3K3-OS-20a-6

コンテスト型クラウドソーシングにおける勝敗予測

Winning Prediction for Crowdsourcing Contests

馬場 雪乃 *1　木下 慶 *2　鹿島 久嗣 *1
Yukino Baba　Kei Kinoshita　Hisashi Kashima

*1京都大学 Kyoto University　*2ランサーズ株式会社 Lancers Inc.

We propose a novel participation recommendation approach for crowdsourcing contests including probabilistic contest participation and winner determination models. Our method estimates the winning and participation probability of each worker and offers ranked lists of recommended contests. Since there is only one winner in most contests, standard recommendation techniques fail to estimate the accurate winning probability using only the extremely sparse winning information of completed contests. Our solution is to utilize contest participation information and features of workers and contests as auxiliary information. We use the concept of a transfer learning method for matrices and a feature-based matrix factorization method. Experiments conducted using real crowdsourcing contest datasets show that the use of auxiliary information is crucial for improving the performance of contest recommendation.

1. はじめに

コンテスト型は，デザインやソフトウェア開発のような創造性が求められる複雑な仕事をクラウドソーシングする際に用いられる発注方法である．この発注方法では，一つの発注に対して複数のワーカが成果物を提出し，発注者によって選ばれた（主に一人の）ワーカ，つまりコンテストの勝者だけが報酬を獲得する．報酬獲得の可否が他の参加者が提出した成果物の品質に依存し，ワーカが報酬を得るためには勝利可能性が高いコンテストの選択が重要となる．

本稿では，コンテスト型クラウドソーシングにおける確率的な勝者決定モデルとその推定方法を提案する．また，コンテスト推薦への応用を示す．提案手法は，完了したコンテストでの勝敗情報を用いて，未完了コンテストでの各ワーカの勝利確率を予測する．ワーカの勝利確率は，ワーカの能力とワーカとコンテストの相性によって決まるものとしてモデル化する．各コンテストの勝者は一人だけであり，未完了コンテストでは勝者が決定していないというコンテスト型クラウドソーシングの性質上，勝敗情報は非常に疎となりモデル推定が困難である．

観測情報の疎性に対処するため，我々は補助情報を利用した推定方法を提案する．具体的には，コンテストの参加者情報と，コンテストとワーカの特徴を補助情報として利用する．多くの場合，一つのコンテストには複数人が参加するため参加者情報は勝者情報よりも密となり，ワーカ間・コンテスト間の類似度推定に利用できる．また，ワーカとコンテストの特徴もワーカ間・コンテスト間の類似度推定に有効と考えられる．我々は，参加者行列の分解結果からの転移学習 [Pan 10] と，特徴に基づく行列分解手法 [Chen 12] をモデル推定に適用する．

2. コンテスト型クラウドソーシングにおける勝敗予測問題

2.1 問題設定

J 個のコンテストと，これらのコンテストに参加可能な I 人のワーカがいるとする．勝者行列 $\mathbf{\Sigma} \in \{0,1,?\}^{I\times J}$ は，i 番目のワーカが j 番目のコンテストの勝者か $(\Sigma_{ij}=1)$ 否か $(\Sigma_{ij}=0)$ を表す．“?” は欠損値を表す．勝者決定済みの完了コンテストの集合を $\mathcal{J}_f$ で表す．すべての $j \in \mathcal{J}_f$ について $\Sigma_{ij} \in \{0,1\}$ となる．勝者未決定の未完了コンテストの集合を $\mathcal{J}_u$ で表す．すべての $j \in \mathcal{J}_u$ について Σ_{ij} には “?” が割り当てられている．各コンテストの勝者は一人とする．従って，$\mathbf{\Sigma}$ の各列は，最大でも一つの要素だけが “1” となる．各 $j \in \mathcal{J}_f$ について，勝者のインデックスを σ_j で表す．

