

jh221002

## 医療・介護領域の人材マッチングに最適化された

### 大規模グラフニューラルネットワーク

鈴木 豊太郎 (東京大学)

#### 概要

本研究では、人材マッチングシステム用の強化学習を用いた推薦システムを探求した。ジョブ推薦システムは求人推薦のため、ものやサービスを推薦するシステムと比べ大きく異なる性質が必要とされる。一般の推薦システムの場合、ユーザがどのアイテムに興味があるかを予測することで、ユーザのクリック率や購入数を最大化する。一方ジョブ推薦システムでは、各求人採用枠上限などの制約の中でユーザと求人間のマッチング数を最大化することが求められている。従来のジョブ推薦システムでは、ユーザの応募先の求人を分散させたり、数理最適化アルゴリズムを用いたりする手法が提案されているが、特定の仮定や条件を必要としており、マッチング数最大化を直接的に行っているわけではない。本研究では、マッチング数を報酬に組み込んだ強化学習を利用することで、直接的にマッチング数を最大化する手法を提案する。人工データを用いたシミュレーション実験で、既存手法と比べ少なくともマッチング数を 131.2%まで向上させ、性能の優位性を確認した。

#### 1. 共同研究に関する情報

- (1) 共同利用・共同研究を実施している拠点名 (該当するものを残す)

mdx

- (2) 課題分野 (該当するものを残す)

データ科学・データ利活用課題分野

- (3) 共同研究分野 (HPCI 資源を利用している研究課題のみ、該当するものを残す)

- (4) 参加研究者の役割分担

研究者の種類	氏名	所属機関・部署名	役割・担当分野等
代表者	鈴木豊太郎	東京大学・大学院情報理工学系研究科	モデル設計

共同研究者	脇聡志 金刺宏樹 華井雅俊	東京大学・情報基盤センター	モデル開発・実験、評価、論文執筆
共同研究者	小林秀 長谷川正彦 小貝洸希	エス・エム・エス社	医療・介護データに関するデータ処理、ドメイン知識提供

#### 2. 研究の目的と意義

- (1) 研究の目的

本研究では、人材マッチングシステム用の強化学習を用いた推薦システムを探求する。推薦システムは、現在様々な応用先があり、活発に研究されている分野の一つである。ジョブ推薦システムは求人/仕事を推薦するため、ものやサービスを推薦するシステムと比べ大きく異なる性質を必要とされる。One worker, one jobはその性質の一つであり、我々は強化学習を用いることでこの性質を

満たす推薦システムを考案する。

推薦システムとは、ユーザ、情報、およびその相互作用に関する情報を利用して、個々のユーザの関心や興味に合わせたアイテムを提示するシステムである。推薦システムは、E コマース、ストリーミングサービス、観光サービスなど、実世界の様々な分野で広く活用されている [Ko 22]。ジョブ推薦システムは、求職者に仕事を推薦することに焦点を当てた推薦システムの一つであり、重要な応用分野の一つである。多くの人々は人生の大半の時間を仕事に割いていることから、仕事は人々の生活に長期的かつ多大な影響を与えることは明らかである。また、人が有する知識やスキルは企業の業績に深く関わっており、優れた人材を獲得することは企業にとって必要不可欠である [Crook 11]。このように、ジョブ推薦システムは他の推薦システムと比較して、ユーザの人生や企業の競争力に影響を与えるという点で重要な研究分野である。

ジョブ推薦システムでは、求人の募集人数の制限を考慮する必要がある。一般的な推薦システムの場合、推薦対象となるアイテムの在庫が大量にあり、一部のアイテムを頻繁に推薦しても問題がない。しかし、求人は通常、数人の求職者しか募集しないため、一部の求人を多数のユーザに推薦しても、最終的にはごくわずかしかマッチングしない。そのため、ジョブ推薦システムは、ユーザの好みに基づく推薦に加えて、全体のマッチング数を最大化する推薦を行う必要がある。

