

潜在クラスを導入した学生の就職活動エントリー予測モデルに関する研究

情報数理応用研究

5212C043-9 峯吉和史
指導教員 後藤正幸

Entry Prediction Model of Students' Job Hunting with Latent Classes

MINETOMA Kazushi

1 研究背景・目的

近年、数多くの学生がインターネット上での就職ポータルサイトを利用した就職活動を行っている。就職ポータルサイトでは掲載企業から利益を得るビジネスモデルをとるため、学生の個々の企業への入社志願（以下、エントリー）数の確保が重要となる。しかし、就職ポータルサイトの掲載企業間のエントリー数には特定の大企業にエントリー数が集中し、その他大多数の企業ではエントリー数が少ない傾向がある。そのため、エントリー数の少ない企業へのエントリー数向上を目的として、膨大に蓄積された学生や企業のエントリー情報からエントリー数を予測するモデルの構築が望まれている。エントリー数を予測モデルとしては、決定木 [1] などを用いて、学生や企業の属性データから予測モデルを構築する方法が考えられる。しかし、このような予測モデルでは学生のエントリー数に寄与する属性や企業の被エントリー数を向上させる属性を知ることは出来るが、学生のエントリーモデルと企業の被エントリーモデルが独立にモデル化されるため、両者の関係性がモデルに取り込まれない。そのため、「どのような学生クラスにアプローチすることがどの分野の企業のエントリー数の増加につながるか」といった有効な施策を検討するための議論に結びつかない。そこで、マーケティング分野において消費者や商品の異質性を表現するモデルとして有用性が示されている潜在クラスモデル [2] を導入することで、学生と企業の背後に潜在的な嗜好を表す変数を仮定し、学生と企業の潜在的な意味関係を捉えたエントリー数の予測モデルの構築を考える。

このような関係を分析する確率的潜在クラスモデルとしては、Aspect Model の一種として Hofmann によって考案された PLSI [3] が挙げられる。しかし、PLSI は本来情報検索や文書モデルを対象として提案されたものであり、潜在クラスから各学生と各企業への確率分布に多項分布を仮定している。これは現実の学生の就職活動を考慮した場合「いずれかの学生がいずれかの企業にエントリーする」という事象を条件としたもとの学生と企業の共起の条件付確率をモデル化している、そのため学生間のエントリー数の差異は、学生の多項分布に取り込まれてしまい、エントリー数に差はあっても類似した企業にエントリーしていれば同じ潜在クラスにまとめられてしまうという傾向がある。また、各潜在クラスに所属する学生数や平均エントリー数を陽に計算できるモデルとはなっていないため、「どの潜在クラスに所属する学生に対してアプローチを行うことで企業へのエントリー数に最も効果的に影響を与えるか」といった検討のためにエントリー数の予測シミュレーションを行うことが出来ない。

そこで本研究では、Aspect Model の一種である潜在クラスを考慮したモデルを取り入れつつ、潜在クラスから各企業への確率分布として「エントリーの有無」という事象に対して二項分布を仮定することで、各潜在クラスに所属する企業

と学生の傾向に加え、潜在クラスに所属する学生数や平均エントリー数を陽に計算できるモデルを提案する。また、提案モデルを実際の就職活動データに適用し、学生への効果的なアプローチ法の検討などのシナリオシミュレーションを行い、その有効性を示す。

2 就職ポータルサイトを用いた就職活動

就職ポータルサイトは学生の就職活動を支援する情報サイトであり、学生はこの就職ポータルサイトを用いることで企業情報や採用情報の検索、自身の興味のある企業へのエントリーやセミナーの予約を行うことができる。

この就職ポータルサイトを利用した就職活動の普及に伴い、学生は多くの企業情報や採用情報を容易に獲得し、個々の嗜好に適合する企業を発見するようになった。そこで、就職ポータルサイト運営会社では学生の個々の嗜好をより詳細に分析するための就職活動の構造化や、エントリー数下位企業に対してエントリーする可能性の高い学生への効果的なアプローチを検討するためのエントリー数予測モデル構築など、様々な統計モデルを用いた分析を進めている。

