

# 就職ポータルサイトにおける企業のアピールポイントと学生の志望理由のマッチング分析モデルに関する一考察

坂元 哲平<sup>1</sup> 山下 遥<sup>2,a)</sup> 萩原 大陸<sup>3</sup> 後藤 正幸<sup>1,b)</sup>

受付日 2017年1月17日, 採録日 2017年6月6日

**概要:** 近年, 企業の採用活動や学生の就職活動において, 就職ポータルサイトが多く活用されている. サイト上で, 企業は自社の広報活動を行うことで学生からのエントリーを募ることができ, 学生は多種多様な企業を志望することが可能である. つまり, 採用・就職活動は企業と学生のある種のマッチング活動ととらえられ, 就職ポータルサイトがそのマッチングをサポートする機能を提供していると考えられる. 一方, 企業が自社の強みとして掲載するアピールポイントと, 学生が感じるその企業の魅力 (企業に対する志望理由) は必ずしもマッチしていない可能性を指摘することができる. そのようなマッチングは, 選考段階や採用後の企業と学生の相性不良につながる恐れがあるため, 何らかの解決策が望まれる. そのため, 企業のアピールポイントに対して学生がどのような理由で企業を志望するかを分析する必要がある. そこで本研究では, 企業のアピールポイントと学生の志望理由の関係性を分析するためのモデルを提案する. 両者の関係性の背景には, 各企業特有の情報やアピールの仕方, また, 各学生の嗜好には互いに異なるグループが混在していると考えられるため, 潜在クラスモデルを用いて関係モデルを構築する. また, 対象データ構造がベクトルである場合に対応するため, 学習アルゴリズムを改良する. 最後に, 実際に大手就職ポータルサイト A のデータに提案モデルを適用し, 有効性を示す.

**キーワード:** 就職活動, 採用活動, ビジネスアナリティクス, 潜在クラスモデル, EM アルゴリズム

## A Latent Class Model to Analyze the Relationship Between Companies' Appeal Points and Students' Reasons for Application

TEPPEI SAKAMOTO<sup>1</sup> HARUKA YAMASHITA<sup>2,a)</sup> TAIRIKU OGIHARA<sup>3</sup> MASAYUKI GOTO<sup>1,b)</sup>

Received: January 17, 2017, Accepted: June 6, 2017

**Abstract:** Recently, many Japanese university students use Internet portal sites that help them in their job-hunting activities. Companies are able to use these sites for public relations activities in order to gather applications from students, and students are able to get information about a wide variety of companies and apply to companies that they want to work for. For companies, in the part of public relation activities, they post their appeal points such as welfare and technological strength; however, they do not always match with students' reasons for their application. This matching relationship between companies and students may lead to incompatibility in the screening stage and its solution is desired. In this study, we propose a model to analyze their matching relationship. We introduce a latent class model because it is considered that information specific to each company, how company appeals, and different preferences for each student are mixed in the background of the relationship. Also, we improve a parameter estimation algorithm in order to adopt cases which the target data structure is a co-occurrence of vectors. Finally, we demonstrate the analysis based on the proposed model by applying to a practical data.

**Keywords:** job-hunting, recruit activity, business analytics, latent class model, EM algorithm

<sup>1</sup> 早稲田大学  
Waseda University, Shinjuku, Tokyo 169-0072, Japan

<sup>2</sup> 上智大学  
Sophia University, Chiyoda, Tokyo 102-0094, Japan

<sup>3</sup> 株式会社リクルートキャリア  
Recruit Career Co. Ltd., Chiyoda, Tokyo 100-6640, Japan  
a) h-yamashita-1g8@sophia.ac.jp  
b) masagoto@waseda.jp

## 1. はじめに

近年、企業の採用活動や学生の就職活動を支援するためのツールとして、就職ポータルサイトが活用されている。就職ポータルサイトを通して、企業は自社の採用活動の広報を行い、学生からのエントリを募ることができ、学生は多種多様な企業を志望すること<sup>\*1</sup>が可能である。企業は広報活動の一環として、就職ポータルサイト内の個社ページで自社の基本情報、採用情報、アピールポイント等を掲載することで、就職活動中の学生に対して広く企業情報を提供する。また、学生は得られた企業情報をもとに、企業に魅力を感じ、それが理由となってその企業へと志望する場合があると考えられる。このように、採用活動・就職活動は企業と学生のある種のマッチング活動ととらえることができ、就職ポータルサイトがそのマッチングをサポートする機能を提供していると考えられる。

就職ポータルサイトの登場により、サイト上での学生の行動履歴や企業の採用活動についてのデータが蓄積されるようになり、様々な分析を行うことが可能となった。これは、社会学的な立場からの定性的な議論や、学生へのインタビュー調査等のアプローチが主流であった既存の就職活動に関する研究に対し、新たなアプローチとなっている。たとえば、学生の就職活動終了日予測を学生の属性情報や行動情報から行う研究 [1], [2] や、サイト内での企業の採用活動と被エントリ数の関係の研究 [3] 等があり、サイトに蓄積されたデータの有効活用の可能性が示されている。

一方、企業が自社の強みとして掲載するアピールポイントと、学生が感じるその企業の魅力（企業に対する志望理由）は必ずしもマッチしていない可能性がある。たとえば、企業が技術力をアピールポイントとしているにもかかわらず、学生は給与水準を志望理由としている場合や、そもそもアピールポイントが学生の志望に影響を与えていない場合である。このようなマッチングは、選考段階や採用後の企業と学生の相性不良につながることも考えられるため、何らかの解決策が望まれている。そのため、企業のアピールポイントに対して学生がどう魅力を感じ、企業を志望するかを分析し、その関係性を把握することが必要がある。

これまで、このような関係性の分析は企業ごとに学生の志望理由とその企業のアピールポイントの一致度を見るというような分析しかできていなかった。しかし、就職ポータルサイトにおける各企業へのエントリ数は一部の人気企業に集中しており、多くの中小規模の企業に対しては、1社ごとに別々に集計すると単純にデータ数が不足してしまう可能性がある。ゆえに、それらの一致度が統計的に意味をなさないような数値が多くなってしまふという問題点がある。このような場合に対して、企業ごとに個別に集計す

るのではなく、類似性の高い企業をまとめて集計することにより、信頼性の高い推定値を得ることができると考えられる。ここでさらに、そのまとめ方について一般に様々な方法が考えられるが、分析者の恣意的な層別に依存してしまうという課題がある。この点で、データ間の統計的な関係の類似性によって自動的にまとめ方を学習するような手法の提案が望まれる。また、就職ポータルサイト運営会社から見ても、各企業というミクロな観点以上に、より全体的な観点で「どのような企業クラスがアピールポイントと志望理由がマッチしている（いない）のか」を明らかにすることで、サイト設計や企業・学生へのサポート面で役立てることが可能と考えられる。つまり、マッチングの関係による企業のクラスタリング結果から新たな知見を見出すことができる可能性がある。

ここで、本研究で対象とする就職ポータルサイト A（以下、就活サイト A）では、学生が企業を志望する際に、その志望理由を登録できるシステムが提供されている。また、企業が掲載するアピールポイントについても情報が蓄積されている。そのため、企業側のアピールポイントに対し学生が志望した理由をひも付けることが可能であり、その共起関係をもとに両者のマッチングの分析が可能である。ここで、両者の関係性の背後には、各企業特有の情報やアピールの仕方、学生ごとに異なる嗜好が混在していると考えられる。このような学生の嗜好の異質性を表現できるモデルとして、潜在クラスモデル [4], [5], [6] が広く知られている。

潜在クラスモデルは、観測データの背後に観測できない潜在的な変数を仮定したモデルであり、異質のデータが混在していると仮定できる対象の分析が可能である手法である。本研究においても、潜在クラスモデルの導入により、マッチングの傾向をとらえることが可能となることが予想される。潜在クラスモデルには、データがある潜在クラスより同時生起すると仮定する混合ユニグラムモデル [7] や、複数の潜在クラスから生起することを許容する Aspect Model [8] 等があり、分析対象によって様々なモデルが提案されている。

そこで本研究では、企業のアピールポイントと学生の志望理由の関係性を分析するための潜在クラスモデルを提案する。具体的には、学生は1つの企業に対して複数の志望理由を選択することができ、また、企業は複数のアピールポイントを掲載しているため、ベクトルどうしの共起を潜在クラスによってモデル化する。一般に、潜在クラスモデルのパラメータ学習は、観測できない変数が存在するが、EM アルゴリズム [9], [10], [11] を用いて探索的に求めることができる。その際にベクトルの次元の差異によって、パラメータの学習が一方のベクトルの特徴を重視して行われてしまうという問題が発生する。そこで本研究では、学習アルゴリズムの改良を提案することにより、この問題を軽減する。

<sup>\*1</sup> 本研究ではいわゆる「エントリ」だけでなく、「企業に興味を持つこと」等を含めて広く「企業を志望する」事象として扱う。

最後に、提案モデルを用いて、就活サイト A に蓄積されたデータを分析し、提案モデルの有効性を示すとともに、結果の分析により新たな知見を得る。

## 2. 準備

### 2.1 就職活動と就職ポータルサイト

日本における就職活動は、諸外国とは大きく異なり、大学生の卒業時期に合わせた新卒一括採用という慣行に従っていっせに行われる。企業は Web 上での企業情報・採用情報の掲載や、説明会の開催、その他採用活動によって学生に自社を広く周知し、それに対して学生は働きたい企業に対して入社試験の申込み（エントリー）を行う。ここで、本研究で対象とする就職ポータルサイトとは、上述の企業の採用活動と学生の就職活動を Web 上でサポートするサービスであり、近年では多くの企業と学生に利用されている。

