

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7593473号
(P7593473)

(45)発行日 令和6年12月3日(2024.12.3)

(24)登録日 令和6年11月25日(2024.11.25)

(51)国際特許分類 F I
G 0 6 N 20/00 (2019.01) G 0 6 N 20/00

請求項の数 8 (全16頁)

(21)出願番号	特願2023-503257(P2023-503257)	(73)特許権者	000004237 日本電気株式会社 東京都港区芝五丁目7番1号
(86)(22)出願日	令和3年3月3日(2021.3.3)	(74)代理人	100107331 弁理士 中村 聡延
(86)国際出願番号	PCT/JP2021/008149	(74)代理人	100104765 弁理士 江上 達夫
(87)国際公開番号	WO2022/185444	(74)代理人	100131015 弁理士 三輪 浩誉
(87)国際公開日	令和4年9月9日(2022.9.9)	(72)発明者	坂井 智哉 東京都港区芝五丁目7番1号 日本電気 株式会社内
審査請求日	令和5年8月16日(2023.8.16)	審査官	坂庭 剛史

最終頁に続く

(54)【発明の名称】 互換性評価装置、互換性評価方法、及び、プログラム

(57)【特許請求の範囲】

【請求項1】

評価データに対する第1の予測器及び第2の予測器の出力を取得する取得手段と、
前記第1の予測器の出力と前記第2の予測器の出力との関係を示す複数の関係式の組み合わせにより規定される一般化後方互換性指標を決定する指標決定手段と、
前記第1の予測器の出力と、前記第2の予測器の出力と、前記一般化後方互換性指標とを用いて、前記第1の予測器と前記第2の予測器との互換性を示すスコアを算出する演算手段と、
を備える互換性評価装置。

【請求項2】

前記一般化後方互換性指標は、重み付けされた複数の関係式の四則演算により表される請求項1記載の互換性評価装置。

【請求項3】

互換性指標の指定を受け取る指定手段を備え、
前記指標決定手段は、前記指定に基づいて前記複数の関係式の各々に対する重みを設定して、前記一般化後方互換性指標から評価用指標を決定し、
前記演算手段は、前記評価用指標を用いて前記スコアを算出する請求項2に記載の互換性評価装置。

【請求項4】

前記関係式は、

10

20

前記第 1 の予測器の出力と前記第 2 の予測器の出力が共に正解である割合を示す第 1 式と、

前記第 1 の予測器の出力と前記第 2 の予測器の出力が共に不正解である割合を示す第 2 式と、

前記第 1 の予測器の出力が不正解であり、前記第 2 の予測器の出力が正解である割合を示す第 3 式と、

前記第 1 の予測器の出力が正解であり、前記第 2 の予測器の出力が不正解である割合を示す第 4 式と、を含む請求項 1 乃至 3 のいずれか一項に記載の互換性評価装置。

【請求項 5】

前記第 1 の予測器及び前記第 2 の予測器は回帰分析を行い、

前記演算手段は、前記第 1 の予測器及び前記第 2 の予測器の出力である予測値と、当該予測値に対応する実績値との差が所定の閾値以下である場合、当該出力は正解であるとみなし、前記差が前記閾値より大きい場合、当該出力は不正解であるとみなす請求項 4 に記載の互換性評価装置。

【請求項 6】

前記関係式は、2 つの評価データに対する前記第 1 の予測器の出力の大小関係、及び、前記 2 つの評価データに対する前記第 2 の予測器の出力の大小関係を示し、

前記演算手段は、前記第 1 の予測器の出力の大小関係と、前記第 2 の予測器の出力の大小関係とが一致する期待値を前記スコアとして算出する請求項 1 に記載の互換性評価装置。

【請求項 7】

コンピュータにより実行される互換性評価方法であって、

評価データに対する第 1 の予測器及び第 2 の予測器の出力を取得し、

前記第 1 の予測器の出力と前記第 2 の予測器の出力との関係を示す複数の関係式の組み合わせにより規定される一般化後方互換性指標を決定し、

前記第 1 の予測器の出力と、前記第 2 の予測器の出力と、前記一般化後方互換性指標とを用いて、前記第 1 の予測器と前記第 2 の予測器との互換性を示すスコアを算出する互換性評価方法。

【請求項 8】

評価データに対する第 1 の予測器及び第 2 の予測器の出力を取得し、

前記第 1 の予測器の出力と前記第 2 の予測器の出力との関係を示す複数の関係式の組み合わせにより規定される一般化後方互換性指標を決定し、

前記第 1 の予測器の出力と、前記第 2 の予測器の出力と、前記一般化後方互換性指標とを用いて、前記第 1 の予測器と前記第 2 の予測器との互換性を示すスコアを算出する処理をコンピュータに実行させるプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本開示は、予測器を評価する技術に関する。

【背景技術】

【0002】

AI (Artificial Intelligence) の運用においては、環境の変化などに対して AI の性能を適応、向上させるため、新たなデータを用いて再学習を行い、AI を更新することが必須である。AI を更新する際には、更新後の AI の精度が更新前より向上することが求められる。特許文献 1 は、機械学習により生成したモデルの更新に際し、モデルの改悪を低減する手法を開示している。また、特許文献 2 は、予測モデルの再学習時に、再学習の前後の予測モデルの構造の近さを、予測モデルの性質の近さとして評価する手法を開示している。

