

A Study of Prediction of Number of Students' Applications on Internet Portal Site for Job-Hunting Based on Mixture of Regression Model

NAGAMORI Seiya

1 研究背景・目的

日本の大学生の就職活動において、近年、就職ポータルサイトの活用が一般的なものとなりつつある。企業は、効率的な採用活動のために就職ポータルサイトを通じて学生に求人情報を提供している。就職ポータルサイトを用いて採用活動を行うことで多くの学生に情報を提供できるようになり、その結果多くの学生からのエントリーを期待できることが企業側のメリットとなっている。そのため、就職ポータルサイトを活用しようとする企業は、就職ポータルサイト上での採用活動に対して獲得できる被エントリー数とその変動要因に関心があるといえる。

ここで、企業が獲得できる被エントリー数と、採用活動における行動の関係性を分析する最も基本的な手法として線形回帰モデル [1] が考えられる。しかしながら、被エントリー数と企業が行う採用活動の関係には就職ポータルサイトに顕在化している情報のみならず、企業の学生からの認知度や景気などの企業の潜在的要因も影響すると考えられ、単一の回帰モデルでは精度の高いモデルの推定が困難である。この問題に対し混合回帰モデル [2] を用いた要因分析が可能である。しかしながら、この混合回帰モデルは新たに就職ポータルサイトを利用しようとする企業の被エントリー数や採用活動のインパクトの予測に用いることができない。一方で、業種や従業員規模などの企業の基本情報は上記の関係性に影響を与える要因の一つであり、新規掲載企業に対し予測可能なモデルへの拡張に貢献するものと考えられる。

本稿では、新たに就職ポータルサイトを利用する企業に対する予測を可能とする混合回帰モデルをベースとした新たな予測モデルの提案を行う。この提案モデルを用いて新規掲載企業に対しても被エントリー数や採用活動のインパクトの予測を可能とすることを目的とする。その際に、企業の基本情報を潜在クラスに反映させることで、予測が可能となり、汎化能力の高いモデルが得られると期待できる。すなわち、本稿の提案モデルは、企業の基本情報を考慮したうえで、企業の採用活動における行動情報と被エントリー数の関係性を表現し、獲得可能な被エントリー数と採用活動における行動情報のインパクトの予測を可能とする。本研究の提案モデルは、補助変数を用いた回帰モデルの混合に関する研究の枠組みの中で、文献 [3] の一般モデルを拡張した特殊なケースとなっている。また実データに対して提案するモデルを適用しモデルの評価を行い、推定されたモデルを分析することで有用な知見が得られることを示す。

2 準備

2.1 混合回帰モデル

混合回帰モデル [2] とは目的変数 y と説明変数 \mathbf{x} の線形構造の背後に潜在クラスを仮定したモデルである。このモデルはそれぞれの潜在クラスに対し回帰モデルを仮定しており、それらの混合により表現される。 K 個の潜在クラスを仮定したとき、混合回帰モデルのモデル式は以下の式 (1), (2) で表現される。

$$h(y|\mathbf{x}, \boldsymbol{\psi}) = \sum_{k=1}^K \pi_k g_k(y|\mathbf{x}, \boldsymbol{\theta}_k) \quad (1)$$

$$\pi_k \geq 0 \quad \text{and} \quad \sum_{k=1}^K \pi_k = 1 \quad (2)$$

ここで、 $\boldsymbol{\psi}$ は混合回帰モデルのすべてのパラメータを表すベクトルであり、 π_k は混合割合、 $g_k(\cdot)$ は平均 $\boldsymbol{\beta}_k^T \mathbf{x}$ 、分散 σ_k^2 の正規分布を示し、 $\boldsymbol{\theta}_k = (\boldsymbol{\beta}_k^T, \sigma_k^2)^T$ である。