就職ポータルサイトを利用するためには、企業はサイト運営会社へ登録の申請をすることが必要である。また、学生もサイトにアカウント登録することが必要であり、登録段階では大学等の基本情報を登録する。さらに、企業が説明会やインターンシップを実施した情報や、学生が企業にエントリーした情報といった行動情報も随時蓄積されていく。そのため、サイト運営会社には多様なデータが膨大に蓄積され、この膨大なデータをビジネスに活かすことが期待されている。

### 2.2 関連研究

これまで、就職活動を対象とした研究は幅広く存在する。しかし、それらの多くは就職活動・採用活動の在り方や支援方法についての研究等、社会科学的な立場からのアプローチを取っており、定性的な議論やマクロ統計データを用いた分析、もしくは比較的小人数を対象としたアンケート調査に基づいている。以下では、それらの先行研究についてその事例を紹介する。

就職活動に関する分析について、就職先が決まるまでのプロセスを学生の活動量や情報源の違い等から分析する研究 [12] がある。アンケートをもとに分析を行う中で、就職ポータルサイトが学生にとって役に立つという結果が示され、就職ポータルサイトを利用していない学生は 5% に満たないと報告がある。ただし、就職ポータルサイトを利用しているか否かのみを焦点としており、蓄積されたデータについては触れられていない。ほかにも、学生の就職活動への意識調査に関する研究 [13] や、学生の就職活動での情報源と情報内容に関する研究 [14]、就職活動中の学生のストレスとそのサポートに関する研究 [15] 等、就職活動に関する研究が多角的に行われている。

採用活動に関する分析についても研究が行われている。採用人数や採用方法を雇用動向調査やインタビューから分

析する研究 [16], [17] では、新卒採用を重視する中で、企業と学生のミスマッチを軽減する 1 つの方策としてインターンシップに注目している。また、採用時点でのミスマッチを軽減する採用方針に関する研究 [18] では、インタビューをもとにインターンシップがミスマッチを軽減することに有効であると結論づけている。このように、採用活動に関する研究ではミスマッチを軽減することを 1 つの課題として設定し、様々なアプローチで解決を試みていることが分かる。しかし、Web 上での採用活動や掲載情報についての議論はさかんではない。

近年、就職ポータルサイトに蓄積されたデータを用いた研究も行われている。先に紹介した研究 [1], [2], [3] のほかに、文献 [19] では、エントリー数に注目して学生と企業を潜在クラスを用いてクラスタリングし、エントリー数を増加させるクラスをシミュレーションにより発見している。

以上のように、就職活動に関する研究は広く展開されているが、企業と学生のマッチング傾向を数理的にモデル化した研究はなく、本研究は新規性があるといえる。

### 2.3 対象事例

本研究では企業のアピールポイントに対する学生の志望理由の関係をモデル化する。両本研究で対象とする就活サイト A で両データがどのように扱われるかを説明し、事前分析について述べる。

#### 2.3.1 企業のアピールポイント

一般に、就職ポータルサイトには、個社ページと呼ばれる各企業のページが用意されている。各企業は個社ページ内で自社の基本的な情報（業種、所在地、沿革等）、採用情報（採用予定人数、選考スケジュール等）、そして自社のアピールポイントを掲載し、就職活動中の学生に対して広く自社の情報を提供する。就職ポータルサイトは、学生が就職活動にあたってははじめに見るものであるから、個社ページは学生の志望に対して影響を与えていると考えられる。その中でも、アピールポイントは各企業が自由に掲載することができるため、分析の結果の活用が容易な項目と考えられる。就活サイト A では、各企業は個社ページ内で、就活サイト A の運営会社が決めた 11 のアピールポイントの中から 3 つまでのアピールポイントを掲載し、それに沿った企業紹介文や画像を掲載することができる。アピールポイント群は次の 11 項目で構成される。

企業理念、ビジョン、事業の特徴、事業優位性、  
仕事内容、魅力的人材、社風、制度、職場環境、  
技術力、採用方針

#### 2.3.2 学生の志望理由

就活サイト A の特徴として、学生が企業を志望する際、その志望理由を登録できる機能がある。学生が志望理由を登録する際には、学生は就活サイト A の運営会社が定めた 20 の志望理由の中から、複数選択可で理由を登録する。ま

た、同じ学生が複数の企業に対して別々の理由を登録することができる。志望理由群は次の20項目で構成される。

仕事内容, 業界, 勤務地, 職場雰囲気, ビジョン,  
ステータス, 経験の活用, 商品の内容, 事業戦略,  
事業成長性, 給与水準, 職場環境, 福利厚生,  
勤務形態, 教育制度, 能力主義, 昇進の早さ,  
自己成長性, 内定期待度, 社会貢献度

### 2.3.3 事前分析

上記で説明した企業のアピールポイントと学生の志望理由について、統計的関係性があるかについての確認を行った。その結果、項目間の相関がほとんどなく、数量化III類等の集約化を行う手法を適用した場合の寄与率は低いことが確認された。つまり、学生の志望理由や企業のアピールポイントには、変数間での直接的な統計的相関は低いと考えられる。これは、様々なマーケティングデータで見られるように、項目間の統計的関係性が学生や企業によって異なっており、これらの統計的性質の異なるサンプルの集合によって全体が構成されているためと考えられる。そのため、分析の第一段階として集約化を行うことは適切ではなく、すべての項目を用いてクラスタリングを行うことで両者の関係性をモデル化すべきと考えられる。

## 2.4 潜在クラスモデル

潜在クラスモデルは、観測されたデータの背後に観測できない潜在的な変数の存在を仮定したモデルである。潜在的な変数の仮定は、異質のデータが混ざったような現実的な複雑な問題の分析が可能とする。いい換えれば潜在クラスモデルは、データ全体は特徴の異なるグループが混在した集合体であるとして仮定したモデルである。潜在クラスモデルには、データがある潜在クラスのもとで生起すると仮定する混合ユニグラムモデルや、データが複数の潜在クラスから確率的に生起することを許容するAM (Aspect Model) 等があり、対象とする事象やデータ構造に合わせて、仮定する確率分布や考慮する事象の特徴をモデルに組み込むことで拡張される。たとえば、文書分類 [20], [21] や購買履歴分析 [22], [23], [24], 推薦システム [25], [26], [27] といった情報科学の分野に限らず、社会計量学の分野 [28] まで幅広く適用されている。たとえば文献 [21] では生成分布に多項分布ではなく、ディリクレ分布を用い、文献 [24] ではロジットモデルに潜在クラスを仮定している。また文献 [27] では、購買履歴だけでなく評価履歴をモデルに取り込むことで、推薦システムの精度を向上させている。

## 3. 提案モデル

### 3.1 概要

先に述べたとおり、本研究の目的は、企業のアピールポイントと学生の志望理由の関係性を分析するモデルの提案である。具体的には、両者の関係の背後には、各企業特有

の情報やアピールの仕方、学生ごとに異なる嗜好が混在していると考えられるため、潜在クラスモデルを新たに提案する。

企業のアピールポイントと学生の志望理由を分析するためのモデルが対象とする事象の持つ2つの特徴を述べる。1つ目の特徴として、学生は複数の企業を志望するが、そのたびに学生が企業にどのように魅力を感じるかは異なると考えられることがあげられる。そこで、AMのように、同じ学生・同じ企業でも志望理由登録の度に異なる潜在クラスに所属することが可能なモデルの構築が望まれる。2つ目の特徴として、1回の志望理由登録においては、ある関係性をもとにして企業のアピールポイントと学生の志望理由が同時生起していると考えられる。この特徴は、混合ユニグラムモデルでの「潜在クラスをもとにデータが同時生起する」という仮定と一致する。この2つの特徴を表現するため、すなわち、AMと混合ユニグラムモデルの双方の特徴を有したモデルを構築するため、志望理由登録ごとに潜在クラスを仮定したモデルを提案する。そこで、各理由登録に対して、その登録された企業が掲載しているアピールポイントと、学生が登録した志望理由がベクトルの共起で与えられたデータ構造を仮定する。

提案モデルは潜在クラスモデルの一種であるので、潜在変数の観測は不可能である。そこで、EMアルゴリズムを用いて各パラメータを算出するというアプローチを取る。しかしながら、従来のAMで対象とされてきたデータ構造と異なり、対象データ構造がベクトルとなる。よって、パラメータ推定アルゴリズム中の確率計算において、アピールポイントベクトルおよび志望理由ベクトルはそれぞれのベクトルの要素の生起確率の積の形で表される。そのため、ベクトルの次元数とそのベクトルの生起確率の高低に影響する。結果として、次元数の大きいベクトルが軽視されてしまいやすいという問題点が存在を生じさせてしまう。本研究では、この問題点を解決する学習アルゴリズムについても提案する。

### 3.2 変数の定義

まず、提案モデルで用いる変数を定義する。 $J$  個のアピール集合を  $\mathcal{A} = \{a_j : 1 \leq j \leq J\}$ ,  $I$  個の理由集合を  $\mathcal{R} = \{r_i : 1 \leq i \leq I\}$  と定義する。全  $N$  件の理由登録のうち、 $n$  番目に理由登録された企業のアピールポイントベクトルを  $\mathbf{x}_n = (x_1^n, x_2^n, \dots, x_J^n)$ ,  $n$  番目の理由登録に対する志望理由ベクトルを  $\mathbf{y}_n = (y_1^n, y_2^n, \dots, y_I^n)$  と表す。ただし、 $x_j^n, y_i^n$  はそれぞれ  $n$  番目の理由登録でアピールポイント  $a_j$ , 志望理由  $r_i$  が選択されていれば1, そうでなければ0を取る二値変数である。また、 $K$  個の潜在クラス集合を  $\mathcal{Z} = \{z_k : 1 \leq k \leq K\}$  と定義する。

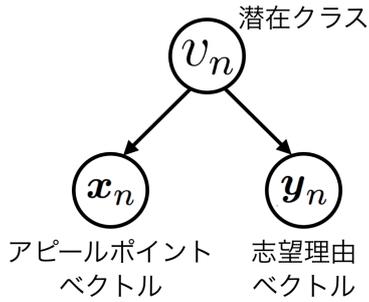


図 1 提案モデルのグラフィカルモデル  
Fig. 1 Graphical model of proposed model.

