

学生データを用いたキャリア向上 情報推薦機構

滝沢 光介

富山県立大学 電子・情報工学科

June 25, 2021

背景

近年、コロナウイルスの影響から売り手市場で会った就活情勢が買い手市場にシフトしつつある。そこで学生は就職を希望する会社に入社するにはその会社が求める人材になるために効率的に勉強を行う必要がある。

目的

過去の卒業生の就職先や成績などのデータベースからクラスタリングを行い、その学生がどうしたら希望する企業に効率よく就職できるかのフィードバックを行う。また、学生が希望する企業の分析も行い、より細かいフィードバックを行えるようなシステムを作成する。

卒業後、就職したい企業は決まっているが
その企業に就職するために何を学べば効率が良いかわからない

クラスタリング分析を行い、就職のために
学生がやるべき勉強をアドバイス

自分の取り組むべきことがはっきりと理解でき
モチベーションアップに繋がる

図 1: システム完成目標

現在考えている機能について

- 目標とする企業を指定し、過去の学生の成績や行動を分析
- 目標とする企業を指定し、その会社のホームページなどからその会社がどのような分野を重要視しているかを分析



これらの分析結果から学生がどの分野、どの勉強を行えば効率的にその企業へ就職できるかをフィードバック

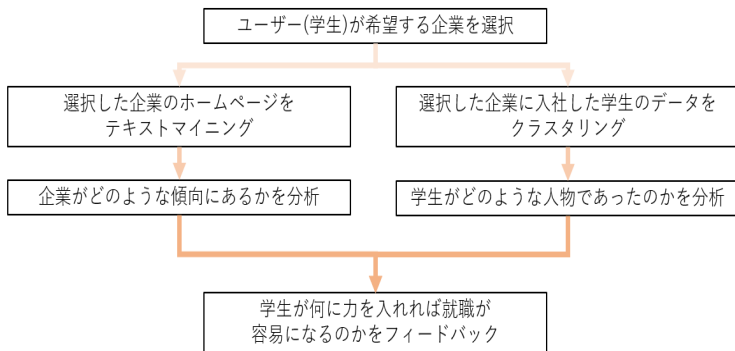


図 2: システム全体の流れ

アドミッション

センター試験の得点を科目ごとに変更

学籍番号	センター国	数学1A	数学2B	物理	化学	生物	英語	センター数	二次得点
1815001	42	85	86	48	72	0	131	464	(
1815002	69	73	69	52	70	0	144	477	290
1815003	65	46	49	52	64	0	133	409	(
1815004	49	55	78	73	41	0	137	433	(
1815005	69	66	66	60	90	0	128	479	(
1815006	64	67	67	59	66	0	136	459	290
1815007	49	56	70	62	0	64	133	434	(
1815008	61	57	66	74	81	0	129	468	(
1815009	58	50	66	66	62	0	154	456	(
1815010	69	53	69	63	53	0	132	439	(

図 3: 入試得点

関係性マイニング

ラフ集合論に基づき, 属性間の相互関係を表す新たな属性を相互関係表にまとめ, 項目間の比較により特徴を抽出する手法.

ユーザ	アイテム					
	A	B	C	D	E	F
ユーザ1	3	5	1	*	5	*
ユーザ2	3	4	*	1	5	5
ユーザ3	4	*	2	4	4	3
ユーザ4	2	4	2	5	4	3
ユーザ5	*	4	4	3	*	1

Table 1: 評価値行列

ユーザ	$A \leq_2 B$	$A \leq_2 E$	$C \leq_2 A$	$C \leq_2 B$	$C \leq_2 E$
ユーザ1	1	1	1	1	1
ユーザ2	0	1	*	*	*
ユーザ3	*	0	1	*	1
ユーザ4	1	1	0	1	1
ユーザ5	*	*	*	0	*

Table 2: 選好パターン

類似度

評価値行列から作成された選好パターン行列を用いて、類似度の高い学生を抽出する。

```

日本国憲法<=社会学 日本国憲法<=総合英語1 日本国憲法<=コンピュータ工学 ... 伝送工学2<=ネットワーク設計論 伝送工
学2<=大規模通信システム工学 伝送工学2<=電磁気学
0      0.0      NaN      0.0 ...      0.0      NaN      0.0
1      NaN      NaN      NaN ...      NaN      0.0      NaN
2      NaN      0.0      0.0 ...      NaN      0.0      0.0
3      1.0      1.0      1.0 ...      1.0      1.0      1.0
4      0.0      0.0      0.0 ...      NaN      0.0      NaN
...
995     NaN      NaN      NaN ...      0.0      0.0      0.0
996     NaN      0.0      NaN ...      0.0      0.0      0.0
997     0.0      0.0      0.0 ...      NaN      0.0      NaN
998     0.0      0.0      NaN ...      0.0      NaN      0.0
999     0.0      0.0      0.0 ...      0.0      0.0      0.0

[1000 rows x 107 columns]
類似度の高い学生は
1915124
1915339
1915681
1915771
1915951

```

図 4: 類似度の高い学生

未完成

類似度の高い学生を抽出した後, 対象ユーザーの未評価のアイテムに対して評価値予測を行い, 値の高い物を推薦するが, 現状評価値予測の部分がまだできていない。

自己組織化マップ (SOM)

ニューラルネットワークの一種で与えられた入力情報の類似度をマップ上での距離で表現するモデル.

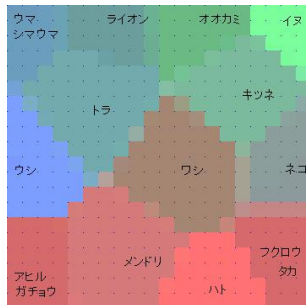


図 5: 自己組織化マップ例

扱うデータ

動物の名前とその動物に関する情報

CaseID	small	medium	large	two_legs	four_legs	body_hair	hoof	mane	win
pigeon	1	0	0	1	0	0	0	0	
fowl	1	0	0	1	0	0	0	0	
duck	1	0	0	1	0	0	0	0	
wild_duck	1	0	0	1	0	0	0	0	
owl	1	0	0	1	0	0	0	0	
hawk	1	0	0	1	0	0	0	0	
eagle	0	1	0	1	0	0	0	0	
fox	0	1	0	0	1	1	0	0	
dog	0	1	0	0	1	1	0	0	
wolf	0	1	0	0	1	1	0	1	
cat	1	0	0	0	1	1	0	0	
tiger	0	0	1	0	1	1	0	0	

図 6: データセット

結果

動物のデータから, 似たような特性を持つ動物がマップ上で近い距離に配置されている.

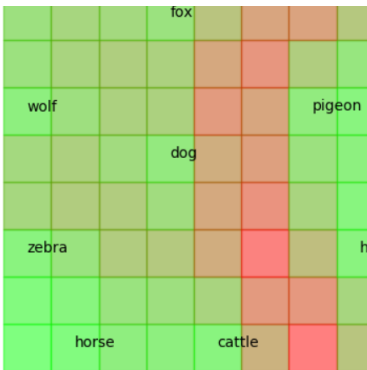


図 7: 実行結果

進捗

- デモデータの修正
- 関係性マイニングを用いた情報推薦の実装 (未完成)
- 簡単な SOM の例の実行

今後の課題

- 未完成の物を完成させる
- SOM をどう使っていくのか考える.SOM の勉強.
- 本研究の有効性の評価をどう行うかを考える