

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6208552号
(P6208552)

(45) 発行日 平成29年10月4日 (2017. 10. 4)

(24) 登録日 平成29年9月15日 (2017. 9. 15)

(51) Int. Cl.

F I

G 0 6 N 99/00 (2010.01)

G 0 6 N 99/00 1 5 3

請求項の数 7 (全 17 頁)

(21) 出願番号	特願2013-235810 (P2013-235810)	(73) 特許権者	502324066
(22) 出願日	平成25年11月14日 (2013. 11. 14)		株式会社デンソーアイティーラボラトリ
(65) 公開番号	特開2015-95212 (P2015-95212A)		東京都渋谷区渋谷2-15-1 渋谷クロ
(43) 公開日	平成27年5月18日 (2015. 5. 18)		スタワー28F
審査請求日	平成28年3月18日 (2016. 3. 18)	(73) 特許権者	000004260
			株式会社デンソー
			愛知県刈谷市昭和町1丁目1番地
		(74) 代理人	100113549
			弁理士 鈴木 守
		(74) 代理人	100115808
			弁理士 加藤 真司
		(72) 発明者	佐藤 育郎
			東京都渋谷区渋谷二丁目15番1号 渋谷
			クロスタワー28F 株式会社デンソーア
			イティーラボラトリ内
			最終頁に続く

(54) 【発明の名称】 識別器、識別プログラム、及び識別方法

(57) 【特許請求の範囲】

【請求項 1】

教師あり学習に基づく識別器であって、
 識別の対象となる未知データに対してデータ拡張を行って複数の疑似未知データを生成するデータ拡張部と、
 前記複数の疑似未知データを識別モデルに適用し、それらの結果を統合してクラス分類を行う識別部と、
 を備えたことを特徴とする識別器。

【請求項 2】

前記データ拡張部は、前記識別モデルを生成する際に訓練データに対して行ったデータ
 拡張と同じ方法で前記未知データに対してデータ拡張を行うことを特徴とする請求項 1 に
 記載の識別器。

【請求項 3】

前記識別部は、前記複数の疑似未知データを前記識別モデルに適用した結果の期待値に
 基づいてクラス分類を行うことを特徴とする請求項 1 又は 2 に記載の識別器。

【請求項 4】

前記識別部は、前記未知データを前記識別モデルに適用することなく、前記クラス分類
 を行うことを特徴とする請求項 1 ないし 3 のいずれか一項に記載の識別器。

【請求項 5】

前記データ拡張部は、乱数を用いて前記未知データに対してデータ拡張を行うことを特

10

20

徴とする請求項 1 ないし 4 のいずれか一項に記載の識別器。

【請求項 6】

コンピュータを、
教師あり学習に基づく識別器であって、
識別の対象となる未知データに対してデータ拡張を行って複数の疑似未知データを生成するデータ拡張部と、
前記複数の疑似未知データを識別モデルに適用し、それらの結果を統合してクラス分類を行う識別部と、
を備えた識別器として機能させることを特徴とする識別プログラム。

【請求項 7】

教師あり学習に基づく識別方法であって、
識別の対象となる未知データに対してデータ拡張を行って複数の疑似未知データを生成するデータ拡張ステップと、
前記複数の疑似未知データを識別モデルに適用し、それらの結果を統合してクラス分類を行う識別ステップと、
を含むことを特徴とする識別方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、教師あり学習に基づく識別装置、識別プログラム、及び識別方法に関し、より詳しくは、訓練データに対してデータ拡張を行って学習をして生成した識別モデルを用いる識別装置、識別プログラム、及び識別方法に関するものである。

【背景技術】

【0002】

教師あり学習に基づく識別器を構築するには、ターゲット値を伴った訓練データ（教師データともいう）を収取して、それらの入出力関係を機械学習の枠組みによって学習する必要がある。ターゲット値とは、訓練データの出力のことであり、学習時にはある訓練データを入力させたときに識別器の出力がその訓練データに対応するターゲット値に近づくように学習パラメタの探索が行われる。

【0003】

このような学習を経て得られた識別器は、その運用時において、学習データには含まれないもののパターンの似ている未知のデータに対して識別を行うことになる。このような識別の対象となる未知のデータ（以下、単に「未知データ」という。）に対する識別能力を汎化能力という。識別器は、高い汎化能力を有することが望まれる。

【0004】

一般に、訓練データは多ければ多いほど、そのような訓練データを用いて学習された識別器の汎化能力は高くなる。しかし、訓練データの収集には人的コストが発生するため、少ない量の訓練データで高い汎化能力を持たせたいという要求がある。つまり訓練データの分布の密度の低さに対する対処が必要となる。

【0005】

そこで、P. Y. Simard, D. Steinkraus, J. C. Platt, "Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis", ICDAR 2003.（非特許文献 1）やCiresan et al., "Deep Big Simple Neural Nets Excel on Handwritten Digit Recognition", Neural Computation 2010.（非特許文献 2）に記載されているようなデータ拡張と呼ばれる発見的方法が提案されてきた。データ拡張とは、サンプルとして与えられているデータに対し、パラメトリックな変形を施すことにより、データの種類を増幅することである。ただし、これらの変形は、元となるデータの所属するクラスに特有の特徴を損なうものであってはならない。

【0006】

非特許文献 1 には、畳み込みニューラルネットワーク（CNN：Convolutional Neural

10

20

30

40

50

Network)を使った手書き数字認識の研究が記載されている。非特許文献1では、訓練データに対し「弾性歪み」という変換を施すことにより人為的に大量のデータを生成し(データ拡張)、それを学習する。非特許文献1には、このような学習により、データ拡張のない場合と比較して飛躍的に高い識別性能が得られることが記載されている。

