深層強化学習型麻雀 AI の行動選択に対する Feature Attribution

大倉 功士　和賀 正樹　池渕 未来　末永 幸平

深層強化学習を用いて実装された麻雀 AI は，高い競技性能を達成するようになっている一方で，選択した行動の根拠を人間が理解することは困難な場合が多い．この問題を解決するために，本論文では麻雀 AI Mortal の行動選択の根拠を説明する以下の 2 手法を提案する: (1) 強化学習 AI のための摂動に基づく feature attribution 手法 SARFA を Mortal の入力表現に適用する手法，(2) SARFA の特異度と関連度の定義から導出される近似を用いて，入力特徴量の勾配からスコアを計算する手法．M リーグ 2025–2026 シーズンの牌譜から抽出した 1,004 行動時点で両手法を評価した結果，勾配法は説明単位全体の重要度順位とスコア符号において摂動法の近似に利用できる一方，最大スコアを持つ説明単位の一致率は 17.5%にとどまった．勾配法は全 1,004 行動時点で摂動法より高速であり，平均計算時間を約 8 分の 1 に短縮した．

1　はじめに

チェス，将棋，囲碁などのボードゲームで，機械学習により実装された AI が人間のトッププレイヤーを上回る性能を達成するようになっている [18]．不完全情報ゲームであるリーチ麻雀も例外ではなく，例えば深層強化学習によって実装された AI である Suphx は，オンライン麻雀サイト天鳳で公式順位を持つプレイヤーの 99.99%を上回る成績を収めている [11]．

しかし，多くの麻雀 AI が深層強化学習によって実装されているために，AI が各局面における行動に与える評価値の根拠を人間が理解することは困難である．これは特に，牌譜の分析に AI を利用したり，AI の判断を人間が学習や解説に利用する場合に問題となる．

本論文では，深層強化学習によって実装された麻雀 AI である Mortal [4] について，与えられた局面における各合法行動の softmax 相対評価に対する入力特徴量の符号付き feature attribution スコアを計算する 2 つの手法を提案する．麻雀 AI には多くの実装が存在するが，(1) 局面の入力特徴量への符号化方法とモデルの実装が公開されていること，(2) 入力特徴量が手牌，河，点数などの意味的情報に基づいて与えられるため，feature attribution の結果を人間が理解しやすいことから，本論文では Mortal を麻雀 AI として選択した．

第一の手法は，強化学習 AI の ための feature attribution 手法として提案された SARFA [15] を Mortal に適用する**摂動法**である．この手法においては，行動 a に対する説明単位 f の feature attribution スコアを，a の Q 値の相対的な softmax 値が f をマスクすることによってどの程度変化するかを表す値（**特異度**）と，a 以外の合法行動の Q 値の softmax 分布が f をマスクすることによってどの程度変化しづらいかを表す値（**関連度**）から計算するものである．前者は a の相対評価が f にどの程度依存しているかを表し，後者は f がどの程度 a のみに影響を与えているかを表す．スコアが正の場合は f が行動 a の相対評価を支持し，負の場合は抑制すると解釈する．

摂動法はモデルの内部パラメタに対するアクセス

Feature Attribution for Action Selection by a Deep Reinforcement Learning-Based Mahjong AI
This is an unrefereed paper. Copyrights belong to the Authors.

Koji Okura, Masaki Waga, Mirai Ikebuchi, Kohei Suenaga, 京都大学大学院情報学研究科, Graduate School of Informatics, Kyoto University.

図 1　可視化用アプリケーションによる符号付きスコアの表示例．手牌の を打牌するという行動に対してのスコアが可視化されている．赤はその情報があることによって を打牌することが支持されることを示し，青は抑制されることを示す． を打牌するという行動が **(1)** リーチ者である下家の宣言牌が であるという情報によって支持されていることと **(2)** （おそらく の打牌で面子が崩れることにより） が手牌にあるという情報によって抑制されていることが伺える．

がなくとも実行できるという利点があるものの，入力特徴量テンソルを摂動させて繰り返し推論を行う必要があるため，計算時間がかかるという欠点がある．そこで，第二の手法として，特異度と関連度を，各合法行動の Q 値の入力勾配を用いて近似する**勾配法**を提案する．具体的には，Q 値の入力特徴量に対する勾配によって特異度を一次近似し，関連度の計算に用いる KL ダイバージェンスを Q 値の softmax 分布の Fisher 情報行列が誘導する二次形式によって近似する．

提案手法と，スコアを可視化しつつ牌譜を再生する GUI アプリケーションとを実装した．図 1 に，局面を可視化した例を示す．†1

この実装を用いて，実際の牌譜に対する feature attribution スコアを両手法で計算した．M リーグ 2025–2026 シーズンの牌譜から抽出した，1,004 個の行動時点を含む牌譜を用いて両手法を比較した結果，定性的分析では手牌または他家の河と Mortal の選択行動との関係を読み取れることを確認した．定量評価では，勾配法は説明単位全体の重要度順位と符号の近似に利用できる一方，最大スコアを持つ説明単位の一致率は 17.5%にとどまった．また，勾配法は全行動時点で摂動法より高速であり，平均計算時間は摂動法の約 8 分の 1 であった．

本論文の貢献は以下の通りである．

1. 特異度の一次近似と，関連度を定める KL ダイバージェンスの Fisher 情報行列が誘導する二次形式を導出し，各合法行動の Q 値の入力 Jacobian

†1　本図のキャプションおよび 6 節には，実装によって出力された feature attribution スコアを著者らが解釈した記述が含まれている．参考までに，著者中にはプロ資格を持つものはおらず，執筆時点での著者ごとの雀魂でのランクは以下の通りである: 第 1 著者=雀傑 1，第 2 著者=雀士 2，第 3 著者=雀傑 3，第 4 著者=雀士 3.

から SARFA 型スコアを近似的に計算する勾配法を提案した．

2. 両手法と，スコアを可視化しつつ牌譜を再生する GUI アプリケーションとを実装した．
3. M リーグの牌譜から抽出した 1,004 行動時点で両手法を比較し，勾配法が説明単位全体の重要度順位と符号の近似に利用できる一方，最大スコアを持つ説明単位の一致率は 17.5%にとどまること，ならびに全行動時点で摂動法より高速であり，平均計算時間を 87.6%短縮することを明らかにした．

本論文の構成は以下の通りである．2 節では feature attribution とゲーム AI の説明手法に関する関連研究を議論する．3 節では Mortal の入力特徴量と行動空間について説明する．4 節では提案する二つの feature attribution 手法を説明する．5 節では提案手法の実装について説明する．6 節では実験について述べる．

2 関連研究

2.1 麻雀 AI の行動説明手法

これまでに麻雀 AI の行動説明手法がいくつか提案されている．岡本らは CNN による打牌選択モデルの判断根拠を可視化し，復習手法による説明と比較を行い [25]，続く研究で Kernel SHAP を適用した手法を提案した [26]．高野らは SHAP と Grad-CAM による入力特徴量の帰属を自然言語で提示する対話型牌譜検討支援システムを提案した [20]．Fang らは，打牌判断の根拠を言語化するニューロシンボリックなモデルである NeSy-Mahjong を提案した [5]．また，広く用いられている麻雀 AI の NAGA に対して，Guided Backpropagation を用いて入力特徴量に対する feature attribution を行う機能が実装されている [3]．本論文は Mortal v4 の入力特徴量テンソルを対象とし，SARFA 型の摂動法と勾配法によって，合法行動の softmax 相対評価に対する入力特徴量の符号付きスコアを計算する点において，これらの研究とは異なっている．

2.2 強化学習エージェントの行動説明手法

Greydanus らは，入力映像の局所領域へ摂動を加えたときの方策と価値関数の変化から saliency map を構成し，Atari エージェントの行動を説明する手法を提案した [8]．SARFA は特異度と関連度とを組み合わせた指標により，説明対象の行動に固有の特徴量を調べる手法である [15]．廣瀬らは，SARFA を負の特異度も扱えるように拡張し，将棋 AI である dlshogi のための feature attribution 手法を提案した [23]．その後廣瀬らは，解釈しやすい特徴量集合で attribution を行う手法を設計し，dlshogi 以外にも適用可能な説明手法を提案した [24]．本論文で提案する手法は SARFA に基づく説明手法を Mortal の入力特徴量へ適用したものである．SARFA はこれまで入力特徴量に対する摂動から計算されてきた．著者らの知る限り，説明対象行動を除く合法行動の相対評価分布に対する Fisher 二次形式を用いて SARFA の関連度を局所近似し，これを特異度の一次近似と符号付き調和平均によって統合した，勾配に基づく feature attribution 手法はこれまで提案されていない．

2.3 深層モデルのための feature attribution 手法

Feature attribution は AI の推論結果の解釈性を向上させる手法として広く用いられている．Simonyan らは，クラススコアの入力勾配を saliency map として用いる手法を提案した [19]．Ancona らは，Gradient × Input を含む勾配に基づく feature attribution 手法を理論的および実験的に整理し，摂動に基づく feature attribution 手法との関係を論じた [1]．LIME は対象入力の近傍で解釈可能な代理モデルを学習し，これに基づいて feature attribution を行う手法である [16]．また，SHAP は協力ゲームにおける公平な利得配分である Shapley 値を，特徴量ごとの推論結果の寄与として用いる手法である [12]．Covert らは，特徴量をマスクすることで feature attribution を行う手法を removal-based explanation として統一的に理解する枠組みを与えた [2]．特徴量のマスク方法について，単に特徴量の値を 0 に置換するだけでは意味的に妥当な入力になるとは限らない問題を解決するために，

ぼかしやノイズなどの解釈可能な摂動を用いる手法も提案されている [6]．畠山らは，black-box 画像分類モデルを対象として，色成分ごとの saliency を可視化する手法を提案した [9]．菊池らは，画像および動画分類モデルの saliency map をベイズ最適化によって精緻化する BOREx を提案した [10]．

本研究で用いた Fisher 情報行列をモデル解釈に利用する研究も行われている．Gat らは，functional entropy を functional Fisher information によって上方近似し，入力特徴量および特徴量集合の情報寄与を測る feature attribution 手法を提案した [7]．Martino らは，解釈可能性向上手法 Integrated Gradients の計算に用いる積分経路として，分類器の softmax 出力が入力空間に誘導する引き戻し Fisher 計量を用いて，予測分布空間上の Fisher–Rao 測地線に対応する入力空間上の経路を利用する手法 FRInGe を提案した [13]．Sengupta らは，入力に依存する Fisher 情報行列の固有空間を用いて，入力画像を予測分布の局所変化に対する感度が高い成分とその直交成分に分解する手法を提案した [17]．Zhao らは，分類器が入力空間に誘導する Fisher 計量の最大固有値に対応する方向を用いる敵対的攻撃手法と，固有値を特徴として用いる敵対的入力の検出手法を提案した [22]．Miyato らは，入力摂動の前後における出力分布間の KL ダイバージェンスについて，摂動 0 で一次微分が 0 になることから Hessian による二次近似を用い，分布の局所平滑性を学習する Virtual Adversarial Training を提案した [14]．本論文は，(1) 局所的な KL ダイバージェンスの近似を学習時の正則化ではなく，SARFA の関連度を勾配から近似するために用いる点，(2) Fisher 二次形式によって SARFA の関連度を局所近似する点で，これらの研究と異なる．

3 Mortal

3.1 アーキテクチャ

Mortal は深層強化学習によって実装されたオープンソースのリーチ麻雀用 AI である [4]．本論文では Mortal v4 を単に Mortal と呼ぶ．また，本論文ではリポジトリ中のコミット `d5e80bff` の実装を参照している．

表 1　Mortal v4 の入力特徴量の概要

チャネル	内容
0–34	自家の手牌，点数，順位，局，本場，供託，風，局進行度，ドラ表示牌
35–716	自家と他家 3 人の河の時系列情報
717–750	残りツモ可能牌数，ドラ情報，各家の捨て牌集計
751–835	各家の副露，暗槓，観測済み牌数
836–870	他家の直近手出し牌と立直情報，自家の手牌状態
871–888	応答対象牌，打牌候補，行動可能フラグ
889–1011	受け入れ枚数や期待値の情報

Mortal は局面の評価を行うために DNN を用いている．DNN は局面に対応する入力特徴量テンソルを受け取り，その局面における各行動の Q 値を出力する．Mortal は出力された Q 値のうち合法行動でかつ最大のものを選択することで，局面における行動を決定する．DNN は特徴抽出を行う 1 次元畳み込み層からなる ResNet と，その出力を受け取って Q 値を計算する Dueling DQN [21] からなる．

3.2 入力特徴量テンソル

Mortal の DNN は局面から生成される大きさ $(1012, 34)$ の入力特徴量テンソルを入力として受け取る．テンソルの第 1 次元を**チャネル**，第 2 次元を**牌種列**と呼ぶ．34 列の牌種列は から，から ， から ， ， ， ， ， ， ， ， の各牌に対応する．赤牌 ， ， は，黒牌の ， ， と同一の牌種列で表現されるが，赤牌フラグを保持するチャネルの値で区別される．プレイヤーに関する特徴量は，自家を 0 とする相対席順で並ぶ．

表 1 に入力特徴量テンソルの概要を示す．詳細は appendix A で説明する．

3.3 Q 値

Mortal の Dueling DQN は，スカラーの状態価値 $V(\mathbf{x}_t)$ と 46 次元の advantage ベクトル $A(\mathbf{x}_t)$ を計算し，両者から各行動の Q 値を構成する [21]．Q 値ベクトルの 46 次元は，(1) 打牌やチー・ポン・カンにおける牌種選択を行うための 27 種の黒数牌、7 種

の字牌、3 種の赤牌に対応する 37 次元，(2) リーチ宣言行動に対応する 1 次元，(3) チー・ポン・カンの行動を表す 5 次元 †2，(4) 和了行動に対応する 1 次元，(5) 九種九牌における流局行動を表す 1 次元，および (6) 鳴きをパスすることを表す 1 次元からなる．

局面 t の合法な行動の集合を $\mathcal{A}_t$ とすると，局面 t における行動 $a \in \mathcal{A}_t$ の Q 値 $Q_a(\mathbf{x}_t)$ は

$$Q_a(\mathbf{x}_t) := V(\mathbf{x}_t) + A_a(\mathbf{x}_t) - \frac{1}{|\mathcal{A}_t|}\sum_{b\in\mathcal{A}_t} A_b(\mathbf{x}_t)$$

で定義され，DQN は出力がこれを近似するように学習される．ここで $V(\mathbf{x}_t)$ は局面 t の状態価値，$A_a(\mathbf{x}_t)$ は局面 t における行動 a の advantage である．Mortal は DQN の出力のうち，現在の局面で合法ではない行動の Q 値を $-\infty$ へ置換した上で，Q 値が最大の行動を選択する [4]．以下では a が非合法な行動であった場合に，$Q_a(\mathbf{x}_t)$ の値はあらかじめ $-\infty$ となっているものとする．また，$Q(\mathbf{x}_t)$ で局面 t における全行動の Q 値ベクトルを表す．

4 提案手法

局面 t における行動 a の**相対評価値** $P(a \mid \mathbf{x}_t)$ を

$$P(a \mid \mathbf{x}_t) := \frac{\exp Q_a(\mathbf{x}_t)}{\sum_{b\in\mathcal{A}_t}\exp Q_b(\mathbf{x}_t)}$$

