

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2019-148877
(P2019-148877A)

(43) 公開日 令和1年9月5日(2019.9.5)

(51) Int. Cl. F I テーマコード (参考)
G06N 20/00 (2019.01) G06N 99/00 150
G06N 7/00 (2006.01) G06N 7/00 150

審査請求 未請求 請求項の数 5 O L (全 17 頁)

(21) 出願番号 特願2018-31875 (P2018-31875)
 (22) 出願日 平成30年2月26日 (2018.2.26)

(出願人による申告) 平成29年度、国立研究開発法人科学技術振興機構、戦略的創造研究推進事業(AC T-1) 研究領域「情報と未来」研究課題名「適応型制限ボルトマンマシンの複素拡張に基づくボコーダー不要な非パラレル声質変換」委託研究、産業技術力強化法第19条の適用を受ける特許出願

(71) 出願人 504133110
 国立大学法人電気通信大学
 東京都調布市調布ケ丘一丁目5番地1
 (74) 代理人 110000925
 特許業務法人信友国際特許事務所
 (72) 発明者 中鹿 亘
 東京都調布市調布ケ丘一丁目5番地1 国立大学法人電気通信大学内

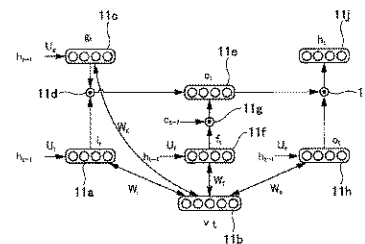
(54) 【発明の名称】 符号化装置、符号化方法およびプログラム

(57) 【要約】

【課題】 時系列データに対して重要な記憶を確率的に伝播させることのできるRBMの拡張モデルを適用することで、良好な符号化ができるようにする。

【解決手段】 パラメータ学習ユニットと符号化ユニットとを備え、入力データを表現する可視素子と、制限ボルトマンマシンによる確率モデルを適用して隠れ素子を推定して、学習および符号化を行う。ここで学習ユニットと符号化ユニットの入力データは時系列データであり、学習ユニットと符号化ユニットは、可視素子に基づいて時系列のデータを記憶する記憶セルと、記憶セルに記憶されたデータおよび入力データに基づいてデータを一時記憶する短期記憶セルと、可視素子に基づいて短期記憶セルの出力を制御する忘却ゲートとを有する。そして、忘却ゲートに基づいて制御された短期記憶セルの出力に基づいて、時刻ごとに伝播する隠れ素子を得る。

【選択図】 図3



【特許請求の範囲】**【請求項 1】**

入力データを表現する可視素子と、潜在的な情報を表現する隠れ素子との間に結合重みが存在すると仮定した制限ボルツマンマシンによる確率モデルを適用して、学習用入力データに対して、前記隠れ素子および前記結合重みを推定する処理を行うパラメータ学習ユニットと、

符号化用入力データに対して、前記パラメータ学習ユニットで推定した前記制限ボルツマンマシンによる確率モデルを適用して、前記隠れ素子を推定し、推定した前記隠れ素子に基づいて符号化データとしての出力データを得る符号化ユニットとを備え、

前記学習用入力データおよび前記符号化用入力データは時系列の入力データであり、前記パラメータ学習ユニットおよび前記符号化ユニットは、前記可視素子に基づいて時系列のデータを記憶する記憶セルと、前記記憶セルに記憶されたデータおよび入力データに基づいてデータを一時記憶する短期記憶セルと、前記可視素子に基づいて前記短期記憶セルの出力を制御する忘却ゲートとを有し、

前記忘却ゲートに基づいて制御された前記短期記憶セルの出力に基づいて、時刻ごとに伝播する前記隠れ素子を得るようにした

符号化装置。

【請求項 2】

前記短期記憶セルが記憶する、入力データについての特定時刻における短期記憶データは、その特定時刻の1つ前の時刻における短期記憶データに前記忘却ゲートを通したものと、前記記憶セルのデータに入力データを通したものととの和で取得し、

得られた短期記憶データに出力データを掛け合わせて、非線形関数を乗算したものを、時刻ごとに伝播する前記隠れ素子とする

請求項 1 に記載の符号化装置。

【請求項 3】

さらに、前記符号化ユニットで得られた符号化データを復号化する復号化処理ユニットを備えた

請求項 1 又は 2 に記載の符号化装置。

【請求項 4】

入力データを表現する可視素子と、潜在的な情報を表現する隠れ素子との間に結合重みが存在すると仮定した制限ボルツマンマシンによる確率モデルを適用して、時系列のデータである学習用入力データに対して、前記隠れ素子および前記結合重みを推定するパラメータ学習処理と、

時系列のデータである符号化用入力データに対して、前記パラメータ学習処理で推定した前記制限ボルツマンマシンによる確率モデルを適用して、前記隠れ素子を推定し、推定した前記隠れ素子に基づいて符号化データとしての出力データを得る符号化処理とを行う符号化方法であり、

