

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6832783号
(P6832783)

(45) 発行日 令和3年2月24日(2021.2.24)

(24) 登録日 令和3年2月4日(2021.2.4)

(51) Int.Cl.

G06N 3/08 (2006.01)

F I

G06N 3/08 120

請求項の数 10 (全 22 頁)

(21) 出願番号	特願2017-83608 (P2017-83608)	(73) 特許権者	000005108
(22) 出願日	平成29年4月20日 (2017.4.20)		株式会社日立製作所
(65) 公開番号	特開2018-181187 (P2018-181187A)		東京都千代田区丸の内一丁目6番6号
(43) 公開日	平成30年11月15日 (2018.11.15)	(74) 代理人	110001678
審査請求日	令和1年10月17日 (2019.10.17)		特許業務法人藤央特許事務所
		(72) 発明者	柴原 琢磨
			東京都千代田区丸の内一丁目6番6号 株式会社日立製作所内
		審査官	渡部 博樹
		(56) 参考文献	特開2018-026040 (JP, A)
)
		(58) 調査した分野(Int.Cl., DB名)	
			G06N 3/02

(54) 【発明の名称】 データ分析装置、データ分析方法、およびデータ分析プログラム

(57) 【特許請求の範囲】

【請求項1】

第1入力層と、第1出力層と、前記第1入力層と前記第1出力層との間において前段の層からのデータと第1学習パラメータとを第1活性化関数に与えて計算して後段の層に計算結果を出力する2層以上の第1中間層と、により構成される第1ニューラルネットワークを用いたデータ分析装置であって、

前記各第1中間層からの出力データを受け付けて、前記各出力データと第2学習パラメータとに基づいて、前記各第1中間層の重みを設定して、前記第1出力層に出力する設定部と、

前記第1出力層に含まれ、前記各出力データを、前記設定部によって設定された前記第1中間層ごとの重みにより、それぞれ重み付けする重み付け処理部と、

前記第1出力層に含まれ、前記重み付け処理部によって重み付けされた各出力データと第3学習パラメータとに基づいて、予測データを算出する算出部と、

を有することを特徴とするデータ分析装置。

【請求項2】

請求項1に記載のデータ分析装置であって、

前記設定部は、前記各第1中間層からの出力データに替えて、前記第1入力層からの出力データを受け付けて、当該出力データと前記第2学習パラメータとに基づいて、前記各第1中間層の重みを設定して、前記第1出力層に出力することを特徴とするデータ分析装置。

10

20

【請求項 3】

請求項 1 に記載のデータ分析装置であって、

前記各第 1 中間層からの出力データを受け付けて、前記各出力データの次元数を縮退させて、縮退後の各出力データを前記設定部に出力する第 1 縮退部を有し、

前記設定部は、前記第 1 縮退部からの縮退後の各出力データを受け付けて、前記縮退後の各出力データと前記第 2 学習パラメータとに基づいて、前記各第 1 中間層の重みを設定して、前記第 1 出力層に出力することを特徴とするデータ分析装置。

【請求項 4】

請求項 1 に記載のデータ分析装置であって、

前記第 1 入力層に訓練データが与えられた場合に、前記第 1 学習パラメータ、前記第 2 学習パラメータ、および前記第 3 学習パラメータを調整する学習部を有することを特徴とするデータ分析装置。

10

【請求項 5】

請求項 3 に記載のデータ分析装置であって、

前記各第 1 中間層からの出力データを受け付けて、前記各出力データの次元数を縮退させて、縮退後の各出力データを前記重み付け処理部に出力する第 2 縮退部を有し、

前記重み付け処理部は、前記第 2 縮退部からの縮退後の各出力データを、前記第 1 中間層ごとの重みにより、それぞれ重み付けすることを特徴とするデータ分析装置。

【請求項 6】

請求項 4 に記載のデータ分析装置であって、

前記学習部は、

前記訓練データを受け付ける第 2 入力層と、前記第 1 ニューラルネットワークのハイパーパラメータを出力する第 2 出力層と、前記第 2 入力層と前記第 2 出力層との間において前段の層からのデータと第 4 学習パラメータとを第 2 活性化関数に与えて計算して後段の層に計算結果を出力する第 2 中間層と、により構成される第 2 ニューラルネットワークを用い、前記第 2 入力層に前記訓練データが与えられた場合に、前記第 4 学習パラメータを調整し、

20

前記訓練データを、前記第 4 学習パラメータが調整された前記第 2 ニューラルネットワークの前記第 2 入力層に与えることにより、前記第 2 出力層から前記ハイパーパラメータを出力し、

30

前記ハイパーパラメータにより前記第 1 ニューラルネットワークの構造を決定し、

前記構造が決定された前記第 1 ニューラルネットワークの前記第 1 入力層に前記訓練データが与えられた場合に、前記第 1 学習パラメータ、前記第 2 学習パラメータ、および前記第 3 学習パラメータを調整することを特徴とするデータ分析装置。

【請求項 7】

請求項 6 に記載のデータ分析装置であって、

前記ハイパーパラメータは、前記第 1 活性化関数の種類、損失関数の種類、正則化関数の種類、または前記第 1 ニューラルネットワークの種類を示すパラメータであることを特徴とするデータ分析装置。

【請求項 8】

請求項 6 に記載のデータ分析装置であって、

前記ハイパーパラメータは、前記第 1 中間層の層数または前記第 1 中間層内のユニット数を示すパラメータであることを特徴とするデータ分析装置。

40

【請求項 9】

第 1 入力層と、第 1 出力層と、前記第 1 入力層と前記第 1 出力層との間において前段の層からのデータと第 1 学習パラメータとを第 1 活性化関数に与えて計算して後段の層に計算結果を出力する 2 層以上の第 1 中間層と、により構成される第 1 ニューラルネットワークを用いたデータ分析装置によるデータ分析方法であって、

前記データ分析装置は、プロセッサと、前記第 1 ニューラルネットワークを記憶する記憶デバイスと、を有し、

50

前記プロセッサは、

前記各第 1 中間層からの出力データを受け付けて、前記各出力データと第 2 学習パラメータとに基づいて、前記各第 1 中間層の重みを設定して、前記第 1 出力層に出力する設定処理と、

前記第 1 出力層に含まれ、前記各出力データを、前記設定処理によって設定された前記第 1 中間層ごとの重みにより、それぞれ重み付けする重み付け処理と、

前記第 1 出力層に含まれ、前記重み付け処理によって重み付けされた各出力データと第 3 学習パラメータとに基づいて、予測データを算出する算出処理と、

を実行することを特徴とするデータ分析方法。

【請求項 10】

10

第 1 入力層と、第 1 出力層と、前記第 1 入力層と前記第 1 出力層との間において前段の層からのデータと第 1 学習パラメータとを第 1 活性化関数に与えて計算して後段の層に計算結果を出力する 2 層以上の第 1 中間層と、により構成される第 1 ニューラルネットワークが記憶された記憶デバイスにアクセス可能なプロセッサに、

前記各第 1 中間層からの出力データを受け付けて、前記各出力データと第 2 学習パラメータとに基づいて、前記各第 1 中間層の重みを設定して、前記第 1 出力層に出力する設定処理と、

前記第 1 出力層に含まれ、前記各出力データを、前記設定処理によって設定された前記第 1 中間層ごとの重みにより、それぞれ重み付けする重み付け処理と、

前記第 1 出力層に含まれ、前記重み付け処理によって重み付けされた各出力データと第 3 学習パラメータとに基づいて、予測データを算出する算出処理と、

20

を実行させることを特徴とするデータ分析プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、データを分析するデータ分析装置、データ分析方法、およびデータ分析プログラムに関する。

【背景技術】

【0002】

入力データから出力データを予測する手法として、パーセプトロンと呼ばれる手法がある。パーセプトロンは、入力ベクトルと重みベクトルの線形結合の演算結果により、予測値を出力する。ニューラルネットワークは、別名、マルチパーセプトロンとも呼ばれ、複数のパーセプトロンを多層的に重ねることで、線形分離不可能問題を解く能力をもたせた技術であり、1980年代に登場した。2012年頃から、ドロップアウト（下記特許文献1を参照）等の新しい技術を導入したニューラルネットワークはdeep learningと呼ばれている。ドロップアウトは、ニューラルネットワークの学習の際に、各層の学習パラメータをランダムに欠損させる技術であり、下層から上層までの学習が行われることを狙っている。