参加者行列 $\mathbf{\Gamma} \in \{0,1,?\}^{I\times J}$ は，i 番目のワーカが j 番目のコンテストの参加者か $(\Gamma_{ij}=1)$ 否か $(\Gamma_{ij}=0)$ を表す．コンテストが完了した時点ではすべての参加者がわかっているため，すべての $j \in \mathcal{J}_f$ について $\Gamma_{ij} \in \{0,1\}$ となる．また，勝者は参加者から選ばれるため $\Gamma_{\sigma_j j}=1$ となる．

各ワーカとコンテストについて，それぞれ N, M 次元の特徴ベクトルが与えられているとする．ワーカの特徴行列を $\mathbf{\Phi} \in [0,1]^{I\times N}$，コンテストの特徴行列を $\mathbf{\Psi} \in [0,1]^{J\times M}$ と表す．

我々の目的は，$\mathbf{\Sigma}$，$\mathbf{\Gamma}$，$\mathbf{\Phi}$，$\mathbf{\Psi}$ が与えられたときに未完了の各コンテスト $j \in \mathcal{J}_u$ について，各ワーカの勝利確率 $\Pr[\Sigma_{ij}=1]$ を予測することである．

2.2 データセット

我々は，2008 年 12 月から 2013 年 4 月までにランサーズ *1 で開催されたコンテストのデータを収集し，「プロジェクト方式」と「コンペ方式」の二種類のコンテストのデータセットを作成した．プロジェクト方式はソフトウェア開発などで主に用いられる発注方式である．ワーカは作業計画書を提出してコンテストに参加する．仕事の発注者は計画書を見て，仕事を発注するワーカを決定する．選ばれたワーカが仕事を完了した時点で報酬が支払われる．コンペ方式はデザインなどで用いられる発注方式である．この方式では，ワーカはほとんど完成した成果物を提出してコンテストに参加する．リクエスタは成果物を評価し，最良の成果物を選択する．その成果物の作成者に報酬が支払われる．

我々は，それぞれの発注方式において，6 個以上のコンテストで勝者となったワーカを選択してデータセットを構築した．プロジェクト方式のデータセットは 5,703 個のコンテストと 458 人のワーカから成る．コンテスト方式は，12,384 個のコ

連絡先: 馬場 雪乃, ybaba@nii.ac.jp

*1 http://www.lancers.jp/

ンテストと 602 人のワーカから成る．各データセットにおいて，コンテストの平均参加人数は，プロジェクト方式で 2.16 人，コンペ方式で 18.77 人である．プロジェクト方式データセットはウェブアプリケーション開発や文章執筆等の多様な仕事のコンテストを含むのに対してコンテスト方式データセットは，84.6% がデザインのコンテストである．二つのデータセットは，それぞれコンテストとワーカの特徴行列を含む．コンテストの特徴は，仕事のカテゴリ・発注者 ID・報酬額・仕事の説明中の単語等から成る．ワーカの特徴は，年齢・性別・職業・ワーカの自己紹介文中の単語等から成る．コンテストの特徴はプロジェクト方式では 6,571 次元，コンペ方式では 16,365 次元である．ワーカの特徴はプロジェクト方式では 2,610 次元，コンペ方式では 2,636 次元である．

3. コンテスト型クラウドソーシングの確率モデル

3.1 勝者決定モデル

ワーカ i がコンテスト j で発揮する性能を $X_{ij} \in \mathbb{R}$ で表す．性能 X_{ij} は，ワーカ i とコンテスト j の相性 A_{ij} と，ワーカ i の能力 b_i の和 $X_{ij} = A_{ij} + b_i$ で決まるとする．

ワーカ i がコンテスト j の勝者になる確率を，性能に対するソフトマックス関数で表現する．参加者行列が全て観測されているとき，つまり，すべての j について $\Gamma_{ij} \in \{0, 1\}$ のときは，勝利確率を

$$\Pr\left[\Sigma_{ij} = 1\right] = \frac{\Gamma_{ij} \exp\left(X_{ij}\right)}{\sum_p \Gamma_{pj} \exp\left(X_{pj}\right)} \tag{1}$$

