

機械学習を用いた進路意思決定支援 レコメンドシステムの開発と評価

小川 哲司

Development and Evaluation of Career Decision Support Recommendation System Using Machine Learning

OGAWA, Tetsuji

名古屋経済大学経営学部准教授

要旨：学生の進路指導の現場において、データに基づく就職支援活動が取り組まれているとは言い難い。本研究の目的は、大学生の就職活動において、教育ビッグデータを活用して、効率的にエントリー企業を決定支援するシステムを開発して評価をすることである。成績データから導出した学生間の類似度よりレコメンドした企業は、エントリー候補企業に繋がるという考え方にに基づき、利用者ベースの協調フィルタリングを用いたレコメンドシステムを開発して、評価を行った。本研究の成果として、成績データから導出される学生間の類似性に基づくレコメンドシステムは、学生の進路意思決定支援に繋がることが明らかになった。

キーワード：機械学習、協調フィルタリング、レコメンドシステム、教育ビッグデータ

1. はじめに

近年、インターネットの普及によって、Web や SNS のようなデジタルメディアが登場した。これによって、消費者の購買行動に変化が起きて、企業のマーケティング活動においても大きな変化がもたらされている。Web や SNS のようなデジタルメディアでは、サーバー上でアクセス数や購買履歴などがログとして取得できるため、データに基づくマーケティングが台頭するようになった。このようなデジタルメディアを通じて取得したデータに基づくマーケティングのことを、一般的にはデジタルマーケティングと呼ぶ。

デジタルマーケティングでは、デジタルメディアから取得できる消費者の属性や購買履歴などのビックデータを、AI(Artificial Intelligence: 人工知能) によって分析することで、消費者ごとの嗜好などをパーソナライズすることが可能となった。パーソナライズとは、消費者一人一人に最適化された情報を提示するもので、パーソナライズを活用するものとしては、レコメンドが代表的な手法となる。

レコメンドとは、消費者の閲覧履歴や購入履歴などビックデータの分析に基づいて、購入する可能性が高い商品などを勧めるものである。特にデジタルメディアでは、

リアル店舗のように接客などの営業活動ができないため、このような仕組みを利用して、消費者の購入頻度や購入単価を高める取り組みが必要となる。

例えば、米 Amazon 社では、購買履歴などをビックデータとして収集して、AI で分析することで、消費者ごとにパーソナライズ化された購入予想モデルを作る。このモデルを通じて、精度の高いレコメンドを実施して、売り上げを拡大している。このように、デジタルマーケティングにおいては、ビックデータを活用したレコメンドが必要不可欠な手法となっている。

一方で、大学教育の質保証のあり方は、転換期を迎えている。大学には、今まで以上に教育の内部質保証が求められるようになった。内部質保証とは、「高等教育機関が、自らの責任で自学の諸活動について点検・評価を行い、その結果をもとに改革・改善に努め、これによって、その質を自ら保証すること」(大学改革支援・学位授与機構, 2016) であり、組織的に継続的な自己点検評価を実施し、改善を図ることである。さらに、2018 年より 18 歳以下の人口が減少期に入る「2018 年問題」が到来した中で、今後多くの私立大学では厳しい経営状況になることが予想されている。ゆえに、継続的な大学教育

の質の向上は喫緊の課題となっている。

このような背景から近年、大学における IR (Institution Research) が注目されている。大学における IR とは、大学内のデータを活用して、様々な成果や課題を可視化して、経営改善、教育改善、情報提供などを行う取り組みのことである。特に、教育改善を図ろうとする取り組みのことを「教学 IR」と呼び、学生の学習履歴や行動履歴などのいわゆる教育ビッグデータが分析対象となる。