前記パラメータ学習処理と前記符号化処理のそれぞれは、

前記可視素子に基づいて時系列のデータを記憶する記憶処理と、

前記記憶処理により記憶されたデータおよび入力データに基づいてデータを一時記憶する短期記憶処理と、

前記短期記憶処理で一時記憶されたデータの出力を前記可視素子に基づいて制御する忘却処理と、

前記忘却処理に基づいて制御された前記短期記憶処理の出力に基づいて、時刻ごとに伝播する前記隠れ素子を得る隠れ素子取得処理と、を含む

符号化方法。

【請求項 5】

入力データを表現する可視素子と、潜在的な情報を表現する隠れ素子との間に結合重みが存在すると仮定した制限ボルツマンマシンによる確率モデルを適用して、時系列のデータである学習用入力データに対して、前記隠れ素子および前記結合重みを推定するパラメ

10

20

30

40

50

ータ学習ステップと、

時系列のデータである符号化用入力データに対して、前記パラメータ学習ステップで推定した前記制限ボルツマンマシンによる確率モデルを適用して、前記隠れ素子を推定し、推定した前記隠れ素子に基づいて符号化データとしての出力データを得る符号化ステップとを、コンピュータに実行させるプログラムであり、

前記パラメータ学習ステップと前記符号化ステップのそれぞれは、

前記可視素子に基づいて時系列のデータを記憶する記憶ステップと、

前記記憶ステップにより記憶されたデータおよび入力データに基づいてデータを一時記憶する短期記憶ステップと、

前記短期記憶ステップで一時記憶されたデータの出力を前記可視素子に基づいて制御する忘却ステップと、

前記忘却ステップに基づいて制御された前記短期記憶ステップの出力に基づいて、時刻ごとに伝播する前記隠れ素子を得る隠れ素子取得ステップと、

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】

本発明は、符号化装置および符号化方法、並びに符号化方法を実行するプログラムに関する。

【背景技術】

【0002】

近年、ディープラーニングを用いた機械学習手法が飛躍的に高い精度を上げ、画像認識や音声認識など、幅広い分野において盛んに研究され、利用が進んでいる。これまでに数多くの機械学習手法が提案されているが、最も代表的なモデルとして、制限ボルツマンマシン (restricted Boltzmann machine: 以下、「RBM」と称する) が用いられている。

RBMは、可視ユニットと隠れユニットからなる2層構造の確率モデルであり、その応用は特徴抽出や識別器、DNN (deep neural network) の事前学習など多岐に渡る。

【0003】

一方で、音声認識や自然言語処理など、時系列データを取り扱う場合には、過去の情報を伝播させる機構を持つRNN (recurrent neural networks) やCNN (convolutional neural networks)、LSTM (long short term memory)、GRU (gated recurrent unit) などが有効である。中でもLSTMは記憶セルによる長・短期記憶構造によって重要な情報を保持させることができ、音声認識、音声合成、機械翻訳など様々なタスクにおいて高い精度を得ることができる。

非特許文献1には、ボルツマンマシンで特徴量を抽出する技術が記載されている

【先行技術文献】

【非特許文献】

【0004】

【非特許文献1】Y. Freund and D. Haussler, "Unsupervised learning of distributions of binary vectors using two layer networks," Computer Research Laboratory, pp. 912-919, 1994

【発明の概要】

【発明が解決しようとする課題】

【0005】

ところで、RBMはフレームごとに独立して処理を行うため、時系列データを表現するのに不十分である。時系列データを表現するRBMの拡張モデルとして、既にTRBM (temporal RBM) が提案されているが、RNNと同様に、勾配が消失する、過去の長期的な依存関係を表現できない、といった問題がある。

【0006】

本発明は、LSTMにある長・短期記憶構造に着目し、重要な記憶を確率的に伝播させ

10

20

30

40

50

ることのできるRBMの拡張モデルを適用して、良好な符号化ができる符号化装置、符号化方法およびプログラムを提供することを目的とする。

【課題を解決するための手段】

【0007】

本発明の符号化装置は、パラメータ学習ユニットと符号化ユニットとを備える。

パラメータ学習ユニットは、入力データを表現する可視素子と、潜在的な情報を表現する隠れ素子との間に結合重みが存在すると仮定した制限ボルツマンマシンによる確率モデルを適用して、学習用入力データに対して、隠れ素子および結合重みを推定する処理を行う。

符号化ユニットは、符号化用入力データに対して、パラメータ学習ユニットで推定した制限ボルツマンマシンによる確率モデルを適用して、隠れ素子を推定し、推定した隠れ素子に基づいて符号化データとしての出力データを得る。

ここで、学習用入力データおよび符号化用入力データは時系列の入力データであり、パラメータ学習ユニットおよび符号化ユニットは、可視素子に基づいて時系列のデータを記憶する記憶セルと、記憶セルに記憶されたデータおよび入力データに基づいてデータを一時記憶する短期記憶セルと、可視素子に基づいて短期記憶セルの出力を制御する忘却ゲートとを有する。

そして、忘却ゲートに基づいて制御された短期記憶セルの出力に基づいて、時刻ごとに伝播する隠れ素子を得るようにした。

【0008】

また、本発明の符号化方法は、パラメータ学習処理と符号化処理とを行うものである。

パラメータ学習処理は、入力データを表現する可視素子と、潜在的な情報を表現する隠れ素子との間に結合重みが存在すると仮定した制限ボルツマンマシンによる確率モデルを適用して、時系列のデータである学習用入力データに対して、隠れ素子および結合重みを推定する。

符号化処理は、時系列のデータである符号化用入力データに対して、パラメータ学習処理で推定した制限ボルツマンマシンによる確率モデルを適用して、隠れ素子を推定し、推定した隠れ素子に基づいて符号化データとしての出力データを得る。