この問題に対し、Borisyuk ら [Borisyuk 17] は求人間で応募数が均等になるよう、ユーザが求人に応募する確率を予測するモデルの出力を決められたアルゴリズムで調整する定式化された関数によってユーザの応募数を再分配する手法を提案した。Su ら [Su 21] はマッチングプロセスをモデル化し、期待マッチング数を数式として表現することで、マッチング数を最大化する推薦システムを数理最適化により求めた。

しかしながら、推薦してから採用に至るまでのプロセスは複雑で長期にわたることが多いため、既存の手法では間接的な最適化やマッチングプロセ

スに特定の仮定を置いた場合での最適化しか実現されていない。実際、Borisyuk らの手法では求職者の応募を分散させることで遠回しに全体のマッチング数を増加させており、Su らの手法ではユーザの応募プロセスや雇用主の採用プロセスを Position-based model (PBM) [Richardson 07, Chuklin 15, Joachims 17] で数理化した場合に限定して最適化を行っている。PBM ではユーザが求人に応募する確率を、ユーザに求人を表示する確率と表示された場合にユーザが求人に応募する確率の積で表現する。ここでユーザが求人に応募する確率として様々な関数が仮定されている。

本論文では、ジョブマッチングサービスにおいて、ユーザが訪れてから採用に至るまでの一連のプロセスを再現することでより現実に即したシミュレーション環境を構築し、強化学習を用いて end-to-end にマッチング数最大化を最適化する手法を提案する。強化学習を用いることで、シミュレーション条件が変更されても同様の枠組みで最適化でき、また長期的なマッチング数最大化を達成することができる。ユーザの好みに基づく推薦システムを強化学習で求めた研究 [Chen 19, Chen 22] はあるが、強化学習を用いてマッチング数を最大化する試みは初めてである。人工データによる実験の結果、提案手法は応募数最大化の手法や応募再分配の手法に比べてマッチング数が向上したことが確認された。

## (2) 研究の意義

ジョブ推薦システムで One worker, one job を考慮することの重要性は前述したとおりである。また、特に本共同研究で扱うジョブ推薦システムは医療・介護領域のシステムであり、日本においてはより重要な研究である。日本は高齢化に伴い、医療や介護の需要が増大する一方で、生産年齢人口の減少により、これらのサービスを支える従事者の不足が深刻な課題となっている。国の推計によると、医療・介護従事者の需給ギャップは 2025 年に看護師で 6~27 万人、介護職で 22 万人にまで拡大する見込みである。高齢者や患者のケアを担う働き手の不足により、質の高い医療・介護サー

ビスの提供が難しくなると予想される。現在、政府はこの需給ギャップ解消に向けて、賃金の改善などを行い、是正に向けて動いている。一方、医療・介護従事者の就職活動・転職活動は、他の業種と比較して、就職から転職までの期間が2、3年と短い。そのため結婚や子育てなどのライフステージ、居住地、正規職員か非正規職員、給与等の条件に合う職場と的確にマッチングさせ、面接など求人・求職にかかわるコストを下げつつ離職率を低くすることが喫緊の課題となっている。本研究で提案した手法により、高齢化社会に向けた医療・介護領域の人材不足と需給のギャップを緩和することを目指す。

### 3. 当拠点の公募型研究として実施した意義

本研究の実施は、我々の拠点が掲げる課題解決型の学際的な研究へのアプローチに見事に符合している。我々は常に社会的課題を解決するための技術開発とその適用に焦点を当てており、今回の公募型研究はその目的を具現化したものである。

特に、医療・介護領域での人材マッチングの改善というテーマは、我々社会が直面している深刻な問題の一つに対応している。日本の高齢化が進む中、医療・介護領域の人材不足という問題はますます深刻化しており、この問題解決に向けた努力として、強化学習を用いた新たな人材マッチングシステムの開発は、非常に大きな意義を持つ。

さらに、強化学習という先進的な技術を活用して具体的な社会問題に取り組んだ点も、この公募型研究の特筆すべき意義である。我々の拠点では、最新の技術を社会問題解決のために用いることを常に重視しており、その精神が本研究で見事に具現化されたと言える。加えて本研究では実証的な研究手法を用い、人工データを活用した具体的な実験によって、提案手法の有効性を確認した。