3 決定木モデル

本研究ではエントリー数の予測モデルの基本的な予備実験として、決定木を用いた予測モデルを構築した。

3.1 決定木モデル概要

決定木モデルは、分類と予測の双方で用いられ、サンプルの属性情報を利用してルールを作成し、そのルールをもとに母集団を層別するモデルである。決定木モデルの構造はノード（節）とそこから分岐した（葉）ノードで示される。ノードはそこからの分岐条件を示し、データは次のノード、もしくは葉ノードへと引き渡される。葉ノードは、これ以上分岐する必要がない最終的な分類を意味し、頂点（根）ノードから葉ノードまでの構造を決定木モデルと呼ぶ。

例として学生プロフィールとその学生のエントリー回数について性別、文理、学種が示されたデータベースから作成された「エントリー回数」に関する決定木を示す。

表 1. 学生プロフィール (例)

	性別	文理	学種	エントリー数
001	男性	理系	院生	23
002	女性	理系	学部生	42
003	女性	文系	院生	30
004	男性	文系	学部生	67
...

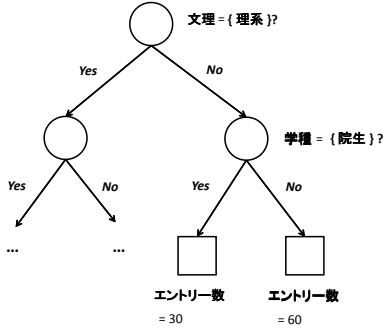


図 1. エントリー回数に関する決定木モデル

木の各中間ノードにはデータを分類するための分岐条件が示されている。また、各中間ノードには分岐条件を適用した結果の取り得る値に対応する下向きの枝をもつ。データベース上の各学生は、この決定木を用いることで、いずれかの葉ノードに割り当てられ、そのノードに分類される学生集合のエントリー数の平均値が割り振られる。

3.2 決定木モデルの限界

決定木モデル等の一般的な予測モデルでは、学生の性別、文理、学種などを入力属性とした学生のエントリー数予測モデルや、企業の業種、規模、所在地などの属性を入力した企業の被エントリー数予測モデルを構築することができる。しかし、学生のエントリーの嗜好や学生と企業間の関連性を構造化したい場合には、意味のあるモデルとなっていない。

4 潜在クラスモデル

本研究では学生のエントリーの嗜好や学生と企業間の関連性の構造化を行うため、潜在クラスモデルの一つである AspectModel[3] の適用を考える。

4.1 潜在クラスモデル概要

潜在クラスモデルは文書分類や推薦システムにおいて幅広く利用されているモデルである。特に推薦システムにおける Aspect Model は、ユーザとアイテムの間に潜在クラスを仮定することで、ユーザの嗜好やアイテムの特徴の類似性を考慮することができる。本研究では、学生と企業の背後に潜在的な嗜好を表す変数を仮定し、学生と企業の潜在的な意味関係を捉えたエントリー数の予測モデルの構築を考える。

4.2 PLSI

Aspect Model の一種である PLSI とは、Hofmann [3] によって提案された確率潜在クラスモデルであり、ユーザとアイテムが潜在クラスに基づいて生成される過程を確率的に表現している。いま、 M 個のアイテムからなるアイテム集合を $\mathcal{Y} = \{y_j : 1 \leq j \leq M\}$ 、 N 人のユーザからなるユーザ集合を $\mathcal{X} = \{x_i : 1 \leq i \leq N\}$ と定義する。ここで (x_i, y_j) はユーザ x_i がアイテム y_j を購入する事象を表している。また、ユーザとアイテムの間に潜在クラスの存在を仮定し、その集合を $\mathcal{Z} = \{z_k : 1 \leq k \leq K\}$ とする。このとき、PLSI のモデルの一般式は式 (1) で表現される。

