ポアソン混合効果モデルを用いた就職ポータルサイトにおける被エントリ数の予測モデルの構築に関する研究

情報数理応用研究

5214C030-8　野津琢登
指導教員　後藤正幸

A Study on Prediction Model of the Number of Applications on Internet Portal Sites for Job Hunting Using Poisson Mixed-effects Models

Takuto Notsu

1　研究背景・目的

近年，就職ポータルサイトを利用して就職活動を行う学生の割合は高まっており，多くの企業もまたその利用により採用活動を行っている．企業は就職ポータルサイトを有効活用することで，自社の情報を広く学生側に伝えることができると共に，学生からの多くのエントリの確保に結びつくと考えられる．そのため，掲載企業にとってどの程度の被エントリ数が期待できるか，加えてどのような企業行動 (就職ポータルサイト上でインターンシップや説明会の実施情報を掲載するといった行動) が被エントリ数増加に影響を与えるかは大きな関心対象となる．

そこで本研究では，被エントリ数の予測モデルの構築，ならびに就職ポータルサイト上での企業行動と被エントリ数の関係性のモデル化を研究対象とする．

しかし，被エントリ数の予測モデルの構築をする上で，以下の 2 つの問題がある．一点目は，就職ポータルサイト以外の顕在化されていない外部要因が予測に悪影響を及ぼす可能性があることである．二点目は，変数選択，すなわちモデル選択の問題である．一点目の問題では，被エントリ数には，その企業のブランド力や認知度といった就職ポータルサイト以外の顕在化されていない外部要因が大きく影響することが考えられる．そのため，就職ポータルサイト上での企業行動のみから被エントリ数を正確に説明しようとすることは極めて困難であり，外部要因による変動を予測モデルに組み込む必要がある．また，企業行動が被エントリ数に与える影響は，従業員規模や業種，企業ごとに異なると考えられるため，より正確な予測のためには企業ごとに回帰パラメータを推定する必要がある．二点目の問題では，一般に回帰モデル構築する上で，変数選択は重要な問題であり，一度に多くの説明変数を利用すると、過学習などの問題が生じる可能性が高くなる．さらに，目的を予測に限定した場合，単一のモデルを選択することが予測の面から最適であるとは限らず，より優れた予測法が存在する可能性がある．

本研究では，上記の一点目の問題に対して，ポアソン混合効果モデル [1], [2] を導入することで解決を図る．ポアソン混合効果モデルを用いて，説明変数に企業の行動情報を設定し，企業間の差や業種間の差を表すランダム効果を導入することによって，各企業ごとに回帰パラメータを算出でき，就職ポータルサイト上での企業行動，外部要因の双方を考慮した被エントリ数の予測モデルを構築できる．二点目の問題に対しては，少数の説明変数ごとで複数の回帰モデルを構築し，それらを混合する手法を提案する．この提案手法は，単一のモデルを選択せず，各回帰モデルの構築段階では，変数選択をせずに，モデル構築に用いることのできる全ての説明変数から少数の説明変数を選択し，複数のモデルを構築した後に混合する．異なる少数の説明変数の組み合わせで構築した各回帰モデルを混合することにより，過学習のリスクの軽減にも繋がると考えられる．

これらの 2 つの手法を組み合わせ，本研究では複数のポアソン混合効果モデルを少数の説明変数を用いて構築した後，それらを混合することによる被エントリ数の予測モデルの構築法を提案する．提案手法の有効性を検証するため，当該サイトに蓄積されている実データを用いて実験を行う．加えて，提案モデルの応用法として，構築された予測モデルを分析することで企業行動と被エントリ数の増加の関係性が明らかになることを示す．

2　準備

本研究では，ポアソン混合効果モデルを基に被エントリ数の予測モデルを構築する．以下では準備として，ポアソン回帰モデル [2]，ポアソン混合効果モデル [1], [2] とそのパラメータ推定方法 [3] について述べる．