混合回帰モデルの拡張モデルとして、補助変数を用いたモデル [3] が提案されている。このモデルでは混合回帰モデルの混合割合が補助変数 \mathbf{v} に依存するモデルとなっている。補助変数を用いた混合回帰モデルは以下の式 (3), (4) で表現される。

$$h(y|\mathbf{x}, \mathbf{v}, \boldsymbol{\phi}) = \sum_{k=1}^K \pi_k(\mathbf{v}, \boldsymbol{\alpha}) g_k(y|\mathbf{x}, \boldsymbol{\theta}_k) \quad (3)$$

$$\pi_k(\mathbf{v}, \boldsymbol{\alpha}) \geq 0 \quad \text{and} \quad \sum_{k=1}^K \pi_k(\mathbf{v}, \boldsymbol{\alpha}) = 1 \quad (4)$$

ここで、 $\boldsymbol{\phi}$ は補助変数を用いた混合回帰モデルのすべてのパラメータを表すベクトルであり、 $\boldsymbol{\alpha}$ は補助変数に対するパラメータである。このモデルの混合割合 $\pi_k(\mathbf{v}, \boldsymbol{\alpha})$ は制約式 (4) を満たすように関数を設定すればよい。

2.2 被エントリー数に対する企業行動モデルの定式化

本研究では企業の行動情報と被エントリー数の関係性をモデル化している。ここでいう行動情報とは、就職ポータルサイト上で行われる企業のインターンシップ募集の有無など、企業が選択実施可能な採用活動オプションのことを指す。上記のモデル化を達成するための最も基本的なモデルは重回帰モデルである。しかし被エントリー数と行動情報の関係性には企業の特徴により異なる構造が混在していると考えられる。そのため単一の重回帰モデルでは企業ごとの採用における行動情報と被エントリー数の関係性の違いを表現することができない。被エントリー数の増減に影響する効果は異なるものと考えられ、これは企業の持っている基本情報や外的要因などの潜在的要因によって決められると考えられる。

そこで本研究では、企業の行動情報と被エントリー数の関係性が類似した企業群は同じ潜在クラスに所属し、同じ回帰式が当てはまることを仮定した混合回帰モデル [2] を導入する。これにより個々の企業の混在の特徴を考慮して被エントリー数と採用活動における行動との関係性を分析することが可能となる。また、適切な企業群の潜在クラスを確率的に推定することで、より推定精度が高く、説明能力の高い回帰モデルが構築され、より正確な解釈を与えることを可能とする。

いま、 K 個の潜在クラス集合を $\mathcal{Z} = \{z_k : 1 \leq k \leq K\}$ としたとき、混合回帰モデルは各潜在クラスにおける確率密度関数 $P_k(y_l|\mathbf{x}_l)$ の線形結合によりモデル化される。ここで、 L 社の企業のうち l 番目の企業の行動情報

を表す説明変数ベクトルは $\mathbf{x}_l = (x_{l0}, x_{l1}, x_{l2}, \dots, x_{lI})^T$, 目的変数である被エントリ数は y_l である。また潜在クラス z_k における $I + 1$ 個の回帰モデルのパラメータを $\beta_k = (\beta_{0k}, \beta_{1k}, \beta_{2k}, \dots, \beta_{Ik})^T$ としたとき、混合回帰モデルは以下の式 (5) で示される。ただし, $x_{l0} = 1$ とする。

$$P(y_l|\mathbf{x}_l) = \sum_{k=1}^K w_{lk} P_k(y_l|\mathbf{x}_l) \quad (5)$$

ここで, w_{lk} は k に関する和が 1 となる第 l 企業の各潜在クラスへの重みであり, 潜在クラス z_k ごとに仮定される y_l の確率密度関数は分散を σ_k^2 としたとき, 次式で表される。

$$P_k(y_l|\mathbf{x}_l) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left\{-\frac{(y_l - f_k(\mathbf{x}_l))^2}{2\sigma_k^2}\right\} \quad (6)$$

$$f_k(\mathbf{x}_l) = \sum_{i=0}^I \beta_{ik} x_{li} \quad (7)$$

混合回帰モデルのグラフィカルモデルは図 1 で表される。

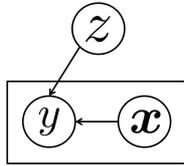


図 1: 混合回帰モデルのグラフィカルモデル

混合回帰モデルは, EM アルゴリズム [4] によって学習可能である。混合回帰モデルの β_k のパラメータ推定は潜在クラス z_k に対して大きい重みをもつ企業を重点的に学習し, 企業の特徴を回帰モデルのパラメータに反映させることができる。これは学習データの目的変数に対する推定精度を向上させるようなパラメータ推定となっている。すなわち, 行動情報 \mathbf{x}_l と被エントリ数 y_l の組み合わせで潜在クラスが構築されるため, これから就職ポータルサイトをしようとしている (被エントリ数 y_l のデータがない) 新規企業に対しては単純に予測を行うことができない。

