

# 教学データからのモチベーション向上・キャリアパス支援のための情報推薦機構

平松 楓也

富山県立大学 情報基盤工学講座

August 5, 2020

はじめに

データ分析の種類

協調フィルタリング

進捗状況

3.1 教学データの構成・内容

3.2 教学データから得られる知見

おわりに

## 背景

近年の就職活動は売り手市場と言われていたが、コロナウイルスの影響で世界的に経済状況が悪化しており、買い手市場に推移していく可能性が考えられる。また、大手企業へ就職を考えた場合、応募人数が多く狭き門であることが多いため、企業は企業がより求めている人材を採用すると思われる。そのため、学生の間にも、より効率的に企業が求める人材になるための勉強が必要になると思われる。

## 目的

過去の卒業生の就職先や、学業成績、野外活動のデータをクラスタリングし、在校生がより効率的に就職活動を行えるよう対話型の情報推薦機構の基礎技術を開発する。

はじめに

データ分析の種類

協調フィルタリ  
ング

進捗状況

3.1 教学データ  
の構成・内容

3.2 教学データ  
から得られる知見

おわりに



入学したばかりの大学1年生、卒業後就職したい企業が決まっている学生  
その企業に就職するためにはどうすれば効率的に動けるかわからない



協調フィルタリングで過去の卒業生の  
データからやるべきことを推薦



やるべき勉強、活動が明確になりモチベーション向上  
就職したかった企業への内定

はじめに

データ分析の種類

協調フィルタリ  
ング

進捗状況

3.1 教学データ  
の構成・内容

3.2 教学データ  
から得られる知見

おわりに

## 説明

事実を説明する  
見つける

## 例

どんな人が何を買っているか？  
ある広告がどれだけ売り上げに貢献しているか？

## 手法

BI、クラスタリング、アソシエーション分析

はじめに

データ分析の種類

協調フィルタリ  
ング

進捗状況

3.1 教学データ  
の構成・内容

3.2 教学データ  
から得られる知見

おわりに

## 説明

未来や欠測値を予想する

## 例

ある商品群を閲覧した人の性別は？  
広告を出稿したらどれだけ売り上げが上がるのか？

## 手法

分類・回帰、統計的機械学習、協調フィルタリング

## 説明

最適解を探す

## 例

利益を最大化するための、最適な仕入れ量は？  
売上を最大化するには、どこに広告を出稿するべきか？

## 手法

最適化、実験結果

はじめに

データ分析の種類

協調フィルタリ  
ング

進捗状況

3.1 教学データ  
の構成・内容

3.2 教学データ  
から得られる知見

おわりに

協調フィルタリングとは、Amazon が開発したレコメンドエンジンで、多くのユーザの嗜好情報を蓄積し、あるユーザと嗜好の類似した他のユーザの情報を用いて自動的に推論を行う方法論である。

また、協調フィルタリングには二種類あり、ユーザベース協調フィルタリングとアイテムベース協調フィルタリングがある。

ユーザベース協調フィルタリングでは「ユーザ A は未評価アイテム I に対して、当該ユーザと似たような嗜好をしている他ユーザと同じような評価をするだろう」という仮定に基づいている。

## ユーザベース協調フィルタリング

履歴から  
類似ユーザ  
を見つける

	商品A	商品B	商品C	商品D
ユーザA	○	-	○	○
ユーザB	×	○	-	×
ユーザC	○	○	×	-
ユーザD	○	×	○	?