と表す．ここで，ワーカはコンテストに参加していなければ勝者にはならないため，$\Gamma_{ij} = 0$ のとき $\Pr\left[\Sigma_{ij} = 1\right] = 0$ である．このモデルでは，他の参加者の性能が低いほど，あるいは参加者数が少ないほど勝利確率が高くなる．

次に，参加者行列が部分的に観測されている場合の勝利確率を考える．式 (1) において Γ_{ij} を参加確率 $\Pr\left[\Gamma_{ij} = 1\right]$ で置き換え，勝利確率を次式で表す：

$$\Pr\left[\Sigma_{ij} = 1\right] = \frac{\Pr\left[\Gamma_{ij} = 1\right] \exp\left(X_{ij}\right)}{\sum_p \Pr\left[\Gamma_{pj} = 1\right] \exp\left(X_{pj}\right)}. \tag{2}$$

3.2 参加モデル

次に，参加確率 $\Pr\left[\Gamma_{ij} = 1\right]$ のモデルを考える．ワーカのコンテストへの参加判断は，各ワーカが独立に行い他の参加者の影響されないと仮定し，ロジスティック関数で参加確率を表現する．すなわち，

$$\Pr\left[\Gamma_{ij} = 1\right] = \frac{1}{1 + \exp\left(-\tilde{X}_{ij}\right)} \tag{3}$$

となる．ここで $\tilde{X}_{ij} \in \mathbb{R}$ はコンテスト j に対するワーカ i の参加意欲を表すとする．性能モデルと同様にして，ワーカの参加意欲 $\tilde{X}_{ij}$ は複数の要素の和 $\tilde{X}_{ij} = \tilde{A}_{ij} + \tilde{b}_i + \tilde{c}_j + \tilde{d}$ で表現されるとする．

4. モデル推定

4.1 勝者決定モデルの推定

4.1.1 MAP 推定

$\boldsymbol{A} = [A_{ij}]_{i,j}$ と $\boldsymbol{b} = (b_1, \ldots, b_I)^\top$ をパラメータとする勝者決定モデルの対数尤度は，

$$L\left(\boldsymbol{A}, \boldsymbol{b}\right) = \sum_{j \in \mathcal{J}_f} \left(A_{\sigma_j j} + b_{\sigma_j} - \log \sum_{p \in \mathcal{I}_j} \exp\left(A_{pj} + b_p\right) \right)$$

で与えられる．$\|\boldsymbol{A}\|_{\mathrm{F}}^2 + \|\boldsymbol{b}\|_2^2$ を正則化項とする MAP 推定により，完了コンテストのパラメータは推定することができる．しかし，勝利確率の予測対象である未完了コンテスト ($j \in \mathcal{J}_u$) では勝者が決まっていないため，コンテストのパラメータは推定することができない．また，一度も，あるいは少ない回数しか勝者になっていないワーカのパラメータを推定するのは困難である．我々は，補助情報を用いてこの二つの問題に対処する．

4.1.2 参加者行列からの転移学習

我々は，あるコンテストにおけるワーカの性能は，そのワーカの当該コンテストへの参加意欲と関連があると考えた．そこで，参加者情報を用いて勝者情報の不足を補う．各コンテストの参加者は多くの場合複数であるため，参加者行列は勝者行列よりも密である．参加者行列からの転移学習のために，行列分解にもとづく転移学習手法の一つ CST 法 [Pan 10] を利用する．CST 法では，(比較的密な) 転移元行列を分解し，補助行列の分解結果に近づくように転移先行列を分解する．

