

人材ネットワーク

—株式会社ビズリーチにおける事例を中心に—

大概 知貴, 菅谷 信介

株式会社ビズリーチは転職サービスに限らない、数多くのサービスを運営しているが、その多くは第一部が「人材」、第二部が求人などの「アイテム」からなる二部グラフである、という共通点をもっている。本稿では、こうした状況下での機械学習、とりわけベイズ推論の応用と、数理最適化によって「大人の事情」を考慮する、という取り組みについて論ずる。

キーワード：協調フィルタリング、二部マッチング、ベイズ推論、潜在変数モデル、線形計画法

1. はじめに

株式会社ビズリーチは 2019 年に創業 10 周年を迎えた。その間、

- ビズリーチやキャリアトレなどの転職サービス
- ニクリーチやビズリーチ・キャンパスなどの新卒学生向けサービス
- HRMOS 採用管理や HRMOS 評価などの（広い意味での）クラウド型人事サービス
- そのほか、ビズリーチ・サクシード（事業承継および M&A サービス）やビズヒント（ビジネスパーソン向けコンテンツ・メディア）など、非人材業界領域でのプラットフォーム

など、さまざまなプラットフォームを立ち上げてきた。これらのサービスの多くには、それをネットワークという観点から眺めたとき、登場人物が二部グラフから成る、という共通点がある。二部グラフとは、グラフの頂点を重複なく二つに分割することができ、同じ分割内にはリンクが存在しない、という特殊なグラフである。

たとえば、ビズリーチのような転職サービスであれば、登場人物は大まかに言って求職者と採用企業である。SNS のように求職者同士がビズリーチ上でお互いに知り合う、ということはないが、採用企業と求職者はお互いに接触をもつ（すなわち両者の間にリンクが存在しうる）ことができる（図 1）。

このような構造のネットワークにおいては、人と人との間に直接的なコネクションはないものの、

- 求職者 A と求職者 B は同じ職種の書類選考に応

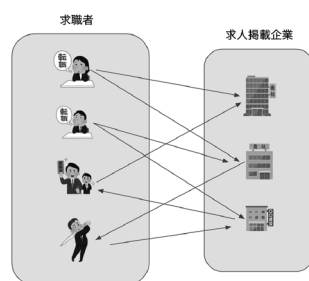


図 1 ビズリーチサービスにおける二部グラフの実現の例
頂点集合が「求職者」と「求人掲載企業・ヘッドハンター」の二部に分割される。求職者と企業とのリンクの種類は、「求職者から企業への応募」や、「企業から求職者へのスカウト」などのパターンが存在する。

募、あるいは選考通過をしているので、似たスキルセットや嗜好をもっている可能性が高い

- 求人 C は、他の求人 A, B の書類選考には通過している候補者をリジェクトしていることから、要求する水準が A, B と比して高そうである

といったことはわかる。すなわち、二部グラフの頂点集合のもう一方を通じて、人と人、あるいは企業と企業との間の繋がりが関係性といったものが見えてくるのである。

ビズリーチサービスを始めとするプラットフォームとしての使命は、いかに良質なマッチングを提供するか、ということに集約されるが、その目的のためにはこうした形の推論アルゴリズムが欠かせない。本稿では、株式会社ビズリーチの諸サービス（特に人材系のもの）における、機械学習や数理最適化を用いた推薦システム構築の取り組みの一端を紹介する。

2 節では、このような状況下での推薦アルゴリズムについて、機械学習的なアプローチ、中でもベイズ推論に基づいた方法について、株式会社ビズリーチにお

ける事例を交えながら論ずる。

3節では、2節とほぼ同様の問題設定を扱うが、ここでの興味はむしろ予測性能よりも、各種の係数の事後分布にある点が異なっている。

4節では、二部グラフ的な構造を利用し、種々のビジネス上の制約を考慮にいたれたマッチング最大化を、(拘束条件が完全単模である) 整数計画問題として定式化する、という取り組みについて紹介する。

2. 機械学習モデルによる推薦アルゴリズム

1節でも述べたように、推薦システムでは、特定のユーザーの振る舞いを、それと似たユーザーの振る舞いから推論すること、すなわち協調フィルタリングが欠かせない。一般的な用語として、行動履歴のみに基づく推論アルゴリズムは“協調フィルタリングベース”(あるいは“履歴ベース”)であると言われる。他方、行動履歴を加味せず、ユーザー/アイテムの属性値のみからマッチングの成否を予測するアルゴリズムは“コンテンツベース”と称される。

広く言われているとおり、

- 協調フィルタリングベースの方法は、行動/閲覧履歴の多いユーザー/アイテムに対しては高い精度を発揮するが、それらが少ない(あるいは皆無である)ユーザー/アイテムに対しては精度が低下する、ないし全く予測が不可能になってしまう(これはコールドスタート問題と呼ばれる)
- コンテンツベースの方法では、行動/閲覧履歴の多少によらず推薦が可能である一方、データの増大に伴う恩恵はそれほど得られない

というトレードオフが存在する。そこで肝要なのは、いかに自然に両極端のアプローチの間を補間する、ハイブリッドな推薦を実現するか、である。

本節では、このようなアプローチのうち、特にベイズ推論を利用した確率モデルに基づくアプローチについて、株式会社ビズリーチにおける適用例を交えながら議論する。推薦アルゴリズムに関する一般的な教科書としては文献 [1] を参照されたい。

2.1 ロジスティック回帰による分析

まず単純な例として、本節ではユーザーとアイテムの相性を考慮しないモデルを考える。この中でも最もシンプルな確率モデルとして、線型なロジスティック回帰モデルを例に、協調フィルタリングベースの手法とハイブリッドなものの違いについて説明していこう。