と定義する．$P(a \mid \mathbf{x}_t)$ は合法行動の Q 値に温度 1 の softmax を適用した値であり，$\mathcal{A}_t$ 上の確率分布をなす．以下では，$\mathcal{A}_t \neq \varnothing$ である局面を対象とし，局面 t に対応する合法行動集合 $\mathcal{A}_t$ を各手法の計算中に固定する．

Feature attribution を行う際に，入力特徴量テンソルの各要素のスコアを個別に計算すると，入力特徴量の数が多く，また入力特徴量の複数の要素が同一の情報を冗長にエンコードする場合があるため解釈が困難になる．そのため本論文では，入力特徴量の要素をいくつかのグループに分けて，そのグループごとにスコアを計算するアプローチを取る．この各グループを**説明単位**と呼ぶ．より形式的には，入力特徴量テンソルの添字の集合を $\mathcal{I} := \{0,\dots,1011\} \times \{0,\dots,33\}$ と定義する．説明単位のラベル集合を $\mathcal{F} = \{f_1,\dots,f_n\}$ とし，各ラベルに対応する添字集合を与える写像を

$$S : \mathcal{F} \to 2^{\mathcal{I}}$$

と定義する．以下では，説明単位 $f \in \mathcal{F}$ に対応する添字集合 $S(f)$ を S_f と略記する．

4.1 摂動法

提案手法の一つである摂動法について説明する．摂動法は，SARFA [15] に基づいて feature attribution を行う手法である．まず SARFA の拡張である廣瀬ら [23] の説明単位の符号付きスコアを定義し，この定義に基づくスコアを，入力テンソルに対する摂動とそれに伴う Q 値の変化から計算するアルゴリズムを導入する．

4.1.1 符号付きスコア

説明単位 f を 0 でマスクしたテンソル $\mathbf{x}_t^{(-f)}$ を次式で定義する．

$$[\mathbf{x}_t^{(-f)}]_{c,j} = \begin{cases} 0 & ((c,j) \in S_f), \\ [\mathbf{x}_t]_{c,j} & (\text{otherwise}). \end{cases}$$

説明対象行動 $a \in \mathcal{A}_t$ に対する単位 f の**特異度**を

$$\Delta p^{\mathrm{pert}}_{t,f,a} = P(a \mid \mathbf{x}_t) - P(a \mid \mathbf{x}_t^{(-f)})$$

と定義する．特異度は，説明単位 f をマスクしたときに行動 a の相対評価値がどれだけ低下するかを表す．特異度が正の場合，説明単位 f は行動 a を支持し，負の場合は抑制すると解釈できる．

また，$b \in \mathcal{A}_t \setminus \{a\}$ に対して，行動 a を除く合法行動の相対評価分布を

$$R_{t,a}(b \mid \mathbf{x}) := \frac{\exp Q_b(\mathbf{x})}{\sum_{c\in\mathcal{A}_t\setminus\{a\}}\exp Q_c(\mathbf{x})}$$

で定義する．これを用いて，$D^{\mathrm{pert}}_{t,f,a}$ を

$$D^{\mathrm{pert}}_{t,f,a} := D_{\mathrm{KL}}\Big(R_{t,a}(\cdot \mid \mathbf{x}_t^{(-f)}) \,\Big\|\, R_{t,a}(\cdot \mid \mathbf{x}_t)\Big)$$

と定義する．ここで $D_{\mathrm{KL}}(P\|Q)$ は確率分布 P と Q の間の KL ダイバージェンスであり，確率分布 P と確率分布 Q の間の差を測る指標である．この値を用いて，説明単位 f の**関連度**を

$$K^{\mathrm{pert}}_{t,f,a} := \begin{cases} \frac{1}{1+D^{\mathrm{pert}}_{t,f,a}} & (|\mathcal{A}_t| \geq 2) \\ 1 & (|\mathcal{A}_t| < 2) \end{cases}$$

で定義する．$K^{\mathrm{pert}}_{t,f,a}$ は $(0,1]$ の値を取る．局面 t における合法行動集合 $\mathcal{A}_t$ の要素数が 2 未満の場合には，関連度は 1 と定義する．

$D^{\mathrm{pert}}_{t,f,a}$ は，説明単位 f をマスクすることで，局面 t

†2 ポンとカンがそれぞれ 1 次元，チーは対象牌を最小牌にするか，中央牌にするか，最大牌にするかに対応して 3 次元が使用されている．

における行動 a 以外の合法行動の相対評価分布がどれだけ変化するかを表す指標として解釈できる．もしこの値が大きい場合，説明単位 f は行動 a 以外の合法行動の相対評価分布を大きく変化させるため，単位 f は行動 a に特有の特徴量ではないと解釈できる．関連度は $D^{\mathrm{pert}}_{t,f,a}$ が大きいほど小さくなるため，関連度が大きい場合には，説明単位 f は行動 a に特有の特徴量であり有用性が高いと解釈できる．

特異度と関連度を用いて，説明単位 f の**符号付きスコア**を

$$S^{\mathrm{pert}}_{t,f,a} = \mathrm{sgn}(\Delta p^{\mathrm{pert}}_{t,f,a}) \frac{2K^{\mathrm{pert}}_{t,f,a} |\Delta p^{\mathrm{pert}}_{t,f,a}|}{K^{\mathrm{pert}}_{t,f,a} + |\Delta p^{\mathrm{pert}}_{t,f,a}|}$$

で定義する．符号付きスコアは，廣瀬ら [23] によって導入された，特異度の絶対値と関連度の調和平均に特異度の符号を付与した値である．$S^{\mathrm{pert}}_{t,f,a}$ が正の場合で絶対値が大きい場合は説明単位 f が行動 a を支持しかつ説明の有用性が高いと解釈し，負で絶対値が大きい場合は f が a を抑制しかつ説明の有用性が高いと解釈できる．また，その絶対値

$$I^{\mathrm{pert}}_{t,f,a} := |S^{\mathrm{pert}}_{t,f,a}|$$

を説明単位 f の**重要度**と定義する．

4.1.2 アルゴリズム

Algorithm 1 に入力特徴量テンソル $\mathbf{x}_t$，説明単位 f，モデル Q を受け取って，$S^{\mathrm{pert}}_{t,f,a}$ を計算するアルゴリズムを示す．以下で挙動を説明する．

- 1–4 行目では，元の入力特徴量テンソル $\mathbf{x}_t$ に対する各合法行動の相対評価分布を表すベクトル $\mathbf{p}$ を計算する．
- 5–8 行目では，説明単位 f をマスクした入力特徴量テンソル $\mathbf{x}_t^{(-f)}$ に対する各合法行動の相対評価分布を表すベクトル $\mathbf{p}'$ を計算する．
- 9–23 行目では，各合法行動 a に対して，符号付きスコア $S^{\mathrm{pert}}_{t,f,a}$ を計算する．まず，10 行目で特異度 Δp を計算する．11–12 行目では，合法行動が 2 個未満の場合の関連度 K を 1 とする．合法行動が 2 個以上の場合，14–17 行目で，説明対象行動 a を除く合法行動の相対評価分布 $\mathbf{r}_a$ と $\mathbf{r}'_a$ をマスク前後の入力特徴量テンソルについて計算する．その後，18 行目で KL ダイバージェンス D を計算し，19 行目で関連度 K を計算する．最後に，21 行目で特異度と関連度の調和平均 H を計算し，22 行目で符号付きスコア $S^{\mathrm{pert}}_{t,f,a}$ を計算する．
- 24 行目で，計算した全合法行動に対する符号付きスコア S^{pert} を返す．

Algorithm 1 摂動に基づく符号付きスコアの計算

Given: 入力特徴量テンソル $\mathbf{x}_t$，説明単位 f，モデル Q
Returns: $\left(S^{\mathrm{pert}}_{t,f,a}\right)_{a \in \mathcal{A}_t}$

▷ 元入力に対する基準分布を計算する
1: $\mathbf{q} \leftarrow Q(\mathbf{x}_t)$
2: **for all** $b \in \mathcal{A}_t$ **do**
3: $\quad \mathbf{p}[b] \leftarrow \dfrac{\exp(\mathbf{q}[b])}{\sum_{c \in \mathcal{A}_t} \exp(\mathbf{q}[c])}$
4: **end for**
▷ 説明単位をマスクした入力に対する分布を計算する
5: $\mathbf{q}' \leftarrow Q(\mathrm{mask}(\mathbf{x}_t, S_f))$
6: **for all** $b \in \mathcal{A}_t$ **do**
7: $\quad \mathbf{p}'[b] \leftarrow \dfrac{\exp(\mathbf{q}'[b])}{\sum_{c \in \mathcal{A}_t} \exp(\mathbf{q}'[c])}$
8: **end for**
▷ 各合法行動について特異度と関連度を統合する
9: **for all** $a \in \mathcal{A}_t$ **do**
10: $\quad \Delta p \leftarrow \mathbf{p}[a] - \mathbf{p}'[a]$
11: $\quad$ **if** $|\mathcal{A}_t| < 2$ **then**
12: $\quad\quad K \leftarrow 1$
13: $\quad$ **else**
14: $\quad\quad$ **for all** $b \in \mathcal{A}_t \setminus \{a\}$ **do**
15: $\quad\quad\quad \mathbf{r}_a[b] \leftarrow \dfrac{\exp(\mathbf{q}[b])}{\sum_{c \in \mathcal{A}_t \setminus \{a\}} \exp(\mathbf{q}[c])}$
16: $\quad\quad\quad \mathbf{r}'_a[b] \leftarrow \dfrac{\exp(\mathbf{q}'[b])}{\sum_{c \in \mathcal{A}_t \setminus \{a\}} \exp(\mathbf{q}'[c])}$
17: $\quad\quad$ **end for**
18: $\quad\quad D \leftarrow D_{\mathrm{KL}}(\mathbf{r}'_a[\cdot] \,\|\, \mathbf{r}_a[\cdot])$
19: $\quad\quad K \leftarrow 1/(1+D)$
20: $\quad$ **end if**
21: $\quad H \leftarrow 2K|\Delta p|/(K + |\Delta p|)$
22: $\quad S^{\mathrm{pert}}_{t,f,a} \leftarrow \mathrm{sgn}(\Delta p) H$
23: **end for**
▷ 全合法行動の符号付きスコアを返す
24: **return** S^{pert}

4.2 勾配法

Algorithm 1 に示した摂動法は，説明単位をマスクした入力特徴量テンソルに対する推論を必要とするため，説明単位の数が多い場合に計算コストが高い．本節では，摂動法の特異度と関連度の勾配による近似式を導出し，より低コストに符号付きスコアを計算する手法を提案する．式変形の詳細は appendix B に示す．

4.2.1 特異度の近似式の導出

$\boldsymbol{\delta}_{t,f}$ を

$$\boldsymbol{\delta}_{t,f} := \mathbf{x}_t - \mathbf{x}_t^{(-f)}$$

で定義する．$\boldsymbol{\delta}_{t,f}$ は，局面 t の入力特徴量テンソル $\mathbf{x}_t$ において，説明単位 f に属する要素を 0 に置換する入力特徴量の空間での方向を表すベクトルである．また，$\mathbf{x}_{t,f}(\varepsilon)$ を

$$\mathbf{x}_{t,f}(\varepsilon) := \mathbf{x}_t - \varepsilon\boldsymbol{\delta}_{t,f}, \qquad 0 \leq \varepsilon \leq 1$$

で定義する．$\mathbf{x}_{t,f}(\varepsilon)$ は入力特徴量テンソル $\mathbf{x}_t$ の説明単位 f に属する要素を，基準値 0 に向かって ε の強度で線形補間して得られる入力特徴量テンソルである．$\mathbf{x}_{t,f}(0) = \mathbf{x}_t$ であり，$\mathbf{x}_{t,f}(1) = \mathbf{x}_t^{(-f)}$ である．

行動 $b \in \mathcal{A}_t$ の Q 値のマスク方向に沿った方向微分 $d_{t,f,b}$ は

$$d_{t,f,b} := (\nabla_{\mathbf{x}} Q_b(\mathbf{x}_t))^\top \boldsymbol{\delta}_{t,f}$$

と定義できる．ただし，$\nabla_{\mathbf{x}} Q_b(\mathbf{x}_t)$ は

$$\nabla_{\mathbf{x}} Q_b(\mathbf{x}_t) := \left(\left. \frac{\partial Q_b(\mathbf{x})}{\partial [\mathbf{x}]_{c,j}} \right|_{\mathbf{x}=\mathbf{x}_t} \right)_{(c,j)\in\mathcal{I}}$$

で定義される，局面 t における行動 b の Q 値の入力特徴量テンソル $\mathbf{x}_t$ に関する勾配である．各合法行動の Q 値が $\varepsilon = 0$ の近傍で十分に滑らかであると仮定すると，Taylor の定理により

$$Q_b(\mathbf{x}_{t,f}(\varepsilon)) = Q_b(\mathbf{x}_t) - \varepsilon d_{t,f,b} + O(\varepsilon^2)$$

が成り立つ．

ここで，特異度 $\Delta p^{\mathrm{pert}}_{t,f,a}$ の定義により，

$$\Delta p^{\mathrm{pert}}_{t,f,a} = P(a \mid \mathbf{x}_{t,f}(0)) - P(a \mid \mathbf{x}_{t,f}(1))$$

が成り立つ．この特異度を近似するため，マスク強度 ε に対する特異度を

$$\Delta p^{\mathrm{pert}}_{t,f,a}(\varepsilon) = P(a \mid \mathbf{x}_{t,f}(0)) - P(a \mid \mathbf{x}_{t,f}(\varepsilon))$$

と定義する．この定義から，$\Delta p^{\mathrm{pert}}_{t,f,a}(1) = \Delta p^{\mathrm{pert}}_{t,f,a}$ が成り立つ．

$p_{t,b} = P(b \mid \mathbf{x}_t)$ と置く．softmax の偏微分は

$$\frac{\partial P(a \mid \mathbf{x}_t)}{\partial Q_b(\mathbf{x}_t)} = p_{t,a}(\delta_{ab} - p_{t,b})$$

である．ここで，δ_{ab} は Kronecker のデルタである．また，Q 値の Taylor 展開から，

$$\left. \frac{\mathrm{d}}{\mathrm{d}\varepsilon} Q_b(\mathbf{x}_{t,f}(\varepsilon)) \right|_{\varepsilon=0} = -d_{t,f,b}$$

が成り立つ．したがって，連鎖律により，次式が成り立つ．

$$\begin{aligned} &\left. \frac{\mathrm{d}}{\mathrm{d}\varepsilon} P(a \mid \mathbf{x}_{t,f}(\varepsilon)) \right|_{\varepsilon=0} \\ &= \sum_{b\in\mathcal{A}_t} p_{t,a}(\delta_{ab} - p_{t,b})(-d_{t,f,b}) \\ &= -p_{t,a}\left(d_{t,f,a} - \sum_{b\in\mathcal{A}_t} p_{t,b} d_{t,f,b} \right) \end{aligned}$$

ここで

$$\alpha_{t,f,a} := p_{t,a}\left(d_{t,f,a} - \sum_{b\in\mathcal{A}_t} p_{t,b} d_{t,f,b} \right)$$

と定義する．このとき，$P(a \mid \mathbf{x}_{t,f}(0)) = p_{t,a}$ から，

$$\Delta p^{\mathrm{pert}}_{t,f,a}(\varepsilon) = \varepsilon\alpha_{t,f,a} + O(\varepsilon^2)$$

