就職ポータルサイトにおける個社ページ閲覧とエントリーの関係分析モデルに関する研究

1X13C062-5　杉山裕貴
指導教員　後藤正幸

1　研究背景と目的

近年，採用活動を行う企業や就職活動を行う学生の多くが就職ポータルサイトを活用している．企業は，個社ページに自社の基本情報や採用情報等を掲載し，学生ユーザ（以下，ユーザ）からのエントリーを募集することができる．一方，ユーザは，このサイトを通して，個社ページを閲覧することで企業の魅力を知り，興味のある企業に対してエントリーを行うことができる．就職ポータルサイト上には，これらのユーザの行動履歴データが大量に蓄積されており，これらのデータを有効活用することで，ユーザの行動情報と企業の関係性を分析し，企業側に様々な施策を提案できる可能性がある．

就職ポータルサイトのデータを用いた研究においては，ユーザのエントリー履歴データを用いることで，ユーザの嗜好と企業の関係性に着目した統計的分析モデルが提案されている [1]．しかしながら，ユーザがある企業にエントリーするという行動の背景には，個社ページを閲覧し興味を持ってエントリーした場合と，個社ページを閲覧せず企業名や業種のみを見てエントリーした場合の 2 通りが想定されるが，従来研究ではエントリーという事象を一様に扱っているため，それらの差異が考慮できていない．エントリー履歴のみを用いた従来研究に対して，閲覧履歴データを新たに考慮することで，ユーザの特性をより的確に捉えられると考えられる．これにより，企業にとって，個社ページ閲覧とエントリーの両方を行いやすいユーザと，個社ページ閲覧のみでエントリーを行いにくいユーザといった行動傾向の差異を把握することが可能である．このように，ユーザの閲覧とエントリーの関係性を分析することで，企業は，個社ページを閲覧したにも関わらずエントリーに結びつかないユーザ層に対して，閲覧からエントリーへとつなげる何らかの施策を行うことができる．

そこで本研究では，ユーザの企業に対する個社ページ閲覧とエントリーの関係を分析するためのモデルを提案する．提案モデルにおいては，ユーザと企業にそれぞれ潜在クラスを独立に仮定し，それらの組み合わせによってユーザの企業に対する行動（個社ページ閲覧，エントリー）をモデル化する．そして，提案手法により得られた 2 つの潜在クラスの特徴を分析することで，個社ページ閲覧とエントリーの共起関係のクロス分析が可能となる．本研究の提案モデルの有効性を確認するため，大手就職ポータルサイト（以下，サイト A）における実データの分析を行い，ユーザの閲覧とエントリー行動の関係性が表現されていることを示す．

2　準備

Goto ら [2] は，EC サイトの顧客のアイテムに対する閲覧と購買の 2 種類の行動履歴を用いた潜在クラスモデルを提案している．この手法では，EC サイトにおける顧客のアイテムに対する行動を，購買回数 w_1，閲覧回数 w_2 で構成される二次元ベクトル $\boldsymbol{w}=(w_1,w_2)$ で表現する．ここで，G 個からなるアイテム集合を $\mathcal{A}=\{a_g:1\leq g\leq G\}$，$H$ 人からなる顧客集合を $\mathcal{B}=\{b_h:1\leq h\leq H\}$，アイテム a_g に仮定する M 個の潜在クラス集合を $\mathcal{V}_d=\{d_m:1\leq m\leq M\}$，顧客 b_h に仮定する N 個の潜在クラス集合を $\mathcal{V}_e=\{e_n:1\leq n\leq N\}$ と定義する．Goto らのモデルでは，顧客の潜在クラス $e_n\in\mathcal{V}_e$ とアイテムの潜在クラス $d_m\in\mathcal{V}_d$ を独立に仮定し，それらの組み合わせにより，顧客のアイテムに対する行動を表現している．顧客 $b_h\in\mathcal{B}$ のアイテム $a_g\in\mathcal{A}$ に対する購買と閲覧の回数 $\boldsymbol{w}=(w_1,w_2)$ の確率モデルは以下の式 (1) で表される．

$$P(a_g,b_h,\boldsymbol{w})=\sum_{m,n}P(d_m)P(e_n)P(b_h|e_n)P(a_g|d_m)P(\boldsymbol{w}|d_m,e_n) \quad (1)$$

