

GRPO を用いたスタッフスケジューリングにおける代替出勤依頼方策の獲得手法の提案 Proposal of a Method for Obtaining Alternative Attendance Request Policy for Staff Scheduling Using GRPO

横山想一郎¹⁾ 山下 倫央¹⁾ 川村 秀憲¹⁾
Soichiro Yokoyama Tomohisa Yamashita Hidenori Kawamura

1 はじめに

スタッフスケジューリングは、勤務シフトにおいて、業務量や従業員の希望を考慮しながら、人員を適切に割り当てる問題であり、オペレーションズリサーチの分野で、Nurse Scheduling Problem (NSP) などの形で広く研究されてきた。従業員の欠勤といった勤務に関する不確実性への対処法として、欠勤シナリオを想定し、人員のリスクを許容範囲内に抑える勤務シフトを作成する方法が提案されてきた。一方で、実務では欠勤の発生時に従業員に追加出勤を依頼することが一般的であるが、その方法は十分に検討されていない。

本研究では、出勤の有無や業務量が不確実な環境下で、人員を適切に確保する出勤依頼の方策を求める。依頼される出勤は本来の勤務シフトと異なるため、必ず受諾されるわけではない。この点を踏まえ、必要な人員を確保しつつ、依頼回数を最小化することで従業員と管理者双方の負担を軽減し、依頼を特定の従業員に偏らせず公平性を考慮した方策を検討する。

スタッフスケジューリングにおける出勤依頼の方策を評価するため、勤務シフト決定後に発生する業務量・勤務量の変動に対し、従業員に出勤を依頼する過程をモデル化する。従業員の出勤依頼の応答を確率分布により表現し、従業員の応答傾向を反映した多様な環境下で機能する方策を求める。方策の探索には深層強化学習を用いる。依頼候補の従業員の特徴量を入力とし、依頼の優先度を出力する深層モデルを構築し、勤務可能日程の希少性のような従業員間の関係性を注意機構により捉える。

本稿ではその初期的な検討として、出勤依頼の応答に確率分布を仮定し、数値計算実験を行い、既存のルールベースによる方策と結果を比較した。

本稿は次の通り構成される。第 2 章で本研究が取り扱う不確実性を持つスタッフスケジューリングの関連研究と強化学習によるスケジューリングへのアプローチを示す。第 3 章では、スタッフスケジューリングの定式化を行い、第 4 章で強化学習による方策の獲得方法を説明する。第 5 章で数値計算実験を行い、第 6 章に結論をまとめる。

2 関連研究

勤務シフト作成の問題は、オペレーションズリサーチにおいて広く研究されてきた。本節では、ナーススケジューリング問題 (Nurse Scheduling Problem, NSP) の定式化と、本稿で扱う不確実性の考慮、深層強化学習によるアプローチの関連研究を示す。

2.1 ナーススケジューリング問題

日付ごとの業務量と従業員の勤務日程に関する希望を考慮した勤務シフトの作成は、組合せ最適化の一種である NSP として定式化され、NP 困難な問題であることから、ヒューリスティクスによる解法が研究されてきた。

池上ら [1] は日本の病院における実業務をもとに、看護師の技術レベルや勤務時間帯を踏まえた定式化を行い、タブー探索による解法を示した。2010 年に開催された NSP のベンチマーク問題に関するコンペティション [2] では、多くのアプローチが比較され、長期的な人員配置の決定と勤務時間帯などの詳細の決定の 2 段階を分け、小問題に分割し整数計画法を繰り返し適用する手法 [3] が有望とされている。

2.2 業務量や欠勤についての不確実性の考慮

実業務においては、必要な業務量の変動や従業員の欠勤による労働力の欠如などにより、あらかじめ作成された勤務シフトに従った業務の遂行が難しい場合が考えられる。こうした不確実性を考慮したスケジューリングとして、NSP を拡張した確率計画法の適用やリアクティブスケジューリングが提案されている。

確率計画法では、一般に 2 段階の決定が行われる。第 1 段階では通常の勤務シフトと同様に人員配置を決定し、第 2 段階で、顕在化した不確実性への対応コストを計算する。発生するシナリオを多数生成し、第 1 段階の決定を全シナリオで共通としつつ、第 2 段階で発生するコストの平均値などを最小化する。医療分野では、El-Rifai ら [4] が医師と看護師の最適配置を提案した。小売業では、Parisio ら [5] が隠れマルコフモデルを用いて顧客需要を予測し、確率計画法で最適な人員配置を求めた。コールセンターでは、サービスレベル契約の遵守を目的として、Whitt [6] が受電と従業員の欠勤の不確実性を考慮した人員配置法を提案した。

勤務シフト作成後に人員不足が発生した場合、最小限の変更で対応するリアクティブスケジューリングが用いられる。Pato ら [7] は、充足の度合いと変更の少なさを考慮した二目的最適化を提案し、遺伝的アルゴリズムによる解法を示した。Otero-Caicedo ら [8] は、確率計画法を用いた勤務シフト作成とともに、代替出勤の依頼先となる従業員を決定するルールを比較し有効性を検証した。

著者らの研究グループでは、勤務シフト変更を依頼する際の負担に着目し、不確実性が高く頻繁に代替出勤の依頼が行われるコールセンターを想定した代替出勤依頼モデルにおいて、依頼順が与える影響を明らかとした [9, 10]。本稿では、深層強化学習により、依頼順を学習する手法を提案する。

