

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

富山県立大学工学部電子・情報工学科
1715059 平松楓也

指導教員：奥原浩之

1 はじめに

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、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。

そこで、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。

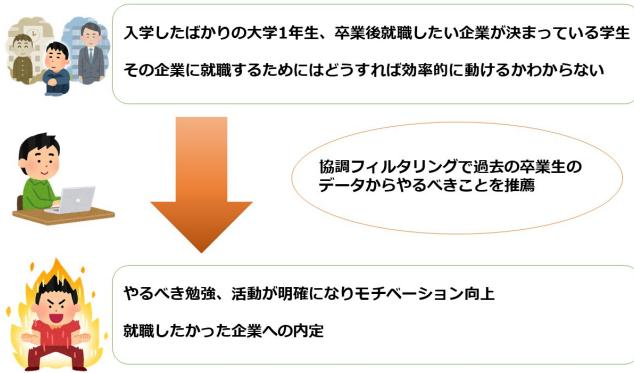


図1 完成目標

2 データ分析の種類

2.1 説明的データ分析

データ分析の中で一番シンプルなものでデータ分析により何か特徴を見つけたり、事実を説明するときに使われる。例えば、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？といったことに使われ、その手法はBI、クラスタリング、アソシエーション分析などが挙げられる [1]。

2.2 予測的データ分析

未来や欠測値の予測に使われる。例えば、株価やドル円の予測を行ったりすることができる。その手法には、分類・回帰、統計的機械学習、協調フィルタリングのなどが挙げられる [2]。

2.3 指示的データ分析

主に最適解を探すことを使われる。例えば、利益を最大化するための、最適な仕入れ量は？などの問題に対し、シミュレーションを行う。また、AI教育の分野では個別最適化を行い教育の効率化なども取り組まれている。手法としては、最適化やシミュレーション実験などが挙げられる [3]。

3 協調フィルタリング

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、多くのユーザーの嗜好情報を蓄積し、あるユーザーと嗜好の類似した他のユーザー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。

3.1 ユーザベース協調フィルタリング

ユーザベース協調フィルタリングでは「ユーザー A は未評価アイテム I に対して、当該ユーザーと似たような嗜好をしている他ユーザー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。つまりユーザー A と似ている（=類似度の高い）ユーザーの未評価アイテム I への評価点を元に

ユーザ A の評価点を予測する、というアプローチになる。

3.2 アイテムベース協調フィルタリング

今回用いる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ー 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」というアプローチになり、この手法の方がよりオフライン処理しやすく、かつ計算速度という面で優れていることからより使われている [1]。

4 今後行うアイテムベース協調フィルタリングについて

一般に使われる協調フィルタリングは全ユーザーのデータを基にフィルタリングを行うのに対し、今回では、ユーザー A が就職を希望している企業に就職したユーザー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。

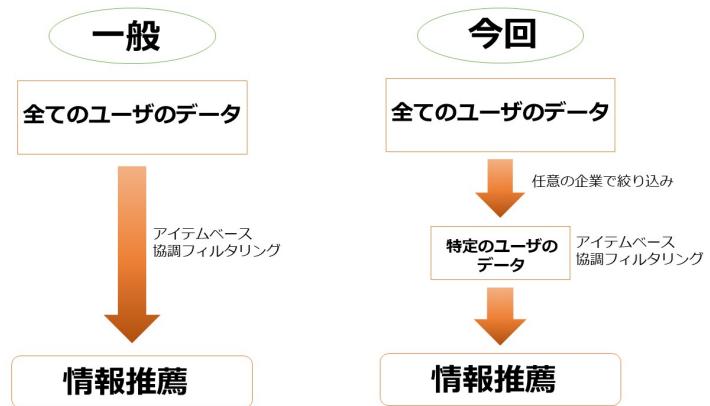


図2 今回の協調フィルタリング

5 進捗状況

卒論の 2.2 にあたる教学データから得られる知見のうちの一つであるクラスター分析に取り掛かった。クラスター分析とは、異なる性質のものが混ざり合った集団から、互いに似た性質を持つものを集め、クラスターを作る方法である。クラスター分析には、大きく分けて 2 種類あり、階層クラスター分析と非階層クラスター分析に分けられる。階層クラスター分析は、結合過程が理解しやすいことがメリットである。非階層クラスター分析は、大量サンプルであってもクラスタの分類が可能である点がメリットである。

今回クラスター分析を行うにあたって python のデータ分析に特化した jupyter notebook というオープンソースで利用できるものを使う。今回はセンター試験と二次試験の各科目的結果から、その人が進学か就職するのか分類したい。

今回のクラスター分析では、クラスター間の距離の計算 (method) は単純平均距離 (average) にした。また、距離計量 (metric) はデフォルトのユークリッド距離を使う。

6 おわりに

本来進学か就職かを分類したかったが、入試種別の分類になってしまっている。この原因是、前期、後期、推薦で点数の差が大きいのが問題だと思うので、正規化するなどして差を縮める必要があったかもしれない。しかし、一応正しく分類出来ていることは確認出来た。今後は、クラスター分析を使って面白い結果が得られそうなることを考える必要がある。卒論 2.2 ではもっと他の手法で得られる知見もやっていきたいが、

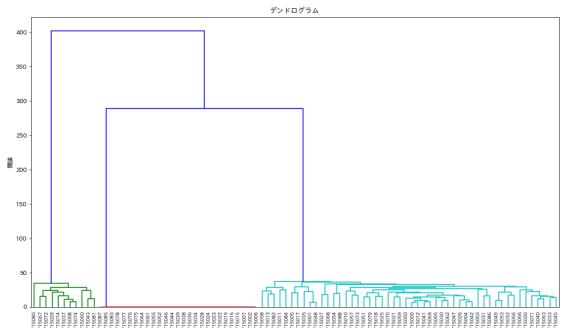


図 3 クラスター分析結果

教学 IR の分野だけで調べてもあまりなかったので、他の分野で自分の研究に応用できそうなものを探していくみたい。また、自分でできるか進めていきたい **参考文献**

- [1] <https://www.slideshare.net/takemikami/ss-76817490>
- [2] <https://www.dhbr.net/articles/-/1578?page=3>
- [3] <https://www.digital-knowledge.co.jp/product/edu-ai/edu-ai-merit/>
- [4] 教学 IR での決定木分析の活用 一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— 関西大学高等教育研究 第 8
号 2017 年 3 月