【0007】

また、非特許文献2には、ニューラルネットワークを使った手書き数字認識の研究が記載されている。非特許文献2では、弾性歪みに加え、回転やスケールの変換を施すことによりデータ拡張を行うことで、非常に高い認識性能を持つことが記載されている。

【0008】

このように、非特許文献1や非特許文献2では、手書き数字認識の問題において、局所的な弾性歪み、微小回転、微小スケール変化といった変形を適用することにより、数字の特徴を失わないデータ拡張を可能とし、拡張のない場合と比べて高い汎化能力を持たせることに成功している。なお、データ拡張による学習をした上で未知データの識別を行うことは、特に画像認識の分野において慣用されている手法である。

【先行技術文献】

【非特許文献】

【0009】

【非特許文献1】P. Y. Simard, D. Steinkraus, J. C. Platt, "Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis", ICDAR 2003.

【非特許文献2】Ciresan et al., "Deep Big Simple Neural Nets Excel on Handwritten Digit Recognition", Neural Computation 2010.

【発明の概要】

【発明が解決しようとする課題】

【0010】

本発明は、訓練データを拡張して学習をした上で、未知の入力データの識別をするに際して、入力データをどのクラスに割り当てるかに関するルール(決定則)を改良して、識別性能を向上することを目的とする。

【課題を解決するための手段】

【0011】

従来、データ拡張を行った場合においても、データ拡張を行わない場合においても、決定則そのものは同一のものを使用してきた。本発明では、データ拡張を行った場合とデータ拡張を行わない場合とでは、理論的に最適な決定則が異なるとの知見に基づき、データ拡張時の改良された識別手法を提供する。

【0012】

本発明の第1の態様は、教師あり学習に基づく識別器であって、未知データに対してデータ拡張を行うデータ拡張部と、前記データ拡張部にて拡張された未知データを識別モデルに適用し、それらの結果を統合してクラス分類を行う識別部とを備えたことを特徴とする。

【0013】

この構成により、未知データが拡張されて複数の疑似未知データが生成され、それらの識別結果が統合されてクラス分類がされるので、未知データそのものを識別する場合と比較して、識別能力が向上する。

【0014】

本発明の第2の態様の識別器は、第1の態様の識別器において、前記データ拡張部は、前記識別モデルを生成する際に訓練データに対して行ったデータ拡張と同じ方法で前記未知データに対してデータ拡張を行うことを特徴とする。

【0015】

この構成により、識別モデルを生成する際の訓練データの拡張と同じ方法で未知データの拡張が行われるので、その分布がクラスの事後分布に重なる可能性が高まり、識別モデルを生成する際に訓練データについてデータ拡張を行った場合における識別能力が向上す

10

20

30

40

50

る。

【0016】

本発明の第3の態様の識別器は、第1又は第2の態様の識別器において、前記識別部は、前記拡張された未知データを前記識別モデルに適用した結果の期待値に基づいてクラス分類を行うことを特徴とする。

【0017】

この構成により、識別モデルを生成する際の目的関数を最小にすることを決定則としてクラス分類が行われるので、識別モデルを生成する際に訓練データについてデータ拡張を行った場合における識別能力が向上する。

【0018】

本発明の第4の態様の識別器は、第1ないし第3のいずれかの態様の識別器において、前記識別部は、前記未知データを前記識別モデルに適用することなく、前記クラス分類を行うことを特徴とする。

【0019】

この構成により、未知データ自体は識別に用いられることなく、当該未知データのクラス分類が行われる。

【0020】

本発明の第5の態様の識別器は、第1ないし第4のいずれかの態様の識別器において、前記データ拡張部は、乱数を用いて前記未知データに対してデータ拡張を行うことを特徴とする。

【0021】

この構成により、未知データが乱数を用いてデータ拡張されるので、その分布がクラスの事後分布に重なる可能性が高まり、識別モデルを生成する際に訓練データについてデータ拡張を行った場合における識別能力が向上する。

【0022】

本発明の第6の実施の形態は、識別プログラムであって、コンピュータを、教師あり学習に基づく識別器であって、未知データに対してデータ拡張を行うデータ拡張部と、前記データ拡張部にて拡張された未知データを識別モデルに適用し、それらの結果を統合してクラス分類を行う識別部とを備えた識別器として機能させることを特徴とする。

【0023】

この構成によっても、未知データが拡張されて複数の疑似未知データが生成され、それらの識別結果が統合されてクラス分類がされるので、未知データそのものを識別する場合と比較して、識別能力が向上する。

【0024】

本発明の第7の実施の形態は、教師あり学習に基づく識別方法であって、未知データに対してデータ拡張を行うデータ拡張ステップと、前記データ拡張部にて拡張された未知データを識別モデルに適用し、それらの結果を統合してクラス分類を行う識別ステップとを含むことを特徴とする。

【0025】

この構成によっても、未知データが拡張されて複数の疑似未知データが生成され、それらの識別結果が統合されてクラス分類がされるので、未知データそのものを識別する場合と比較して、識別能力が向上する。