この問題に対し, 補助変数を用いた手法が適用可能である。文献 [3] では補助変数を潜在クラスに反映させるモデルの一般式が与えられている。しかし具体的なモデルの例としては, 式 (4) の混合割合が補助変数を用いたロジットモデルで与えられるケースが示されているのみである。一方, 補助変数として複数の離散データの基本情報を想定する本研究では, 企業の特徴は基本情報の組み合わせに依存すると考えられる。そこで, 本提案モデルは基本情報それぞれに対して多項分布を仮定し, 基本情報間の交互作用を考慮したモデル化を行う。次節以降, 混合回帰モデルを基礎とし, 扱う問題に適した形で被エントリ数の予測モデルへの拡張を行う。

3 基本情報を考慮した予測モデル (提案モデル)

企業の行動情報と獲得できる被エントリ数の関係性をモデル化する際に, 同時に新たな企業が就職ポータルサイトをしようとしていることで獲得できる被エントリ数を予測することが望ましい。また, 就職ポータルサイトをしようとする企業は被エントリ数を向上させる上で効果的な採用活動における行動に関心があり, その効果を定量的に判断できることが求められる。

そこで, 本節では, 学習データにおける行動情報 \mathbf{x}_l と被エントリ数 y_l の関係性を表す回帰モデルと企業の基本情報 \mathbf{d}_l により潜在クラスを構築するモデルを提案する。このモデルは回帰式により表現される被エントリ数と行

動情報の関係性と, 企業の基本情報の共起を表しており, 基本情報からその企業の潜在クラスへの所属確率を推定することができる。すなわち, 潜在クラスに基本情報の特徴を反映させることで, 被エントリ数が学習データにない新規企業に対しても, 潜在クラスへの重みを推定することが可能となる。

これにより, 就職ポータルサイトを利用する新たな企業に対して, その企業の基本情報を用いることで被エントリ数を予測することが可能となる。さらに, 混合回帰モデルのパラメータを分析することで, 採用活動における行動情報が与える被エントリ数への影響も定量的に把握することができる。

3.1 定式化

いま, 企業に関する j 番目の基本情報 ($1 \leq j \leq J$) の要素集合を $\mathcal{D}^j = \{d_{n_j}^j : 1 \leq n_j \leq N_j\}$, $d_{n_j}^j$ を j 番目の基本情報の n_j 番目の要素, N_j を j 番目の基本情報の要素数とし, l 番目の企業の基本情報を表す変数ベクトルを $\mathbf{d}_l = (d_{l1}, d_{l2}, \dots, d_{lJ})^T$, d_{lj} を l 番目の企業の j 番目の基本情報の要素とする。このとき提案する確率モデルは, 式 (8) で示される。

$$P(y_l, \mathbf{x}_l, \mathbf{d}_l) = \sum_{k=1}^K P(z_k) P_k(y_l|\mathbf{x}_l) \prod_{j=1}^J \prod_{n_j=1}^{N_j} P(d_{n_j}^j|z_k)^{\delta(d_{n_j}^j, d_{lj})} \quad (8)$$

ここで, $\delta(a, b)$ は a と b が一致していれば 1, さもなくば 0 とする指示関数とする。提案モデルのグラフィカルモデルは図 2 で表される。