CST 法を用いて，ワーカ・コンテスト間の相性パラメータ $\boldsymbol{A}$ とワーカの能力パラメータ $\boldsymbol{b}$ を推定する．$\boldsymbol{A}, \boldsymbol{b}$ を，それぞれ $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{G}\boldsymbol{V}^\top$，$\boldsymbol{b} = \boldsymbol{U}\boldsymbol{h}$ と分解することを考える．ここで，$\boldsymbol{U} \in \mathbb{R}^{I \times K}$, $\boldsymbol{V} \in \mathbb{R}^{J \times K}$, $\boldsymbol{G} \in \mathbb{R}^{K \times K}$, $\boldsymbol{h} \in \mathbb{R}^K$ であり，K は自然数とする．まず，参加者行列の K ランク分解により $\boldsymbol{\Gamma} \sim \tilde{\boldsymbol{U}}\tilde{\boldsymbol{G}}\tilde{\boldsymbol{V}}^\top$ を得る．ここで，$\tilde{\boldsymbol{U}} \in \mathbb{R}^{I \times K}$, $\tilde{\boldsymbol{V}} \in \mathbb{R}^{J \times K}$ である．$\tilde{\boldsymbol{G}}$ は $K \times K$ の対角行列とする．次に，次式の最大化により $\boldsymbol{h}$, $\boldsymbol{G}$, $\boldsymbol{U}$, $\boldsymbol{V}$ を推定する：

$$\begin{aligned} &L\left(\boldsymbol{h}, \boldsymbol{G}, \boldsymbol{U}, \boldsymbol{V}\right) \\ &= \sum_{j \in \mathcal{J}_f} \left(X_{\sigma_j j} - \log \sum_{i \in \mathcal{I}_j} \exp\left(X_j\right) \right) \\ &- \lambda \left(\|\boldsymbol{h}\|^2 + \|\boldsymbol{G}\|_{\mathrm{F}}^2 \right) - \eta \|\boldsymbol{U} - \tilde{\boldsymbol{U}}\|_{\mathrm{F}}^2 - \kappa \|\boldsymbol{V} - \tilde{\boldsymbol{V}}\|_{\mathrm{F}}^2. \end{aligned} \tag{4}$$

ここで $X_{ij} = \sum_{k,l} U_{ik} G_{kl} V_{jl} + \sum_k U_{ik} h_k$ であり，$\lambda > 0, \eta > 0, \kappa > 0$ は正則化定数である．CST 法は，正則化項 $\|\boldsymbol{U} - \tilde{\boldsymbol{U}}\|_{\mathrm{F}}^2$ と $\|\boldsymbol{V} - \tilde{\boldsymbol{V}}\|_{\mathrm{F}}^2$ の導入により，$\boldsymbol{U}$ と $\boldsymbol{V}$ をそれぞれ転移元の $\tilde{\boldsymbol{U}}$ と $\tilde{\boldsymbol{V}}$ に近づける．

4.2 参加モデルの推定

$\tilde{\boldsymbol{A}} = [\tilde{A}_{ij}]_{i,j}$, $\tilde{\boldsymbol{b}} = (\tilde{b}_1, \ldots, \tilde{b}_I)^\top$, $\tilde{\boldsymbol{c}} = (\tilde{c}_1, \ldots, \tilde{c}_J)^\top$, $\tilde{d}$ をパラメータとする参加モデルの対数尤度は次式で与えられる：

$$\begin{aligned} &L\left(\tilde{\boldsymbol{A}}, \tilde{\boldsymbol{b}}, \tilde{\boldsymbol{c}}, \tilde{d}\right) \\ &= -\sum_{i,j \in \mathcal{O}_P} l\left(\tilde{X}_{ij}\right) - \sum_{i,j \in \mathcal{O}_N} \left(\tilde{X}_{ij} + l\left(-\tilde{X}_{ij}\right) \right). \end{aligned}$$

ここで，$l\left(x\right) = \log\left(1 + \mathrm{e}^x\right)$, $\tilde{X}_{ij} = \tilde{A}_{ij} + \tilde{b}_i + \tilde{c}_j + \tilde{d}$ であり，$\mathcal{O}_P$ は $\Gamma_{ij} = 1$ となる (i, j) ペアの集合，$\mathcal{O}_N$ は $\Gamma_{ij} = 0$ となる (i, j) ペアの集合である．