30

【0003】

機械学習分野では、入力ベクトルから得られた予測値と実際の値（真値）の誤差が最小となるように、学習パラメータ（パーセプトロンにおける重みベクトル等）を計算することを学習と呼ぶ。学習のプロセスが完了すると、学習に用いなかったデータ（以降、テストデータと呼ぶ）から新しい予測値を計算することができる。ニューラルネットワークでは、テストデータの予測値と真値の誤差が大きくなる傾向にあることが知られており、この問題は過学習と呼ばれる。過学習が発生する要因の1つとして、パーセプトロンを多層化したことで、下層の学習パラメータの学習が十分に行われない場合がある。

40

【先行技術文献】

【特許文献】

【0004】

【特許文献1】国際公開WO2014/105866号公報

50

【発明の概要】

【発明が解決しようとする課題】

【0005】

しかしながら、ドロップアウトを用いた場合でも、10層程度のネットワークで過学習が発生する場合があります、本質的な改善には至っていない。

【0006】

本発明は、上記に鑑みてなされたものであって、ニューラルネットワークの過学習の抑制を図ることを目的とする。

【課題を解決するための手段】

【0007】

本願において開示される発明の一側面となるデータ分析装置は、第1入力層と、第1出力層と、前記第1入力層と前記第1出力層との間において前段の層からのデータと第1学習パラメータとを第1活性化関数に与えて計算して後段の層に計算結果を出力する2層以上の第1中間層と、により構成される第1ニューラルネットワークを用いたデータ分析装置であって、前記各第1中間層からの出力データを受け付けて、前記各出力データと第2学習パラメータとに基づいて、前記各第1中間層の重みを設定して、前記第1出力層に出力する設定部と、前記第1出力層に含まれ、前記各出力データを、前記設定部によって設定された前記第1中間層ごとの重みにより、それぞれ重み付けする重み付け処理部と、前記第1出力層に含まれ、前記重み付け処理部によって重み付けされた各出力データと第3学習パラメータとに基づいて、予測データを算出する算出部と、を有することを特徴とする。

【発明の効果】

【0008】

本発明の代表的な実施の形態によれば、ニューラルネットワークの過学習の抑制を図ることができる。前述した以外の課題、構成及び効果は、以下の実施例の説明により明らかにされる。

【図面の簡単な説明】

【0009】

【図1】図1は、実施例1にかかるニューラルネットワークの構造例を示す説明図である。

【図2】図2は、データ分析システムのシステム構成例を示すブロック図である。

【図3】図3は、データ分析装置の機能的構成例を示すブロック図である。

【図4】図4は、データ分析装置によるデータ分析処理手順例を示すフローチャートである。

【図5】図5は、図4に示したハイパーパラメータ算出処理（ステップS402）の詳細な処理手順例を示すフローチャートである。

【図6】図6は、データ分析結果の比較例を示す説明図である。

【図7】図7は、実施例2にかかるニューラルネットワークの構造例を示す説明図である。

【図8】図8は、実施例3にかかるニューラルネットワークの構造例を示す説明図である。

【図9】図9は、実施例4にかかるニューラルネットワークの構造例を示す説明図である。

【発明を実施するための形態】

【実施例1】

【0010】

<ニューラルネットワーク>

図1は、実施例1にかかるニューラルネットワーク100の構造例を示す説明図である。ニューラルネットワーク100は、データユニット群DUと、レポーティングユニット群RUと、サジェスチョンユニットSUと、識別ユニットIUと、を有する。データユニ

10

20

30

40

50

ット群 DU は、複数のデータユニット DU_l (l は、階層番号であり、 $0 \leq l \leq L$ 。 L は最下層の階層番号であり、図 1 では、 $L = 3$) を直列に接続した構成である。最上段である $l = 0$ のデータユニット DU_0 は、ニューラルネットワーク 100 の入力層であり、 $l = 1$ のデータユニット DU_1 は、ニューラルネットワーク 100 の中間層（隠れ層ともいう）に相当する。データユニット DU_l は、前段のデータユニット DU_{l-1} からの入力データを自データユニット DU_l の学習パラメータを用いて計算して出力データを出力するパーセプトロンである。

【0011】

ただし、入力層となるデータユニット DU_0 は、入力データを受け付けず、訓練データやテストデータを出力する。ここで、訓練データは、たとえば、画像 x_n とそのラベル y_n とにより構成される ($n = 1, 2, \dots, N$ 。 N は画像の枚数)。画像 x_n は、2次元のマトリックス構造を持つデータであり、以降、行列として取り扱う。説明の容易さのために、「 x 」と表記した場合は、行列形式の画像 x_n をラスタスキャンした 1次元ベクトルとする。なお、テストデータについても同様の構成である。

10

【0012】

ラベル y_n は、画像 x_n の種類数 K に対して、one hot 表現で種類（たとえば、犬や猫などの動物）を示す K 次元のベクトルである。one hot 表現では、ベクトルのある要素が画像 x_n の種類に対応しており、ただ 1 つの要素だけに 1、0 が格納され、他の要素は全て 0、0 である。1、0 の要素に対応する種類（たとえば、犬）が正解となる種類である。なお、画像 x_n を X 線、CT、MRI、超音波等の医療用画像を入力とした場合、疾病の種類や患者の予後（良いまたは悪い）を予測することができる。

20

【0013】

データユニット DU_l における出力ベクトル h^l_D の計算は、下記式 (1) で実施される。

【0014】

【数 1】

$$h_D^l = g(W_D^l h_D^{l-1}) \dots (1)$$

30

ただし、 $h_D^0 = x$

【0015】

式 (1) 中、添え字 l は、階層番号を示す（以下の式も同様）。ただし、入力層となるデータユニット DU_0 からの出力ベクトル h_D^0 は、訓練データやテストデータとなる。右辺の行列 W_D^l は、データユニット DU_l の学習パラメータの 1 つである。関数 g は、活性化関数である。活性化関数は、たとえば、シグモイド関数、双曲線正接関数（tanh 関数）、ReLU (Rectified Linear Unit) 関数のいずれかの関数である。右辺のベクトル h^{l-1}_D は、データユニット DU_l に入力される入力ベクトル、すなわち、前段のデータユニット DU_{l-1} からの出力ベクトルである。

40

【0016】

レポーティングユニット RU_l は、同一階層のデータユニット DU_l からの出力ベクトル h_D^l を入力し、当該出力ベクトル h_D^l の次元数を縮退させる。レポーティングユニット RU_l からの出力ベクトル h_R^l の計算は、下記式 (2) で実施される。

【0017】

【数 2】

$$h_R^l = g(W_R^l h_D^l) \dots (2)$$

50

【 0 0 1 8 】

式(2)中、行列 W^1_R はレポーティングユニット $RU1$ の学習パラメータである。具体的には、たとえば、データユニット $DU1$ からの出力ベクトル h^1_D の次元数を d とした場合、行列 W^1_R は、たとえば、1行目から m ($m < d$) 行目までの成分を行列 W^1_D と同一成分とし、 $d - m + 1$ 行から k 行目までの成分を0とした行列である。これにより、データユニット $DU1$ からの d 次元の出力ベクトル h^1_D は、 m 次元の出力ベクトル h^1_R に縮退される。

【 0 0 1 9 】

サジェスチョンユニット SU は、各レポーティングユニット $RU1$ からの出力ベクトル h^1_R を用いて、各データユニット $DU1$ の重み(出力ベクトル h_S)を計算する。サジェスチョンユニット SU における出力ベクトル h_S の計算は、下記式(3)で実施される。

10

【 0 0 2 0 】

【数3】

$$h_S = \text{softmax}(W_S h_R) \dots (3)$$

【 0 0 2 1 】

式(3)中、行列 W_S はサジェスチョンユニット SU の学習パラメータである。下記式(4)に示すように、活性化関数の1つである softmax 関数では、層数 L (本例では $L = 3$)に等しい次元のベクトル h_S を計算する。右辺のベクトル h_R は h^1_R を垂直方向にスタックしたベクトルである。