式 (1) における各パラメータの推定は，EM アルゴリズムにより行う．

3　提案モデル

3.1　概要

本研究では，Goto らのモデルを就職ポータルサイトに適用し，ユーザの企業に対する個社ページ閲覧とエントリーの関係をモデル化する潜在クラスモデルを提案する．

提案モデルでは，ユーザの企業に対する個社ページ閲覧とエントリーの 2 種類の行動の有無を，それぞれ 2 値で構成される二次元ベクトルで表現する．例えば，ユーザが個社ページを閲覧したがエントリーしなかった場合は，$\boldsymbol{w}=(w_1,w_2)=(1,0)$ と表される．また，Goto らのモデルと同様に，ユーザと企業にそれぞれ独立に潜在クラスを仮定したもとで，その組み合わせにより，ユーザの企業に対する閲覧とエントリーの関係性をモデル化する．本研究の提案モデルの有効性を確認するため，このモデルをサイト A 上のユーザの行動履歴データに適用し，ユーザと企業の潜在クラスを分析する．

3.2　定式化

P 社からなる企業集合を $\mathcal{C}=\{c_p:1\leq p\leq P\}$，$Q$ 人からなるユーザ集合を $\mathcal{U}=\{u_q:1\leq q\leq Q\}$，$I$ 個からなる企業の潜在クラス集合を $\mathcal{V}_s=\{s_i:1\leq i\leq I\}$，$J$ 個からなるユーザの潜在クラス集合を $\mathcal{V}_t=\{t_j:1\leq j\leq J\}$ と定義する．このときの確率モデルは以下の式 (2) で表される．

$$P(c_p,u_q,\boldsymbol{w})=\sum_{i,j}P(s_i)P(t_j)P(c_p|s_i)P(u_q|t_j)P(\boldsymbol{w}|s_i,t_j) \quad (2)$$

3.3　モデルの学習

与えられた全データの件数を L とし，l 番目のデータにおけるユーザを $y_l\in\mathcal{U}$，企業を $x_l\in\mathcal{C}$，ユーザ y_l の企業 x_l に対する閲覧とエントリーの有無を $\boldsymbol{w}_l=(w_{l1},w_{l2})$ とする．提案モデルにおける式 (2) のパラメータ $P(s_i)$，$P(t_j)$，$P(c_p|s_i)$，$P(u_q|t_j)$，$P(\boldsymbol{w}|s_i,t_j)$ は，EM アルゴリズムにより推定する．具体的には，式 (3) の対数尤度関数 LL が収束するまで，式 (4)–(9) の E-step，M-step を繰り返し，パラメータを更新する．

$$LL=\sum_{l=1}^{L}\log P(x_l,y_l,\boldsymbol{w}_l) \quad (3)$$

【E-step】

$$P(s_i,t_j|x_l,y_l,\boldsymbol{w}_l)=\frac{P(s_i)P(t_j)P(x_l|s_i)P(y_l|t_j)P(\boldsymbol{w}_l|s_i,t_j)}{\sum_{i,j}P(s_i)P(t_j)P(x_l|s_i)P(y_l|t_j)P(\boldsymbol{w}_l|s_i,t_j)} \quad (4)$$

【M-step】

$$P(s_i)=\frac{\sum_{l=1}^{L}\sum_{t_j\in\mathcal{V}_t}P(s_i,t_j|x_l,y_l,\boldsymbol{w}_l)}{L} \quad (5)$$

$$P(t_j)=\frac{\sum_{l=1}^{L}\sum_{s_i\in\mathcal{V}_s}P(s_i,t_j|x_l,y_l,\boldsymbol{w}_l)}{L} \quad (6)$$

$$P(c_p|s_i)=\frac{\sum_{l=1}^{L}\sum_{t_j\in\mathcal{V}_t}P(s_i,t_j|x_l,y_l,\boldsymbol{w}_l)\delta(x_l=c_p)}{L\times P(s_i)} \quad (7)$$

$$P(u_q|t_j)=\frac{\sum_{l=1}^{L}\sum_{s_i\in\mathcal{V}_s}P(s_i,t_j|x_l,y_l,\boldsymbol{w}_l)\delta(y_l=u_q)}{L\times P(t_j)} \quad (8)$$

$$P(\boldsymbol{w}|s_i,t_j)=\frac{\sum_{l=1}^{L}\delta(\boldsymbol{w}_l=\boldsymbol{w})P(s_i,t_j|x_l,y_l,\boldsymbol{w}_l)}{\sum_{l=1}^{L}P(s_i,t_j|x_l,y_l,\boldsymbol{w}_l)} \quad (9)$$