が成り立つ．完全な 0 マスクに対応する $\varepsilon = 1$ にこの一次近似を適用し，勾配法の特異度を

$$\widetilde{\Delta p}_{t,f,a} := \alpha_{t,f,a}$$

と定義する．

4.2.2 関連度の近似式の導出

$|\mathcal{A}_t| \geq 2$ であると仮定する．マスク強度 ε に対する KL ダイバージェンスを

$$D^{\mathrm{pert}}_{t,f,a}(\varepsilon) = D_{\mathrm{KL}}(R_{t,a}(\cdot \mid \mathbf{x}_{t,f}(\varepsilon)) \,\|\, R_{t,a}(\cdot \mid \mathbf{x}_t))$$

と定義し，$b \in \mathcal{A}_t \setminus \{a\}$ に対して

$$r_{t,a,b} := R_{t,a}(b \mid \mathbf{x}_t) = \frac{p_{t,b}}{1 - p_{t,a}}$$

と定義する．また，

$$\mu_{t,f,a} = \sum_{b\in\mathcal{A}_t\setminus\{a\}} r_{t,a,b} d_{t,f,b}$$

と定義する．$\mu_{t,f,a}$ は，局面 t における行動 a を除く合法行動の Q 値の方向微分の平均である．

関連度の近似式を導出するにあたっては，説明対象となる行動を除いた行動の Q 値の相対評価分布の変化を測る KL ダイバージェンスの近似が重要となる．KL ダイバージェンスは，特異度と異なり，2 つの分布が一致する点で一次の項が 0 となるため，関連度の近似には二次の項を用いる必要がある．$\beta_{t,f,a}$ を

$$\beta_{t,f,a} = \frac{1}{2} \sum_{b\in\mathcal{A}_t\setminus\{a\}} r_{t,a,b}(d_{t,f,b} - \mu_{t,f,a})^2$$

と定義すると，$D^{\mathrm{pert}}_{t,f,a}(\varepsilon) = \varepsilon^2\beta_{t,f,a} + O(\varepsilon^3)$ が成り立つ．そこで勾配法における KL ダイバージェンスの近似値を

$$\widetilde{D}_{t,f,a} := \begin{cases} \beta_{t,f,a} & (|\mathcal{A}_t| \geq 2), \\ 0 & (|\mathcal{A}_t| < 2) \end{cases}$$

と定義する．$|\mathcal{A}_t| < 2$ の場合は，行動 a を除く相対評価分布を定義せず，関連度を 1 とするために $\widetilde{D}_{t,f,a} = 0$ と定める．完全な 0 マスクは $\varepsilon = 1$ に対

応するため，勾配法では $\alpha_{t,f,a}$ を特異度の近似値として用い，$|\mathcal{A}_t| \geq 2$ の場合には $\beta_{t,f,a}$ を KL ダイバージェンスの近似値として用いて，勾配法における関連度を

$$\widetilde{K}_{t,f,a} := \begin{cases} \dfrac{1}{1+\widetilde{D}_{t,f,a}} & (|\mathcal{A}_t| \geq 2), \\ 1 & (|\mathcal{A}_t| < 2) \end{cases}$$

と定義する．

4.2.3 勾配に基づく符号付きスコア

摂動法と同じ符号付き調和平均を用いて，説明単位 f の符号付きスコアを

$$S^{\mathrm{grad}}_{t,f,a} = \mathrm{sgn}(\widetilde{\Delta p}_{t,f,a}) \frac{2\widetilde{K}_{t,f,a}|\widetilde{\Delta p}_{t,f,a}|}{\widetilde{K}_{t,f,a} + |\widetilde{\Delta p}_{t,f,a}|}$$

と定義する．$S^{\mathrm{grad}}_{t,f,a}$ が正でかつ絶対値が大きい場合には説明単位 f が行動 a を支持しかつ説明が有用であると，負でかつ絶対値が大きい場合には f が行動 a を抑制しかつ説明が有用であると解釈できる．また，その絶対値

$$I^{\mathrm{grad}}_{t,f,a} := |S^{\mathrm{grad}}_{t,f,a}|$$

を説明単位 f の重要度と定義する．$|\mathcal{A}_t| = 1$ の場合には $P(a \mid \mathbf{x}_t) = 1$ であるため，$\widetilde{\Delta p}_{t,f,a} = 0$ および $S^{\mathrm{grad}}_{t,f,a} = 0$ となる．

4.2.4 $\beta_{t,f,a}$ と Fisher 情報行列との関係

$$\mathbf{d}_{t,f,-a} = (d_{t,f,b})_{b \in \mathcal{A}_t \setminus \{a\}},$$
$$\mathbf{r}_{t,a} = (r_{t,a,b})_{b \in \mathcal{A}_t \setminus \{a\}}$$

と置くと，$\beta_{t,f,a}$ は

$$\beta_{t,f,a} = \frac{1}{2}\mathbf{d}^{\top}_{t,f,-a}\big(\mathrm{Diag}(\mathbf{r}_{t,a}) - \mathbf{r}_{t,a}\mathbf{r}^{\top}_{t,a}\big)\mathbf{d}_{t,f,-a}$$

とも表せる．行列 $\mathrm{Diag}(\mathbf{r}_{t,a}) - \mathbf{r}_{t,a}\mathbf{r}^{\top}_{t,a}$ は，softmax のロジットを微小に変化させたときの確率分布の変化を表す Fisher 情報行列である．したがって，$\beta_{t,f,a}$ は，説明単位 f のマスクによって行動 a 以外の Q 値が変化したとき，それらの相対評価分布がどの程度変化するかを二次近似した量であると言える．行動 a 以外の合法行動の Q 値を並べたベクトルを

$$\mathbf{Q}_{t,-a}(\mathbf{x}) := \big(Q_b(\mathbf{x})\big)_{b \in \mathcal{A}_t \setminus \{a\}}$$

とし，その入力 Jacobian を

$$\mathbf{J}_{t,-a} := \left.\frac{\partial \mathbf{Q}_{t,-a}(\mathbf{x})}{\partial\, \mathrm{vec}(\mathbf{x})^{\top}}\right|_{\mathbf{x}=\mathbf{x}_t}$$

と定義する．ここで，vec は入力特徴量テンソルの要素を一列に並べて列ベクトルにする操作である．したがって，$\mathbf{J}_{t,-a}$ の各行は，行動 $b \in \mathcal{A}_t \setminus \{a\}$ の Q 値の入力勾配 $(\nabla_{\mathbf{x}} Q_b(\mathbf{x}_t))^{\top}$ に対応する．また，$\mathbf{F}_{t,a} := \mathrm{Diag}(\mathbf{r}_{t,a}) - \mathbf{r}_{t,a}\mathbf{r}^{\top}_{t,a}$ と置く．このとき，

$$\mathbf{d}_{t,f,-a} = \mathbf{J}_{t,-a}\,\mathrm{vec}(\boldsymbol{\delta}_{t,f}),$$
$$\beta_{t,f,a} = \frac{1}{2}\,\mathrm{vec}(\boldsymbol{\delta}_{t,f})^{\top}\mathbf{J}^{\top}_{t,-a}\mathbf{F}_{t,a}\mathbf{J}_{t,-a}\,\mathrm{vec}(\boldsymbol{\delta}_{t,f})$$

が成り立つ．$\mathbf{J}^{\top}_{t,-a}\mathbf{F}_{t,a}\mathbf{J}_{t,-a}$ は，行動 a 以外の相対評価分布の Fisher 計量を入力空間へ引き戻した計量である [13]．したがって，$\beta_{t,f,a}$ は，この引き戻し Fisher 計量によってマスク方向 $\boldsymbol{\delta}_{t,f}$ の長さの二乗を測った量の半分と解釈できる．

4.2.5 アルゴリズム

Algorithm 2 に，勾配に基づいて符号付きスコアを計算するアルゴリズムを示す．Algorithm 1 の摂動法では一つの説明単位 f を入力とするが，勾配法では各合法行動の Q 値の入力 Jacobian をすべての説明単位で共有できる．この共有を明示するため，Algorithm 2 は説明単位集合 $\mathcal{F}$ と対応する添字集合 $(S_f)_{f \in \mathcal{F}}$ を受け取り，すべての $f \in \mathcal{F}$ と $a \in \mathcal{A}_t$ に対するスコアをまとめて計算する．

Algorithm 2 の挙動を以下に説明する．

- 1–7 行目では，元の入力特徴量テンソル $\mathbf{x}_t$ に対する Q 値と，合法行動上の softmax 相対評価分布 $\mathbf{p}$ を計算する．$|\mathcal{A}_t| \geq 2$ の場合には，各説明対象行動 $a \in \mathcal{A}_t$ について，a を除いた合法行動上の相対評価分布 $\mathbf{r}_a$ も計算する．$\mathbf{p}$ と $\mathbf{r}_a$ は説明単位 f に依存しないため，すべての説明単位で共通に使用できる．
- 8–10 行目では，各合法行動 $b \in \mathcal{A}_t$ の Q 値について，入力特徴量テンソル $\mathbf{x}_t$ に関する勾配 $\mathbf{J}[b] = \nabla_{\mathbf{x}} Q_b(\mathbf{x}_t)$ を計算する．
- 11–16 行目では，各説明単位 f について，マスクへの方向ベクトル

  $$\boldsymbol{\delta}_{t,f} := \mathbf{x}_t - \mathrm{mask}(\mathbf{x}_t, S_f)$$

  を求める．その後，各合法行動 b について

  $$d[b] := \langle \nabla_{\mathbf{x}} Q_b(\mathbf{x}_t), \boldsymbol{\delta}_{t,f} \rangle = d_{t,f,b}$$

  を計算する．また，方向微分の softmax 加重平均

  $$\overline{d} := \sum_{b \in \mathcal{A}_t} \mathbf{p}[b] d[b]$$

  を計算する．
- 17–30 行目では，各説明対象行動 $a \in \mathcal{A}_t$ の符号

Algorithm 2 勾配に基づく符号付きスコアの計算

Given: 入力特徴量テンソル $\mathbf{x}_t$，説明単位集合 $\mathcal{F}$，対応する添字集合 $(S_f)_{f\in\mathcal{F}}$，モデル Q
Returns: $\left(S^{\mathrm{grad}}_{t,f,a}\right)_{f\in\mathcal{F},\, a\in\mathcal{A}_t}$

▷ 元入力に対する相対評価分布を計算する
1: $\mathbf{q} \leftarrow Q(\mathbf{x}_t)$
2: $\mathbf{p} \leftarrow \mathrm{softmax}\big((\mathbf{q}[b])_{b\in\mathcal{A}_t}\big)$
▷ 各説明対象行動を除いた相対評価分布を計算する
3: **if** $|\mathcal{A}_t| \geq 2$ **then**
4: **for all** $a \in \mathcal{A}_t$ **do**
5: $\mathbf{r}_a \leftarrow \mathrm{softmax}\big((\mathbf{q}[b])_{b\in\mathcal{A}_t\setminus\{a\}}\big)$
6: **end for**
7: **end if**
▷ 各合法行動の Q 値の入力 Jacobian を計算する
8: **for all** $b \in \mathcal{A}_t$ **do**
9: $\mathbf{J}[b] \leftarrow \nabla_{\mathbf{x}} Q_b(\mathbf{x}_t)$
10: **end for**
▷ 各説明単位と各合法行動のスコアを計算する
11: **for all** $f \in \mathcal{F}$ **do**
12: $\boldsymbol{\delta} \leftarrow \mathbf{x}_t - \mathrm{mask}(\mathbf{x}_t, S_f)$
13: **for all** $b \in \mathcal{A}_t$ **do**
14: $d[b] \leftarrow \langle \mathbf{J}[b], \boldsymbol{\delta} \rangle$
15: **end for**
16: $\overline{d} \leftarrow \sum_{b\in\mathcal{A}_t} \mathbf{p}[b] d[b]$
17: **for all** $a \in \mathcal{A}_t$ **do**
18: $\alpha \leftarrow \mathbf{p}[a]\big(d[a] - \overline{d}\big)$
19: **if** $|\mathcal{A}_t| < 2$ **then**
20: $\widetilde{D} \leftarrow 0$
21: $\widetilde{K} \leftarrow 1$
22: **else**
23: $\mu \leftarrow \sum_{b\in\mathcal{A}_t\setminus\{a\}} \mathbf{r}_a[b] d[b]$
24: $\beta \leftarrow \dfrac{1}{2} \sum_{b\in\mathcal{A}_t\setminus\{a\}} \mathbf{r}_a[b] \big(d[b] - \mu\big)^2$
25: $\widetilde{D} \leftarrow \beta$
26: $\widetilde{K} \leftarrow \dfrac{1}{1+\widetilde{D}}$
27: **end if**
28: $H \leftarrow \dfrac{2\widetilde{K}|\alpha|}{\widetilde{K} + |\alpha|}$
29: $S^{\mathrm{grad}}_{t,f,a} \leftarrow \mathrm{sgn}(\alpha) H$
30: **end for**
31: **end for**
32: **return** $\left(S^{\mathrm{grad}}_{t,f,a}\right)_{f\in\mathcal{F},\, a\in\mathcal{A}_t}$

付きスコアを計算する．18 行目で計算する

$$\alpha := \mathbf{p}[a]\big(d[a] - \overline{d}\big) = \alpha_{t,f,a}$$

を，特異度の近似値 $\widetilde{\Delta p}_{t,f,a}$ として用いる．

$|\mathcal{A}_t| \geq 2$ の場合，23–26 行目では，行動 a 以外の Q 値の方向微分の加重平均

$$\mu := \sum_{b\in\mathcal{A}_t\setminus\{a\}} \mathbf{r}_a[b] d[b] = \mu_{t,f,a}$$

と，その加重分散の半分

$$\beta := \frac{1}{2} \sum_{b\in\mathcal{A}_t\setminus\{a\}} \mathbf{r}_a[b] \big(d[b] - \mu\big)^2 = \beta_{t,f,a}$$

を計算する．β を KL ダイバージェンスの近似値 $\widetilde{D}_{t,f,a}$ とし，$\widetilde{K}_{t,f,a} = 1/(1+\widetilde{D}_{t,f,a})$ によって関連度へ変換する．$|\mathcal{A}_t| < 2$ の場合には，20–21 行目で $\widetilde{D}_{t,f,a} = 0$ および $\widetilde{K}_{t,f,a} = 1$ とする．最後に，28–29 行目で特異度の近似値の絶対値と関連度の調和平均 H を計算し，α の符号を付与して $S^{\mathrm{grad}}_{t,f,a}$ を得る．

- 32 行目で，すべての説明単位と合法行動に対して計算した符号付きスコアを返す．

入力 Jacobian $\mathbf{J}$ は説明単位 f に依存しないため，同じ局面に対する全説明単位を処理するとき，8–10 行目の入力 Jacobian を計算する処理は一度だけ実行すればよい．ただし，これは逆伝播を一回だけ実行すればよいことを意味しない．逆モード自動微分によって合法行動ごとの vector–Jacobian product を個別に計算する実装では，合法行動数に応じた回数の逆伝播を要する．これらをまとめて計算できる自動微分機構を用いる場合には，一括して $\mathbf{J}$ を得られる．一度 $\mathbf{J}$ を計算した後は，各説明単位について内積と有限個の和だけを計算し，説明単位ごとの追加の順伝播または逆伝播は行わない．したがって，モデルの微分計算回数は説明単位数 $|\mathcal{F}|$ に依存せず，入力 Jacobian の計算方法と合法行動数 $|\mathcal{A}_t|$ に依存する．