### 3.3 モデルの定式化

ここで、 $n$  番目の志望理由登録に注目し、所属する潜在クラスを  $v_n \in \mathcal{Z}$  とし、 $n$  番目の志望理由登録を表すデータを  $(\mathbf{x}_n, \mathbf{y}_n, v_n)$  と表すものとする。このとき、 $v_n$  を含む完全データの確率モデル  $P(\mathbf{x}_n, \mathbf{y}_n, v_n)$  は式 (1) で表される。

$$\begin{aligned}
 P(\mathbf{x}_n, \mathbf{y}_n, v_n) &= P(v_n)P(\mathbf{x}_n|v_n)P(\mathbf{y}_n|v_n) \\
 &= P(v_n) \prod_{j=1}^J P(a_j|v_n)^{x_j^n} P(\bar{a}_j|v_n)^{1-x_j^n} \\
 &\quad \times \prod_{i=1}^I P(r_i|v_n)^{y_i^n} P(\bar{r}_i|v_n)^{1-y_i^n} \quad (1)
 \end{aligned}$$

ただし、 $P(a_j|v_n)$  は潜在クラス  $v_n$  のもとで、アピールポイント  $a_j$  を掲載する事象の確率を表し、 $P(\bar{a}_j|v_n)$  は掲載しない事象の確率を表す。同様に、 $P(r_i|v_n)$  は潜在クラスのもとで志望理由  $r_i$  が登録する事象の確率を表し、 $P(\bar{r}_i|v_n)$  は登録しない事象の確率を表す。このとき、 $P(a_j|v_n) + P(\bar{a}_j|v_n) = 1$ 、 $P(r_i|v_n) + P(\bar{r}_i|v_n) = 1$  である。

本研究では、各アピールポイントと各志望理由の確率モデルには二項分布を仮定している。これにより、複数の項目を選択できる状況のもとで、各項目が独立に選択されることが表現される。また、得られるパラメータが各潜在クラスにおける各項目の生起確率になり、結果の解釈が容易になる。

提案モデルのグラフィカルモデルは、図 1 で表される。

ここで、全  $N$  件の志望理由登録の同時確率を表現した確率モデルは、式 (1) の積により、次式で表される。

$$\begin{aligned}
 P(\mathbf{X}, \mathbf{Y}, \mathbf{V}) &= \prod_{n=1}^N P(\mathbf{x}_n, \mathbf{y}_n, v_n) \\
 &= \prod_{n=1}^N P(v_n)P(\mathbf{x}_n|v_n)P(\mathbf{y}_n|v_n) \\
 &= \prod_{n=1}^N P(v_n) \prod_{j=1}^J P(a_j|v_n)^{x_j^n} P(\bar{a}_j|v_n)^{1-x_j^n} \\
 &\quad \times \prod_{i=1}^I P(r_i|v_n)^{y_i^n} P(\bar{r}_i|v_n)^{1-y_i^n} \quad (2)
 \end{aligned}$$

ただし、 $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$ 、 $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N)$ 、 $\mathbf{V} = (v_1, v_2, \dots, v_N)$  とする。

### 3.4 パラメータの学習アルゴリズム

提案モデルにおけるパラメータ  $P(z_k)$ 、 $P(a_j|z_k)$ 、 $P(r_i|z_k)$  は、EM アルゴリズムによって推定することができる。EM アルゴリズムは、観測不可能な潜在変数に確率モデルが依存し、直接的に最尤推定ができない場合に、尤度を極大化することによってパラメータを推定する方法の 1 つである。具体的には、対数尤度の期待値をとる操作 (E-step) と、それを最適化するパラメータの算出する操作 (M-step) を繰り返すことで、パラメータを探索的に推定する。各 Step の更新式は以下の式 (3)~(6) となる\*2。

【E-step】

$$P(z_k|\mathbf{x}_n, \mathbf{y}_n) = \frac{P(\mathbf{x}_n, \mathbf{y}_n, z_k)}{\sum_{k=1}^K P(\mathbf{x}_n, \mathbf{y}_n, z_k)} \quad (3)$$

【M-step】

$$P(z_k) = \frac{1}{N} \sum_{n=1}^N P(z_k|\mathbf{x}_n, \mathbf{y}_n) \quad (4)$$

$$P(a_j|z_k) = \frac{1}{NP(z_k)} \sum_{n=1}^N P(z_k|\mathbf{x}_n, \mathbf{y}_n) x_j^n \quad (5)$$

$$P(r_i|z_k) = \frac{1}{NP(z_k)} \sum_{n=1}^N P(z_k|\mathbf{x}_n, \mathbf{y}_n) y_i^n \quad (6)$$

EM アルゴリズムでは、E-step と M-step を合わせて 1 ステップとし、次式で表されるすべての完全データの対数尤度  $LL$  について、更新前のステップで得られた対数尤度と更新後のステップで得られた対数尤度の差が規定値以下になるまで繰り返し計算を行う。

$$\begin{aligned}
 LL &= \log P(\mathbf{X}, \mathbf{Y}, \mathbf{V}) \\
 &= \sum_{n=1}^N \sum_{v_n \in \mathcal{Z}} \log P(\mathbf{x}_n, \mathbf{y}_n, v_n) \quad (7)
 \end{aligned}$$

### 3.5 学習アルゴリズムの改良

本研究で対象となるデータ構造は、各理由登録に対してアピールと理由がそれぞれベクトルで与えられる。そして、式 (1) における  $P(\mathbf{x}_n|v_n)$ 、 $P(\mathbf{y}_n|v_n)$  はベクトル内の各要素の二項分布の積の形で計算される。そのため、 $I \neq J$  のとき、要素数が多い方のベクトルの生起確率は小さくなりやすく、結果として要素数が少ない方のベクトルの特徴に過剰に適合した潜在クラスを構成してしまいやすいという問題が存在する。この問題を解決するために、式 (3) における  $P(\mathbf{x}_n, \mathbf{y}_n, z_k)$  を次式で置き換えた学習アルゴリズムを提案する。

$$P'(\mathbf{x}_n, \mathbf{y}_n, z_k) = P(z_k)P(\mathbf{x}_n|z_k)^\alpha P(\mathbf{y}_n|z_k)^\beta \quad (8)$$

ただし、 $\alpha$ 、 $\beta$  は重みパラメータであり、ベクトルの次元数や各要素の確率分布を考慮して経験的に決定する必要がある。

\*2 更新式の導出に関しては付録を参照

表 1 各アピールポイントの生起確率の相対値の比較

Table 1 Comparison of occurrence probability of each appeal point.

	企業理念	ビジョン	事業の特徴	事業優位性	仕事内容	魅力の人材	社風	制度	職場環境	技術力	採用方針
$P(a_j z_1)/P(a_j)$	<b>1.07</b>	0.74	<b>1.14</b>	<b>1.07</b>	<b>1.29</b>	<b>1.49</b>	0.98	0.91	0.62	0.02	1.00
$P(a_j z_2)/P(a_j)$	<b>1.84</b>	0.64	0.62	0.46	0.27	0.76	<b>1.30</b>	0.36	<b>3.47</b>	<b>2.92</b>	0.35
$P(a_j z_3)/P(a_j)$	0.57	<b>1.18</b>	0.17	<b>1.48</b>	<b>1.34</b>	<b>1.42</b>	0.75	<b>1.21</b>	0.46	<b>1.59</b>	<b>2.43</b>
$P(a_j z_4)/P(a_j)$	<b>1.20</b>	<b>1.44</b>	0.84	0.35	0.03	<b>1.21</b>	<b>1.31</b>	<b>2.05</b>	<b>2.50</b>	0.39	<b>1.12</b>
$P(a_j z_5)/P(a_j)$	0.80	<b>1.22</b>	<b>1.46</b>	1.00	0.97	0.27	0.98	0.98	0.48	<b>1.22</b>	0.44

表 2 各理由の生起確率の相対値の比較

Table 2 Comparison of occurrence probability of each reason.