就職活動に取り組む大学生の中には、進路探索行動が不足して、志望する業界や企業を決定できない学生が一定数存在する。このような場合、学生の指導教員は、学生の性格や実力などと共に、各種データから学生の特性や嗜好などを多面的に把握した上で、関心を持ちそうな業界や職種などを助言して、視野や選択肢を広げる役割が求められる。しかしながら、学生の進路指導の現場において、データに基づく就職支援活動が取り組まれているとは言い難い。これは、データに基づき学生一人一人に丁寧な助言をしようとする、膨大な時間が掛かり、負担が大きいためである。また、学生の嗜好が多様化してきている中で、業界や企業とのマッチングを行うのは困難になってきていることも一因として挙げられる。

そこで本研究の目的は、大学生の就職活動において、教育ビッグデータを活用して、効率的にエントリー企業を決定支援するシステムを開発して、評価することである。学生にとって関心が高い就職活動において、意思決定の支援をすることができれば、学生が認知していなかった企業のリストアップに繋げることができる。これにより、学生の検討の幅が広がり、高い満足度をもたらし、大学教育の質保証にも繋げることが期待できる。

2. 先行研究とリサーチクエスション

2.1 就職活動に関する先行研究

大学生の就職活動に関する研究は、これまで様々な観点から行われてきている。労働政策研究・研修機構は、未就職卒業生における課題として、「エントリーシートが書けない」、「何をしたらいいかわからない」、「就職活動をスタートするのが遅い」、「こだわりが強い」が多いとしている。さらに、学生が会社・仕事選択の際に重視する基準として、「大学での専門分野との関連」は 51.3%、「OB・OG の有無・定着度の高さ」は 33.6% になるとしている (労働政策研究・研修機構, 2010)。この研究より、就職活動にて課題を持つ学生に対して、大学での専門分野との関連性や OB・OG の有無情報を提供することができれば、課題解決のきっかけになると考えることができる。

若松は、進路未決定者は進路決定者よりも、進路探索行動が少ないことを明らかにしている。また、進路意思決定の遅延者の多くは、「どのように考えていったらよ

いか分からない」という初歩的な段階で留まり、進路探索行動をあまり行わずに、直感的に決めているとしている (若松, 2012)。さらに松田らは、就職活動不安が、「アピール不安」、「サポート不安」、「活動継続不安」、「試験不安」、「準備不足不安」から構成されることを明らかにしている (松田ら, 2010)。これらの研究より、進路意思決定の遅延者に対しては、具体的な会社名を提示しながら、進路探索行動を促すようなサポートが望ましいことが示唆される。

教育ビッグデータと就職活動に注目する研究として、赤木らは過去 3,000 件の内定報告書を、ニューラルネットワークによる自然言語処理を実施することで、「面接でされた質問」を可視化して就職支援に役立てようとした (赤木ら, 2017)。西川らは、学生が作成したエントリーシートを、再帰型ニューラルネットワークである LSTM(Long Short Term Memory)にて分析することで、採用可能企業を推定しようとした (西川ら, 2020)。これらの研究は、大学内に存在するデータを教育ビッグデータとして捉え、機械学習を用いて分析する手法の有効性を示すものである。

成績データを活用する研究として、安部は学生の成績や欠席データをナイーブベイズなどの機械学習手法で分析することで、留年・退学や成績不良者の早期発見に繋げることができるとしている (安部, 2019)。大河内らは、数学プレースメントテストの結果より、学生を 5 つに分類することで、未進級者の早期発見に繋げることができるとしている (大河内, 2016)。さらに、成績データと就職に関する研究として、李らは学生の GPA (Grade Point Average: 評定平均) をクラスター分析して、学生を 4 つのクラスに分類することで、就職先との関連性に言及しようとした (李ら, 2019)。半田によると、GPA は 4 年間における学生の行動特性を示すものであるため、就職の際の推薦基準になるとしている (半田, 2008)。

