoTシステムをデザインする手法について学ぶ。<br>また、そのIoTの基盤となるセンサ技術や通信技術について簡易に、集めたデータをもとに分析する手法について実践的に学ぶ。 |
| 学生の到達目標          | IoTの概要と、IoTを実現する基礎技術を理解する。<br>無線LAN/Wi-Fiなどの無線通信技術について理解する。<br>センサネットワークなどの無線技術の応用について理解する。<br>IoTをどのように地域課題解決に結びつけるか、デザイン思考による問題解決手法を身につける。   |
| 授業計画             | ①IoTを実現するための主要技術の概要を理解する。(3週)<br>②無線LAN技術の基礎を理解する。(2週)<br>③センサネットワーク関連技術の基礎を理解する。(2週)<br>④IoTを構成する技術を理解する。(2週)<br>⑤調査、グループディスカッション等による地域課題の発掘(2週)<br>⑥地域課題を解決するためのIoTシステムデザイン手法を学ぶ(4週)   |
| キーワード            | 無線LAN、アドホックネットワーク、センサネットワーク、IoT  |
| 成績評価基準           | 小テスト (30%)、プレゼンテーション (20%)、レポート(課題) (50%) によって総合評価する。  |
| 教科書・教材参考文献等      | 配布資料を用いて行う。参考書：別途提示する。   |
| 開講科目・履修条件等       | 本講義の受講にあたっては、インターネット工学および無線伝送方式を履修しておくことが望ましい。   |
| 履修上の注意事項や学習上の留意点 | 講義中の説明で理解できなかった時には必ず質問する。<br>この授業科目は、実習経験のある教員による授業科目または主として実習的教育から構成される授業科目です。<br>遠隔授業と実習とを両立して履修した履修を定めた授業   |
| 学生からの質問への対応方法    | 質問は授業中に対応する。また、個別の質問があれば教室・文書どちらでも随時受け付けるが、事前に電子メールで相談内容を連絡して予約して下さい。電子メールアドレス：imuro@toyoma.ac.jp  |

図：Webシラバス

### ①シラバスから授業計画をスクレイピング

- ①IoTを実現するための主要技術の概要を理解する。(3週)
- ②無線LAN技術の基礎を理解する。(2週)
- ③センサネットワーク関連技術の基礎を理解する。(2週)
- ④IoTを構成する技術を理解する。(2週)
- ⑤調査、グループディスカッション等による地域課題の発掘(2週)
- ⑥地域課題を解決するためのIoTシステムデザイン手法を学ぶ(4週)

### ②スクレイピングの結果をGoogleとYouTubeで検索



### ③検索結果の上位を取得し教材に

図：教材作成の流れ

|                            |   |
|----------------------------|---|
| Youtube#title              | Youtubeurl  |
| 【中学英語】40分で学べる中1英語全解説授業     | <a href="https://www.youtube.com/watch?v=eBQ8gcXsl_g">https://www.youtube.com/watch?v=eBQ8gcXsl_g</a>                                   |
| 【英語】中1-3一般動詞(基本編)          | <a href="https://www.youtube.com/watch?v=iDfTdcCbiE">https://www.youtube.com/watch?v=iDfTdcCbiE</a>                                     |
| 【高校英語】文法の基礎を総まとめ           | <a href="https://www.youtube.com/watch?v=Ngn-eQdd_3s">https://www.youtube.com/watch?v=Ngn-eQdd_3s</a>                                   |
| 【99%の人が勘違い】英文法20選【完全イメージ化】 | <a href="https://www.youtube.com/watch?v=Nk-UBi8siZ4">https://www.youtube.com/watch?v=Nk-UBi8siZ4</a>                                   |
| HTitle                     | HUrl  |
| 名詞の前に形容詞を並べる順番のルール【ラングランド】 | <a href="https://www.langland.co.jp/english/column/english-column11">https://www.langland.co.jp/english/column/english-column11</a>     |
| 誰でも英語の文法をマスターできる魔法の3ステップと  | <a href="https://www.kobetsu.co.jp/manabi-vitamin/subject/jhs-english">https://www.kobetsu.co.jp/manabi-vitamin/subject/jhs-english</a> |
| 英語 文法 - 東京外国語大学            | <a href="http://www.coelang.tufs.ac.jp/mt/en/gmof/">http://www.coelang.tufs.ac.jp/mt/en/gmof/</a>                                       |
| 英語の文法を効率的に学ぶための順番と確実に定着させ  | <a href="https://www.alohaenglish.jp/english-grammar/">https://www.alohaenglish.jp/english-grammar/</a>                                 |

図：取得された教材



### 3.3 信頼性を考慮した教材の提示

9/16

レビューの信頼性の指標として類似性, 協調性, 集中性, 情報性の4つの指標を定義している<sup>3</sup>.