2.3 深層強化学習によるスケジューリング

近年では、スケジューリングを強化学習の問題と捉え、多層ニューラルネットワークを用いた方策の学習によって解決を試みる深層強化学習が、多様な制約に対応可能な柔軟性から注目されている。

組合せ最適化問題に対しては、在庫配送問題を対象として、ニューラルネットワークにより訪問先の都市を選択する機構を繰り返し適用することで解を作成し、自

1) 北海道大学大学院情報科学研究院

表 1: 代替出勤依頼問題の定式化に用いられる記号

記号	定義
n	対象とする勤務日の総数
m	対象とする従業員の総数
$D = \{d_1, d_2, \dots, d_n\}$	出勤依頼の対象となる日付の集合
$W = \{w_1, w_2, \dots, w_m\}$	出勤依頼の対象となる従業員の集合
$E = (e_{i,j}) \in \{0, 1\}^{n \times m}$	日付 d_i における従業員 w_j への代替出勤の依頼可能性
$A = (a_{i,j}) \in \{0, 1\}^{n \times m}$	日付 d_i における従業員 w_j の代替出勤依頼への応答
$Q = (q_i) \in \mathbb{Z}_{\geq 0}^n$	各日付において代替出勤が必要とされる従業員数
$U = (u_j) \in \mathbb{Z}_{\geq 0}^m$	各従業員の対象期間における最大の代替出勤数
$R = (r_{i,j}) \in \{0, 1\}^{n \times m}$	日付 d_i における従業員 w_j への代替出勤依頼の有無
$S = (s_{i,j}) \in \{0, 1\}^{n \times m}$	日付 d_i における従業員 w_j の代替出勤の有無
$\pi(d_i, q_i, E, R, S) = w_j$	代替出勤依頼の対象とする従業員を選択する方策関数

動生成された問題に対して目的関数を最小化するようにニューラルネットワークを強化学習で訓練し近似解を得る手法 [11] が提案されている。ジョブショップ問題に対しては、Feng ら [12] が各時刻におけるジョブの割り当てを同様に強化学習で訓練する手法を提案し、ルールベースの手法に対する優位性を示している。

NSP に対しては、永吉ら [13] が Q 学習に基づく勤務シフトの交換による制約違反の抑制を提案し、Muklasan ら [14] は強化学習によりヒューリスティクスを選択する手法を提案しているが、いずれも本稿で扱う不確実性を考慮していない。

3 代替出勤依頼問題の定式化

表 1 に示す記号を用いて代替出勤依頼問題を定式化する。 m 名の従業員が存在する職場の n 日間のスケジュールを考え、従業員集合を $W = \{w_j (0 < j \leq m)\}$ 、日付集合を $D = \{d_i (0 < i \leq n)\}$ とする。

E, A は、代替出勤の依頼先候補とその回答内容を表す $n \times m$ 行列であり、各要素は 0 または 1 をとる。 E は本稿において依頼可能性と呼び、日付 d_i において従業員 w_j への代替出勤の依頼が可能である場合に i 行 j 列の要素 $e_{i,j}$ が 1 をとり、それ以外の場合に 0 をとる。実問題において E の要素が 0 をとる場合として、職場の業務シフトが既に従業員に割り当てられている場合や、従業員が指定感染症への罹患を職場に届出済みである場合など、代替出勤が不可能であることが明白である場合が考えられる。同様に、 A は受諾応答と呼び、各従業員が各日付に代替出勤依頼を受けた際に依頼を受諾する場合は 1 をとり、それ以外の場合に 0 をとる。このため、 $e_{i,j} = 1, a_{i,j} = 0$ は、日付 d_i において職場から従業員 w_j への代替出勤が依頼され得るが、依頼を受けた場合に従業員 w_j が代替出勤を拒否することを表す。 $e_{i,j} = 0$ であるときは、 $a_{i,j} = 0$ とする。

Q, U は代替出勤に関する制約を表す。ベクトル Q は

日付数 n の要素を持ち、各日付における代替出勤の必要人数を表す。本研究では、 A などの設定によらず解の存在を保証するため、必要人数に満たない場合を許容するソフト制約とする。ベクトル U は従業員数 m の要素を持ち、各従業員についての最大代替出勤数を表す。 $\sum_{i=1}^n s_i \leq u_j$ を満たす解のみを許容するハード制約とする。

R, S は、代替出勤依頼の回答結果を表す $n \times m$ 行列であり、各要素は 0 または 1 をとる。 R は各日付における代替出勤が各従業員に依頼された場合に 1 を、それ以外の場合に 0 をとり、 S は依頼を受けた従業員が代替出勤を受諾した場合にのみ 1 をとる。