さらに平沢は、就職活動で求められる力と大学で要領よく成績を上げる能力は通底しており、学業成績と就職活動の成功には正の関係性があるとしている (平沢, 2010)。荻谷らも、学業成績が良いと、企業の内定確率は高いとしている (荻谷ら, 2006)。これらの研究より、成績データは学生の特性を示すもので、就職活動成果の説明変数になり得ることを示唆している。

2.2 レコメンドに関する先行研究

レコメンドに関する研究として、Resnick らは GroupLens の実験を通じて、対象ユーザーの嗜好と類似した行動を取った別ユーザーとの類似度に基づいて、商品を提示する協調フィルタリングの手法を自動化した (Resnick, 1994)。類似度とは、嗜好パターンがどれくらい似ているかを定量化したものであり、計算方法として、対象ユーザーと別ユーザーとのベクトルの距離を

求める。この類似度に基づいて、対象ユーザーが好む商品などを予測して、レコメンドを行った。協調フィルタリングは、簡易的な処理で説明変数に欠損値が大量に存在していても高い精度の処理ができるため、その後多くの改良研究が行われている。例えば、ユーザーの嗜好パターンに合わせてレコメンドするだけでなく、レコメンドの目新しさや多様性を加えることで顧客満足度を高める研究などが存在する。

Ziegler は多様性を向上させるために、予測精度を低下させても、より多様な商品をレコメンドするアルゴリズムを提案している(Ziegler, 2005)。Swearingen は、新製品などの意外性があるレコメンドをすると、利用者の満足度を向上させるとしている(Swearingen, 2001)。これらの研究は、ユーザー間の類似度に基づきレコメンドを行う、協調フィルタリングの有効性を示すものである。

2.3 リサーチクエスト

上記の先行研究のレビューより、進路意思決定が遅延している学生には、大学での専門分野との関連性やOB・OGの有無情報に基づき、エントリー候補企業をレコメンドするサポート活動が有効であると考えられる。また学生が履修した科目情報や成績情報が含まれる成績データは学生の特性を示すものであるため、成績データより学生間の類似度を計測できると考えることができる。これらより、成績データに基づく学生間の類似度によって、企業をレコメンドするシステムを構築すれば、学生のサポートに繋げることが期待できる。

以上より、本研究のリサーチクエストは、成績データに基づく進路意思決定を支援するレコメンドシステムの精度はどのくらいか、とする。これを明らかにすることで、教育ビッグデータを活用して、効率的にエントリー企業を決定支援するシステムを構築する足掛かりとする。

3. レコメンドシステム

3.1 レコメンドシステムの概要

レコメンドとは、日本語の「推薦」という意味で、今やデジタルマーケティングにおいて必須の機能である。消費者側のメリットとして、自分で商品などを探さなくても、自分の嗜好にあった商品などを見つけやすくなる点が挙げられる。企業側のメリットとしては、顧客のデジタルメディア上での満足度を向上させることでリピーターを増やして、購入頻度や購入単価を上げることができる点が挙げられる。

神寫によると、レコメンドシステムが求められる背景は2つあるとされている(神寫, 2016)。1つ目は、情報化技術の進展により、個人・団体が容易かつ低コストで発信できるようになったためである。2つ目として、

これら大量の情報の蓄積や流通が容易になり、誰もが大量の情報を得ることができるようになったことである。つまり、計算機の記憶媒体の大規模化や、通信の高速化に起因して、大量に発信された情報を、だれもが大量に取得できる状況が生じたが、自身が必要な情報を取捨選択できないという状況が生じたため、有用な情報を効率よく見つけ出すシステムが求められるようになったのである。

レコメンドシステムは、データ入力、嗜好の予測、レコメンドの提示の3つの段階で成り立っており、O-I-Pモデル(Output-Input-Process model)とも呼ばれる(konstan, 2003)。レコメンド処理の流れは、関心や好みを数値化したデータを入力して、機械学習手法などを利用して類似度を計測して、類似度が近い商品などをレコメンドするものが一般的である。