2.1.1 協調フィルタリングベースの方法

たとえば、転職サービスの書類選考履歴から「あな

たが書類合格しやすい求人」を推薦するシステムを構築したいとする。このとき、各候補者 i に対して「能力値」が、各求人 j には「難易度」が存在するとし、書類合格する/しないは能力値と難易度の兼ね合いから定まると考える。「能力値」や「難易度」といった概念は抽象的なものであるが、ベイズ推論の枠組みでは、これらに対応する潜在変数の存在を(適当な事前分布の元で)仮定すれば、書類選考の可否などの観測可能なデータから、事後分布という形で見積もりを得ることができる(教科書 [2, 3] や、本誌の記事 [4] を参照)。

手始めに、最も単純な確率モデルとして、 $\mathcal{N}(\mu, \sigma^2)$ で平均 μ 、分散 σ^2 の正規分布を表すとき、以下のものを考える:

1. 各求職者 i ごとに「能力値」パラメータ $c_i \sim \mathcal{N}(0, \sigma_c^2)$ をサンプリングする
2. 各求人 j ごとに「難易度」パラメータ $d_j \sim \mathcal{N}(0, \sigma_j^2)$ をサンプリングする
3. 求職者 i が求人 j の書類選考に(応募したとして)通過するか否かは、確率

$$p_{i,j} = \text{sigmoid}(c_i + d_j + b) \quad (1)$$

から定まるベルヌーイ分布であるとする。ここで、バイアス項 b は大きな分散値 Λ^2 をもつ正規分布に従う確率変数 $b \sim \mathcal{N}(0, \Lambda^2)$ で、これによって全体的な合格率(必ずしも平均的に 50%とは限らない)が勘案される。また、sigmoid は $\text{sigmoid}(x) = 1/(1 + \exp(-x))$ で定義されるシグモイド関数である。

上ではあたかも σ_c, σ_j は定数であるかのように書いたが、場合によってはこれも何か確率分布からの標本であると定義してもよい。本稿では、これらの分散値も片側コーシー分布などの無情報事前分布に従う確率変数と考える。このように確率モデルを定義した後は、PyMC3 [5] や Stan [6] に基づいて容易に事後分布の推論を行うことができる。

こうしたモデルでは、「求職者 i が何らかの書類選考に通過する」というイベントがあったとき、「 d が小さい(選考通過が難しい) 求人をパスする場合」と「 d が大きい(選考通過が容易) な求人をパスする場合」とで c の事後分布の姿は異なる。「求職者の優秀さ」を選考通過率だけでなく、「求人難易度」も考慮して見積もろうとすると、今度は「求人を通過している求職者の優秀さ」が必要になってしまう…といったある種のブートストラップ問題は、潜在変数の導入により回避できるのである。

表 1 応募履歴表データのイメージ

求職者	求人	通過	出身校	掲載元	...
1	2	OK	xx 大学	7	...
1	3	NG	xx 大学	3	...
			⋮		
10	2	OK	yy 大学	7	...

2.2 線型モデルのハイブリッド化

さて、バイズ推論的な枠組みにおいては、コールドスタート問題は「行動履歴がないため予測ができなくなる」というよりは「行動履歴がないために、事前分布以上のものが得られておらず、事後分布の分散が大きい」という状況に対応する。ハイブリッド化の目的は、そのほかの属性値を用いることでこの分散をほかの属性に負担させることである。たとえば、引き続き転職サイトの例を用いると、求職者全体での「能力値」パラメータの分散よりも、求職者を（たとえば）年齢や学歴別のセグメントに分けた中で分散のほうが値は小さそうに思える。

この観点から、上の確率モデルを変形していこう。応募履歴を表すデータに求人や求職者の情報をマージし、表 1 のような形のデータが得られたとする。

このとき、上の確率モデルの簡単な拡張として、以下のものが考えられる：

1. 表データの（通過フラグ以外の）各列（col と書く）ごとに、それが取りうる異なる値 v に応じて、スコア値

$$s_v^{(\text{col})} \sim \mathcal{N}(0, \sigma^{(\text{col})2}) \quad (2)$$

をサンプリングする。col は求職者や求人の識別子も含む。

2. 第 i 行 col 列の値を $v_{i,\text{col}}$ と表記するとき、第 i 行の通過フラグが“OK”であるか否かは、確率

$$p_i = \text{sigmoid} \left(b + \sum_{\text{col} \in \text{columns}} s_{v_{i,\text{col}}}^{(\text{col})} \right) \quad (3)$$

から定まるベルヌーイ分布であるとする。ここでも、バイアス項 $b \sim \mathcal{N}(0, \Lambda^2)$ である。

このように書かれてみると、合格確率値を決める部分は、カテゴリ値を説明変数とする単なるロジスティック回帰と同じであることがわかるだろう。ただし、協調フィルタリングの文脈では、しばしば各々の列の取りうるカテゴリの数が膨大になりうることに注意しよう。また、 $(\sigma^{(\text{col})})^2$ は通常の回帰分析の文脈では L2 正則化係数に対応するものであるが、ここではこれらも

（属性ごとに異なる）確率分布に従う、と仮定できる点がやや異なる。ここでは単純化のためにすべての変数をカテゴリ値だとしたが、実数値への拡張も、回帰係数の導入や区分カテゴリへの導入で容易に行うことができる。

それでは、以下に株式会社ビズリーチが運営する“挑戦する 20 代のための転職サービス”である「キャリアトレ」¹における実データに対して、上のモデルの事後分布を求めた例を示そう。データはある期間で取得された、おおよそ 5×10^5 行からなる書類選考合格否データである。このような規模のデータになると、ハミルトニアンモンテカルロ法など、MCMC 系のアルゴリズムによる事後分布の推定は高コストである。そのため、以下の実験では [7] で導入された自動微分変分推論 (Automatic Differentiation Variational Inference, **ADVI**) の PyMC3 [5] における実装を用いた²。ここでは、すべての分散値 $\sigma^{(\text{col})}$ も確率変数とし、その事前分布としては（大きな分散をもつ）片側コーシー分布を仮定した。