$$P(x_i, y_j) = \sum_k P(x_i|z_k)P(y_j|z_k)P(z_k) \quad (1)$$

式 (1) のモデル構造から、PLSI では各ユーザ、アイテムは全ての潜在クラスへの所属確率を持つという特徴を持つ。ここで各パラメータ $P(z_k), P(x_i|z_k), P(y_j|z_k)$ はそれぞれ多項分布に従い、各パラメータは EM アルゴリズム [4] で対数尤度を最大にするパラメータとして推定することができる。

4.3 PLSI の限界

PLSI はアイテム y_1, y_2, \dots, y_M の確率分布 $P(y_j|z_k)$ に多項分布を想定しているため、購入点数の多いユーザは $P(y_j|z_k)$ が大きくなるため、異なる平均購入点数を持つユーザが同じ潜在クラスにまとめられてしまう傾向がある。このため、マーケティングモデルとして PLSI の適用を考えた際に平均購入点数が多いユーザと平均購入点数の少ないユーザといった優良顧客や非優良顧客が上手く分けてセグメント化することが出来ないといった課題がある。

5 提案手法

本研究では Aspect Model としての潜在クラスを考慮したモデルを取り入れつつ、各企業 y_j に対して「 y_j にエントリーする」「 y_j にエントリーしない」という二値の事象に対して確率を付与するモデルを考える。すなわち、パラメータ $P(y_j|z_k)$ に潜在クラス z_k のもとで企業 y_j にエントリーするか否かの二項分布を仮定することで、学生が複数の企業をエントリーする現実の就職活動に則し、潜在クラスの特徴に平均エントリー数の大小を取り込んだモデルを考案し、その学習アルゴリズムを示す。

5.1 提案モデルの構造

いま、 M 個の企業からなる企業集合を $\mathcal{Y} = \{y_j : 1 \leq j \leq M\}$ 、 N 人の学生からなる学生集合を $\mathcal{X} = \{x_i : 1 \leq i \leq N\}$ と定義する。ここで (x_i, y_j) は学生 x_i が企業 y_j をエントリーする事象を表す。いま \mathbf{a}_x を学生 x のエントリー履歴として、次の式 (2) で表現する。

$$\mathbf{a}_x = (a_{x1}, a_{x2}, \dots, a_{xM}) \quad (2)$$

ここで a_{xj} は学生 x の企業 y_j へのエントリーの有無を表し、次の式 (3) で定義する。

$$a_{xj} = \begin{cases} 1 & (\text{学生 } x \text{ が企業 } y_j \text{ にエントリー}) \\ 0 & (\text{学生 } x \text{ が企業 } y_j \text{ に非エントリー}) \end{cases} \quad (3)$$

そのもとで、提案モデルでは PLSI をもとにして、 y を企業 y にエントリーする、 \bar{y} を企業 y にエントリーしないの二値の事象に対して二項分布を仮定し、確率モデルを式 (4) で定義する。

$$P(x, \mathbf{a}_x, z) = P(z)P(x|z) \prod_{j=1}^M P(y_j|z)^{a_{xj}} P(\bar{y}_j|z)^{1-a_{xj}} \quad (4)$$

ここで、 $x \in \mathcal{X}$ 、 $y \in \mathcal{Y}$ 、 $z \in \mathcal{Z}$ であり、またパラメータ $P(y|z_k)$ は潜在クラス z_k のもとで企業 y がエントリーされる確率を表し、以下を満たす。

$$P(y|z_k) = 1 - P(\bar{y}|z_k) \quad (5)$$

このとき、 N 人のユーザ全体の確率モデルは次式で表現される。

$$P(\mathcal{S}, \mathcal{C}, \mathcal{V}) = \prod_{i=1}^N P(v_i)P(x_i|v_i) \prod_{j=1}^M P(y_j|v_i)^{a_{xij}} P(\bar{y}_j|v_i)^{1-a_{xij}} \quad (6)$$