2.1　ポアソン回帰モデル

ポアソン回帰モデルは，目的変数が交通事故件数などの非負整数値をとる計数データである場合の要因系の説明変数との関係を分析するための回帰モデルである．

いま，説明変数が g 個，データが N 組ある場合を考える．i 番目のデータの説明変数ベクトル $\boldsymbol{x}_i = (1, x_{i1}, x_{i2}, \cdots, x_{ig})^{\mathrm{T}} \in \mathcal{R}^{g+1}$ に対する目的変数を $y_i \in \mathcal{Z}^+$，$\boldsymbol{\xi} = (\xi_0, \xi_1, \xi_2, \cdots, \xi_g)^{\mathrm{T}} \in \mathcal{R}^{g+1}$ を $g+1$ 個の回帰パラメータとする．ただし，$\mathcal{Z}^+$ は非負整数の集合を表す．

このとき，目的変数 y_i が式 (1) で表されるポアソン分布に従うと仮定したもとで，その平均 λ_i を式 (2) の非線形モデルとして表す．

$$p(y_i; \lambda_i) = \frac{\lambda_i^{y_i} \exp(-\lambda_i)}{y_i!} \tag{1}$$

$$\lambda_i = \exp(\boldsymbol{\xi}^{\mathbf{T}} \boldsymbol{x}_i) \tag{2}$$

ここでパラメータ $\boldsymbol{\xi}$ は，式 (3) の尤度関数 $L(\lambda)$ を最大化する事で推定される．

$$L(\lambda) = \prod_{i=1}^{N} p(y_i; \lambda_i) = \prod_{i=1}^{N} \frac{\lambda_i^{y_i} \exp(-\lambda_i)}{y_i!} \tag{3}$$

一方，ポアソン回帰モデルでは，過分散となるデータに対して，当てはまりが悪くなるという問題がある．過分散とは，モデルの理論的な分散の値に対し，実際のデータの分散が大きくなることである．ポアソン分布は，期待値と分散値が等しいことを仮定しているが，本研究が対象とするデータでは期待値と比べ，分散の方が遥かに大きな値となる．過分散であるにもかかわらず，目的変数がポアソン分布に従っていると仮定して分析した場合，分散を実際よりも小さいものとして分析することに相当し，説明変数が目的変数の期待値に与える効果の過大評価に結びつく．この過分散を解決するための手法として，次節で示すポアソン混合効果モデルがある．

2.2　ポアソン混合効果モデル

2.2.1　概要

ポアソン混合効果モデルとは，ポアソン回帰モデルの過分散の問題を解決するため，未観測の個体間の差やグループ間の差を表すランダム効果を組み込んだ回帰モデルである．

目的変数 y_i が式 (1) で表されるポアソン分布に従うと仮定したもとで，式 (4) のようにポアソン分布の平均 λ_i に対して，観測されていないグループ間差を表すパラメータ $r_{j(i)}$ を加える．

$$\lambda_i = \exp\left(\boldsymbol{\xi}^{\mathbf{T}}\boldsymbol{x}_i + r_{j(i)}\right) \tag{4}$$

ただし，$j(i)$ は $i \in \{1,\cdots,N\}$ 番目のデータが属するグループ $j \in \{1,\cdots,m\}$ のことを表している．このとき，$\boldsymbol{\xi}^{\mathbf{T}}\boldsymbol{x}_i$ を固定効果，グループ間差 $r_{j(i)}$ をランダム効果と呼ぶ．一般にランダム効果 $r_{j(i)}$ は，平均 0 で分散 s^2 の正規分布に従うと仮定する．このとき，確率密度関数 $p(r_{j(i)}; s^2)$ は式 (5) で表される．

$$p(r_{j(i)}; s^2) = \frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{r_{j(i)}}{2s^2}\right) \tag{5}$$

2.2.2 最尤法を用いたパラメータ推定

ポアソン混合効果モデルでは，$r_{j(i)}$ を積分消去することで式 (6) のように尤度 L_i を定義し，ランダム効果の分散 s^2 を推定する．このとき，ポアソン混合効果モデルの全データに対する尤度は式 (7) で表され，この対数尤度を最大化する $(\xi_0, \xi_1, \cdots, \xi_g, s^2)$ を求めることで，回帰モデルの導出を行う．

$$L_i = \int_{-\infty}^{\infty} p(y_i; \boldsymbol{\xi}, r_{j(i)}) p(r_{j(i)}; s^2)\, dr_{j(i)} \tag{6}$$