表 2 は、完全な協調フィルタリングベース、すなわち求職者と求人のみを説明変数にした場合 (CF-only) と、ハイブリッド型、すなわち求職者や求人の ID 以外の説明変数も取り入れた場合 (Hybrid) の結果比較である。コールドスタート問題の実演のため、ここでは訓練データとテストデータは時間によって 9 : 1 に分割したものをを用いた。ハイブリッド型では、求職者側の変数としてその出身校と年齢を、求人側では求人掲載元企業の識別子を用いた。期待どおり、ハイブリッド化により $\sigma^{(\text{candidate})}$ 、 $\sigma^{(\text{job})}$ の推定値（事後平均）は減少している。また、テストデータに対する ROC-AUC 値の上昇は目覚ましいものがある。なお、興味深い結果として、四年制大学に限定した、出身校のスコア値の事後平均と、その学校の大学入試偏差値³ との対比を図 2 に示す。ここでは、「偏差値」という記法に寄せるため、大学 k の y 軸の値は $50 + 10 \times \mathbb{E}[s_k^{(\text{univ.id})}] / \sigma^{(\text{univ.id})}$ として定義した。

2.3 行列分解による相互作用の導入

さて、前節で考えたモデルはどれもスコア値の構成は線型であったため、「よい求職者はどこでも書類合格しやすい」あるいは「難しい求人は誰でもはねのける」といったことになってしまうが、現実問題としては人、

¹ <https://www.careertrek.com>

² 筆者が用いたのは v3.6 である。

³ <https://www.minkou.jp/university/ranking/deviaton/> より、本稿での使用のみを目的として取得した。

表 2 事後分布の分散と予測性能

	$\sigma^{\text{(candidate)}}$	$\sigma^{\text{(job)}}$	ROC-AUC
CF-only	0.89	1.14	0.58
Hybrid	0.85	0.87	0.84

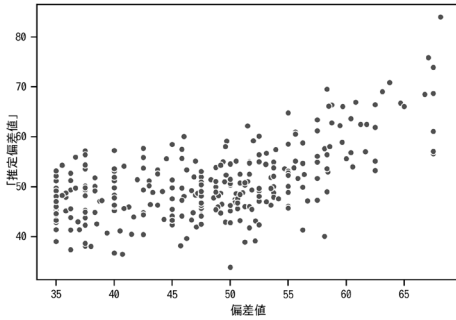


図 2 ロジスティック回帰モデルから読み取った、各出身大学にかかる係数を偏差値風に見せたもの (y 軸) とその大学の偏差値 (x 軸, ここでは全学部全学科についての単純平均とした). 正の相関が見て取れる.

求人の方にも適性が存在し、それらの相性が無視できない可能性がある。以下では、そうした非線形性をいかに捉えるか、という観点について概説する。

このようなコンテキストで「相性」を表現するモデルとしては、早い段階から行列分解 [8][9] の有用性が論じられており、特に Netflix Prize ⁴ のように、ユーザーの嗜好性によって大きくアイテムの評価が変わる問題については強みを発揮してきた。

アウトプットが二値変数であれ連続変数であれ、行列分解の肝は式 (1) でのスコア値 $b + c_i + d_j$ を、

$$b + c_i + d_j + \mathbf{u}_i \cdot \mathbf{v}_j \quad (4)$$

と変更することである。ここで、 $\mathbf{u} \sim \mathcal{N}(0, \sigma_u^2 I_D)$ 、 $\mathbf{v} \sim \mathcal{N}(0, \sigma_v^2 I_D)$ は正規分布 (I_D は D 次元単位行列で、分布は等方的であるとする) に従う D 次元のベクトルで、 $\mathbf{u} \cdot \mathbf{v}$ でもってユーザーと求人の相性を表現するのである。

行列分解の名前の由来は、バイアス項たち (b , c_i と d_j) がなく、かつ観測変数が実数値の場合には、式 (4) はユーザー (求職者) i のアイテム (求人) j へのレーティング行列 R_{ij} を

$$R_{ij} \simeq (U^t V)_{ij}$$

と近似することに相当するためである。

確率モデルによる記述という観点からは、このモデルはロジスティック回帰と大差はなく、引き続き PyMC3 や Stan などで容易に記述ができる。

ヘビーユーザーに対する、行列分解などの協調フィルタリングベースアルゴリズムの強力さを利用しつつ、行動履歴の少ない (ないし皆無な) ユーザーに対するコールドスタート問題を、ユーザー属性値でもって補完するアプローチには多くのものが提案されてきた (近年のサーベイとしては [10] が詳しい)。中でも単純かつ実装が容易なのは [11] で導入された Regression-Based Latent Factor Model (RLFM) とその拡張 [12] である。これは確率モデルとしては 2.2 節のロジスティック回帰モデルのシンプルな拡張であって、

- (ユーザー ID を含む) ユーザーの属性に対応する各列 col ごとに、それが取りうる異なる値 v に応じて、バイアス値

$$c_v^{(\text{col})} \sim \mathcal{N}(0, (\gamma^{(\text{col})})^2) \quad (5)$$

と潜在因子

$$\mathbf{u}_v^{(\text{col})} \sim \mathcal{N}(0, (\sigma^{(\text{col})})^2 I_D) \quad (6)$$