ただし、 v_i は学生 x_i の潜在クラスで $v_i \in \mathcal{Z}$ であり、 $\mathcal{S} = (x_1, x_2, \dots, x_N)^T$ 、 $\mathcal{C} = (\mathbf{a}_{x1}, \mathbf{a}_{x2}, \dots, \mathbf{a}_{xN})$ 、 $\mathcal{V} = (v_1, v_2, \dots, v_N)^T$ である。

5.2 パラメータ学習アルゴリズム

提案手法の式 (4) として表現される確率モデルでは不観測データを含むため、パラメータ $P(z)$, $P(x|z)$, $P(\mathbf{a}_x|z)$ を EM アルゴリズムによって推定する。

5.2.1 E-step

初めに、E ステップの期待値計算を行うために $P(\mathcal{V}|\mathcal{C}, \mathcal{S})$ を用意する。

$$\begin{aligned} P(\mathcal{V}|\mathcal{S}, \mathcal{C}) &= \frac{\prod_{i=1}^N P(v_i)P(x_i|v_i) \prod_{j=1}^M P(y_j|v_i)^{a_{x_i j}} P(\bar{y}_j|v_i)^{1-a_{x_i j}}}{\sum_{z \in \mathcal{Z}} \prod_{i=1}^N P(z)P(x_i|z) \prod_{j=1}^M P(y_j|z)^{a_{x_i j}} P(\bar{y}_j|z)^{1-a_{x_i j}}} \\ &= \prod_{i=1}^N \frac{P(v_i)P(x_i|v_i) \prod_{j=1}^M P(y_j|v_i)^{a_{x_i j}} P(\bar{y}_j|v_i)^{1-a_{x_i j}}}{\sum_{z \in \mathcal{Z}} P(z)P(x_i|z) \prod_{j=1}^M P(y_j|z)^{a_{x_i j}} P(\bar{y}_j|z)^{1-a_{x_i j}}} \\ &= \prod_{i=1}^N P(v_i|x_i, \mathbf{a}_x) \end{aligned} \quad (7)$$

このとき Q 関数は式 (6) の $P(\mathcal{S}, \mathcal{C}, \mathcal{V})$ と式 (7) の $P(\mathcal{V}|\mathcal{C}, \mathcal{S})$ を用いて次のように展開できる。

$$\begin{aligned} Q &= \sum_{\mathcal{V}} P(\mathcal{V}|\mathcal{S}, \mathcal{C}) \log P(\mathcal{S}, \mathcal{C}, \mathcal{V}) \\ &= \sum_{\mathcal{V}} P(\mathcal{V}|\mathcal{S}, \mathcal{C}) \log \left\{ \prod_{i=1}^N P(v_i)P(x_i|v_i) \prod_{j=1}^M P(y_j|v_i)^{a_{x_i j}} P(\bar{y}_j|v_i)^{1-a_{x_i j}} \right\} \\ &= \sum_{\mathcal{V}} P(\mathcal{V}|\mathcal{S}, \mathcal{C}) \sum_{i=1}^N \{ \log P(v_i) + \log P(x_i|v_i) \\ &+ \sum_{j=1}^M (a_{x_i j} \log P(y_j|v_i) + (1 - a_{x_i j}) \log P(\bar{y}_j|v_i)) \} \\ &= \sum_{i=1}^N \sum_{z \in \mathcal{Z}} P(z|x_i, \mathbf{a}_x) (\log P(z) + \log P(x_i|z)) \\ &+ \sum_{k=1}^N \sum_{z \in \mathcal{Z}} P(z|x_i, \mathbf{a}_x) \\ &\sum_{j=1}^M (a_{x_i j} \log P(y_j|z) + (1 - a_{x_i j}) \log P(\bar{y}_j|z)) \end{aligned} \quad (8)$$