20

【 0 0 2 2 】

【数4】

$$h_R = [h^1_R; \dots; h^L_R] \dots (4)$$

L=3の例

$$h^1_R = [0, 1, 0]$$

$$h^2_R = [0, 0, 1]$$

$$h^3_R = [1, 0, 0]$$



$$h_R =$$

$$\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

$$h^1_R$$

$$h^2_R$$

$$h^3_R$$

30

40

【 0 0 2 3 】

したがって、行列 W_S は、 L 行 M 列(M は、ベクトル h_R の要素数)の行列となる。サジェスチョンユニット SU に softmax 関数を採用することにより、要素数 L のベクトル h_S の各要素(全要素の和は1)は、対応するデータユニット $DU1$ の重みをあらわすことになる。

【 0 0 2 4 】

識別ユニット IU は、各層のデータユニット $DU1$ ($1 \sim 1$)の出力ベクトル h^1_D を識別する。具体的には、たとえば、識別ユニット IU は、ベクトル h_S により、中間層のデータユニット $DU1$ ($1 \sim 1$)の各出力ベクトル h^1_D を重み付けする。具体的には、たとえば、識別ユニット IU は、中間層のデータユニット $DU1$ ($1 \sim 1$)の出力ベクトル

50

ル h^1_D を統合して統合ベクトル h を算出する。統合ベクトル h の計算は、下記式 (5) で実施される。

【 0 0 2 5 】

【 数 5 】

$$h = \frac{1}{L} \sum_{l=1}^L h_s[l] h_D^l \dots (5)$$

【 0 0 2 6 】

式 (5) 中、右辺のスカラー $h_s[l]$ は、式 (3) に示したベクトル h_s の 1 次元の要素を示す。スカラー $h_s[l]$ により、出力ベクトル h^1_D に重み付けされることになる。つぎに、識別ユニット IU は、最終的なラベルの予測値を示す予測ベクトル t を式 (6) で計算する。

【 0 0 2 7 】

【 数 6 】

$$t = f(x) = \text{softmax}(Wh) \dots (6)$$

【 0 0 2 8 】

式 (6) 中、行列 W は識別ユニット IU の学習パラメータである。活性化関数の 1 つである softmax 関数では、予測値として K 次元の予測ベクトル t を出力する。 K は、上述したラベル y_n の次元数、すなわち、要素数である。なお、予測ベクトル t の各次元には、それぞれ画像 x_n の種別を表す確率値が格納される。

【 0 0 2 9 】

本実施例では、式 (1) から式 (6) の計算方法と、その際に用いた学習パラメータである行列 W^1_D 、行列 W^1_R 、行列 W_S 、行列 W の値を含めて予測モデルと呼称する。予測モデルの学習パラメータは、ニューラルネットワーク 100 に訓練データ $\{x_n, y_n\}$ を与えることにより生成される。

【 0 0 3 0 】

これにより、テストデータを入力する場合、その予測ベクトル t の要素のうち、最大確率の要素に対応する出力ベクトル h^1_D を出力すればよいことになる。すなわち、学習により、正解率の高い階層 1 の重みが高くなるため、階層 1 ごとの重みを予測に反映することにより、パーセプトロンを多層化した場合であってもテストデータの予測値と真値との誤差を小さくすることができ、過学習を抑制することができる。

【 0 0 3 1 】

< システム構成例 >

図 2 は、データ分析システムのシステム構成例を示すブロック図である。図 2 では、サーバ - クライアント型のデータ分析システム 2 を例に挙げて説明するが、スタンドアロン型でもよい。(A) は、データ分析システム 2 のハードウェア構成例を示すブロック図であり、(B) は、データ分析システム 2 の機能的構成例を示すブロック図である。(A) および (B) において同一構成には同一符号を付す。

【 0 0 3 2 】

データ分析システム 2 は、クライアント端末 200 とサーバであるデータ分析装置 220 とがネットワーク 210 で通信可能に接続される構成である。

【 0 0 3 3 】

(A) において、クライアント端末 200 は、補助記憶装置である HDD (hard disk drive) 201、主記憶装置であるメモリ 202、プロセッサ 203、キーボードやマウスである入力装置 204、モニタ 205 を有する。データ分析装置 220 は、補助記憶装置である HDD 221、主記憶装置であるメモリ 222、プロセッサ 223、キーボードやマウスである入力装置 224、モニタ 225 を有する。なお、主記憶装置、補助記憶装置、および、図示しない可搬型の記憶媒体を総称して、記憶デバイスと称

10

20

30

40

50

す。記憶デバイスは、ニューラルネットワーク 100 や後述するハイパーパラメータ計算用ニューラルネットワーク 300 およびこれらの学習パラメータを記憶する。

【0034】

(B)において、クライアント端末 200 は、クライアントデータベース(DB) 251 を有する。クライアントDB 251 は、HDD 201 やメモリ 202 などの記憶デバイスに格納される。クライアントDB 251 には、テストデータ集合 252 と、予測結果 253 と、が格納される。テストデータ集合 252 は、テストデータの集合である。予測結果 253 は、予測部 262 からネットワーク 210 経由で得られたデータである。なお、サーバ-クライアント型の場合、クライアント端末 200 は 1 台以上存在する。

【0035】

データ分析装置 220 は、学習部 261 と、予測部 262 と、サーバデータベース(DB) 263 と、を有する。学習部 261 は、図 1 および図 2 に示した処理を実行し、学習パラメータ 265 を出力する機能部である。学習パラメータ 265 には、上述した学習パラメータが含まれる。

【0036】

予測部 262 は、学習パラメータ 265 を用いて、ニューラルネットワーク 100 を構築し、ニューラルネットワーク 100 にテストデータが与えられることで、予測処理を実行し、予測結果 253 をクライアント端末 200 に出力する機能部である。学習部 261 および予測部 262 は、HDD 221、メモリ 222 などの記憶デバイスに記憶されたプログラムをプロセッサ 223 に実行させることによりその機能を実現する。

【0037】

サーバDB 263 は、訓練データ集合 264 と、学習パラメータ 265 と、を格納する。訓練データ集合 264 は、上述した画像 x_n と、ラベル y_n と、を含む。学習パラメータ 265 は、学習部 261 からの出力データであり、行列 W^1_D 、行列 W^1_R 、行列 W_S 、行列 W を含む。

【0038】

なお、データ分析装置 220 は複数台で構成されてもよい。たとえば、負荷分散のため、データ分析装置 220 が複数存在してもよい。また、データ分析装置 220 は、機能ごとに複数台で構成されてもよい。たとえば、学習部 261 およびサーバDB 263 を含む第 1 のサーバと、予測部 262 およびサーバDB 263 を含む第 2 のサーバとで構成されてもよい。また、また、学習部 261 および予測部 262 を含む第 1 のデータ分析装置と、サーバDB 263 を含む第 2 のデータ分析装置とで構成されてもよい。また、学習部 261 を含む第 1 のデータ分析装置と、予測部 262 を含む第 2 のデータ分析装置と、サーバDB 263 を含む第 3 のデータ分析装置とで構成されてもよい。

【0039】

<データ分析装置の機能的構成例>

図 3 は、データ分析装置 220 の機能的構成例を示すブロック図である。データ分析装置 220 は、図 2 に示したように、学習部 261 および予測部 262 を有する。学習部 261 は、ニューラルネットワーク 100 およびハイパーパラメータ計算用ニューラルネットワーク 300 の各学習パラメータを学習する。予測部 262 は、学習パラメータ 265 が決定されたニューラルネットワーク 100 にテストデータを与えることにより、予測データ t' を算出する。

【0040】

ここで、ハイパーパラメータ計算用ニューラルネットワーク 300 は、一般的なニューラルネットワークでよい。図 3 では、ハイパーパラメータ計算用ニューラルネットワーク 300 は 1 つであるが、ハイパーパラメータ計算用ニューラルネットワーク 300 はハイパーパラメータごとに設けられる。ハイパーパラメータ計算用ニューラルネットワーク 300 は、入力層となる階層 $l' = 0$ のデータユニット du_0 、中間層となる階層 $l' = 1$ (本例では、層数 $L' = 3$) のデータユニット $du_1 \sim du_3$ 、出力層となる識別ユニット iu が直列に接続されて構成される。各層は、パーセプトロンにより構成される。

【 0 0 4 1 】

また、データ分析装置 2 2 0 は、設定部 3 0 1 と、重み付け処理部 3 0 2 と、算出部 3 0 3 と、縮退部 3 0 4 と、を有する。設定部 3 0 1、重み付け処理部 3 0 2、算出部 3 0 3 および縮退部 3 0 4 は、H D D 2 2 1、メモリ 2 2 2 などの記憶デバイスに記憶されたプログラムをプロセッサ 2 2 3 に実行させることによりその機能を実現する。

