

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

2.2 教学データ
から得られる知見

おわりに

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

July 21, 2020

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

2.2 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

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

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

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

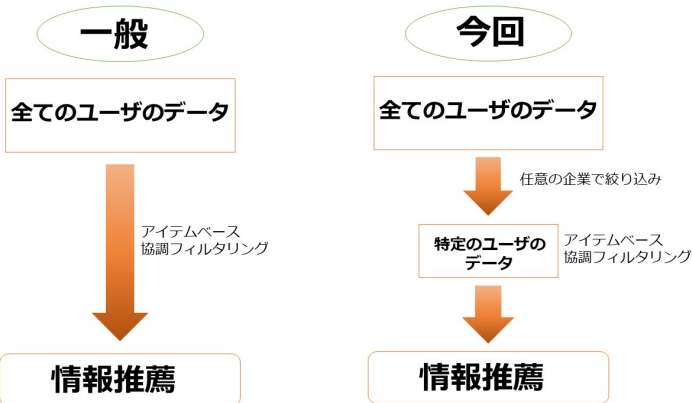
	ユーザA	ユーザB	ユーザC	ユーザD
商品A	○	-	○	○
商品B	×	○	-	×
商品C	○	○	×	-
商品D	○	×	○	?

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

10/21

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

2.2 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 教学データの活用（従来研究）

2.1、教学データの構成・内容（論文 5 本）

2.2、教学データから得られる知見（論文 5 本）

3 章 （従来研究）（数学的な手法）

3.1

3.2 情報推薦の仕組み

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

- ・ 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．

- ・ 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

2.2 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

2.2 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 %

- ・ 前回の決定木は生徒の成績がただの乱数だったため、決定木分析が適切に行われているか判別できなかった。

- ・ 今回は、デモデータを操作しセンター試験の結果がある一定の値より上だった場合進学するようにした。

目的変数は学部卒業後、説明変数は各科目の成績と合計点である。

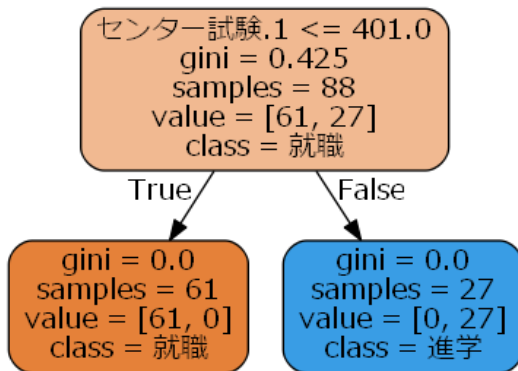


図 8: 決定木分析結果

- ・ NoSQL は Not only SQL の略で，データベースの分類の一つ.

- ・ 聞いたことがあるものでいえば MongoDB がある.

- ・ csv と違って表形式での管理（RDB）じゃなくて，id が割り振られて名前，体重，身長のような情報がくっついてるイメージ

- ・ 決定木分析だったりを行うときは表形式のデータのほうがいいのが難点

【NoSQL と CSV】

- ① NoSQL を使う場合、Python との連携のしやすさを考えると小野田さんも使っていた MongoDB がいいのかもしれない。
- ② 一番管理が楽になるのが科目情報の管理である。

【科目情報の項目】

- ① 学科，必修・選択，履修時期，単位数，科目名の 5 つである。
- ② NoSQL のことを忘れて csv で作業していたため，機械科の部分まで作成済み

学科	共通	共通	共通	共通	共通	共通	共通
必修・選択	必修	必修	選択 1	選択 1	選択 1	選択 1	選択 1
履修時期 1	1年前期	1年後期	1年前期	2年前期	1年前期	2年後期	3年後期
履修時期 2			1年後期	3年後期	1年後期		
単位数		1	1	2		2	2
科目名	教養ゼミ 1	教養ゼミ 2	経済学 1	経済学 2	社会学 1	社会学 2	法学

図 9: 科目情報の csv

まとめ

- ① 決定木分析が適切に動作していることを確認した
- ② クラスタ分析の適切な適用に関してはもう少し考える必要あり
- ③ デモデータの選択科目の部分の処理が大変

今後の課題

- 1 NoSQL(MongoDB) を使うかどうか. また, どう使っていくか考える
- 2 勉強不足の部分が多い