- はじめに
- 教学データ分析  
と情報推薦
- 授業科目の成績  
予測と教材の最  
適化
- 提案手法
- 数値実験並びに  
考察
- おわりに

#### 類似性

- 複製されたレビューには多くのスパムが含まれていることがわかっている
- レビューの文章の類似度を図る指標として類似性スコアを定義する。

bigramで区切ったレビュー $l_i$ と $l_j$ の類似度

$$\text{sim}(l_i, l_j) = \frac{|X_{l_i} \cap X_{l_j}|}{|X_{l_i} \cup X_{l_j}|}$$

列について最大値を類似度とする

$$S.\text{score}(l_i) = \max_j \{\text{sim}(l_i, l_j) \mid j \neq i, j = 1, 2, \dots, n\}$$

$n \times n$ の類似度行列を作成

|       | $l_1$                  | $l_2$                  | ... | $l_n$                  |
|-------|------------------------|------------------------|-----|------------------------|
| $l_1$ | 0                      | $\text{sim}(l_1, l_2)$ | ... | $\text{sim}(l_1, l_n)$ |
| $l_2$ | $\text{sim}(l_2, l_1)$ | 0                      | ... | $\text{sim}(l_2, l_n)$ |
| ⋮     |                        |                        |     |                        |
| $l_n$ | $\text{sim}(l_n, l_1)$ | $\text{sim}(l_n, l_2)$ | ... | 0                      |

類似度をスコア0から5の範囲で正規化

$$S.\text{score}_{\text{norm}}(l_i) = 5 \cdot S.\text{score}(l_i)$$

各行について最大値を求める

#### 集中性

- スパムレビューは時間的に集中して投稿される傾向にある
- レビューが集中的に投稿されているかを測る指標として集中性スコアを定義する。

レビューが密集に投稿されている日を求める



レビュー $l_i$ の集中性スコアを算出

$$T.\text{score}(l_i) = \ln(\text{size}(g_i))$$

レビューが密集に投稿されている日についてレビューが密集に投稿されている期間を求める

集中性スコアをスコア0から5の範囲で正規化

$$T.\text{score}_{\text{norm}}(l_i) = \frac{5 \cdot T.\text{score}(l_i)}{\max_j \{T.\text{score}(l_j) \mid j = 1, 2, \dots, N\}}$$

レビューが密集に投稿されている期間を求めその期間に投稿されたレビューの数 $\text{size}(g_i)$ を求める

#### 情報性

- レビュー本文がinformativeであるほどスパムである可能性が低い
- レビューがinformativeであるかを測る指標として情報性スコアを定義する。

レビュー本文から名詞を抽出

今は夏。彼女はそれを思い出す。 → 今、夏、彼女

情報性スコアをスコア0から5の範囲で正規化

$$I.\text{score}_{\text{norm}}(l_i) = 5 \cdot \left( 1 - \frac{I.\text{score}(l_i)}{\max_j \{I.\text{score}(l_j) \mid j = 1, 2, \dots, N\}} \right)$$

抽出した名詞がほかのレビューで使用されている回数をカウント

今 = 4回, 夏 = 3回, 彼女 = 1回の場合  
[今, 夏, 彼女] = [4, 3, 1]

情報性スコアを算出

$$I.\text{score}(l_i) = \ln \left( 1 + \sum_{j=1}^{|R_i|} \ln \left( \frac{\theta}{d(\text{term}_j)} \right) \right)$$

記号

レビュー $l_i$ と同じジャンルに属するレビュー数:  $\theta$

レビュー $l_i$ に出現する名詞集合:  $R_i$

→ [今, 夏, 彼女]

レビュー $l_i$ と同じジャンルのレビュー集合において $\text{term}_j$ が $R_i$ を含んだレビューの数:  $d(\text{term}_j)$

→ [4, 3, 1]