ここで、パラメータ学習処理と符号化処理のそれぞれは、可視素子に基づいて時系列のデータを記憶する記憶処理と、記憶処理により記憶されたデータおよび入力データに基づいてデータを一時記憶する短期記憶処理と、短期記憶処理で一時記憶されたデータの出力を可視素子に基づいて制御する忘却処理と、忘却処理に基づいて制御された短期記憶処理の出力に基づいて、時刻ごとに伝播する隠れ素子を得る隠れ素子取得処理と、を含む。

【0009】

また、本発明のプログラムは、上述した符号化方法の記憶処理、短期記憶処理、忘却処理、隠れ素子取得処理のそれぞれのステップをコンピュータに実行させるものである。

【発明の効果】

【0010】

本発明によると、時系列データに対して制限ボルツマンマシン(RBM)の機能を適用することができ、適切に特徴量を抽出することができるので、時系列の入力データから高い精度で特徴量を抽出して符号化することが可能になり、効率の良い符号化を行うことができる。

【図面の簡単な説明】

【0011】

【図1】本発明の一実施の形態例による符号化装置の構成例を示すブロック図である。

【図2】図1の符号化装置のハードウェア構成例を示すブロック図である。

【図3】本発明の一実施の形態例に適用される、LSTBM(long short term Boltzmann memory)を模式的に示す図である。

【図4】本発明の一実施の形態例によるパラメータ学習の流れを示すフローチャートである。

10

20

30

40

50

【図 5】本発明の一実施の形態例による符号化の流れを示すフローチャートである。

【図 6】図 4 のステップ S 1 3 の L S T B M の学習処理を示すフローチャートである。

【図 7】図 5 のステップ S 2 3 の符号化処理を示すフローチャートである。

【図 8】本発明の一実施の形態例により符号化されたデータを復号化する復号化装置の構成例を示すブロック図である。

【図 9】本発明の一実施の形態例による復号化の流れを示すフローチャートである。

【図 10】図 9 のステップ S 5 2 の復号化処理を示すフローチャートである。

【図 11】本発明を適用した L S T B M と従来例 (R B M 、 T R B M) とを比較した特性図である。

【図 12】符号化前のオリジナルの信号スペクトル (図 1 2 A) と、本発明を適用した L S T B M で符号化した信号スペクトル (図 1 2 B) とを対比して示す波形図である。

10

【発明を実施するための形態】

【 0 0 1 2 】

以下、本発明の好適な一実施の形態例について説明する。

【 0 0 1 3 】

[1 . 符号化装置の構成例]

図 1 は、本発明の一実施の形態例にかかる符号化装置の構成例を示す図である。図 1 に示すように、コンピュータ (P C) 等により構成される符号化装置 1 は、パラメータ学習ユニット 1 1 と符号化処理ユニット 1 2 とパラメータ記憶ユニット 1 4 とを備える。

パラメータ学習ユニット 1 1 は、符号化を行うデータと同じ種類のデータについて事前に学習処理を行い、符号化に必要なパラメータを得る。パラメータ学習ユニット 1 1 が決定したパラメータは、パラメータ記憶ユニット 1 4 に記憶される。符号化処理ユニット 1 2 は、学習処理でパラメータ記憶ユニット 1 4 に記憶されたパラメータを使って、入力データ (符号化用データ) の符号化を行う。

20

符号化を行う入力データは、音声データ、画像データなどの様々な時系列データを含む。

【 0 0 1 4 】

パラメータ学習ユニット 1 1 は、系列データ取得部 1 1 1 と前処理部 1 1 2 とパラメータ推定部 1 1 3 とを備える。系列データ取得部 1 1 1 には、学習用時系列データが供給される。系列データ取得部 1 1 1 で取得した学習用時系列データは、前処理部 1 1 2 で前処理が行われた後、パラメータ推定部 1 1 3 に供給される。

30

例えば、系列データ取得部 1 1 1 で取得される学習用時系列データが音声データの場合には、前処理部 1 1 2 は、学習用の音声データを単位時間ごと (以下、フレームという) に切り出して、M F C C (Mel Frequency Cepstrum Coefficients : メル周波数ケプストラム係数) やメルケプストラム特徴量などのフレームごとの音声信号のスペクトル特徴量を計算し、これを正規化する。

【 0 0 1 5 】

パラメータ推定部 1 1 3 は、可視素子推定部 1 1 3 1 と隠れ素子推定部 1 1 3 2 とによって構成される確率モデルを持つ。本実施の形態例では、可視素子推定部 1 1 3 1 および隠れ素子推定部 1 1 3 2 で構成される確率モデルとして、可視素子と、潜在的な情報を表現する隠れ素子との間に、結合重みが存在すると仮定した R B M (制限ボルツマンマシン) を使用する。パラメータ推定部 1 1 3 は、この結合重みの情報についても推定して持つ。この R B M の詳細については後述する。

40

【 0 0 1 6 】

符号化処理ユニット 1 2 は、系列データ取得部 1 2 1 と前処理部 1 2 2 と符号化部 1 2 3 とを備える。

系列データ取得部 1 2 1 には、符号化用系列データが供給される。系列データ取得部 1 2 1 で取得された符号化用系列データは、前処理部 1 2 2 で前処理が行われた後、符号化部 1 2 3 に供給される。