最後に、本研究が産学連携の形で進められた点も大きな意義がある。我々の拠点では、産学連携による実践的な研究を推進しており、今回の研究がその一例となった。さらに本研究により得られた成果は、実際の業界への応用が見込まれるため、

産業界との連携を深める上でも大きな役割を果たすと考える。

### 4. 前年度までに得られた研究成果の概要

前年度は実施していないので該当なし

### 5. 今年度の研究成果の詳細

#### 提案手法

#### 強化学習の問題設定と学習アルゴリズム

強化学習はエージェントが環境内で試行錯誤を繰り返して最適な行動を獲得するための学習制御の枠組みである。環境は以下の4組 $(S, A, T, R)$ で定義されるマルコフ決定過程 (Markov Decision Process; MDP) [Bellman 57] を用いて定式化される。

- $S$ : 環境の取りうる有限な状態集合。
- $A$ : エージェントが取りうる有限な行動集合。
- $T: S \times A \times S \rightarrow [0, 1]$ : ある状態からある行動をした後、ある状態に遷移する確率。
- $R: S \times A \times S \rightarrow R$ : ある状態遷移をしたときの報酬を定める関数。

環境の状態  $s \in S$  に対して行動  $a \in A$  を選択する関数  $\pi(s, a)$  を方策と呼び、強化学習の目的は式 (1) で定義される割引累積報酬の期待値を最大化する最適方策  $\pi^*$  を求めることである。

$$V = \sum_{t=0}^{\infty} \gamma^t r_t \quad (1)$$

ここで、 $r_t$  は  $t$  時点での報酬、 $\gamma \in [0, 1]$  は割引率である。方策がパラメータ  $\theta$  の関数であるとき、期待割引累積報酬  $E[V]$  の  $\theta$  に対する勾配を用いて式 (2) のように更新することで局所最適解が求まる。

$$\theta_{i+1} = \theta_i + \alpha \nabla_{\theta_i} \mathbb{E} \left[ \sum_{t=0}^{\infty} \gamma^t r_t \mid \theta_i \right] \quad (2)$$

ただし、 $\alpha$  は学習率である。REINFORCE [Williams 92] では式 (3) で勾配を近似し、最適方策を求める。

$$\nabla_{\theta} \mathbb{E} \left[ \sum_{t=0}^{\infty} \gamma^t r_t \mid \theta \right] \simeq \sum_{t=0}^{\infty} \nabla_{\theta} \ln \pi_{\theta}(s_t, a_t) \left( \sum_{k=0}^{\infty} \gamma^k r_{t+k} \right) \quad (3)$$

### 強化学習を用いたジョブ推薦システム

求人におけるマッチングプロセスは、以下の4つのステップに分割できる。

1. ユーザが求人サイトを訪問する。
2. 推薦システムがユーザに求人を推薦する。
3. ユーザが推薦された求人に対して、応募や無視などのアクションを取る。
4. ユーザが応募した場合、求人を掲載した雇用主がユーザを採用または保留するなどのアクションを取る。

これら4つのステップをMDPにおける1時刻と見なす。t時点で推薦対象となるユーザの集合を  $U_t$ 、求人の集合を  $J_t$ 、t時点までに得られているユーザ・求人間の相互作用に関する情報を  $h_t$  としたとき、MDPの各要素は以下ようになる。ただし実験では相互作用の情報  $h_t$  として t 時点で各求人に応募された回数と各求人の求職者の保留人数を用いた。

- $S$ : t 時点でサイトを訪問したユーザ  $u \in U_t$ 、求人の集合、過去の相互作用の情報の3組  $s_t = (u, J_t, h_t)$ 。
- $A$ : ユーザに対して推薦する求人、あるいは求人の集合。実験では推薦する求人の数は1つとしたため、 $a_t = j$  である。
- $T$ : 推薦された求人に対してユーザが取るアクションの確率、応募された雇用主が取るアクションの確率、次の時刻にサイトを訪問するユーザの確率を定める関数。オンライン実験では実際のユーザや雇用主自身の振る舞いが関数  $T$  となり、定義する必要はないが、オフライン実験やシミュレーションでは適用な関数を定義する必要がある。シミュレーションの実験で用いた関数は第3.1節に記述した。
- $R$ : t-1 時点から t 時点で新たに採用が決まったユーザ・求人のマッチング数  $Match_t$ 。このように報酬を定義するこ