#### 協調性

- サクラはグループを組んでスパムレビューを投稿する傾向にある
- レビューがサクラによって投稿されたかを測る指標として協調性スコアを定義する。

トランザクションA

ID: 0  
ID: 2  
ID: 5

トランザクションC

ID: 3  
ID: 7  
ID: 9

トランザクションB

ID: 0

トランザクションD

ID: 0  
ID: 4  
ID: 10

支持度数 $\text{support}(g_i)$ とユーザーID数 $\text{size}(g_i)$ で協調度を計算

$$\text{collaborate}(g_i) = \text{support}(g_i) \cdot \text{size}(g_i)$$

レビュー $l_i$ の協調性スコアを算出

$$C.\text{score}(l_i) = \begin{cases} \ln(\max_{g_i \in G_{\text{all}}} \{\text{collaborate}(g_i)\}) & |G_{\text{all}}| \neq 0 \\ 0 & |G_{\text{all}}| = 0 \end{cases}$$

協調性スコアをスコア0から5の範囲で正規化

$$C.\text{score}_{\text{norm}}(l_i) = \frac{5 \cdot C.\text{score}(l_i)}{\max_j \{C.\text{score}(l_j) \mid j = 1, 2, \dots, N\}}$$

■ 投稿者グループA

ID: 0 ID: 6  
ID: 3 ID: 10

■ 投稿者グループB

ID: 2 ID: 7  
ID: 1 ID: 8  
ID: 4 ID: 9

■ 投稿者グループC

ID: 3 ID: 4

■ 投稿者グループD

ID: 0 ID: 6  
ID: 3 ID: 10

■ 投稿者グループE

ID: 2 ID: 7  
ID: 1 ID: 8  
ID: 4 ID: 9

■ 投稿者グループF

ID: 3 ID: 4

<sup>3</sup>伊木惇, 亀井清華, 藤田聡, “レビューを対象とした信頼性判断支援システムの提案”, 情報処理学会論文誌, Vol. 55, No. 11, pp. 2461-2475, Nov 2014.

## 4.1 予測評価値からの適切な授業科目の推薦

10/16

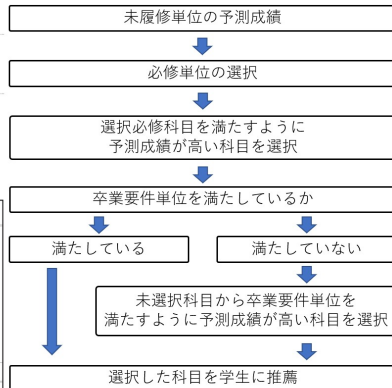
大学には卒業要件単位が存在し卒業要件単位を満たさなければ卒業できない。ここでいう適切とは、「予測成績が高く、卒業要件単位、必修科目、選択必修科目のすべてを満たす」ということである。

|               |   |   |  |  |  |   |   |                     |
|---------------|---|---|--|--|--|---|---|---------------------|
| ○論理回路         | ◇ |   |  |  |  | 半 | 2 | 4 単位以上(※)<br>修得すること |
| ○アルゴリズムとデータ構造 |   | ◇ |  |  |  | 半 | 2 |                     |
| ○インターネット工学    |   | ◇ |  |  |  | 半 | 2 |                     |
| ○情報理論         |   | ◇ |  |  |  | 半 | 2 |                     |
| ○電気回路 1       | ◇ |   |  |  |  | 半 | 2 | 4 単位以上(※)<br>修得すること |
| ○電子回路 1       |   | ◇ |  |  |  | 半 | 2 |                     |
| ○電磁気学 1       |   | ◇ |  |  |  | 半 | 2 |                     |
| ○電子物性         |   | ◇ |  |  |  | 半 | 2 |                     |