【発明の効果】

【0026】

本発明によれば、未知データが拡張されて、それらの識別結果が統合されてクラス分類がされるので、未知データそのものを識別する場合と比較して、識別能力を向上できる。

【図面の簡単な説明】

【0027】

【図1】本発明の実施の形態における学習装置の構成を示すブロック図

【図2】ある多様体上における、あるクラスのデータ分布（確率密度）を示す図

10

20

30

40

50

【図 3】図 2 に示すデータ分布に対する訓練データを示す図

【図 4】手書き数字の訓練データ及びその疑似データの例を示す図

【図 5】疑似データの分布を示す図

【図 6】記号の諸定義を示す図

【図 7】疑似データを学習した結果得られるクラスの事後分布を示す図

【図 8】本発明の実施の形態における識別器の構成を示すブロック図

【図 9】本発明の実施の形態における未知データの例を示す図

【図 10】本発明の実施の形態における疑似未知データの標本分布を示す図

【図 11】本発明の実施の形態の実験結果を示す図

【発明を実施するための形態】

10

【0028】

以下、本発明の実施の形態の学習装置及び識別器について、図面を参照しながら説明する。なお、以下に説明する実施の形態は、本発明を実施する場合の一例を示すものであって、本発明を以下に説明する具体的構成に限定するものではない。本発明の実施にあたっては、実施の形態に応じた具体的構成が適宜採用されてよい。

【0029】

以下では、画像データ等の未知データに対してクラス分類を行うパターン識別器、及びそのパターン識別器にて用いる識別モデルを学習するための学習装置を例に挙げて本発明の実施の形態を説明する。また、識別モデルとして、フィードフォワード型の多層ニューラルネットワークを採用する場合を説明する。なお、識別モデルとしては、畳み込みニューラルネットワーク等の他のモデルを採用してもよい。

20

【0030】

(学習装置)

図 1 は、本発明の実施の形態の学習装置の構成を示すブロック図である。学習装置 100 は、訓練データ記憶部 11 と、データ拡張部 12 と、変換パラメタ生成部 13 と、学習部 14 とを備えている。学習装置 100 は、コンピュータによって実現される。このコンピュータは、補助記憶部、一時記憶部、演算処理部、入出力部等を備えており、訓練データ記憶部 11 は、例えば補助記憶部によって実現される。また、データ拡張部 12、変換パラメタ生成部 13、及び学習部 14 は、演算処理部が学習プログラムを実行することで実現される。

30

【0031】

訓練データ記憶部 11 には、ターゲット値を伴った訓練データ（以下、「データサンプル」ともいう。）が記憶されている。変換パラメタ生成部 13 は、データ拡張部 12 にて訓練データ記憶部 11 に記憶された訓練データを拡張するための変換パラメタを生成する。データ拡張部 12 は、変換パラメタ生成部 13 にて生成された変換パラメタを用いて訓練データ記憶部 11 に記憶された訓練データに対してパナメトリックな変換を施してデータ拡張を行う。

【0032】

学習部 14 は、データ拡張部 13 にて拡張された訓練データを用いて学習を行い、識別器において用いる識別モデルを生成する。学習部 14 は、多層ニューラルネットワークのパラメタである各層の重み W を決定する。

40

【0033】

データ拡張部 12 におけるデータ拡張について説明する。図 2 は、ある多様体上における、あるクラスのデータ分布（確率密度）を示す図である。実際のデータサンプルはこの分布に従う確率変数であり、確率的に生成されることとなる。

【0034】

図 3 は、図 2 に示すデータ分布に対する訓練データを示す図である。図 3 では、図 2 に示すデータ分布に対して、訓練データ記憶部 11 に記憶された訓練データ $t d 1 \sim t d 7$ が示されている。訓練データの個数が無限に近づけば、その確率密度は図 2 の分布に漸近するが、現実的には限られた数の訓練データしか得られないため、分布の近似精度は荒く

50

ならざるを得ない。

【 0 0 3 5 】

データ拡張部 1 2 は、訓練データを変換することでデータ数を増大させる。この変換は、データの多様体上におけるデータ点近傍へのパラメトリックな変換である。この変換には、例えば、画像の局所的な歪み、局所的な輝度変化、アフィン変換、ノイズの重畳等が含まれる。図 4 は、識別モデルが画像による手書き数字認識を行うものである場合の手書き数字の訓練データ（オリジナルデータ）及びその訓練データを拡張して得られた新たなデータ（疑似データ）の例を示す図である。

【 0 0 3 6 】

図 5 は、疑似データの分布を示す図である。図 5 において、疑似データの分布は実線で示されている。与えられている訓練データに対して、クラスの特徴を失わない程度に少し変形を加えたとき、それによって生成される疑似データは、もとの訓練データの近傍に位置することになる。

【 0 0 3 7 】

変換パラメタをまとめたものを θ とし、変換式を $u(x_0; \theta)$ とすると、1つの訓練データから無数に疑似データを生成した場合、疑似データは、下式で表される分布を持つことになる。

【 数 1 】

$$p(u(x^i; \theta)) \begin{cases} u: \mathbb{R}^D \rightarrow \mathbb{R}^D \\ \theta \sim p(\theta) \end{cases}$$

10

20

ここで、 D はデータの次元であり、図 2 に示すデータ分布の空間の次元に対応している。

【 0 0 3 8 】

学習部 1 4 は、拡張された訓練データを学習する。上述のように、本実施の形態では、学習部 1 4 は、識別モデルとして、フィードフォワード型の多層ニューラルネットワークを学習する。

【 0 0 3 9 】

学習部 1 4 は、出力値とターゲット値が近づくほど低い値をとる目的関数を用いて、この目的関数を最小化する識別モデルのパラメタを探索することで、汎化能力の高い識別モデルを決定する。本実施の形態では、目的関数として、クロスエントロピーを採用する。

30

【 0 0 4 0 】

まず、記号の諸定義を図 6 に示す。図 6 において、 a_l 及び x_l は以下のとおり定義される。

【 数 2 】

$$a_l = \bar{W}_l \bar{x}_{l-1} \begin{cases} \bar{W}_l = [W_l | b_l] \\ \bar{x}_l = [x_l^T | 1]^T \end{cases}$$