5 実装

前節で提案した摂動法と勾配法を Mortal に適用するプログラム，牌譜解析ツール，解析結果の可視化ツールを実装した．本実装は Mortal v4 のコミット `d5e80bff` をベースにしており，Mortal v4 用の学習済みモデルを使用した．

Algorithm 1 と Algorithm 2 の擬似コードと実装との違いは以下の通りである．

- 摂動法の実装において、Algorithm 1 においては単一の説明対象行動を受け取りスコアを計算するように定義しているが， 一つの説明単位をマスクしたときの Q 値は全ての説明対象行動で共有できるため，説明単位ごとに推論を一度だけ行い，その結果を各合法行動のスコア計算に再利用した．
- Softmax の計算では数値がオーバーフローしな

いように，合法行動の Q 値から最大値を減じた上で log-sum-exp を計算するように実装した．行動 a を除く相対評価分布を計算する上でも，$1 - P(a \mid \mathbf{x}_t)$ による除算ではなく，a を除いた Q 値へ直接 softmax を適用して求めるように実装した．

- Algorithm 2 では特徴量を完全にマスクすることに対応する $\varepsilon = 1$ での値を求めているが，一方点数，シャンテン数，残り牌数などの入力特徴量においては，これらの値を 0 に設定するマスクだけではなく，数値がどの程度減るごとにスコアがどの程度変化するかを確認することが必要になる場合がある．そのため，これらの特徴量においては ε の値を特徴量ごとに設定し，例えば点数が 1000 点減ることでスコアがどの程度変化するかなどを解釈できるようにした．

牌譜全体を解析するコマンドラインツールは，MJAI 形式の牌譜を受け取り，指定したプレイヤーについて牌譜中の行動を列挙する．各時点での合法行動の Q 値と Mortal が選択した行動に対する説明単位ごとのスコアを JSON 形式で出力する．

可視化用 Web アプリケーションは，MJAI 形式の牌譜と摂動法による解析結果の JSON ファイルを読み込み，ブラウザ上でイベント単位で牌譜を再生する．図 1 に画面のスクリーンショットを示す．解析結果に含まれる各打牌判断では，自家の手牌中の候補へマウスポインタを重ねると，その行動に対する各説明単位のスコアを局面表示へオーバーレイして表示する．自家の手牌，他家の河，ドラ表示牌，および各プレイヤーの点数に対応するスコアは，同じ局面と説明対象行動における最大絶対値で正規化し，正の値を赤，負の値を青，絶対値を色の濃さで表す．シャンテン数，フリテン，残り牌数，局情報，自家の河，待ち牌，他家の副露など，麻雀卓上の一つの要素へ対応付けられない入力特徴量のスコアは，特徴量名と数値を別のパネルに表示する．

6 実験

実装した解析ツールを用いて，摂動法と勾配法を定性的評価および計算時間の観点から比較する．本実験では，次の二つのリサーチクエスチョンを設定する．

RQ1 提案手法が出力する符号付きスコアから，Mortal の行動に対する入力特徴量の寄与を局面に即して解釈できるか．

RQ2 勾配法は摂動法による符号付きスコアをどの程度近似し，その計算時間をどの程度短縮するか．

RQ1 については，複数の局面における符号付きスコアを可視化し，Mortal の行動と入力特徴量との関係を定性的に分析する．この評価は，提案手法から得られる情報を具体例によって確認するものであり，説明に対する人間の納得度を定量的に評価するものではないことに注意されたい．RQ2 については，同一の局面，合法行動集合，および説明単位に対する両手法の符号付きスコアを対応付け，近似の一致度を評価する．さらに，同じ条件で両手法の計算時間を測定し，勾配法による計算時間の短縮率を評価する．

6.1 実験方法

評価対象として，M リーグ 2025–2026 シーズンの 2025 年 9 月 15 日および 16 日の試合を用いる．[†3]9 月 15 日の 11 局と 9 月 16 日の 14 局を合わせた 25 局について，4 プレイヤーすべての視点から解析し，合計 100 件（25 局 ×4 視点）のプレイヤー視点別牌譜を作成した．解析対象は，対象プレイヤーが自摸した時点に限り，チー，ポン，カン，ロンの判断時点は含めない．100 件のプレイヤー視点別牌譜から，自摸後の意思決定点を 1,050 点抽出した．このうち 44 点はリーチ後で合法行動が強制打牌の一つだけであり，合法行動上の softmax が常に 1 となって両手法の特異度と符号付きスコアが自明に 0 となるため，評価対象から除外した．さらに，先頭 2 点をウォームアップ実行とみなして計測結果から除外した．したがって，評価した意思決定点は $1{,}050 - 44 - 2 = 1{,}004$ 点で

†3 M リーグではシーズン終盤になると，セミファイナル進出を見据えたチームの条件，2 年連続でチームがセミファイナルに進出できない場合にチームの構成選手を 1 名以上入れ替えるレギュレーション等，各試合の勝敗以外の影響が打牌選択に影響を与える可能性が強く考えられるため，シーズン序盤の試合を評価対象とした．

表 2　実験で用いた説明単位

説明単位の種類	単位数
自家の手牌中の牌種	h_t
ドラ表示牌の牌種	d_t
河（他家 72 スロットと自家の河 3 成分）	75
自家および局全体の状態	15
副露および他家の状態	26
ドラ情報および派生特徴量	10
合計	$126 + h_t + d_t$

ある．

各意思決定点で Mortal の Q 値が最大となる合法行動を Mortal の選択行動とする．各評価局面に対して，摂動法と勾配法を用いて，Mortal の選択行動を含む全合法行動に対する各説明単位の符号付きスコアを計算する．両手法の比較においては，入力特徴量テンソル，合法行動集合，および説明単位を共通にした．

表 2 に，本実験で用いた説明単位の概要を示す．自家の手牌とドラ表示牌については，入力特徴量テンソル中の各牌種ごとに一つの説明単位とする．ここで，h_t と d_t は，それぞれ意思決定点 t で列挙された自家手牌とドラ表示牌の牌種数である．Mortal v4 の入力特徴量テンソルには，各家の河について，先頭 6 枚をスロットごとに符号化する表現，末尾 18 枚をスロットごとに符号化する表現，各牌種の捨て牌の新しさを符号化する時間減衰表現が含まれる．自家の河については，先頭 6 スロット全体，末尾 18 スロット全体，時間減衰表現全体をそれぞれ一つの説明単位とする．他家の河については，各家の 24 スロットを個別の説明単位とし，三家で 72 単位とする．各他家の時間減衰表現は一つの説明単位にまとめ，表では「副露および他家の状態」にカウントする．この説明単位は，捨て牌，手出し牌，リーチ宣言牌に対応するチャネルをまとめてマスクする．残る 126 単位は，待ち，フリテン，シャンテン数，点数，局情報，残り牌，副露，他家の状態，ドラ情報，Mortal v4 の手牌計算モジュールが計算した打牌候補ごとの期待値，聴牌確率，和了確率などの派生特徴量からなる．

説明単位数は行動時点ごとに異なり，最小 130 単位，中央値 139 単位，最大 173 単位であった．一つの説明単位をマスクしても入力が変化しない説明単位は，1 意思決定点当たり平均 70.4 単位存在した．これらの単位は摂動法のスコアは自明に 0 となるが，計算時間の比較では推論を省略せずにアルゴリズムを実行した．点数などの数値特徴量については，5 節で述べた単位量当たりの変化ではなく，0 へのマスクを近似する方法を用いた．

使用した学習済みモデルは Mortal 実装者の Discord チャネルで配布されていたものであり，SHA-1 は `15d5253f6ba0da3510a658a54b2997497c211209` である．

RQ1 では，解析した複数の行動時点のうち，Q 値が 1 位の行動と 2 位の行動との差が小さい局面を選び，摂動法で解析した結果を可視化し，Mortal の行動との関係を分析する．RQ2 では，各評価局面について摂動法と勾配法の符号付きスコアを比較し，さらに両手法が符号付きスコアを計算するために要した時間を測定する．

近似一致度については，各行動時点 t，説明単位 f，合法行動 a に対して，両手法のスコア，特異度，および関連度を対応付けて評価する．主要な集計対象は，マスクによる変化が非零の説明単位と全合法行動の組であり，Mortal の選択行動だけに限定した結果も併記する．保存済みの 1,004 点のデータに含まれる行動時点，説明単位，合法行動の組は計 1,602,728 組であり，そのうち変化が非零であった 823,185 組を主要集計に用いる．スコアと特異度については絶対誤差を，合法行動が 3 個以上の点では KL ダイバージェンスの絶対誤差を求める．また，摂動法のスコアの絶対値が 10^{-6} 以上である組に対する符号一致率を求める．各行動時点と行動の組について，説明単位間のの Spearman 順位相関係数と，上位説明単位の Overlap@1，Overlap@3，Overlap@5 を求める．絶対誤差は平均，中央値，四分位範囲，および 95 パーセンタイルを報告する．

計算時間の計測にあたっては，一つの行動時点における全合法行動の全説明単位に対する符号付きスコアを計算するまでの時間を，1 行動時点あたりの計算時間とする．計測時間には，入力特徴量テンソルを受け取った後の説明単位の列挙，Q 値の計算，摂動または

Jacobian の計算，および全合法行動のスコア集計の時間を含めるが，両手法に共通する牌譜の読み込み，牌譜の再生，観測テンソルの符号化，および結果のファイルへの書き出しの時間は含めない．共通処理である牌譜の再生と入力特徴量テンソルの符号化には，別に計測したところ 1 行動時点あたり平均 104.9 ms を要した．各行動時点では 1 回のウォームアップ後に 3 回を計測し，各処理成分について 3 回の中央値を求めてから，成分の中央値を合計してその行動時点の計算時間とした．時間の計測には `perf_counter` を用い，乱数シードを 0 に固定し，外れ値は除外しなかった．分散には標本分散を用いた．両手法について計算時間の平均，分散，および中央値を求める．平均計算時間に基づく勾配法の摂動法に対する高速化の倍率を，摂動法の平均計算時間を勾配法の平均計算時間で除した値とし，短縮率を 1−（勾配法の平均計算時間／摂動法の平均計算時間）として定義する．局面ごとの高速化倍率についても，平均，分散，中央値，および四分位範囲を求める．

計測にあたっては PyTorch のスレッド数を 1 に固定した．計測ホストの CPU は Intel Core i7-14700 であり，1 ソケットに 20 個の物理コアと 28 個の論理 CPU を備える．OS から利用可能な総メモリ量は 32,613,196 kB（約 31.1 GiB）である．OS は x86_64 版 Ubuntu 24.04.4 LTS であり, Linux カーネルのバージョンは 6.17.0-35-generic である．ソフトウェア環境は Python 3.12.3，PyTorch 2.11.0+cu130，NumPy 2.4.3 である．

6.2　結果

6.2.1　RQ1：定性的評価

事例 1：2025 年 9 月 15 日第 1 試合における南 2 局 0 本場 6 巡目．

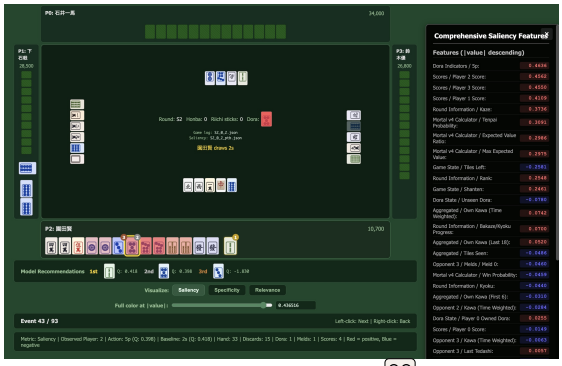
図 2 および図 3 に，それぞれの意思決定点における Mortal の第 1 候補と第 2 候補について，摂動法の saliency を候補間で共通の絶対尺度により可視化した画面を示す．

図 2 の手牌は であり， ， ， ， ， の五つの対子を持つ七対子一向聴であった．対子になっていない は石井一馬氏の河に 1 枚見えており， もドラ表示牌として 1 枚見えているため，いずれも自牌以外に同牌が 1 枚見えている状況である．石井一馬氏の河の に対応する説明単位に対するスコアは， 切りでは 0.099， 切りでは −0.087 であり， がすでに見えていることは 切りを支持した．一方，ドラ表示牌が であることに対応する説明単位に対するスコアは， 切りでは −0.273， 切りでは 0.464 であり， が 1 枚見えていることは 切りを支持した．七対子への進行が考えられる局面では孤立牌が対子になる残り枚数が打牌選択に関係するため，これらは牌譜から読み取れる両打牌候補の比較と整合する．

局開始時の得点は石井一馬氏が 34,000 点，下石戟氏が 28,500 点，園田賢氏が 10,700 点，鈴木優氏が 26,800 点であり，園田氏は 3 着の鈴木氏と 16,100 点差の 4 着であった．下石氏と鈴木氏という他家の得点に加え，園田氏自身の得点に対するスコアは，切りではそれぞれ −0.448，−0.488，−0.486 である一方， 切りでは 0.411，0.455，0.456 であり，いずれも絶対値が 0.4 を超え，重要度が高かった． 切りでこれらのスコアが負であるということは，もし他家の点数がより少なかったならば 切りの相対評価が上がるということである．従って，他家の点数が現状高いことが， 切りの相対評価を下げ， 切りの相対評価を上げることに寄与している．着順に対応する説明単位のスコアも 切りでは −0.277， 切りでは 0.255 であり，園田氏がこの時点で 4 着であるという情報が両打牌候補を異なる向きに評価している．

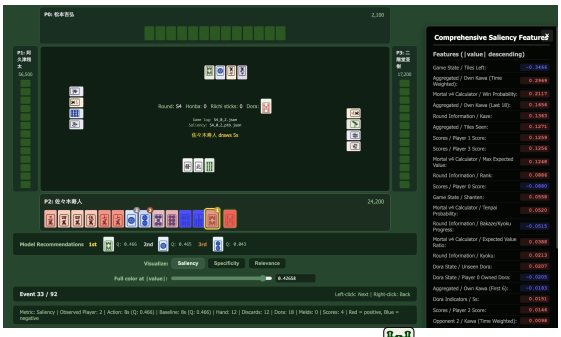
聴牌確率，最大期待値，期待値比の説明単位のスコアは， 切りではそれぞれ −0.344，−0.294，−0.236， 切りでは 0.309，0.297，0.299 であり，得点および着順と同じく 切りを支持した．これらがマイナスであるということは，これらの値がもし 0 に近い値であった場合に 切りの相対評価が上がることを意味するため， を手元に残すことが聴牌確率，最大期待値，期待値比に置いてポジティブであることと整合している．現状では が見えていないため， を

(a) Mortal の第 1 候補である切り

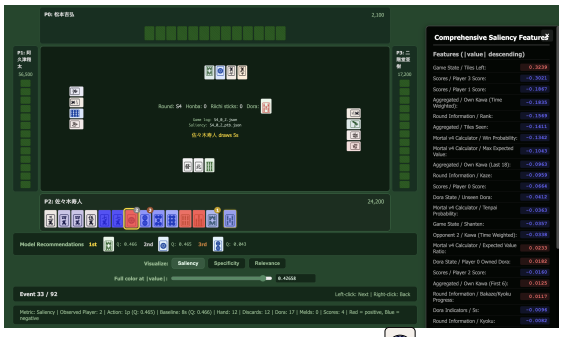


(b) Mortal の第 2 候補である切り

図 2　2025 年 9 月 15 日第 1 試合における南 2 局 0 本場 6 巡目，南家の園田賢氏がをツモった時点の摸動法による可視化画面．切りの Q 値は 0.418，切りの Q 値は 0.398 である．