未完了コンテストの情報不足を補うために，コンテストとワーカの特徴を利用する．特徴に基づく行列分解手法 [Chen 12] により参加者行列を分解する．この分解結果は，勝者モデルの推定に用いることができる．コンテストに対するワーカの嗜好パラメータ $\tilde{\boldsymbol{A}}$，ワーカの活発度パラメータ $\tilde{\boldsymbol{b}}$，コンテストの魅力パラメータ $\tilde{\boldsymbol{c}}$ を，それぞれ $\tilde{\boldsymbol{A}} = \boldsymbol{\Phi}\tilde{\boldsymbol{Q}}\left(\boldsymbol{\Psi}\tilde{\boldsymbol{R}}\right)^\top$, $\tilde{\boldsymbol{b}} = \boldsymbol{\Phi}\tilde{\boldsymbol{s}}$, $\tilde{\boldsymbol{c}} = \boldsymbol{\Psi}\tilde{\boldsymbol{t}}$ と分解することを考える．ここで，$\tilde{\boldsymbol{Q}} \in \mathbb{R}^{N \times K}$, $\tilde{\boldsymbol{R}} \in \mathbb{R}^{M \times K}$, $\tilde{\boldsymbol{s}} \in \mathbb{R}^N$, $\tilde{\boldsymbol{t}} \in \mathbb{R}^M$ であり，K は自然数である．次式の最大化

表 1: 参加者行列が全て観測された場合の，完了コンテストの割合を変化させたときの推薦精度

評価指標	勝利予測	プロジェクト方式データセット					コンペ方式データセット				
		10%	30%	50%	70%	90%	10%	30%	50%	70%	90%
MAP@3	NTSFR	0.920	0.904	0.900	0.897	0.887	**0.357**	0.345	0.347	0.333	0.340
	TSFR-SVD	0.918	0.898	0.903	0.897	0.889	0.337	0.344	**0.351**	**0.345**	0.342
	TSFR-FMF	**0.925**	**0.914**	**0.905**	**0.905**	**0.905**	0.350	**0.346**	0.341	0.341	**0.353**
MAP@5	NTSFR	0.860	0.859	0.867	0.878	0.887	**0.290**	**0.295**	0.301	0.300	0.339
	TSFR-SVD	0.852	0.854	0.864	0.876	0.890	0.273	0.290	**0.304**	0.307	0.342
	TSFR-FMF	**0.868**	**0.872**	**0.871**	**0.884**	**0.902**	**0.290**	0.287	0.295	**0.308**	**0.347**

により $\tilde{\boldsymbol{Q}}$, $\tilde{\boldsymbol{R}}$, $\tilde{\boldsymbol{s}}$, $\tilde{\boldsymbol{t}}$, $\tilde{d}$ を推定する：

$$
\begin{aligned}
&L\left(\tilde{\boldsymbol{Q}}, \tilde{\boldsymbol{R}}, \tilde{\boldsymbol{s}}, \tilde{\boldsymbol{t}}, \tilde{d}\right) \\
&= -\sum_{i,j\in\mathcal{O}_P} l\left(\tilde{X}_{ij}\right) - \sum_{i,j\in\mathcal{O}_N} \left(\tilde{X}_{ij} + l\left(-\tilde{X}_{ij}\right)\right) \\
&\quad - \lambda\left(\|\tilde{\boldsymbol{Q}}\|_F^2 + \|\tilde{\boldsymbol{R}}\|_F^2 + \|\tilde{\boldsymbol{s}}\|^2 + \|\tilde{\boldsymbol{t}}\|^2\right). \qquad (5)
\end{aligned}
$$

ここで，$\tilde{X}_{ij} = \sum_{k,n,m} \Phi_{in}\tilde{Q}_{nk}\Psi_{jm}\tilde{R}_{mk} + \sum_{k,n} \tilde{s}_n\Phi_{in} + \sum_{k,m} \tilde{t}_m\Psi_{jm} + \tilde{d}$ である．