$$L(\xi_0, \xi_1, \cdots, \xi_g, s^2) = \prod_{i=1}^{N} L_i \tag{7}$$

2.2.3 階層ベイズ法を用いたパラメータ推定

ランダム効果が複数あるモデルの場合，式 (6) が多重積分を含むため，最尤推定ではパラメータ推定が困難になる．そこで分散 s^2 に対しても事前分布を導入し，階層ベイズ法によって各パラメータの事後分布を算出し，その事後平均を求めることによりパラメータの推定を行う．階層ベイズ法とは，統計モデルにおける確率分布のパラメータ自体が超パラメータにより規定される確率分布に従うという階層構造を持つベイズモデルの推定法である．パラメータの事後分布は，メトロポリス・ヘイスティングス法を用いて算出することができる．

3 ポアソン混合効果モデルを用いた予測モデルの構築

本研究では被エントリ数予測をするために，ポアソン混合効果モデルを用いて予測モデルを構築する．本節では就職ポータルサイトのデータに対するポアソン混合効果モデルの適用法を示す．また，ポアソン混合効果モデルを就職ポータルサイトの被エントリ数予測に用いることの有効性を検証するために，予備実験を行った．

3.1 概要

本研究では，就職ポータルサイト上での企業行動と被エントリ数の関係を明らかにするため，企業行動を用いて，被エントリ数の予測モデルを構築する．しかし，実際には就職ポータルサイト以外の外部要因が被エントリ数に影響していると考えられるため，これらの外部要因による変動もモデルに考慮する必要がある．そこで本研究では，ポアソン混合効果モデルを用いた予測モデルを構築する．観測することができない外部要因をモデルに反映するため，固定効果に企業の行動情報を，ランダム効果として，企業差，業種差，従業員規模差，株式公開の有無の差を表すパラメータを設定する．

3.2 回帰式の設定

本研究で設定する回帰モデルを以下の式 (8) で与える．

$$y_i \sim Po(\lambda_i) \tag{8}$$

$$\lambda_i = \exp(\boldsymbol{\beta_i}^{\mathbf{T}}\boldsymbol{x}_i) \tag{9}$$

いま，データ数を N，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 H 個の企業行動情報を $\{x_i^{(1)}, x_i^{(2)}, \cdots, x_i^{(H)}\}$ とする．ただし，$x_i^{(h)}$ は $h(h = 1,\cdots,H)$ 番目の行動情報に対し，その行動をとれば 1，そうでなければ 0 をとる 2 値変数とする．このとき，i 番目のデータの説明変数ベクトル $\boldsymbol{x}_i = (1, x_i^{(1)}, x_i^{(2)}, \cdots, x_i^{(H)})^{\mathrm{T}} \in \mathcal{R}^{H+1}$ に対する目的変数 y_i を被エントリ数とした．また $\boldsymbol{\beta}_i = (\beta_i^{(0)}, \beta_i^{(1)}, \beta_i^{(2)}, \cdots, \beta_i^{(H)})^{\mathrm{T}} \in \mathcal{R}^{H+1}$ を $H+1$ 個の回帰パラメータとし，$\beta_i^{(q)}$ $(q = 0, 1, 2, \cdots, H)$ を以下で与える．

$$\beta_i^{(q)} = \begin{cases} \eta_q + t_{k(i)}^{(q)} + r_{j(i)}^{(q)} + u_{l(i)}^{(q)} + v_{m(i)} & (q = 0) \\ \eta_q + t_{k(i)}^{(q)} + r_{j(i)}^{(q)} + u_{l(i)}^{(q)} & (q \neq 0) \end{cases} \tag{10}$$