(a) Mortal の第 1 候補である切り



(b) Mortal の第 2 候補である切り

図 3　2025 年 9 月 16 日の第 1 試合における南 4 局 0 本場の 4 巡目，北家の佐々木寿人氏がをツモった時点の摸動法による saliency 可視化画面．切りの Q 値は 0.466，切りの Q 値は 0.465 であり，両図は絶対値 0.42658 で同じ色の濃さとなる尺度を用いる．

残してを引いた場合の得点上昇がこの判断に影響を与えている可能性がある．

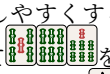
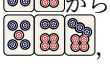
一方，残り牌数に対応する説明単位のスコアは切りで 0.346，切りで −0.258 であり，局の残り打牌可能枚数に関する情報は逆に切りを支持した．すなわち，局の残り打牌可能枚数がまだ多く序盤であるために，まだを残しても良いのではないかという判断が働いた可能性がある．

手牌にが存在するという説明単位に対するスコアは切りでは −0.378，切りでは 0.333 と反転した．定義通りに解釈すれば，が手牌に存在するためにを切った方が良い，という判断をしていることになるが，これに関しては，執筆時点では明確な解釈ができていない．

事例 2：2025 年 9 月 16 日第 1 試合における南 4 局 0 本場 4 巡目．

図 3 の手牌は　　であり，副露者もリーチ者もいない 4 巡目に，切りと切りの Q 値がそれぞれ 0.466 と 0.465 でほぼ等しかった．切りに対するそれぞれの手牌に対応する説明単位のスコアは，が −0.692，が −0.490，が 0.424，が 0.280 であった．

切りでは， が 0.433， が 0.422， が 0.390 である一方， は −0.224， は −0.163 であった． 切りでは， に対応する説明単位のスコアが正であるため， の存在が 切りを支持しており， を切っても が残ることと整合する．一方， と に対応する説明単位のスコアは負であるため，これらの牌の存在は 切りを抑制しており， を残せば を利用できることと整合する． 切りでは， と に対応する説明単位のスコアが正であるため， をそのまま保持する選択を支持している．これに対して， と に対応する説明単位のスコアは負であり，両端のどちらかを整理する選択との競合を示している．

切りは手牌から唯一の幺九牌を除いてタンヤオを目指しやすくする一方，切りは の塔子を維持して  を確定させる．後者では と からそれぞれ と を引くことで， ，  ， の三色同順への発展する可能性も残るため，この打牌選択の背景には，受入れ枚数だけでなく，タンヤオと三色同順を含む打点の比較もあると考えられる．この点について，和了確率，最大期待値，聴牌確率の説明単位のスコアは， 切りではそれぞれ 0.212，0.125，0.052 であったのに対し， 切りでは −0.134，−0.104，−0.036 であった．期待値比の説明単位のスコアも 切りでは 0.039， 切りでは 0.023 であり，Mortal 内部の局収支に関する推定値は，タンヤオを直接目指せる 切りよりも， と を残す 切りを支持した可能性がある．

この局面は南 4 局であり，北家の佐々木寿人氏は 24,200 点の 2 着で，首位の西家の阿久津翔太氏とは 32,300 点差，3 着とは 7,000 点差であり，供託はない状況であった．3 着者の得点の説明単位に対するスコアは 切りで 0.126， 切りで −0.302，順位の説明単位のスコアはそれぞれ 0.089 と −0.157 であり，着順および得点状況に関する情報も 切りを支持し， 切りを抑制した．残り牌数の説明単位のスコアは 切りで −0.347， 切りで 0.324 となり，自己の河についても打牌候補間で符号が反転した．Mortal は手牌の面子構成と役，和了確率および期待値，局の進行，着順状況を用いて僅差の打牌候補を評価していたと解釈できる．

以上の 2 局面では，候補ごとの符号を比較することで，見えている牌の枚数，手牌中の面子候補，得点状況のうち，どの入力が各打牌を支持または抑制したかを局面に即して読むことができる．ただし，各説明単位のスコアは一つの説明単位をマスクしたときの出力差であり，説明単位間の相互作用を含むことには留意が必要であると考えられる．よって RQ1 に対しては，**少なくとも選定した 2 局面では，摂動法が出力する符号付きスコアを打牌候補間で比較することにより，手牌構成，見えている牌，得点・着順などの入力が各打牌候補を支持または抑制する方向を，局面に即して解釈できた**と結論付ける．

なお，ここまでの定性的評価において，feature attribution スコアの上に与えた解釈は著者らの主観に基づくものであることに留意が必要である．また，今回は説明の妥当性に関する評価者間の一致や，説明を提示したときの意思決定への効果は評価していない．これらの評価は今後の課題である．

6.2.2　RQ2：摂動法に対する近似一致度

表 3 に，入力変化が非零の説明単位 823,185 組に対する摂動法と勾配法の各指標の絶対誤差を示す．全合法行動でも Mortal の選択行動に限定しても，平均が中央値を大きく上回っており，小さい誤差を持つ多数の行動時点と，少数の大きい誤差を持つ行動時点が存在する分布になっていることがわかる．

表 4 に，スコアの符号と重要度順位の一致度を示す．全合法行動では，摂動法のスコアの絶対値が 10^{-6} 以上である組に対するその符号一致率は 88.65%であった．説明単位全体の順位に対しては Spearman 順位相関係数は高い正の相関を示した．一方，Overlap@1 は 0.175 であり，最大のスコア絶対値を持つ説明単位が一致した局面と行動の組は全体の 17.5%にとどまった．Overlap@3 は 0.318，Overlap@5 は 0.414 であり，上位の説明単位ほど両手法の相違が大きいことがわかる．Mortal の選択行動に限定しても同じ傾向であった．

表 3　摂動法に対する勾配法の絶対誤差 (平均 / 中央値 [Q1, Q3] / P95)

対象	指標	平均	中央値	[Q1, Q3]	P95
全合法行動	スコア	0.00782	3.87×10^{-5}	$[1.97\times10^{-6}, 5.95\times10^{-4}]$	0.0336
	特異度	0.00484	1.94×10^{-5}	$[9.84\times10^{-7}, 2.99\times10^{-4}]$	0.0177
	KL div.	0.0307	3.30×10^{-5}	$[1.89\times10^{-6}, 0.00225]$	0.1119
Mortal の選択行動	スコア	0.0311	0.00128	$[0.000223, 0.0162]$	0.1837
	特異度	0.0202	0.000649	$[0.000112, 0.00850]$	0.1118
	KL div.	0.0471	5.19×10^{-5}	$[3.19\times10^{-6}, 0.00377]$	0.1887

表 4　摂動法と勾配法の符号および重要度順位の一致度

対象	符号一致率	Spearman 平均	Spearman 中央値	Overlap@1	@3 / @5
全合法行動	0.8865	0.8254	0.8476	0.1754	0.3182 / 0.4142
Mortal の選択行動	0.8852	0.8218	0.8455	0.1683	0.2971 / 0.4052

表 5　1 行動時点あたりの計算時間.

手法	平均 [ms]	分散 [ms^2]	中央値 [ms]
摂動法	1,656.6	18,860.3	1,578.3
勾配法	206.1	875.8	208.4

以上より，*RQ2* の近似一致度に対しては「勾配法は説明単位全体の重要度順位と符号を近似する用途には利用できるが，最大スコアを持つ説明単位またはスコアの数値については摂動法との誤差が大きい」と結論づけられる.

6.2.3　RQ2：計算時間の比較

表 5 に，1,004 行動時点における計算時間の集計結果を示す．平均計算時間に基づく高速化倍率は 8.04 倍であり，勾配法は摂動法から計算時間を 87.6%短縮した．行動時点ごとで比較したときの高速化倍率は，平均 8.21 倍，分散 1.907，中央値 8.04 倍，四分位範囲 7.16–8.69 倍であり，最小 6.22 倍，最大 17.57 倍であった．全 1,004 点で勾配法の計算時間が摂動法を下回った.

計算時間の内訳についても測定した．摂動法では総計算時間の 97.0%を説明単位ごとの順伝播が占め，勾配法では 94.2%を Jacobian 計算が占めた．図 4 に，摂動法の計算時間と説明単位数，および勾配法の計算時間と合法行動数の関係を示す．摂動法の計算時間は説明単位 1 単位当たり 11.28 ms の傾きで増加し，決定係数は $R^2 = 0.994$ であった．勾配法の計算時間は合法行動 1 個当たり 15.57 ms の傾きで増加し，決定係数は $R^2 = 0.996$ であった．図 4 では，各手法の計算時間に影響を与える要素のうち支配的なものとの関係を示すため，摂動法には説明単位数，勾配法には合法行動数を横軸として用いた．反対の組合せでは，摂動法の計算時間に対する合法行動数の決定係数は $R^2 = 0.105$，勾配法の計算時間に対する説明単位数の決定係数は $R^2 = 0.103$ であった．この結果は，摂動法が説明単位ごとに順伝播を行う一方，勾配法の主要な計算コストは合法行動に対する Jacobian の計算によるという観察を裏付ける.

以上より，*RQ2* の計算時間については，「勾配法は評価したすべての局面で摂動法より高速であり，平均では *1* 意思決定点当たりの計算時間をおよそ *8* 分の *1* へ短縮した」と結論づけられる.

7　結論

本論文では，深層強化学習によって実装された麻雀 AI である Mortal を対象として，その判断根拠を解釈するための情報として説明単位ごとの符号付き feature attribution score を計算する二つの手法を提案した．第一の手法である摂動法では，説明単位をマスクした前後の Q 値から，説明対象行動の相対評価の変化を表す特異度と，それ以外の合法行動の相対評価分布の変化を表す関連度を計算し，両者の符号付き調和平均をスコアとした．勾配法では，特異度を Q 値の入力勾配によって一次近似し，関連度に用いる KL ダイバージェンスを softmax 分布の Fisher 情報行列が誘導する二次形式によって近似した．これにより，説明単位ごとにモデルの推論を繰り返すことな

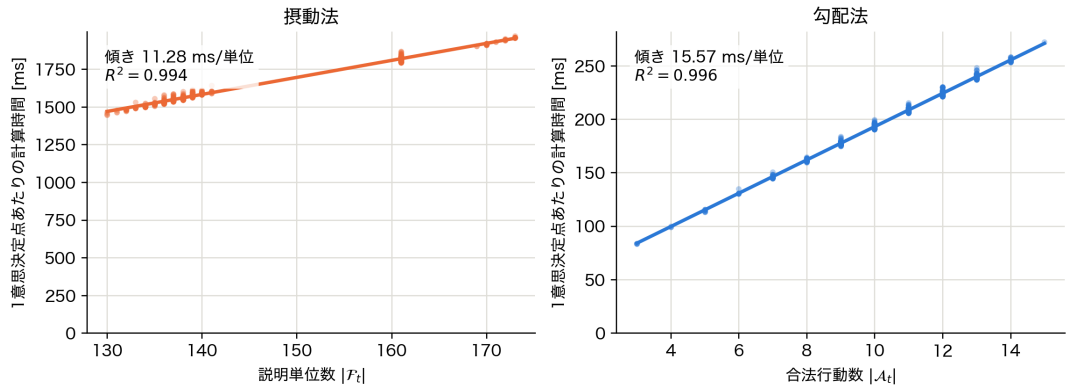


図 4　各手法に支配的なパラメータに対する計算時間の変化．

く，Jacobian を複数の説明単位で共有して SARFA 式のスコアを計算できる．

提案した二手法を Mortal v4 へ適用し，牌譜から合法行動の Q 値と説明単位ごとの符号付きスコアを出力する解析ツールを実装した．解析結果を牌譜の進行と対応付けて確認するため，符号付きスコアを局面表示へ重ねて表示する GUI アプリケーションも実装した．実験では，符号付きスコアから Mortal の行動に対する入力特徴量の寄与を局面に即して解釈できるかを定性的に分析し，摂動法に対する勾配法の近似一致度と計算時間を評価した．定性的評価では，選定した 2 局面について摂動法の符号付きスコアを打牌候補間で比較した．その結果，手牌構成，見えている牌，得点・着順などの入力が各打牌候補を支持または抑制する方向を，局面に即して解釈できた．ただし，一部の説明単位には明確な麻雀上の解釈を与えられず，各スコアには説明単位間の相互作用も含まれるため，個々のスコアを独立した判断理由とみなすべきではないことには留意が必要である．全合法行動に対する符号一致率は 88.65%，説明単位の重要度順位に対する Spearman 順位相関係数は平均 0.825 であったが，Overlap@1 は 0.175 であった．この結果から，勾配法は説明単位全体の符号と重要度順位を近似する用途には利用できるが，最大スコアを持つ説明単位またはスコアの数値を摂動法と同一視できない．計算時間の評価では，1,004 行動時点のすべてで勾配法が摂動法より高速であり，平均計算時間を 1,656.6 ms から 206.1 ms へ短縮した．これは平均計算時間に基づいて 8.04 倍の高速化，すなわち 87.6%の短縮に相当する．

本実験の定性的分析は選定した事例に基づくものであり，説明の妥当性に関する評価者間一致度や，説明を提示したときの意思決定への効果を明らかにするものではない．特に，説明を解釈する者の雀力によっては，解釈の妥当性が変動する可能性がある．また，摂動法のスコアは説明単位を 0 でマスクする方法に基づいており，勾配法のスコアは元の入力近傍における局所的な線形近似に基づいている．前者は持ち点数やシャンテン数など，入力特徴量が数値である場合に問題となり，後者は元の入力近傍での線形近似が不十分な場合に問題となる．今後の課題は，より近似精度の良い勾配法の設計，説明の有用性の人間の評価者による検証やより多くの情報が得られる入力特徴量のマスク方法の検討が挙げられる．また，より精密な定性評価を行うために，著者らの雀力を向上させることも重要な課題である．

謝辞．

本研究は JSPS 科研費 JP25H01113 の助成を受けて実施された。

参 考 文 献

[1] Ancona, M., Ceolini, E., Oztireli, C., and Gross, M.: Towards Better Understanding of Gradient-Based Attribution Methods for Deep Neural Networks, *International Conference on Learning Representations*, 2018.

[2] Covert, I., Lundberg, S. M., and Lee, S.-I.: Explaining by Removing: A Unified Framework for Model Explanation, *Journal of Machine Learning Research*, Vol. 22, No. 209(2021), pp. 1–90.

[3] DWANGO Co., Ltd.: 麻雀 AI「NAGA」の技術解説, DWANGO Media Village, January 2019. `https://dmv.nico/ja/articles/mahjong_ai_naga/`（2026 年 8 月 4 日参照）.

[4] Equim-chan: Mortal: A Japanese Mahjong AI, GitHub repository, commit `d5e80bff`, 2025. `https://github.com/Equim-chan/Mortal`（2026 年 8 月 4 日参照）.

[5] Fang, Z., 李吉屹, and 吉岡真治: 説明可能な麻雀 AI に向けた 2 段階 Dual-LoRA を用いたニューロシンボリック指示学習, 言語処理学会第 *32* 回年次大会発表論文集, 2026, pp. 2828–2833.