40

【 数 3 】

$$x_l = f_l(a_l)$$

なお、 f_l は、（劣）微分可能な単調非減少 / 非増加関数である。

【 0 0 4 1 】

また、出力の次元数はクラス数とする。このときターゲット値は、出力の要素のうちの 1 つが値 1 を持ち、その他の要素は 0 を持つものとする。2 クラス分類の場合は、出力を 1 次元としてもよく、このときターゲット値は 0 または 1 をとる。

50

【 0 0 4 2 】

以下では、まず、データ拡張をしない場合の学習について説明し、それとの比較において、本実施の形態のデータ拡張をする場合の学習について説明する。

【 0 0 4 3 】

データ拡張を行わない場合の目的関数を下式 (1) と (1 ') に記す。

【 数 4 】

$$G_{\text{train}}(W) = \sum_{i \in \text{train}} G_i(W) \rightarrow \min \quad (1)$$

10

【 数 5 】

$$G_i(W) = \sum_{c \in C} -t_c^i \ln y_c(x_0^i; W) \quad (1')$$

【 数 6 】

$$\text{s. t. } \begin{cases} y_c = \text{softmax}(a_L)_c \rightarrow \sum_c y_c = 1, 0 < y_c < 1 \\ \sum_c t_c = 1, t_c \in \{0, 1\} \quad (\text{訓練データのターゲット値}) \\ W = \{\bar{W}_1, \bar{W}_2, \dots, \bar{W}_L\} \end{cases}$$

20

ここで、 i は訓練データのインデクスであり、 C はクラスレベルである。

【 0 0 4 4 】

このように、ニューラルネットワークの出力にソフトマックス関数を施すことにより、ベクトルが正規化されるとともに正の値に変換される。このベクトルに対し、式 (1 ') で定義されるクロスエントロピーを施すことで、ある訓練サンプルの分類の悪さが定量化される。なお、1次元の出力 $y(x_0^i; W)$ の場合は、 $y_1(x_0^i; W) = y(x_0^i; W)$ 、 $y_2(x_0^i; W) = 1 - y(x_0^i; W)$ 、 $t_1 = t$ 、 $t_2 = 1 - t$ と変数を置き換えることで式 (1) と (1 ') を適用できる。

30

【 0 0 4 5 】

この目的関数の勾配

【 数 7 】

$$\frac{\partial G_i}{\partial \bar{W}_l}$$

40

を計算し、複数のデータサンプルの和を取った勾配を使用して、確率的勾配降下法 (S G D : Stochastic Gradient Descent) によって W の要素を下式 (2) のとおりに更新する。

【数 8】

$$\bar{W}_l \leftarrow \bar{W}_l - \epsilon \sum_{i \in RPE} \frac{\partial G_i}{\partial \bar{W}_l} \quad (2)$$

この更新は、重み W の各要素（各層の重み）が収束するまで繰り返される。ここで、式（2）の RPE は、Randomly Picked Example の略であり、データサンプルを反復ごとにランダムに選ぶことを意味している。

【0046】

10

次に、データ拡張を行う本実施の形態の場合について説明する。本実施の形態の目的関数は、下式（3）、（3'）のとおりとなる。

【数 9】

$$G_{\text{train}}(W) = \sum_{i \in \text{train}} G_i(W) \rightarrow \min \quad (3)$$

【数 10】

$$G_i(W) = \sum_{c \in C} E_{\theta} [-t_c^i \ln y_c(u(x_0^i; \theta); W)] \quad (3')$$

20

【数 11】

$$E_{\theta}[\dots] \simeq \frac{1}{M} \sum_{\theta = \theta_1, \theta_2, \dots, \theta_M} \dots, \quad \theta_j \sim p(\theta_j)$$

【数 12】

30

$$\text{s.t.} \begin{cases} y_c = \text{softmax}(a_L)_c \rightarrow \sum_c y_c = 1, 0 < y_c < 1 \\ \sum_c t_c = 1, t_c \in \{0, 1\} \quad (\text{訓練データのターゲット値}) \\ W = \{\bar{W}_1, \bar{W}_2, \dots, \bar{W}_L\} \end{cases}$$

【0047】

式（3'）では、式（1'）のように訓練データそのものを学習部 14 に入力するのではなく、訓練データから変換によって派生した人為的データである疑似データをデータ拡張部 12 で生成して、それらを学習部 14 に入力する。また、式（3'）では、式（1'）と異なり、変換パラメタに対するクロスエントロピーの期待値が取られている。学習部 14 は、この目的関数の最適化方法として、確率的勾配降下法を採用する。

40

【0048】

具体的手順は次の通りである。データ拡張部 12 は、訓練データ記憶部 11 に記憶された訓練データを 1 つ選択し、また、変換パラメタ生成部 14 から変換パラメタを複数個分、適当な確率分布に従う乱数によりサンプリングする。データ拡張部 12 は、これらのパラメタを使って訓練データに対して変換を施すことにより、この 1 つの訓練データを複数個に拡張する。学習部 14 は、これら複数個の疑似データを使って勾配

【数 1 3】

$$\frac{\partial G_i}{\partial \bar{W}_l}$$

を計算し、複数のデータサンプルの和を取った勾配を使用して、確率的勾配降下法によって W の要素を下式 (4) のとおりに更新する。

【数 1 4】

$$\bar{W}_l \leftarrow \bar{W}_l - \epsilon \sum_{i \in \text{RPED}} \frac{\partial G_i}{\partial \bar{W}_l} \quad (4)$$

10

この更新は、重み W の各要素（各層の重み）が収束するまで繰り返される。ここで、式 (4) の R P E R D は、Randomly Picked Example with Random Distortionの略であり、乱数によって変形されたデータサンプルの中からデータサンプルを選ぶことを意味する。