	仕事内容	業界	勤務地	職場雰囲気	ビジョン	ステータス	経験の活用	商品の内容	事業戦略	事業成長性
$P(r_i z_1)/P(r_i)$	0.95	<b>1.13</b>	<b>1.11</b>	<b>1.17</b>	<b>1.14</b>	0.78	0.88	0.81	0.47	0.50
$P(r_i z_2)/P(r_i)$	0.74	<b>1.04</b>	<b>1.06</b>	0.79	0.99	0.88	0.40	0.89	0.93	<b>1.01</b>
$P(r_i z_3)/P(r_i)$	<b>1.07</b>	<b>1.05</b>	0.99	0.13	0.07	0.66	<b>1.06</b>	0.66	0.28	0.39
$P(r_i z_4)/P(r_i)$	<b>1.09</b>	0.31	0.95	<b>2.11</b>	<b>1.38</b>	0.68	0.93	0.65	<b>1.45</b>	<b>1.24</b>
$P(r_i z_5)/P(r_i)$	<b>1.09</b>	0.96	0.87	<b>1.09</b>	<b>1.26</b>	<b>1.54</b>	<b>1.34</b>	<b>1.51</b>	<b>1.90</b>	<b>1.82</b>
	給与水準	職場環境	福利厚生	勤務形態	教育制度	能力主義	昇進の早さ	自己成長性	内定期待度	社会貢献度
$P(r_i z_1)/P(r_i)$	0.60	0.78	0.71	0.79	0.75	0.25	0.44	0.64	0.96	0.86
$P(r_i z_2)/P(r_i)$	0.19	0.83	0.34	0.22	0.52	0.83	0.56	0.94	0.87	0.70
$P(r_i z_3)/P(r_i)$	0.74	0.28	0.62	0.60	0.54	0.32	0.44	0.74	0.74	0.16
$P(r_i z_4)/P(r_i)$	<b>1.56</b>	0.72	<b>1.91</b>	<b>1.95</b>	<b>2.87</b>	<b>2.99</b>	<b>2.83</b>	<b>1.32</b>	<b>1.68</b>	<b>1.25</b>
$P(r_i z_5)/P(r_i)$	<b>1.75</b>	<b>1.76</b>	<b>1.56</b>	<b>1.52</b>	<b>1.27</b>	<b>1.81</b>	<b>1.67</b>	<b>1.49</b>	<b>1.07</b>	<b>1.66</b>

## 4. 分析

提案モデルを用いて、就活サイト A に蓄積された企業のアピールポイントデータおよび学生の志望理由データの関係を分析する。さらに、構成された潜在クラスについて、企業の業種情報等を用いて多角的に分析することで、知見を得ることを試みる。

### 4.1 分析データ

本研究では、就活サイト A における 2016 年卒業学生を対象とするサービスにおいて蓄積されたアピールデータおよび理由データを用いる。アピールポイント数および志望理由数は  $J = 11$ ,  $I = 20$  である。また、分析対象の企業として 2 つの条件を設けた。アピールポイントを 3 つ掲載していること\*3と、従業員規模 1,000 人未満であること\*4である。学生の理由登録データは、2014 年 6 月から 2015 年 9 月までの期間における対象企業への全理由登録データから、ランダムに 100,000 件を抽出した。また、事前実験の結果より、潜在クラス数は 6 以上では解釈が困難であった

\*3 企業が掲載しているアピールポイント数の差異による不公平性を排除するため、全体の約 94% の企業が規定上限の 3 つのアピールポイントを掲載していた。

\*4 従業員規模が 1,000 人以上の所謂大企業の場合、就活サイト A 内の個社ページに掲載しているアピールポイントが、学生の志望理由に与える影響が小さいと考えられるため。

表 3 潜在クラスの生起確率

Table 3 Probability of latent class.

	$z_1$	$z_2$	$z_3$	$z_4$	$z_5$
$P(z_k)$	0.35	0.11	0.16	0.07	0.31

ため、 $K = 5$  とした\*5。式 (8) における重みパラメータは  $\alpha = 1.25$ ,  $\beta = 1.00$  とした\*6。

### 4.2 分析結果・考察

提案手法の適用により得られた結果を表 1, 表 2, 表 3 に示す。各潜在クラスのもとでの各アピール、各理由の生起確率  $P(a_j|z_k)$ ,  $P(r_i|z_k)$  を基準となる  $P(a_j)$ ,  $P(r_i)$  に対する比を表 1, 表 2 に示し、数値が 1 を超える場合に太字にした。ただし、 $P(a_j)$ ,  $P(r_i)$  は 2 値変数  $x_j^n$ ,  $y_i^n$  を用いて次の式 (9), 式 (10) より、アピールポイント  $a_j$ , 志望理由  $r_i$  が登録された頻度から求めた。また、各潜在クラスの生起確率を表 3 に示す。

\*5 潜在クラス数はモデル構造 (モデルの自由度) を定めるため、意味のある潜在クラスの集合が構成されるような数を選ぶ必要がある。一方で、提案法であるパラメータ推定法は、アピールポイントと志望理由の双方の統計的特徴がきちんと推定に反映されるようにするための方法である。したがって、近似式によるパラメータ推定を導入して、アピールポイントと志望理由の双方の統計的特徴がきちんと推定に反映されるように学習された潜在クラスモデルを、異なる潜在クラス数に対して構築して、そのなかで最も解釈可能性の高いモデルを選ぶというのが現実的な方法となる。

\*6 本研究では、次元の大きい  $\mathbf{y}_n$  の重みパラメータ  $\beta$  を 1.00 と固定し、 $\mathbf{x}_n$  の重みパラメータ  $\alpha$  を 0.05 刻みで変化させて学習を行い、潜在クラスの特徴がよく現れる値  $\alpha = 1.25$  を探索した。詳しい設定については付録で解説する。

$$P(a_j) = \frac{1}{N} \sum_{n=1}^N x_j^n \quad (9)$$

$$P(r_i) = \frac{1}{N} \sum_{n=1}^N y_i^n \quad (10)$$

表 1, 表 2 を併せて比較検討することで, 各アピールと各理由の関係性を分析することができる。これらの結果から, それぞれの潜在クラスについて特徴の解釈を行う。

#### 潜在クラス 1

表 1 より, 「企業理念」, 「事業の特徴」, 「事業優位性」, 「仕事内容」, 「魅力的人材」のアピールポイントの生起確率が比較的高い。一方, 表 2 より, 「業界」, 「勤務地」, 「職場雰囲気」, 「ビジョン」の志望理由の生起確率が比較的高い。アピールポイントの「事業の特徴」や「事業優位性」は, 志望理由の「業界」にマッチしており, ほかに「企業理念」と「ビジョン」が, 「魅力的人材」と「職場雰囲気」がそれぞれマッチしていると分かる。これらから, 潜在クラス 1 はアピールポイントと志望理由が総じてマッチしているクラスといえる。特に「企業は事業や魅力的人材についてアピールし, 学生の志望理由もマッチしているクラス」と解釈することができる。また,  $P(z_1)=0.35$  と生起確率が他の潜在クラスと比べて最も高いため, この関係性が最も多く出現していると分かる。

#### 潜在クラス 2

表 1 より, 「企業理念」, 「社風」, 「職場環境」, 「技術力」のアピールポイントの生起確率が比較的高い。一方, 表 2 より, 「業界」, 「勤務地」, 「事業成長性」の志望理由の生起確率がそれぞれ比較的高い。よって, 「企業は社風や環境についてアピールしているが, 学生は(業界や勤務地といった)基本情報を軸として就職活動を行っているクラス」と解釈することができる。また,  $P(z_2)=0.11$  であるため, この共起は潜在クラス 1 と比較して生起確率が低いといえる。

#### 潜在クラス 3

潜在クラス 3 で特徴となるのが, アピールポイントの「採用方針」と志望理由の「内定期待度」の生起確率であり, 両者に注目することで興味深いことが分かる。潜在クラス 3 において, 「採用方針」は全潜在クラス中で最も生起確率が高いが, 「内定期待度」は最も生起確率が低い。「内定期待度」は学生が主観的に「内定が取りやすそう」と考えて選択しているものだが, 潜在クラス 3 においては学生のこの主観的な考えが, 企業が採用方針に触れることで志望理由としては弱くなるものと示唆される。内定期待度で志望する(志望される)ことに対して, 一概に良し悪しを決めることはできないが,  $P(z_3)=0.16$  と生起確率が低いわけではなく, この結果は, 各企業が掲載するアピールポイン

トを決めるために参考となりうる結果と思われる。

#### 潜在クラス 4

表 1 より, 「企業理念」, 「ビジョン」といった面や, 「社風」, 「制度」, 「職場環境」といった面のアピールポイントの生起確率が比較的高い。一方, 表 2 より, 志望理由は全体的にどの項目も生起確率が比較的高いが, 特に「職場雰囲気」, 「福利厚生」, 「勤務形態」, 「教育制度」, 「能力主義」, 「昇進の早さ」の生起確率が高い。どちらも社風や制度面での項目が高いことから, 潜在クラス 4 は「企業がアピールする社風や制度面のポイントと, 学生の志望理由がマッチしているクラス」と解釈することができる。また,  $P(z_4)=0.07$  と生起確率が低いため, 社風や制度面でマッチすることは少ないといえる。

#### 潜在クラス 5

表 1 より, 潜在クラス 5 の特徴を見ると, 「ビジョン」, 「事業の特徴」, 「技術力」のアピールポイントの生起確率が比較的高いが, 全体で見ると 1 に近い項目が多い。すなわち, 潜在クラス 5 は, アピールポイントではあまり特徴づけられていないと解釈ができる。一方, 表 2 より, 志望理由の生起確率はどれも比較的高い。これらから, 潜在クラス 5 は, 「企業がアピールする事業についてのポイントと学生の志望理由がマッチしていることに加え, 全体的に学生の志望理由が多いクラス」と解釈することができる。また,  $P(z_5)=0.31$  と生起確率が比較的高く, このクラスは学生が一度の理由登録において, 項目は問わず多数の項目を選択することで構成されたクラスと考えられる。

以上のように, 提案モデルによって企業のアピールポイントと学生の志望理由の関係性を定量的に分析し, 関係性を抽出することが可能となった。具体的には, 潜在クラス 1, 4 は, 企業のアピールポイントと学生の志望理由がマッチしているクラスであり, その他のクラス 2, 3, 5 は上述のとおりでマッチしていないことが分かった。これらのマッチしていないクラスに属する確率の高い企業は, アピールポイントが学生の志望理由からやや外れていると考えられるため, 所属する潜在クラスの特徴を把握したうえで, その原因を考察することで対応を検討する必要がある。