代替出勤依頼問題の目的は、依頼対象の従業員を選択する方策関数 $\pi(d_i, q_i, E, R, S)$ を求めることである。 $\pi(d_i, q_i, E, R, S)$ は、日付 d_i において、代替出勤者の必要人数 q_i が明らかであるとき、依頼可能性 E および過去の依頼回答結果 R, S に基づいて代替出勤の依頼対象従業員 w_j を返す関数である。ただし、 w_j は次の制約を満たすものとする。

$$e_{i,j} = 1 \quad (1)$$

$$r_{i,j} = 0 \quad (2)$$

$$\sum_{i=1}^n s_i < u_j \quad (3)$$

式 (1) は従業員 w_j が依頼可能であることを、式 (2) は日付 d_i において従業員 w_j に代替出勤が依頼済みでないことを、式 (3) は従業員ごとの最大代替出勤数が制約を満たすことをそれぞれ保証する。

D, W, E, A, Q, U により表される代替出勤依頼環境において、方策 $\pi(d_i, q_i, E, R, S)$ による代替出勤依頼の結果 R, S を、アルゴリズム 1 により得る。 R, S および調整対象の日付を表す i を初期化し、方策 $\pi(d_i, q_i, E, R, S)$ により代替出勤対象の従業員 w_j を決定する。従業員 w_j が代替出勤を受諾する場合には、代替出勤が割り当てられる。日付 d_i の必要代替出勤数が充足された場合、あるいは、日付 d_i において代替出勤を依頼可能な従業員が存在しない場合は、次の日付の代替出勤の依頼を開始し、すべての日付の代替出勤依頼が完了した場合に停止する。

方策 $\pi(d_i, q_i, E, R, S)$ により得られた代替出勤依頼結果 R, S は、次の 2 つの指標により評価される。

$$U_f = \sum_{i=1}^n \left(q_i - \sum_{j=1}^m s_{i,j} \right) \quad (4)$$

$$R_{tot} = \sum_{i=1}^n \sum_{j=1}^m r_{i,j} \quad (5)$$

式 (4) は、代替出勤依頼後の未充足勤務者数を表し、各日付ごとの必要代替出勤数と依頼受諾数の差分についての総和をとる 0 以上の整数である。職場から従業員に代替出勤依頼を行う目的は、必要出勤数の充足であるため、未充足勤務者数は小さい値をとることが望ましい。式 (5) は、対象期間中に代替出勤が依頼された累計回数を表す 0 以上の整数である。職場からの代替出勤の依頼と、依頼を受けた従業員からの回答は、それぞれ負担を伴うと考えられるため、代替出勤依頼の累計回数につい

Algorithm 1 Algorithm to determine substitution

```

1: function SUBSTITUTION( $D, W, E, A, Q, \pi$ )
2:    $R \leftarrow \mathbf{0}_{n \times m}$  ▷ Initialization
3:    $S \leftarrow \mathbf{0}_{n \times m}$ 
4:    $i \leftarrow 1$ 
5:   while  $i < n$  do
6:      $w_j \leftarrow \pi(d_i, q_i, E, R, S)$  ▷ Select a worker to request
7:      $r_{i,j} \leftarrow 1$ 
8:     if  $a_{i,j} = 1$  then ▷ Request is approved
9:        $s_{i,j} \leftarrow 1$ 
10:    end if
11:    if  $\sum_{k=1}^m s_{i,k} = q_i$  (Enough substitutions found) or  $\sum_{k=1}^m e_{i,k}(1 - r_{i,k}) = 0$  (No available worker) then
12:       $i \leftarrow i + 1$  ▷ Proceed to next day
13:    end if
14:  end while
15:  return  $R, S$ 
16: end function

```

ても、小さい値をとることが望ましい。代替出勤の依頼回数が増加することに伴う負担とその悪影響の例として、従業員が頻繁に職場からの代替出勤依頼を受け取り、自身の予定を考慮しながら回答する手間を負担に感じることや、その手間を避けるため、代替出勤の依頼に対して一律に拒否の回答を返す可能性などが考えられる。

本稿で提案する定式化は、 $E = A$ であるとき、方策 $\pi(d_i, q_i, E, R, S)$ は従業員の代替出勤の受諾に関する完全な情報を受け取る。このとき、定式化された問題は A, Q, U により表される単純な制約をもつ NSP と等価とみなすことができ、方策 $\pi(d_i, q_i, E, R, S)$ は従業員の割り当てを逐次的に決定することで NSP の解を導出するといえる。

4 深層強化学習による代替出勤依頼方策

本稿では、定式化で示した代替出勤の依頼先従業員を決定する方策 $\pi(d_i, q_i, E, R, S)$ をニューラルネットワークにより構成し、深層強化学習により充足率が高く依頼回数の小さい方策を獲得する手法を提案する。

4.1 代替出勤依頼方策のニューラルネットワーク

代替出勤依頼方策 $\pi(d_i, q_i, E, R, S)$ を表現するニューラルネットワークは、従業員ごとに特徴量を算出し、各従業員に対する依頼を実施する確率を計算する。注意機構 [15] により、勤務可能日程の希少性のような従業員間の関係性を捉える。

4.1.1 従業員および依頼環境の特徴量の算出

従業員 $w_j \in W$ の特徴および必要人数 $q_i \in Q$ に対する充足率などの依頼環境の特徴を計算し、 d_{hidden} 次元の特徴ベクトルとしてニューラルネットワークへの入力とする。