3.2 レコメンドアルゴリズム

レコメンドを実現するアルゴリズムは、大きく分けると「内容ベースフィルタリング」と「協調フィルタリング」の2種類がある(神寫, 2008)。

内容ベースフィルタリングとは、商品の特性を特徴ベクトルとして利用して、利用者の嗜好パターンと合うものをレコメンドするものである。特徴ベクトルとは、商品の様々な特性を集約してベクトルの形にしたものである。つまり、利用者が過去に高評価した商品と、似た特徴を持つ商品をレコメンドするのである。例えば、車を購入しようとしている消費者に対して、色、デザイン、性能といった車の特徴ベクトルと、過去の消費者の嗜好データを比較して、それらが最も一致するものをレコメンドする。

協調フィルタリングとは、商品の特性は考慮せず、対象利用者の嗜好パターンが類似している他の利用者を見つけ、その利用者が好むものを対象利用者にレコメンドするものである。これは、対象利用者の嗜好パターンと類似している利用者は、同じ商品群を好むという考えに基づいている。例えば、ユーザーAとユーザーBの嗜好パターンが類似しているとして、ユーザーAが購入して高い評価であった商品をユーザーBにレコメンドすると、購入確率が上がるのである。

さらに協調フィルタリングは、メモリベースとモデルベースに分類することができる。メモリベースは、利用者や商品の類似性に注目して、その類似性からレコメンドをする手法であり、データベースを直接利用することが可能である。メモリベースにはさらに、利用者ベースと商品ベースが存在する。利用者ベースは、対象利用者の嗜好パターンが類似する他の利用者をデータベースから見つけて、その利用者が高い評価を付けるものをレコメンドする。商品ベースは、対象利用者の評価が高いものと類似するものをデータベースより見つけて、レコ

メンドするものである。

一方、モデルベースはデータベースから事前にモデルを構築して、利用者の嗜好パターンを推定してレコメンドするものである。

3.3 k近傍法

データベース内にて、対象者と類似する利用者を特定するには、k近傍法(K Nearest Neighbor: 以下kNN)を用いることができる。kNNは、機械学習における教師あり学習のアルゴリズムの一つである。古典的かつ単純なアルゴリズムで、分類と回帰の両方に用いることができる。

分類の場合、分類したいデータから学習データまでの距離が近傍する順にk個のデータを取得し、多数決でデータが属するクラスを推定することができる。kは距離が近いデータを何個参照するかという、パラメータである。

例えば、図1のように緑色の丸をクラス分類したいデータとして、パラメータが3(k=3)の場合、学習データまでの距離が最も近いデータを3つピックアップして、多数決でクラスを予測する。この場合、青い丸が2つ、赤い丸が1つであるため、青い丸と同じクラスに分類される。また、パラメータが5(k=5)の場合、赤い丸のクラスに分類されることになる。

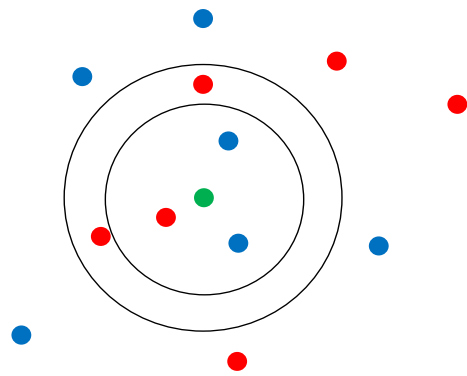


図1 kNNによる分類

回帰の場合、未知のデータを与えた際、特徴空間において近傍するデータをk個抽出することができる。このようにkNN回帰を利用すると、特徴ベクトルによって表現される特徴空間において、対象データに最も近傍するデータを見つけ出すことが可能となる。

kNNでは、近傍するデータを抽出する指標として、特徴空間上での距離を用いる。距離を算出する距離関数は、ユークリッド距離を用いることが多い。ユークリッド距離とは、特徴空間における2点間の距離のことであり、(1)式のようにピタゴラスの公式によって与えられる。

$$d(p, q) = \sqrt{(q_1 - p_1)^2 + \dots + (q_n - p_n)^2} = \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \quad (1)$$