[6] Fong, R. C. and Vedaldi, A.: Interpretable Explanations of Black Boxes by Meaningful Perturbation, *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 3449–3457.

[7] Gat, I., Calderon, N., Reichart, R., and Hazan, T.: A Functional Information Perspective on Model Interpretation, *Proceedings of the 39th International Conference on Machine Learning*, Proceedings of Machine Learning Research, Vol. 162, 2022, pp. 7266–7278.

[8] Greydanus, S., Koul, A., Dodge, J., and Fern, A.: Visualizing and Understanding Atari Agents, *Proceedings of the 35th International Conference on Machine Learning*, Proceedings of Machine Learning Research, Vol. 80, 2018, pp. 1792–1801.

[9] Hatakeyama, Y., Sakuma, H., Konishi, Y., and Suenaga, K.: Visualizing Color-Wise Saliency of Black-Box Image Classification Models, *Computer Vision – ACCV 2020*, Lecture Notes in Computer Science, Vol. 12624, 2021, pp. 189–205.

[10] Kikuchi, A., Uchida, K., Waga, M., and Suenaga, K.: BOREx: Bayesian-Optimization-Based Refinement of Saliency Map for Image- and Video-Classification Models, *Computer Vision – ACCV 2022*, Lecture Notes in Computer Science, Vol. 13847, 2023, pp. 274–290.

[11] Li, J., Koyamada, S., Ye, Q., Liu, G., Wang, C., Yang, R., Zhao, L., Qin, T., Liu, T.-Y., and Hon, H.-W.: Suphx: Mastering Mahjong with Deep Reinforcement Learning, arXiv:2003.13590, 2020.

[12] Lundberg, S. M. and Lee, S.-I.: A Unified Approach to Interpreting Model Predictions, *Advances in Neural Information Processing Systems 30*, 2017, pp. 4765–4774.

[13] Martino, G. and Tschiatschek, S.: FRInGe: Distribution-Space Integrated Gradients with Fisher–Rao Geometry, arXiv:2605.06404, 2026.

[14] Miyato, T., Maeda, S., Koyama, M., and Ishii, S.: Virtual Adversarial Training: A Regularization Method for Supervised and Semi-Supervised Learning, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 41, No. 8(2019), pp. 1979–1993.

[15] Puri, N., Verma, S., Gupta, P., Kayastha, D., Deshmukh, S., Krishnamurthy, B., and Singh, S.: Explain Your Move: Understanding Agent Actions Using Specific and Relevant Feature Attribution, *International Conference on Learning Representations*, 2020.

[16] Ribeiro, M. T., Singh, S., and Guestrin, C.: Why Should I Trust You?: Explaining the Predictions of Any Classifier, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 1135–1144.

[17] Sengupta, S. and Anastasio, M. A.: Input-Dependent Fisher Information for Local Sensitivity Analysis of Medical Image Classifiers, arXiv:2606.16362, 2026.

[18] Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T., Simonyan, K., and Hassabis, D.: A General Reinforcement Learning Algorithm that Masters Chess, Shogi, and Go through Self-Play, *Science*, Vol. 362, No. 6419(2018), pp. 1140–1144.

[19] Simonyan, K., Vedaldi, A., and Zisserman, A.: Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps, ICLR Workshop, arXiv:1312.6034, 2014.

[20] 髙野 智椰, 宮本友樹, and 片上大輔: 深層学習麻雀 AI の打牌選択への寄与度可視化による対話型牌譜検討支援システムの提案, 第 *41* 回ファジィシステムシンポジウム講演論文集, 2025, pp. 334–339.

[21] Wang, Z., Schaul, T., Hessel, M., van Hasselt, H., Lanctot, M., and de Freitas, N.: Dueling Network Architectures for Deep Reinforcement Learning, *Proceedings of the 33rd International Conference on Machine Learning*, Proceedings of Machine Learning Research, Vol. 48, 2016, pp. 1995–2003.

[22] Zhao, C., Fletcher, P. T., Yu, M., Peng, Y., Zhang, G., and Shen, C.: The Adversarial Attack and Detection under the Fisher Information Metric, *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33, No. 01(2019), pp. 5869–5876.

[23] 廣瀬雄一, 和賀正樹, 末永幸平: Feature Attribution を用いた dlshogi の指し手の解釈可能性向上手法, 情報処理学会研究報告ゲーム情報学（*GI*）, *Vol. 2023-GI-49*, *No. 13*, 2023, pp. 1–8.

[24] 廣瀬雄一, 和賀正樹, 末永幸平: 解釈が容易な特徴量を用いた Feature Attribution に基づく将棋 AI の指し手の解釈可能性向上手法, 日本ソフトウェア科学会第 *40* 回大会講演論文集, 2023. 講演番号 43-R-S.

[25] 岡本龍馬, 永井秀利, 中村貞吾: CNN を用いた麻雀の打牌選択における判断根拠の説明, *2023* 年度電気・情報関係学会九州支部連合大会講演論文集, 2023.

[26] 岡本龍馬, 永井秀利, 中村貞吾: Kernel SHAP を用いた麻雀の打牌選択モデルにおける判断根拠の説明, *2024* 年度電気・情報関係学会九州支部連合大会講演論文集, 2024.

A　Mortal v4 の入力特徴量

本付録は，Mortal のコミット `d5e80bff` の実装が用いる入力特徴量とエンコーダについて詳述する [4].

A.1　テンソルの構造

入力は大きさ $(1012, 34)$ の実数値テンソルであり，チャネル番号と牌種列番号は 0 から始まる．牌種列 0–8 は ，9–17 は ，18–26 は ，27–33 は に対応する． ， ， は通常の ， ， と牌種列を共有し，赤牌フラグ用チャネルで区別する．牌種に対応しない数値またはフラグに関するチャネルでは，対応するチャネルの 34 列すべてを同じ値で埋める．プレイヤーに関する特徴量は，自家を $p = 0$ とする席順で表現する．

A.2　手牌と局の基本情報

0–6：自家の手牌．

チャネル 0–3 は，牌種ごとの手牌枚数を表す．牌種 t を n 枚持つ場合，チャネル 0 から $n-1$ までの列 t の値を 1 にする．例えば， を 3 枚持つ場合， 列のチャネル 0–3 は $[1, 1, 1, 0]$ となる．チャネル 4, 5, 6 は，それぞれ ， ， を手牌に持つ場合に 34 列すべてを 1 にする．

7–14：各家の点数．

プレイヤー p の点数が s_p であるとき，チャネル $7+2p$ の 34 列すべてを $\mathrm{clamp}(s_p, 0, 100000)/100000$ で埋める．ここで $\mathrm{clamp}(x, a, b) = \max(a, \min(x, b))$ である．また，チャネル $8+2p$ の 34 列すべてを $\mathrm{clamp}(s_p, 0, 30000)/30000$ で埋める．

15–18：自家の現在順位．

自家の順位を 4 チャネルの one-hot 表現で符号化する．1 位ならチャネル 15，2 位なら 16，3 位なら 17，4 位なら 18 の 34 列すべてを 1 にする．

19–22：現在の局番号．

場風とは独立に，現在が第 1 局から第 4 局のいずれであるかを 4 チャネルの one-hot 表現で符号化する．第 1 局ならチャネル 19，第 2 局なら 20，第 3 局なら 21，第 4 局なら 22 の 34 列すべてを 1 にする．

23–24：本場数と供託数．

本場数を h，供託数を k とする．チャネル 23 を $\min(h, 10)/10$，チャネル 24 を $\min(k, 10)/10$ で埋める．

25–26：場風と自風．

チャネル 25 は場風，チャネル 26 は自風を表す． ， ， ， のうち該当する牌種列を 1 にする．

27：局進行度．

局番号 $q \in \{0, 1, 2, 3\}$ と，場風の からの差 w を用いる．チャネル 27 の 34 列すべてを

$$\frac{4\min(w, 1) + q}{7}$$

で埋める．例えば，南 2 局は 5/7 となる．$\min(w, 1)$ を用いるため，西入以降は対応する南場と同じ値に

表 6　入力特徴量の全チャネル割当て

チャネル	数	内容
0–6	7	自家の手牌
7–14	8	各家の点数
15–18	4	自家の現在順位
19–22	4	現在の局番号
23–24	2	本場数と供託数
25–26	2	場風と自風
27	1	局進行度
28–34	7	ドラ表示牌
35–131	97	自家の河
132–716	585	他家 3 人の河
717	1	残りツモ可能牌数
718–721	4	各家のドラ保有数
722	1	未観測ドラ枚数
723–750	28	各家の捨て牌集計
751–830	80	各家の副露
831–834	4	各家の暗槓
835	1	観測済み牌数
836–844	9	他家の直近手出し牌
845–853	9	他家の立直宣言牌
854–859	6	他家の立直宣言と立直成立
860	1	自家の待ち牌
861	1	自家のフリテン状態
862–868	7	自家のシャンテン数
869	1	自家の立直成立
870	1	槓牌種選択フェーズ
871–873	3	鳴きまたはロンの対象牌
874–878	5	打牌候補とシャンテン変化
879–888	10	行動可能フラグ
889–1011	123	v4 専用の単独手牌計算
合計	1012	

表 7　自家の河スロットの構成

オフセット	内容
0	直前の槓に関係する牌の牌種列
1	捨て牌の牌種列
2	捨て牌が赤牌なら 34 列すべてを 1
3	捨て牌が通常ドラなら 34 列すべてを 1

なる．

28–34：ドラ表示牌．

チャネル 28–31 は，牌種ごとの表示枚数を手牌と同じ形式のフラグで表す．チャネル 32–34 は，それぞれ , , が表示牌に含まれる場合に 34 列すべてを 1 にする．

A.3　河の時系列情報

各家について河の先頭 6 スロットを古い順に保持し，末尾 18 スロットを新しい順に保持する．河が短い場合，同じ捨て牌が先頭 6 スロットと末尾 18 スロットの双方に現れることになる．また，Mortal は，ポンまたは大明槓による手番のずれを揃えるため，牌を持たない空スロットを河へ挿入する場合がある．

35–131：自家の河．

チャネル 35–58 は先頭 6 スロットの捨て牌，59–130 は末尾 18 スロットの捨て牌である．131 は各牌種がどの程度最近捨てられたかを指数減衰値で表す時間減衰特徴量であり，全プレイヤーの河の最大長を L，対象牌が捨てられたスロットを r としたときに，その牌種列に $\exp\{-0.2(L-1-r)\}$ が格納される．各河スロットは表 7 の 4 チャネルで表す．

先頭 6 スロット捨て牌の番号を $j \in \{0,\dots,5\}$ とすると，スロット j はチャネル $35+4j$ から $38+4j$ に対応する．末尾 18 スロットの番号を $j \in \{0,\dots,17\}$ とすると，スロット j はチャネル $59+4j$ から $62+4j$ に対応する．

132–716：他家の河．

$p \in \{1,2,3\}$ の捨て牌はチャネル $b_p = 132+195(p-1)$ から格納される．チャネル b_p から b_p+47 は先頭 6 スロット，b_p+48 から b_p+191 は末尾 18 スロット，b_p+192 から b_p+194 は時間減衰特徴量である．

表 8　他家の河スロットの構成

オフセット	内容
0–1	直前のチーまたはポンで手牌から消費した 2 牌
2	直前の槓に関係する牌
3	捨て牌
4	捨て牌が赤牌
5	捨て牌が通常ドラ
6	捨て牌が手出し
7	捨て牌が立直宣言牌

各河スロットは表 8 の 8 チャネルで表す．

先頭 6 スロットの捨て牌については，スロットの番号を $j \in \{0,\dots,5\}$ とすると，スロット j はチャネル b_p+8j から b_p+8j+7 に対応する．末尾 18 スロットの番号を $j \in \{0,\dots,17\}$ とすると，スロット j はチャネル $b_p+48+8j$ から $b_p+55+8j$ に対応する．時間減衰特徴量の 3 チャネルは，捨て牌，手出し牌，立直宣言牌をそれぞれ表す．

全プレイヤーの河スロット数の最大値を L，対象捨て牌のスロット位置を r としたときに，時間減衰特徴量として対象牌の牌種列に

$$\exp\{-0.2(L-1-r)\}$$

を格納する．同じ牌種が複数回現れた場合は，後の捨て牌による値が上書きされる．

A.4　牌数，ドラ，捨て牌集計，副露

717：残りツモ可能牌数．

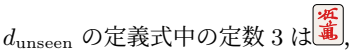
山からツモ可能な残り牌数を n とし，チャネル 717 の 34 列すべてを $n/69$ で埋める．この値は残り牌数の合計であり，山に残る牌種別の内訳ではない．

718–721：各家のドラ保有数．

プレイヤー p のドラ保有数を d_p とし，チャネル $718+p$ の 34 列すべてを $\min(d_p,12)/12$ で埋める．赤牌は 1 枚のドラとして数える．自家は手牌，副露，暗槓を含む正確な保有数を用いる．他家は副露と暗槓から観測できるドラだけを数える．

722：未観測ドラ枚数．

ドラ表示牌数を m，観測済みドラ枚数を d_{seen} とする．未観測ドラ枚数 $d_{\text{unseen}} := 4m+3-d_{\text{seen}}$ を求め，チャネル 722 を $\min(d_{\text{unseen}},23)/23$ で埋める．

d_{unseen} の定義式中の定数 3 は  を表す．

723–750：各家の捨て牌集計．

プレイヤー p にチャネル $723+7p$ から $729+7p$ を割り当て，牌種ごとの捨て牌枚数を集計する．先頭 4 チャネルは捨てた牌の牌種ごとの累積枚数を表す．末尾 3 チャネルは を捨てたかを表す．後に他家から鳴かれた捨て牌も集計に含む．

751–830：各家の副露．

プレイヤー p にチャネル $751+20p$ から $770+20p$ を割り当てて，副露の内容を記録する．1 人あたり最大 4 面子を副露順に記録し，1 面子を 5 チャネルで表す．先頭 4 チャネルは面子内の牌種ごとの累積枚数であり，5 番目は赤牌を含む場合に 34 列すべてを 1 にする．

831–834：各家の暗槓．

プレイヤー p の暗槓をチャネル $831+p$ に記録する．暗槓した牌種の列を 1 にし，複数の暗槓があれば一つのチャネル内で複数列を 1 にする．

835：観測済み牌数．

牌種 t の観測済み牌数を s_t とし，チャネル 835 の列 t に $s_t/4$ を格納する．自家の手牌，過去のツモ，捨て牌，表示牌，公開面子などから観測した牌を含む．赤牌は対応する通常の 5 と同じ牌種として数える．

A.5　直近牌，立直，手牌状態

836–844：他家の直近手出し牌．

プレイヤー 1，2，3 に，それぞれチャネル 836–838，839–841，842–844 を割り当てて，直近の手出し牌を格納する．各 3 チャネルはそれぞれ牌種，赤牌フラグ，通常ドラフラグを表す．直近の捨て牌がツモ切りでも，それ以前の最後の手出し牌を保持する．

845–853：他家の立直宣言牌．

プレイヤー 1，2，3 に，それぞれチャネル 845–847，848–850，851–853 を割り当てる．各 3 チャネルはそれぞれ牌種，赤牌フラグ，通常ドラフラグを表す．