をサンプリングする。

- (アイテム ID を含む) アイテムの属性に対応する各列 col ごとに、それが取りうる異なる値 v に応じて、バイアス値

$$d_v^{(\text{col})} \sim \mathcal{N}(0, (\gamma^{(\text{col})})^2) \quad (7)$$

と潜在因子

$$\mathbf{v}_v^{(\text{col})} \sim \mathcal{N}(0, (\sigma^{(\text{col})})^2 I_D) \quad (8)$$

をサンプリングする。

- 第 i 行 col 列の値を $v_{i,\text{col}}$ と表記するとき、第 i 行の 0/1 目的変数が 1 である確率は

$$p_i = \text{sigmoid}(b + c_i + d_j + \mathbf{u}_i \cdot \mathbf{v}_j) \quad (9)$$

から定まるベルヌーイ分布であるとする。ここで、

$$\begin{aligned} c_i &= \sum_{\text{col} \in \text{user_columns}} c_{v_{i,\text{col}}}^{(\text{col})} \\ d_i &= \sum_{\text{col} \in \text{item_columns}} d_{v_{i,\text{col}}}^{(\text{col})} \\ \mathbf{u}_i &= \sum_{\text{col} \in \text{user_columns}} \mathbf{u}_{v_{i,\text{col}}}^{(\text{col})} \\ \mathbf{v}_i &= \sum_{\text{col} \in \text{item_columns}} \mathbf{v}_{v_{i,\text{col}}}^{(\text{col})} \end{aligned}$$

というモデルである。Factorization Machine [13] と非常に似たモデルであるが、ユーザーの特徴因子同士、

⁴ 背景については Wikipedia の記事 (https://en.wikipedia.org/wiki/Netflix_Prize) を参照

表3 キャリトレにおける「興味あり」予測精度

モデル	ROC-AUC	対数尤度
Linear, CF-Only	0.879	0.389
Linear, Hybrid	0.879	0.389
MF	0.889	0.379
RLFM	0.896	0.369

アイテムの特徴因子同士では相互作用がない点が異なっている。

さて、キャリトレでは、求職者が表示された求人に対して「興味あり」「興味なし」を選択することで、採用企業側に自分の存在を知らせる機能がある。「興味あり」を1、「興味なし」を0と定義すれば、書類合格の場合と全く同様に、未知の組合せに対して興味の有無を予測することができる。表3にこれまで述べてきた種々のモデルによる、テストデータに対する性能を記す。

表中の Linear, CF-Only は 2.1 節で考えたモデルに、Linear, Hybrid は 2.2 節で考えたモデルに相当する。MF は (4) の素朴な行列分解アルゴリズムを用いた結果である。

表3のとおり、RLFMが精度でほかを上回っている。また、単純な行列分解も属性値を加味したロジスティック回帰を上回っていることに、本データにおける行列分解の優位性がうかがえる。なお、前述の書類合格データに対しても行列分解を適用したところ、ロジスティック回帰と大きく精度は変わらなかった。書類合格する／しないよりも、求人に対する好き嫌いのほうが、より相性が重視される、ということは直感に即しているのではないだろうか。

近年では深層学習によるさまざまな方法が提案されているが、単純な行列分解でもチューニング次第ではまだまだ最高の精度を出す場合も多いようだ [14]。

最後に、学習されたベクトル \mathbf{u} と \mathbf{v} はユーザー／アイテムの特徴量ベクトルとして用いることができるので、これらをもとにした類似ユーザー／類似アイテムの推薦に使われることも多い。キャリトレの採用企業向けの機能である類似会員推薦には、2019年9月以降は行列分解と K-means 法に基づくアルゴリズムが採用されている。

3. 推薦アルゴリズム以外への応用

前節では、行動履歴と属性値からいかに「求職者が合格しやすい求人」や「求職者が興味をもちそうな求人」をバイズ推論の枠組みから見つけ出すか、について論

じた。「推薦アルゴリズム」という観点からは、最も重要な指標は推薦の予測性能であり、この向上のためには時として深層学習などの非線形かつブラックボックスな手法に頼ることも必要である。しかし、場合によっては推薦そのものよりも、その過程で得られる (図2のような) 各属性に紐づくスコアの推論値に重要な示唆が含まれていることがある。本節では、2.1 節のロジスティック回帰モデルの発展系として、

- 複数の面接官による候補者評価、というノイズなデータから「面接官の厳しさ」を差し引いて候補者の能力値を推定する
- サービスを運営側の期待に反した、偏った方法で利用しているユーザーを発見する

という応用事例について紹介する。

3.1 面接評価からの能力推定

これまでは可否データ (あるいは “like” or “hate”) のように、観測変数が 0/1 で表せるような場合を考えてきたが、これが実数値出力である場合への拡張は容易であろう。ここでは、もう少し捻りが必要な例として、ビズリーチが運営するクラウド型採用管理サービス「HRMOS 採用管理」における面接評価データの事例について紹介する。

ここで取り扱うデータは株式会社ビズリーチにおいて 2017 年に実施された面接記録であり、

1. 中途／新卒にかかわらず、採用面接は一人の候補者に対して複数人の面接官によって行われる
2. 一回の面接ごとに、面接官は候補者を S, A, B, C の 4 グレードで評価する (それぞれ優, 良, 可, 不可に相当する)
3. 面接には、通常の面接に加えて「最終面接」などのカテゴリが存在する

となっている。人事担当者にとって悩ましいのは、面接官によって S, A, B, C の基準が異なることが多々あることである。また、面接が「最終面接」であるか通常の面接であるか、などによっても基準は異なると信じられている。そこで、ここでも上の書類合格の例と同様に、候補者の「能力値」パラメータ $s^{(\text{candidate})}$ と面接官の「優しさ」パラメータ $s^{(\text{interviewer})}$ および面接ステージごとのバイアス項 $s^{(\text{round})}$ を導入し、

$$s^{(\text{total})} = b + s^{(\text{candidate})} + s^{(\text{interviewer})} + s^{(\text{round})} \quad (10)$$