一方, 企業の事業というアピールポイントと学生の業界といった志望理由がマッチしている潜在クラス 1 の生起確率が 35% と比較的高く, このマッチングは比較的上手くいっていると考えられる。このクラスに属する企業は, 期待通りの結果が得られていることを確認し, 継続的にこのクラスへの所属から外れないようにモニタリングすべきである。これに対し, 潜在クラス 4 では確かに企業がアピールする社風や制度面に対して学生の志望理由がマッチしているクラスであるが, その生起確率が 7% と低い値になっている。つまり, 社風や制度面でマッチすることは少ない

と考えられ、サイト上の限られたスペースにおいて社風や制度面をアピールしようとする企業は、この点に注意すべきだと考えられる。

### 4.3 潜在クラスと企業を併せた分析

ここまではアピールと理由の条件付き生起確率から各潜在クラスの解釈を行ったが、本項では実際に各潜在クラスにはどのような企業が所属しているかを分析する。分析には、企業の業界、従業員規模を用いる。

分析に先立ち、提案モデルにおいて、企業がどのように潜在クラスに所属するかを説明する。パラメータを推定するためのEMアルゴリズム中で、 $P(z_k|\mathbf{x}_n, \mathbf{y}_n)$ が算出される。すなわち、アピールポイントに対する各理由登録データが所属する潜在クラスが算出される。また、各理由登録データは、計算においてはアピールポイントベクトルおよび志望理由ベクトルのみを用いているが、企業のデータも関連付けられている。よって、各企業が所属する潜在クラスも容易に求めることが可能である。本分析においては、企業の潜在クラスへの所属確率を理由登録回数としてソフトクラスタリングし、業界や従業員規模によって潜在クラスへの所属割合がどのように異なるかについて分析を行う。ただし、企業は複数の学生から志望されるため、そのたびに潜在クラスへの所属度合いが変わることを許容している。

いま、 $n$ 番目の理由登録データの企業を $c_n$ とする。また、業界 $g_l$ の集合を $\mathcal{G} = \{g_l : 1 \leq l \leq L\}$ 、従業員規模カテゴリ $s_m$ の集合を $\mathcal{S} = \{s_m : 1 \leq m \leq M\}$ と定義する。ここで、潜在クラス $z_k$ に所属し、業界が $g_l$ 、従業員規模が $s_m$ の企業への志望理由登録数をそれぞれ $\mathcal{N}(z_k, g_l)$ 、 $\mathcal{N}(z_k, s_m)$ とし、以下の式で集計する。

$$\mathcal{N}(z_k, g_l) = \sum_{n=1}^N \delta(c_n = g_l)(z_k|\mathbf{x}_n, \mathbf{y}_n) \quad (11)$$

$$\mathcal{N}(z_k, s_m) = \sum_{n=1}^N \delta(c_n = s_m)(z_k|\mathbf{x}_n, \mathbf{y}_n) \quad (12)$$

ただし、 $\delta$ は引数が真のときに1を、偽のときに0を返すインジケータ関数である。

#### 4.3.1 業界と併せた分析

業界と潜在クラスの関係を、表4に示す。表4は式(11)によって集計されたクロス表の要素を、その列の合計理由登録数に対する割合で表したものである。また、各列の値が右の全体列の値を超える場合には太字とした。

表4の結果より、潜在クラスおよび業界の特徴を分析する。潜在クラス1, 3, 4, 5では、潜在クラス内の各業界の理由登録数の割合にそれほど変わりはないが、潜在クラス2における各業界の割合が特徴的である。潜在クラス2は他の潜在クラスに比べて、「メーカー」と「ソフトウェア」の割合が高く、「商社」「サービス」の割合が低い。こ

表4 業界と潜在クラスのクロス表 (各列で正規化)

Table 4 Cross table of industry and latent class (normalized in each column).

	$z_1$	$z_2$	$z_3$	$z_4$	$z_5$	$P(g_l)$
メーカー	0.20	<b>0.38</b>	0.22	0.24	<b>0.36</b>	0.27
商社	<b>0.15</b>	0.09	<b>0.16</b>	<b>0.19</b>	0.12	0.14
小売店	0.03	0.03	0.02	<b>0.06</b>	<b>0.04</b>	0.03
金融	0.02	0.01	0.02	<b>0.03</b>	0.01	0.02
メディア	<b>0.18</b>	0.15	<b>0.20</b>	0.12	0.15	0.17
ソフトウェア	0.06	<b>0.14</b>	0.07	<b>0.10</b>	0.07	0.08
サービス	<b>0.36</b>	0.20	<b>0.32</b>	0.27	0.26	0.30
合計	1.00	1.00	1.00	1.00	1.00	1.00

表5 従業員規模と潜在クラスのクロス表 (各列で正規化)

Table 5 Cross table of employee size and latent class (normalized in each column).

	$z_1$	$z_2$	$z_3$	$z_4$	$z_5$	$P(s_m)$
1-49人	<b>0.10</b>	0.05	<b>0.10</b>	0.07	0.05	0.08
50-99人	<b>0.15</b>	<b>0.21</b>	0.10	0.13	0.10	0.13
100-299人	<b>0.30</b>	0.24	<b>0.29</b>	0.27	<b>0.29</b>	0.28
300-499人	0.17	0.17	<b>0.22</b>	<b>0.19</b>	0.16	0.18
500-999人	0.27	0.32	0.30	<b>0.35</b>	<b>0.40</b>	0.33
合計	1.00	1.00	1.00	1.00	1.00	1.00

ここで、潜在クラス2は「社風や職場環境をアピールポイントとしても、学生は基本情報を軸に就職活動をしているクラス」であった。すなわち、「メーカー」「ソフトウェア」といった「ものづくり」の企業は、マッチングの観点からすると、社風や環境面以外についてのアピールポイントを掲載することが良い傾向にある可能性があることと指摘することができる。また、潜在クラス1と潜在クラス3は、潜在クラス内の業界の割合が類似していることが分かる。

#### 4.3.2 従業員規模と併せた分析

従業員規模と潜在クラスの関係を、表5に示す。表は表4と同様に示す。

表5を検討し、各潜在クラスの従業員規模の特徴を分析する。潜在クラス1を見ると、従業員規模が小さい場合には平均と比べて多くの企業が所属していることが分かる。潜在クラス1は「企業は事業や魅力的人材についてアピールし、学生の志望理由もマッチしているクラス」であったが、規模の小さい企業の方がこのようなマッチングの傾向が多く現れると分かる。逆に、規模が大きくなるにつれて所属する企業が比較的多くなる潜在クラスとして、潜在クラス4, 5があげられる。潜在クラス4は「企業がアピールする社風や制度面のポイントと、学生の志望理由がマッチしているクラス」であったので、規模が大きな企業は社風や制度をアピールすることがマッチングの観点からは好ましいと考えることができる。また、潜在クラス5は「企業のアピールによらず、学生が多くの志望理由を持つクラス」であった。すなわち、従業員規模が大きくなるにつれ、就

職ポータルサイト上でのアピール以外の外部要因が働き、それが学生の志望理由につながっていると考えられる。

## 5. 考察

### 5.1 モデルのパラメータ決定

提案モデルでは、潜在クラス数  $K$  と学習アルゴリズム内での重みパラメータ  $\alpha$ ,  $\beta$  は分析者によって事前に決定される。

前者について、事前実験により、潜在クラス数が5よりも大きい場合には、解釈が困難なクラスが生成されることや、類似する意味を持つクラスが重複して生成されることが確認された。つまり、潜在クラス数が5のとき（企業のアピールポイントと学生の志望理由の関係性を大きく5つのパターンと考えるとき）で、実社会における現象をある程度表現できているのではないかと考えられる。そこで、本論文では、潜在クラス数を5とし、モデルの適用および分析を行った。ただし、クラスタリングの結果を予測等の別問題に適用する場合には、その予測精度等をもとにクラス数を決定することも考えられる。また、AIC, BIC等の情報量基準を用いることも考えられる。

後者について、提案モデルでは、ベクトルの確率計算においての問題点を改善するため、学習アルゴリズム内でのベクトルの生起確率に重みパラメータを導入した。前章の分析では  $\alpha=1.25$ ,  $\beta=1.00$  としたが、これらはヒューリスティックな探索により定めた値である。パラメータの値を定める1つの指針として、ベクトルの生起確率に対してそのベクトルの次元数の累乗根をとることが考えられる。ただし、両ベクトルの各要素の確率分布の違いから、一般的な方法でバランスをとることは難しいといえる。また、本研究で提案した重みを用いた式(8)は、本研究に限らず、ベクトルの各要素の積により確率を求める場合には汎用的に有効であると考えられる。その際には、ベクトルの次元数やベクトルの要素の確率分布等を考慮して、パラメータの値を定める必要がある。しかし、ベクトルの要素の確率分布は、潜在クラスが与えられたもとの確率分布となるため、試行錯誤的な調整が必要である。また、EMアルゴリズムによる繰り返し計算を考慮し、繰り返しごとに重みを変更する等の改善案も考えられる。

### 5.2 対象事例の解析と他データへの適用

企業のアピールポイントと学生の志望理由について関係性の分析を行い、マッチングの傾向を把握することが可能となった。得られた潜在クラスの解釈や、潜在クラスごとの企業の特徴を見ると、実社会の現象による経験的知識とおおかた一致しており、大局的にアピールポイントと志望理由の関係をとらえられていると考えられる。しかし、いずれも「何かをアピールしているが、業界や勤務地といった基本情報が志望理由になる」というものであった。1章