ただし，$\delta(k{=}k')$ は，$k{=}k'$ のとき 1，$k\neq k'$ のとき 0 をとるインジケータ関数とする．式 (9) で算出された $P(\boldsymbol{w}|s_i,t_j)$ を用いて，潜在クラス s_i，t_j の組み合わせにより個社ページ閲覧とエントリーの関係性を分析することができる．

4 分析

4.1 分析データ

提案モデルを用いて，ポータルサイト A における 2016 年 3 月卒業の学生の個社ページ閲覧履歴データとエントリー履歴データを分析する．本分析では，データの対象期間を 2015 年 3 月 1 日〜3 月 31 日とする．また，分析対象ユーザは対象期間におけるエントリー件数が 10 件以上のユーザ，分析対象企業は対象ユーザによる被エントリー，個社ページ被閲覧の回数がそれぞれ 1 件以上の企業とする．事前分析により，企業の潜在クラス数を $I=4$，ユーザの潜在クラス数を $J=3$ と設定した．

4.2 分析結果

分析の結果として，各潜在クラスにおける閲覧とエントリーの関係を表す $P(w_1{=}1,w_2{=}0|s,t)$，$P(w_1{=}0,w_2{=}1|s,t)$，$P(w_1{=}1,w_2{=}1|s,t)$ を表 1–3 に示す．

表 1. $P(w_1{=}1$(閲覧あり$),w_2{=}0$(エントリーなし$)|s,t)$

ユーザ＼企業	s_1	s_2	s_3	s_4
t_1	**0.989**	**1.000**	**1.000**	**0.718**
t_2	0.000	**0.727**	0.056	0.000
t_3	0.017	**0.996**	**0.540**	0.004

表 2. $P(w_1{=}0$(閲覧なし$),w_2{=}1$(エントリーあり$)|s,t)$

ユーザ＼企業	s_1	s_2	s_3	s_4
t_1	0.000	0.000	0.000	0.000
t_2	**0.840**	0.057	0.000	0.270
t_3	0.000	0.000	0.000	0.000

表 3. $P(w_1{=}1$(閲覧あり$),w_2{=}1$(エントリーあり$)|s,t)$

ユーザ＼企業	s_1	s_2	s_3	s_4
t_1	0.011	0.000	0.000	0.282
t_2	0.160	0.216	**0.944**	**0.730**
t_3	**0.983**	0.004	0.460	**0.996**

まず，企業側の潜在クラス s に着目すると，表 1 より，潜在クラス s_2 に所属する企業は閲覧のみでエントリーが行われにくい傾向がある．一方で，表 3 より，潜在クラス s_4 に所属する企業は閲覧とエントリーの両方が行われやすい傾向があることが確認できる．次にユーザ側の潜在クラス t に着目すると，潜在クラス t_1 に所属するユーザは閲覧のみでエントリーを行いにくい傾向がある．一方で，表 2 より，潜在クラス t_2 に所属するユーザは，個社ページ閲覧を行わずエントリーのみを行いやすい傾向があることが確認できる．これらの結果から，本研究の提案モデルを用いることで，ユーザと企業の潜在クラスの組み合わせにより，閲覧とエントリーの関係性が示せているといえる．

次に，企業 c_p, ユーザ u_q の各潜在クラスへの所属確率 $P(s_i|c_p)$，$P(t_i|u_q)$ により，企業，ユーザを各潜在クラスに割り当てたときの，各潜在クラスに所属する企業，ユーザの割合を以下の表 4,5 に示す．