【 0 0 4 2 】

設定部 3 0 1 は、各中間層のデータユニット D U 1 からの出力データを受け付けて、各出力データと学習パラメータ（行列 W_s ）とに基づいて、各中間層のデータユニット D U 1 の重みを設定して、出力層である識別ユニット I U に出力する。設定部 3 0 1 は、サジェスチョンユニットが実現する機能である。

10

【 0 0 4 3 】

重み付け処理部 3 0 2 は、中間層のデータユニット D U 1 からの出力データを、設定部 3 0 1 によって設定された中間層のデータユニット D U 1 ごとの重みにより、それぞれ重み付けする。算出部 3 0 3 は、重み付け処理部 3 0 2 によって重み付けされた各出力データと学習パラメータ（行列 W ）とに基づいて、予測データ t 、 t' を算出する。重み付け処理部 3 0 2、および算出部 3 0 3 は、識別ユニット I U が実現する機能である。

【 0 0 4 4 】

縮退部 3 0 4 は、各第 1 中間層からの出力データを受け付けて、各出力データの次元数を縮退させて、縮退後の各出力データを設定部 3 0 1 に出力する。縮退部 3 0 4 は、レポータリングユニット R U 1 が実現する機能である。

20

【 0 0 4 5 】

< データ分析処理手順例 >

図 4 は、データ分析装置 2 2 0 によるデータ分析処理手順例を示すフローチャートである。図 4 のうち、ステップ S 4 0 1 ~ S 4 0 4 が学習部 2 6 1 による学習処理であり、ステップ S 4 0 5 ~ S 4 0 7 が予測部 2 6 2 による予測処理である。まず、データ分析装置 2 2 0 は、訓練データ集合 2 6 4 を読み込み（ステップ S 4 0 1）、ハイパーパラメータ算出処理を実行する（ステップ S 4 0 2）。ハイパーパラメータとは、ニューラルネットワークの構造を決定するパラメータである。ハイパーパラメータには、パターン（種類）を決定する第 1 ハイパーパラメータと、順序を決定する第 2 ハイパーパラメータがある。

【 0 0 4 6 】

30

第 1 ハイパーパラメータには、たとえば、活性化関数の種類、損失関数の種類、正則化（関数）の種類、ニューラルネットワーク種などがある。本例では、第 1 ハイパーパラメータとして活性化関数の種類に関するハイパーパラメータ v_A を用いる。第 2 ハイパーパラメータには、中間層の層数 L 、中間層内のユニット数などがある。本例では、第 2 ハイパーパラメータとして中間層の層数 L に関するハイパーパラメータ v_L を用いる。ハイパーパラメータ算出処理（ステップ S 4 0 2）の詳細は後述する。

【 0 0 4 7 】

データ分析装置 2 2 0 は、ハイパーパラメータ算出処理（ステップ S 4 0 2）で算出されたハイパーパラメータを用いて、ニューラルネットワーク 1 0 0 の構造を決定する（ステップ S 4 0 3）。たとえば、データ分析装置 2 2 0 は、活性化関数の種類に関するハイパーパラメータ v_A と中間層の層数 L に関するハイパーパラメータ v_L とを設定して、ニューラルネットワーク 1 0 0 の構造として、識別ユニット I U の関数 $f(x; v_A, v_L)$ を決定する。

40

【 0 0 4 8 】

データ分析装置 2 2 0 は、構造決定されたニューラルネットワーク 1 0 0 に、訓練データ $\{x_n, y_n\}$ を与えることにより学習をおこない、学習パラメータ 2 6 5（行列 W_D 、行列 W_R 、行列 W_s 、行列 W ）を生成する（ステップ S 4 0 4）。学習（ステップ S 4 0 4）では、たとえば、誤差逆伝播法により、学習パラメータ 2 6 5 を調整する。データ分析装置 2 2 0 は、生成した学習パラメータ 2 6 5 をサーバ D B 2 6 3 に保存する。

【 0 0 4 9 】

50

つぎに、データ分析装置 220 は、テストデータ集合 252 を読み込み（ステップ S 405）、各テストデータの画像 x'_n を、学習パラメータ 265 が反映されたニューラルネットワーク 100 に与えて、予測ベクトル t' を算出する（ステップ S 406）。そして、データ分析装置 220 は、予測ベクトル t' の集合である予測結果 253 をクライアント端末 200 に出力する（ステップ S 407）。クライアント端末 200 は、予測結果 253 をモニタ 205 に表示する。

【0050】

<ハイパーパラメータ算出処理（ステップ S 402）>

ハイパーパラメータ算出処理（ステップ S 402）では、データ分析装置 220 は、図 1 に示したニューラルネットワーク 100 とは別のハイパーパラメータ計算用ニューラルネットワーク 300 を用いて、ハイパーパラメータとして、たとえば、ハイパーパラメータ v_A 、 v_L を算出する。

【0051】

ハイパーパラメータ v_A は、活性化関数の種類として、シグモイド関数、双曲線正接関数、ReLU 関数のいずれかを示す値が one hot ベクトルで格納されたハイパーパラメータである。たとえば、

$v_A = (1, 0, 0)$ であれば、シグモイド関数、

$v_A = (0, 1, 0)$ であれば、双曲線正接関数、

$v_A = (0, 0, 1)$ であれば、ReLU 関数を示す。ハイパーパラメータ v_A は、下記式（7）で計算される。

【0052】

【数 7】

$$v_A = f_A(x) = \text{softmax}(W_A^{l'} \cdots g(W_A^{l'} \cdots g(W_A^0 x) \cdots) \cdots) \quad \dots (7)$$

【0053】

式（7）中、行列 $W^{1'}_A$ は、学習パラメータ 265 の 1 つである。行列 $W^{1'}_A$ の初期値は、ハイパーパラメータ算出処理（ステップ S 402）の実行前にあらかじめ設定されているものとする。 L' は、ハイパーパラメータ v_A 計算用ニューラルネットワーク 300 の層数を表す。本実施例では、 $L' = 3$ とする。層数 L' は任意に設定可能である。右辺の $g(W^0_A)$ は、ハイパーパラメータ v_A 計算用ニューラルネットワーク 300 の入力層（データユニット du_0 ）のパーセプトロン（ $g(\cdot)$ は活性化関数）である。 $g(W^{1'}_A)$ は、ハイパーパラメータ v_A 計算用ニューラルネットワーク 300 の第 $1'$ 層である中間層（データユニット $du_{1'}$ ）のパーセプトロンである。 $f_A(x)$ は、ハイパーパラメータ v_A 計算用ニューラルネットワーク 300 の出力層（識別ユニット iu ）のパーセプトロン（softmax 関数）である。

【0054】

ハイパーパラメータ v_L は、ニューラルネットワーク 100 の中間層の層数 L が区間 $[0, 1]$ に正規化した連続値として格納されたハイパーパラメータである。たとえば、中間層の層数 L が $L = 10$ 、すなわち、最大で中間層を 10 層まで構築可能とする場合、 $1 = 1$ は 0.1、 $1 = 2$ は 0.2、...、 $1 = 10$ は 1.0 として正規化される。ハイパーパラメータ v_L は、下記式（8）で計算される。

【0055】

【数 8】

$$v_L = f_L(x) = W_L^{L'} \cdots g(W_L^{L'} \cdots g(W_L^0 x) \cdots) \quad \dots (8)$$

【0056】

式（8）中、行列 $W^{1'}_L$ は、学習パラメータ 265 の 1 つである。行列 $W^{1'}_L$ の初期値は、ハイパーパラメータ算出処理（ステップ S 402）の実行前にあらかじめ設定さ

10

20

30

40

50

れているものとする。L'は、ハイパーパラメータ v_L 計算用ニューラルネットワーク300の層数を表す。本実施例では、 $L' = 3$ とする。層数 L' は任意に設定可能である。右辺の $g(W_L^0)$ は、ハイパーパラメータ v_L 計算用ニューラルネットワーク300の入力層（データユニット du_0 ）のパーセプトロン（ $g()$ は活性化関数）である。 $g(W^{1'}_L)$ は、ハイパーパラメータ v_L 計算用ニューラルネットワーク300の第 $1'$ 層である中間層（データユニット $du_{1'}$ ）のパーセプトロンである。 $f_L(x)$ は、ハイパーパラメータ v_L 計算用ニューラルネットワークの出力層（識別ユニット iu ）のパーセプトロン（softmax関数）である。