$r_{j(i)}^{(q)}$ は企業ごと，$t_{k(i)}^{(q)}$ は業種ごと，$u_{l(i)}^{(q)}$ は従業員規模ごと，$v_{m(i)}$ は株式公開の有無ごとのばらつきを表すパラメータとする．ここで，$j(i)$ は i 番目のデータの企業 ID，$k(i)$ は i 番目のデータの業種 ID，$l(i)$ は i 番目のデータの従業員規模 ID，$m(i)$ は i 番目のデータの株式公開 ID を表す．$r_{j(i)}^{(q)}$，$t_{k(i)}^{(q)}$，$u_{l(i)}^{(q)}$ は式 (9) の切片項 $\beta_i^{(0)}$，係数項 $\{\beta_i^{(1)}, \beta_i^{(2)}, \cdots, \beta_i^{(H)}\}$ に影響を与える効果，$v_{m(i)}$ は切片項にのみ影響を与える効果である．$r_{j(i)}^{(q)}$，$t_{k(i)}^{(q)}$，$u_{l(i)}^{(q)}$，$v_{m(i)}$ は平均が 0，分散がそれぞれ $\sigma_{r^{(q)}}, \sigma_{t^{(q)}}, \sigma_{u^{(q)}}, \sigma_v$ の正規分布に従うとする．また，これらの分散の事前分布は，一般的に $[0, 10^4]$ の一様分布とされている [2], [4] ため，本研究でも同様にした．η_q は全データに対して共通の回帰パラメータであり，事前分布は平均 0，分散 10^2 の正規分布とする．

3.3 予備実験

ポアソン混合効果モデルを被エントリ数の予測に用いることの有効性，ならびに回帰モデルの説明変数に用いる最適な企業行動情報の設定を検証するために予備実験を行った．

3.3.1 実験条件

本実験では，本社所在地が東京かつ 50 件以上のエントリを獲得していた企業を対象とした．学習データは，2013 年度から 2015 年度にかけての 3 年間継続して掲載している企業のデータのうち，2013 年度と 2014 年度のデータとした．テストデータに，継続して就職ポータルサイトに掲載している企業の被エントリ数予測，新規掲載企業の被エントリ数予測をそれぞれ評価するため 2 つのデータセットを用いる．すなわち，テストデータ 1 として 2013 年度から 2015 年度にかけての 3 年間継続して掲載している企業のデータにおける，2015 年度の被エントリデータ，テストデータ 2 として，2015 年度に新規掲載された企業の被エントリデータを用意した．

また本研究で扱うことのできる企業行動情報は 5 つであり，これを行動 A～行動 E と呼ぶことにする．そこで，ポアソン混合モデルの有効性ならびに説明変数の数と予測精度の関係を検討するために，表 1 のように説明変数を設定したポアソン回帰モデル（M_0）とポアソン混合効果モデル（$M_1 \sim M_5$）との比較を行った．

表 1. 説明変数の設定

モデル名	説明変数
M_0	行動 A, 行動 B, 行動 C, 行動 D, 行動 E
$M_1(H=5)$	行動 A, 行動 B, 行動 C, 行動 D, 行動 E
$M_2(H=3)$	行動 A, 行動 D, 行動 E
$M_3(H=3)$	行動 A, 行動 B, 行動 D
$M_4(H=2)$	行動 A, 行動 B
$M_5(H=1)$	行動 A

3.3.2 実験結果と考察

実験結果を図 1 に示す．評価指標は平均二乗誤差である．

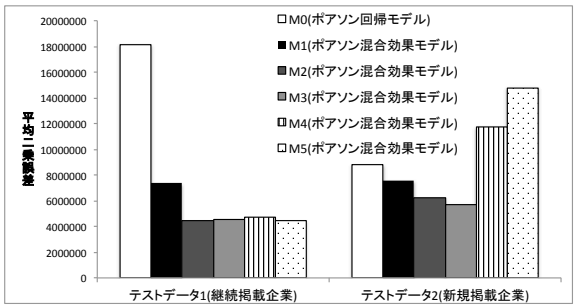

図 **1.** 平均二乗誤差

図 1 から，ポアソン混合効果モデルを用いた $M_1 \sim M_5$ の予測精度はポアソン回帰モデルを用いた M_0 と比べ，継続掲載企業に対して大幅に向上している．そのためポアソン混合効果モデルは継続的に就職ポータルサイトを利用している掲載企業の被エントリ数予測に対して有効であることがわかる．また新規掲載企業に対する予測においては，説明変数が 3 つ以上の場合には，比較手法より予測精度がわずかながら優れていることがわかる．

以上よりポアソン混合効果モデルを用いて，各企業間の差を表すランダム効果を導入し，各企業ごとの回帰パラメータを推定することは，被エントリ数の予測に有効であることがわかる．また既存掲載企業に対しては M_2 が最も予測精度が高い．新規掲載企業に対しては M_3 が最も予測精度が高く，説明変数の数が少なすぎると精度が大幅に悪くなる．さらに各企業によっても，予測精度が高いモデルは異なった．このことから，企業によって最適なモデルは異なることが示唆される．