【0 0 4 9】

なお、通常、多層ニューラルネットワークの重みパラメタの更新には勾配法を出力層側から入力層側へ順番に適用する誤差逆伝搬法が使われる。誤差逆伝搬法も勾配法の一様であるので、確率的勾配降下法が適用可能である。誤差逆伝搬法については、C・M・ビショップ, “パターン認識と機械学習”, シュプリンガー・ジャパンに詳しく記載されている。

20

【0 0 5 0】

図 7 は、以上のようにしてデータ拡張部 1 2 にて生成された疑似データを学習した結果得られるクラスの事後分布を示す図である。図 7 において、クラスの事後分布は実線で示されている。識別器では、このクラスの事後分布を識別モデルとして使用して識別を行うことになる。このデータ拡張により、図 2 に示す本来の分布に対する網羅性を高めることができる。

【0 0 5 1】

(識別器)

次に、本実施の形態の識別器について説明する。図 8 は、本実施の形態の識別器の構成を示すブロック図である。識別器 2 0 0 は、データ入力部 2 1 と、データ拡張部 2 2 と、変換パラメタ生成部 2 3 と、識別部 2 4 とを備えている。識別器 2 0 0 は、コンピュータによって実現される。このコンピュータは、補助記憶部、一時記憶部、演算処理部、入出力部等を備えており、データ入力部 2 1 は、例えば入出力部によって実現される。また、データ拡張部 2 2、変換パラメタ生成部 2 3、及び識別部 2 4 は、演算処理部が本発明の実施の形態の識別プログラムを実行することで実現される。

30

【0 0 5 2】

データ入力部 2 1 には、学習に使用していないデータであって、未知データが入力される。図 9 は、未知データ $u d 1 \sim u d 5$ の例を示す図である。図 9 のような未知データが入力されたとき、データ拡張によって網羅性を高めたことによって正答できる場合も多いが、未知データ $u d 5$ のように、もとの分布の近似精度の限界から誤答する点もあり得る。

40

【0 0 5 3】

そこで、本実施の形態の識別器 2 0 0 では、識別時においても、学習時に行ったものと同様の方法でデータ拡張し、それらに対しての識別結果を適切に統合する。このように、識別時においても乱数を用いてデータを拡張することで、その分布がクラスの事後分布に重なる可能性が高まることから、従来正答できなかった点が正答できる可能性が高くなる。この理由について、以下に詳しく説明する。

【0 0 5 4】

データ拡張を行っていない場合には、あるデータが入力されたとき、最も適切なクラス

50

分類方法は、下式（５）を満たすクラス c を選択することである。

【数１５】

$$y_c(x_0; W) \geq y_{c' \neq c}(x_0; W) \quad (5)$$

この決定則は、データ拡張を行っていない場合の目的関数（１'）を最小化しており、理論上最適である。

【００５５】

従来、データ拡張を行った場合においても、データ拡張を行わない場合と同じ決定則が使用されてきた。即ち、学習の際には式（３'）によって学習をしたとしても、識別の際には、データ拡張を行っていない場合に理論上最適となる式（５）の決定則を用いて識別（クラス分類）が行われてきた。しかしながら、データ拡張を行った場合と行わない場合とでは、理論的に最適な決定則が異なっている。即ち、上式（５）の決定則は、データ拡張をしていない場合の目的関数（１'）を最小化しているが、データ拡張をした場合の目的関数（３'）を最小化するものではない。

10

【００５６】

データ拡張が行われた場合に、最も適切なクラス分類方法は、下式（６）を満たすクラス c を選択することである。

【数１６】

$$E_{\theta}[\ln y_c(u(x_0; \theta); W)] \geq E_{\theta}[\ln y_{c' \neq c}(u(x_0; \theta); W)] \quad (6)$$

20

この決定則は目的関数（３'）を最小化しており、理論上最適である。

【００５７】

以上のように、従来法では、データ拡張用の目的関数を学習時に最小化しているにも関わらず、式（５）の決定則を適用しているため、理論上最適なクラス分類が行えていない。これに対して、本実施の形態の識別器２００は、識別時においても変換パラメタに対する出力の対数の期待値をとって識別を行う。

【００５８】

具体的には、データ拡張部２２は、変換パラメタ生成部２３にて生成された変換パラメタを用いてデータ入力部２１に入力された未知データを変換して、複数の疑似未知データを生成する。このときデータ拡張部２２で用いる変換パラメタは、識別モデルを生成するための学習の際に使用した分布 $p(\theta_j)$ から確率的に生成する。図１０は、未知データ $q d 5$ から生成された疑似未知データの標本分布を示す図である。図１０において、未知データ $q d 5$ の標本分布は実線で示されている。

30

【００５９】

識別部２４は、式（６）の勾配計算を行って、変換パラメタに対する出力の対数の期待値が最大となるクラスレベルを選択する。このように、データ拡張時の最適な決定則を用いることにより、データ収取量が同じ場合で、かつ同じ要領でデータ拡張をした場合であっても、従来よりも高い識別能力を獲得することが可能となる。

40

【００６０】

上記の実施の形態では、目的関数として、クロスエントロピーを採用したが、目的関数はクロスエントロピーに限られない。以下では、目的関数を誤差の二乗の総和とした場合の決定則について説明する。データの拡張がない場合の目的関数は、下式（７）及び（７'）によって表される。

【数 1 7】

$$G_{\text{train}}(W) = \sum_{i \in \text{train}} G_i(W) \rightarrow \min \quad (7)$$

【数 1 8】

$$G_i(W) = \sum_{c \in C} (y_c(x_0^i; W) - t_c^i)^2 \quad (7')$$

10

【数 1 9】

$$\text{s. t. } \begin{cases} y_c = \text{softmax}(a_L)_c \rightarrow \sum_c y_c = 1, 0 < y_c < 1 \\ \sum_c t_c = 1, t_c \in \{0,1\} \\ W = \{\bar{W}_1, \bar{W}_2, \dots, \bar{W}_L\} \end{cases} \quad (\text{訓練データのターゲット値})$$