で述べたような、技術力をアピールしているが、給与水準が志望理由になるといったような、極端な関係性は見られなかった。そのような例は、実社会において存在はするが、傾向として抽出できるほど多く存在していないことが考えられる。

また、対象データとして扱ったアピールポイントデータおよび志望理由データの設計項目について改善の余地があるといえる。たとえばアピールポイントとして「ビジョン」と「企業理念」が別のものとして扱われていることに対し、志望理由としては「企業理念・ビジョン」とまとめられている。その他にも、同等の意味を持つ事象が別の項目として扱われていることがある。そのため、データ収集の段階の設計を改善することで、企業のアピールポイントと学生の志望理由の関係性がよりとらえやすくなると思われる。

本研究の提案モデルは、企業のアピールポイントと学生の志望理由に限らず、複数の項目が同時に選択されるという状況の下で、その1つ1つの項目についてそれぞれ関係性を把握したい場合に有効であると考えられる。特に、マッチング問題と考えれば、就職ポータルサイトにおける企業と学生という事象にとらわれず、お見合いのような男性と女性のマッチング活動にも本提案モデルは適用可能と考えられる。

### 5.3 モデルの評価

提案モデルの客観的な有用性を考察するため、評価指標を導入してモデルの評価を行う。本論文では、潜在クラスモデルやトピックモデルで一般的に評価指標として用いられる Coherence [29] を評価指標として採用し、代表的なクラスタリング手法である  $k$ -means 法を比較対象として性能の比較を行う。

Coherence は潜在クラス内の生起確率トップ10の項目を対象に算出することとする。ここで、Coherence は潜在クラスごとに算出されるが、すべての潜在クラスに対して足し合わせてモデルの評価とする。具体的には式(13)で計算し、初期値をランダムに変更して10回それぞれのモデルの学習を行った結果の平均値を算出する。

$$\begin{aligned} \text{Coherence} &= \sum_{k=1}^K \sum_{(w_p, w_q) \in \text{Top}M_{z_k}} \log \frac{D(w_p, w_q) + \epsilon}{D(w_p)} \quad (13) \end{aligned}$$

ただし、 $w_p$  はアピールポイントもしくは志望理由の1つの項目 ( $w_p \in \mathcal{A} \cup \mathcal{R}$ ) を表し、 $\text{Top}M_{z_k}$  は潜在クラス  $z_k$  における生起確率の高い  $M$  の項目と定義する。また、 $D(w_p)$  は  $w_p$  が含まれる志望理由登録数、 $D(w_p, w_q)$  は  $w_p, w_q$  が共起する志望理由登録数とする。ここで、 $M = 10$ ,  $\epsilon = 1$  とした。

表6より、提案モデルが従来法である  $k$ -means 法に比べて優れた結果を示していることが分かる。これは、確率

表 6 Coherence による性能の比較\*7

Table 6 Comparison of Performance by Coherence.

手法	Coherence
$k$ -means 法	-581.82
提案モデル ( $\alpha = 1.00, \beta = 1.00$ )	-535.08**
提案モデル ( $\alpha = 1.25, \beta = 1.00$ )	-559.59**
提案モデル ( $\alpha = 1.50, \beta = 1.00$ )	-564.52*

的な潜在変数の導入により、マッチングの傾向をより柔軟に表現できることに起因していると考えられる。この結果より、クラスタリング手法という観点での提案モデルの有用性の一側面が示されたといえる。

次に、提案学習アルゴリズムにおける重みパラメータの導入について考察を行う。重みパラメータを導入した場合、導入しない場合（従来の EM アルゴリズム）に比べて Coherence の値が悪くなっている。これは、重みパラメータがない場合が最もデータにフィットする（尤度が大きくなる）ように確率変数の推定を行っていることからこのような結果となると考えられる。しかし、すでに述べているように、重みパラメータを導入しない場合は、次元数の少ないアピールポイントどうしの関係性にクラスタリング結果が引っ張られることになり、アピールポイントと志望理由の関係性を見ることが難しくなってしまうという問題が発生する。これは、アピールポイントの方に過度にフィッティングすることで、学生の志望理由についてはフィッティングしなくても全体としての尤度が大きくなってしまふことが原因である。一方で、Coherence は「アピールポイント」と「志望理由」の関係性の抽出に関係なく、クラスタリングの結果自体を評価している基準であるため、このような関係性抽出の程度を評価することはできないといえる。そのため、目的的分析と Coherence が直接結び付かないという意味では、Coherence のみで提案モデルの有用性を評価することは難しいと考えられ、提案モデルを評価する指標を設定する必要がある。

## 6. まとめと今後の課題

本研究では、就職ポータルサイトにおける企業のアピールポイントと学生の志望理由の関係性を分析することを目的とし、企業別の単純集計では把握することができなかった分析を可能とする潜在クラスモデルを新たに提案した。また、提案モデルのパラメータ推定において、よりマッチング分析に適した潜在クラスの特徴が得られるように重みパラメータを導入した学習アルゴリズムを提案した。さらに、提案モデルを実データに対して適用し、アピールポイントと志望理由の関係性について定量的な分析を行った。その結果として、潜在クラスごとにマッチングの傾向が異

なることが確認され、潜在クラスの解釈や、潜在クラスと企業の基本情報の掛け合わせにより、4章で示したような新たな知見を得ることができた。以上のことから、提案モデルの有効性が示されたといえる。得られた結果や議論を実務に活用していくことが期待される。

今後の課題として、得られた結果の要因をさらに分析する必要があると考えられる。本研究では、潜在クラスに対して企業の基本情報から分析を行ったが、アピールポイントに付随するアピール本文やそれを表現する画像、その他依存していると考えられる学生や企業の諸情報を分析する必要があり、それらを考慮したモデルの拡張が望まれる。特に、アピール本文をテキストマイニングし分析することで、マッチングに効果的な単語を抽出することも考えられ、ビジネス上でも有益な情報につながると思われる。また、本研究に対して適切な評価指標を定め、それに対して最適な潜在クラス数  $K$  の決定方法や、重みパラメータ  $\alpha, \beta$  の設定方法を考えることも必要である。

**謝辞** 本研究にあたり、貴重なデータの提供、並びに熱心な議論をいただいた就職ポータルサイト A の運営会社の皆様に深く感謝いたします。本研究の一部は、科学研究費 (26282090, 26560167) の助成を受けたものである。

## 参考文献

- [1] 早川真央, 三川健太, 荻原大陸, 後藤正幸: 層別木と混合ワイブル分布に基づく就職活動終了時期の分析モデルの構築, 情報処理学会論文誌, Vol.58, pp.1189–1206 (2017).
- [2] Yamagami, K., Mikawa, K., Goto, M. and Ogihara, T.: A Statistical Prediction Model of Students' Finishing Date on Job Hunting Using Internet Portal Sites Data, *Proc. APIEMS2015* (2015).
- [3] Nagamori, S., Yamashita, H., Goto, M. and Ogihara, T.: An Analytic Model of Relation between Companies' Recruitment Activities and Number of Students' Application based on Mixture Regression Model, *Proc. APIEMS2016* (2016).
- [4] Collins, L.M. and Lanza, S.T.: Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences, *John Wiley & Sons*, Vol.718 (2013).
- [5] Hagenaaers, J.A., McCutcheon, A.L., et al.: Applied latent class analysis, *Cambridge University Press* (2002).
- [6] Christopher Bishop: Pattern Recognition and Machine Learning (Information Science and Statistics), 2nd printing, Springer (2010).
- [7] Nigam, K., McCallum, A.K., Thrun, S. and Mitchell, T.M.: Text Classification from Labeled and Unlabeled Documents using EM, *Machine Learning*, Vol.39, pp.103–134 (2000).
- [8] Hofmann, T.: Probabilistic Latent Semantic Analysis, *Proc. UAI'99*, pp.289–296 (1999).
- [9] Dempster, A., Laird, N. and Rubin, D.: Maximum Likelihood from Incomplete Data via the EM Algorithm, *J. Royal Statist. Soc., Series B*, Vol.39, No.1, pp.1–38 (1977).
- [10] McLachlan, G. and Krishnan, T.: The EM Algorithm and Extensions, *John Wiley & Sons*, Vol.382 (2007).

\*7 \*\*は  $k$ -means 法に対して 1%有意, \*は  $k$ -means 法に対して 5%有意を表す

[11] 宮川雅巳：EM アルゴリズムとその周辺，応用統計学，Vol.16, No.1, pp.1–21 (1987).

[12] 堀 健志，濱中義隆，大島真夫，荻谷剛彦：大学から職業へⅢ その2—就職活動と内定獲得の過程，東京大学大学院教育学研究科紀要，Vol.46, pp.43–74 (2007).

[13] 永野 仁：就職活動成功要因としての就職意識—大学生調査の分析，政経論叢，Vol.73, No.5-6, pp.93–113 (2005).

[14] 下村英雄，堀 洋元：大学生の就職活動における情報探索行動：情報源の影響に関する検討，社会心理学研究，Vol.20, No.2, pp.93–105 (2004).

[15] 下村英雄，木村 周：大学生の就職活動ストレスとソーシャルサポートの検討，進路指導研究（日本進路指導学会研究紀要），No.15, pp.11–19 (1997).

[16] 永野 仁：企業の人材採用の変化，日本労働研究雑誌，No.567 (2007).