【先行技術文献】

【特許文献】

【0003】

10

20

30

40

50

【文献】特開2019-204190号公報

【文献】国際公開WO2016/151618号公報

【発明の概要】

【発明が解決しようとする課題】

【0004】

AIの更新により精度が向上した場合であっても、更新の前後でAIの挙動が違ってくることがある。例えば、運用中のAIが正解できるデータを更新後のAIが正解できないという現象が起こりうる。この場合、更新後のAIの癖を把握するのにAI運用者が労力や時間を費やす必要が生じたり、AIの予測に対する業務運用に変更が必要となったりすることもある。

10

【0005】

本開示の1つの目的は、予測器の互換性を評価する手法を提供することにある。

【課題を解決するための手段】

【0006】

本開示の一つの観点では、互換性評価装置は、
評価データに対する第1の予測器及び第2の予測器の出力を取得する取得手段と、
前記第1の予測器の出力と前記第2の予測器の出力との関係を示す複数の関係式の組み合わせにより規定される一般化後方互換性指標を決定する指標決定手段と、
前記第1の予測器の出力と、前記第2の予測器の出力と、前記一般化後方互換性指標とを用いて、前記第1の予測器と前記第2の予測器との互換性を示すスコアを算出する演算手段と、を備える。

20

【0007】

本開示の他の観点では、互換性評価方法は、
評価データに対する第1の予測器及び第2の予測器の出力を取得し、
前記第1の予測器の出力と前記第2の予測器の出力との関係を示す複数の関係式の組み合わせにより規定される一般化後方互換性指標を決定し、
前記第1の予測器の出力と、前記第2の予測器の出力と、前記一般化後方互換性指標とを用いて、前記第1の予測器と前記第2の予測器との互換性を示すスコアを算出する。

【0008】

本開示のさらに他の観点では、プログラムは、
評価データに対する第1の予測器及び第2の予測器の出力を取得し、
前記第1の予測器の出力と前記第2の予測器の出力との関係を示す複数の関係式の組み合わせにより規定される一般化後方互換性指標を決定し、
前記第1の予測器の出力と、前記第2の予測器の出力と、前記一般化後方互換性指標とを用いて、前記第1の予測器と前記第2の予測器との互換性を示すスコアを算出する処理をコンピュータに実行させる。

30

【発明の効果】

【0009】

本開示によれば、予測器の互換性を評価することができる。

【図面の簡単な説明】

40

【0010】

【図1】更新前AIと更新後AIの評価データに対する予測結果の例を示す。

【図2】第1実施形態に係る互換性評価装置の全体構成を示すブロック図である。

【図3】第1実施形態に係る互換性評価装置のハードウェア構成を示すブロック図である。

【図4】第1実施形態に係る互換性評価装置の機能構成を示すブロック図である。

【図5】第1実施形態の互換性評価処理のフローチャートである。

【図6】第2実施形態に係る互換性評価装置の機能構成を示すブロック図である。

【図7】第2実施形態に係る互換性評価装置による処理のフローチャートである。

【発明を実施するための形態】

【0011】

50

以下、図面を参照して、本開示の好適な実施形態について説明する。

< 互換性評価指標 >

(予測器の互換性)

新たなデータを用いて A I の更新 (再学習) を行う場合、精度が向上するように更新を行うが、その際に A I の互換性が問題となる。互換性とは、更新前 A I の正解 / 不正解と、更新後 A I の正解 / 不正解との一致度合いを言う。

【 0 0 1 2 】

互換性を示す指標の 1 つとして、後方信頼互換 (Backward Trust Compatibility ; B T C) スコア (以降、「 B T C 」と呼ぶ。) がある。 B T C は、更新前 A I が正解できるデータを、更新後 A I も正解できる割合を言い、 B T C が高いと、互換性が高いとされる。

10

【 0 0 1 3 】

図 1 は、更新前 A I と、2 つの更新後 A I の評価データに対する予測結果の例を示す。更新前 A I は現在運用中の A I である。2 つの更新後 A I は、更新前 A I を再学習して得た A I であるが、ハイパーパラメータを変えるなどして生成した異なる A I である。図 1 において、チェックマークは予測結果が正解であることを示す。

【 0 0 1 4 】

図示のように、更新前 A I は、評価データ 1 ~ 7 のうち 4 つを正解しており、精度は 4 / 7 である。これに対し、第 1 の更新後 A I と第 2 の更新後 A I は共に精度が 5 / 7 であり、更新前 A I よりも精度が向上している。一方で、第 1 の更新後 A I は、更新前 A I が正解していた 4 つの評価データのうち星印 (\star) で示す 3 つの評価データを正解しており、 B T C スコアは 3 / 4 である。これに対し、第 2 の更新後 A I は、更新前 A I が正解していた 4 つの評価データのうち 2 つしか正解できておらず、 B T C スコアは 2 / 4 である。よって、2 つの更新後 A I は精度が同一であるが、互換性 (B T C スコア) が高い第 1 の更新後 A I の方が良いと評価される。

20

【 0 0 1 5 】

互換性を示す別の指標として、後方誤り互換 (Backward Error Compatibility ; B E C) スコア (以降、「 B E C 」と呼ぶ。) がある。 B E C は、更新後 A I が間違えるデータを更新前 A I も間違える割合であり、 B E C スコアが高いと、互換性が高いとされる。

