

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. 予測評価値から  
の適切な授業科目  
の推薦
5. 提案手法
6. 数値実験並びに  
考察
7. おわりに

# 卒業生の就職先データを活用したキャリア パス指向の履修科目支援システム

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

山本 藤也 (Touya Yamamoto)  
t815043@st.pu-toyama.ac.jp

富山県立大学 工学部 情報システム工学科

Teams, 13:50-14:05 Wednesday, February 16, 2022.

# 1.1 研究の背景

2/20

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る。

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. 予測評価値からの適切な授業科目の推薦
5. 提案手法
6. 数値実験並びに考察
7. おわりに

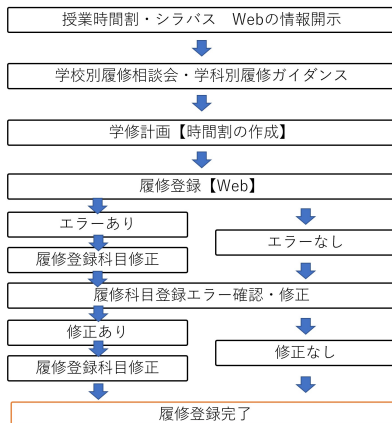


表2：成績からGPへ変換

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

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

図1：履修登録までの流れ

## 1.2 研究の目的

3/20

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。

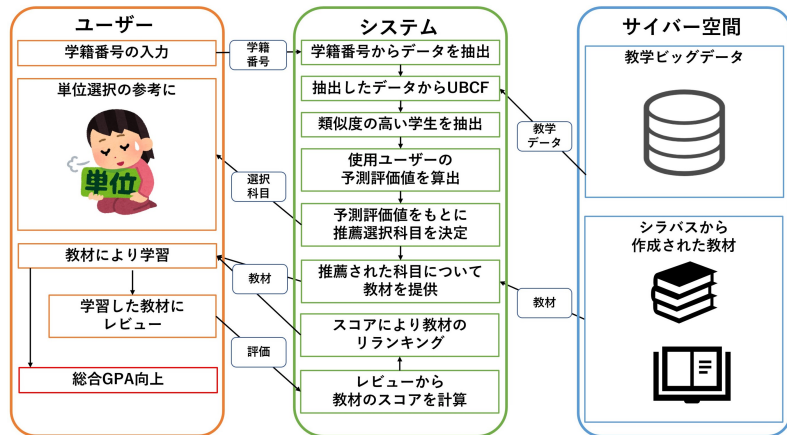


図2：提案システムの流れ

## 2.1 教学におけるビッグデータ・アナリティクス

4/20

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>。

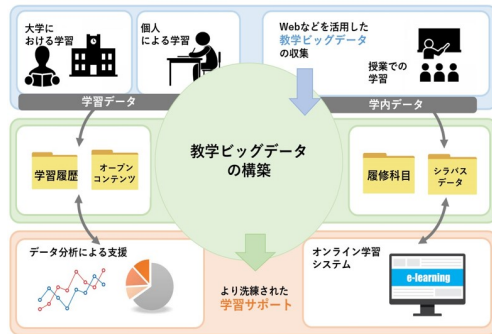


表2：教学データの例

取得時期	教学データ	内容
入学前	出身高校 入試情報 入学前学習	課程差別、判定値、etc. 入試区分、成績 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 授業 学生生活 成績	履修科目 出欠状況、課題提出、etc. 部活、アルバイト、etc. 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

### 6の特徴

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

図3：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

5/20

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>。



図4：情報推薦

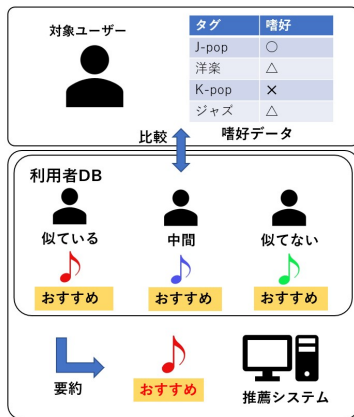


図5：協調フィルタリング

<sup>2</sup>神島敏弘, “推薦システムのアルゴリズム”, <https://www.kamishima.net/archive/recsysdoc.pdf>, 2021.10.28