から面接の 4 段階グレードが定まる、と仮定しよう。

グレードはカテゴリ値ではあるものの、順序関係がある値であり、それがスコアからどのように定まるか

は非自明である。単純な対処法としては、たとえば S, A, B, C を整数値 3, 2, 1, 0 と読み替えて実数値の予測問題とすることも考えられるが、必ずしも S と A の差分が B と C の差分と等しいとは限らない。このような状況で有用なのは、順序ロジット分布である。この分布では、まず S と A, A と B, B と C の間にそれぞれ閾値 $\theta_S > \theta_A > \theta_B$ が存在するとする。そして、「観測グレードが g 以上である確率」を

$$p(\text{grade} \geq g | s^{(\text{total})}) = \text{sigmoid}(s^{(\text{total})} - \theta_g) \quad (11)$$

によって定めるのである。これがわかれば、自明な関係式

$$\begin{aligned} p(\text{grade} = S) &= p(\text{grade} \geq S) \\ p(\text{grade} = A) &= p(\text{grade} \geq A) - p(\text{grade} \geq S) \\ p(\text{grade} = B) &= p(\text{grade} \geq B) - p(\text{grade} \geq A) \\ p(\text{grade} = C) &= 1 - p(\text{grade} \geq B) \end{aligned}$$

によって、各々のグレードの出現確率が求まる。よりグレードの数が多い場合も同様にして、グレード数 - 1 個の閾値を導入することで分布を定義できる。

上で新たに導入した閾値 $\{\theta\}$ は、これ自体確率変数であるとして推論を行うことで、自然な閾値を求めることができる。今回は、バイアス項と分散のスケール自由度により θ_S と θ_B はそれぞれ +1 と -1 に固定できるので、 θ_A のみを推論するとしよう。 $\theta_A \sim \text{Uniform}(-1, 1)$ として、前節と同じく ADVI によって事後分布の推論を行ったところ、 $\theta_A = 0.22 \pm 0.01$ という結果が得られた。 $(\theta_S - \theta_A) < (\theta_A - \theta_B)$ であるので、標語的に言うと、「S と A の差は、A と B の差よりも小さい」ということになる。

図 3 には、候補者の平均グレード (S→3, A→2, B→1, C→0 として算出した) を x 軸に、候補者の推定能力値 (事後分布の平均値) を y 軸に示す。どちらもグレードをもとに算出しているものであるため、両者の相関は極めて強いものの、時として外れ値が見られ、これは厳しい (甘い) 面接官ばかりに遭遇したためと考えられる。

また、図 4 には、図 3 の中のボーダーライン、すなわち平均グレードが B であるところを切り出し、合否に応じて推定能力値をプロットしたものである。微妙ながら実際に合格しているほうが推定能力値は高く (そこまで有意水準が大きいわけではなかったが)、おおむね「同じ平均 B でも、面接官の厳しさを差し引いて合否判断がなされていたようだ」と思われる。

図 5 には、面接官が与えた評価の平均値とその推定

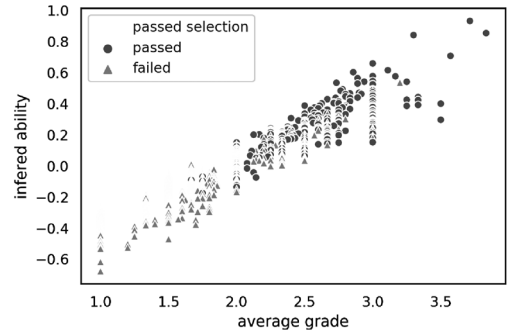


図 3 面接官の厳しさを考慮に入れても、平均グレードとの相関は強い。が、稀に外れ値があるのは見逃せない。

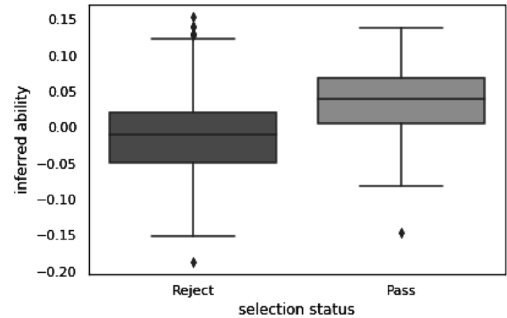


図 4 図 3 を、面接の平均グレード = 'B' に絞ったもの。同じ 'B' でも、合格している場合の方が推定能力値は高いように見える。

寛容さの関係が示されている。こちらの相関は図 3 より小さい。これは、面接官側にはある程度「担当範囲」があるため、たとえばマネージャークラスの面接官には普通は二次面接以降でしか遭遇しないため、と考えられる。

この単純なモデルを発展させ、文献 [15] では、面接ごとのパフォーマンスの揺らぎを考慮したり、また、甘い面接官と厳しい面接官でどのようにグレードに応じて事後分布が異なるか、などが詳細に論じられているので、興味のある読者はそちらも参照されたい。

3.2 人材ネットワークに潜む“偏り”の検出

本節で対象とするのは OB/OG 訪問マッチングサービス「ビズリーチ・キャンパス」における事例である。2019 年初頭、本サービスの競合サービスを悪用した社会人による不祥事が相次いで発覚し、大きな社会問題となったことは記憶に新しい。ビズリーチ・キャンパスにおいても、この事件は他人事ではなく、さまざまな取り組みを実施している。本節では、これまで述べてきたベイズ推論の手法を変形し、OB/OG 会員のリスク因子を推定する、という事例について紹介する。