【 0 0 1 6 】

このように、再学習により A I を更新する際には、精度のみならず、更新前 A I との互換性を考慮する必要がある。以下では、様々なタスクに適用することができる一般化後方互換性指標を提案する。

30

【 0 0 1 7 】

(一般化後方互換性指標)

一般化後方互換性指標は、前述の B T C や B E C などの互換性指標を一般化した指標である。以下に、一般化後方互換性指標の例を説明する。

【 0 0 1 8 】

(第 1 例)

第 1 例は、最も基本的な一般化後方互換性指標の例である。予測器 h 及び入出力の組 (X, Y) を、

40

【 数 1 】

- ・ 予測器 : $h: \mathcal{X} \rightarrow \mathcal{Y}$
- ・ 入出力の組 : $(X, Y) \in \mathcal{X} \times \mathcal{Y}$

とすると、第 1 例の一般化後方互換性 (Generalized Backward Compatibility ; GBC) スコアは、以下のような線形分数指標により定義される。

【 0 0 1 9 】

【数 2】

$$GBC(h_1, h_2) = \frac{a_0 + a_{00}EC(h_1, h_2) + a_{01}IC_1(h_1, h_2) + a_{10}IC_2(h_1, h_2) + a_{11}CC(h_1, h_2)}{b_0 + b_{00}EC(h_1, h_2) + b_{01}IC_1(h_1, h_2) + b_{10}IC_2(h_1, h_2) + b_{11}CC(h_1, h_2)} \quad (1)$$

上記の式 (1) は、評価データに対する予測器 h_1 の出力と予測器 h_2 の出力との間の関係を示す 4 つの関係式 $CC(h_1, h_2)$ 、 $EC(h_1, h_2)$ 、 $IC_1(h_1, h_2)$ 、 $IC_2(h_1, h_2)$ を含む。「 a_0 」、「 a_{00} 」、「 a_{01} 」、「 a_{10} 」、「 a_{11} 」、「 b_0 」、「 b_{00} 」、「 b_{01} 」、「 b_{10} 」、「 b_{11} 」はそれぞれ係数 (重み) である。

10

【 0 0 2 0 】

4 つの関係式は以下の意味を有する。

- ・ CC (Correct Compatibility) (h_1, h_2) は、全評価データのうち、予測器 h_1 が正解を出力し、予測器 h_2 が正解を出力する評価データが占める割合を示す。
- ・ EC (Error Compatibility) (h_1, h_2) は、全評価データのうち、予測器 h_1 が不正解を出力し、予測器 h_2 が不正解を出力する評価データが占める割合を示す。
- ・ IC_1 (Imcompatibility-1) (h_1, h_2) は、全評価データのうち、予測器 h_1 が正解を出力し、予測器 h_2 が不正解を出力する評価データが占める割合を示す。
- ・ IC_2 (Imcompatibility-2) (h_1, h_2) は、全評価データのうち、予測器 h_1 が不正解を出力し、予測器 h_2 が正解を出力する評価データが占める割合を示す。

20

【 0 0 2 1 】

具体的に、上記 4 つの関係式は以下のように与えられる。

【数 3】

$$CC(h_1, h_2) = \mathbb{E}_{XY} [\mathbb{1}[h_1(X) = Y, h_2(X) = Y]] \quad (2)$$

$$EC(h_1, h_2) = \mathbb{E}_{XY} [\mathbb{1}[h_1(X) \neq Y, h_2(X) \neq Y]] \quad (3)$$

$$IC_1(h_1, h_2) = \mathbb{E}_{XY} [\mathbb{1}[h_1(X) \neq Y, h_2(X) = Y]] \quad (4)$$

$$IC_2(h_1, h_2) = \mathbb{E}_{XY} [\mathbb{1}[h_1(X) = Y, h_2(X) \neq Y]] \quad (5)$$

30

\mathbb{E}_{XY} は (X, Y) の期待値

$$\mathbb{1}[cond] = \begin{cases} 1 & \text{if } cond \text{ is true} \\ 0 & \text{otherwise} \end{cases}$$

40

【 0 0 2 2 】

式 (1) において、係数 a_{11} 、 b_{10} 、 b_{11} を「1」に設定し、他の係数を「0」に設定すると、式 (1) の GBC スコアは BTC スコアと一致する。よって、上記の GBC は BTC を包含している。

【 0 0 2 3 】

また、式 (1) において、係数 a_{00} 、 b_{00} 、 b_{10} を「1」に設定し、他の係数を「0」に設定すると、式 (1) の GBC スコアは BEC スコアと一致する。よって、上記の GBC は BEC を包含している。

50

【 0 0 2 4 】

このように、上記の一般化後方互換性指標（GBC）を利用すると、式（1）の係数（重み）を変更することにより、予測器のタスクに応じて適切な互換性指標を定義することができる。

【 0 0 2 5 】

次に、第1例のGBCを用いたスコアの計算式の例を示す。いま、入力を以下のように設定する。

【数4】

- ・互換性を評価する2つの予測器： h_1, h_2
- ・評価データ： $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$
- ・設定した重み： $a_0, a_{00}, a_{01}, a_{10}, a_{11}$
 $b_0, b_{00}, b_{01}, b_{10}, b_{11}$