類似ユーザAはDを評価高くしているのでおすすめできそう



アイテムベース協調フィルタリングでは「アイテム同士の類似度とあるユーザ A の過去に評価したアイテムの評価点を用いて未評価アイテム I の評価点を予測する」というアプローチである。

## アイテム間協調フィルタリング

似た評価の  
商品を見つける。  
商品Aと商品Dは  
似た人に  
買われやすい

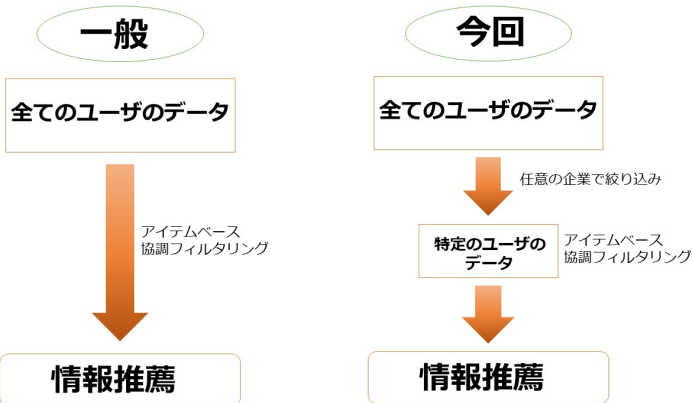
	ユーザA	ユーザB	ユーザC	ユーザD
商品A	○	-	○	○
商品B	×	○	-	×
商品C	○	○	×	-
商品D	○	×	○	?

商品Aの購買者評価と似た評価の商品Dを勧めよう

# 今後行うアイテムベース協調フィルタリングについて

10/26

一般に使われる協調フィルタリングは全ユーザのデータを基にフィルタリングを行うのに対し、今回は、ユーザ A が就職を希望している企業に就職したユーザのみでフィルタリングを行い情報推薦を行おうと考えている。



はじめに

データ分析の種類

協調フィルタリング

進捗状況

3.1 教学データの  
構成・内容

3.2 教学データ  
から得られる知見

おわりに

1 章 はじめに

2 章 QOL の向上のための仕組み（従来研究）

2.1

2.2 モチベーションに繋がるコーチング（やり方、方法）

2.3 情報推薦の仕組み（ガイド、導く）

3 章 教学データの利活用（従来研究）

3.1 教学データの構成・内容（論文 5 本）

3.2 情報推薦の仕組み

3.3 データマイニングにおける評価基準（ジニ不純度とか） 4 章  
提案手法（オリジナリティ）

5 章 数値実験並びに考察

6 章 まとめと今後の課題

はじめに

データ分析の種類

協調フィルタリ  
ング

進捗状況

3.1 教学データ  
の構成・内容

3.2 教学データ  
から得られる知見

おわりに

- ・ csv ファイルを処理，管理しやすいように複数に分けることにした．

はじめに

データ分析の種類

協調フィルタリ  
ング

進捗状況

3.1 教学データ  
の構成・内容

3.2 教学データ  
から得られる知見

おわりに

- ・ アドミッション\_\_出身高校・入試種別.csv
- ・ アドミッション\_\_受験科目・成績.csv
- ・ カリキュラム\_\_履修・評価.csv
- ・ カリキュラム\_\_科目情報.csv
- ・ キャリア関連\_\_インターンシップ.csv
- ・ キャリア関連\_\_就職情報.csv
- ・ キャリア関連\_\_資格・免許.csv
- ・ 課外活動\_\_サークル.csv
- ・ 課外活動\_\_アルバイト.csv
- ・ 卒業後\_\_企業側.csv
- ・ 卒業後\_\_学生.csv

はじめに

データ分析の種類

協調フィルタリ  
ング

進捗状況

3.1 教学データ  
の構成・内容

3.2 教学データ  
から得られる知見

おわりに

	A	B	C	D	E
1	StudentN	出身高校	入試種別		
2	1713001	宇都宮中	前期		
3	1713002	四日市西	前期		
4	1713003	菟道	前期		
5	1713004	片山学園	後期		
6	1713005	恵那	前期		
7	1713006	藤枝明誠	推薦		
8	1713007	氷見	前期		
9	1713008	四日市西	後期		
10	1713009	松阪	前期		