[17] 永野 仁：企業の人材採用の動向，日本労働研究雑誌，No.19 (2012).

[18] 堀田聡子：採用時点におけるミスマッチを軽減する採用のあり方 -RJP (Realistic Job Preview) を手がかりにして，政経論叢，日本労働研究雑誌，Vol.567, pp.60–75 (2007).

[19] Goto, M., Minetoma, K., Mikawa, K., et al.: A Modified Aspect Model for Simulation Analysis, *Proc. IEEE International Conference on Systems Man and Cybernetics*, pp.1306–1311 (2014).

[20] 北 研二：確率的言語モデル，東京大学出版会 (1999).

[21] 貞光九月，三品拓也，山本幹雄：混合ディリクレ分布を用いたトピックに基づく言語モデル，電子情報通信学会論文誌，Vol.J88-D-II, No.9, pp.1771–1779 (2005).

[22] 佐藤栄作，廣松 毅，椿 広計：潜在クラスモデルを利用したデータセグメンテーション，行動計量学，No.30-1, pp.121–133 (2003).

[23] 石垣 司，竹中 毅，本村陽一：日常購買行動に関する大規模データの融合による顧客行動予測システム，人工知能学会論文誌，Vol.26-6-D, pp.670–681 (2011).

[24] 守口 剛：潜在クラス・ロジット・モデルを利用したロイヤルティ・セグメンテーション，オペレーションズ・リサーチ：経営の科学，Vol.48, No.10, pp.747–752 (2003).

[25] Hofmann, T. and Puzicha, J.: Latent Class Models for Collaborative Filtering, *Proc. 16th International Joint Conference on Artificial Intelligence*, pp.688–693 (1999).

[26] Hofmann, T.: Latent Semantic Models for Collaborative Filtering, *ACM Trans. Inf. Syst.*, Vol.22, No.1, pp.89–115 (2004).

[27] 大井貴裕，三川健太，後藤正幸：評価と購買の両履歴データの学習による確率的潜在クラスモデルの推定精度向上に関する一考察，日本経営工学会論文誌，Vol.6, No.4, pp.286–293 (2015).

[28] 藤原 翔，伊藤理史，谷岡 謙：潜在クラス分析を用いた計量社会的アプローチ 地位の非一貫性，格差意識，権威主義的伝統主義を例に，年報人間科学，No.33, pp.43–68 (2012).

[29] Mimno, D., Wallach, H.M., Talley, E., et al.: Optimizing Semantic Coherence in Topic Models, *Proceedings of the Conference on Empirical Methods in Natural Language Processing Association for Computational Linguistics*, pp.262–272 (2011).

## 付 録

### A.1 EM アルゴリズムの更新式の導出

ここでは，式 (3)～(6) の導出を解説する．

#### • E-step

$$\begin{aligned}
 P(\mathbf{V}|\mathbf{X}, \mathbf{Y}) &= \frac{P(\mathbf{V}, \mathbf{X}, \mathbf{Y})}{P(\mathbf{X}, \mathbf{Y})} \\
 &= \frac{P(\mathbf{X}, \mathbf{Y}, \mathbf{V})}{\sum_{v_n \in \mathcal{Z}} P(\mathbf{X}, \mathbf{Y}, v_n)} \\
 &= \frac{\prod_{n=1}^N P(\mathbf{x}_n, \mathbf{y}_n, v_n)}{\prod_{n=1}^N \sum_{v_n \in \mathcal{Z}} P(\mathbf{x}_n, \mathbf{y}_n, v_n)} \\
 &= \prod_{n=1}^N \frac{P(\mathbf{x}_n, \mathbf{y}_n, v_n)}{\sum_{v_n \in \mathcal{Z}} P(\mathbf{x}_n, \mathbf{y}_n, v_n)} \\
 &= \prod_{n=1}^N P(v_n|\mathbf{x}_n, \mathbf{y}_n) \tag{A.1}
 \end{aligned}$$

このとき Q 関数は式 (2) と式 (A.1) を用いて次のように表すことができる．

$$\begin{aligned}
 Q &= \sum_{\mathbf{V}} P(\mathbf{V}|\mathbf{X}, \mathbf{Y}) \log P(\mathbf{X}, \mathbf{Y}, \mathbf{V}) \\
 &= \sum_{n=1}^N \left\{ \sum_{k=1}^K P(z_k|\mathbf{x}_n, \mathbf{y}_n) \log P(z_k) \right. \\
 &\quad + \sum_{k=1}^K P(z_k|\mathbf{x}_n, \mathbf{y}_n) \sum_{j=1}^J x_j^n \log P(a_j|z_k) \\
 &\quad + \sum_{k=1}^K P(z_k|\mathbf{x}_n, \mathbf{y}_n) \sum_{j=1}^J x_j^n \log P(\bar{a}_j|z_k) \\
 &\quad + \sum_{k=1}^K P(z_k|\mathbf{x}_n, \mathbf{y}_n) \sum_{i=1}^I y_i^n \log P(r_i|z_k) \\
 &\quad \left. + \sum_{k=1}^K P(z_k|\mathbf{x}_n, \mathbf{y}_n) \sum_{i=1}^I y_i^n \log P(\bar{r}_i|z_k) \right\} \tag{A.2}
 \end{aligned}$$

#### • M-step

次に，M ステップでは式 (A.2) で表された Q 関数を最大化することを考える．以下の条件のもとで，ラグランジュの未定乗数法を用いてパラメータ  $P(z_k)$ ， $P(a_j|z_k)$ ， $P(r_i|z_k)$  を算出する．

##### 制約条件

$$\sum_{k=1}^K P(z_k) = 1 \tag{A.3}$$

$$P(a_j|z_k) + P(\bar{a}_j|z_k) = 1 \tag{A.4}$$

$$P(r_i|z_k) + P(\bar{r}_i|z_k) = 1 \tag{A.5}$$

以上の条件式 (A.3)，(A.4)，(A.5) のもとでラグランジュ関数  $L$  を次式で定義する．ただし， $\iota$ ， $\kappa_{z_k a_j}$ ， $\lambda_{z_k r_i}$  はラグランジュ定数である．

$$\begin{aligned}
 L &= Q - \iota \left( \sum_{k=1}^K P(z_k) - 1 \right) \\
 &\quad - \sum_{k=1}^K \sum_{j=1}^J \kappa_{z_k a_j} \left( P(a_j|z_k) + P(\bar{a}_j|z_k) - 1 \right)
 \end{aligned}$$

$$-\sum_{k=1}^K \sum_{i=1}^I \lambda_{z_k r_i} \left( P(r_i|z_k) + P(\bar{r}_i|z_k) - 1 \right) \quad (\text{A.6})$$

式 (A.6) を求めたいパラメータ  $P(z_k)$ ,  $P(a_j|z_k)$ ,  $P(r_i|z_k)$  について偏微分して 0 と置くことで、解を導出する。

**(1)  $P(z_k)$  について**

式 (A.6) を  $P(z_k)$  で偏微分し、0 と置くと、

$$\frac{\partial L}{\partial P(z_k)} = \sum_{n=1}^N P(z_k|\mathbf{x}_n, \mathbf{y}_n) \frac{1}{P(z_k)} - \iota = 0 \quad (\text{A.7})$$

となる。これに対して、

$$P(z_k) = \frac{1}{l} \sum_{n=1}^N P(z_k|\mathbf{x}_n, \mathbf{y}_n) \quad (\text{A.8})$$

と変形すると、条件式 (A.3) より、

$$\sum_{k=1}^K P(z_k) = \sum_{k=1}^K \frac{1}{l} \sum_{n=1}^N P(z_k|\mathbf{x}_n, \mathbf{y}_n) = 1 \quad (\text{A.9})$$

であることから、

$$\iota = \sum_{k=1}^K \sum_{n=1}^N P(z_k|\mathbf{x}_n, \mathbf{y}_n) = N \quad (\text{A.10})$$

となる。したがって、 $P(z_k)$  は次式で表される。

$$P(z_k) = \frac{1}{N} \sum_{n=1}^N P(z_k|\mathbf{x}_n, \mathbf{y}_n) \quad (\text{A.11})$$

**(2)  $P(a_j|z_k)$  について**

式 (A.6) を  $P(a_j|z_k)$ ,  $P(\bar{a}_j|z_k)$  でそれぞれ偏微分し、0 と置くと、

$$\begin{aligned} \frac{\partial L}{\partial P(a_j|z_k)} &= \sum_{n=1}^N P(z_k|\mathbf{x}_n, \mathbf{y}_n) x_j^n \frac{1}{P(a_j|z_k)} - \kappa_{z_k a_j} \\ &= 0 \end{aligned} \quad (\text{A.12})$$

$$\begin{aligned} \frac{\partial L}{\partial P(\bar{a}_j|z_k)} &= \sum_{n=1}^N P(z_k|\mathbf{x}_n, \mathbf{y}_n) (1 - x_j^n) \frac{1}{P(\bar{a}_j|z_k)} - \kappa_{z_k \bar{a}_j} \\ &= 0 \end{aligned} \quad (\text{A.13})$$

となる。これに対して、

$$P(a_j|z_k) = \frac{1}{\kappa_{z_k a_j}} \sum_{n=1}^N P(z_k|\mathbf{x}_n, \mathbf{y}_n) x_j^n \quad (\text{A.14})$$

$$P(\bar{a}_j|z_k) = \frac{1}{\kappa_{z_k \bar{a}_j}} \sum_{n=1}^N P(z_k|\mathbf{x}_n, \mathbf{y}_n) (1 - x_j^n) \quad (\text{A.15})$$