前処理部 1 2 2 は、パラメータ学習ユニット 1 1 の前処理部 1 1 2 と同じ構成である。

50

【 0 0 1 7 】

符号化部 1 2 3 は、パラメータ学習ユニット 1 1 のパラメータ推定部 1 1 3 と同じ構成であり、可視素子推定部 1 2 3 1 で推定される可視素子と隠れ素子推定部 1 2 3 2 で推定される隠れ素子とによって構成される R B M の確率モデルを備える。可視素子推定部 1 2 3 1 および隠れ素子推定部 1 2 3 2 で、可視素子および隠れ素子を推定する際には、パラメータ学習ユニット 1 1 のパラメータ推定部 1 1 3 で推定されたパラメータが利用される。

【 0 0 1 8 】

符号化装置 1 は、符号化部 1 2 3 の隠れ素子推定部 1 2 3 2 で推定された隠れ素子を、符号化データとして外部に出力する。

なお、図 1 に示す構成では、学習処理を行うパラメータ推定部 1 1 3 と、入力データの符号化処理を行う符号化部 1 2 3 を個別の構成としたが、パラメータ推定部 1 1 3 と符号化部 1 2 3 は、ほぼ同じ機能を有しており、パラメータ推定部 1 1 3 で符号化部 1 2 3 の処理を行うようにしてもよい。系列データ取得部 1 1 1 , 1 2 1 や前処理部 1 1 2 , 1 2 2 についても共通化することが可能である。

【 0 0 1 9 】

図 2 は、符号化装置 1 のハードウェア構成例を示す図である。ここでは、符号化装置 1 をコンピュータ (P C) で構成した例を示す。

図 2 に示すように、符号化装置 1 は、バス 1 0 7 を介して相互に接続された C P U (中央制御ユニット : Central Processing Unit) 1 0 1 、 R O M (Read Only Memory) 1 0 2 、 R A M (Random Access Memory) 1 0 3 、 H D D (Hard Disk Drive) / S S D (Solid State Drive) 1 0 4 、接続 I / F (Interface) 1 0 5 、通信 I / F 1 0 6 を備える。C P U 1 0 1 は、R A M 1 0 3 をワークエリアとして R O M 1 0 2 または H D D / S S D 1 0 4 等に格納されたプログラムを実行することで、符号化装置 1 の動作を統括的に制御する。接続 I / F 1 0 5 は、符号化装置 1 に接続される機器とのインターフェースである。通信 I / F は、ネットワークを介して他の情報処理機器と通信を行うためのインターフェースである。

【 0 0 2 0 】

学習用データや符号化用データの入出力および設定は、接続 I / F 1 0 5 または通信 I / F 1 0 6 を介して行われる。図 1 で説明した符号化装置 1 の機能は、C P U 1 0 1 において所定のプログラムが実行されることで実現される。プログラムは、記録媒体を経由して取得してもよく、ネットワークを経由して取得してもよく、R O M に組み込んで使用してもよい。また、一般的なコンピュータとプログラムの組合せでなく、A S I C (Application Specific Integrated Circuit) や F P G A (Field Programmable Gate Array) などの論理回路を組むことで、符号化装置 1 の構成を実現するためのハードウェア構成にしてもよい。

【 0 0 2 1 】

[2 . R B M の拡張モデル (L S T B M) の例]

次に、符号化装置 1 のパラメータ学習ユニット 1 1 および符号化処理ユニット 1 2 が可視素子と隠れ素子を推定する処理について説明する。

本実施の形態においては、R B M (制限ボルツマンマシン) の拡張モデルである、L S T B M (long short term Boltzmann memory) を新規に開発して、この L S T B M を実装して、時系列データの符号化および復号化を適切に行えるようにした点を特徴とする。L S T B M は、R B M と同様に可視素子と隠れ素子で構成される確率モデルであるが、入力ゲート、出力ゲート、忘却ゲート (forget gate) 、および記憶セルを備えて、これらから計算される短期記憶が次の時刻 (フレーム) のデータに伝播する構造を持つ。

【 0 0 2 2 】

まず、本実施の形態で特徴となる L S T B M を説明する前に、L S T B M の前提となる処理である R B M について説明する。

R B M は入力データを表現する可視素子と、潜在的な情報を表現する隠れ素子の間に双

10

20

30

40

50

方向の接続重みが存在する（但し可視素子間または隠れ素子間には接続はない）と仮定したボルツマンマシン的一种である。ここで説明するRBMは、実数値データを扱うものとする。

I次元の実数ベクトルである可視素子 $\mathbf{v} \in \mathbb{R}^I$ と、J個のバイナリーベクトルである隠れ素子 $\mathbf{h} \in \mathbb{B}^J$ （但し \mathbb{B} は0と1の集合）を表現するRBMは、以下のように定義される。

【0023】

【数1】

$$p(\mathbf{v}, \mathbf{h}) = \frac{1}{Z} e^{-E(\mathbf{v}, \mathbf{h})}$$

【0024】

【数2】

$$E(\mathbf{v}, \mathbf{h}) = \frac{1}{2} \left\| \frac{\mathbf{v} - \mathbf{b}}{\sigma} \right\|_2^2 - \mathbf{c}^\top \mathbf{h} - \left(\frac{\mathbf{v}}{\sigma^2} \right)^\top \mathbf{W} \mathbf{h}$$