5. コンテスト推薦への応用

5.1 参加者行列が全て観測された場合

提案手法をコンテスト推薦に利用する方法を示す．まず，未完了コンテストも含めすべてのコンテストで参加者行列が全て観測されている場合を考える．すなわち，すべての i, j について $\Gamma_{ij} \in \{0, 1\}$ の場合である．これは，コンテストの参加者全員が参加表明を出しているが実際の成果物はまだ作成していない状況に相当する．この状況で，提案手法により勝利確率を予測することで，ワーカに対して労力を掛けるべきコンテストを推薦できる．この場合は，我々は参加確率を推定する必要はないが，参加者行列からの転移学習を行うために参加モデルを推定する必要がある．推薦手順は以下となる：

Step 1. 式 (5) を最大化することで特徴に基づく行列分解手法を参加者行列 $\boldsymbol{\Gamma}$ に対して適用し，参加モデル $\tilde{\boldsymbol{Q}}, \tilde{\boldsymbol{R}}, \tilde{\boldsymbol{s}}, \tilde{\boldsymbol{t}}, \tilde{d}$ を推定する．

Step 2. $\tilde{\boldsymbol{U}} = \boldsymbol{\Phi}\tilde{\boldsymbol{Q}}$ と $\tilde{\boldsymbol{V}} = \boldsymbol{\Psi}\tilde{\boldsymbol{R}}$ を用いて式 (4) を最大化し，勝者決定モデル $\boldsymbol{h}, \boldsymbol{G}, \boldsymbol{U}, \boldsymbol{V}$ を推定する．

Step 3. $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{G}\boldsymbol{V}^\top$, $\boldsymbol{b} = \boldsymbol{U}\boldsymbol{h}$ を用いて，式 (1) により，ワーカ i の未完了コンテスト j における勝利確率 $\Pr[\Sigma_{ij} = 1]$ を予測する．

Step 4. 各ワーカについて，勝利確率降順で未完了コンテスト $j \in \mathcal{J}_u$ を整列し上位 n 個のコンテストを推薦する．

5.2 参加者行列が部分的に観測された場合

次に，参加者行列が部分的に観測された場合を考える．この場合，我々は完了コンテストについては参加者を全て知っているが，未完了コンテストについては一部の参加者と不参加者しかわからないとする．提案手法により，ワーカに対して参加すべきコンテストを推薦することができる．推薦手順は以下となる：

Step 1. 式 (5) を最大化することで特徴に基づく行列分解手法を参加者行列 $\boldsymbol{\Gamma}$ に対して適用し，参加モデル $\tilde{\boldsymbol{Q}}, \tilde{\boldsymbol{R}}, \tilde{\boldsymbol{s}}, \tilde{\boldsymbol{t}}, \tilde{d}$ を推定する．

Step 2. $\tilde{\boldsymbol{A}} = \boldsymbol{\Phi}\tilde{\boldsymbol{Q}}\left(\boldsymbol{\Psi}\tilde{\boldsymbol{R}}\right)^\top$, $\tilde{\boldsymbol{b}} = \boldsymbol{\Phi}\tilde{\boldsymbol{s}}$, $\tilde{\boldsymbol{c}} = \boldsymbol{\Psi}\tilde{\boldsymbol{t}}$ を用いて，式 (3) により，ワーカ i が未完了コンテスト $j \in \mathcal{J}_u$ に参加する確率 $\Pr[\Gamma_{ij} = 1]$ を予測する．

Step 3. $\tilde{\boldsymbol{U}} = \boldsymbol{\Phi}\tilde{\boldsymbol{Q}}$ と $\tilde{\boldsymbol{V}} = \boldsymbol{\Psi}\tilde{\boldsymbol{R}}$ を用いて式 (4) を最大化し，勝者決定モデル $\boldsymbol{h}, \boldsymbol{G}, \boldsymbol{U}, \boldsymbol{V}$ を推定する．

Step 4. $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{G}\boldsymbol{V}^\top$, $\boldsymbol{b} = \boldsymbol{U}\boldsymbol{h}$ と $\Pr[\Gamma_{ij} = 1]$ を用いて，式 (2) によりワーカ i の未完了コンテスト j における勝利確率 $\Pr[\Sigma_{ij} = 1]$ を予測する．