従業員の特徴は、方策が代替出勤依頼対象を決定する以前になされた依頼に対する応答と、代替出勤の可能日程から計算される。応答に関する特徴として、依頼回数 $\sum_{x=0}^i r_{x,j}$ および受諾回数 $\sum_{x=0}^i s_{x,j}$ を全結合層により変換する。また、代替出勤の可能日程については、各日付に対する n 個の埋め込み表現を保持し、 $e_{x,j}$ が 1 をとる場合は、 x 番目の埋め込みベクトルを加算することで表現する。全結合層および埋め込みベクトルは、すべての従業員の間で共有し、学習により更新される。

後述する従業員ごとの特徴量の注意機構で参照するため、 $e_{i,j}$ が 0 をとる従業員についても、同様に特徴量の計算を行う。

4.1.2 従業員ごとの特徴量の処理

Transformer[15] のデコーダ部を参考とした構造により、各従業員の特徴量を処理する。提案する構造は、注意機構および多層パーセプトロン (MLP) により構成され、パラメータはすべての従業員の間で共有される。

注意機構は Multi-Head Attention 機構を用いる。Head 数を n_{head} とし、従業員数 m の特徴量系列に対して Self Attention を適用する。Transformer による言語モデルと異なり、時刻埋め込みのような各従業員を一意的に特定可能な特徴量は明示的に付加されない。

MLP 部では、全結合層により d_{hidden} 次元の特徴量を d_{inner} 次元に拡大し、得られた特徴量を再び d_{hidden} 次元に縮小する。Skip Connection により、MLP 部の入力特徴量と MLP 部の出力特徴量の加算結果を全体としての特徴量とする。

上記の構造を 1 ブロックとして、異なるパラメータのもとで n_{block} だけ繰り返し適用する。

4.1.3 従業員に対する代替出勤依頼の優先度の算出

各従業員の特徴量を全結合層を用いて 1 次元に圧縮し、 $e_{i,j}$ が 1 をとる従業員について、ソフトマックス関数を適用することで優先度を得る。提案手法では、この優先度を、それぞれの従業員に代替出勤を依頼する確率とみなし、 $\pi(w_i|d_i, q_i, E, R, S)$ を確率の方策として扱う。

4.2 深層強化学習のアルゴリズム

確率の方策として表現された $\pi(w_i|d_i, q_i, E, R, S)$ を、方策勾配法の一つである Group Relative Policy Optimization[16] に基づく手法により訓練する。

強化学習を行う環境として、 D, W および E, A, Q に関する確率分布をあらかじめ決定し、ランダムに生成された環境を用いる。生成された環境と訓練中の方策のもとでアルゴリズム 1 を実行し、得られた結果 R, S から式 (5, 5) を用いて次の通り報酬を求める。

$$\text{reward} = U_f + \alpha R_{\text{tot}}$$

従業員ごとに代替出勤の依頼対象となる確率を示す確率の方策 $\pi(w_i|d_i, q_i, E, R, S)$ の更新は、Proximal Policy

Optimization[17] で提案されたサロゲート関数を用いて

$$L_{\text{CLIP}}(\theta) = \hat{\mathbb{E}}_r \left[\min \left(r_t(\theta) \hat{A}_t, \text{clip}(r_t(\theta), 1 - \epsilon, 1 + \epsilon) \hat{A}_t \right) \right]$$

$$r_t = \frac{\pi_{\theta}(w_i | d_i, q_i, E, R, S)}{\pi_{\theta_{\text{old}}}(w_i | d_i, q_i, E, R, S)}$$

の最小化により行う。ただし、 θ_{old} および θ は確率的方策を表すニューラルネットワークの更新前後のパラメータを表し、 $\hat{\mathbb{E}}_r$ は有限回の試行の平均値により勾配の値を推定していることを示す。 \hat{A}_t は確率的方策によるアドバンテージであり、特定の 1 回の依頼先従業員の選択により、報酬が改善される見込みを示す。

アドバンテージ \hat{A}_t の推定のため、本稿では GRPO に基づく手法を提案し、その適用結果を報告する。同一の環境 D, W, E, A, Q および確率的方策 $\pi(w_i | d_i, q_i, E, R, S)$ のもとで G 回の独立したアルゴリズム 1 の実行を行い、得られる報酬を $\text{reward}_1, \text{reward}_2, \dots, \text{reward}_G$ とする。このとき、 y 回目の適用で得られた確率的方策の実行結果に対するアドバンテージは、同一の方策が同一の環境から得た報酬の平均値に対する差分として、

$$\hat{A}_t = \text{reward}_y - \sum_{x=1}^G \text{reward}_x$$

により求める。

確率的方策の更新は、 $G \times P$ 回のアルゴリズム 1 実行結果から $L_{\text{CLIP}}(\theta)$ を求め、ミニバッチ法によりこれを最小化することで実施する。

5 数値計算実験

本稿では、30 名の従業員に対する 14 日間の代替出勤依頼問題を対象として、提案手法の有効性を検証する。

代替出勤依頼の応答可能性 E および受諾応答 A の環境変動を考慮した方策の評価を行うため、代替出勤依頼の環境 E, A を確率的に生成し、得られた複数の環境インスタンスで方策の平均性能を評価する。

応答可能性 E は、日付および従業員ごとに独立なベルヌーイ分布に従う確率変数として生成する。これは、従業員の休暇申請、病欠、および既存の勤務割当などの要因が互いに独立に発生するという仮定に基づいている。