とで直接マッチング数に対して最適化できるが、ジョブマッチングでは推薦から採用に至る確率が非常に小さく、報酬が非常に疎になり学習が困難になるため、マッチングに至らない場合であっても、ユーザが求人に応募すれば小さな報酬を与える。具体的には、ユーザ  $u$  が推薦した求人のうち応募した数を  $Apply_t$  とおいて  $r_t = Match_t + 0.01 \cdot Apply_t$  とする。

方策は図1のように設計した。まず、パラメータ  $\theta$  を持つ関数  $\phi_U \theta$  を用いてユーザの埋め込み  $e_u = \phi_U \theta(u)$  を計算する。次に、推薦対象となる各求人  $j \in J_t$  に対し、同様にパラメータ  $\theta$  を持つ関数  $\phi_J \theta$  を用いて、求人の情報  $j$  と相互作用の情報  $h_{t,j}$  から求人の埋め込み  $e_j = \phi_J \theta(j, h_{t,j})$  を得る。続いてユーザの埋め込みと求人の埋め込みの内積を求め、これを推薦の優先度を表すマッチングスコア  $su_{j} = e_u \cdot e_j$  とする。最後にマッチングスコアの最も高い求人  $j = \operatorname{argmax}_{j \in J_t} (e_u \cdot e_j)$  を行動として出力する。

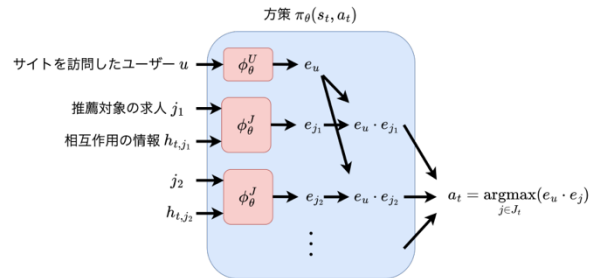


図1: 方策のアーキテクチャ。ユーザーと求人・相互作用の情報に対する埋め込み  $e_u, e_j$  の内積が大きい求人を推薦する。

### 実験

推薦システムの評価方法は、オンライン評価、オフライン評価、人工データを用いたシミュレーションでの評価に分けられる。オンライン評価は、実世界のユーザや雇用主がシステムに対し直接フィードバックを与えるため、最も正確で現実的な評価ができる。一方で実験をスケールさせるのが困難であり、特にジョブ推薦システムでは採用まで至る確率が低く、長期間に渡る実験が必要となる。オンライン評価はユーザの個人情報、求人の情報、これまで得られた行動履歴で評価するため、ユーザ

や雇用主の特性をある程度反映できる。しかし、収集されているデータが現状の推薦システムに偏っており、既存の推薦システムで推薦されていないデータに対する評価が困難である。人工データを用いたシミュレーションでの評価では、ユーザの情報、求人の情報、ユーザや雇用主の行動をアルゴリズムで模擬的に計算することで、任意の推薦に対する応答を取得でき、実運用から独立して評価が可能一方、実データがもつ特性をすべて組み入れることが困難である。

このため、オンライン評価やオフライン評価をする前段階として実験コストが低く、データにバイアスのないシミュレーションでの評価がしばしば行われており [Su 21, Bills 21], 我々もこれに倣い人工データを用いたシミュレーションでの評価を行う。

### 遷移関数の設定

今回はシミュレーションでの評価であるため、遷移関数に適切な関数を定める必要がある。実験では、ユーザが完全にランダムにサイトを訪問し、推薦された求人との関連度が 0.2 を超えた場合応募するとする。ここで、ユーザ  $u$  と求人  $j$  の関連度  $p(u, j)$  は式 (4) で定義する。

$$p(u, j) = \frac{f_u \cdot f_j - \max_{(u', j') \in U_0 \times J_0} (f_{u'} \cdot f_{j'})}{\max_{(u', j') \in U_0 \times J_0} (f_{u'} \cdot f_{j'}) - \min_{(u', j') \in U_0 \times J_0} (f_{u'} \cdot f_{j'})} \quad (4)$$