しかしながら、特徴量が多く、高次元のデータを扱う場合、ユークリッド距離を用いると信頼性が低くなる傾向がある。そこで、高次元データを扱う場合、コサイン類似度を用いることが多い。

コサイン類似度とは、特徴空間においてn次元ベクトルの類似度を、それらがなす角度に基づいて計測するものである。これは、特徴空間におけるベクトル同士の成す角度の近さを表現するため、ユーザー間の角とみなすことができ、2本のベクトルがどれくらい同じ向きを向いているのかを表す指標になる。よって、コサイン類似度の値が1に近いほど、強い類似度を示し、2本のベクトルは同じ向きに近いことになる。

コサイン類似度を用いたaとbの類似度は、(2)式で定義される。 $|\vec{a}|$ はベクトルのユークリッド距離であり、「 \cdot 」はベクトルの内積を示す。

$$\cos(\vec{a}, \vec{b}) = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| |\vec{b}|} = \frac{a_1 b_1 + \dots + a_n b_n}{\sqrt{a_1^2 + \dots + a_n^2} \sqrt{b_1^2 + \dots + b_n^2}} \quad (2)$$

3.4 レコメンドシステムの評価方法

レコメンドシステムの評価とは、レコメンドされたものが被験者の情報要求にどれだけ適合しているかを示すものである。レコメンドシステムの評価方法には、オフライン評価とオンライン評価の2つがある。

オフライン評価は、学習データとテストデータに分割して、学習データからレコメンドのモデルを構築して、テストデータに適用して評価するものである。オンライン評価は、実際に運用されているシステムにレコメンドを適用して評価するものである。

オフライン評価の指標には、MAE: 予測値と実測値の差の絶対値を平均したものや、RMSE: 予測値と実測値の差を二乗した値の平均値の平方根を求めるものなどが用いられる。さらにPrecision(適合率): レコメンドされたもののうち、ユーザーの嗜好に合致した割合の指標や、Recall(再現率): ユーザーが実際に嗜好した商品のうち、レコメンドされた割合を示す指標も存在する。

オンライン評価の指標には、CTR(クリック率): レコメンドしたアイテムが実際にどの程度クリックされたかを数値化するものや、CVR(コンバージョン率): レコメンドしたアイテムが、実際にどの程度購入されたかを数値化するものがある。

4. 研究方法

4.1 提案システム

本研究では、成績データから対象学生の特性と類似する卒業生を導出して、その卒業生が就職した企業をエントリー候補企業としてレコメンドするシステムを提案する。本システムは、履修する科目の傾向が似ていて、各科目の成績も同等レベルであれば、学生の特性が類似するため、エントリー候補企業が似たものになるという考え方に基づいている。

このようなエントリー候補企業をレコメンドするシステム（以下、企業レコメンドシステム）を構築するた

めに、利用者ベースの協調フィルタリングを導入することとする。アルゴリズムとして、kNN 回帰アルゴリズムを用いて、距離関数は、コサイン類似度を用いることとする。

企業レコメンドシステムの処理の流れは、図2の通り、①卒業生の成績データと就職データを機械学習の学習用データとして入力する、②企業レコメンドモデルを構築する、③特徴空間上にて対象学生との距離を計測して類似度を算出する、④類似度が高い卒業生を抽出する、⑤抽出された卒業生の就職先をエントリー候補企業としてレコメンドする、である。