表 4. 各潜在クラスに所属する企業の割合

潜在クラス	s_1	s_2	s_3	s_4
割合	13.6%	33.1%	35.6%	17.6%

表 5. 各潜在クラスに所属するユーザの割合

潜在クラス	t_1	t_2	t_3
割合	26.2%	16.6%	57.1%

表 4,5 より，閲覧のみでエントリーが行われにくい潜在クラス s_2 に所属する企業と，企業の潜在クラスによって異なる行動傾向を示す潜在クラス t_3 に所属するユーザの割合が高いことがわかる．これより，「閲覧あり，エントリーなし」となりやすいユーザ・企業は多く存在するといえる．また，最も多くのユーザが所属する潜在クラス t_3 のユーザは，最も多くの企業が所属する潜在クラス s_3 の企業に対して，「閲覧あり，エントリーなし」,「閲覧あり，エントリーあり」の行動を同程度の確率で行う傾向にあることがわかる．

4.3 考察

まず，ユーザの潜在クラス間の差異を確認するため，潜在クラスごとのユーザのエントリー数と個社ページ閲覧数の平均の比率を以下の表 6 に示す．

表 6. 各潜在クラスの平均エントリー数と平均閲覧数の比率

潜在クラス	平均エントリー数の比率	平均閲覧数の比率
t_1	1.000	3.109
t_2	1.412	1.000
t_3	1.337	1.616

閲覧のみでエントリーを行いにくい潜在クラス t_1 のユーザと，閲覧とエントリーの両方を行いやすい潜在クラス t_3 のユーザの間には，平均エントリー数と平均閲覧数に，Welch の t 検定の有意水準 5%で有意な差がある．

また，企業の潜在クラスごとの特徴を確かめるため，各潜在クラスに所属する企業の従業員規模ごとの割合を以下の表 7 に示す．

表 7. 従業員規模ごと各潜在クラスに所属する企業の割合

従業員規模＼潜在クラス	s_1	s_2	s_3	s_4
50 人未満	6.2%	**59.1%**	30.7%	4.0%
50〜100 人未満	8.3%	**46.4%**	36.6%	8.6%
100〜300 人未満	10.0%	37.2%	**40.7%**	12.2%
300〜500 人未満	12.4%	27.8%	**40.8%**	19.0%
500〜1000 人未満	15.4%	21.6%	**38.6%**	24.4%
1000〜3000 人未満	19.8%	17.9%	30.1%	**32.3%**
3000〜5000 人未満	27.2%	12.1%	22.4%	**38.3%**
5000 人以上	**38.3%**	13.3%	18.6%	29.8%

表 1 と表 7 より，従業員規模が小さい企業ほど，閲覧のみでエントリーが行われにくい潜在クラス s_2 に所属する割合が高いことがわかる．一方，従業員規模が大きい企業ほど，閲覧とエントリーの両方が行われやすい潜在クラス s_4 に所属する割合が高いことがわかる．

分析結果より，例えば，潜在クラス s_2 に所属する企業は個社ページの改善や企業イメージの向上により，個社ページを閲覧したユーザのエントリー数増加が期待できる．一方，潜在クラス s_1 に所属する企業はユーザの潜在クラスによって，行動傾向の違いが顕著であるため，それぞれのユーザに応じた施策を行うべきであると考えられる．

5 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ユーザの企業に対する閲覧行動とエントリー行動を同時に分析するモデルを提案し，そのモデルによる分析結果を示した．

今後の課題として，個社ページ閲覧とエントリーの前後関係を考慮したモデル化，およびその他の行動情報を取り入れたモデルの検討などが挙げられる．

参考文献

[1] 大森悠矢, 三川健太, 石田崇, 後藤正幸, 小川晋一郎, “就職ポータルサイトにおけるレコメンデーションモデルに関する一考察,” 第 36 回情報理論とその応用シンポジウム, SITA2013, No.7.3.1, 2013

[2] M. Goto, K. Mikawa, S. Hirasawa, M. Kobayashi, T. Suko, S. Horii, “A New Latent Class Model for Analysis of Purchasing and Browsing Histories on EC Sites,” *Industrial Engineering & Management Science*, Vol.14, No.4, pp.335-346, 2015.