【0025】

【数3】

$$Z = \int \sum_{\mathbf{h}} e^{-E(\mathbf{v}, \mathbf{h})} d\mathbf{v}$$

20

【0026】

これらの式において、モデルのパラメータ集合を、 $\Theta = \{\mathbf{b}, \mathbf{c}, \mathbf{W}, \sigma\}$ としたとき、それぞれのパラメータである、 $\mathbf{b} \in \mathbb{R}^I$ は可視素子のバイアス、 $\mathbf{c} \in \mathbb{R}^J$ は隠れ素子のバイアス、 $\mathbf{W} \in \mathbb{R}^{I \times J}$ は可視素子と隠れ素子の間の双方向結合重み、 $\sigma \in \mathbb{R}^I$ は可視素子の偏差を表す。また、式中の分数での表記は除算を示し、 $[\cdot]^2$ は2乗を示し、 E はエネルギーを示し、 \top は転置を示す。

30

このとき、隠れ素子 \mathbf{h} が与えられたときの可視素子の条件付き確率 $p(\mathbf{v} | \mathbf{h})$ 、および可視素子 \mathbf{v} が与えられたときの隠れ素子の条件付き確率 $p(\mathbf{h} | \mathbf{v})$ は、それぞれ以下の単純な式で計算することができる。

【0027】

【数4】

$$p(\mathbf{v} | \mathbf{h}) = \mathcal{N}(\mathbf{v}; \mathbf{W} \mathbf{h} + \mathbf{b}, \text{diag}(\sigma^2))$$

【0028】

【数5】

$$p(\mathbf{h} | \mathbf{v}) = \mathcal{B}(\mathbf{h}; \rho(\mathbf{W}^\top \left(\frac{\mathbf{v}}{\sigma^2} \right) + \mathbf{c}))$$

【0029】

ここで、 $\mathcal{N}(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ は、平均 $\boldsymbol{\mu}$ と分散共分散行列 $\boldsymbol{\Sigma}$ の多変量正規分布、 $\mathcal{B}(\cdot; \boldsymbol{\mu})$ は、

50

)は成功確率の多次元ベルヌイ分布、 (\cdot) は要素ごとのシグモイド関数、 $\text{diag}(\cdot)$ は入力ベクトルを対角成分とする対角行列を返す関数を表す。

【0030】

本実施の形態においては、このRBMを拡張したモデルであるLSTBM(long short term Boltzmann memory)のモデルを実装する。

LSTBMは、背景技術の欄で説明したLSTMと同様に、時刻 t における短期記憶データ $c_t \in \mathbb{R}^J$ は、1つ前の時刻 $t-1$ の短期記憶データ c_{t-1} に忘却ゲート $f_t \in \mathbb{R}^J$ を通したものと、記憶コア(長期記憶)のデータ $g_t \in \mathbb{R}^J$ に入力ゲート $i_t \in \mathbb{R}^J$ を通したものととの和で表現される。すなわち、次式で示される。

【0031】

【数6】

$$c_t \triangleq c_{t-1} \circ f_t + g_t \circ i_t.$$

【0032】

また、次の[数7]式に示すように、この短期記憶データ c_t に出力ゲート $o_t \in \mathbb{R}^J$ を掛け合わせ、非線形関数 ρ を通したものを、時刻毎に伝播する隠れ状態 $h_t \in [0, 1]^J$ と定義する。

【0033】

【数7】

$$h_t \triangleq \rho(c_t \circ o_t)$$

【0034】

LSTBMのモデルでは、各ゲートおよび記憶セルが入力 v_t と作用し合う潜在変数であると、次の[数8]式に示すように、潜在変数に関してマルコフ過程を仮定する。次の式において、 $p(v_1^T, s_1^T)$ は確率密度を示し、[数8]式は確率密度が最大になるように学習することを示す。

【0035】

【数8】

$$\begin{aligned} p(v_1^T, s_1^T) &= \prod_{t=1}^T p(v_t, s_t | s_{t-1}) \\ &= \prod_{t=1}^T p(v_t, s_t | h_{t-1}) \end{aligned}$$

【0036】

但し、[数8]式においては、 $s_t = \{f_t, i_t, o_t, g_t\}$ とする。ここで、時刻 t における条件付き確率 $p(v_t, s_t | h_{t-1})$ を以下の[数9]式~[数11]式のように定める。

【0037】

10

20

30

【数 9】

$$p(\mathbf{v}_t, \mathbf{s}_t | \mathbf{h}_{t-1}) = \frac{1}{Z(\mathbf{h}_{t-1})} e^{-E(\mathbf{v}_t, \mathbf{s}_t | \mathbf{h}_{t-1})}$$

【0038】

【数 10】

$$Z(\mathbf{h}_{t-1}) = \int \sum_{\mathbf{s}_t} e^{-E(\mathbf{v}_t, \mathbf{s}_t | \mathbf{h}_{t-1})} d\mathbf{v}_t$$

【0039】

【数 11】

$$\begin{aligned} E(\mathbf{v}_t, \mathbf{s}_t | \mathbf{h}_{t-1}) &= \frac{1}{2} \left\| \frac{\mathbf{v}_t - \mathbf{b}_v}{\sigma} \right\|_2^2 \\ &\quad - \sum_l \mathbf{b}^{(l)\top} \mathbf{s}_t^{(l)} \\ &\quad - \sum_l \left(\frac{\mathbf{v}_t}{\sigma^2} \right)^\top \mathbf{W}^{(l)} \mathbf{s}_t^{(l)} \\ &\quad - \sum_l \mathbf{h}_{t-1}^\top \mathbf{U}^{(l)} \mathbf{s}_t^{(l)} \end{aligned}$$