受諾応答 A は、従業員を高受諾確率群と低受諾確率群の 2 つに分け、それぞれ異なる確率でベルヌーイ分布に従う確率変数として生成する。著者らの先行研究 [18] におけるヒアリングにより、積極的に代替出勤に応じる従業員と、そうでない従業員が存在することが確認されているためである。

確率パラメータの組み合わせごとに N 件の環境インスタンスを独立に生成し、各方策に対してアルゴリズム 1 を適用する。得られた結果から、未充足勤務者数 U_f および依頼回数 R_{tot} を算出し、それぞれの平均値を評価指標とする。これにより、方策の違いが代替出勤の決定に与える影響を明らかにするとともに、環境設定ごとの感度を分析する。感度分析に用いる環境インスタンスの件数 N は 1000 とし、計算時間の制約からルールベースによる方策のみを用いた比較を行う。

最後に、感度分析により特定された、深層強化学習による方策の獲得が有効性を持ちうるパラメータのインスタンスにおいて、提案手法により学習した方策を獲得し、ルールベースの代替出勤依頼方策と比較する。

5.1 比較対象のルールベース方策

比較するルールベースの代替出勤依頼方策は、次の通りである。

ランダム選択 w_j を候補従業員の中から一様分布によりランダムに選択する。各従業員の状況を考慮せず、完全にランダムに依頼対象を決定する方策であり、依頼に対する受諾傾向や過去の履歴を考慮しないため、基準となる比較手法として用いる。

代替出勤可能日数の残数の昇順 各従業員 w_j について、日付 d_i 以降に代替出勤候補となる日数の総和を計算し、その値が最も小さい従業員を選択する。同数の従業員が複数存在した場合にはランダムに 1 名を選択する。このルールは、今後の出勤依頼の余地が少ない従業員を優先する方策であり、将来的な選択肢の枯渇を回避することを目的としている。

残り代替出勤回数の昇順 w_j に対して残り代替出勤可能回数 $u_j - \sum_{i=1}^n s_{i,j}$ の最も小さい従業員を選択する。同数の従業員が複数存在した場合には、ランダムに 1 名を選択する。このルールは受諾頻度の高い従業員に優先して依頼を行う傾向を持ち、高受諾率の従業員が存在することを前提に、依頼回数を減少させることを目的とする。

残り代替出勤回数の降順 w_j に対して残り代替出勤可能回数 $u_j - \sum_{i=1}^n s_{i,j}$ の最も大きい従業員を選択する。同数の従業員が複数存在した場合には、ランダムに 1 名を選択する。このルールは、特定の従業員への代替出勤依頼の集中を避けることを目的としており、将来的な選択肢の枯渇を回避することを目的としている。

整数計画ソルバによる下限 比較のため、 A, U, Q に関する制約のもとで未充足勤務者数 U_f を最小化する解を、整数計画ソルバにより算出し、下限値 (Lower bound) として提示する。この解はアルゴリズム 1 において直接の参照ができない受諾情報 A を用いて算出されるため、方策として実装可能ではないが、近似最適性を評価するベンチマークとして用いる。

5.2 依頼可能性・受諾応答パラメータの感度分析

本節では、代替出勤依頼における応答可能性 E および受諾応答 A の生成に用いる確率パラメータを変化させ、各方策の性能指標に対する感度を分析する。

まず、すべての従業員が同質であると仮定し、 E および A の各要素を同一の確率パラメータに基づくベルヌーイ分布から独立に生成する設定を検討する。このとき、単純なルールベースの選定方策によって得られる結果と、比較対象として設定された理論的な充足率の上界を比較し、方策の性能が環境条件により大きく左右されるパラメータ領域を特定する。

次に、特定された領域において、従業員を高受諾率群と低受諾率群の 2 群に分割し、受諾応答 A の群間の差の大きさと方策性能との関係を検証する。この比較により、従業員間の異質性が代替出勤依頼方策に与える影響を明らかにする。

5.2.1 同質な従業員に対する感度分析

本節では、提案手法および比較対象となるルールベース方策の有効性を評価するために、シミュレーションに

よる数値計算実験を行う。

対象とする環境は、従業員数 30 名、期間 14 日間とし、各日付において必要とされる代替出勤者数を 3 名、従業員ごとの最大代替出勤回数を 2 回に設定した。

環境は、従業員全体が同質であると仮定し、出勤可能性 E および代替出勤依頼に対する受諾応答 A の要素を、それぞれ独立なベルヌーイ分布により確率的に生成する。応答可能性および受諾確率はともに 0.10 から 1.00 まで 0.05 刻みで変化させ、計 $19 \times 19 = 361$ 通りの環境パラメータの組み合わせを設定した。各組み合わせに対して独立に 1000 件のインスタンスを生成し、平均性能を比較することで、環境条件に対する感度分析を実施する。

比較対象とする方策は、ランダム選択、代替出勤可能日数の残数による昇順選択、残り代替出勤回数の昇順・降順選択の 4 種であり、さらに提案手法である深層強化学習に基づく方策と、整数計画法により導出される最適解をあわせて比較する。