10

【 0 0 2 6 】

GBCスコアの推定値 \widehat{GBC} は、以下の式で与えられる。なお、便宜上、文字「X」の上に「 $\widehat{}$ 」を付した記号を「 \widehat{X} 」と表記する。

20

【数5】

$$\widehat{GBC}(h_1, h_2) = \frac{a_0 + a_{00}\widehat{EC}(h_1, h_2) + a_{01}\widehat{IC}_1(h_1, h_2) + a_{10}\widehat{IC}_2(h_1, h_2) + a_{11}\widehat{CC}(h_1, h_2)}{b_0 + b_{00}\widehat{EC}(h_1, h_2) + b_{01}\widehat{IC}_1(h_1, h_2) + b_{10}\widehat{IC}_2(h_1, h_2) + b_{11}\widehat{CC}(h_1, h_2)} \quad (6)$$

【 0 0 2 7 】

なお、各関係式 CC 、 EC 、 IC_1 、 IC_2 は、式（2）～（5）における期待値を標本平均に置き換え、以下の式で与えられる。

30

【数6】

$$\widehat{CC}(h_1, h_2) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}[h_1(\mathbf{x}_i) = y_i, h_2(\mathbf{x}_i) = y_i] \quad (7)$$

$$\widehat{EC}(h_1, h_2) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}[h_1(\mathbf{x}_i) \neq y_i, h_2(\mathbf{x}_i) \neq y_i] \quad (8)$$

$$\widehat{IC}_1(h_1, h_2) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}[h_1(\mathbf{x}_i) \neq y_i, h_2(\mathbf{x}_i) = y_i] \quad (9)$$

40

$$\widehat{IC}_2(h_1, h_2) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}[h_1(\mathbf{x}_i) = y_i, h_2(\mathbf{x}_i) \neq y_i] \quad (10)$$

【 0 0 2 8 】

（第2例）

50

上記の第 1 例では、式 (1) に示すように、4 つの関係式 CC 、 EC 、 IC_1 、 IC_2 に対して係数 (重み) を設定している。これに対し、第 2 例では予測器 h_1 、 h_2 が予測するクラス y 毎に係数 (重み) を設定する。第 2 例に係る GBC スコアは以下の式で与えられる。

【 0 0 2 9 】

【 数 7 】

$$GBC(h_1, h_2) = \frac{a_0 + CC(h_1, h_2; \mathbf{a}_{11}) + EC(h_1, h_2; \mathbf{a}_{00}) + IC_1(h_1, h_2; \mathbf{a}_{01}) + IC_2(h_1, h_2; \mathbf{a}_{10})}{b_0 + CC(h_1, h_2; \mathbf{b}_{11}) + EC(h_1, h_2; \mathbf{b}_{00}) + IC_1(h_1, h_2; \mathbf{b}_{01}) + IC_2(h_1, h_2; \mathbf{b}_{10})} \quad (11)$$

10

また、4 つの関係式は以下のように与えられる。

【 0 0 3 0 】

【 数 8 】

$$CC(h_1, h_2; \mathbf{a}_{11}) = \sum_{y \in \mathcal{Y}} a_{11,y} \mathbb{E}_{XY} [\mathbb{1}[h_1(X) = y, h_2(X) = y, Y = y]] \quad (12)$$

$$EC(h_1, h_2; \mathbf{a}_{00}) = \sum_{y \in \mathcal{Y}} a_{00,y} \mathbb{E}_{XY} [\mathbb{1}[h_1(X) \neq y, h_2(X) \neq y, Y = y]] \quad (13)$$

20

$$IC_1(h_1, h_2; \mathbf{a}_{01}) = \sum_{y \in \mathcal{Y}} a_{01,y} \mathbb{E}_{XY} [\mathbb{1}[h_1(X) \neq y, h_2(X) = y, Y = y]] \quad (14)$$

$$IC_2(h_1, h_2; \mathbf{a}_{10}) = \sum_{y \in \mathcal{Y}} a_{10,y} \mathbb{E}_{XY} [\mathbb{1}[h_1(X) = y, h_2(X) \neq y, Y = y]] \quad (15)$$

* 分母に関しても同様に定義

30

なお、式 (1 1) において、 $a_{11} = a_{11,1} = \dots = a_{11,|y|}$ というように重みを一定にすると、第 1 例の式 (1) と一致する。

【 0 0 3 1 】

第 2 例の GBC では、線形分数式で表せる既存の様々な二値分類指標を後方互換性の文脈で構成することが可能となる。例えば、式 (1 1) に示す GBC の重みを調整し、不均衡二値分類に有効な互換性指標を構成することができる。互換性を考慮しない場合、二値分類 $Y \in \{0, 1\}$ における F 値 ($Y = 1$ が正クラス、 $Y = 0$ が負クラス) は以下のようになる。

【 0 0 3 2 】

【 数 9 】

$$F(h_1) = \frac{2TP(h_1)}{2TP(h_1) + FP(h_1) + FN(h_1)} = \frac{2\mathbb{E}_{XY} [\mathbb{1}[h_1(X) = 1, Y = 1]]}{2\mathbb{E}_{XY} [\mathbb{1}[h_1(X) = 1, Y = 1]] + \mathbb{E}_{XY} [\mathbb{1}[h_1(X) = 1, Y = 0]] + \mathbb{E}_{XY} [\mathbb{1}[h_1(X) = 0, Y = 1]]} \quad (16)$$