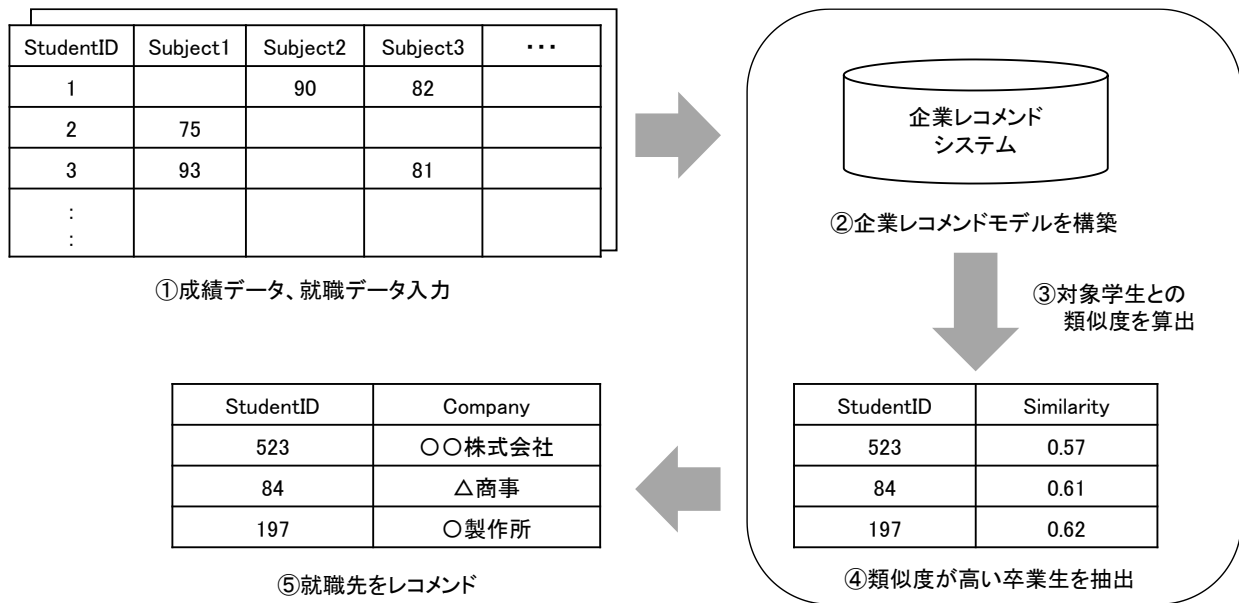


図2 企業レコメンドシステムの処理

4.2 データセット

本研究では、当大学経営学部の2014年度から2018年度の5年間に入学した学生の成績データと、2014年度から2017年度に入学した卒業生の就職データを用いた。成績データのレコード数として、学生数が644名、対象科目数は114科目である。成績データには、学生別の履修科目情報と成績情報が含まれており、学生が履修した科目にて0~100点の範囲で評価されている。履修科目情報には、必修科目と選択科目が含まれる。退学等した学生は除き、在籍している学生と卒業した学生のみを対象とした。対象とした科目は、対象期間を通じて開講していた科目である。就職データには、卒業生の就職先情報が含まれている。

データの取得対象期間を5年間にした理由は、この期間は履修科目の改編が比較的少なく、対象期間内で最も多くの成績データが取得できるためである。成績情報は、

全体的に高い評価を付けがちな教員と、低い評価を付けがちな教員が存在するため、科目ごとに標準化処理を行う。これにより、各科目の成績評価の偏りを解消することができるため、各科目の評価を同等に扱うことが可能となる。

4.3 評価方法

企業レコメンドシステムの精度評価は、適合率(Precision)を用いることとする。適合率は、レコメンドされたもののうち、対象者の情報要求に合致した割合である。レコメンドされたものの中で対象者の情報要求に合った数をTP、合わなかった数をTFとすると、式(3)のように表される。

$$Precision = \frac{TP}{TP + TF} \quad (3)$$

4.4 実験方法

4.1 章で提案した企業レコメンドシステムを実装して、精度を評価するために実験を行った。

実験は、就職活動に取り組んでいる大学4年生11名に対して、2021年4月16日～23日の期間で実施した。被験者の選定方法は、進路探索活動が少なく、まだエントリーする業界や企業が具体的に決まっていな大学生で、すでにエントリーする業界や企業が決まっている学生は除外した。

