

教学データにおけるクラスタリングを用いた情報推薦における最適化

滝沢 光介

富山県立大学工学部電子・情報工学科情報基盤工学講座

要約

本研究では、大学の単位取得をより効率的に行うために過去の教学データを参照し、分析、それらの結果からその人、学年にあった単位取得の意思決定支援を行う。

1 はじめに

近年、労働者が不足しており就活生にとって有利な状況である売り手市場であった就活情勢がコロナウイルスの影響から就活生に不利な買い手市場にシフトしつつある。そこで、学生は自分の希望する企業に就職するために効率的に大学での勉学を納める必要がある。現状、様々な大学や機関で就職支援と称したシステムが実装されているが、その実態は求人情報や就職情報、学生に対するアンケートの結果をまとめただけのものが多い。これらの機能を不要だとは言えないが、その情報を生徒個人の価値として利用するには生徒が再び情報の取扱いを行わなければならず、生徒が自発的にこの機能を使うとは考えづらい。また、学生の成績の向上を目的としたシステム「M2B」などが存在するがあくまでも授業中の理解度などから学習支援、教育改善を図るためにシステムであり、就職に対する支援としては十分な役割を果たせていない[1]。そこで本研究では、過去の卒業生の就職先や教学データなどのデータ分析を行い、システムを使用する学生がどのように学校での単位を取得すれば効率の良い学習を行えるかをフィードバックすることを目的としたシステムの実装を行う。

2 ユーザー間類似度

2.1 教学IRとラーニングアナリティクス

IRとはInstitutional Researchの略であり、浅野はIRを「ある特定の目的に沿って情報を収集し、それらを加工・統合して分析し、計画立案や意思決定を支援するために展開される活動の総称」と定義しており、それらを教学データに応用させたものを教学IRと呼んでいる[2]。教学IRと類似した研究分野として、ラーニングアナリティクス(LA)が存在するが浜田らは教学IRの目的は「教育機関全体あるいは学部・学科の意思決定支援や質保証に必要な分析結果の提供」であり、LAの目的は「各授業やeラーニングコースの改善」としており両者は目的による違いであるとしている[3]。本研究では、教学IRを広義に解釈し就職活動に有効に活用できるような意思決定システムを提案する。

2.2 教学データとクラスタリング

2.3 情報推薦と協調フィルタリング

3 情報推薦に基づく取得科目の最適化

3.1 大規模不完全AHP

AHPとは階層分析手法と呼ばれ、世の中の事象を目的、評価基準、代替案の3つの階層に整理し、一対比較表を作成し、各代替案における比較対象者の主観的な重要度を算出する手法である。

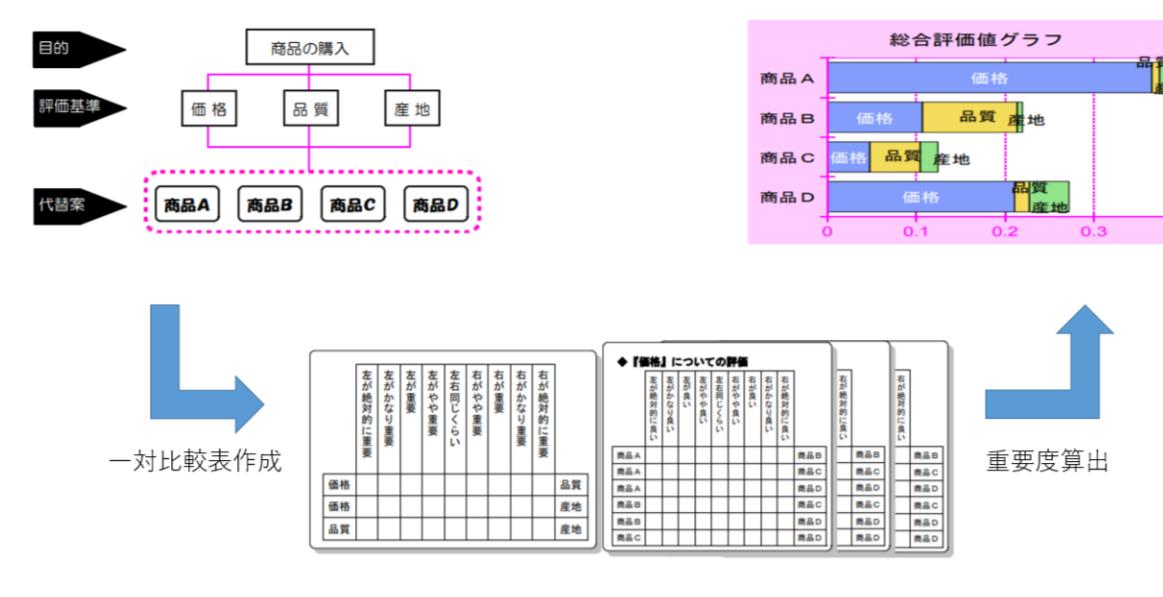


図1 AHPの流れ

大規模とは代替案に対する評価を行う評価対象者が複数人であることを示しており、不完全とはデータの欠損を示している。本来のAHPならば全ての代替案同士についての一対比較を行うが、現実では全ての代替案について一対比較を行うことは難しい。そこで、全ての一対比較が行われていなくとも、AHPのようにそれぞれの代替案について重要度を算出する手法を不完全AHPと呼び、二つの特性を合わせたAHPを大規模不完全AHPと呼ぶ[4]。