5.2.2 M-step

次に、M ステップでは式 (8) の潜在変数以外を固定し潜在変数以外のパラメータを算出する。ここで、各パラメータの制約条件を次の条件式 (9),(10),(11) を仮定する。

$$\sum_{x \in \mathcal{X}} P(x|z) = 1 \quad (9)$$

$$P(y|z) + P(\bar{y}|z) = 1 \quad (10)$$

$$\sum_{z \in \mathcal{Z}} P(z) = 1 \quad (11)$$

条件式 (9),(10),(11) のもとで、ラグランジュ未定乗数法 [5] を用いて最適化を行うためラグランジュ乗数 α , β_z , γ_{yz} をおき、ラグランジュ関数 $g(x, y, \alpha, \beta_z, \gamma_{yz})$ を以下で定義する。

$$\begin{aligned} g(x, y, \alpha, \beta_z, \gamma_{yz}) &= Q + \alpha(1 - \sum_{z \in \mathcal{Z}} P(z)) + \sum_{z \in \mathcal{Z}} \beta_z(1 - \sum_{x \in \mathcal{X}} P(x|z)) \\ &+ \sum_{y \in \mathcal{Y}} \sum_{z \in \mathcal{Z}} \gamma_{yz}(1 - P(y|z) - P(\bar{y}|z)) \end{aligned} \quad (12)$$

式 (12) を $P(x_k|z)$, $P(y_j|z)$, $P(z)$ に関して偏微分して 0 とおくことで次の解を導出することが出来る。

$$P(x_i|z) = \frac{P(z|x_i, \mathbf{a}_x)}{\sum_{i=1}^N P(z|x_i, \mathbf{a}_x)} \quad (13)$$

$$P(y_j|z) = \frac{\sum_{i=1}^N P(z|x_i, \mathbf{a}_x) a_{x_i j}}{\sum_{i=1}^N P(z|x_i, \mathbf{a}_x)} \quad (14)$$

$$P(z) = \frac{1}{N} \sum_{i=1}^N P(z|x_i, \mathbf{a}_x) \quad (15)$$

E-step の式 (7) と M-step の式 (13),(14),(15) の更新式を各パラメータが収束するまで繰り返し、パラメータの推定を行う。

5.2.3 予測値算出

提案モデルを用いて実データ分析を行う際に、推定されたパラメータを用いて各潜在クラスに所属する学生数 $\hat{N}(z_k)$ 、平均エントリー数 $\hat{N}(e_k)$ 、各企業への予測被エントリー回数 $\hat{N}(y_j)$ を次の式で与える。

$$\hat{N}(z_k) = NP(z_k) \quad (16)$$

$$\hat{N}(e_k) = \sum_{j=1}^M P(y_j|z_k) \quad (17)$$

$$\begin{aligned} \hat{N}(y_j) &= P(y_j|z_k) \hat{N}(z_k) \\ &= NP(y_j|z_k)P(z_k) \end{aligned} \quad (18)$$

6 実データ分析

以下では、就職ポータルサイトにおける実データとして学生のエントリーデータを使用した実データ分析を行う。そのもとで、各潜在クラスに所属する学生や企業の傾向に加え、各潜在クラスに所属する学生数と平均エントリー数の大小を捉えることのできるエントリー数予測モデルを構築する。構築された提案モデルを用いることでどの潜在クラスに所属する学生に対してアプローチを行うことが企業へのエントリー数に最も効果的に影響を与えるかといったエントリー数の予測シミュレーションを行う。