図：選択必修単位

| 区 分      | 卒業要件単位  | 卒業研究履修条件単位                | 指定科目履修条件単位  |
|----------|---|---------------------------|---|
| 総 合 科 目  | 人 間 2 単位以上<br>社 会・環 境 6 単位以上<br>言 語・文 化 4 単位以上<br>精 神・身 体 3 単位以上<br>総合科目計 19 単位(注3) | 教養小計 44 単位<br>15 単位以上(注3) | 教養小計 38 単位<br>13 単位以上(注1)<br>6 単位以上(注1)<br>2 単位以上 |
| 基 礎 科 目  | 13 単位   | 13 単位                     | 70 単位   |
| 外国語科目    | 英 語 10 単位<br>第 2 外国語 2 単位   | 10 単位<br>2 単位以上           |   |
| キャリア形成科目 | 7 単位  | 3 単位(注2)                  |   |
| 専門基礎科目   | 卒業研究以外 71 単位  | 69 単位(注2)                 |   |
| 専門共通科目   | 専門小計 79 単位  |                           |   |
| 専門科目     | 卒業研究 8 単位   | —                         | —   |
| 合 計      | 130 単位  | 110 単位                    | 70 単位   |

図：卒業要件単位



図：適切な授業科目の選択

- はじめに
- 教学データ分析  
と情報推薦
- 授業科目の成績  
予測と教材の最  
適化
- 提案手法
- 数値実験並びに  
考察
- おわりに

## 4.2 関連資料のアップデート

11/16

学生はこのような Web 学習システムにおいて互いに協調し合ってスパムレビューを投稿するとは考えにくいといった理由から、学生のレビューに対して類似性、集中性、情報性の 3 つの指標で各レビューのスパム性の判断を行い、信頼性スコアを算出する。

### 類似性

表 4.1: n-gram における文章の区切り

|                |                                    |
|----------------|------------------------------------|
| n = 1: unigram | 今日/は/いい/天/気/で/す/ね/。                |
| n = 2: bigram  | 今日/日は/はいい/いい/天/気/で/す/すね/ね。         |
| n = 3: trigram | 今日は/日はいい/はいい/いい/天/気/天/気/で/す/すね/すね。 |

- ・ Bigramで区切った文字を比較

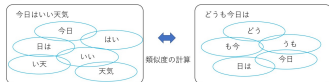


図4.2: bigramによる類似度の計算

### 情報性

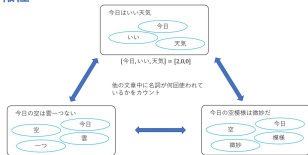


図4.3: 名詞の使用回数からの情報性の計算

### 集中性

Step 1: 日足によりバーストしている日を求める

Step 2: バーストした日について15分足でバーストしている時間帯を求める

Step 3: バーストしている時間帯に投稿されたレビューの投稿数を調べる

Step 4: 投稿数によって集中性を計算

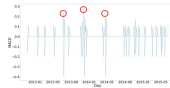


図4.4: 日足によるMACDの計算



図4.5: 15分足によるMACDの計算

### 信頼性

- ・ 指標からスパム性スコアを計算

$$\left. \begin{array}{l} \text{類似性: } S\_score_{norm} \\ \text{情報性: } I\_score_{norm} \\ \text{集中性: } C\_score_{norm} \end{array} \right\} \text{スパム性スコア: } F\_score$$

- ・ スパム性スコアとレビュー評価値から信頼性スコアを計算

$$\frac{\text{教材に対するレビュー評価値の平均}}{\text{レビュー評価値の平均}} = \text{信頼性スコア: } T\_score$$

- はじめに
- 教学データ分析と情報推薦
- 授業科目の成績予測と教材の最適化
- 提案手法
- 数値実験並びに考察
- おわりに

## 4.3 提案手法のアルゴリズム

12/16

1. はじめに
2. 教学データ分析  
と情報推薦
3. 授業科目の成績  
予測と教材の最  
適化
4. 提案手法
5. 数値実験並びに  
考察
6. おわりに

動画

# 5.1 数値実験の概要

13/16

推薦の個人化, 情報推薦の有効性, レビュー信頼性の評価, システム全体の成績向上の有効性においてシステムの検証を行っていく。

- はじめに
- 学術データ分析と情報推薦
- 授業科目の成績予測と教材の最適化
- 提案手法
- 数値実験並びに考察
- おわりに

## 推薦の個人化

表: 取得科目が同一の学生データ

| 学番号     | 教養ゼミ1 | 教養ゼミ2 | 経済学1 | 社会学1 | 環境論1 | 日本語表現法 | 芸術学 | 健康科学演習 |
|---------|-------|-------|------|------|------|--------|-----|--------|
| 1915001 | 2     | 1     |      | 2    | 1    | 4      |     | 4      |
| 1915002 | 1     | 4     |      | 3    | 1    | 1      |     | 2      |