性能評価指標としては、未充足勤務者数（必要人数と受諾者数の差分の総和） U_f および累積依頼回数 R_{tot} の平均値を用いる。

図 2 は、代替出勤依頼における環境条件と各方策の性能の関係を示している。図 1a は、応答可能性および受諾確率の各組合せにおいて、整数計画ソルバにより算出された未充足勤務者数 U_f の下限値の平均を示している。下限値は、応答可能性および受諾確率が共に低い左上の領域で高く、右下に向かうにつれて低下する傾向が確認される。これは、代替出勤の依頼対象者が少ない環境や、受諾が得られにくい環境では、構造的に代替出勤者の充足が困難であることを反映している。

図 1b は、最良のルールベース方策によって得られた代替出勤者数の受諾総数を、同条件下での整数計画ソルバによる代替出勤者数で割った比率を示す。0 除算を避けるため、未充足数ではなく、代替出勤の受諾総数を用いていることに注意されたい。比率が 1.0 に近いほど、ルールベース方策が理想的な水準に近い性能を発揮していることを意味する。一方で、比率が小さい領域では、単純なルールベース方策では対応が困難であり、改善の余地が大きいことを示す。特に、応答可能性および受諾確率が中程度である条件では、代替出勤依頼の難易度が高く、ルールベースによる依頼戦略の限界が現れる。

これらの分析結果から、本研究では応答可能性 0.5、受諾確率 0.4 の条件を、提案手法の学習および評価における代表的な環境設定として選定した。

5.2.2 受諾確率の異なる従業員に対する感度分析

次に、従業員の代替出勤依頼に対する応答傾向に異質性が存在する場合の方策性能について分析を行う。本実験では、従業員を受諾確率の異なる 2 群に分割した上で、その構成の変化に対する方策の感度を検証する。

具体的には、低受諾確率群の受諾確率を一定間隔で変化させるとともに、高受諾確率群の従業員数を変化させ、その両者の組合せに対して環境を生成する。ただし、全体の平均受諾確率が先の実験でも求められた 0.4 となるように、高受諾確率は内挿的に設定される。このとき、低受諾確率の従業員数が増えるほど、平均を保つために高受諾確率はさらに高くなる必要がある。このような条件下で、ルールベースの方策が十分に適応できる

かどうかを分析する。

この分析においても、先の分析と同様に、ルールベースの方策によって得られた代替出勤の受諾総数と、整数計画ソルバによる下限値との比を指標とし、受諾確率の異質性が高まる場合（すなわち 2 群の性能差が顕著な条件）における方策の性能限界と、学習方策による改善の余地を明らかにする。

図 2 は、受諾確率の異なる 2 つの従業員群の構成によって、各方策の性能がどのように変化するかを示している。

図 2a は、環境条件ごとに整数計画ソルバによって算出された未充足勤務者数 U_f の下限値平均を示している。縦軸は高受諾確率群の従業員数、横軸は低受諾確率群の受諾確率を表す。図の左端では、低受諾確率群の応答確率がほぼゼロであるため、高受諾確率群の従業員のみで代替出勤を賄う必要がある。この場合、高受諾群に依頼が集中し、最大代替出勤回数の制約により未充足が発生しやすくなる。このような条件では、構造的に勤務者の充足が困難であることが明らかとなる。

図 2b は、最良のルールベース方策により得られた代替出勤の受諾総数を、整数計画ソルバによる受諾総数で割った比率を示している。この比率は 1.0 に近いほど、ルールベース方策が理想に近い性能を発揮していることを意味し、値が小さい領域では改善の余地があることを示す。なお、未充足数の直接比較ではゼロ除算の懸念があるため、本比較では受諾数を使用している。特に、図 2a において理論的に未充足が発生し始める条件付近では、単純なルールベース方策の性能低下が顕著となり、工夫された依頼戦略の重要性が高まる。

これらの結果を踏まえ、本研究では、低受諾確率群の受諾確率を 0.26、高受諾確率群の従業員数を 6 とする条件を、提案手法の学習および評価における代表的な環境として選定した。このとき、平均受諾確率 0.4 を維持するために、高確率群の受諾確率は 0.96 と設定されている。