40

この F 値は、不均衡二値分類において、データが少ない正クラスを重視する精度の指標となる。

50

【 0 0 3 3 】

一方、互換性を考慮した F 値 (「BC-F」と呼ぶ。)は、GBCにおいて、 $a_{11,1} = b_{11,1} = 2$ 、 $b_{11,0} = b_{00,1} = 1$ とし、残りの係数を「0」とすると、以下のようになる。

【数 1 0】

$$BC-F(h_1, h_2) = \frac{2\mathbb{E}[\mathbb{1}[h_1(X) = 1, h_2(X) = 1, Y = 1]]}{2\mathbb{E}[\mathbb{1}[h_1(X) = 1, h_2(X) = 1, Y = 1]] + \mathbb{E}[\mathbb{1}[h_1(X) = 1, h_2(X) = 1, Y = 0]] + \mathbb{E}[\mathbb{1}[h_1(X) = 0, h_2(X) = 0, Y = 1]]} \quad (17)$$

10

このBC-F値は、不均衡二値分類において、データが少ない正クラスを重視する互換性の指標となる。このように、GBCの重みを調整することにより、様々な二値分類における互換性指標を生成することができる。

【 0 0 3 4 】

(第3例)

第3例は、第1例や第2例のような線形分数式以外の互換性指標の例である。二値分類において、更新前の予測器のスコアランキングが更新後の予測器でも一致して欲しいタスクを考える。予測器が実数を「-1」と「+1」に割り当てるものとする、以下のよう

20

【 0 0 3 5 】

【数 1 1】

- ・ 予測器: $h(X) = \text{sign}(g(X)) \quad g: \mathcal{X} \rightarrow \mathbb{R}$
- ・ クラス: $\mathcal{Y} = \{-1, +1\}$
- ・ 互換性指標の例:

$$\mathbb{E}_{X|Y=1} [\mathbb{E}_{X|Y=-1} [\mathbb{1}[g_1(X) > g_1(X')] \mathbb{1}[g_2(X) > g_2(X')]]] \quad (18)$$

30

【 0 0 3 6 】

この互換性指標は、正解が「+1」の評価データXと、正解が「-1」の評価データX'を入力したときの更新前の予測器の出力の大小関係を示す関係式

【数 1 2】

$$\mathbb{1}[g_1(X) > g_1(X')]$$

40

と、更新後の予測器の出力の大小関係を示す関係式

【数 1 3】

$$\mathbb{1}[g_2(X) > g_2(X')]$$

を含み、更新前のX、X'に対する出力の大小関係が更新後にも維持される期待値がGBC

50

スコアとして得られる。即ち、GBCスコアは、入力に対する更新前後の予測器の出力傾向が一致しているか否かを示す値となる。この互換性指標では、AUC (Area under the ROC curve) のような効果が見込まれる。

【0037】

(回帰タスクへの適用)

上記の第1例及び第2例では、予測器が分類タスクを実行するものとしているが、回帰タスクを実行する予測器に対してもGBCを適用することができる。その場合には、評価データに対して予測器が出力する予測値と、その評価データに対応する実績値との差が予め定めた閾値以下であれば予測値は正解であるとみなし、閾値より大きければ予測値は不正解であるとみなして、第1例又は第2例のGBCを適用すればよい。

10

【0038】

<第1実施形態>

[全体構成]

図2は、第1実施形態に係る互換性評価装置の全体構成を示すブロック図である。互換性評価装置100は、2つの予測器の互換性を評価し、互換性スコアを出力する。図示のように、2つの予測器 h_1 、 h_2 には同一の評価データが入力される。典型的な例では、予測器 h_1 は現在運用中の予測器、即ち、更新前予測器であり、予測器 h_2 は更新後予測器である。

【0039】

予測器 h_1 及び予測器 h_2 は、入力された評価データに対する予測値を互換性評価装置100へ出力する。互換性評価装置100は、上記の一般化後方互換性指標(GBC)を用いて、予測器 h_1 の出力と予測器 h_2 の出力との互換性を示す互換性スコアを出力する。

20

【0040】

[ハードウェア構成]

図3は、互換性評価装置100のハードウェア構成を示すブロック図である。互換性評価装置100は、インタフェース101と、プロセッサ102と、メモリ103と、記録媒体104と、入力部105と、表示部106とを備える。

【0041】

インタフェース(IF)101は、予測器 h_1 、 h_2 から予測値を受け取る。また、IF101は、互換性評価装置100が計算した互換性スコアを外部装置へ出力する。IFは取得手段の一例である。

30

【0042】

プロセッサ102は、CPUなどのコンピュータであり、予め用意されたプログラムを実行することにより、互換性評価装置100の全体を制御する。なお、プロセッサ102は、GPU又はFPGA(Field-Programmable Gate Array)であってもよい。具体的に、プロセッサ102は、後述する互換性評価処理を実行する。

【0043】

メモリ103は、ROM(Read Only Memory)、RAM(Random Access Memory)などにより構成される。メモリ103には、一般化後方互換性指標の情報、指標番号毎の係数(重み)などが記憶される。また、メモリ103は、プロセッサ102による各種の処理の実行中に作業メモリとしても使用される。