- 取得している科目は同じであるが成績が異なる2人の学生データを使用して推薦を行い推薦結果を比較



## レビュー信頼性の評価

- わざとスパム性が疑われるレビューを投稿しその時の信頼性スコアの確認を行う

表:  $df(term_i)$  と信頼性スコア

| $df(term_i)$ | 1/24 1816 | 1/25 1448 | 1/26 2148  | 1/28 2220  | 1/29 3293  |
|--------------|-----------|-----------|------------|------------|------------|
| 信頼性スコア       | 0.1       | 0.1       | 0.3 2.2    | 0.3 2.2    | 0.1        |
|              | 2.588     | 2.588     | 1.585      | 2.267      | 3.055      |
| 1/29 3/30    | 1/29 3/31 | 1/29 3/34 | 1/29 10/36 | 1/29 16/59 | 1/27 16/12 |
| 0.1          | 0.1       | 0.1       | 0.1        | 0.1        | 0.1        |
| 3.055        | 3.055     | 3.055     | 3.055      | 3.055      | 3.055      |

表: 類似度と類似性スコア

| 類似度       | 1/31 10/45 | 1/25 12/48 | 1/26 21/48 | 1/28 1/26  | 1/29 3/30  |
|-----------|------------|------------|------------|------------|------------|
| 類似性スコア    | 0.108      | 0.108      | 0.108      | 0.108      | 0.1        |
|           | 0.789      | 0.789      | 0.814      | 0.814      | 0.0        |
| 1/29 3/30 | 1/29 3/31  | 1/29 3/34  | 1/29 10/36 | 1/29 16/59 | 1/27 16/12 |
| 1.0       | 1.0        | 1.0        | 1.0        | 0.129      | 0.163      |
| 5.0       | 5.0        | 5.0        | 5.0        | 0.009      | 0.714      |

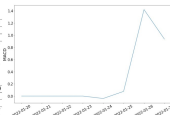


図: MACDヒストグラム

## 情報推薦の有効性

|     | 教養ゼミ1 | 経済学1 | 社会学1 | ... | 電気回路1 | 経済学2 | 社会学2 |
|-----|-------|------|------|-----|-------|------|------|
| 学生A | 3     | 4    | ...  | 2   |       |      |      |
| 学生B | 4     | 4    | 3    | ... | 4     |      |      |
| 学生C | 2     | 4    | 4    | ... |       |      |      |
| ... | ...   | ...  | ...  | ... | ...   | ...  | ...  |
| 学生N | 3     | 4    | 1    | ... | 3     |      |      |

1年後期までのデータを使用

2年前期以降を予測

図: 学生データ

- 実際の学生データを使用し1年後期までの履修科目から2年前期までの科目の予測成績を算出しMAEとCatalogue Coverageで情報推薦の有効性の検証を行う

## システムの有効性

- システム使用前と使用後で英語の文法テストを行う
- システムを使用することで成績が向上することの確認を行う



以下の目文に合うように語句を選びなさい。

例1

目文: バンタシアは、古生代後期の時代に、唯一の大陸でした。

英文: Pangaea ( ) the only continent during the late Paleozoic era.

語句: is are was were

例2

目文: 太陽は、太陽の周りを、公転します。

英文: The Earth ( ) around the Sun.

語句: go goes is going have gone

図: 英語例題

## 5.2 実験結果と考察

14/16

科目の推薦の個人化度合いと 2 つの指標について実験結果の考察を述べる。

表: 推薦科目の被覆率

|      | 2 年前期 | 2 年後期 | 3 年前期 | 3 年後期 |
|------|-------|-------|-------|-------|
| 1 回目 | 10/23 | 3/5   | 11/18 | 5/8   |
| 2 回目 | 9/19  | 9/14  | 7/16  | 1/2   |
| 3 回目 | 8/17  | 5/9   | 3/11  | 7/15  |
| 4 回目 | 5/7   | 13/15 | 13/20 | 10/19 |
| 5 回目 | 13/23 | 11/15 | 9/17  | 1/2   |
| 平均   | 53%   | 68%   | 50%   | 52%   |