4 提案手法

4.1 概要

一般に回帰モデルを考える上で，変数選択は重要な問題である．説明変数が少なすぎる場合には対象とする問題の構造を推定できず，逆に一度に多くの説明変数を利用すると，過学習が生じる可能性が高くなる．また，データによっては予測に悪影響を及ぼす変数がある可能性があり，最適なモデルはデータごとで異なると考えられる．予備実験では，既存掲載企業, 新規掲載企業双方に対して最も精度が高いモデルは使用可能な行動情報を全て用いた M_1 ではなく，3 つの行動情報を用いた M_2 や M_3 であり，企業ごとでも予測精度が高いモデルが異なった．すなわち予測精度の面では，企業によって回帰モデルの構築に用いる最適な変数は異なると考えられるため，全ての場合に最適な変数を選択することは困難である．このような問題を解決するため，提案手法では少数の変数の組み合わせでモデルを構築し，それらのモデルを混合することでこの問題の解決し，予測精度の向上を図る．

4.2 ポアソン混合効果モデルの混合

前述の通り，本研究では，少数の変数を用いたモデルを混合することで予測精度の向上を図る．具体的には，ポアソン混合効果モデルを用いて少数の説明変数でモデルを複数構築した後，適切な重みを用いて，各モデルの予測値との重み付け和とることで混合モデルを構成する．すなわち，混合する回帰モデルの数を D，入力 $\boldsymbol{x}$ に対して，d $(d = 1, 2, \cdots, D)$ 番目の回帰モデル（モデル d）により出力される予測値を $f_d(\boldsymbol{x})$, モデル d の重みを w_d としたとき，混合モデルは式 (11) で表される．

$$y = \sum_{d=1}^{D} w_d f_d(\boldsymbol{x}) \tag{11}$$

ただし，モデル d における回帰式は 3.2 節の方法で推定する．以下では与えられた D 個のモデルから最適な重み $w_1, w_2, \cdots, w_D$ を求めることを考える．いま i 番目のデータの被エントリ数を y_i, 式 (8) により算出されるモデル d の i 番目のデータの予測値を $\hat{y}_i^{(d)} = f_d(\boldsymbol{x_i})$ とする．被エントリ数 y_i に対し，各回帰モデルの予測値を説明変数ベクトル $\boldsymbol{p}_i = (\hat{y}_i^{(1)}, \hat{y}_i^{(2)}, \cdots, \hat{y}_i^{(D)})^{\mathrm{T}} \in \mathcal{R}^D$ として回帰モデルを構成する．すなわち，目的変数ベクトル $\boldsymbol{y} = (y_1, y_2, \cdots, y_N)^{\mathrm{T}} \in \mathcal{R}^N$ と全データの説明変数の行列 $\boldsymbol{P} = (\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots, \boldsymbol{p}_N)^{\mathrm{T}} \in \mathcal{R}^{N \times D}$ を用いて，重みベクトル $\boldsymbol{w} = (w_1, w_2, \cdots, w_D)^{\mathrm{T}} \in \mathcal{R}^D$ を以下の最小二乗法により推定する．

$$\underset{\boldsymbol{w}}{\text{minimize}} \quad \|\boldsymbol{y} - \boldsymbol{P}\boldsymbol{w}\|^2 \tag{12}$$

$$\text{s.t.} \quad \sum_{d=1}^{D} w_d = 1 \tag{13}$$

$$0 \leq w_d \leq 1 \tag{14}$$

上記の制約付き最小二乗問題に対する最適な $\boldsymbol{w}$ は内点法を用いて解くことで得られる．

5 評価実験

提案手法の有効性を評価するため，就職ポータルサイトの実データを用いた実験を行った．

5.1 実験条件

実験ではモデルの数 D を 10 個とし，それらを混合することで回帰モデルの構築を行うものとした．各モデルの回帰式の設定は 3.2 節における $H = 3$ のときと同様であり，各モデルの説明変数の設定は表 2 に示す．$H = 3$ としたのは 3.3 節の予備実験において，継続掲載企業，新規掲載企業双方への予測精度を考えた場合，$H = 2$ や $H = 1$ よりも適切であると判断したためである．