ただし、 $f_u$  はユーザ  $u$  の特徴量、 $f_j$  は求人  $j$  の特徴量である。また、雇用主は 5 時刻ごとに応募されているユーザの内、最も関連度の高いユーザのユーザが 0.5 以上の場合、そのユーザを採用するとする。

### 人工データの生成

ユーザ数と求人数はいずれも 100 とするまたユーザの情報は 5 次元のベクトルとし一様分布  $U(0, 1)$  から生成する。求人の情報も 5 次元のベクトルとするが、人気の求人を生成するため、 $k$  番目に生成する求人の情報は一様分布  $U(0, \exp(\frac{\max(0, 21-k)}{40}))$  から生成する。また、汎化性能を見るため、訓練用の環境とテスト用の環境で、含まれるユーザと求人は完全に独立に生成した。図 2 に実際に生成

したユーザ・求人の関連度行列を示した。区間の広い一様分布から生成した求人ほど、多くのユーザとの関連度が大きくなる傾向となる。

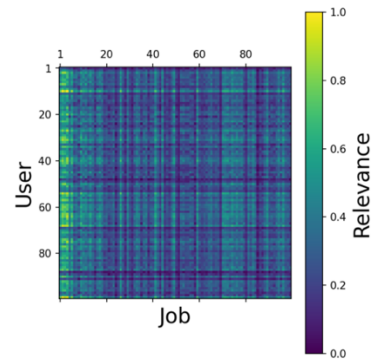


図 2: ユーザー・求人の関連度の行列  $(p(u, j))_{u,j}$ .

### 方策のモデル

方策内部の関数  $\phi_\theta^U$  と  $\phi_\theta^J$  は図 3 で表されるニューラルネットワークでモデル化した。ニューラルネットワークの入力は  $\phi_\theta^U$  では  $u, \phi_\theta^J$  では  $j \oplus h_{t,j}$  とする。ニューラルネットワークは 8 個のユニットを持つ 4 層から構成され、中間層でスキップ接続させる。また活性化関数は ReLU を用い、ドロップアウト率は 0.1 とした。

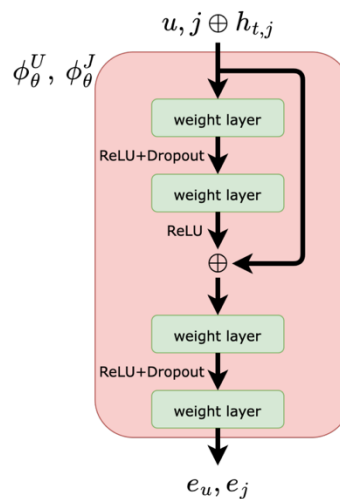


図 3:  $\phi_\theta^U$  および  $\phi_\theta^J$  で使用するモデル。

### 比較手法

提案手法の有効性を評価するために、応募数最大化の手法と 応募再分配の手法と比較する。応募数最大化の手法は、報酬が応募数の報酬のみで学習し、方策のアーキテクチャや学習アルゴリズムは同じものを用いる。また応募再分配の手法は応募

数 最大化の手法に加え， 方策内部にマッチングスコアを変更する Borisyuk らの手法 [Borisyuk 17] と同様の機構を加え， 特定の 推薦を抑制・促進させる． 具体的には， 求人  $j$  への応募数を  $n_j$  と置いたとき， 応募数が平均値  $|J_t| \sum_{j \in J_t} n_j$  より多い場合は

$$s'(u, j) = s(u, j) \cdot \exp\left(-\frac{n_j}{softness}\right) \quad (5)$$

によりマッチングスコアを小さくし， 応募数が平均値より少なくかつマッチングスコアが  $h$  より大きい場合は

$$s'(u, j) = \min\{s(u, j) \cdot boostFactor, 1.0\} \quad (6)$$

によってマッチングスコアを大きくする． ここで，  $softness$ ,  $boostFactor$ ,  $h$  はハイパーパラメータである． 実験では  $softness \in \{1, 5, 10\}$ ,  $boostFactor \in \{1.1, 2, 5\}$ ,  $h \in \{0, 0.01, 0.1\}$  の中でグリッドサーチし最適なパラメータを求めた．