6.1 データ概要

提案手法の有効性を示すために、2013 年卒業学生を対象とする 1 年間のエントリーデータで学習を行った。学習データとして東京在住の学部生を対象として学生数 $N=5,000$ 人をランダムに抽出し、対象とする学生がエントリーした被エントリー企業数 $M=9,065$ 社を対象とし、学習データ数 $L=301,401$ 件を扱う。また業種数 $S=126$ 、従業員規模数 $T=8$ 、本社所在地 $U=49$ 、潜在クラス数 $Z=10$ として提案モデルを学習し、エントリー数の予測学習と学生の企業への潜在的エントリー嗜好を示す。

6.2 提案モデルの学習データへの当てはまり

学習データにおける各企業 y_j へのエントリー回数 $N(y_j)$ と提案モデルから得られた企業 y_j への予測エントリー回数 $\hat{N}(y_j)$ との平均二乗誤差損失 $Loss(N(y_j), \hat{N}(y_j))$ を次の式 (19) で与える .

$$Loss(N(y_j), \hat{N}(y_j)) = \frac{\sum_{j=1}^M (N(y_j) - \hat{N}(y_j))^2}{M} \quad (19)$$

学習データに対して $Loss(N(y_j), \hat{N}(y_j))$ が 3.072×10^{-8} となった . これは EM アルゴリズムによるパラメータ学習が行われたことで提案モデルが学習データに対して当てはまりの良い適切なモデルとして得られたと考えられる .

6.3 学生の企業への潜在的エントリー嗜好

表 2 に各潜在クラス z_k のもとでの企業 y_j への条件付きエントリー確率 $P(y_j|z_k)$ 上位 10 社の最多一致企業属性と各潜在クラス z_k のもとでの学生 x_i への条件付きエントリー確率 $P(x_i|z_k)$ 上位 10 人の最多一致学生属性の傾向と各潜在クラスに所属する学生数 $\hat{N}(z_k)$ と平均エントリー数 $\hat{N}(e_k)$ を示す .

表 2. 各潜在クラスの所属学生傾向と主なエントリー先企業

潜在クラス	所属学生傾向	潜在クラス平均エントリー数	潜在クラス所属学生数	所属企業傾向
z_1	私立・服飾 中心	48.26	321.46	服飾・アパレル 業界
z_2	私国立・文系 中心	51.15	554.00	大規模 業界
z_3	私国立・理系 中心	34.55	233.91	情報処理・ソフトウェア 業界
z_4	私立・文系 中心	38.35	198.46	冠婚葬祭・ホテル 業界
z_5	私立・理系 中心	59.14	360.00	ゲームソフト・インターネット 業界
z_6	私立・文系 中心	67.34	829.75	広告・インターネット 業界
z_7	私立・文系 中心	77.26	1294.90	商社(総合) 業界
z_8	私立・理系 中心	63.03	446.01	半導体・電子部品 大規模 業界
z_9	私立・文系 中心	45.96	360.71	食品 大規模 業界
z_{10}	私立・文系 中心	49.00	400.80	不動産 業界

表 2 から各潜在クラスに所属する学生の特徴を私国立や文系といった大学の属性によって捉えることができる . また各潜在クラスのユーザがよくエントリーする企業の特徴を業種や規模 , 所在地といった業界の属性によってとらえることができる . これは提案モデルでは Aspect Model と同様に学生と企業間に潜在クラスを設けることで , 学生から企業への潜在的なエントリー嗜好を捉えることができたと考えられる .

また各潜在クラスに対して平均エントリー数の差異と所属する学生数の差異を捉えることができる . これは提案モデルでは AspectModel をもとに各潜在クラスから各企業への確率分布として「エントリーの有無」という事象に対して二項分布を仮定することで潜在クラスの特徴として平均エントリー数の大小を取り込んだモデルを構築できたと考えられる .

6.4 学生 1,000 人増加時のインパクト

学習データに対して学習を行った提案モデルに対して , 各潜在クラス z_k に所属する学生数 $\hat{N}(z_k)$ を 1,000 人増加させた際の全企業に対するエントリー数の総エントリー数 , エントリー数上位 5,000 企業のエントリー数 , エントリー数上位 5,000 企業以外のエントリー数の変化を示す .