20

【0 0 6 1】

この目的関数の勾配

【数 2 0】

$$\frac{\partial G_i}{\partial \bar{W}_l}$$

を計算し、複数のデータサンプルの和を取った勾配を使用して、確率的勾配降下法によって W の要素を下式 (8) のとおりに更新する。この更新は、 W の各要素が収束するまで繰り返される。

30

【数 2 1】

$$\bar{W}_l \leftarrow \bar{W}_l - \epsilon \sum_{i \in RPE} \frac{\partial G_i}{\partial \bar{W}_l} \quad (8)$$

【0 0 6 2】

次に、上記の例においてデータ拡張を行う本実施の形態の場合について説明する。本実施の形態の目的関数は、下式 (9)、(9') のとおりとなる。

【数 2 2】

$$G_{\text{train}}(W) = \sum_{i \in \text{train}} G_i(W) \rightarrow \min \quad (9)$$

40

【数 2 3】

$$G_i(W) = \sum_{c \in C} E_{\theta} [(y_c(u(x_0^i; \theta); W) - t_c^i)^2] \quad (9')$$

【数 2 4】

$$E_{\theta}[\cdots] \simeq \frac{1}{M} \sum_{\theta=\theta_1, \theta_2, \cdots, \theta_M} \cdots, \theta_j \sim p(\theta_j)$$

【数 2 5】

$$\text{s. t. } \begin{cases} y_c = \text{softmax}(a_L)_c \rightarrow \sum_c y_c = 1, 0 < y_c < 1 \\ \sum_c t_c = 1, t_c \in \{0,1\} \quad (\text{訓練データのターゲット値}) \\ W = \{\bar{W}_1, \bar{W}_2, \cdots \bar{W}_L\} \end{cases} \quad 10$$

【0 0 6 3】

このように、式(9')では、式(7')と異なり、変換パラメタに対する誤差の二乗の総和の期待値が取られている。

【0 0 6 4】

従来は、決定則として、下式(10)が用いられていた。

20

【数 2 6】

$$y_c(x_0; W) \geq y_{c' \neq c}(x_0; W) \quad (10)$$

しかしながら、この決定則は、データ拡張なしの目的関数(7')を最小化しているが、データ拡張をした場合の目的関数(9')を最小化するものではない。そこで、データ拡張をした場合には、下式(11)のように、変換パラメタに対する誤差の二乗の総和の期待値を最小化する決定則を採用する。

【数 2 7】

$$E_{\theta}[y_c(u(x_0; \theta); W)] \geq E_{\theta}[y_{c' \neq c}(u(x_0; \theta); W)] \quad (11)$$

30

【0 0 6 5】

データ拡張を行った場合に、式(11)の決定則を用いることにより、上記の実施の形態と同様に、従来よりも高い識別能力を獲得することが可能となる。

【0 0 6 6】

以上のように、本実施の形態の識別器200において、データ拡張部22は、未知データに対して、学習時のデータ拡張と同様の方法でデータ拡張を行って疑似未知データを生成し、識別部24は、疑似未知データの期待値に基づいてクラス分類を行う。換言すれば、識別器200は、未知データそのものについてクラス分類を行うのではなく、未知データを拡張してそれらのクラス分類の結果を統合して、クラス分類を行う。即ち、識別器200は、学習を行う際に用いた目的関数を最小化するという決定則によってクラス分類を行う。これにより、与えられた訓練データに対してデータ拡張を行って学習をして識別モデルを生成した場合に、訓練データの収取量が同じであり、かつ同じ要領で訓練データをデータ拡張をした場合の従来法よりも高い識別能力を実現できる。

40

【0 0 6 7】

(実験例)

以下、本実施の形態の学習装置及び識別器を用いて行った実験を説明する。この実験では、以下のように条件を設定した。データセットとして、手書き数字データセット(MNIST、<http://yann.lecun.com/exdb/mnist/>、図4参照)を用いた。訓練データとして

50

MNISTの訓練データセット(60,000セット)のうちの6,000セットを用い、テストデータとしてMNISTのテストデータセット(10,000セット)のうちの1,000セットを用いた。識別モデルとしては、フィードフォワード・全結合型6層ニューラルネットワークを用いた。評価基準として、誤識別率を評価した。

【0068】

学習装置における学習条件として、従来法による識別を行う場合についても、本発明の実施の形態の識別を行う場合にも同一のデータ拡張を適用した。また、生成されたサンプルからは1度しか微分を計算せず、オリジナルサンプルからは一切微分を計算しないこととした。識別器における識別条件として、従来法では、オリジナルサンプルのみを識別し、本発明の実施の形態の識別器では、複数生成されたサンプルより期待値を評価した。なお、本発明の実施の形態の識別器では、オリジナルサンプルそのものは期待値に使用しなかった。

10

【0069】

実験結果を図11に示す。図11では、横軸に変換パラメタの種類の数Mをとり、縦軸に誤識別率をとり、従来法を用いた場合の誤識別率も合わせて示している。なお、上述のように、変換パラメタの種類の数Mについては、下式が成り立つ。

【数28】

$$E_{\theta}[\ln y_c(u(x_0; \theta); W)] \simeq \frac{1}{M} \sum_{\theta=\theta_1}^{\theta_M} \ln y_c(u(x_0; \theta); W)$$

20

【0070】

図11の結果から、変換パラメタの種類がM=16以上では、従来法でオリジナルサンプルのみを識別した場合よりも誤識別率が低くなっており、本発明の実施の形態の識別時における期待値演算が有効であることがわかる。