## 2.3 レビューの信頼性の判断支援

6/20

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6：国内の主なECサイト



図7：ECサイトにおけるレビュー



図9：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

7/20

ユーザーベース協調フィルタリング (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3:評価値行列

	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

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

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

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

Pearson相関でユーザー同士の嗜好の類似度を求める

予測評価が高い食べ物を優先的に推薦

### 教学データにおけるUBCF

表4: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	

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

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

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

Pearson相関で学生同士の成績の類似度を求める

予測成績が高い科目を優先的に推薦

- 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
- 予測評価値からの適切な授業科目の推薦
- 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

8/20

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
- 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
- 予測評価値からの適切な授業科目の推薦
- 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		
職階	氏名	所属
准教授	◎ 根本 健朗	情報システム工学科
関連学期	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、モノや環境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、これからの次世代社会の基盤となる技術である。 一方で、IoT(感知、認識)を駆使するIoT(技術)は社会の発展と技術者としてのIoTを社会課題の解決に生かすことはできない。そのため本授業では、社会課題の発見から解決手段の提案、そしてシステムの設計まで、社会課題の解決手段としてのIoTシステムをデザインする技術について学ぶ。 また、そのIoTの発展となるIoT技術や通信技術について概観し、集めたデータをどのように分析するべきかについて実践的に学ぶ。	
学生の到達目標	IoTの概要と、IoTを支える基礎技術を理解する。 無線LANやWi-Fiなどの無線通信技術について理解する。 センサネットワークなどの情報技術の応用について理解する。 IoTをどのように社会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。(2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見(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)	
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・教材参考書等	教科書資料を用いて行う。参考書：別途指示する。	
関連科目・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。 この授業科目は、実務経験のある教員による授業科目または主として実務経験者から構成される授業科目です。 遠隔授業で授業開始前に開講した授業は通知される。	
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、履修の質問があれば来室・文書でもらっても随時受け付けるが、事前に電子メールで授業内容等を連絡して予約して下さい。電子メールアドレス：kamoto@ipc.toyama.ac.jp	

図10: 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。(2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見(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11: 教材作成の流れ

Youtube title	Youtube url
【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=iDFtdgcCbiE">https://www.youtube.com/watch?v=iDFtdgcCbiE</a>
【高校英語】文法の基礎を総まとめ	<a href="https://www.youtube.com/watch?v=Npn-eQdd_3s">https://www.youtube.com/watch?v=Npn-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>
HP title	HP url
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-column1/">https://www.langland.co.jp/english/column/english-column1/</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	<a href="https://www.kobetsu.co.jp/manabi-vitamin-journal/jhs-english">https://www.kobetsu.co.jp/manabi-vitamin-journal/jhs-english</a>
英語 文法 - 東京外国語大学	<a href="http://www.coelang.tufs.ac.jp/mt/en/gmod/">http://www.coelang.tufs.ac.jp/mt/en/gmod/</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12: 取得された教材



## 3.3 レビュー信頼性の指標

9/20

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
- 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
- 予測評価値からの適切な授業科目の推薦
- 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て
- 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

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

$n \times n$ 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。

レビューが頻繁に投稿されている日を知る



レビュー $l_i$ の集中性スコアを算出

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

レビューが頻繁に投稿されている日についてレビューが頻繁に投稿されている時間を求める

集中性スコアをスコア0から5の範囲で正規化

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

レビューが頻繁に投稿されている時間を求めるその時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

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

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回の時  
[今, 夏, 彼女] = [4, 3, 1]

記号

レビュー $l_i$ と同じジャンルに属するレビュー数:  $o$

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

$\text{term}_j \in R_i$ を含んだレビューの数:  $df(\text{term}_j)$

→ [4, 3, 1]

情報性スコアを算出

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

### 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



トランザクション



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

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

レビュー $l_i$ の協調性スコアを算出

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



投票者グループA

ID: 0 ID: 6 ID: 7

ID: 3 ID: 10 ID: 8

投票者グループB

ID: 0 ID: 4 ID: 9

支持度数 4 IDの数 4

投票者グループB

ID: 1 ID: 7 ID: 8

ID: 2 ID: 4 ID: 9

支持度数 3 IDの数 6

ID: 10 投票者の決定方法

協調性スコアをスコア0から5の範囲で正規化

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

<sup>3</sup> 伊木淳, 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

10/20

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	
○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	
○電子回路1				◇				半	2	
○電磁気学1					◇			半	2	
○電子物性				◇				半	2	