不完全なデータから各代替案の重要度を算出するために、「単独の評価者($L = 1$)による全代替案に対する一対比較から得た一対比較値 $x_{ij}^l (i, j = 1, \dots, n)$ が完全整合した場合、それに限り、 $w_i/w_j = x_{ij}^l (i, j = 1, \dots, n)^T$ となる代替案の重要度ベクトル $w = (w_1, \dots, w_n)^T$ が存在する。」という事実から、ある重要度ベクトル $w_i/w_j = x_{ij}^l (i, j = 1, \dots, n)^T$ に対して比較者による一対比較値 $x_{ij}^l (i, j = 1, \dots, n)$ は $w_i/w_j = x_{ij}^l$ であるときが理想と仮定し、(1)式で偶然誤差的に生じていると仮定する。

$$x_{ij}^l = \frac{w_i}{w_j} e_{ij}^l \quad (1)$$

(1)式を両辺について対数変換を行うと(2)式となる。

$$x_{ij}^l = \tilde{w}_i - \tilde{w}_j + \tilde{\epsilon}_{ij}^l \quad (2)$$

各評価者 $l (l = 1, \dots, L)$ が代替案 (i, j) に対して与えた評価 x_{ij}^l から一対比較グラフ G を作成し、それを一対比較ネットワーク N と呼び、一対比較ネットワークの接続行列 A とカットベクトル \tilde{x} を利用することで(2)式は(3)式のように表せることができる。

$$\tilde{x} = A^T \tilde{w} + \tilde{\epsilon} \quad (3)$$

重要度ベクトルを求めるために、誤差最小化問題(4)を考える。

$$\min \|A^T \tilde{w} - \tilde{x}\| \quad (4)$$

$\|\cdot\|$ をユークリッドノルム $\|\cdot\|_2$ とすると問題(4)は最小ノルム問題の特殊形として解くことができ、 I を単位行列とすると[4]の定理3.1より $\min \|A^T \tilde{w} - \tilde{x}\|_2$ の一般解は

$$\tilde{w} = M \tilde{x} + (I - MA^T) y \quad (5)$$

で与えられる。 $\mathbf{1}$ を全ての要素が1であるベクトルとしたとき、[4]の系3.2より $Q = (AA^T + \mathbf{1}\mathbf{1}^T)^{-1}A$ とすると、問題(4)の一般解 \tilde{w} は次の式で表すことができる。

$$\tilde{w} = Q \tilde{x} + \frac{\tilde{a}}{n} \mathbf{1} \quad (6)$$

式(6)で与えられる \tilde{w} を \tilde{a} -重要度ベクトルと呼び、逆対数変換したベクトル w を a -重要度ベクトルと呼ぶ。[4]の系3.3より、全ての評価者が全一对比較を行う場合、 \tilde{a} -重要度ベクトルは(7)式で与えられる。

$$\tilde{w} = \frac{1}{n} \sum_{j=1}^n \left(\frac{1}{L} \sum_{l=1}^L \tilde{x}_{ij}^l \right) + \frac{\tilde{a}}{n} \cdot (i = 1, \dots, n) \quad (7)$$

(7)式の \tilde{w} を対数逆変換すると、

$$\tilde{w} = \left(a \prod_{j=1}^n z_{ij} \right)^{\frac{1}{L}} \cdot (i = 1, \dots, n) \quad (8)$$

となる。ただし、

$$z_{ij} = \left(\prod_{l=1}^L x_{ij}^l \right)^{\frac{1}{L}} \cdot (i, j = 1, \dots, n) \quad (9)$$

である。(8)式において、 $w^T \mathbf{1} = 1$ となるように a を決定すれば、幾何平均法となる。重複評価がなく、一对比較されていない代替案が存在する場合は、不完全AHPであり、

3.2 多目的最適化

3.3

3.4

4 提案手法

4.1 前提データ

4.2 先行研究によるブルウェイップ効果の算定と所見

5 実験並びに考察

参考文献

- [1] "M2B(みつば)学習支援システム", Learning Analytics Center, Kyusyu University, Japan
- [2] 浅野茂, "データベースの構築とIRの課題", 高等教育研究 第19集(2016)
- [3] 松田岳士, 渡辺雄貴, "教学IR, ラーニング・アナリティクス, 教育工学", 日本教育工学会論文誌 41(3), 199-208, 2017
- [4] 八巻直一, 関谷和之, "複数の評価者を想定した大規模AHPの提案と人事評価への適用", Journal of the Operations Research Society of Japan Vol. 42, No.4, December 1999