【0057】

なお、ハイパーパラメータ v_L の算出に際し、中間層の層数 L が区間 $[0, 1]$ に正規化されているため、正規化したときの倍率の逆数で非正規化されて、ハイパーパラメータ v_L が得られる。たとえば、中間層の層数 L を $L = 10$ の場合、 $1/10$ として正規化した場合、ハイパーパラメータ v_L は10倍されることで、層数を示すことになる。なお、浮動小数は、切捨てる。このように、ハイパーパラメータ v_A, v_L により識別ユニットIUの関数 $f(x; v_A, v_L)$ を構成することで、予測ベクトル t の正解値が上昇する方向に学習パラメータ（行列 $W^{1'}_A, W^{1'}_L$ ）を決定することができる。

【0058】

また、ハイパーパラメータ算出処理（ステップS402）では、データ分析装置220は、下記式（9）の行動価値関数を用いて、 $Q(v)$ を計算する。

【0059】

【数9】

$$Q(v) = W_Q^{L'} \cdots g(W_Q^{L'} \cdots g(W_Q^0 v) \cdots) \cdots \quad (9)$$

【0060】

式（9）中、行列 $W_Q^{1'}$ 、学習パラメータ265の1つである。行列 $W_Q^{1'}$ の初期値は、ハイパーパラメータ算出処理（ステップS402）の実行前にあらかじめ設定されているものとする。左辺のベクトル v は、下記式（10）に示すように、ハイパーパラメータ v_A, v_L, x を垂直方向にスタックしたベクトルである。

【0061】

【数10】

$$v = [v_A; v_L; x] \quad (10)$$

$$\begin{array}{l} v_A = [0.7, 0.2, 0.1] \\ v_L = 0.6 \\ x = [1, 2, 3, 4] \end{array} \Rightarrow v = \begin{bmatrix} 0.7 \\ 0.2 \\ 0.1 \\ 0.6 \\ 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} \begin{array}{l} \left. \begin{array}{c} 0.7 \\ 0.2 \\ 0.1 \end{array} \right\} v_A \\ \left. \begin{array}{c} 0.6 \end{array} \right\} v_L \\ \left. \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} \right\} x \end{array}$$

【0062】

図5は、図4に示したハイパーパラメータ算出処理（ステップS402）の詳細な処理手順例を示すフローチャートである。データ分析装置220は、ステップS401で読み

10

20

30

40

50

込んだ訓練データ集合 264 を分割する（ステップ S501）。具体的には、たとえば、データ分析装置 220 は、訓練データ集合 264 をランダムに 9：1（分割比率は一例）に分割する。すなわち、訓練データ集合 264 の 90% をハイパーパラメータ用訓練データ集合とし、残余の 10% をハイパーパラメータ用テストデータ集合とする。ハイパーパラメータ用訓練データ集合内のハイパーパラメータ用訓練データを u_t とし、ハイパーパラメータ用テストデータ集合内のハイパーパラメータ用テストデータを s_t とする。

【0063】

つぎに、データ分析装置 220 は、乱数 r を計算し（ステップ S502）、乱数 r がより小さいか否かを判断する（ステップ S503）。 r は、区間 $[0, 1]$ 内の値であり、例として $r = 0.3$ とする。乱数 r がより小さい場合（ステップ S503：Yes）
10、データ分析装置 220 は、ハイパーパラメータ v_A をランダムに生成して、ステップ S506 に移行する（ステップ S504）。一方、乱数 r が以上である場合（ステップ S503：No）、式（7）の $f_A(x)$ により、ハイパーパラメータ v_A を算出して、ステップ S506 に移行する（ステップ S505）。

【0064】

つぎに、データ分析装置 220 は、上記式（8）を用いて、ハイパーパラメータ v_L を、下記式（11）により計算する（ステップ S506）。

【0065】

【数 11】

$$v_L = f_L(x) + N_t \quad \dots (11)$$

20

【0066】

式（11）中、右辺第 2 項の N_t は、区間 $[0, 1]$ の間で発生された乱数である。これにより、ハイパーパラメータ v_A 、 v_L が決まるため、ニューラルネットワーク 100 の構造として、識別ユニット IU の関数 $f(x; v_A, v_L)$ が仮決定する。

【0067】

つぎに、データ分析装置 220 は、仮決定された識別ユニット IU の関数 $f(x; v_A, v_L)$ を学習する（ステップ S507）。具体的には、たとえば、データ分析装置 220 は、下記式（12）に示すように、仮決定された $f(x; v_A, v_L)$ に関する交差エ
30ントロピー E に、ハイパーパラメータ用訓練データ集合の各ハイパーパラメータ用訓練データ u_t を与え、交差エントロピー E を最小化する処理を実行する。当該最小化する処理には、確率的最急降下法が用いられる。これにより、仮決定された識別ユニット IU の関数 $f(x; v_A, v_L)$ の学習パラメータ 265（行列 W^1_D 、行列 W^1_R 、行列 W_S 、行列 W ）が更新される。

【0068】

【数 12】

$$E = - \sum_{n=1}^N \sum_{k=1}^K t_n[k] \log(y_n[k]) \quad \dots (12)$$

40

【0069】

式（12）中、 $t_n[k]$ は、ハイパーパラメータ用訓練データ u_t である訓練データ x_n についての予測ベクトル t の k 次元目の要素である。たとえば、予測ベクトル t_n が $t_n = [0, 0, 1]$ である場合、 $t_n[k=1] = 0$ 、 $t_n[k=2] = 0$ 、 $t_n[k=3] = 1$ となる。同様に、 $y_n[k]$ は、ハイパーパラメータ用訓練データ u_t である訓練データ x_n に対応するラベル y_n の k 次元目の要素である。たとえば、ラベル y_n が $t_n = [0, 0, 1]$ である場合、 $y_n[k=1] = 0$ 、 $y_n[k=2] = 0$ 、 $y_n[k=3] = 1$ となる。

【0070】

50

つぎに、データ分析装置 220 は、予測値を示す予測ベクトル t と真値を示すラベル y_n とから報酬 r を AUC (Area Under Curve) として計算する (ステップ S508)。なお、ステップ S507 の予測ベクトル t は、学習 (ステップ S507) 後の $f(x; v_A, v_L)$ にハイパーパラメータ用テストデータ s_t を与えることで計算される。

【0071】

具体的には、たとえば、データ分析装置 220 は、縦軸を真陽性 (True Positive) の確率、横軸を偽陽性 (False Positive) としたグラフ上の ROC (Receiver Operating Characteristic) 曲線を生成し、曲線下の面積を報酬 r として算出する。

10

【0072】

真陽性は、陽性者を陽性と判定する性質であり、この場合は、ラベル y_n 内の要素である真値、すなわち、正解を示す値 (1.0) に対応する予測ベクトル t の要素が正解であると判定することである。たとえば、正解を示す値 (1.0) に対応する予測ベクトル t の要素が「0.6」であり、しきい値が「0.5」であるとする、当該要素「0.6」は正解を示すため、真陽性を示す。

【0073】

偽陽性は、陰性者を陽性と判定する性質であり、ラベル y_n 内の要素である偽値、すなわち、不正解を示す値 (0.0) に対応する予測ベクトル t の要素が正解であると判定することである。たとえば、不正解を示す値 (0.0) に対応する予測ベクトル t の要素が「0.6」であり、しきい値が「0.5」であるとする、当該要素「0.6」は正解を示すため、偽陽性を示す。

20

【0074】

なお、横軸は、偽陰性 (False Negative) の確率でもよい。偽陰性とは、陽性者を陰性と判定する性質であり、ラベル y_n 内の要素である真値、すなわち、正解を示す値 (1.0) に対応する予測ベクトル t の要素が不正解であると判定することである。たとえば、正解を示す値 (1.0) に対応する予測ベクトル t の要素が「0.4」であり、しきい値が「0.5」であるとする、当該要素「0.4」は不正解を示すため、偽陰性を示す。

【0075】

つぎに、データ分析装置 220 は、訓練データ集合 264 をランダムに 9:1 に分割することで、90% のデータをハイパーパラメータ用訓練データ u_{t+1} とし、10% のデータをハイパーパラメータ用テストデータ s_{t+1} をとし、データセット (s_t, v_t, r_t, s_{t+1}) をメモリ 222 に保存する (ステップ S509)。