4単位以上(※)  
修得すること

4単位以上(※)  
修得すること

図14: 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
人 間 関 係 学	2単位以上	2単位以上	
社会・環境	6単位以上	6単位以上	
言語・文化	4単位以上	4単位以上	
精神・身体	3単位以上	3単位以上	
総合科目計	19単位(注3)	15単位以上(注3)	
基礎科目	13単位	13単位以上	
外国語科目	10単位	8単位以上(注1)	
第2外国語	2単位	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	専門研究 8単位	—	—
合 計	130単位	110単位	70単位

図15: 卒業要件単位

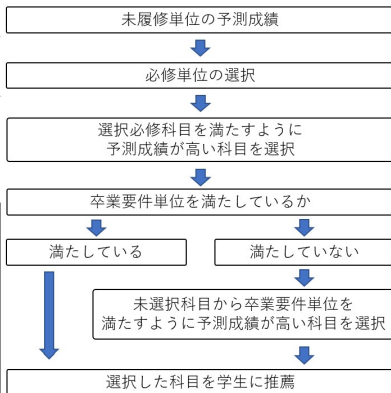


図16: 適切な授業科目の選択

- 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
- 予測評価値から  
の適切な授業科目  
の推薦
- 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

11/20

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5: n-gramによる文字の区切り

n = 1: unigram	今日/日は/いい/いい/天気/で/す/ね/。
n = 2: bigram	今日/日は/いい/いい/いい/天気/で/す/ね/。
n = 3: trigram	今日/日は/いい/いい/いい/いい/天気/で/す/ね/。

・ Bigramで区切った文字を比較



図17: bigramによる類似度の計算

### 集中性

Step 1: 日足によりバーストしている日を求める

Step 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 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 4: 投稿数によって集中性を計算



図19: 日足によるMACDの計算



図20: 15分足によるMACDの計算

### 情報性

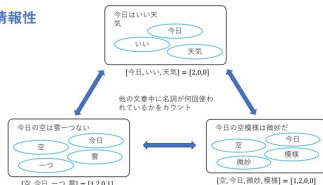


図18: 名詞の使用回数からの情報性の計算

### 信頼性

・指標からスパム性スコアを計算

類似性:  $S\_score_{norm}$   
 情報性:  $I\_score_{norm}$   
 集中性:  $C\_score_{norm}$

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

教材に対するスパム性スコアの平均  
 レビュー評価値の平均 = 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
- 予測評価値からの適切な授業科目の推薦
- 提案手法
- 数値実験並びに考察
- おわりに

## 5.1 ロジスティック回帰とキャリア推薦

12/20

### 目的：どの科目がキャリアに有効かを知ること

過去の卒業生の 成績データ と 就職先データ を用い、特定の業種に進むためにどの科目が重要だったのかを分析する。

そのために、学生ごとに 適性スコア を計算し、合格の可能性を予測するモデルを構築する。

(ここにフロー図などの画像を挿入)

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. 予測評価値から  
の適切な授業科目  
の推薦
5. 提案手法
6. 数値実験並びに  
考察
7. おわりに

### ステップ1：適性スコア $z$ の計算

学生の成績と科目の 重要度 を掛け合わせ 適性スコア  $z$  を計算する。このスコアが高いほど、その業種への適性が高いと判断される。

$$z = (w_1 \cdot x_1) + (w_2 \cdot x_2) + \cdots + (w_p \cdot x_p) + b$$

これをベクトルで書くと次のようになる。

$$z = \mathbf{w}^\top \mathbf{x} + b$$

- $z$ ：ある学生の 適性スコア
- $\mathbf{x}$ ：学生の成績ベクトル ( $x_1$ : 科目 1 の成績, ...)
- $\mathbf{w}$ ：各科目の 重要度 (重み) ベクトル (モデルが学習する値)
- $b$ ：全体のスコアを調整する 基本点 (バイアス)

## 5.2 予測の仕組み (2)

14/20

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. 予測評価値からの適切な授業科目の推薦
5. 提案手法
6. 数値実験並びに考察
7. おわりに

### ステップ2：合格可能性 $P$ への変換

適性スコア  $z$  を 0% から 100% の 合格可能性  $P$  に変換するため、シグモイド関数 という便利な関数を用いる。

$$P(y = 1 \mid \mathbf{x}) = \frac{1}{1 + e^{-z}}$$

- $P(y = 1 \mid \mathbf{x})$ ：成績  $\mathbf{x}$  の学生がターゲット業種に就職する ( $y = 1$ ) 確率
- $e$ ：ネイピア数 (約 2.718)
- $z$ ：前スライドで計算した適性スコア