854–859：他家の立直宣言と立直成立．

チャネル 854–856 はプレイヤー 1–3 が立直を宣言したかどうかを表すフラグである．このフラグは宣言イベントの瞬間だけでなく，局が終了するまで保持される．チャネル 857–859 は各プレイヤーの立直が成立した後に 1 となる．

表 9　応答対象牌のチャネル

チャネル	内容
871	対象牌の牌種列
872	対象牌が赤牌なら 34 列すべてを 1
873	対象牌が通常ドラなら 34 列すべてを 1

860：自家の待ち牌．

内部状態が保持する待ち牌の牌種列を 1 にする．役の有無を考慮しないため，このフラグだけでは直ちに和了を宣言できることを保証しない．

861：自家のフリテン状態．

自家がフリテンなら，チャネル 861 の 34 列すべてを 1 にする．

862–868：自家のシャンテン数．

シャンテン数 $n \in \{0,\dots,6\}$ を 7 チャネルの one-hot 表現で符号化し，チャネル $862+n$ の 34 列すべてを 1 にする．最大シャンテン数は 6 なので，7 チャネルを用いる．

869：自家の立直成立．

自家の立直が成立していれば，チャネル 869 の 34 列すべてを 1 にする．

870：槓牌種選択フェーズ．

槓を行うかどうかの選択後に，槓をする牌種を選ぶステップにいる場合，チャネル 870 の 34 列すべてを 1 にする．

A.6　行動候補

871–873：鳴きまたはロンの対象牌．

応答対象となる他家の捨て牌を表 9 の 3 チャネルで表す．

874–878：打牌候補とシャンテン変化．

打牌の候補とそれぞれがシャンテン数をどう変化させるかの情報をこれらのチャネルに格納する．表 10 に 5 チャネルの意味を示す．チャネル 875 と 876 は互いに排他的である．チャネル 877 は内部シャンテン数が 1 以下の場合に計算され，局面によって 875 または 876 の部分集合となる．

表 10　打牌候補とシャンテン変化のチャネル

チャネル	内容
874	合法な打牌候補
875	シャンテン数を維持する打牌
876	シャンテン数を 1 減らす打牌
877	打牌後に無条件で聴牌となる候補
878	自家が立直を宣言済み

879–888：行動可能フラグ．

チャネル 879 は立直，880–882 は 3 種類のチー，883 はポン，884 は大明槓が可能であることを表す．チャネル 885 と 886 は，それぞれ暗槓候補と加槓候補の牌種列を 1 にする．チャネル 887 はツモ和了またはロン和了，888 は九種九牌による途中流局が可能であることを表す．

A.7　手牌計算情報のブロック

チャネル 889 以降には，他家からの和了や放銃を考慮しない手牌計算器が求めた有効牌，聴牌確率，和了確率，期待値の情報を格納する．打牌可能な局面では，各牌種列がその牌種を捨てた場合の値に対応する．

889–890：最大期待値．

手牌計算器が計算した期待値のうちで，現時点での期待値が最大となる打牌候補の期待値を $E_{\max}$ とする．チャネル 889 を $\mathrm{clamp}(E_{\max}, 0, 100000)/100000$ で埋める．チャネル 890 を $\mathrm{clamp}(E_{\max}, 0, 30000)/30000$ で埋める．自家に今後のツモが残っていない場合や、手牌がすでに和了形である場合など、手牌計算ができない場合は，裏ドラを仮定しない最小ツモ和了点を代わりに用い，チャネル 891–1011 を 0 にする．

891–958：打牌候補ごとの有効牌．

打牌候補の牌種を d，有効牌種を r とする．向聴戻しを伴わない候補では，チャネル $891 + d$ の列 r を 1 にする．向聴戻しを伴う候補では，チャネル $925 + d$ の列 r を 1 にする．この範囲は有効牌の種類を表し，残り枚数は格納しない．

959–960：受入最大候補または現在手牌の有効牌．

打牌可能な局面では，向聴戻しを伴わない候補を優先し，有効牌枚数が最大となる打牌候補の牌種列をチャネル 959 で 1 にする．同数の場合は実装上の打牌優先順位を用いる．打牌できない局面では，現在の手牌の有効牌種列をチャネル 960 で 1 にする．

961–1011：将来 17 巡分の確率と期待値．

現在から数えた巡目 $i \in \{0, \dots, 16\}$ ついて，その巡目の聴牌確率をチャネル $961 + i$，和了確率をチャネル $978 + i$ に格納する．期待値を E_i とすると，$E_{\max} \geq 1$ の場合にチャネル $995 + i$ へ $\min(E_i / E_{\max}, 1)$ を格納する．$E_{\max} < 1$ の場合は期待値チャネルを 0 にする．シャンテン数が 4 以上で確率を計算しない場合や，先頭候補の聴牌確率が 0 の場合はチャネル 961–1011 を 0 にする．

B　勾配法の近似式の詳細な導出

4.2 節で用いた特異度の一次係数 $\alpha_{t,f,a}$ と，KL ダイバージェンスの二次係数 $\beta_{t,f,a}$ を以下で導出する．以下では，局面 t，説明単位 f，説明対象行動 $a \in \mathcal{A}_t$ を固定する．

B.1　前提と記号

マスク強度 ε に対する入力特徴量テンソルを

$$\mathbf{x}_{t,f}(\varepsilon) = \mathbf{x}_t - \varepsilon \boldsymbol{\delta}_{t,f}, \qquad \boldsymbol{\delta}_{t,f} = \mathbf{x}_t - \mathbf{x}_t^{(-f)} \tag{1}$$

とする．ここで，$0 \leq \varepsilon \leq 1$ であり，$\varepsilon = 0$ は元の入力，$\varepsilon = 1$ は説明単位 f を 0 でマスクした入力に対応する．

本導出では，合法行動集合 $\mathcal{A}_t$ をマスク経路上で固定する．また，各 $b \in \mathcal{A}_t$ について，$Q_b(\mathbf{x}_{t,f}(\varepsilon))$ が $\varepsilon = 0$ の近傍で 3 回連続微分可能であると仮定する．ReLU を用いるモデルでは，$\varepsilon = 0$ の近傍でマスク経路が活性化領域の境界を横切らない場合に，この仮定の下での局所展開を適用できる．以下の $\varepsilon = 0$ における微分と Taylor 展開は，$\varepsilon \geq 0$ からの右側微分と右側 Taylor 展開として用いる．

記号を簡単にするため，各合法行動 b について

$$q_b(\varepsilon) := Q_b(\mathbf{x}_{t,f}(\varepsilon)), \qquad p_b(\varepsilon) := \frac{\exp q_b(\varepsilon)}{\displaystyle\sum_{c \in \mathcal{A}_t} \exp q_c(\varepsilon)} \tag{2}$$

と置く．本文の記号との対応は $p_b(0) = p_{t,b}$ である．

行動 a 以外の合法行動集合を

$$\mathcal{B}_{t,a} := \mathcal{A}_t \setminus \{a\}$$

と置く．$|\mathcal{A}_t| \geq 2$ の場合には，$b \in \mathcal{B}_{t,a}$ に対して

$$r_b(\varepsilon) := \frac{\exp q_b(\varepsilon)}{\sum_{c \in \mathcal{B}_{t,a}} \exp q_c(\varepsilon)}, \qquad r_b(0) = r_{t,a,b} \tag{3}$$

と定義する．有限の Q 値に softmax を適用しているため，$p_b(\varepsilon)$ と $r_b(\varepsilon)$ は正であり，それぞれの定義域上で総和が 1 となる．

B. 2　マスク方向に沿った Q 値の変化

説明単位 f を 0 でマスクするため，$\boldsymbol{\delta}_{t,f}$ の各成分は

$$[\boldsymbol{\delta}_{t,f}]_{c,j} = \begin{cases} [\mathbf{x}_t]_{c,j} & ((c,j) \in S_f), \\ 0 & ((c,j) \notin S_f) \end{cases} \tag{4}$$

となる．したがって，Q 値のマスク方向に沿った方向微分 $d_{t,f,b}$ は

$$\begin{aligned} d_{t,f,b} &= (\nabla_{\mathbf{x}} Q_b(\mathbf{x}_t))^\top \boldsymbol{\delta}_{t,f} \\ &= \sum_{(c,j) \in \mathcal{I}} \left.\frac{\partial Q_b(\mathbf{x})}{\partial [\mathbf{x}]_{c,j}}\right|_{\mathbf{x}=\mathbf{x}_t} [\boldsymbol{\delta}_{t,f}]_{c,j} \\ &= \sum_{(c,j) \in S_f} [\mathbf{x}_t]_{c,j} \left.\frac{\partial Q_b(\mathbf{x})}{\partial [\mathbf{x}]_{c,j}}\right|_{\mathbf{x}=\mathbf{x}_t} \end{aligned} \tag{5}$$

と展開できる．

式 (1) の各成分は

$$[\mathbf{x}_{t,f}(\varepsilon)]_{c,j} = [\mathbf{x}_t]_{c,j} - \varepsilon [\boldsymbol{\delta}_{t,f}]_{c,j}$$

である．したがって，各成分を ε で微分すると，

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon} [\mathbf{x}_{t,f}(\varepsilon)]_{c,j} = -[\boldsymbol{\delta}_{t,f}]_{c,j}$$

となる．

$q_b(\varepsilon) = Q_b(\mathbf{x}_{t,f}(\varepsilon))$ では，ε は $\mathbf{x}_{t,f}(\varepsilon)$ の各成分を通じて Q_b に影響する．

この合成関数に多変数関数の連鎖律を適用すると，

$$\begin{aligned} q_b'(\varepsilon) &= \sum_{(c,j) \in \mathcal{I}} \left.\frac{\partial Q_b(\mathbf{x})}{\partial [\mathbf{x}]_{c,j}}\right|_{\mathbf{x}=\mathbf{x}_{t,f}(\varepsilon)} \\ &\quad \times \frac{\mathrm{d}}{\mathrm{d}\varepsilon} [\mathbf{x}_{t,f}(\varepsilon)]_{c,j} \\ &= -\sum_{(c,j) \in \mathcal{I}} \left.\frac{\partial Q_b(\mathbf{x})}{\partial [\mathbf{x}]_{c,j}}\right|_{\mathbf{x}=\mathbf{x}_{t,f}(\varepsilon)} [\boldsymbol{\delta}_{t,f}]_{c,j} \end{aligned}$$

を得る．第 1 式は，各入力成分の変化率に，その成分に対する Q_b の偏微分を掛けて足し合わせたものである． $\varepsilon = 0$ では $\mathbf{x}_{t,f}(0) = \mathbf{x}_t$ であるため，

$$\begin{aligned} q_b'(0) &= -\sum_{(c,j) \in \mathcal{I}} \left.\frac{\partial Q_b(\mathbf{x})}{\partial [\mathbf{x}]_{c,j}}\right|_{\mathbf{x}=\mathbf{x}_t} [\boldsymbol{\delta}_{t,f}]_{c,j} \\ &= -(\nabla_{\mathbf{x}} Q_b(\mathbf{x}_t))^\top \boldsymbol{\delta}_{t,f} \\ &= -d_{t,f,b} \end{aligned} \tag{6}$$

が成り立つ．1 つ目の等号では，$\mathbf{x}_{t,f}(0) = \mathbf{x}_t$ を用いて，偏微分を評価する点を $\mathbf{x}_t$ に置き換えた．2 つ目の等号は，勾配ベクトルと $\boldsymbol{\delta}_{t,f}$ の内積の成分表示である．最後の等号では，式 (5) で定義した $d_{t,f,b}$ を用いた．負号が付くのは，$\boldsymbol{\delta}_{t,f}$ を元の入力から減算する向きに ε を増やすためである．Taylor の定理と式 (6) から，

$$q_b(\varepsilon) = q_b(0) - \varepsilon d_{t,f,b} + O(\varepsilon^2) \tag{7}$$

を得る．

B. 3　特異度の一次近似

まず，softmax の偏微分を確認する．$Z = \sum_{c \in \mathcal{A}_t} \exp q_c$ と置くと，$p_a = \exp q_a / Z$ である．Kronecker のデルタを δ_{ab} とすると，商の微分から

$$\begin{aligned} \frac{\partial p_a}{\partial q_b} &= \frac{\delta_{ab} \exp q_a Z - \exp q_a \exp q_b}{Z^2} \\ &= \frac{\exp q_a}{Z} \left(\delta_{ab} - \frac{\exp q_b}{Z}\right) \\ &= p_a(\delta_{ab} - p_b) \end{aligned} \tag{8}$$

が成り立つ．

次に，マスク経路に沿って $p_a(\varepsilon)$ を微分する．連鎖律，式 (6) および式 (8) から，

$$\begin{aligned} p_a'(0) &= \sum_{b \in \mathcal{A}_t} \left.\frac{\partial p_a}{\partial q_b}\right|_{\varepsilon=0} q_b'(0) \\ &= \sum_{b \in \mathcal{A}_t} p_{t,a}(\delta_{ab} - p_{t,b})(-d_{t,f,b}) \\ &= -p_{t,a} \sum_{b \in \mathcal{A}_t} \delta_{ab} d_{t,f,b} + p_{t,a} \sum_{b \in \mathcal{A}_t} p_{t,b} d_{t,f,b} \\ &= -p_{t,a} \left(d_{t,f,a} - \sum_{b \in \mathcal{A}_t} p_{t,b} d_{t,f,b} \right) \end{aligned} \tag{9}$$

を得る．第 3 式から第 4 式では，$\sum_b \delta_{ab} d_{t,f,b} = d_{t,f,a}$

を用いた．

マスク強度 ε に対する特異度を

$$\Delta p^{\mathrm{pert}}_{t,f,a}(\varepsilon) := p_a(0) - p_a(\varepsilon) \tag{10}$$

と置く．この定義から，$\Delta p^{\mathrm{pert}}_{t,f,a}(0) = 0$ であり，

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\Delta p^{\mathrm{pert}}_{t,f,a}(\varepsilon)\Big|_{\varepsilon=0} &= -p'_a(0)\\ &= p_{t,a}\left(d_{t,f,a} - \sum_{b\in\mathcal{A}_t} p_{t,b} d_{t,f,b}\right)\\ &=: \alpha_{t,f,a}\end{aligned} \tag{11}$$

となる．したがって，$\varepsilon = 0$ のまわりで特異度を Taylor 展開すると，

$$\Delta p^{\mathrm{pert}}_{t,f,a}(\varepsilon) = \varepsilon\alpha_{t,f,a} + O(\varepsilon^2) \tag{12}$$

を得る．

B.4 KL ダイバージェンスの二次近似

$|\mathcal{A}_t| \geq 2$ とし，行動 a を除いた相対評価分布の KL ダイバージェンスを

$$D^{\mathrm{pert}}_{t,f,a}(\varepsilon) := \sum_{b\in\mathcal{B}_{t,a}} r_b(\varepsilon)\log\frac{r_b(\varepsilon)}{r_b(0)} \tag{13}$$

と書く．式 (3) は $\mathcal{B}_{t,a}$ 上の softmax であるため，式 (8) と同じ計算から

$$\frac{\partial r_b}{\partial q_c} = r_b(\delta_{bc} - r_c)$$

が成り立つ．よって，$r_b(\varepsilon)$ の $\varepsilon = 0$ における微分は

$$\begin{aligned}r'_b(0) &= \sum_{c\in\mathcal{B}_{t,a}} r_{t,a,b}(\delta_{bc} - r_{t,a,c})(-d_{t,f,c})\\ &= -r_{t,a,b}d_{t,f,b} + r_{t,a,b}\sum_{c\in\mathcal{B}_{t,a}} r_{t,a,c}d_{t,f,c}\\ &= -r_{t,a,b}(d_{t,f,b} - \mu_{t,f,a})\end{aligned} \tag{14}$$

