

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。

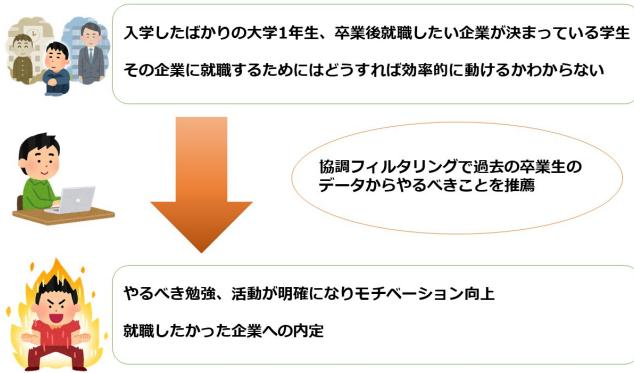


図1 完成目標

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に使われる。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 ユーザーベース協調フィルタリング

ユーザー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。

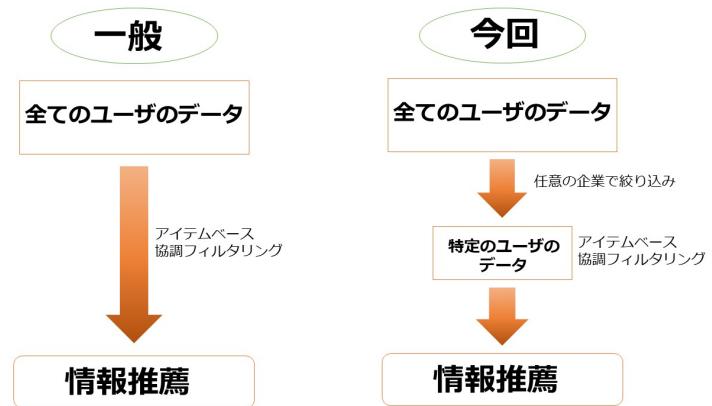


図2 今回の協調フィルタリング

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決定木分析に取り掛かった。決定木分析とは、分類木(雨の日と晴れの日などの条件で分ける)と回帰木(○○円などの変わらう値で分ける)を組み合わせたもので、ツリーによってデータを分析する。決定木分析のメリットは、分析結果の可読性が高い点デメリットは過学習を起こしやすい点である。

今回決定木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。決定木分析を行った教学 IR の先行研究 [4] では、大学での目的、入学時不安、高校での成績から一年次 GPA にどのような関係があるのかを調べていたが、今回は、入試種別、センター試験、二次試験から学部卒業後の進路にどのような関係があるのかを調べる。

今回の決定木では目的変数を学部卒業後、説明変数を入試種別とセンター試験点数、二次試験点数にする。また、木の最大の深さを 4 とした。

6 おわりに

本当は 1 年次 GPA との関係性を決定木分析したかったが GPA の用意が間に合わなかった。また、擬似データが乱数で発生させたものなので分析結果があまり良くなかった。これに関しては実際のデータを説明変数に入れるしかないと思われる。しかし、実際のデータがもらえるのであれば決定木分析の基盤は完成しているので適用は楽に行けそう。また、一度本当にこの決定木分析が正しく行われているかを検証するためにも、進学する人の入試成績が高いようにするなどして、もう一度決定木分析をやってみたほうがいいかもしれない。今後は、卒論 2.2 ではもっ

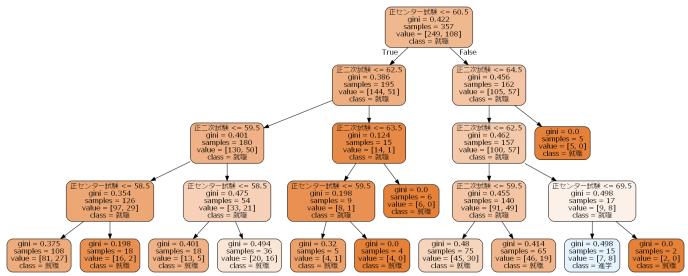


図3 決定木分析結果

と他の手法で得られる知見もやっていきたいので次はクラスター分析などの分析をIRで行っている論文を探し、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