表 **2.** 各モデルの説明変数の設定

モデル名	説明変数
モデル 1（d=1）	行動 A, 行動 B, 行動 C
モデル 2（d=2）	行動 A, 行動 B, 行動 D
モデル 3（d=3）	行動 A, 行動 B, 行動 E
モデル 4（d=4）	行動 A, 行動 C, 行動 D
モデル 5（d=5）	行動 A, 行動 C, 行動 E
モデル 6（d=6）	行動 A, 行動 D, 行動 E
モデル 7（d=7）	行動 B, 行動 C, 行動 D
モデル 8（d=8）	行動 B, 行動 C, 行動 E
モデル 9（d=9）	行動 B, 行動 D, 行動 E
モデル 10（d=10）	行動 C, 行動 D, 行動 E

実験条件は予備実験と同様であり，テストデータ 1 とテストデータ 2 の二つのテストデータに対する予測精度を平均二乗誤差で評価する．また比較手法は，予備実験で構築した行動 A, 行動 B, 行動 C, 行動 D, 行動 E の 5 個の企業行動情報を説明変数としたポアソン混合効果モデル M_1 とした．

5.2 実験結果と考察

実験結果を図 2 に示し，重み $\boldsymbol{w}$ の推定値を表 3 に示した．

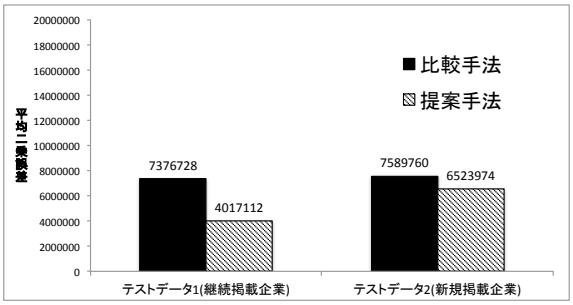

図 **2.** 平均二乗誤差

表 3. 重み $\boldsymbol{w}$ の推定値

w_1	w_2	w_3	w_4	w_5
0.00000	0.14998	0.06647	0.02495	0.12723
w_6	w_7	w_8	w_9	w_{10}
0.01975	0.15219	0.00000	0.32713	0.13228

図 2 から，提案手法はテストデータ 1, テストデータ 2 双方に対して，比較手法よりも精度が向上していることがわかる．したがって，予測の面では，単一のモデルを選択するよりも，モデル選択をせずに，複数のモデルを混合することが有効であると考えられる．

6 提案モデルを用いた企業行動の効果分析

提案手法では，複数のモデルの混合により予測精度の向上を図った．一方で，提案モデルを分析することで，各行動が被エントリ数に与える影響を分析することもまた可能となる．以下では，行動 B に着目し，行動 B を説明変数としているモデルを対象にその分析を行う．これにより提案モデルの実際の利用法について検討を行う．

いま，混合する前のモデル d で推定された行動 B の回帰係数 $\beta_{di}^{(B)}$ としたときに，$\beta_{di}^{(B)}$ を表 3 のモデルの重み $\boldsymbol{w}$ で重み付き平均した値を $\beta_i^{(B)}$ とする．$\beta_i^{(B)}$ は以下の式 (15) で表される．

$$\beta_i^{(B)} = w_1\beta_{1i}^{(B)} + w_2\beta_{2i}^{(B)} + w_3\beta_{3i}^{(B)} + w_7\beta_{7i}^{(B)} + w_8\beta_{8i}^{(B)} + w_9\beta_{9i}^{(B)} \quad (15)$$

表 4 に 15 年度に行動 B を起こしていなかった企業のうち，$\beta_i^{(B)}$ が大きい 10 社の予測誤差率と期待上昇率を示す．また表 5 に式 (10) における行動 B の全データに共通する回帰パラメータ η_q の各モデルの推定値を示す．

表 4. 予測誤差率と期待上昇率

	$\beta_i^{(B)}$	予測誤差率	期待上昇率
企業 1	0.60163	25	135
企業 2	0.58253	33	114
企業 3	0.56793	45	103
企業 4	0.35695	39	193
企業 5	0.32509	22	168
企業 6	0.31656	8	148
企業 7	0.28773	5	121
企業 8	0.27885	3	132
企業 9	0.26514	56	203
企業 10	0.26360	13	138