となる．ここで，

$$\mu_{t,f,a} := \sum_{c\in\mathcal{B}_{t,a}} r_{t,a,c}d_{t,f,c} \tag{15}$$

と置いた．

KL ダイバージェンスの一階微分を計算する．$\sum_b r_b(\varepsilon) = 1$ を微分すると，

$$\sum_{b\in\mathcal{B}_{t,a}} r'_b(\varepsilon) = 0 \tag{16}$$

が成り立つ．一方，$r_b(\varepsilon) > 0$ であるため，

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\log r_b(\varepsilon) = \frac{r'_b(\varepsilon)}{r_b(\varepsilon)}$$

である．積の微分，対数の微分，式 (16) を順に用いると，

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\varepsilon}D^{\mathrm{pert}}_{t,f,a}(\varepsilon) &= \sum_{b\in\mathcal{B}_{t,a}} r'_b(\varepsilon)\log\frac{r_b(\varepsilon)}{r_b(0)}\\ &\quad + \sum_{b\in\mathcal{B}_{t,a}} r_b(\varepsilon)\frac{r'_b(\varepsilon)}{r_b(\varepsilon)}\\ &= \sum_{b\in\mathcal{B}_{t,a}} r'_b(\varepsilon)\log\frac{r_b(\varepsilon)}{r_b(0)}\end{aligned} \tag{17}$$

を得る．$\varepsilon = 0$ では対数の中身が 1 となるため，

$$D^{\mathrm{pert}}_{t,f,a}(0) = 0, \qquad \frac{\mathrm{d}}{\mathrm{d}\varepsilon}D^{\mathrm{pert}}_{t,f,a}(\varepsilon)\Big|_{\varepsilon=0} = 0 \tag{18}$$

が成り立つ．KL ダイバージェンスの一次の項が消えるのは，二つの分布が $\varepsilon = 0$ で一致するためである．

次に，式 (17) をもう一度微分する．

$$\begin{aligned}\frac{\mathrm{d}^2}{\mathrm{d}\varepsilon^2}D^{\mathrm{pert}}_{t,f,a}(\varepsilon) &= \sum_{b\in\mathcal{B}_{t,a}} r''_b(\varepsilon)\log\frac{r_b(\varepsilon)}{r_b(0)}\\ &\quad + \sum_{b\in\mathcal{B}_{t,a}} \frac{(r'_b(\varepsilon))^2}{r_b(\varepsilon)}\end{aligned} \tag{19}$$

となる．$\varepsilon = 0$ では第 1 項が 0 となるので，式 (14) を第 2 項へ代入すると，

$$\begin{aligned}\frac{\mathrm{d}^2}{\mathrm{d}\varepsilon^2}D^{\mathrm{pert}}_{t,f,a}(\varepsilon)\Big|_{\varepsilon=0} &= \sum_{b\in\mathcal{B}_{t,a}} \frac{(r'_b(0))^2}{r_{t,a,b}}\\ &= \sum_{b\in\mathcal{B}_{t,a}} r_{t,a,b}(d_{t,f,b} - \mu_{t,f,a})^2\end{aligned} \tag{20}$$

を得る．式 (18) および式 (20) を Taylor の定理へ代入すると，

$$\begin{aligned}D^{\mathrm{pert}}_{t,f,a}(\varepsilon) &= D^{\mathrm{pert}}_{t,f,a}(0) + \varepsilon\left.\frac{\mathrm{d}}{\mathrm{d}\varepsilon}D^{\mathrm{pert}}_{t,f,a}(\varepsilon)\right|_{\varepsilon=0}\\ &\quad + \frac{\varepsilon^2}{2}\left.\frac{\mathrm{d}^2}{\mathrm{d}\varepsilon^2}D^{\mathrm{pert}}_{t,f,a}(\varepsilon)\right|_{\varepsilon=0} + O(\varepsilon^3)\\ &= \frac{\varepsilon^2}{2}\sum_{b\in\mathcal{B}_{t,a}} r_{t,a,b}(d_{t,f,b} - \mu_{t,f,a})^2 + O(\varepsilon^3)\\ &=: \varepsilon^2\beta_{t,f,a} + O(\varepsilon^3)\end{aligned} \tag{21}$$

となる．したがって，二次係数は

$$\beta_{t,f,a} = \frac{1}{2}\sum_{b\in\mathcal{B}_{t,a}} r_{t,a,b}(d_{t,f,b} - \mu_{t,f,a})^2 \tag{22}$$

である．

完全な 0 マスクに対応する $\varepsilon = 1$ へ式 (21) の二次近似を適用し，勾配法における KL ダイバージェンスの近似値を

$$\widetilde{D}_{t,f,a} := \begin{cases}\beta_{t,f,a} & (|\mathcal{A}_t| \geq 2),\\ 0 & (|\mathcal{A}_t| < 2)\end{cases}$$

と定義する．$|\mathcal{A}_t| < 2$ の場合には行動 a を除いた相対評価分布を定義できないため，KL ダイバージェンスの近似値を 0 とする．

この近似値を関連度へ変換すると，

$$\widetilde{K}_{t,f,a} := \begin{cases} \dfrac{1}{1+\widetilde{D}_{t,f,a}} & (|\mathcal{A}_t| \geq 2), \\ 1 & (|\mathcal{A}_t| < 2) \end{cases}$$

となる．

B.5 係数の表示とその解釈

B.5.1 説明対象の行動以外の行動群による一次の係数

$|\mathcal{A}_t| \geq 2$ の場合には，$\alpha_{t,f,a}$ を説明対象の行動以外の行動との比較と解釈できる形に書き換えられる．$p_{t,b} = (1-p_{t,a})r_{t,a,b}$ を用いると，

$$\begin{aligned} \sum_{b\in\mathcal{A}_t} p_{t,b}d_{t,f,b} &= p_{t,a}d_{t,f,a} + \sum_{b\in\mathcal{B}_{t,a}} p_{t,b}d_{t,f,b} \\ &= p_{t,a}d_{t,f,a} \\ &\quad + (1-p_{t,a}) \sum_{b\in\mathcal{B}_{t,a}} r_{t,a,b}d_{t,f,b} \end{aligned} \tag{23}$$

が成り立つ．式 (23) を式 (11) へ代入すると，

$$\begin{aligned} \alpha_{t,f,a} &= p_{t,a}(1-p_{t,a})d_{t,f,a} \\ &\quad - p_{t,a}(1-p_{t,a}) \sum_{b\in\mathcal{B}_{t,a}} r_{t,a,b}d_{t,f,b} \\ &= p_{t,a}(1-p_{t,a}) \left(d_{t,f,a} - \sum_{b\in\mathcal{B}_{t,a}} r_{t,a,b}d_{t,f,b} \right) \end{aligned} \tag{24}$$

を得る．この式では，行動 a の Q 値の変化と競合行動群の Q 値の変化の加重平均との差に，$p_{t,a}(1-p_{t,a})$ を掛けている．特に，すべての合法行動について $d_{t,f,b}$ が同じ場合には，この差が 0 となるため $\alpha_{t,f,a} = 0$ である．

B.5.2 二次の係数の重み付き分散表示

式 (22) の総和は，$r_{t,a,b}$ を重みとする $d_{t,f,b}$ の分散である．実際，$\sum_b r_{t,a,b} = 1$ と式 (15) から，

$$\begin{aligned} &\sum_{b\in\mathcal{B}_{t,a}} r_{t,a,b}(d_{t,f,b} - \mu_{t,f,a})^2 \\ &\quad = \sum_{b\in\mathcal{B}_{t,a}} r_{t,a,b}d_{t,f,b}^2 - 2\mu_{t,f,a} \sum_{b\in\mathcal{B}_{t,a}} r_{t,a,b}d_{t,f,b} \\ &\qquad + \mu_{t,f,a}^2 \sum_{b\in\mathcal{B}_{t,a}} r_{t,a,b} \\ &\quad = \sum_{b\in\mathcal{B}_{t,a}} r_{t,a,b}d_{t,f,b}^2 - \mu_{t,f,a}^2 \end{aligned} \tag{25}$$

が成り立つ．このため，$\beta_{t,f,a}$ は重み付き分散の半分であり，常に非負である．

B.5.3 Fisher 情報行列との関係

次に，この分散と Fisher 情報行列の関係を導く．$\mathcal{B}_{t,a}$ 上のカテゴリ変数 Y が $\Pr(Y=b) = r_{t,a,b}$ に従うとする．softmax のロジットベクトルを $\mathbf{q}$，$Y=b$ を表す one-hot ベクトルを $\mathbf{e}_b$ とすると，score は

$$\begin{aligned} \frac{\partial}{\partial q_c}\log r_{t,a,b} &= \frac{1}{r_{t,a,b}}\frac{\partial r_{t,a,b}}{\partial q_c} \\ &= \delta_{bc} - r_{t,a,c} \end{aligned} \tag{26}$$

である．したがって，ベクトル表示では

$$\nabla_{\mathbf{q}} \log \Pr(Y=b) = \mathbf{e}_b - \mathbf{r}_{t,a} \tag{27}$$

となる．

Fisher 情報行列は score の外積の期待値であるため，

$$\begin{aligned} \mathbf{F}_{t,a} &:= \mathbb{E}_{Y\sim\mathbf{r}_{t,a}} \left[(\mathbf{e}_Y - \mathbf{r}_{t,a})(\mathbf{e}_Y - \mathbf{r}_{t,a})^\top \right] \\ &= \mathbb{E}[\mathbf{e}_Y\mathbf{e}_Y^\top] - \mathbb{E}[\mathbf{e}_Y]\mathbf{r}_{t,a}^\top \\ &\quad - \mathbf{r}_{t,a}\mathbb{E}[\mathbf{e}_Y]^\top + \mathbf{r}_{t,a}\mathbf{r}_{t,a}^\top \\ &= \mathrm{Diag}(\mathbf{r}_{t,a}) - \mathbf{r}_{t,a}\mathbf{r}_{t,a}^\top \end{aligned} \tag{28}$$

を得る．最後の変形では，$\mathbb{E}[\mathbf{e}_Y] = \mathbf{r}_{t,a}$ と $\mathbb{E}[\mathbf{e}_Y\mathbf{e}_Y^\top] = \mathrm{Diag}(\mathbf{r}_{t,a})$ を用いた．

$\mathbf{d}_{t,f,-a} = (d_{t,f,b})_{b\in\mathcal{B}_{t,a}}$ と置く．式 (28) が誘導する二次形式は

$$\begin{aligned} &\mathbf{d}_{t,f,-a}^\top \mathbf{F}_{t,a}\mathbf{d}_{t,f,-a} \\ &\quad = \sum_{b\in\mathcal{B}_{t,a}} r_{t,a,b}d_{t,f,b}^2 - \left(\sum_{b\in\mathcal{B}_{t,a}} r_{t,a,b}d_{t,f,b} \right)^2 \\ &\quad = \sum_{b\in\mathcal{B}_{t,a}} r_{t,a,b}(d_{t,f,b} - \mu_{t,f,a})^2 \end{aligned} \tag{29}$$

となる．Q 値のロジットベクトルの $\varepsilon = 0$ における変化方向は $-\mathbf{d}_{t,f,-a}$ であるが，二次形式では符号が相殺されるため，$(-\mathbf{d})^\top\mathbf{F}_{t,a}(-\mathbf{d}) = \mathbf{d}^\top\mathbf{F}_{t,a}\mathbf{d}$ である．したがって，式 (22) および式 (29) から

$$\beta_{t,f,a} = \frac{1}{2}\mathbf{d}_{t,f,-a}^\top \mathbf{F}_{t,a}\mathbf{d}_{t,f,-a} \tag{30}$$

を得る．

任意のベクトル $\mathbf{v}$ に対して，$\mathbf{v}^\top\mathbf{F}_{t,a}\mathbf{v}$ は $\mathbf{v}$ の各成分の重み付き分散であるため，$\mathbf{F}_{t,a}$ は半正定値である．また，$\mathbf{1}$ を全成分が 1 のベクトルとすると，

$$\mathbf{F}_{t,a}\mathbf{1} = \mathbf{r}_{t,a} - \mathbf{r}_{t,a}(\mathbf{r}_{t,a}^\top\mathbf{1}) = \mathbf{0}$$

が成り立つ．したがって，行動 a 以外のすべての Q 値が同じだけ変化する場合には $\beta_{t,f,a} = 0$ となる．これは，全ロジットへの共通加算によって softmax 分布が変化しないことに対応する．

B. 6　マスクへの適用と境界事例

式 (12) および式 (21) は，$\varepsilon = 0$ の近傍で成り立つ局所展開である．勾配法では，完全に入力特徴量テンソルの要素を 0 でマスクすることに対応する $\varepsilon = 1$ へ特異度の一次近似を適用し，

$$\widetilde{\Delta p}_{t,f,a} := \alpha_{t,f,a}$$

と定義する．同様に, Appendix B. 4 で定義した $\widetilde{D}_{t,f,a}$ と $\widetilde{K}_{t,f,a}$ も, KL ダイバージェンスの二次近似を $\varepsilon = 1$ へ適用して得られる．$\widetilde{\Delta p}_{t,f,a}$ と $\widetilde{D}_{t,f,a}$ は完全に 0 でマスクすることによる変化と正確に成り立つ恒等式ではない．マスク経路上の非線形性や ReLU の活性化領域の変化が大きいほど，高次の項を無視することによる近似誤差が生じ得る．

勾配法の符号付きスコアは，近似した特異度と関連度を摂動法と同じ符号付き調和平均へ代入して

$$S^{\mathrm{grad}}_{t,f,a} = \mathrm{sgn}(\widetilde{\Delta p}_{t,f,a}) \frac{2\widetilde{K}_{t,f,a}|\widetilde{\Delta p}_{t,f,a}|}{\widetilde{K}_{t,f,a} + |\widetilde{\Delta p}_{t,f,a}|}$$

と定義する．この定義は，符号関数と絶対値を含むスコア全体を Taylor 展開したものではなく，特異度と KL ダイバージェンスをそれぞれ近似してから SARFA の合成式へ代入したものであることに注意されたい．

最後に，合法行動数に関するエッジケースを確認する．$|\mathcal{A}_t| = 0$ の場合には説明対象行動 a が存在しないため，スコアの定義域外である．$|\mathcal{A}_t| = 1$ の場合には，唯一の合法行動 a について $p_a(\varepsilon) = 1$ であるため，特異度は厳密に 0 である．この場合，式 (11) から $\alpha_{t,f,a} = 0$ となり，定義により $\widetilde{D}_{t,f,a} = 0$, $\widetilde{K}_{t,f,a} = 1$，$S^{\mathrm{grad}}_{t,f,a} = 0$ となる．$|\mathcal{A}_t| = 2$ の場合には，$\mathcal{B}_{t,a}$ が一つの行動だけからなるため，$r_{t,a,b} = 1$ かつ $\mu_{t,f,a} = d_{t,f,b}$ である．したがって，式 (22) から $\beta_{t,f,a} = 0$ となり，行動 a 以外の相対評価分布はマスクの前後で変化しない．