表 3. 各潜在クラスの学生 1,000 人増加時のインパクト

潜在クラス	全企業増加エントリー数	エントリー数上位 5000 企業増加エントリー数	エントリー数上位 5000 以外企業増加エントリー数
z_1	49259.07	47242.62	2016.45
z_2	51143.90	49712.98	1430.92
z_3	34544.24	34205.04	339.20
z_4	38352.83	34067.80	4285.04
z_5	59137.83	57607.47	1529.96
z_6	67343.99	65167.64	2176.36
z_7	77255.02	74024.96	3230.05
z_8	63030.21	60041.98	2988.22
z_9	45959.10	43016.32	2942.78
z_{10}	49001.44	47711.48	1289.96

表 3 から各潜在クラス z_k に対してエントリー数上位・下位企業に対して最もエントリー数にインパクトを与える潜在クラス z_7, z_4 が存在することが分かる . これにより , 提案モデルを用いることで全体のエントリー数を向上させるためにアプローチすべき潜在クラスやエントリー数の少ない特定の企業のエントリー数を向上させるためにアプローチすべき潜在クラスを抽出することが可能となる .

6.5 考察

提案モデルを用いたエントリー予測モデルの有用性として , 1 つ目に潜在クラスモデルとしての学生のエントリー嗜好の抽出が挙げられる . 具体的には実データである学習データに対して当てはまりの良いモデルを構築し , 表 2 のように各潜在クラスの特徴を学生の属性とエントリー傾向によって捉えることができる . 2 つ目にエントリー数下位企業のエントリー数向上のための学生への効果的なアプローチ法を検討可能であることが挙げられる . 学習した提案モデルを用いて各潜在クラスに所属する学生 1,000 人増加時のインパクトを計ることで , 表 3 より潜在クラス z_4 に所属する学生数を増加させることが , エントリー数下位企業に対して最もインパクトが高いことが示せる . 一方 , 表 2 より , 潜在クラス z_4 に所属する学生は私立・文系学生である . これらの各大学・各学生に対してよくエントリーする冠婚葬祭ホテル業界へのプロモーションを行うことでエントリー数下位企業に対して最もエントリー数の向上が考えられる .

7 まとめと今後の課題

本研究では , 相互に排他的な事象のみを扱う従来の Aspect Model に対して , 現実の就職活動モデルを想定し , パラメータ $P(y_j|z_k)$ に二項分布を用いるモデルを提案し , 実データ分析によって提案手法の有効性を示した . 具体的には , 提案モデルを用いることで学生の企業へのエントリー志向の抽出に加えて , 特定の企業クラスタに対してのエントリー数向上を図るためにどのような潜在クラスに所属する学生をターゲットとすべきかを把握することを可能とした .

今後の課題として本研究の提案手法では潜在クラスに所属する学生の企業へのエントリー志向を $P(y_j|z_k)$ の上位 10 件により業種・規模・所在地を判断していたが , より学生の企業へのエントリー志向を定量的なスコアとして表現できるモデルとして拡張することが挙げられる .

参考文献

- [1] J. R. Quinlan, “C.4.5: Problems for Machine Learning,” *Morgan Kaufmann*, 1993.
- [2] K. E. Train, “Discrete Choice Methods with Simulation - Second edition,” *Cambridge University Press, Cambridge*, 2009
- [3] Thomas, H., “Probabilistic Latent Semantic Indexing”, *In Proceedings of the 22nd Annual International SIGIR Conference on Research and Development in Information Retrieval*, pp.289-296, 1999.
- [4] C.M.Bishop, “Pattern Recognition and Machine Learning,” *Springer*, 2006.
- [5] 金谷健一., “これなら分かる最適化数学,” 共立出版, pp.64-77, 2005.