表 5. 各モデルの行動 B の全データに共通する回帰パラメータ η_q

モデル 2	モデル 3	モデル 7	モデル 9
-0.64433	-0.40790	-0.28946	-0.43209

ここで表 4 における予測誤差率は，予測値が実測値とどれだけずれているかを表す指標である．また期待上昇率とは,「もし各行動を起こしていなかった企業が各行動を起こした際に，実測値からどれだけ上昇するか」を表す指標である．予測誤差率と期待上昇率は以下の式で表される．

$$\text{予測誤差率}\,[\%] = \frac{|\ \text{実測値}–\text{予想エントリ数}\ |}{\text{実測値}} \times 100 \quad (16)$$

$$\text{期待上昇率}\,[\%] = \frac{\text{期待獲得エントリ数}}{\text{実測値}} \times 100 \quad (17)$$

式 (16),(17) における実測値とは各企業が 15 年度に獲得した被エントリ数であり，予想エントリ数とは提案モデルを用いて算出した各企業の 15 年度の被エントリ数の予測値のことである．期待獲得エントリ数とは，15 年度に行動 B を起こしていなかった企業が，起こしたとした場合に提案モデルで算出される被エントリ数の予測値である．表 5 から，予測誤差率が 10 %未満の企業もあり，企業ごとでばらつきの多い被エントリ数のデータに対して，部分的にではあるが高精度な予測ができていることが分かる．また表 5 を見ると，行動 B に関しては全データ共通のパラメータが全てのモデルでマイナス値をとっているものの，表 4 より，上位の企業ではプラス値となっており，企業ごとのばらつきをランダム効果で適切に表現できていることが推察される．また各行動の偏回帰係数が大きな企業群に対して，業種や従業員規模といった観点から，各企業の特徴に関して分析を行った結果，業種や従業員規模に規則性は見られなかった．このことから，業種や従業員規模よりも企業ごとのばらつきを表すランダム効果 $r_{j(i)}^{(q)}$ が被エントリ数に大きく影響しており，各企業行動が被エントリ数に与える影響度は企業ごとで大きく異なることが推察される．また，提案モデルを用いることにより，表 4 における期待上昇率のように各企業がある行動を起こした場合，現在獲得している被エントリ数からどれだけ新たなエントリを獲得できるかなどの予測が可能となる．

例えば，継続して掲載しているある企業が次年度の採用戦略を立てる場合を考える．提案モデルを用いることで,「この企業が今まで起こしていなかった行動を起こすことにより，現在よりもどれだけ被エントリ数が上昇するのか」,「今まで起こしていた行動を起こさなくなった場合，どれだけ被エントリ数が減少するのか」を定量的に把握することができる．これにより，各企業の各行動に対する費用対効果が算出でき，次年度その行動を起こすべきかどうかの合理的な判断材料となる．以上の点から提案モデルは，採用戦略を考えるにあたって実用性の高いモデルであると考えられる．

7 まとめと今後の課題

本研究では，就職ポータルサイト上での行動と外部要因の双方を考慮した被エントリ数の予測モデルを構築するため, ポアソン混合効果モデルを基にした予測モデルを提案し, 実データを用いた検証実験によりその有効性を示した. また，提案モデルを用いた分析を通じ，就職ポータルサイト上での行動と被エントリ数との関係性を明らかにした．

今後の課題としては, 新規掲載企業に対する精度向上が考えられる．継続掲載企業に対しての予測精度は大幅に向上したが，新規掲載企業に対しては大幅な向上は見られなかったため，改善の余地があると考える．そのために残差の大きかった企業に対する分析を進めていく必要がある．

参考文献

[1] N. E. Breslow and D. G. Clayton, “Approximate Inference in Generalized Linear Mixed Models” *Journal of the American Statistical Association*, Vol. 88, No. 421, pp. 9–25(1993)

[2] 久保拓弥: データ解析のための統計モデリング入門, 岩波書店 (2012)

[3] 伊庭幸人，石黒真木夫，松本隆，乾敏郎，田邊國士: 階層ベイズモデルとその周辺, 岩波書店 (2004)

[4] A.Gelman, “Prior distributions for variance parameters in hierarchical models” *Bayesian analysis*, 1, Number 3, pp. 515–533(2006)