30

【0076】

s_t は、ハイパーパラメータ用テストデータである。 v_t は、式 (10) において、1 次元ベクトル x をハイパーパラメータ用訓練データ u_t に置き換えて生成されたベクトルである。 r_t は、予測ベクトル t の算出元となった報酬 r である。

【0077】

つぎに、データ分析装置 220 は、 I 個のデータセット (s_i, v_i, r_i, s_{i+1}) をメモリ 222 からランダムに取得する (ステップ S510)。

40

【0078】

データ分析装置 220 は、下記式 (13) によりロス関数 E_Q を計算する (ステップ S511)。

【0079】

【数 13】

$$E_Q = \frac{1}{I} \sum_{i=1}^I (q_i - Q(v_i))^2 \dots (13)$$

50

【 0 0 8 0 】

式 (1 3) 中、右辺の $Q (v_i)$ は、上述した式 (9) を、ステップ S 5 1 0 で取得したデータセット (s_i, v_i, r_i, s_{i+1}) 内の v_i に適用した値である。また、右辺の q_i は、下記式 (1 4) により計算される。

【 0 0 8 1 】

【 数 1 4 】

$$q_i = r_i + \gamma Q(v_i) \dots (14)$$

【 0 0 8 2 】

式 (1 4) 中、右辺の γ は、割引率であり、区間 [0 . 1] の間で設定される値である。

10

【 0 0 8 3 】

データ分析装置 2 2 0 は、確率的勾配降下法を式 (1 3) のロス関数 E_Q に適用することで、式 (9) の学習パラメータ W^1_Q を更新する (ステップ S 5 1 2)。

【 0 0 8 4 】

そして、データ分析装置 2 2 0 は、ロス関数 E_Q をハイパーパラメータ v_A で偏微分した勾配を用いて、勾配法により式 (7) の $f_A (x)$ の学習パラメータ W_A を更新するとともに、ロス関数 E_Q をハイパーパラメータ v_L で偏微分した勾配を用いて、勾配法により式 (8) の $f_L (x)$ の学習パラメータ W_L を更新する (ステップ S 5 1 3)。

【 0 0 8 5 】

20

このあと、データ分析装置 2 2 0 は、ステップ S 5 0 1 からの処理を所定の繰り返し回数 (たとえば、1 0 0 0 回) 分実行したか否かを判断する (ステップ S 5 1 4)。所定の繰り返し回数分実行していない場合 (ステップ S 5 1 4 : N o)、ステップ S 5 0 1 に戻る。一方、所定の繰り返し回数分実行した場合 (ステップ S 5 1 4 : Y e s)、データ分析装置 2 2 0 は、訓練データ x_n を式 (7) の $f_A (x)$ に与えることで、最適なハイパーパラメータ v_A を算出し、訓練データ x_n を式 (7) の $f_L (x)$ に与えることで、最適なハイパーパラメータ v_L を算出する (ステップ S 5 1 5)。

【 0 0 8 6 】

具体的には、たとえば、データ分析装置 2 2 0 は、各訓練データ x_n を式 (7) の $f_A (x)$ に与えることで、訓練データ x_n ごとのハイパーパラメータ v_A を算出する。そして、データ分析装置 2 2 0 は、訓練データ x_n ごとのハイパーパラメータ v_A の統計値を算出し、当該統計値を最適なハイパーパラメータ v_A とする。統計値とは、訓練データ x_n ごとのハイパーパラメータ v_A の中の中央値、平均値、最小値、または最大値である。最適なハイパーパラメータ v_L についても同様である。これにより、最適なハイパーパラメータ (v_A, v_L) が決定され、ステップ S 4 0 3 に移行する。

30

【 0 0 8 7 】

< データ分析結果 >

図 6 は、データ分析結果の比較例を示す説明図である。図 6 のグラフの縦軸は、A U C、すなわち、真陽性の確率である。具体的には、ラベル y_n 内の要素である正解を示す値 (1 . 0) に対応する予測ベクトル t の要素が正解である確率である。手法 A は、従来法、たとえば、特許文献 1 のニューラルネットワークを用いたデータ分析であり、手法 B は、本実施例のニューラルネットワーク 1 0 0 を用いたデータ分析である。なお、手法 A および手法 B とともに、ハイパーパラメータとして、層数 L を 1 0 層、活性化関数をシグモイド関数とする。

40

【 0 0 8 8 】

ここでは、クレジットカードの入会信用審査結果を予測するため、 x_n として、画像ではなく、クレジットカードの審査情報とした。審査情報 x_n は、被審査者の収入情報などを格納した 1 5 次元のベクトルデータとした。また、ラベル y_n は、審査の合否 (0 : 不合格、1 : 合格) を one h o t 表現で表したベクトルである。なお、訓練データ数 N は、 $N = 6 2 1$ 、テストデータ数 N は、 $N = 6 9$ とした。手法 B は、手法 A と比較して、約

50

30%の精度向上を実現する。

【実施例2】

【0089】

実施例2は、実施例1のニューラルネットワーク100の構造の変形例を示す。実施例2のニューラルネットワークは、実施例1に比べて簡素な構成となるため、計算量が低減し、学習処理の高速化を実現する。なお、実施例2では、実施例1との相違点を中心に説明するため、実施例1と同一内容については説明を省略する。

【0090】

図7は、実施例2にかかるニューラルネットワークの構造例を示す説明図である。ニューラルネットワーク700には、レポーティングユニットRULが存在しない。このため、実施例2では、式(2)は不要となる。したがって、実施例2では、式(3)のかわりに、下記式(15)が適用される。

【0091】

【数15】

$$h_s = \text{softmax}(W_s h_D) \dots (15)$$

【0092】

このあと、実施例1と同様、式(5)および式(6)を適用することにより、データ分析装置220は、予測ベクトルtを算出することができる。このように、レポーティングユニットRULによる計算処理が不要となるため、学習処理の高速化を実現することができる。

【実施例3】

【0093】

実施例3は、実施例1のニューラルネットワーク100の構造の変形例を示す。実施例3のニューラルネットワークは、実施例1に比べて簡素な構成となるため、計算量が低減し、学習処理の高速化を実現する。なお、実施例3では、実施例1との相違点を中心に説明するため、実施例1と同一内容については説明を省略する。

【0094】

図8は、実施例3にかかるニューラルネットワークの構造例を示す説明図である。ニューラルネットワーク800には、レポーティングユニットRULが存在しない。このため、実施例3では、式(2)は不要となる。

【0095】

また、中間層のデータユニットDUL(1__1)からの出力データは、識別ユニットIUにのみ入力される。入力層のデータユニットDU0からの訓練データ x_n などの出力データは、サジェスチョンユニットSUに入力される。したがって、実施例2では、式(3)、式(4)のかわりに、下記式(16)が適用される。

【0096】

【数16】

$$h_s = \text{softmax}(W_s x) \dots (16)$$

【0097】

このあと、実施例1と同様、式(5)および式(6)を適用することにより、データ分析装置220は、予測ベクトルtを算出することができる。このように、レポーティングユニットRULによる計算処理が不要となるため、学習処理の高速化を実現することができる。また、サジェスチョンユニットSUへの入力、入力層のデータユニットDU0からの訓練データ x_n のみとなるため、サジェスチョンユニットSUの計算負荷の低減を図り、これにより、学習処理の効率化を図ることができる。

【実施例4】

【0098】

実施例4は、実施例1のニューラルネットワーク100の構造の変形例を示す。実施例

10

20

30

40

50

4のニューラルネットワークは、実施例1にさらにレポーティングユニットR U 1と同様、次元縮退させるユニットを追加した構成となるため、学習処理の効率化を実現する。なお、実施例4では、実施例1との相違点を中心に説明するため、実施例1と同一内容については説明を省略する。

【0099】

図9は、実施例4にかかるニューラルネットワークの構造例を示す説明図である。ニューラルネットワーク900は、実施例1のニューラルネットワークに、ハーモナイジングユニットが追加された構成となる。ハーモナイジングユニットH U 1 (1__1) は、中間層のデータユニットD U 1 (1__1) と識別ユニットとの間に、中間層のデータユニットD U 1 (1__1) ごとに設けられる。ハーモナイジングユニットH U 1 は、レポーティングユニットR U 1 と同様、中間層のデータユニットD U 1 (1__1) からの出力データを次元縮退させる縮退部901である。したがって、識別ユニットI U には、ハーモナイジングユニットH U 1 で次元縮退されたデータが入力される。