Step 5. 各ワーカについて，勝利確率降順で未完了コンテスト $j \in \mathcal{J}_u$ を整列し上位 n 個のコンテストを推薦する．

6. 評価実験

コンテストの推薦精度で提案手法を評価する．評価実験においては，提案手法の正則化定数は $\lambda = 0.001, \eta = 100, \kappa = 100$ とし，行列分解のランクは $K = 10$ とする．推薦精度の評価指標として，n 件のコンテストをワーカに推薦した場合の Mean Average Precision (MAP@n) を用いる．また，参加予測精度の評価指標として ROC 曲線下の面積（ROC-AUC) を用いる．

6.1 参加者行列が全て観測された場合

最初に，参加者行列が全て観測されている場合の推薦精度を評価する．開始時間が早い方から x% を完了コンテストとし，残りを未完了コンテストとして実験を行った．

6.1.1 ベースライン手法

提案手法を TSFR-FMF と呼ぶ．提案手法で特徴に基づく行列分解手法を用いた部分を特異値分解に置き換えた手法を TSFR-SVD と呼ぶ．TSFR-SVD と TSFR-FMF の比較により，特徴利用の有効性を評価する．ワーカの能力パラメータ $\boldsymbol{b} \in \mathbb{R}^I$ を直接推定する手法を NTSFR と呼ぶ．NTSFR と TSFR-SVD・TSFR-FMF の比較により，転移学習の有効性を評価する．

6.1.2 結果

結果を表 1 に示す．ほとんどの場合に，転移学習を用いた手法 (TSFR-SVD・TSFR-FMF) が，転移学習を用いない手法 (NTSFR) よりも高い推薦精度を示した．これにより，転移学習が，勝利確率予測とその応用としてのコンテスト推薦で有効であることを確認した．また，プロジェクト方式のデータセットの方が，コンペ方式のデータセットよりも，TSFR-FMF による精度向上が大きい．プロジェクト方式においてはワーカの

表 2: 参加者行列が部分的に観測された場合の，観測要素の割合を変化させたときの推薦精度．5 回の試行における平均と標準偏差を示す．

評価指標	参加予測	勝利予測	プロジェクト方式データセット					コンペ方式データセット				
			10%	30%	50%	70%	90%	10%	30%	50%	70%	90%
MAP@3	AF	NTSFR	0.200 ±0.005	0.400 ±0.013	0.514 ±0.014	0.603 ±0.004	0.758 ±0.005	**0.086** ±0.012	**0.152** ±0.011	**0.200** ±0.006	**0.255** ±0.010	**0.305** ±0.005
	FMF	TSFR -FMF	**0.209** ±0.008	**0.454** ±0.012	**0.604** ±0.015	**0.742** ±0.006	**0.858** ±0.004	0.062 ±0.008	0.099 ±0.012	0.146 ±0.012	0.197 ±0.009	0.253 ±0.004
MAP@5	AF	NTSFR	0.138 ±0.003	0.324 ±0.013	0.452 ±0.016	0.556 ±0.003	0.708 ±0.005	**0.065** ±0.009	**0.126** ±0.008	**0.170** ±0.004	**0.223** ±0.010	**0.269** ±0.006
	FMF	TSFR -FMF	**0.145** ±0.005	**0.363** ±0.011	**0.523** ±0.015	**0.678** ±0.003	**0.816** ±0.005	0.049 ±0.005	0.083 ±0.009	0.128 ±0.009	0.173 ±0.007	0.227 ±0.004