図 1: 出身高校・入試種別

	A	B	C	D	E
1	StudentN	センター	二次試験		
2	1713001	390	277		
3	1713002	389	292		
4	1713003	400	281		
5	1713004	874	0		
6	1713005	389	281		
7	1713006	0	0		
8	1713007	395	254		
9	1713008	885	0		
10	1713009	388	283		

図 2: 受験科目・成績

・出身高校のリストは県大の主な出身校を載せたサイトがあったのでそれを参考にした・点数は平均点を富山県立大学の H29 の平均点を参考に標準偏差 10 でランダムに発生

	A	B	C	D	E	
1	StudentN	経済学 1	富山と日本	環境論 1	環境論 2	E
2	1713001	1	2	1	5	
3	1713002	3	5	2	3	
4	1713003	5	5	5	2	
5	1713004	2	3	4	5	
6	1713005	2	3	5	1	
7	1713006	4	3	4	4	
8	1713007	3		3	5	
9	1713008	2	5	2	2	
10	1713009	5	3	5	4	

図 3: 履修・評価

	A	B	C	D	E
1	StudentN	学部卒業後就職先			
2	1713001	就職	デンソーテクノ		
3	1713002	就職	スギノマシン		
4	1713003	就職	コマツNTC		
5	1713004	就職	富山村田製作所		
6	1713005	進学	YKK AP		
7	1713006	進学	スズキ		
8	1713007	就職	立山科学グループ		
9	1713008	進学	豊田合成		
10	1713009	進学	澁谷工業		

図 4: 就職情報

	A	B	C	D	E	F	G	H
1	StudentN	普通自動車運転免許	TOEIC公開試験	MOS Word	MOS Excel	簿記2級	基本情報技術者	
2	1713001	1	0	0	0	0	1	
3	1713002	1	0	0	0	0	0	
4	1713003	1	0	0	0	0	0	
5	1713004	1	0	0	0	0	0	
6	1713005	0	0	0	0	0	0	
7	1713006	1	0	0	0	0	0	
8	1713007	0	0	0	0	0	0	
9	1713008	1	0	0	0	0	0	
10	1713009	1	0	0	0	0	0	

図 5: 資格・免許

・就職先は富山県立大学の HP を参考にし、学科、学部、院ごとに別ランダム振り分け

・0 が未取得、1 が取得済み

はじめに

データ分析の種類

協調フィルタリング

進捗状況

3.1 教学データの構成・内容

3.2 教学データから得られる知見

おわりに

はじめに

データ分析の種類

協調フィルタリ  
ング

進捗状況

3.1 教学データの  
構成・内容

3.2 教学データ  
から得られる知見

おわりに

	A	B	C	D	E
1	StudentN	サークル			
2	1713001	Chat Box			
3	1713002	無所属			
4	1713003	無所属			
5	1713004	アカペラサークル			
6	1713005	フットサルサークル			
7	1713006	PDC			
8	1713007	軟式野球部			
9	1713008	無所属			
10	1713009	TRPG・映画研究会			

図 6: サークル

	A	B	C	D	E
1	StudentN	アルバイト			
2	1713001	串網			
3	1713002	アルビス			
4	1713003	秋吉			
5	1713004	アルバイトなし			
6	1713005	個別指導塾スタンダード			
7	1713006	串網			
8	1713007	アルバイトなし			
9	1713008	アルバイトなし			
10	1713009	セブンイレブン			

図 7: アルバイト

・サークルも県大に実在するサークルを全て同じ確率でランダムに振り分け、所属率は 50 %

・アルバイト先は県大周辺にあるアルバイト募集している店舗をランダムで振り分け、アルバイトしている確率は 90 %



## 主な教学 I R データ

### ■ 入学時点のデータ

- 入試成績、調査書、入学時アンケート（志望動機、学習歴、本意／不本意、不安など）、入学前教育状況、プレースメントテスト、通学時間、保証人情報など

### ■ 在学中のデータ

- 履修、成績、修得単位数、出席、ゼミ、学籍異動、学習履歴（学習時間や提出物など）、課外活動、在学中アンケート（学生調査、授業アンケートなど）、学習支援／相談利用状況、留学、奨学金状況など