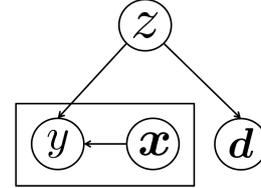


図 2: 提案モデルのグラフィカルモデル

3.2 パラメータの推定方法

提案モデルのパラメータを EM アルゴリズムを用いて推定する方法を示す。学習データに対する対数尤度関数 LL は以下の式 (9) のように示される。

$$LL = \sum_{l=1}^L \log P(y_l, \mathbf{x}_l, \mathbf{d}_l) \quad (9)$$

EM アルゴリズムは対数尤度を最大化するパラメータを E-step と M-step の繰り返し計算を行うことによって求める。以下に, 提案モデルのパラメータである w_{lk} , $P(z_k)$, σ_k^2 , β_k , $P(d_{n_j}^j|z_k)$ を EM アルゴリズムを用いて推定する方法を示す。ここでは, 「 w_{lk} の推定」と 「 $P(z_k)$, σ_k^2 , β_k , $P(d_{n_j}^j|z_k)$ の推定」を繰り返すことでパラメータの学習を行う。

【E-step】

まず E-step では以下の式 (10) で w_{lk} が計算され更新される。

$$w_{lk} = \frac{P(z_k) P_k(y_l|\mathbf{x}_l) \prod_{j=1}^J \prod_{n_j=1}^{N_j} P(d_{n_j}^j|z_k)^{\delta(d_{n_j}^j, d_{lj})}}{\sum_{k=1}^K P(z_k) P_k(y_l|\mathbf{x}_l) \prod_{j=1}^J \prod_{n_j=1}^{N_j} P(d_{n_j}^j|z_k)^{\delta(d_{n_j}^j, d_{lj})}} \quad (10)$$

【M-step】

次に, M-step では w_{lk} を固定した元で, 各パラメータを更新する。混合割合 $P(z_k)$ および各潜在クラス z_k における分散 σ_k^2 はそれぞれ式 (11) および式 (12) で更新される。

$$P(z_k) = \frac{\sum_{l=1}^L w_{lk}}{L} \quad (11)$$

$$\sigma_k^2 = \frac{\sum_{l=1}^L w_{lk}(y_l - f_k(\mathbf{x}_l))^2}{\sum_{l=1}^L w_{lk}} \quad (12)$$

これらの更新式では各企業を K 個の潜在クラスに確率的に所属させ、その重みを用いて各潜在クラスで回帰モデルを構築することを考えている。潜在クラス z_k におけるパラメータ β_k は、式 (13) を用いて更新する。

$$\beta_k = \arg \min_{\beta_k} \sum_{l=1}^L w_{lk}(y_l - f_k(\mathbf{x}_l))^2 \quad (13)$$

企業の基本情報に関するパラメータについては以下の式 (14) で更新する。

$$P(d_{n_j}^j | z_k) = \frac{\sum_{l=1}^L \delta(d_{n_j}^j, d_{lj}) w_{lk}}{\sum_{l=1}^L w_{lk}} \quad (14)$$

3.3 新規データに対する被エントリ数の予測

本モデルは新規企業に対して被エントリ数の予測が可能である。企業の特徴は基本情報の組み合わせで表現できると考えられる。ここで、 $\mathbf{d}'_t = (d'_{t1}, d'_{t2}, \dots, d'_{tJ})^T$ を t 番目 ($t = 1, 2, \dots, T$) の予測対象企業の基本情報を表す変数ベクトル、 \hat{y}_t を t 番目の予測対象企業の被エントリ数の予測値とすると、基本情報から潜在クラスへの所属確率が計算可能であり、それぞれの潜在クラスでの回帰モデルを混合することで新規企業に対して被エントリ数を予測する。ここで、 \hat{w}_{tk} を予測対象企業の潜在クラスへの所属確率の予測値、 $\hat{\beta}_{ik}$ を回帰パラメータの推定値とすると、 \hat{w}_{tk} および被エントリ数の予測値 \hat{y}_t は、以下の式 (15) および式 (16) で推定される。

$$\hat{w}_{tk} = \frac{P(z_k) \prod_{j=1}^J \prod_{n_j=1}^{N_j} P(d_{n_j}^j | z_k)^{\delta(d_{n_j}^j, d'_{tj})}}{\sum_{k=1}^K P(z_k) \prod_{j=1}^J \prod_{n_j=1}^{N_j} P(d_{n_j}^j | z_k)^{\delta(d_{n_j}^j, d'_{tj})}} \quad (15)$$

$$\hat{y}_t = \sum_{k=1}^K (\hat{w}_{tk} \sum_{i=0}^I \hat{\beta}_{ik} x'_{ti}) \quad (16)$$