表 3: 参加者行列が部分的に観測された場合の，観測要素の割合を変化させたときの参加予測精度．5 回の試行における平均と標準偏差を示す．

評価指標	参加予測	プロジェクト方式データセット					コンペ方式データセット				
		10%	30%	50%	70%	90%	10%	30%	50%	70%	90%
AUC	AF	0.476 ±0.001	0.541 ±0.003	0.598 ±0.003	0.635 ±0.008	0.657 ±0.011	**0.713** ±0.000	**0.776** ±0.001	**0.804** ±0.001	**0.819** ±0.001	**0.828** ±0.002
	FMF	**0.678** ±0.032	**0.771** ±0.012	**0.765** ±0.021	**0.777** ±0.014	**0.830** ±0.012	0.619 ±0.003	0.640 ±0.008	0.656 ±0.011	0.680 ±0.014	0.673 ±0.013

参加コンテスト数が少ないため，ワーカとコンテストの特徴がより有効だったと考えられる．

6.2 参加者行列が部分的に観測された場合

次に，参加者行列が部分的に観測された場合について評価実験を行った．この実験では，完了コンテストの割合は 50% とした．未完了コンテスト $j \in \mathcal{J}_u$ については，参加者行列 Γ_{ij} からランダムに選んだ y% の要素が観測済みであるとした．各 y について，ランダムサンプリングを 5 回実施しそれぞれについて手法を適用した．

6.2.1 ベースライン手法

提案手法 (TSFR-FMF) と，転移学習を用いない手法 (NTSFR) を比較する．提案手法で用いている，特徴に基づく行列分解による参加予測手法を (FMF) と呼ぶ．参加予測のベースラインとして，平均補間法 (AF) を用いた．この手法では，ワーカ i のコンテストに対する参加意欲 $\tilde{X}_{ij}$ はワーカの活発さだけで決まるものとし，活発さを参加率で推定している．すなわち，$\tilde{X}_{ij} = (J_i + 1) / (J + 2)$ として推定する．ここで，J_i はワーカ i が参加したコンテストの数である．

6.2.2 結果

表 2 に結果を示す．プロジェクト方式のデータセットにおいては，参加者行列が部分的にしか観測できていなくても，提案手法が転移学習を用いない手法よりも高い推薦精度を示した．しかし，コンペ方式では，単純なベースライン手法の方が高い精度を示した．この理由として表 3 に示す，コンペ方式における参加予測の精度の低さが考えられる．参加者行列が部分的にしか観測できていない場合は，参加予測の精度が勝利予測の精度に直接影響を与えてしまう．

プロジェクト方式では，特徴に基づく行列分解手法 (FMF) がベースライン手法よりも高い参加予測精度を示したが，コンペ方式では単純な手法の方が優れていた．このことは，ワーカとコンテストの特徴が，プロジェクト方式では有効に機能したがコンペ方式ではそうではなかったことを示している．プロジェクト方式データセットは多様な仕事を含んでおり，それぞれ異なるスキルをワーカに要求する．そのため，ワーカは自分自身でスキルに適した仕事を選んでおり，また特徴がそのスキルを適切に表現していると考えられる．そのため，プロジェクト方式では特徴を用いた参加予測手法が有効だった．一方でコンペ方式データセットはほとんどがデザインに関する仕事である．プロジェクト方式と比べて仕事の多様性が低いため，特徴が有効ではなかったと考えられる．

7. むすび

コンテスト型クラウドソーシングにおける勝利予測手法を提案し，そのコンテスト推薦への応用を示した．勝者決定モデルと参加モデルの二つの確率モデルを提案し，データの疎性への解決策として転移学習と特徴に基づく行列分解によるモデル推定手法を提案した．評価実験により，参加者情報からの転移学習が勝利予測に有効であることと，特に参加者情報が疎である場合にワーカとコンテストの特徴が有効であることを示した．

参考文献

[Chen 12] Chen, T., Zhang, W., Lu, Q., Chen, K., Zheng, Z. and Yu, Y. SVDFeature: A Toolkit for Feature-based Collaborative Filtering, Journal of Machine Learning Research, pp. 3619–3622 (2012)

[Pan 10] Pan, W., Xiang, E. W., Liu, N. N. and Yang, Q. Transfer Learning in Collaborative Filtering for Sparsity Reduction, In Proceedings of the 24th Conference on Artificial Intelligence (AAAI) (2010)