### ■ 卒業時点のデータ

- 進路、就職先、資格・免許取得、最終成績、卒業論文、卒業時アンケート、学習成果ルーブリックなど

図 8: データ構成例 1

はじめに

データ分析の種類

協調フィルタリング

進捗状況

3.1 教学データの  
構成・内容

3.2 教学データ  
から得られる知見

おわりに

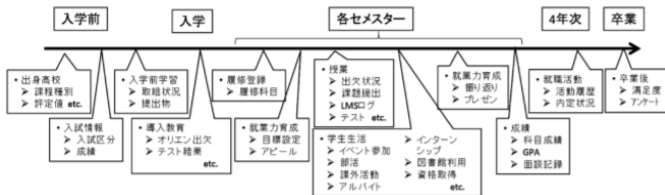


図 9: データ構成例 2

はじめに

データ分析の種類

協調フィルタリ  
ング

進捗状況

3.1 教学データ  
の構成・内容

3.2 教学データ  
から得られる知見

おわりに

測りたいもの		実際に用いる変数	データの種類
従属変数			
	入学後の修学状況	3 年次前期までの単位取得状況	教務データ
独立変数			
入学時の状況	入試形態		入試データ
	本意不本意入学意識・居住形態		入学時アンケート
	ブレースメントテストの結果		A 分野独自データ
	補完補習教育やメンター制度の利用		センター独自データ
	1 年次の修学状況	英語・初修外国語・人文社会科学系教養科目・理系教養科目・自分野の専門基礎教育科目の取得単位数	教務データ

図 10: データ構成例 3

はじめに

データ分析の種類

協調フィルタリング

進捗状況

3.1 教学データの  
構成・内容

3.2 教学データ  
から得られる知見

おわりに

【決定木分析】

【クラスター分析】

【クロス分析】

アンケート結果の集計でよく使われる手法。特定の2つもしくは3つの情報に限定しデータの集計・分析を行う手法

【アソシエーション分析】

マーケティングで利用される代表的なデータ分析手法。顧客が商品を購入する際の購入パターンや売買履歴を分析することで、ある商品Aと商品Bの売れ行きについて、その関連性を抽出する。

【ロジスティクス回帰分析】

ある現象の発生確率を、複数の因子の組み合わせとそれらの程度からモデル化する方法。

## 【先行研究】

- ① 【タイトル】 教学 IR での決定木分析の活用：初年次の学修成果に影響する入学時の学生特徴の探索を例として
- ② 【発行日】 2017 年
- ③ 【分析詳細】 説明変数は「大学での目的」、「入学時不安」、「高校時成績」で目的変数は「GPA」

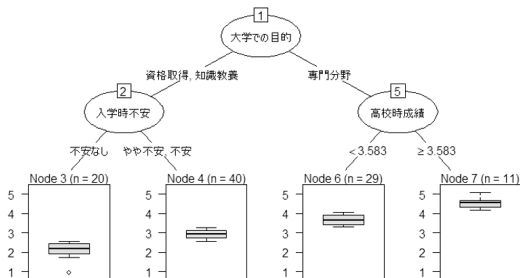


図 11: 決定木分析

## 【先行研究】

- ① 【タイトル】「実践報告」学生調査とeポートフォリオならびに成績情報の分析について—大阪府立大学の教学IR実践から—
- ② 【発行日】2014 年
- ③ 【分析詳細】授業学習・授業外学習（1 週間あたりの活動時間数のうち、授業に出ている時間と授業時間外学習の時間）から3 クラスターの構造であると見なした