被験者に対して、企業レコメンドシステムよりレコメンドされた10企業分のリストを提示して、エントリー候補企業として「興味がある」、「興味がない」という質問の回答を求めた。この結果より、適合率を求めて、被験者の情報要求にどれだけ適合しているかを評価す

る。

さらに企業レコメンドシステムの精度を評価するために、就職データよりランダムでリストアップした10企業分のリストを作成して、同一被験者に同様のアンケート調査を実施して回答を得た。

5. 結果

実験の被験者に企業レコメンドシステムとランダムにそれぞれレコメンドしたエントリー候補企業について、適合率を表2に示す。被験者にレコメンドした企業のうち、「興味がある」と回答した割合を適合率として算出した。企業レコメンドシステムを用いてレコメンドした企業の平均適合率は38.2%、ランダムでレコメンドした企業の平均適合率は27.3%であった。

表2 実験結果

被験者		1	2	3	4	5	6	7	8	9	10	11	平均
適合率(%)	企業レコメンドシステム	60	40	20	40	20	20	20	70	20	40	70	38.2
	ランダム	40	30	20	40	30	20	20	20	30	20	30	27.3

次に、この結果に対してt検定を行い、企業レコメンドシステムとランダムでレコメンドした場合の平均適合率に、有意差があるかを検定する。t検定は、2つの結果に差があるかどうかを統計的に示すものである。ここでは、片側検定、有意水準5%と

する。

t検定の結果を表3に示す。企業レコメンドシステムと、ランダムにレコメンドした条件間に有意な差が得られた ($t(10)=1.83, p=0.048$)。

表3 t検定の結果

	企業レコメンドシステム	ランダム
平均	38.18181818	27.27272727
分散	416.3636364	61.81818182
観測数	11	11
ピアソン相関	0.277657165	
仮説平均との差異	0	
自由度	10	
t	1.834253522	
P(T<=t) 片側	0.048251008	
t境界値 片側	1.812461123	
P(T<=t) 両側	0.096502016	
t境界値 両側	2.228138852	

6. 結論

本研究では、進路探索行動が少なく進路意思決定が遅延している学生に対して、成績データから導出

した卒業生との類似度よりレコメンドした企業は、エントリー候補企業に繋がるのではという問題意識から、利用者ベースの協調フィルタリングを用いたレコメンドシステムを構築して、評価を行ってきた。

本研究の成果は、以下の2点に集約される。

第1に、企業レコメンドシステムを用いることで、ランダムでレコメンドする場合より、エントリー候補企業の適合率が高まることが明らかになった点である。利用者ベースの協調フィルタリング手法を用いて構築した企業レコメンドシステムでは、レコメンドの平均適合率が38.2%であり、ランダムでレコメンドした場合より、有意に高いことが示された。これより、本研究で構築した企業レコメンドシステムを通じてレコメンドした企業は、学生のエントリー企業の決定支援に繋がることを示された。

第2に、学生の特性が成績データから定量的に導出されることが明らかなった点である。本研究では、成績データより作られる特徴空間において、対象学生と類似度が近傍する卒業生を効率的に見出すことができ、エントリー候補企業のレコメンドに活用することができた。これより、成績データから導出される学生の特性は、エントリー候補企業の決定支援のみならず、他の用途にも活用できる可能性が示唆される。

本研究の課題として、本研究で提案した企業レコメンドシステムの精度はあまり高くないため、今後精度の改善が求められる。本研究では、成績データ

のみから学生間の類似度を計測したが、より精度の高いレコメンドシステムを目指すために、成績データ以外のデータを活用することが今後の改善点として挙げられる。今後は、学生の行動履歴など他の教育ビッグデータを活用して、適切な特徴量を追加していきたい。