ここで、 T 社の予測対象企業のうち t 番目の企業の行動情報を表す説明変数ベクトルを $\mathbf{x}'_t = (x'_{t0}, x'_{t1}, x'_{t2}, \dots, x'_{tI})^T$ とし、 $x'_{t0} = 1$ とする。

3.4 提案モデルのアルゴリズム

提案モデルは以下のアルゴリズムで構築される。

- Step1** 各パラメータの初期値をランダムに設定する。
- Step2** E-step: w_{lk} を式 (10) を用いて推定する。
- Step3** M-step: $P(z_k)$, σ_k^2 , β_k および $P(d_{n_j}^j | z_k)$ を式 (11) ~ 式 (14) を用いて更新する。
- Step4** 収束条件を満たしていれば Step5 へ。さもなければ Step2 にもどる。
- Step5** 新規データに対して式 (15) を用いて潜在クラスへの重みを推定し、式 (16) を用いて被エントリ数の予測を行う。

□

4 提案モデルの実データへの適用

4.1 提案モデルの評価実験

本節では、実データを用いて提案モデルの推定を行い、学習データに対しての当てはまりと予測対象企業に対する予測精度の2つの観点から結果を考察する。学習データへの当てはまりが良いほど表現能力が高いモデルと言えるが、一方、予測精度が高いほど汎化能力の高いモデルが得られていると判断できる。

4.1.1 実験条件

実験データとして、2014年度卒業の学生に対する就職ポータルサイト上で100件以上1000件以下の被エントリを獲得した企業約5000社 ($L \approx 5000$) を学習データ、2015年度卒業の学生に対する就職ポータルサイトで同様の被エントリを獲得した企業約5000社 ($T \approx 5000$) を予測対象データとして用いた。目的変数を各企業の被エントリ数とし、就職ポータルサイトに蓄積されているデータから利用可能な企業の採用活動における4つの行動オプションを説明変数として用いた ($I = 4$)。また、基本情報として4変数を用いた ($J = 4$)。潜在クラス数 K は2~10, 15, 20として実験を行った。比較モデルとして潜在クラスモデルの一つである Aspect Model を多変量に拡張したモデルを用いて企業の基本情報で潜在クラスを推定し、それぞれに回帰モデルを構築するモデル (AM+回帰) と潜在クラスや企業の基本情報を用いない単一の回帰モデル (単一回帰) の結果も示す。この実験ではそれぞれのモデルで、各学習データの企業に対しては被エントリ数の推定を行い、算出された推定値と実測値の平均二乗誤差により推定精度を評価した。また各予測対象データの企業に対しては被エントリ数の予測を行い、算出された予測値と実測値の平均二乗誤差を評価することにより、予測精度を評価した。また異なる初期値で100回実験を繰り返し、それらの平均を用いて評価を行う。

4.1.2 結果と考察

実験結果を表1に示す。

表1. 平均二乗誤差の比較結果

K	提案モデル		AM+回帰		単一回帰	
	学習	予測	学習	予測	学習	予測
2	54238.0	60010.8	59326.3	60800.8	59836.3	61148.1
3	28091.1	55607.4	58249.6	59759.4		
4	28647.7	55204.6	57909.5	59544.0		
5	27534.0	55017.3	57461.5	59119.1		
6	26169.1	54943.6	57160.2	58806.6		
7	25278.0	54947.0	57018.2	58718.5		
8	22386.3	55084.1	56842.5	58612.4		
9	20928.9	55235.1	56694.9	58617.7		
10	20789.2	55612.2	56570.1	58593.2		
15	16444.5	56833.0	55926.2	58545.9		
20	14401.6	58090.0	55469.9	58640.6		

表1より複数の回帰モデルを混合することによるモデルの推定精度の向上が確認できる。この結果より被エントリ数と企業の行動情報の関係性の構造は複数存在し、潜在クラスを仮定し混在的要因を考慮しているモデルの方が本データの分析に適していることが示唆される。