まず、前提として、OB/OG 訪問フローは

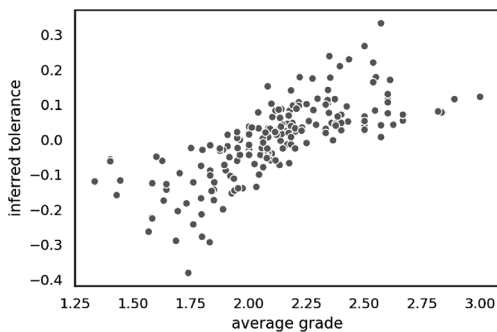


図5 面接官が与えた評価の平均値と推定寛容さ $s_g(\text{interviewer})$

1. 学生会員の側から、訪問対象企業のOB/OGへ訪問依頼メッセージを送付する
2. OB/OG会員は、訪問を許可するのであれば返信メッセージを送付する
3. 以降、日程調整などのやりとりが続く

という形をとっている。メッセージの文面から自然言語処理ベースで問題のあるメッセージを発見するには、教師データがアノテーション（そもそもアノテーション自体も個人情報保護という観点から問題がある）を行わなければ得られない、という大問題があった。そのため、自然言語処理よりも行動履歴ベースで問題のあるユーザーを見つけ出す方向に舵が切られた。

根本となるアイデアは、仮に不適切な目的で性的マジョリティに属するユーザーが本サービスを使うのであれば、同性の後輩からの訪問依頼と異性からのそれとで返信率が異なるはず、という仮定である（性的マイノリティに属するユーザーの分析は別途必要となる）。ただ、ここで単純に返信率を指標として用いるには

- メッセージの件数が少ない、新しい会員に対しては返信率の分母が0になってしまいがちで、対処法がわからない
- たまたま「性別にかかわらずぜひ応援したくなる異性の後輩」からメッセージが来てOKしたが、同性の後輩からのメッセージは来なかった場合に、「高リスク会員」と判定されてしまう

という課題がある。そこで有用なのがベイズ推論である。ここでは、以下の確率モデルを考える：

1. 学生会員 i ごとに「能力値」 $s_i \sim \mathcal{N}(0, \sigma_s^2)$ をサンプリングする
2. OB/OG 会員 j ごとに「寛容さ」 $o_j \sim \mathcal{N}(0, \sigma_o^2)$ と「異性優先係数」 $g_j \sim \mathcal{N}(0, \sigma_g^2)$ をサンプリングする。
3. 学生会員 i の OB/OG 会員 j への訪問依頼が受

諾されるか否かは、確率

$$p_{i,j} = \text{sigmoid}(s_i + o_j + I(i,j)g_j + b) \quad (12)$$

から定まるベルヌーイ分布に従うとする。 b は式 (1) と同様のバイアス項である。 $I(i,j)$ は、学生会員 i と OB/OG 会員 j が異性同士の組合せであれば 1、そうでなければ 0 という値を取るものとする。

すなわち、異性優先係数 g_j は、その真意はさておき「OB/OG ユーザー j がどの程度異性に対してゲタを履かせるか」という意味合いをもつ。このモデルによれば、 s_i の高い異性のユーザーからの訪問を OK して、 s_i の低い同性のユーザーからの訪問を断った場合でも、 g_j の増分はさほど大きくならない。また、行動履歴のないユーザーに対しては、事前分布を信じるという明快な指針がある。

こうして得られた異性優先係数の事後分布をリスクの目安とすることができるが、折角事後分布の中央値だけでなく、分散値も求まっているので、リスク値には

$$\mathbb{E}[g_j] + a \times \sqrt{\text{Var}[g_j]}$$

を目安としたいところである。ここで $a > 0$ はグローバルな定数で、これによって「分散が大きいところもいったん高リスクとして警戒すべし」という意味合いをもたせることができる。 a の値はなるべく大きく取るほうが安全であるが、あまりに大きい値を設定すると、今度は期待値よりも分散が大きいものが優先されてしまう、というトレードオフがある。

そこで、分散の低減のために、2.2 節での議論を踏襲し、ほかの属性値を用いることを考えよう。ここでは、以下の分解

$$\begin{aligned} s_i &\leftarrow s_i^{(\text{all})} = s_{\text{university}[i]}^{(\text{university})} + s_i \\ o_j &\leftarrow o_j^{(\text{all})} = o_{\text{company}[j]}^{(\text{company})} + o_j \\ g_j &\leftarrow g_j^{(\text{all})} = g_{\text{company}[j]}^{(\text{company})} + g_j \end{aligned}$$

を考える。これは、学生の能力値を大学ごとの中心値 + 個人ごとの分散に、OB/OG の寛容さと異性優先係数を企業ごとの中心値 + 個人ごとの分散と考えることに相当する。

これにより、異性優先係数の分散 σ_G の推定値は $0.900 \pm 0.006 \rightarrow 0.634 \pm 0.006$ へと減少した。また、異性優先係数全体 $g^{(\text{all})}$ の事後分散の平均も、 $0.732 \rightarrow 0.706$ と減少が見られた。この事実から、「企業ごとのカラー」というものは存在していて、それとなく行動履歴にも反映されている、ということが察せ

られるのではないだろうか⁵。

4. 数理最適化を用いたマッチングの最大化

ビズリーチのような転職サービスでは、映画の推薦などユーザーの嗜好性が大きく反映されるものと比べ、人気企業の求人に応募が集中しやすい。また、逆の立場からすると、企業はなるべく優秀だとわかりやすい求職者に求人を露出しようとするため、そこでも偏りが生じてしまう。この偏りにより、以下のような現象が生じがちである：

1. 書類合格率をもとにする推薦システムでは、新規求職者や新規参入企業に推薦が回ってこない(データを取得する機会もいつまでも生じない)
2. 人気企業に大量の応募が殺到し、処理に大きな手間がかかる