10

【0100】

ハーモナイジングユニットH U 1 は、同一階層のデータユニットD U 1 からの出力ベクトル h^1_D を入力し、出力ベクトル h^1_D の次元数を縮退させる。ハーモナイジングユニットH U 1 からの出力ベクトル h^1_H の計算は、下記式(17)で実施される。

【0101】

【数17】

$$h^l_H = g(W^l_H h^l_D) \dots (17)$$

20

【0102】

式(17)中、行列 W^1_H はハーモナイジングユニットH U 1 の学習パラメータである。具体的には、たとえば、データユニットD U 1 からの出力ベクトル h^1_D の次元数をdとした場合、行列 W^1_H は、たとえば、1行目からm ($m < d$) 行目までの成分を行列 W^1_D と同一成分とし、d - m + 1行からk行目までの成分を0とした行列である。これにより、データユニットD U 1 からのd次元の出力ベクトル h^1_D は、m次元の出力ベクトル h^1_H に縮退される。mは、次元縮退を決定するハイパーパラメータである。dおよびmは、レポーティングユニットR U 1 のdおよびmと異なる値でもよい。

30

【0103】

また、識別ユニットI U は、ハーモナイジングユニットH U 1 からの出力ベクトル h^1_H を入力するため、式(5)のかわりに、下記式(18)が適用される。

【0104】

【数18】

$$h = \frac{1}{L} \sum_{l=1}^L h_s[l] h^l_H \dots (18)$$

【0105】

このあと、実施例1と同様、式(6)を適用することにより、データ分析装置220は、予測ベクトルtを算出することができる。このように、ハーモナイジングユニットH U 1 を適用することにより、識別ユニットI U に入力されるデータユニットからの出力ベクトル h^1_D が次元縮退されるため、識別ユニットI U の計算負荷の低減を図り、学習処理の効率化を図ることができる。

40

【実施例5】

【0106】

実施例5は、実施例1～4において、式(6)のf(x)の計算方法として回帰を実行する例である。実施例5の訓練データおよびテストデータは医療用画像とし、データ分析装置220は、手術後の患者の再発までの期間を予測するものとする。

50

【 0 1 0 7 】

訓練データは、医療用画像 x_n と再発までの期間 y_n とから構成されている ($n = 1, 2, \dots, N$ 。 N は画像の枚数)。医療用画像 x_n は 2 次元のマトリックス構造を持つデータであり、以降、行列として取り扱う。単に「 x 」と表記した場合は、行列形式の医療用画像 x_n をラスタスキャンした 1 次元ベクトルを示す。なお、テストデータについても同様の構成である。再発までの期間 y_n は、年齢層別の再発までの期間を格納した K 次元のベクトルである。 K は、年齢層の種類数を示す。

【 0 1 0 8 】

実施例 5 では、式 (6) に替えて、下記式 (1 9) が適用される。 t はスカラーであり、再発までの予測期間を示す。

10

【 0 1 0 9 】

【数 1 9 】

$$t = f(x) = Wh \dots (19)$$

【 0 1 1 0 】

また、実施例 5 では、図 5 のステップ S 5 0 7 において、データ分析装置 2 2 0 は、ハイパーパラメータ v_A, v_L により構成された式 (1 9) の $f(x)$ 、すなわち、仮決定された識別ユニット I U の関数 $f(x; v_A, v_L)$ を学習する (ステップ S 5 0 7)。

【 0 1 1 1 】

具体的には、たとえば、データ分析装置 2 2 0 は、下記式 (2 0) に示すように、仮決定された識別ユニット I U の関数 $f(x; v_A, v_L)$ に関する二乗和誤差 E に、ハイパーパラメータ用訓練データ集合の各ハイパーパラメータ用訓練データ u_t を与え、二乗和誤差 E を最小化する処理を実行する。当該最小化する処理には、確率的最急降下法が用いられる。これにより、式 (1 9) で仮決定された識別ユニット I U の関数 $f(x; v_A, v_L)$ の学習パラメータ 2 6 5 (行列 W^1_D , 行列 W^1_R , 行列 W^1_H , 行列 W_S , 行列 W) が更新される。

20

【 0 1 1 2 】

【数 2 0 】

$$E = \sum_{n=1}^N (y_n - t_n)^2 \dots (20)$$

30

【 0 1 1 3 】

また、ステップ S 5 0 8 の報酬 r の計算では、データ分析装置 2 2 0 は、下記式 (2 1) により、予測値である再発までの予測期間 t と真値である再発までの期間 y_n とから、報酬 r を決定係数 R^2 として計算する。

【 0 1 1 4 】

【数 2 1 】

$$r = R^2 = 1 - \frac{\sum (y_n - f_n)^2}{\sum_n (y_n - y_{av})^2} \dots (21)$$

40

【 0 1 1 5 】

以上説明したように、本実施例にかかるデータ分析装置 2 2 0 は、第 1 入力層 (データユニット D U 0) と、第 1 出力層 (識別ユニット I U) と、第 1 入力層と第 1 出力層との間において前段の層からのデータと第 1 学習パラメータ (行列 W^1_D) とを第 1 活性化関数 ($g()$) に与えて計算して後段の層に計算結果を出力する 2 層以上の第 1 中間層 (データユニット D U 1) と、により構成される第 1 ニューラルネットワーク (ニューラルネットワーク 7 0 0) を用いる。

【 0 1 1 6 】

50

データ分析装置 220 は、設定部 301 と、重み付け処理部 302 と、算出部 303 と、を有する。設定部 301 は、各第 1 中間層からの出力データ（出力ベクトル h^1_D ）を受け付けて、各出力データと第 2 学習パラメータ（行列 W_s ）とに基づいて、各第 1 中間層の重み（出力ベクトル h_s ）を設定して、第 1 出力層に出力する（サジェスチョンユニット SU）。

【0117】

重み付け処理部 302 は、第 1 出力層に含まれ、各出力データを、設定部 301 によって設定された第 1 中間層ごとの重みにより、それぞれ重み付けする（識別ユニット IU）。算出部 303 は、第 1 出力層に含まれ、重み付け処理部 302 によって重み付けされた各出力データ（ベクトル h ）と第 3 学習パラメータ（行列 W ）とに基づいて、予測データ（予測ベクトル t ）を算出する。

10

【0118】

これにより、階層 1 ごとの重みを予測に反映することにより、中間層を多層化した場合であってもテストデータの予測値と真値との誤差を小さくすることができ、過学習を抑制することができる。

【0119】

また、設定部 301 は、ニューラルネットワーク 800 に示したように、第 1 入力層からの出力データ（出力ベクトル $h^0_D = x_n$ ）を受け付けて、当該出力データと第 2 学習パラメータとに基づいて、各第 1 中間層の重みを設定して、第 1 出力層に出力してもよい。これにより、設定部 301 であるサジェスチョンユニット SU の計算負荷の低減を図り、学習処理の効率化を図ることができる。

20

【0120】

また、ニューラルネットワーク 100 の場合、データ分析装置 220 は、第 1 縮退部 304 を有する（レポーティングユニット RU1）。第 1 縮退部 304 は、各第 1 中間層からの出力データを受け付けて、各出力データの次元数を縮退させて、縮退後の各出力データを設定部 301 に出力する。そして、設定部 301 は、第 1 縮退部 304 からの縮退後の各出力データを受け付けて、縮退後の各出力データと第 2 学習パラメータとに基づいて、各第 1 中間層の重みを設定して、第 1 出力層に出力する。これにより、各第 1 中間層からの出力データのデータ量を削減することで、設定部 301 での重み設定の負荷低減を図ることができる。

30

【0121】

また、データ分析装置 220 は、第 1 入力層に訓練データが与えられた場合に、たとえば、誤差逆伝播法により、第 1 学習パラメータ、第 2 学習パラメータ、および第 3 学習パラメータを調整する学習部 261 を有する。したがって、学習パラメータの学習により、正解率の高い階層 1 の重みが高くなる。これにより、階層 1 ごとの重みを予測に反映することにより、中間層を多層化した場合であってもテストデータの予測値と真値との誤差を小さくすることができ、過学習を抑制することができる。