【0040】

[数 11] 式右辺第三項は、可視素子 \mathbf{v}_t と潜在変数 \mathbf{s}_t 間のポテンシャルを表し、 \mathbf{v}_t と \mathbf{s}_t は双方向に演算していることを示す。

ここで、 $l = 1, 2, 3, 4$ は、いずれかの潜在変数を指すインデックスであり、次の [数 12] 式のように定義する。

【0041】

【数 12】

$$(\mathbf{s}_t^{(l)}, \mathbf{b}^{(l)}, \mathbf{W}^{(l)}, \mathbf{U}^{(l)}) = \begin{cases} (\mathbf{f}_t, \mathbf{b}_f, \mathbf{W}_f, \mathbf{U}_f) & (l = 1) \\ (\mathbf{i}_t, \mathbf{b}_i, \mathbf{W}_i, \mathbf{U}_i) & (l = 2) \\ (\mathbf{o}_t, \mathbf{b}_o, \mathbf{W}_o, \mathbf{U}_o) & (l = 3) \\ (\mathbf{g}_t, \mathbf{b}_g, \mathbf{W}_g, \mathbf{U}_g) & (l = 4) \end{cases}$$

【0042】

[数 11] 式および [数 12] 式において、 \mathbf{b}_v および $\mathbf{b}^{(1)}$ は可視素子および潜在変数のバイアスパラメータ、 $\mathbf{W}^{(1)}$ は可視素子と各潜在変数との同じ時間内での双方向

接続重み行列パラメータ、 $U^{(l)}$ は時間方向の各潜在変数の再起接続重み行列パラメータ、 σ^2 は可視素子の偏差パラメータを表す。これらのパラメータは、勾配法に基づく最尤推定によって最適化が可能である。各パラメータの勾配には計算困難な項が含まれるが、通常の R B M と同様に、C D (contrastive divergence) 法によって近似することができる。勾配計算には、時刻ステップごとのサンプリングが必要になるが、R T R B M (recurrent T R B M) のように期待値を伝播させて効率よく近似計算させることもできる。

本実施の形態例では、一度、可視素子系列 v_1^T から、各フレームの隠れ状態を潜在変数の期待値 $E[s_t]$ から計算して伝播させた後、全てのフレーム $t = 1, \dots, T$ について、[数 1 3] 式および [数 1 4] 式に示すようにサンプリングし、勾配を計算することでパラメータ更新を行う。

10

【 0 0 4 3 】

【数 1 3】

$$\begin{aligned} \forall l, \tilde{s}_t^{(l)} &\sim p(s_t^{(l)} | v_t, h_{t-1}) \\ &= \mathcal{B}(s_t^{(l)}; \rho(\mathbf{W}^{(l)\top} \left(\frac{v_t}{\sigma^2}\right) + \mathbf{U}^{(l)\top} h_{t-1} + b^{(l)})) \end{aligned}$$

【 0 0 4 4 】

【数 1 4】

$$\begin{aligned} \tilde{v}_t &\sim p(v_t | \tilde{s}_t) \\ &= \mathcal{N}(v_t; \sum_l \mathbf{W}^{(l)} \tilde{s}_t^{(l)} + b_v, \text{diag}(\sigma^2)) \end{aligned}$$

20

【 0 0 4 5 】

正規化した可視素子 v_t を既知の入力としたときの各潜在変数の期待値は、[数 1 5] 式 ~ [数 1 8] 式で示される。

30

【 0 0 4 6 】

【数 1 5】

$$\mathbb{E}[f_t] = \rho(\mathbf{W}_f^\top v_t + \mathbf{U}_f^\top h_{t-1} + b_f)$$

【 0 0 4 7 】

【数 1 6】

$$\mathbb{E}[i_t] = \rho(\mathbf{W}_i^\top v_t + \mathbf{U}_i^\top h_{t-1} + b_i)$$

【 0 0 4 8 】

【数 1 7】

$$\mathbb{E}[o_t] = \rho(\mathbf{W}_o^\top v_t + \mathbf{U}_o^\top h_{t-1} + b_o)$$

【 0 0 4 9 】

【数 1 8】

$$\mathbb{E}[g_t] = \rho(\mathbf{W}_g^T \mathbf{v}_t + \mathbf{U}_g^T \mathbf{h}_{t-1} + b_g)$$

【0050】

これら【数 1 5】式～【数 1 8】式と【数 6】式および【数 7】式とから、LSTMと同様の順伝播の式になることが分かる。すなわち、本実施の形態で説明しているLSTMは、LSTMの拡張表現ということができ、その順伝播は、本来各要素が0または1の値を取る潜在変数を期待値として近似させる場合に、LSTMと一致する。

10

【0051】

図3は、本実施の形態例における、LSTMのモデルを示す。

本実施の形態例のLSTMのモデル構造は、入力ゲート11a、可視素子セル11b、記憶セル11c、演算器11d、11g、11i、短期記憶セル11e、忘却ゲート11f、出力ゲート11h、および隠れ素子セル11jを有する。