これらはユーザー体験を損ねるだけではなく、企業の基本使用料金が一定というビジネスモデルを採用している場合は、公平性の観点からも問題となる。

本節では、ユーザー(求職者) i とアイテム(求人あるいは採用企業) j のマッチング確率 $p_{i,j}$ がなんらかの方法で求まっている場合に、線形計画法でもって種々の制約条件を課してユーザー体験を損なわないようにしつつ、マッチングの最大化を行う取り組みについて簡単に紹介する。これらはよく知られた二部マッチング最大化問題であるが、線形不等式の工夫により、さまざまなビジネス上の制約を考慮することができるのである。

4.1 単純な二部マッチングの例

まず、このような状況の単純な例として、就活生のためのマッチングサイトである「ニクリーチ」⁶の機能の一つを考えよう。この機能では、

1. 運営から求職者(学生会員)へ演算した推薦企業を配信する
2. 求職者は推薦された企業の中から興味ある企業に対してアプローチする
3. アプローチされた企業からOKが出れば、学生と企業が交流が実現する

という動線が提供される。

こうした場合、求職者 i に企業 j を推薦するか否かを表す二値変数 $x_{i,j} \in \{0, 1\}$ を導入するのが定石である(もちろん $x_{i,j} = 1$ が「推薦を行う」に相当する)。

この変数を用いると、学生会員 i が受け取る推薦企業数は $\sum_{j \in U^{(i)}} x_{i,j}$ 、企業 j が学生会員に推薦される総数は $\sum_{i \in I^{(j)}} x_{i,j}$ で与えられる。ただし、 $U^{(i)}$ はユーザー i とマッチしうる企業IDの集合、 $I^{(j)}$ は企業 j とマッチしうるユーザーIDの集合であり、たとえばすでに応募があった組合せについてはこれらから当該ペアが除外される。

こうして、企業ごとの推薦数は区間 $[C_1, C_2]$ におさまる「ほどほど」の値にし、ユーザーには一定数の求人 R を推薦しつつ期待値を最大化する問題は、以下のような二値変数の整数計画問題として定式化される。

$$\begin{aligned} & \text{maximize} && \sum_i \sum_{j \in U^{(i)}} p_{i,j} x_{i,j} \\ & \text{subject to} && \sum_{j \in U^{(i)}} x_{i,j} = R, \quad i = 1, \dots, N_{\text{user}} \\ & && \sum_{i \in I^{(j)}} x_{i,j} \geq C_1, \quad j = 1, \dots, N_{\text{company}} \\ & && \sum_{i \in I^{(j)}} x_{i,j} \leq C_2, \quad j = 1, \dots, N_{\text{company}}. \end{aligned}$$

このような問題は、整数計画問題ではあるが、二部グラフ構造のお陰で拘束条件が完全単模行列(すなわち任意の部分行列の行列式が $\{0, 1, -1\}$ のいずれか)であるため、単体法による最適解が整数性を満たす。完全単模性(total unimodularity)については、たとえば[16]の第8章を参照されたい。

当機能の開発時点では、ユーザー数は数万単位、導入企業数はおおよそ 10^2 であるため、場合の数は 10^7 におよばなかった。この規模であればソルバーで数分で求解できるので、日次でバッチ実行を走らせている。

4.2 キャリトレにおける求人案内問題

2019年9月にリリース予定のキャリトレの新バージョンにおける「求人案内」機能にも数理計画法を応用した推薦アルゴリズムが搭載される予定である。ここでの問題設定は、

- 求人案内は、企業ごとではなく、求人ごとに送付される。一つの企業の複数求人が単一の求職者に送付されることも許容される。
- 各企業は、自社の各々の求人ごとに必要であればフィルタ条件(たとえば「フリーテキストに以下の文言を含む求職者」や「在籍企業名に以下を含めない」など)を指定する
- 企業の掲載可能な求人数には基本的には制限がなく、企業によって振れ幅は大きい

という点でニクリーチのそれとは異なっている。このように、「求職者」「求人」「求人掲載企業」の三者が存在

⁵ 企業名と異性優先係数との照合結果は納得感の得られるものであったが、極めて秘匿性の高い情報でもあるため、公開は差し控えた。

⁶ <https://29reach.com>

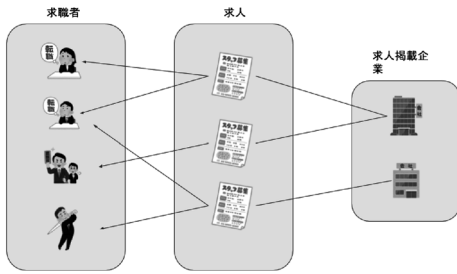


図6 キャリトレにおける「三層グラフ」構造

するという観点からは、むしろここでのネットワークは三層グラフと呼ぶほうがより正確であろう(図6)。

そこで、今回はユーザー i への求人 j 配信を表す二値変数 $x_{i,j}$ に加え、各求人 j ごとに一つずつ正整数の補助変数 $y_j \geq 0$ を導入し、以下のように問題を定式化する：

$$\begin{aligned}
 & \text{maximize} && \sum_i \sum_{j \in U^{(i)}} p_{i,j} x_{i,j} \\
 & \text{subject to} && \sum_{j \in U^{(i)}} x_{i,j} \leq R, \quad i = 1, \dots, N_{\text{user}} \\
 & && \sum_{i \in I^{(j)}} x_{i,j} = y_j, \quad j = 1, \dots, N_{\text{job}} \\
 & && \sum_{j \in J^{(k)}} y_j \leq C, \quad k = 1, \dots, N_{\text{company}}
 \end{aligned}$$