また表1より、提案モデルでは各潜在クラス数において学習データへの当てはまり、予測精度の観点から優れたモデルが推定されていることがわかる。提案モデルでは学習データに対する当てはまりを良くする混合回帰モデルを推定すると同時に企業の基本情報のクラスタリングが適切に行われたことで予測精度が向上したと考えられる。すなわち、企業の基本情報に加え、行動情報の被エントリ数への効果を考慮した新たな企業クラスタリングが可能となっている。

また潜在クラス数の増加に伴い、学習データへの当てはまりが向上していく一方、ある K 以上では予測精度の低下が見られる。これは潜在クラス数に応じて、パラメー

タ数が増加することで学習データへの過度なフィッティングが起きていると考えられる。本提案モデルを適用する際には目的に応じてモデルのパラメータ数、学習データ数を考慮し、適切に潜在クラス数を決定する必要がある。

4.2 提案モデルを用いた分析

本節では構築された提案モデルの応用として実データを用いて結果の分析を行う。潜在クラスに着目した分析と各企業に対する分析に焦点を当てる。本稿では提案モデルにおいて最も良い予測精度結果となった潜在クラス数 $K = 6$ のときに推定されたパラメータを用いて分析を行うこととする。分析データは前節と同様のデータを用いている。

4.2.1 潜在クラスに着目した分析

提案モデルの各潜在クラスにおける回帰モデルのパラメータ推定値を表2に示す。表2における $\hat{P}(z_k)$ は混合割合の推定値を示す。

表2. 提案モデルによる分析結果 (1)

	z_1	z_2	z_3	z_4	z_5	z_6
$\hat{P}(z_k)$	0.10	0.21	0.21	0.19	0.11	0.18
$\hat{\beta}_{0k}$	177.65	218.73	411.70	462.90	513.88	615.85
$\hat{\beta}_{1k}$	12.22	8.26	68.91	25.01	-15.82	40.24
$\hat{\beta}_{2k}$	52.41	19.95	160.83	104.91	61.75	68.24
$\hat{\beta}_{3k}$	31.93	65.28	148.78	129.31	105.59	49.00
$\hat{\beta}_{4k}$	23.93	47.21	99.23	83.26	116.09	58.81

表2より、各潜在クラスにおいて異なる特徴が抽出されていることがわかる。推定されたパラメータは潜在クラスごと、行動ごとに異なり、まずそれぞれの行動が効果的な潜在クラスは異なることがわかる。例えば、行動1, 2, 3に関しては、潜在クラス3が最も効果的な潜在クラスであるが、行動4に関しては潜在クラス5が最も効果的な潜在クラスである。次に、それぞれの行動ごとに被エンタリ数への効果の大きさが異なることがわかる。例えば推定されたパラメータのばらつきから行動2は他の行動と比較し、潜在クラスごとの実施の効果が大きく異なるが、行動1は潜在クラスごとに実施の効果の変動が小さいといえる。また、各潜在クラスが異なる特徴を持っていることがわかる。例えば、潜在クラス5に関しては、他の潜在クラスと比較して行動1の効果が最も低い潜在クラスであるのに対して、行動4の効果が最も高い潜在クラスである。以上のような分析からも、企業の混在的特徴を潜在クラスによって表現できていることが示唆される。

次に各潜在クラスにおける行動を起こしている割合に着目し、結果を表3に示す。

表3. 提案モデルによる分析結果 (2)

	z_1	z_2	z_3	z_4	z_5	z_6
行動1	0.798	0.817	0.754	0.822	0.752	0.743
行動2	0.009	0.037	0.013	0.010	0.044	0.022
行動3	0.024	0.034	0.032	0.033	0.073	0.050
行動4	0.573	0.678	0.676	0.659	0.673	0.710

表3より、潜在クラスごとに行動を起こしている割合が異なることがわかる。また表2と合わせて考察することで各潜在クラスにおける行動に効果があるかどうかを解釈することが可能である。例えば、潜在クラス6は他の潜在クラスと比較し行動1を起こす割合が最も低い潜在クラスであるが、行動1の効果は比較的高い潜在クラスである。よって、潜在クラス6に所属する企業は行動1に対してより積極的な行動を行うことで被エンタリ数の効果的な獲得が期待される。