入力ゲート11aには時刻tの入力データ i_t が得られる。この入力データ i_t の正規化により、可視素子セル11bに時刻tの可視素子 v_t が得られる。

記憶セル11cには、セルの根幹であるローデータ(生データ) g_t が得られ、このローデータ g_t が記憶処理で記憶される。入力ゲート11aに得られた入力データ i_t と記憶セル11cに記憶されたローデータ g_t とが、演算器11dでドット積され、短期記憶セル11eで短期記憶データ c_t を得るための処理が行われる。

20

【0052】

短期記憶データ c_t は、【数 6】式で説明したように、1つ前の時刻t-1の短期記憶データ c_{t-1} に、忘却ゲート11fのデータ $f_t \cdot B^j$ を、演算器11gで通したものと、記憶コアのデータ $g_t \cdot R^j$ に演算器11dで入力ゲート11aのデータ $i_t \cdot R^j$ を通したものととの和で表現される。このように表現される短期記憶データ c_t が、短期記憶処理で短期記憶セル11eに記憶される。忘却ゲート11fのデータ $f_t \cdot B^j$ は、忘却処理で得られたデータである。

【0053】

さらに、この短期記憶データ c_t に出力ゲート11hのデータ o_t を演算器11iで掛け合わせ、非線形関数を通したものが、時刻ごとに伝播する隠れ状態のデータ $h_t \cdot [0, 1]^j$ となって、隠れ素子セル11jに得られる。

30

なお、 W_i 、 W_g 、 W_f 、 W_o は可視素子 v_t と各潜在変数との双方向接続重み行列パラメータ、 U_i 、 U_g 、 U_f 、 U_o は各潜在変数の再起接続重み行列パラメータである。

【0054】

【3. 学習処理動作および符号化処理動作】

次に、本実施の形態例のLSTMを適用して行われる符号化処理について説明する。

図4は、パラメータ学習ユニット11が行うパラメータ学習動作の流れを示すフローチャートである。

まず、系列データ取得部111が学習用の系列データを取得し(ステップS11)、前処理部112がその系列データの前処理を実行する(ステップS12)。例えば、学習用のデータが音声データである場合には、系列データ取得部111は、学習用音声データを単位時刻(フレーム)ごと(例えば、5 msecごと)に切り出し、切り出された学習用音声信号にFFT処理などを施すことでスペクトル特徴量(例えば、MFCCやメルケプストラム特徴量)を算出する。

40

【0055】

次に、前処理が施された系列データがパラメータ推定部113に供給され、パラメータ推定部113は、系列データのパラメータ学習処理を行う(ステップS13)。ステップS13で行われるパラメータ学習処理の詳細については後述する(図6)。

このパラメータ学習処理でLSTMのモデルが持つ各パラメータが決定され、パラメ

50

ータ記憶ユニット 1 4 に記憶される。そして、パラメータ記憶ユニット 1 4 に記憶されたパラメータが、パラメータを符号化部 1 2 3 に引き渡されて、符号化部 1 2 3 で符号化される（ステップ S 1 4）。

【 0 0 5 6 】

図 5 は、符号化処理ユニット 1 2 が行う符号化処理の流れを示すフローチャートである。

まず、系列データ取得部 1 2 1 が符号化用系列データを取得し（ステップ S 2 1）、前処理部 1 2 2 がその系列データの前処理を実行する（ステップ S 2 2）。ここでの前処理は、前処理部 1 1 2 が行うステップ S 1 2 での前処理と同じである。

【 0 0 5 7 】

前処理が施された系列データは符号化部 1 2 3 に供給され、符号化部 1 2 3 は、ステップ S 1 4 で引き渡された L S T B M のモデルが持つパラメータを使って隠れ素子を推定するとともに、符号化処理を行う（ステップ S 2 3）。ステップ S 2 3 で行われる符号化処理の詳細については後述する（図 7）。そして、符号化処理ユニット 1 2 は、ステップ S 2 3 で得られた隠れ素子を符号化データとして出力する（ステップ S 2 4）。

【 0 0 5 8 】

図 6 は、図 4 のステップ S 1 3 で行われるパラメータ学習処理の詳細を示すフローチャートである。

まず、パラメータ推定部 1 1 3 は、L S T B M のモデルが持つパラメータ、セル、隠れ素子の初期値として任意の値を設定し、設定した値の時刻を初期値（例えば時刻 $t = 0$ ）とする（ステップ S 3 1）。なお、時刻は、フレーム番号などで表現される。次に、パラメータ推定部 1 1 3 の可視素子推定部 1 1 3 1 に、前処理が施された学習用系列の t 番目（時刻 t ）のデータを入力する（ステップ S 3 2）。

その後、パラメータ推定部 1 1 3 は、 t 番目の各隠れ状態の確率値を計算し、計算値をサンプリングする（ステップ S 3 3）。なお、ここで「サンプリングする」とは、条件付き確率密度関数に従うデータをランダムに 1 つ生成することをいい、以下、同じ意味で用いる。

【 0 0 5 9 】

また、パラメータ推定部 1 1 3 は、L S T B M のモデルの t 番目の可視素子の確率値を計算し、計算値をサンプリングする（ステップ S 3 4）。その後、L S T B M のモデルの隠れ状態の確率値を再度計算し、計算値を再サンプリングする（ステップ S 3 5）。そして、パラメータ推定部 1 1 3 は、 t 番目の隠れ素子、セルの値を計算し、時刻 t を一つ進めた時刻 $t + 1$ に更新する（ステップ S 3 6）。