## 結果

MDP の終了時刻を 50, 割引率を  $\gamma = 0.99$  とし, 300 エピソード学習させ, シード値を 1 から 5 まで変更して実験した． テスト環境での平均マッチング数の遷移を図 4 に示した． 提案手法はマッチング数が 9.32 組, 応募数最大化の手法は 5.92 組, 応募再分配の手法は ( $softness$ ,  $boostFactor$ ,  $h$ ) = (1.0, 1.1, 0.01) のとき 7.10 組で最大となり, ( $softness$ ,  $boostFactor$ ,  $h$ ) = (1.0, 5.0, 0.01) のとき 2.72 組で最小となった． また応募再分配の手法の全 27 通りのハイパーパラメータのうち, 応募数最大化の手法より優れていたパラメータの組み合わせは 9 通りであった． 以上から, 提案手法は応募数最大化の手法に比べてマッチング数が 157.4%に増加するのに対し, 応募再分配の手法はマッチング数が最大でも 120.0%しか増加せず, 提案手法は既存手法と比べ最低でも 131.2%以上, マッチング数を向上させることが確認された． さらに応募再分配の手法は応募数最大化の手法と比べて性能が改善する割合が 33.3%であり, 最悪ケースでは, マッチング数が 46.0%まで

減少していることから, 応募再分配の手法を用いる場合, ハイパーパラメータを慎重に調整しなければならない． その点, end-to-end に最適化できる提案手法は安定的に性能を向上させることができるため, 実用面も有効である． 実験では, mdx [mdx 2022] のインスタンスを 1GPU (NVIDIA A100) 分使用した．

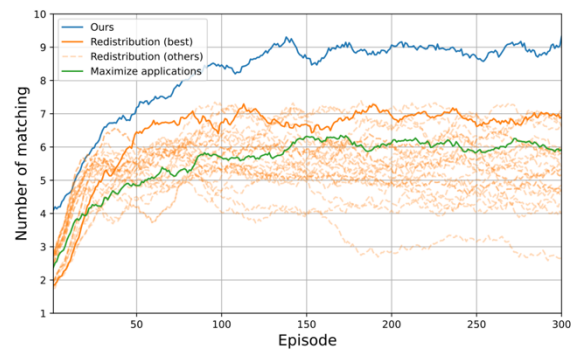


図 4: テスト環境での 5 シード分の平均マッチング数の遷移, 15 エピソードの移動平均を表している． 提案手法が青の実線, 応募再分配の手法で最も良かったパラメータでのモデルが橙色の実線, その他のパラメータのモデルが橙色の点線, 応募数最大化の手法が緑色の実線である．

## 6. 進捗状況の自己評価と今後の展望

我々の目的は, 高齢化社会における医療・介護分野の人材不足と供給と需要のギャップを緩和することでした． このために, 我々はグラフに基づく評価ではなく, 強化学習モデルに着目した． 人工データを用いたシミュレーションでの評価では, 提案手法と比較して 57.4%高いパフォーマンスを示した． 次のステップとして, 実際のデータを用いたモデルの有効性を評価する予定である．

本研究では, ジョブ推薦システムにおいて, マッチング数最大化を実現するための強化学習を利用した最適化手法を提案した． 従来の研究では間接的なマッチング数の最大化, 採用までのプロセスに仮定を置いた場合での最適化しか考慮されていなかったが, 提案手法では強化学習を用いることで環境に対し仮定を置くことなく, end-to-end に最適化することができる． 人工データでのシミュレーションの実験結果, 応募再分配の手法と比べマッチング数を安定して最大化できることが確認できた． 今回はシミュレーションにより評価実験を行ったが, 実際のデータでの有用性を示すために