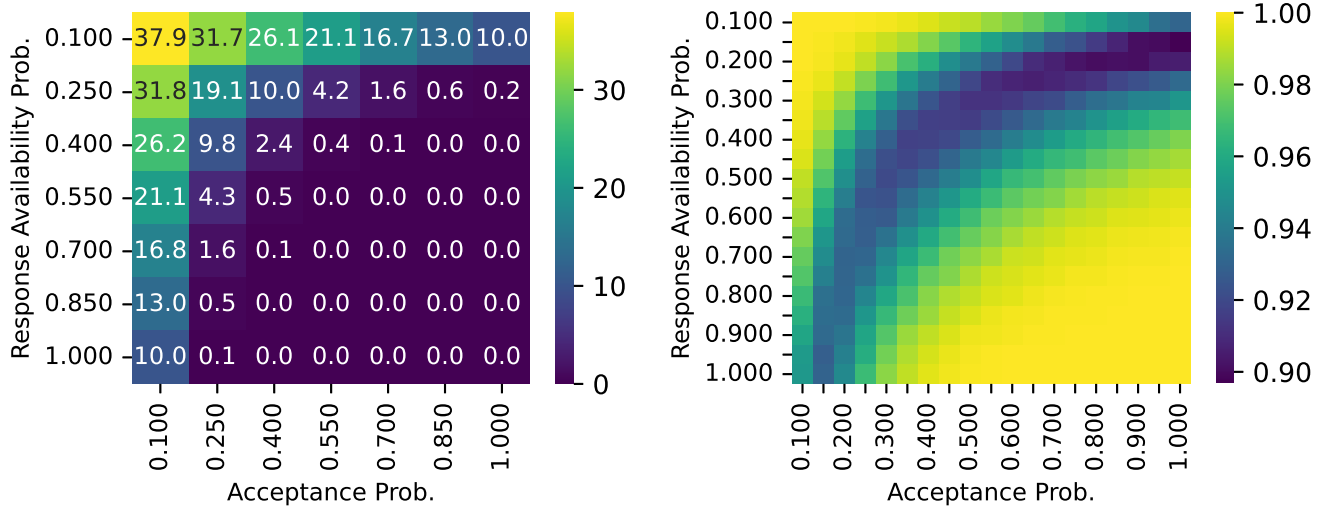
5.3 実験方法

予備実験と同様に、 E, A, Q を確率的に生成して強化学習を行うとともに、得られた方策とルールベースの方策を固定された 1000 の環境のもとで比較する。

E, A, Q は次の通り設定する。7 日間のうちランダムに 5 日を選択し、30 名の各従業員に対して 70% の確率で $e_{i,j} = 1$ とする。残る 2 日については、30% の確率で $e_{i,j} = 1$ とする。これ以外の場合は、 $e_{i,j} = a_{i,j} = 0$ とする。次に、6 名の従業員について、 $e_{i,j} = 1$ を満たす i, j について、80% の確率で $a_{i,j} = 1$ 、それ以外の場合には $a_{i,j} = 0$ とする。残る 24 名の従業員については、10% の確率で $a_{i,j} = 1$ とする。 Q はすべての日付に対して 2、 U はすべての従業員に対して 2 とする。

得られた方策およびルールベースの方策を固定された 1000 の環境に適用し、 U_f, R_{tot} を示す。また、 A, U, Q に関する制約のもとで、整数計画ソルバにより U_f を最小化した厳密解の値を求める。

表 2 に示す設定で方策の学習を行い、先述の方法でランダム生成された 1000 の環境において、 U_f の値を算出する。方策の学習は独立して 5 試行だけを行い、その平均値を示す。注意機構の貢献を示すため、提案手法から注意機構を取り除いた方策についても、同様に学習し比較



(a) 未充足勤務者数 U_f の下限値平均 (b) 最良ルールベース手法の代替出勤者数の比率

図 1: 同質の従業員群における方策の性能比較



(a) 未充足勤務者数 U_f の下限値平均 (b) 最良ルールベース手法の代替出勤者数の比率

図 2: 2つの従業員群における方策の性能比較

する。また、同じ環境に対して、 $E = A, U, Q$ に関する制約のもとで、 U_f を最小化する厳密解を整数計画ソルバにより求め比較する。

表 3 に手法ごとの U_f, R_{tot} の値を示す。提案手法に関しては、5 回試行の平均値を示した。

表 3 の結果から、本実験設定においては、注意機構を用いた提案手法が U_f (未充足勤務者数) およびにおいて、注意機構を除いた同手法と比較してわずかに優れた性能を示した。ただしその差は小さく、 R_{tot} (依頼回数) についてはわずかながら注意機構を除いた手法が優れた性能を示した。これは従業員の一部 (6 名) のみが高い受諾確率を持つ設定において、主に依頼先の選択タイミングが重要であり、従業員間の関係性を捉える注意機構の寄与が相対的に限定的であったためと考えられる。

一方、ルールベースの各手法と比較した場合、提案手法は R_{tot} の増大と引き換えに U_f を減少させ、パレート最適な方策を実現した。残り代替出勤回数の昇順による

ルールでは、受諾確率の高い従業員への依頼が集中するため依頼回数は少なくなるが、期間後半では最大代替出勤回数制約により高確率従業員が依頼不能となり、未充足が顕著となる傾向が見られた。逆に、残り回数の降順におけるルールでは、未充足は抑えられる一方、依頼回数が増加する傾向にあり、いずれの手法も一方の指標に偏った結果となっている。これに対して提案手法は、未充足数を顕著に抑えることに成功しており、現実的な運用上でも選択肢となりうる性能を発揮しているといえる。

また、整数計画ソルバにより得られた理論的な下限と比較すると、提案手法は U_f において二倍以上の差であった。ソルバが全従業員の受諾可否をあらかじめ知っている前提に基づく理想解であることによるものか、提案手法の改善の余地を示唆しているものかを判断するために、今後の検討が求められる。なお、提案手法では依頼回数と未充足のトレードオフを調整するパラメータ α

表 2: 実験設定

Parameter	Value
n	7
m	30
d_{hidden}	16
d_{inner}	64
n_{block}	2
n_{head}	1
α	0.01
ϵ	0.2
P	256
G	16
Optimizer	Adam
Learning rate	1e-3
Minibatch size	128
Number of policy updates	100

表 3: 実験結果

Method	U_f	R_{tot}
提案手法	2.51	119.4
提案手法 (注意機構を除く)	2.56	118.7
ランダム選択	4.96	109.3
代替出勤可能日数の残数の昇順	3.96	112.6
残り代替出勤回数の降順	3.47	115.0
残り代替出勤回数の昇順	5.73	106.6
整数計画ソルバによる下限	0.94	-