【産業上の利用可能性】

【0071】

本発明は、未知データが拡張されて、それらの識別結果が統合されてクラス分類がされるので、未知データ自体を識別する場合と比較して、識別能力を向上できるという効果を有し、訓練データに対してデータ拡張を行って学習をして生成した識別モデルを用いる識別装置等として有用である。

30

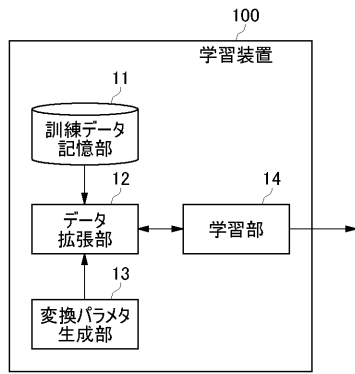
【符号の説明】

【0072】

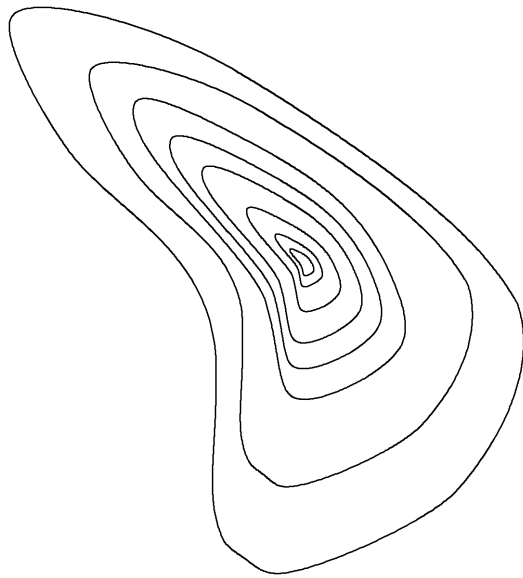
- 100 学習装置
- 11 訓練データ記憶部
- 12 データ拡張部
- 13 変換パラメタ生成部
- 14 学習部
- 200 識別装置
- 21 データ入力部
- 22 データ拡張部
- 23 変換パラメタ生成部
- 24 識別部