と変形すると、条件式 (A.4) より、

$$\begin{aligned} &P(a_j|z_k) + P(\bar{a}_j|z_k) \\ &= \frac{1}{\kappa_{z_k a_j}} \sum_{n=1}^N P(z_k|\mathbf{x}_n, \mathbf{y}_n) x_j^n \\ &\quad + P(\bar{a}_j|z_k) \frac{1}{\kappa_{z_k \bar{a}_j}} \sum_{n=1}^N P(z_k|\mathbf{x}_n, \mathbf{y}_n) (1 - x_j^n) \end{aligned}$$

$$= \frac{1}{\kappa_{z_k a_j}} \sum_{n=1}^N P(z_k|\mathbf{x}_n, \mathbf{y}_n) = 1 \quad (\text{A.16})$$

であることから、

$$\kappa_{z_k a_j} = \sum_{n=1}^N P(z_k|\mathbf{x}_n, \mathbf{y}_n) = NP(z_k) \quad (\text{A.17})$$

となる。したがって、 $P(a_j|z_k)$  は次式で表される。

$$P(a_j|z_k) = \frac{1}{NP(z_k)} \sum_{n=1}^N P(z_k|\mathbf{x}_n, \mathbf{y}_n) x_j^n \quad (\text{A.18})$$

**(3)  $P(r_i|z_k)$  について**

$P(r_i|z_k)$  は、 $P(a_j|z_k)$  についての式展開と同一となるため、導出は省略する。 $P(r_i|z_k)$  は次式で表される。

$$P(r_i|z_k) = \frac{1}{NP(z_k)} \sum_{n=1}^N P(z_k|\mathbf{x}_n, \mathbf{y}_n) y_i^n \quad (\text{A.19})$$

**A.2 重みパラメータの設定**

ここでは、4章の分析で設定した重みパラメータ  $\alpha = 1.25$ ,  $\beta = 1.00$  について解説する。分析対象データでは、アピールポイントの項目数が志望理由の項目数よりも多いため、次元の大きい  $\mathbf{y}_n$  の重みパラメータ  $\beta$  を  $\beta = 1.00$  と固定し、 $\mathbf{x}_n$  の重みパラメータ  $\alpha$  を 0.05 刻みで変化させて学習を行い、潜在クラスの特徴がよく現れる値  $\alpha = 1.25$  を探索した。以下に、 $\alpha$  の値を変えた  $\alpha = 1.00$ ,  $\alpha = 1.50$  場合についての結果を以下に示す。

$\alpha = 1.00$  の場合、尤度を極大化する従来の EM アルゴリズムによる推定値となるが、この場合、1つのクラス（潜在クラス 2）において、すべての志望理由の登録が平均より多くなっている。しかし、実際に一度に登録される志望理由は数個であるため、すべてが平均より多いということは考えにくく、このクラスタリング結果は、次元数の低い（項目数の少ない）アピールポイントベクトルの生起確率に依存していると考えられる。次に、 $\alpha = 1.50$  の場合、表の値がどれも（特に潜在クラス 3, 4, 5 において）1 に近く、アピールポイントとの対応を見ることが難しい。以上の結果から、解釈容易性という観点より、論文では  $\alpha = 1.25$  の結果が実務において有用な知見を得られているという結論に至った。

表 A・1 各志望理由の生起確率の比較 ( $\alpha = 1.00, \beta = 1.00$ )

Table A.1 Comparison of occurrence probability of each reason ( $\alpha = 1.00, \beta = 1.00$ ).

	$\frac{P(r_i z_1)}{P(r_i)}$	$\frac{P(r_i z_2)}{P(r_i)}$	$\frac{P(r_i z_3)}{P(r_i)}$	$\frac{P(r_i z_4)}{P(r_i)}$	$\frac{P(r_i z_5)}{P(r_i)}$
仕事内容	<b>1.11</b>	<b>1.19</b>	0.77	0.90	<b>1.08</b>
業界	0.36	<b>1.07</b>	<b>1.22</b>	<b>1.11</b>	0.96
勤務地	0.91	<b>1.54</b>	0.91	0.78	0.84
職場雰囲気	0.92	<b>2.91</b>	0.31	0.37	0.58
ビジョン	0.82	<b>3.21</b>	0.40	0.33	0.34
ステータス	0.50	<b>1.83</b>	0.83	<b>1.17</b>	0.62
経験の活用	0.76	<b>2.12</b>	0.29	0.72	<b>1.06</b>
商品の内容	0.59	<b>2.08</b>	0.60	<b>1.04</b>	0.68
事業戦略	0.42	<b>3.88</b>	0.14	0.22	0.32
事業成長性	0.24	<b>3.79</b>	0.21	0.38	0.28
給与水準	0.57	<b>3.37</b>	0.54	0.37	0.17
職場環境	0.36	<b>4.05</b>	0.19	0.25	0.13
福利厚生	0.60	<b>3.60</b>	0.40	0.33	0.12
勤務形態	0.44	<b>3.74</b>	0.41	0.31	0.09
教育制度	0.37	<b>4.18</b>	0.06	0.16	0.21
能力主義	0.24	<b>4.39</b>	0.08	0.05	0.16
昇進の早さ	0.21	<b>4.46</b>	0.09	0.05	0.1
自己成長性	0.6	<b>2.89</b>	0.29	0.46	0.71
内定期待度	<b>1.22</b>	<b>2.71</b>	0.43	0.34	0.51
社会貢献度	0.25	<b>3.39</b>	0.34	0.55	0.36

表 A・2 各志望理由の生起確率の比較 ( $\alpha = 1.50, \beta = 1.00$ )

Table A.2 Comparison of occurrence probability of each reason ( $\alpha = 1.50, \beta = 1.00$ ).

	$\frac{P(r_i z_1)}{P(r_i)}$	$\frac{P(r_i z_2)}{P(r_i)}$	$\frac{P(r_i z_3)}{P(r_i)}$	$\frac{P(r_i z_4)}{P(r_i)}$	$\frac{P(r_i z_5)}{P(r_i)}$
仕事内容	<b>1.01</b>	<b>1.14</b>	0.97	0.98	0.94
業界	0.97	0.84	<b>1.02</b>	<b>1.08</b>	<b>1.05</b>
勤務地	<b>1.35</b>	<b>1.14</b>	0.94	0.73	<b>1.01</b>
職場雰囲気	<b>1.61</b>	0.67	<b>1.02</b>	0.93	0.98
ビジョン	<b>1.52</b>	0.39	<b>1.07</b>	<b>1.05</b>	<b>1.07</b>
ステータス	0.56	0.56	<b>1.26</b>	<b>1.01</b>	<b>1.21</b>
経験の活用	<b>1.34</b>	0.52	<b>1.13</b>	0.84	<b>1.16</b>
商品の内容	0.82	0.36	<b>1.21</b>	0.96	<b>1.33</b>
事業戦略	<b>1.22</b>	0.30	<b>1.07</b>	<b>1.21</b>	<b>1.15</b>
事業成長性	<b>1.34</b>	0.43	<b>1.11</b>	0.98	<b>1.13</b>
給与水準	<b>1.32</b>	0.93	<b>1.10</b>	0.88	0.85
職場環境	<b>1.92</b>	0.52	1.00	0.97	0.93
福利厚生	<b>1.65</b>	0.97	<b>1.05</b>	0.75	0.84
勤務形態	<b>1.81</b>	<b>1.07</b>	1.00	0.74	0.75
教育制度	<b>2.12</b>	0.50	0.97	0.87	0.96
能力主義	<b>2.21</b>	0.39	0.88	0.98	<b>1.02</b>
昇進の早さ	<b>1.98</b>	0.40	0.97	<b>1.17</b>	0.88
自己成長性	<b>1.41</b>	0.58	<b>1.01</b>	<b>1.05</b>	<b>1.06</b>
内定期待度	<b>1.48</b>	0.91	<b>1.03</b>	0.75	0.99
社会貢献度	<b>1.27</b>	0.63	<b>1.20</b>	0.78	<b>1.06</b>



坂元 哲平

1993年生。2015年早稲田大学創造理工学部経営システム工学科卒業。現在、同大学大学院創造理工学研究科経営システム工学専攻在学。機械学習を用いたデータ分析に関する研究に興味を持つ。



山下 遥

1987年生。2010年東京理科大学理工学部経営工学科卒業。2012年慶應義塾大学大学院修士課程修了。2015年同大学大学院博士課程修了。博士（工学）。2015年早稲田大学創造理工学部助手。2017年より上智大学理工学部助教。品質管理，統計学，情報工学を融合させた新たなデータ解析方法に関する研究に従事。応用統計学会，日本経営工学会，日本品質管理学会等，各会員。



荻原 大陸

1989年生。2014年早稲田大学大学院修士課程修了。2014年株式会社リクルートキャリア入社。就職支援サイトの企画職として，機械学習を用いたサービス開発に従事。



後藤 正幸（正会員）

1969年生。1994年武蔵工業大学大学院修士課程修了。2000年早稲田大学博士課程修了。博士（工学）。1997年早稲田大学理工学部助手。2000年東京大学大学院工学系研究科助手。2002年武蔵工業大学環境情報学部助教授。2008年早稲田大創造理工学部経営システム工学科准教授。2011年同大教授。情報数理応用とデータサイエンスの研究に従事。著書に、『入門パターン認識と機械学習』，コロナ社（2014），『ビジネス統計 統計基礎とエクセル分析』，オデッセイコミュニケーションズ（2015）等。IEEE，電子情報通信学会，人工知能学会，日本経営工学会，経営情報学会等，各会員。