4.2.2 各企業に着目した分析

本モデルでは各企業に対して採用活動における行動の効果量を定量化することが可能である。行動に対する t 番目の予測対象企業の効果ベクトルを $\beta_t^* = (\beta_{t1}^*, \beta_{t2}^*, \dots, \beta_{tK}^*)^T$ とすると行動の効果の定量化は以下の式 (17) で推定され

る。また、学習データの企業に対しても同様に推定可能である。

$$\beta_t^* = \left(\sum_{k=1}^K \hat{w}_{tk} \hat{\beta}_{1k}, \sum_{k=1}^K \hat{w}_{tk} \hat{\beta}_{2k}, \dots, \sum_{k=1}^K \hat{w}_{tk} \hat{\beta}_{Kk} \right)^T \quad (17)$$

この定量化式により、各企業に対して特徴に応じ、個々に行動の効果量を推定することができる。

ここで、例として2つの企業に着目し分析を行った。着目した2企業は提案モデルによって以下の表のように企業から潜在クラスへの重み \hat{w}_{tk} が推定された。

表4. 提案モデルにより推定された重み

	\hat{w}_{t1}	\hat{w}_{t2}	\hat{w}_{t3}	\hat{w}_{t4}	\hat{w}_{t5}	\hat{w}_{t6}
企業A	0.00	0.29	0.00	0.00	0.31	0.40
企業B	0.76	0.00	0.00	0.24	0.00	0.00

表4より企業Aは潜在クラス2, 5, 6に比較的大きな重みを持っている企業であり、企業Bは潜在クラス1に大きな重み、潜在クラス4に小さな重みを持っている企業である。これは企業Aは潜在クラス2, 5, 6の特徴を有しており、企業Bは潜在クラス1, 4の特徴を有した企業であるといえる。この特徴が異なる2企業に対し、式(17)を用いることにより行動の効果量を以下のように計算できる。

表5. 提案モデルにより計算された行動の効果

	β_{t1}^*	β_{t2}^*	β_{t3}^*	β_{t4}^*
企業A	13.65	52.31	71.44	73.17
企業B	15.33	65.17	55.59	38.35

表5より着目した2企業はそれぞれ行動に対する効果が異なることがわかる。この2企業を比較すると行動1, 2に関しては企業Bの方が企業Aより効果的な行動である一方で行動3, 4に関しては企業Aの方が企業Bより効果的な行動であるといえる。

この分析モデルの活用により、被エンタリ数の推定、また行動を起こした際の効果を定量的に推定可能であり、どのように採用活動を変化させていけば、被エンタリ数を向上させることができるかという観点から採用活動の計画を立てるための一助となることが期待される。

5 まとめと今後の課題

本研究では、企業の採用における行動情報と被エンタリ数の関係性を混合回帰モデルを基に基本情報を考慮することで、被エンタリ数を予測可能なモデルとして提案した。提案したモデルの有効性を示すために就職ポータルサイトの実データを用いて実験を行った。また推定されたモデルを用いて実際のエンタリデータの分析を行い、有用な知見が得られることを示した。今後の課題として、予測精度の向上、具体的な企業の採用における行動計画のサポート手法の検討、潜在クラス数の決定方法の検討などが挙げられる。

参考文献

- [1] Bishop, C. M., "Pattern Recognition and Machine Learning." Springer, 2006.
- [2] Faria, S., and Soromenho, G., "Fitting Mixtures of Linear Regressions." *Journal of Statistical Computation and Simulation*, Vol. 80, No. 2, pp.201-225, 2010.
- [3] Grun, B., and Leisch, F., "FlexMix Version 2: Finite Mixtures with Concomitant Variables and Varying and Constant Parameters." *Journal of Statistical Software*, Vol. 28, Issue 4, pp. 1-35, 2008.
- [4] Dempster, A. P., Laird, N. M., and Rubin, D. B., "Maximum Likelihood from Incomplete Data via the EM Algorithm." *J. Royal Statistical Society. Series B*, Vol. 39, No. 1, pp. 1-38, 1977.