40

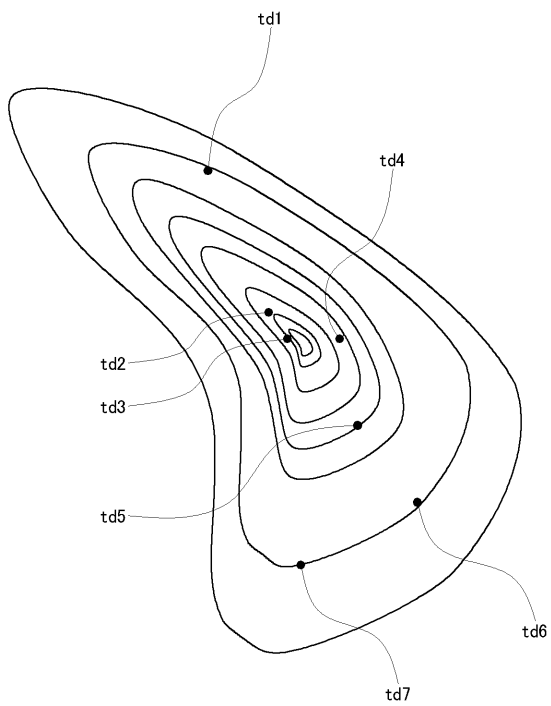
【図 1】



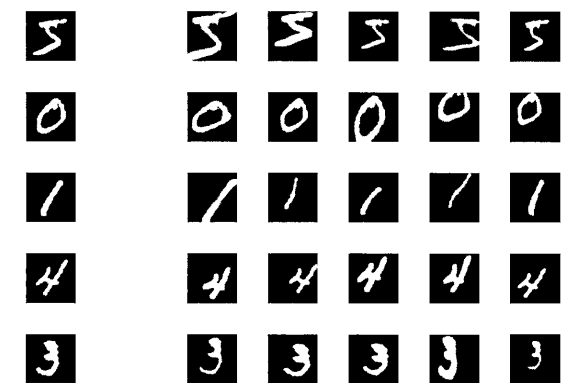
【図 2】



【図 3】



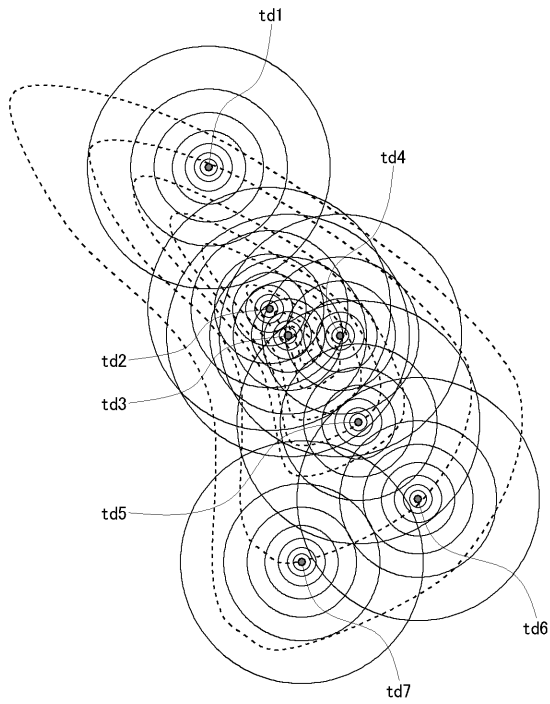
【図 4】



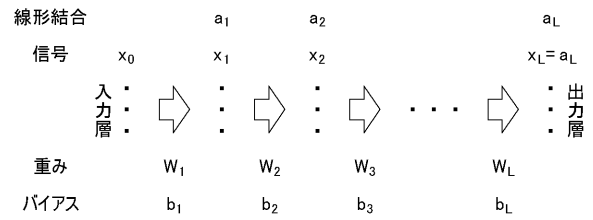
オリジナルデータ
 x_0

疑似データ
 $u(x_0; \theta)$

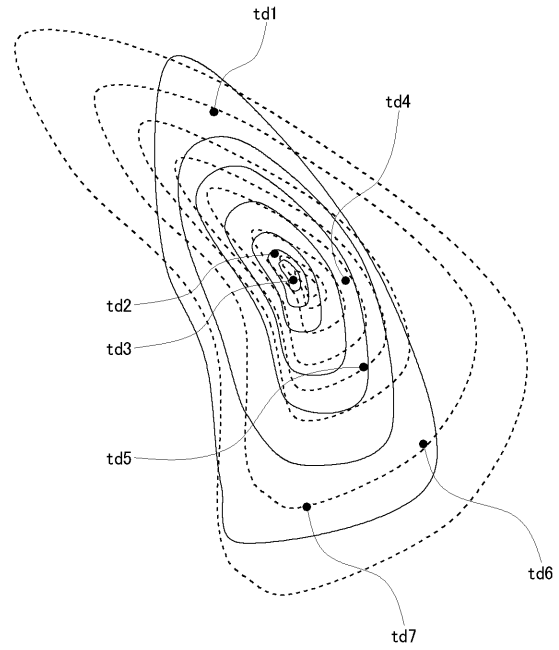
【図 5】



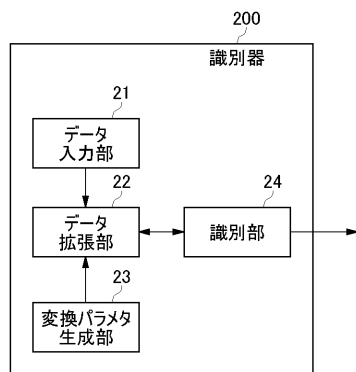
【図 6】



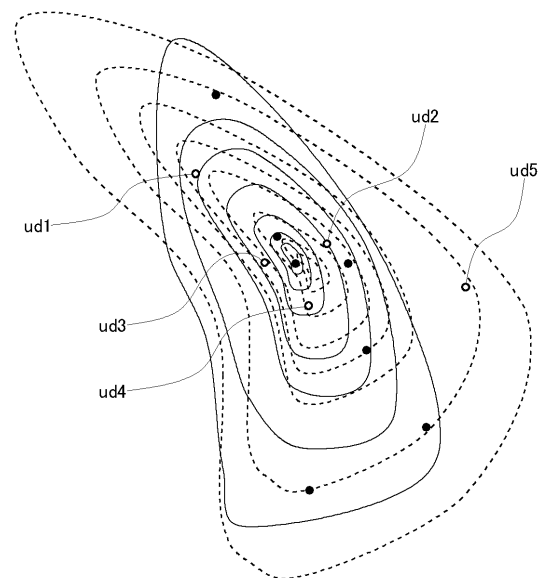
【図 7】



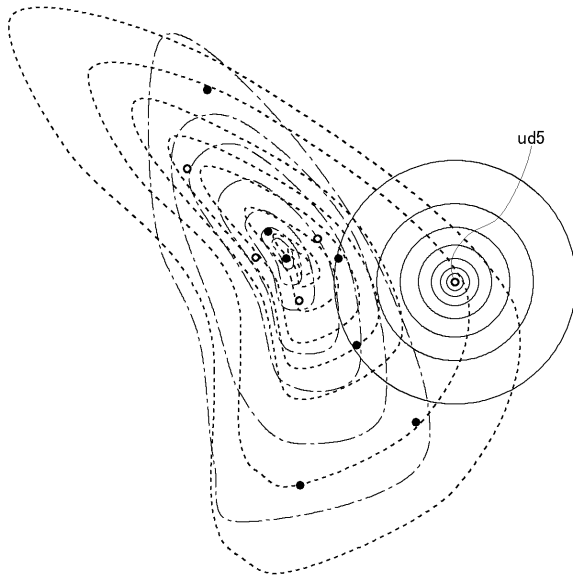
【図 8】



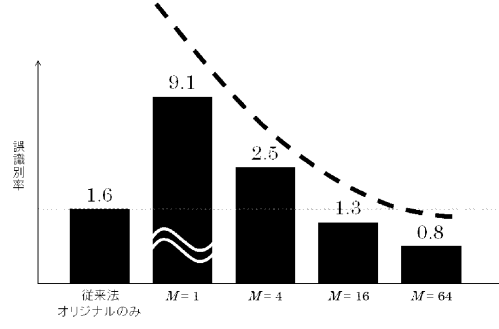
【図 9】



【図 10】



【図 11】



フロントページの続き

(72)発明者 玉津 幸政

愛知県刈谷市昭和町一丁目1番地 株式会社デンソー内

審査官 多胡 滋

(56)参考文献 特表2003-500766(JP,A)

堀田創, 外1名, 文字デザインを例とした感性情報からのファジールール自動抽出法, 電子情報通信学会論文誌 情報・システム, 日本, 社団法人電子情報通信学会, 2006年10月1日, 第10号, pp.2361-2369

福田拓也, 外1名, モンテカルロ法を用いた隠れマルコフモデルに基づく分かち書き, Journal of the DBSJ, 日本, 日本データベース学会, 2008年6月27日, 第7巻, 第1号, pp.73-78

(58)調査した分野(Int.Cl., DB名)

G06N 99/00