次に、成績データへの重み付けすることが挙げられる。本研究では、各科目の成績データをそのまま利用したため、科目の特性を考慮できていない。例えば、共通科目より専門科目の方が、より学生の特性を表すものとして、点数に重み付けすることなどが考えられる。よって、特徴量の特性に応じて、重み付けを行うことを検討していきたい。

さらに、エントリー候補企業として、意外性や多様性がある企業をレコメンドすることが挙げられる。意外性があるものをレコメンドすると、利用者の満足度を向上させることができるため(Swearingen, 2001)、意図的に意外性や多様性のある企業をレコメンドするアルゴリズムを加えることを検討したい。

以上のような研究課題を踏まえて、学生が就職活動にて、効率的にエントリーする企業を決定支援するシステムを開発する研究を深化させていきたい。

参考文献

- [1] Konstan, J. A., and Riedl, J., *Recommender systems: Collaborating in commerce and communities*, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2003.
- [2] Swearingen, K., and Sinha, R., *Beyond Algorithms: An HCI Perspective on Recommender Systems*, ACM SIGIR Workshop on Recommender Systems, 2001.
- [3] Resnick, P., Iacovou, N., Suchak, M., Bergstorm, P., Riedl, J., *GroupLens: An Open Architecture for Collaborative Filtering of Netnews*, Proceeding of Computer Supported Cooperative Work, pp.175-186, 1994.
- [4] Ziegler, C. N., Mcnee, S. M., Konstan, J. A., and Lausen, G., *Improving Recommendation Lists Through Topic Diversification*, Proceedings of World Wide Web Conference, pp.22-32, 2005.
- [5] 赤木里騎, 徐海燕 「就職支援ツールのデータ分析と可視化」, 平成29年度電気・情報関係学会九州支部連合大会(第70回連合大会)講演論文集, 2017.
- [6] 安部恵介「成績不良の早期予測における機械学習手法の有効性の検証」, 第46回計測自動制御学会知能システムシンポジウム論文集, 2019.
- [7] 大河内佳浩, 山中明生 「プレースメントテストや高校の履修状況などのデータを用いた初年時成績不振者の早期発見」, 日本教育工学会論文誌 40(1), pp45-55, 2016.
- [8] 神畷 敏弘 「推薦システムのアルゴリズム (2)」, 人工知能学会誌 vol.23 No.1, pp.89-103, 2008.
- [9] 荻谷剛彦, 平沢和司, 本田由紀, 中村高康, 小山治 「大学から職業へⅢその1」, 東京大学大学院教育学研究科紀要 第46巻, pp43-74, 2006.
- [10] 大学改革支援・学位授与機構 「高等教育に関する質保証関係用語集 第4版」, 大学評価・学位授与機構, 2016.
- [11] 西川浩平, 松村敦, 宇陀則彦 「エントリーシートを用いた機械学習による採用企業推定」, 情報処理学会第82回全国大会講演論文集, pp405-406, 2020.
- [12] 半田智久, 「機能するGPAとは何か」, 静岡大学教育研究 Vol4, 2008.
- [13] 平沢和司, 「確率抽出データを用いた大卒就職機会に関する検討」, 東京大学社会科学研究所, 2010.
- [14] 松田侑子, 永作稔, 新井邦二郎 「大学生の就職活動不安が就職活動に及ぼす影響 -コーピングに注目して-」, 心理学研究 Vol80 No6, pp.512-519, 2010.

- [15] 李雅文, 能上慎也 「学生の GPA の推移、得意科目、及び就職先分野に関するクラスタリング分析」, FIT2019, pp365-366, 2019.
- [16] 労働政策研究・研修機構 「高校・大学における未就職卒業生支援に関する調査」, 労働政策研究・研修機構, 2010.
- [17] 若松養亮 「大学生におけるキャリア選択の遅延 –そのメカニズムと支援–」, 風間書房, 2012.