40

【0044】

記録媒体104は、ディスク状記録媒体、半導体メモリなどの不揮発性で非一時的な記録媒体であり、互換性評価装置100に対して着脱可能に構成される。記録媒体104は、プロセッサ102が実行する各種のプログラムを記録している。互換性評価装置100が処理を実行する際には、記録媒体104に記録されているプログラムがメモリ103にロードされ、プロセッサ102により実行される。

【0045】

入力部105は、例えばキーボード、マウスなどであり、利用者が各種の指示、入力を行う際に使用される。表示部106は、例えば液晶表示装置などであり、利用者に各種の

50

情報を表示する。

【 0 0 4 6 】

[機能構成]

図 4 は、互換性評価装置 1 0 0 の機能構成を示すブロック図である。互換性評価装置 1 0 0 は、機能面では、評価用指標決定部 1 1 0 と、スコア演算部 1 2 0 とを備える。評価用指標決定部 1 1 0 には、指標番号が入力される。指標番号は、互換性の評価に使用する互換性指標を指定する番号である。指標番号は、例えば更新の対象となる予測器のタスクなどに基づいて決定される。評価用指標決定部 1 1 0 は、入力された指標番号に基づいて、式 (1) や式 (1 1) など示す一般化後方互換性指標 (G B C) を基にして、実際に評価に使用する互換性指標 (以下、「評価用指標」とも呼ぶ。) を決定し、スコア演算部 1 2 0 へ出力する。

10

【 0 0 4 7 】

指標番号は、式 (1) に含まれる係数 (重み) の組み合わせに対応付けて予め決定されている。例えば、互換性指標番号「 1 」が B T C に対応する場合、互換性指標番号「 1 」に対しては、係数の組み合わせ「係数 $a_{11} = b_{10} = b_{11} = 1$ 、他の係数 = 0」が予め対応付けされている。よって、利用者が互換性指標番号「 1 」を入力した場合、評価用指標決定部 1 1 0 は、「係数 $a_{11} = b_{10} = b_{11} = 1$ 、他の係数 = 0」を式 (1) に代入し、B T C スコアを示す評価用指標を生成する。

【 0 0 4 8 】

スコア演算部 1 2 0 は、決定された評価用指標を用いて、予測器 h_1 、 h_2 が出力した予測値から互換性スコアを算出し、出力する。例えば、スコア演算部 1 2 0 は、予測器が出力した予測値を式 (7) ~ (1 0) に代入して 4 つの関係式 $C C (h_1, h_2)$ 、 $E C (h_1, h_2)$ 、 $I C_1 (h_1, h_2)$ 、 $I C_2 (h_1, h_2)$ の値を求め、それらを式 (6) などの評価用指標に代入して G B C スコアを計算し、出力する。

20

【 0 0 4 9 】

なお、評価用指標決定部 1 1 0 は指標決定手段の一例であり、スコア演算部 1 2 0 は演算手段の一例である。

【 0 0 5 0 】

[互換性評価処理]

図 5 は、互換性評価装置 1 0 0 が実行する互換性評価処理のフローチャートである。この処理は、図 3 に示すプロセッサ 1 0 2 が予め用意されたプログラムを実行し、図 4 に示す各要素として動作することにより実現される。

30

【 0 0 5 1 】

まず、互換性評価装置 1 0 0 は、利用者による指標番号の入力を受け取る (ステップ S 1 1)。次に、評価用指標決定部 1 1 0 は、入力された指標番号に基づいて、評価用指標を決定する (ステップ S 1 2)。例えば、評価用指標として前述した第 1 例又は第 2 例の G B C を使用する場合、評価用指標決定部 1 1 0 は、指標番号に対応する各係数 (重み) を取得し、式 (1) 又は式 (1 1) に代入して評価用指標を決定する。

【 0 0 5 2 】

次に、スコア演算部 1 2 0 は、評価データに対して予測器 h_1 、 h_2 が出力した予測値を取得し (ステップ S 1 3)、ステップ S 1 2 で決定された評価用指標に入力して互換性スコア (G B C スコア) を算出し、出力する (ステップ S 1 4)。こうして、予測器 h_1 と予測器 h_2 の互換性を示す互換性スコアが得られる。そして、処理は終了する。

40

【 0 0 5 3 】

[コースケース]

G B C は、予測器の更新時にハイパーパラメータやシードが異なる複数の更新後予測器を生成した際に、それらの互換性を評価する指標として使用することができる。生成された複数の更新後予測器のうち、更新前予測器と互換性の高い予測器を選択することで、更新後の A I の挙動変化に伴う手続き変更などのコストを削減することができる。

【 0 0 5 4 】

50

また、季節性が原因となるようなデータの変化が発生した場合、GBCを用いて、過去の予測モデルの中に現在の予測モデルと互換性の高い予測モデルが無いかを調べることができる。現在の予測モデルと互換性が高く、かつ、精度の高い過去の予測モデルがある場合には、現在の予測モデルをその予測モデルに切り替えることにより、再学習のコストをかけることなく、その季節に適した予測モデルへの切り替えが可能となる。

【0055】

また、AIの運用時に、ビジネス側のKPI（Key Performance Indicator：重要業績評価指標）が変わった場合には、GBCを用いて、新しいKPIが重視する項目（例えば正解したいクラス）などを重視した互換性指標を構築し、継続的なAI運用に役立てることができる。