## 5.3 学習の仕組み (1)

15/20

### 最適な 重要度 $w$ の発見方法

モデルは学習の際、以下の コスト関数  $J$  を最小化するような最適な 重要度  $w$  を探す。

$$J(w, b) = \underbrace{-\frac{1}{m} \sum_{i=1}^m [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)]}_{\text{予測のズレ}} + \underbrace{\alpha \sum_{j=1}^p |w_j|}_{\text{複雑さへのペナルティ}}$$

- 第 1 項 (予測のズレ)：モデルの予測  $p_i$  と正解  $y_i$  が違うほど、この値は大きくなる。この部分は 対数尤度損失 と呼ばれる。
- 第 2 項 (複雑さへのペナルティ)：多くの科目の重要度  $|w_j|$  が大きいほど、この値は大きくなる。

モデルはこの両方の合計が最小になるような、最もバランスの良い 重要度  $w$  を見つけ出す。

## 5.3 学習の仕組み (2)

16/20

### 特別なルール：L1 正則化の効果

コスト関数の第 2 項は L1 正則化項 と呼ばれ、モデルに特別な制約を与える。

$$\text{ペナルティ項} = \alpha \sum_{j=1}^p |w_j|$$

コスト関数全体を最小化する過程で、このペナルティ項を小さくしようとする力が働く。その結果、予測への貢献度が低い科目の重要度  $w_j$  は、積極的に 0 に設定される。

- $\sum_{j=1}^p |w_j|$  : 全科目の重要度（重み）の絶対値の合計。
- $\alpha$  : ペナルティの強さを決めるパラメータ。この値が大きいほど、より多くの科目の重みが 0 になる。

この仕組みのおかげで、本当に意味のある キー科目 だけが選び抜かれ、質の高い推薦リストが作成される。

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. 予測評価値からの適切な授業科目の推薦
5. 提案手法
6. 数値実験並びに考察
7. おわりに



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. 予測評価値から  
の適切な授業科目  
の推薦
5. 提案手法
6. 数値実験並びに  
考察
7. おわりに

## 4.3 ロジスティック回帰と業種の予測

18/20

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. 予測評価値からの適切な授業科目の推薦
5. 提案手法
6. 数値実験並びに考察
7. おわりに

学生の成績や履修情報などの特徴量から、どの業種に就職する可能性が高いかを予測するために、ロジスティック回帰モデルを用いる。確率的に分類するため、以下のような条件付き確率を用いる：

$$P(\text{業種} = y \mid \mathbf{X} = \mathbf{x})$$

各記号の意味は以下のとおりである。

- $P(\text{業種} = y \mid \mathbf{X} = \mathbf{x})$ ：入力  $\mathbf{x}$  に対し、業種が  $y$ （例：IT 業、製造業など）である確率
- 業種：学生の就職先の業界カテゴリ（目的変数）
- $y$ ：業種の 1 カテゴリ（例： $y = \text{IT}$  など）
- $\mathbf{X}$ ：学生に関する説明変数ベクトル（履修科目や GPA など）
- $\mathbf{x}$ ：ある学生の実際の特徴量

## 4.3 ロジスティック回帰と業種の予測

19/20

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. 予測評価値から  
の適切な授業科目  
の推薦
5. 提案手法
6. 数値実験並びに  
考察
7. おわりに

実際のロジスティック回帰モデルでは、次のようなソフトマックス関数を使い、複数業種に対して分類を行う：

$$P(\text{業種} = y \mid \mathbf{x}) = \frac{\exp(\mathbf{w}_y^\top \mathbf{x})}{\sum_{k=1}^K \exp(\mathbf{w}_k^\top \mathbf{x})}$$

- $\mathbf{w}_y$ ：業種  $y$  に対応する重みベクトル（学習されるパラメータ）
- $K$ ：分類対象の業種数
- $\exp$ ：指数関数

## おわりに

**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。**

- ・ 推薦された科目について推薦の有効性を示した。
- ・ 信頼性スコアを反映させることで教材のランキングが信頼できるものにした。

## 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。
- ・ 有効性検証の際に同一サンプルを使用し、有効性を示した。  
➡ サンプル数を増やし、改めて検証を行う必要がある。