を固定して学習を行っているが、この値を適切に調整することで、他のパレート最適点への誘導も可能であると考えられる。

6 おわりに

本研究では、スタッフスケジューリングにおける不確実性を考慮した代替出勤依頼問題を定式化し、深層強化学習による依頼方策の獲得手法を提案した。提案手法では、注意機構を活用して従業員間の関係性を考慮しつつ、確率的方策に基づく代替出勤依頼を決定するニューラルネットワークを構築し、Group Relative Policy Optimization (GRPO) を用いた強化学習により訓練を行った。

次に、提案手法を用いて生成したランダム環境に対し、ルールベース手法や整数計画ソルバによる下限解と比較する数値実験を行い、未充足勤務者数 U_f と累積依頼回数 R_{tot} の両面でパレートの優位性を示した。特に、注意機構を導入することで未充足数にわずかながら性能向上が確認でき、ルールベースが達成できない低い未充足数が実現されたことから、実運用における有用性が示唆された。

今後の課題として、実際の職場環境における代替出勤依頼の動向をより詳細に分析し、実務に適用可能なモデルの精緻化を進めることが挙げられる。また、未充足勤務者数と依頼回数のトレードオフを調整するパラメータの最適化や、動的な環境における適応性の向上についても検討を行う。

謝辞

本研究は 2025 年度 (株) トリニティ・ドゥ奨学寄付金の支援を受けて実施されました。

参考文献

- [1] Atsuko Ikegami and Akira Niwa. A subproblem-centric model and approach to the nurse scheduling problem. Vol. 97, No. 3, pp. 517–541.

- [2] Stefaan Haspeslagh, Patrick De Causmaecker, Andrea Schaefer, and Martin Stølevik. The first international nurse rostering competition 2010. Vol. 218, No. 1, pp. 221–236.
- [3] Christos Valouxis, Christos Gogos, George Goulas, Panayiotis Alefragis, and Efthymios Housos. A systematic two phase approach for the nurse rostering problem. Vol. 219, No. 2, pp. 425–433.
- [4] Omar EL-Rifai, Thierry Garaix, Vincent Augusto, and Xiaolan Xie. A stochastic optimization model for shift scheduling in emergency departments. Vol. 18, No. 3, pp. 289–302.
- [5] Alessandra Parisio and Colin Neil Jones. A two-stage stochastic programming approach to employee scheduling in retail outlets with uncertain demand. Vol. 53, pp. 97–103.
- [6] Ward Whitt. Staffing a call center with uncertain arrival rate and absenteeism. Vol. 15, No. 1, pp. 88–102. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1937-5956.2006.tb00005.x>.
- [7] Margarida Vaz Pato and Margarida Moz. Solving a bi-objective nurse rostering problem by using a utopic pareto genetic heuristic. Vol. 14, No. 4, pp. 359–374.
- [8] Ricardo Otero-Caicedo, Carlos Eduardo Montoya Casas, Carolina Barajas Jaimes, Cristian Felipe Guzmán Garzón, Edwin Andrés Yáñez Vergel, and Julián Camilo Zabala Valdés. A preventive-reactive approach for nurse scheduling considering absenteeism and nurses' preferences. Vol. 38, p. 100389.
- [9] 幡本 昂平, 横山 想一郎, 山下 倫央, 川村 秀憲. 代替従業員確保のためのメッセージングアプリを利用した効率的な依頼手法の開発. Vol. 60, No. 10, pp. 1757–1768.
- [10] Soichiro Yokoyama, Tomoya Chisaka, Tomohisa Yamashita, and Hidenori Kawamura. Evaluation of request order sequencing methods in the selection process of substitute employees. Vol. 3, No. 2, pp. 1–18. Number: 2.
- [11] Wouter Kool, Herke van Hoof, and Max Welling. Attention, learn to solve routing problems!
- [12] Yi Feng, Lu Zhang, Zhile Yang, Yuanjun Guo, and Dongsheng Yang. Flexible job shop scheduling based on deep reinforcement learning. In *2021 5th Asian Conference on Artificial Intelligence Technology (ACAIT)*, pp. 660–666.
- [13] Masato Nagayoshi and Hisashi Tamaki. Constructive nurse scheduling using reinforcement learning considering variations in nurse work patterns. Vol. 29, pp. 325–328.
- [14] Ahmad Muklason, Shindu Dimas Rizal Kusuma, Edwin Riksakomara, I. Gusti Agung Premananda, Wiwik Anggraeni, Faizal Mahananto, and Raras Tyasnurita. Solving nurse rostering optimization problem using reinforcement learning - simulated annealing with reheating hyper-heuristics algorithm. Vol. 234, pp. 486–493.
- [15] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, Vol. 30. Curran Associates, Inc.
- [16] DeepSeek-AI et. al. DeepSeek-r1: Incentivizing reasoning capability in LLMs via reinforcement learning.
- [17] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms.
- [18] Kohei Hatamoto, Soichiro Yokoyama, Tomohisa Yamashita, and Hidenori Kawamura. Development of efficient request method using messaging app for substitute fulfillment (in japanese). Vol. 60, No. 10, pp. 1757–1768.