【0056】

[GBCを活用した予測器の構築]

上記の例では、GBCを更新時などにおける予測器の互換性評価に使用しているが、その代わりに、GBCを予測器の学習において利用することもできる。この場合、予測モデルの学習時に、通常の学習時に用いる誤差関数にGBCを正則化として加える。具体的には、既存の一般化二値分類指標と同様に、指示関数を損失関数（二乗損失やヒンジ損失）に置き換えることにより、GBCの上界を構成することができる。そして、構成した上界と通常の二値分類の誤差関数を合わせたものを最小化するように予測モデルを学習する。更新前の予測器と追加収集したデータを入力とし、GBCを正則化にすることで、対象タスクに適した後方互換性の高い新たな予測器を構築することができる。

【0057】

<第2実施形態>

次に、本開示の第2実施形態について説明する。図6は、第2実施形態に係る互換性評価装置70の機能構成を示すブロック図である。互換性評価装置70は、取得手段71と、指標決定手段72と、演算手段73とを備える。

【0058】

図7は、互換性評価装置70による処理のフローチャートである。取得手段71は、評価データに対する第1の予測器及び第2の予測器の出力を取得する（ステップS41）。指標決定手段72は、第1の予測器の出力と第2の予測器の出力との関係を示す複数の関係式の組み合わせにより規定される一般化後方互換性指標を決定する（ステップS42）。演算手段73は、第1の予測器の出力と、第2の予測器の出力と、一般化後方互換性指標とを用いて、第1の予測器と第2の予測器との互換性を示すスコアを算出する（ステップS43）。

【0059】

第2実施形態の互換性評価装置70によれば、予測器のタスクに応じた適切な互換性指標を用いて、予測器の互換性を評価することができる。

【0060】

上記の実施形態の一部又は全部は、以下の付記のようにも記載されうるが、以下には限られない。

【0061】

（付記1）

評価データに対する第1の予測器及び第2の予測器の出力を取得する取得手段と、前記第1の予測器の出力と前記第2の予測器の出力との関係を示す複数の関係式の組み合わせにより規定される一般化後方互換性指標を決定する指標決定手段と、

前記第1の予測器の出力と、前記第2の予測器の出力と、前記一般化後方互換性指標とを用いて、前記第1の予測器と前記第2の予測器との互換性を示すスコアを算出する演算手段と、

を備える互換性評価装置。

【0062】

（付記2）

10

20

30

40

50

前記一般化後方互換性指標は、重み付けされた複数の関係式の四則演算により表される付記 1 記載の互換性評価装置。

【 0 0 6 3 】

(付記 3)

互換性指標の指定を受け取る指定手段を備え、

前記指標決定手段は、前記指定に基づいて前記複数の関係式の各々に対する重みを設定して、前記一般化後方互換性指標から評価用指標を決定し、

前記演算手段は、前記評価用指標を用いて前記スコアを算出する付記 2 に記載の互換性評価装置。

【 0 0 6 4 】

(付記 4)

前記関係式は、

前記第 1 の予測器の出力と前記第 2 の予測器の出力が共に正解である割合を示す第 1 式と、

前記第 1 の予測器の出力と前記第 2 の予測器の出力が共に不正解である割合を示す第 2 式と、

前記第 1 の予測器の出力が不正解であり、前記第 2 の予測器の出力が正解である割合を示す第 3 式と、

前記第 1 の予測器の出力が正解であり、前記第 2 の予測器の出力が不正解である割合を示す第 4 式と、を含む付記 1 乃至 3 のいずれか一項に記載の互換性評価装置。

【 0 0 6 5 】

(付記 5)

前記第 1 の予測器及び前記第 2 の予測器は回帰分析を行い、

前記演算手段は、前記第 1 の予測器及び前記第 2 の予測器の出力である予測値と、当該予測値に対応する実績値との差が所定の閾値以下である場合、当該出力は正解であるとみなし、前記差が前記閾値より大きい場合、当該出力は不正解であるとみなす付記 4 に記載の互換性評価装置。

【 0 0 6 6 】

(付記 6)

前記関係式は、2つの評価データに対する前記第 1 の予測器の出力の大小関係、及び、前記 2 つの評価データに対する前記第 2 の予測器の出力の大小関係を示し、

前記演算手段は、前記第 1 の予測器の出力の大小関係と、前記第 2 の予測器の出力の大小関係とが一致する期待値を前記スコアとして算出する付記 1 に記載の互換性評価装置。

【 0 0 6 7 】

(付記 7)

評価データに対する第 1 の予測器及び第 2 の予測器の出力を取得し、

前記第 1 の予測器の出力と前記第 2 の予測器の出力との関係を示す複数の関係式の組み合わせにより規定される一般化後方互換性指標を決定し、

前記第 1 の予測器の出力と、前記第 2 の予測器の出力と、前記一般化後方互換性指標とを用いて、前記第 1 の予測器と前記第 2 の予測器との互換性を示すスコアを算出する互換性評価方法。

【 0 0 6 8 】

(付記 8)