【0122】

また、ニューラルネットワーク 900 の場合、データ分析装置 220 は、第 2 縮退部 901 を有する（ハーモナイジングユニット HU1）。第 2 縮退部 901 は、各第 1 中間層からの出力データを受け付けて、各出力データの次元数を縮退させて、縮退後の各出力データを重み付け処理部 302 に出力する。そして、重み付け処理部 302 は、第 2 縮退部 901 からの縮退後の各出力データを、第 1 中間層ごとの重みにより、それぞれ重み付けする。これにより、各第 1 中間層からの出力データのデータ量を削減することで、重み付け処理部 302 での重み付け処理の負荷低減を図ることができる。

40

【0123】

また、学習部 261 は、第 2 ニューラルネットワーク（ハイパーパラメータ計算用ニューラルネットワーク 300）を用いる。第 2 ニューラルネットワークは、訓練データを受け付ける第 2 入力層（データユニット du_0 ）と、第 1 ニューラルネットワークのハイパーパラメータ（ v_A, v_L ）を出力する第 2 出力層（識別ユニット iu ）と、第 2 入力層

50

と第2出力層との間において前段の層からのデータと第4学習パラメータ（行列 $W^{1 \rightarrow A}$ 、 $W^{1 \rightarrow L}$ ）とを第2活性化関数（ $g(\cdot)$ ）に与えて計算して後段の層に計算結果を出力する第2中間層（データユニット d_{ul} ）と、により構成される。

【0124】

学習部261は、第2ニューラルネットワークを用い、第2入力層に訓練データが与えられた場合に、たとえば、確率的最急降下法により、第4学習パラメータを調整し、訓練データを、第4学習パラメータが調整された第2ニューラルネットワークの第2入力層に与えることにより、出力層からハイパーパラメータを出力し、ハイパーパラメータにより第1ニューラルネットワークの構造を決定し、構造が決定された第1ニューラルネットワークの第1入力層に訓練データが与えられた場合に、たとえば、誤差逆伝播法により、第1学習パラメータ、第2学習パラメータ、および第3学習パラメータを調整する。これにより、過学習を抑制する最適なハイパーパラメータを自動設定することができる。

10

【0125】

また、ハイパーパラメータは、第1ニューラルネットワークを構成する要素のパターンを決定するハイパーパラメータでもよい。これにより、過学習を抑制する第1ニューラルネットワークを要素のパターンを自動設定することができる。特に、第1活性化関数の種類を決定することにより、ユーザは、どの活性化関数を適用すべきかを考慮する必要がなくなり、ニューラルネットワークの構築の容易化を図ることができる。

【0126】

また、ハイパーパラメータは、第1ニューラルネットワークを構成する要素の順序を決定するハイパーパラメータでもよい。これにより、過学習を抑制する第1ニューラルネットワークを要素の順序を自動設定することができる。特に、第1中間層の層数を決定することにより、中間層をどの程度増減すればよいかを考慮する必要がなくなり、ニューラルネットワークの構築の容易化を図ることができる。

20

【0127】

なお、本発明は前述した実施例に限定されるものではなく、添付した特許請求の範囲の趣旨内における様々な変形例及び同等の構成が含まれる。例えば、前述した実施例は本発明を分かりやすく説明するために詳細に説明したものであり、必ずしも説明した全ての構成を備えるものに本発明は限定されない。また、ある実施例の構成の一部を他の実施例の構成に置き換えてもよい。また、ある実施例の構成に他の実施例の構成を加えてもよい。また、各実施例の構成の一部について、他の構成の追加、削除、または置換をしてもよい。

30

【0128】

また、前述した各構成、機能、処理部、処理手段等は、それらの一部又は全部を、例えば集積回路で設計する等により、ハードウェアで実現してもよく、プロセッサがそれぞれの機能を実現するプログラムを解釈し実行することにより、ソフトウェアで実現してもよい。

【0129】

各機能を実現するプログラム、テーブル、ファイル等の情報は、メモリ、ハードディスク、SSD（Solid State Drive）等の記憶装置、又は、IC（Integrated Circuit）カード、SDカード、DVD（Digital Versatile Disc）の記録媒体に格納することができる。

40

【0130】

また、制御線や情報線は説明上必要と考えられるものを示しており、実装上必要な全ての制御線や情報線を示しているとは限らない。実際には、ほとんど全ての構成が相互に接続されていると考えてよい。

【符号の説明】

【0131】

2 データ分析システム

100 ニューラルネットワーク

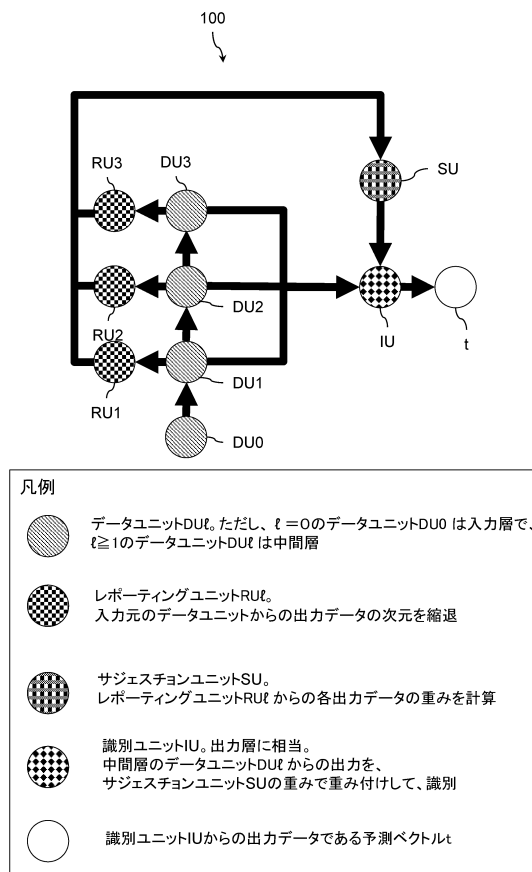
50

200 クライアント端末
 220 データ分析装置
 261 学習部
 262 予測部
 300 ハイパーパラメータ計算用ニューラルネットワーク
 301 設定部
 302 重み付け処理部
 303 算出部
 304 縮退部
 700 ニューラルネットワーク
 800 ニューラルネットワーク
 900 ニューラルネットワーク
 901 縮退部
 DU0 データユニット
 DU1 データユニット
 HU1 ハーモナイジングユニット
 IU 識別ユニット
 RU1 レポートニングユニット
 SU サジェスションユニット
 V_A, V_L ハイパーパラメータ

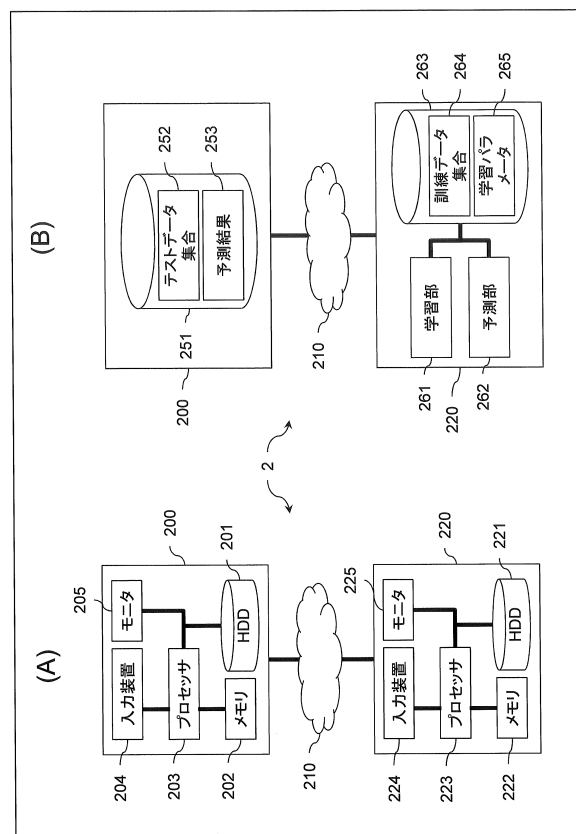
10

20

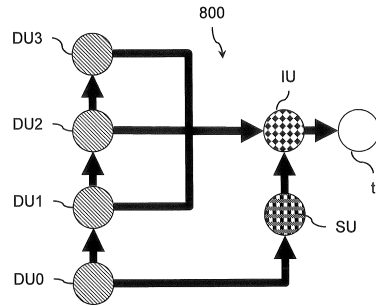
【図1】



【図2】



【 図 8 】



【 図 9 】