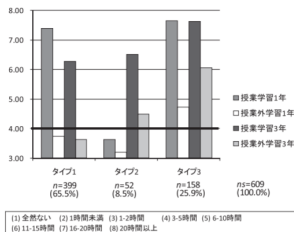


図 12: クラスター分析

## 【先行研究】

- ① 【タイトル】 2018 年度の教学 IR における入試選抜区分による追跡調査
- ② 【発行日】 2018 年
- ③ 【分析詳細】 入試種別と卒業率でクロス分析

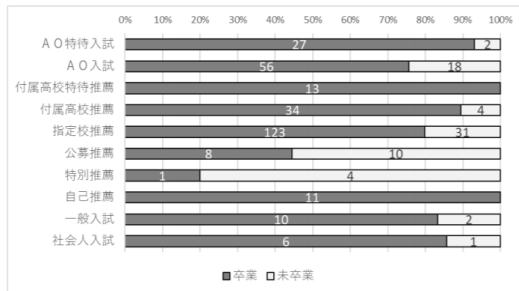


図 13: クロス分析

## 【先行研究】

- ① 【タイトル】 タグマイニングによるテレビ視聴率への影響要因分析
- ② 【発行日】 2020 年
- ③ 【分析詳細】 各テレビドラマの放送回ごとに、取得したツイートをテキストマイニングして、共起タグセット頻度リストを作成する
- ④ 【応用案】 就職活動した企業の関連性、〇〇に就活した人は□□にも就活する傾向がある

表 2 共起タグセット頻度リストの作成例

共起タグ1 (番組名タグ)	共起タグ2 (番組名と共起関係 があるタグ)	共起頻度
#シャーロック	#ディーンフジオカ	9
#同期のサクラ	#高畑充希	99
#モトカレマニア	#新木優子	999



## 【先行研究】

- ① 【タイトル】 体育の授業からみた人間力モデルの構築とその検証——ロジスティック回帰による分析——
- ② 【発行日】 2013 年
- ③ 【分析詳細】 自己充実的達成動機と競争的達成動機の 23 項目のアンケート調査の結果から根子岳登山を選択する学生を予測する
- ④ 【応用案】 進学と就職の発生確率の予測、GPA ○○以上で□□研究室に所属している人は進学しやすい

回帰式とオッズ比							
	偏回帰係数	標準偏回帰係数	標準誤差	ワルド統計量	p値	判定	調整されたオッズ比
質問11前	0.813	0.364	0.40	4.05	0.044 [ < * ]		2.25
質問13前	-0.137	-0.056	0.37	0.13	0.715 [ ]		0.87
質問14前	0.319	0.125	0.41	0.62	0.432 [ ]		1.36
質問16前	-0.953	-0.337	0.44	4.64	0.031 [ < * ]		0.39
質問17前	0.274	0.137	0.30	0.82	0.366 [ ]		1.32
質問18前	0.339	0.175	0.30	1.43	0.231 [ ]		1.43
質問20前	0.823	0.277	0.49	1.61	0.205 [ ]		1.86
質問22前	-0.230	-0.101	0.38	0.37	0.542 [ ]		0.79
質問24前	-0.416	-0.213	0.30	1.98	0.159 [ ]		0.66
質問26前	-0.310	-0.111	0.48	0.46	0.000 [ ]		0.73
質問29前	-0.026	-0.017	0.37	0.01	0.925 [ ]		0.97
質問31前	-0.174	-0.065	0.44	0.15	0.694 [ ]		0.84
質問33前	-0.018	-0.007	0.47	0.00	0.969 [ ]		0.98
定数	0.213		2.66	0.01	0.936 [ ]		

図 15: ロジスティクス回帰分析

## まとめ

- ① 3.1 の先行研究などの例を調べた
- ② 3.2 の先行研究を調べた

はじめに

データ分析の種類

協調フィルタリ  
ング

進捗状況

3.1 教学データ  
の構成・内容

3.2 教学データ  
から得られる知見

おわりに