はオフライン評価やオンライン評価での検証が必要となる。また、ユーザと求人間の相互作用の情報を追加することでより性能の高いモデルを開発することを目指す。

## 7. 研究業績

### (1) 学術論文 (査読あり)

脇 聡志, 鈴木 豊太郎, 金刺 宏樹, 華井 雅俊, 小林 秀, 強化学習によるマッチング数を最大化するジョブ推薦システム, 2023 年度 人工知能学会 全国大会 (第 37 回)

### (2) 国際会議プロシーディングス (査読あり)

Satoshi Waki, Toyotaro Suzumura, Hiroki Kanezashi, Shu Kobayashi “Matching Maximization in Job Recommendation System with Reinforcement Learning”, Thirty-seventh Conference on Neural Information Processing Systems (NeurIPS), 2023 (submitted).

### (3) 国際会議発表 (査読なし)

### (4) 国内会議発表 (査読なし)

### (5) 公開したライブラリなど

### (6) その他 (特許, プレスリリース, 著書等)

## 1. 参考文献

[Ko 22] Ko, H. Lee, S., Park, Y., and Choi, A.: A Survey of Recommendation Systems: Recommendation Models, Techniques, and Application Fields. Electronics, Vol. 11, No. 141 (2022).

[Crook 11] Crook, T.R., Todd, S.Y., Combs, J.G., Woehr, D.J., and Ketchen, D.J.: Does human capital matter? A meta-analysis of the relationship between human capital and firm performance. The Journal of applied psychology, Vol. 96 No. 3, pp. 443-456 (2011).

[Mashayekhi 22] Mashayekhi, Y., Li, N., Kang, B., Lijffijt, J., and Bie, T.D.: A challenge-based survey of recruitment

recommendation systems. arXiv preprint arXiv:2209.05112 (2022).

[Borisjuk 17] Borisjuk, F., Zhang, L., and Kenthapadi, K.: LiJAR: A System for Job Application Redistribution towards Efficient Career Marketplace. Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2017).

[Su 21] Su, Y., Bayoumi, M., and Joachims, T.: Optimizing Rankings for Recommendation in Matching Markets. Proceedings of the ACM Web Conference 2022 (2021).

[Richardson 07] Richardson, M., Dominowska, E., and Ragno, R.: Predicting clicks: estimating the clickthrough rate for new ads. In Proceedings of the 16th international conference on World Wide Web, pp. 521- 530 (2007).

[Chuklin 15] Chuklin, A., Markov, I., and Rijke, M. D.: Click models for web search. synthesis lectures on information concepts, retrieval, and services, Vol. 7, No. 3, pp. 1-115 (2015).

[Joachims 17] Joachims, T., Swaminathan, A., and Schnabel, T.: Unbiased learning-to-rank with biased feedback. In Proceedings of the tenth ACM international conference on web search and data mining, pp. 781- 789 (2017).

[Chen 19] Chen, M., Beutel, A., Covington, P., Jain, S., Belletti, F., and Chi, E. H.: Top-k off-policy correction for a REINFORCE recommender system. In Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, pp. 456-464 (2019).

[Chen 22] Chen, M., Xu, C., Gatto, V., Jain, D., Kumar, A., and Chi, E.: Off-Policy Actor-critic for Recommender Systems. In Proceedings of the 16th ACM Conference on Recommender Systems, pp. 338-349 (2022).

[Bellman 57] Bellman, R: A Markovian Decision Process. Indiana University Mathematics Journal, Vol.6, pp. 679–684 (1957).

[Williams 92] Williams, R. J.: Simple statistical gradientfollowing algorithms for connectionist reinforcement learning. Reinforcement learning, pp. 5–32 (1992).

[Bills 21] Bills, J.T., and Ng, Y.: Looking for Jobs? Matching Adults with Autism with Potential Employers for Job Opportunities. Proceedings of the 25th International Database Engineering and Applications Symposium (2021).

[mdx 2022] Toyotaro Suzumura, et.al., “mdx: A Cloud Platform for Supporting Data Science and Cross-Disciplinary Research Collaborations,” Intl Conf on Cloud and Big Data Computing, Falerna, Italy, 2022, pp. 1–7, doi: 10.1109/DASC/PiCom/CBDCCom/Cy55231.2022.9927975.