- ・被覆率は全体を通して55.75%
- ・卒業要件単位を考慮した推薦結果で被覆率は50%前後
  - ➡ パーソナライゼーションができています

表: MAEとCatalogue Coverage

|      | MAE        | Catalogue Coverage |
|------|------------|--------------------|
| 学生 A | 0.780708   | 0.771739           |
| 学生 B | 0.8632     | 0.858696           |
| 学生 C | 0.687943   | 0.836957           |
| 学生 D | 0.741116   | 0.771739           |
| 平均   | 0.76824175 | 0.80978275         |

- ・ MAEは0.77であり従来の嗜好データにおけるMAEの値と大きな変化はない
  - ➡ 成績データにおいても協調フィルタリングを適用できる
- ・ Catalogue Coverageは0.81となった
  - ➡ 偏った科目ばかりを推薦しているわけではない

- はじめに
- 教学データ分析と情報推薦
- 授業科目の成績予測と教材の最適化
- 提案手法
- 数値実験並びに考察
- おわりに

## 5.2 実験結果と考察

15/16

開発したシステムによりレビューのスパム性の考慮，システムの有効性について実験結果の考察を述べる．

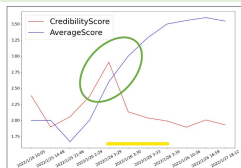


図 5.4: 信頼性スコアと評価値の推移

- ・ 初回のスパムレビュー投稿は信頼性スコア，レビュー評価値ともに上昇
- ・ 2回目以降の投稿では信頼性スコアが減少しているがレビュー評価値は上昇を続けている

➡ レビューのスパム性を考慮できている

### サンプル数5

表 5.8: テスト結果による t 検定

|                  | 使用前                | 使用后  |
|------------------|--------------------|------|
| 平均               | 24.4               | 32.6 |
| 分散               | 117.3              | 35.3 |
| 観測数              | 5                  | 5    |
| ピアソン相関           | 0.780130681        |      |
| 仮説平均との差          | 0                  |      |
| 自由度              | 4                  |      |
| t                | -2.53783628841971  |      |
| $P(T \leq t)$ 片側 | 0.0320640738618854 |      |
| t 境界値 片側         | 2.13184678632665   |      |
| $P(T \leq t)$ 両側 | 0.0641281477237708 |      |
| t 境界値 両側         | 2.77644510519779   |      |

- ・  $\alpha < p$  となり有効性は示せない
- ・ サンプル数の少なさが問題だと考えられる

### サンプル数10

表 5.8: テスト結果による t 検定

|                  | 使用前                | 使用后  |
|------------------|--------------------|------|
| 平均               | 24.4               | 32.6 |
| 分散               | 117.3              | 35.3 |
| 観測数              | 5                  | 5    |
| ピアソン相関           | 0.780130681        |      |
| 仮説平均との差          | 0                  |      |
| 自由度              | 4                  |      |
| t                | -2.53783628841971  |      |
| $P(T \leq t)$ 片側 | 0.0320640738618854 |      |
| t 境界値 片側         | 2.13184678632665   |      |
| $P(T \leq t)$ 両側 | 0.0641281477237708 |      |
| t 境界値 両側         | 2.77644510519779   |      |

- ・ 同一サンプルを使用しサンプル数を2倍に
- ・  $\alpha > p$  となり有効性を示すことができた
- ・ システムを使用することで成績の向上を確認

- はじめに
- 教学データ分析と情報推薦
- 授業科目の成績予測と教材の最適化
- 提案手法
- 数値実験並びに考察
- おわりに

## おわりに

**本研究では、学生が高いGPAを修めることができるよう、情報推薦、教材による学習を受けられるシステムの開発を行った。**

教材につけられるスコアはレビューのスパム性を考慮しており、学生がよい加減につけたレビューや不当に評価を上げようとするレビューを判断し、最終的なレビューのスコアである信頼性スコアに反映させることでより教材のランキングが信頼できるものにした。  
最後に、学生がシステムを使用することで成績が向上することを示した。

## 今後の課題

- ・単純な授業計画による検索では参考になりにくいサイトが多くヒットするときがある。  
➡ 第十五回目「まとめ」など教材の作成を見直す必要がある教材が存在する。
- ・HTML内で上位3件から外れてしまうとなかなか再表示がされない。  
➡ 学生に見せるHTMLの表示の仕方を改善する必要がある。
- ・学生に推薦する科目の決定方法の改善。  
➡ 単純な協調フィルタリングのみでの推薦で良いのかを考える必要がある。