ここでも、 $U^{(i)}$ はユーザー i とマッチしうる求人 ID の集合、 $I^{(j)}$ は各求人 j とマッチしうるユーザー ID の集合であり(まだ応募がない、かつフィルタ条件を満たすユーザーの ID の集合)であるが、それに加えてここではさらに $J^{(k)}$ で企業 k が掲載している求人 ID の集合を表す。

この拘束条件の設定により、1 ユーザー当たりの配信数は R 以下、一企業(求人ではなく)当たりの最大配信数は C 以下であることが保証される。この新しい問題でも完全単模性が保たれることは、上の三層グラフ構造を変形し、企業 k に対応するダミーユーザーが存在すると考えることによってわかる。(ただし、拘束条件に一部負符号が現れることによって、若干数値的安定性が損なわれてしまう)。

こうして、あとは線形緩和した問題を解けばよい訳であるが、キャリトレの配信対象ユーザー数と求人数はともに $O(10^5)$ であり、場合の数はニクリーチと比べて膨大で、とてもソルバーの手に負える問題ではない。そこで、ここではアドホックではあるが求職者・求人を M 個にランダムに分割し、分割 m に属する求職者には分割 m の求人のみがレコメンドされる、として場

合の数の低減を図っている。また、 m 番目の問題における拘束条件では、もとの拘束条件 $\sum_{j \in J^{(k)}} y_j \leq C$ を、

$$\sum_{j \in J^{(k)}} y_j \leq \text{floor} \left(C \times \frac{N_{k,m}^{(\text{job})}}{N_k^{(\text{job})}} \right)$$

と置き換える。ここで、 $N_{k,m}^{(\text{job})}$ は企業 k の求人のうち、分割 m に属するものの数、 $N_k^{(\text{job})}$ は企業 k の求人の総数 ($N_k^{(\text{job})} = \sum_{m=1}^M N_{k,m}^{(\text{job})}$) である。

5. おわりに

本稿では、「人材ネットワーク」、一すなわちユーザー(求職者や面接官など)とアイテム(求人やまた別のカテゴリの人材)からなる二部グラフで、一方が一方を評価するという構造—に対する、株式会社ビズリーチの AI 室での取り組みについて述べた。序論でも述べたとおり、筆者の担当範囲と趣味により、扱うモデルはベイズ推論による単純なモデルが大半を占めてしまったが、目的をきちんと設定すれば、単純なモデルでも多くのことを教えてくれる、ということが伝われば幸いである。

最後に、4 節で述べたように、人材業界のような、それほど個人の嗜好が効いてこない領域においては、人気が特定のユーザー/アイテムに一極集中しがちな点は大きな問題となり、この解決のためには数理最適化を中心とする OR 技術で種々の制約を盛り込んでいくことが不可欠である、と考える。株式会社ビズリーチで OR 技術を遺憾なく発揮されたい方は、ぜひ面談に来ていただければ、ベイズ推論で能力値を推定させていただく所存である(冗談です)。

謝辞 本稿執筆の機会を提供して下さった、中央大学の生田目先生、草稿について多くの有益なコメントをいただいた Retty 株式会社の岩永二郎氏に感謝します。また、大西佑紀氏にはベイズ推論について多くのことを教わりました。この場を借りて御礼申し上げます。

参考文献

- [1] C. Aggarwal, *Recommender Systems*, Springer, 2016.
- [2] R. McElreath, *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*, Chapman and Hall/CRC, 2016.
- [3] J. Kruschke, *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*, Academic Press, 2014.
- [4] 岩田具治, “確率的潜在変数モデルに基づくデータマイニ

- ング,” オペレーションズ・リサーチ：経営の科学, **64**(5), pp. 272–277, 2019.
- [5] J. Salvatier, T. V. Wiecki and C. Fonnesbeck, “Probabilistic programming in Python using PyMC3,” *Peer J Computer Science*, **2**, article number: e55, 2016.
- [6] B. Carpenter, A. Gelman, M. D. Hoffman, D. I. Lee, B. Goodrich, M. I. Betancourt, M. Brubaker, J. Guo, P. Li and A. Riddell, “Stan: A probabilistic programming language,” *Journal of Statistical Software*, **76**, Issue 1, 2017.
- [7] A. Kucukelbir, D. Tran, R. Ranganath, A. Gelman and D. M. Blei, “Automatic differentiation variational inference,” *The Journal of Machine Learning Research*, **18**, pp. 430–474, 2017.
- [8] A. Mnih and R. R. Salakhutdinov, “Probabilistic matrix factorization,” *Advances in Neural Information Processing Systems*, pp. 1257–1264, 2007.
- [9] T. Hofmann, “Probabilistic latent semantic analysis,” In *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, pp. 289–296, 1999.
- [10] W.-H. Chen, C.-I. Hsu, Y.-A. Lai, V. Liu, M.-Y. Yeh and S.-D. Lin, “Attribute-aware collaborative filtering: Survey and Classification,” arXiv: 1810.08765, 2018.
- [11] D. Agarwal and B. C. Chen, “Regression-based latent factor models,” In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 19–28, 2009.
- [12] S. Park, Y. D. Kim and S. Choi, “Hierarchical Bayesian matrix factorization with side information,” In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*, pp. 1593–1599, 2013.
- [13] S. Rendle, “Factorization machines,” In *Proceedings of the IEEE International Conference on Data Mining*, pp. 995–1000, 2010.
- [14] S. Rendle, Z. Li and Y. Koren, “On the difficulty of evaluating baselines: A study on recommender systems,” arXiv: 1905.01395, 2019.
- [15] Y. Ohnishi and S. Sugaya, “2019 Applying Bayesian hierarchical probit model to interview,” In *Proceedings of International Workshop on Talent and Anagement Computing*, to appear.
- [16] A. Schrijver, “A course in combinatorial optimization,” <https://homepages.cwi.nl/~lex/files/dict.pdf> (2019年10月8日閲覧)