評価データに対する第 1 の予測器及び第 2 の予測器の出力を取得し、

前記第 1 の予測器の出力と前記第 2 の予測器の出力との関係を示す複数の関係式の組み合わせにより規定される一般化後方互換性指標を決定し、

前記第 1 の予測器の出力と、前記第 2 の予測器の出力と、前記一般化後方互換性指標とを用いて、前記第 1 の予測器と前記第 2 の予測器との互換性を示すスコアを算出する処理をコンピュータに実行させるプログラムを記録した記録媒体。

【 0 0 6 9 】

10

20

30

40

50

以上、実施形態及び実施例を参照して本開示を説明したが、本開示は上記実施形態及び実施例に限定されるものではない。本開示の構成や詳細には、本開示の範囲内で当業者が理解し得る様々な変更をすることができる。

【符号の説明】

【0070】

- 100 互換性評価装置
- 101 インタフェース
- 102 プロセッサ
- 103 メモリ
- 104 記録媒体
- 105 入力部
- 106 表示部
- 110 評価用指標決定部
- 120 スコア演算部

10

20

30

40

50

【図面】

【図 1】

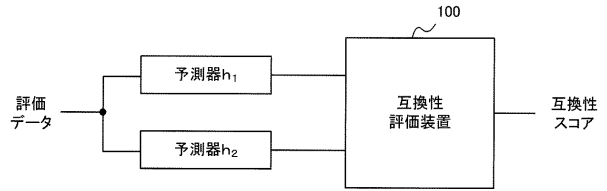
評価データ

ID	更新前AI	第1の更新後AI	第2の更新後AI
1	✓	✓ ★	✓ ★
2	✓	✓ ★	
3		✓	✓
4			✓
5	✓		✓ ★
6		✓	✓
7	✓	✓ ★	
精度	4/7	5/7	5/7
BTC	—	3/4	2/4

「✓」: 正解

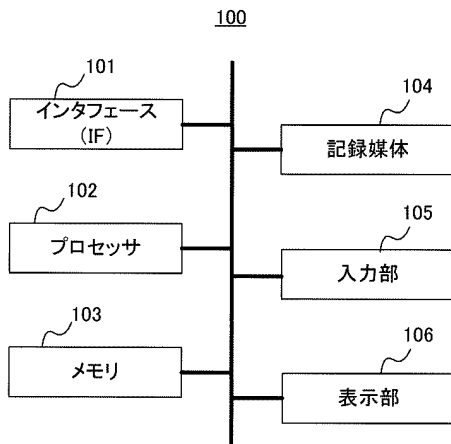
「★」: 更新前後ともに正解

【図 2】

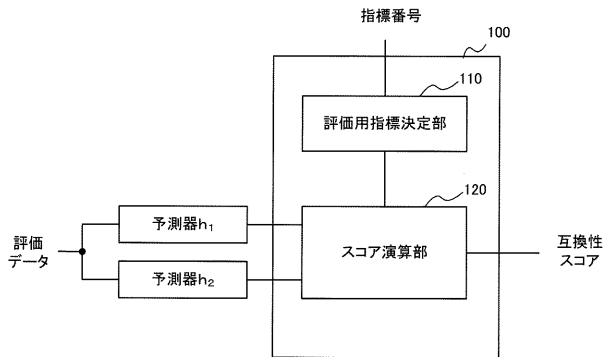


10

【図 3】



【図 4】



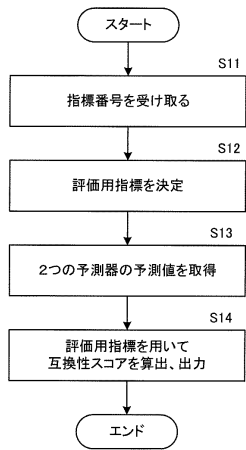
20

30

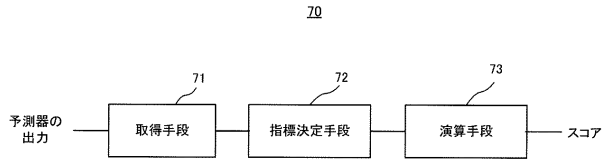
40

50

【 図 5 】



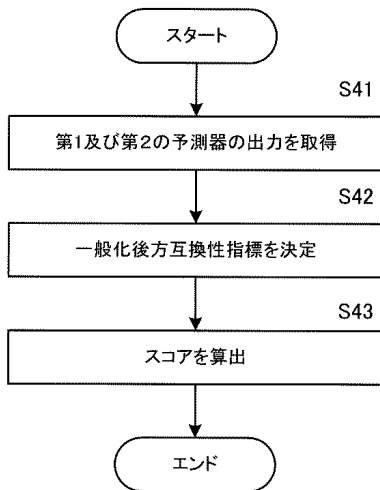
【 図 6 】



10

20

【 図 7 】



30

40

50

フロントページの続き

- (56)参考文献 特開 2 0 2 0 - 0 0 4 1 7 8 (J P , A)
米国特許第 8 2 9 6 2 5 7 (U S , B 1)
SRIVASTAVA, Megha et al. , An Empirical Analysis of Backward Compatibility in Machine Learning Systems , arXiv.org [online] , 米国 , Cornell University , 2020年08月11日 , [検索日: 2021.04.23] , pp.1-9 , インターネット < URL : <https://arxiv.org/pdf/2008.04572.pdf> >
- (58)調査した分野 (Int.Cl. , D B 名)
G 0 6 N 2 0 / 0 0