【 0 0 6 0 】

その後、パラメータ推定部 1 1 3 は、現在の時刻 t が終了時刻 T よりも小さいか否か判断し（ステップ S 3 7）、終了時刻 T よりも小さい場合には（ステップ S 3 7 の Y E S）、ステップ S 3 2 に戻り、ここまでの処理を繰り返す。

また、終了時刻 T に到達したと判断したとき（ステップ S 3 7 の N O）、パラメータ推定部 1 1 3 は、各種パラメータを計算した値で更新する（ステップ S 3 8）。そして、各種パラメータの更新後にパラメータ学習処理の終了条件を満足したか否か判断する（ステップ S 3 9）。

【 0 0 6 1 】

ステップ S 3 9 で、終了条件を満足しないと判断した場合には（ステップ S 3 9 の N O）、ステップ S 3 1 に戻り、ここまでの処理を繰り返す。また、ステップ S 3 9 で終了条件を満足したと判断した場合には（ステップ S 3 9 の Y E S）、パラメータ推定部 1 1 3 は、パラメータ学習処理を終了する。なお、ステップ S 3 9 で終了条件を満たさず場合としては、例えば、これら一連のステップの繰り返し数が予め設定された繰り返し数になった場合がある。

【 0 0 6 2 】

図 7 は、図 5 のステップ S 2 3 で行われる符号化処理の詳細を示すフローチャートであ

10

20

30

40

50

る。

まず、符号化部 1 2 3 は、パラメータ推定部 1 1 3 から引き渡されたパラメータを設定する（ステップ S 4 1）。次に、符号化部 1 2 3 の可視素子推定部 1 2 3 1 に、前処理が施された符号化用系列データを入力する（ステップ S 4 2）。

その後、符号化部 1 2 3 の隠れ素子推定部 1 2 3 2 は、L S T B M のモデルの隠れ素子を計算し、推定した隠れ素子を符号化データとして出力する（ステップ S 4 3）。

【 0 0 6 3 】

[4 . 復号化装置の構成および動作]

図 8 は、本発明の一実施形態例に係る符号化装置 1 に対応する復号化装置 2 の構成例を示したものである。

復号化装置 2 は、符号化装置 1 で得られた符号化データを復号化するものであり、例えばコンピュータで構成される。なお、復号化装置 2 は、符号化装置 1 と一体化してもよい。

復号化装置 2 は、パラメータ学習ユニット 1 1 とパラメータ記憶ユニット 1 4 と復号化処理ユニット 1 3 とを備える。

パラメータ学習ユニット 1 1 は、符号化装置 1 のパラメータ学習ユニット 1 1 と同じであり、パラメータ推定部 1 1 3 として、学習処理で得た可視素子および隠れ素子を推定する可視素子推定部 1 1 3 1 および隠れ素子推定部 1 1 3 2 を備える。

パラメータ学習ユニット 1 1 で得たパラメータは、パラメータ記憶ユニット 1 4 に記憶される。

【 0 0 6 4 】

復号化処理ユニット 1 3 には、符号化装置 1 で得られた符号化データが供給される。復号化処理ユニット 1 3 は、復号化部 1 3 1 を備える。復号化部 1 3 1 は、可視素子推定部 1 3 1 1 と隠れ素子推定部 1 3 1 2 とを有し、L S T B M のモデルのパラメータをパラメータ記憶ユニット 1 4 から取得する。

隠れ素子推定部 1 3 1 2 は、入力した符号化データを隠れ素子の推定値とする。そして、可視素子推定部 1 3 1 1 は、L S T B M のモデルのパラメータを使った演算により、可視素子の推定値を得る。この可視素子の推定値は、後処理部 1 3 2 に供給され、後処理部 1 3 2 で後処理が行われる。後処理部 1 3 2 では、例えば符号化装置 1 の前処理部 1 2 2 での前処理を元に戻す処理が行われる。

そして、出力部 1 3 3 は、後処理が行われた復号化データを出力する。

【 0 0 6 5 】

図 9 は、復号化装置 2 での復号化の流れを示すフローチャートである。

復号化装置 2 は、復号化する符号化データを取得すると（ステップ S 5 1）、復号化処理ユニット 1 3 により復号化処理を行う。復号化処理の詳細は後述する（図 1 0）。

復号化処理ユニット 1 3 での処理で得られたデータは、後処理部 1 3 2 に供給されて後処理が行われ（ステップ S 5 2）、後処理されたデータが出力部 1 3 3 から復号化データとして出力される（ステップ S 5 3）。

【 0 0 6 6 】

図 1 0 は、図 9 のフローチャートのステップ S 5 2 での復号化処理の詳細を示す。

まず、復号化部 1 3 1 は、パラメータ記憶ユニット 1 4 から引き渡された L S T B M のモデルの各種パラメータを設定する（ステップ S 6 1）。ここでは、復号化する符号化データを符号化する際に用いたパラメータをパラメータ記憶ユニット 1 4 から取得して設定する。このパラメータは、図 1 に示す符号化装置 1 での符号化時に使用したパラメータである。そして、復号化部 1 3 1 の隠れ素子推定部 1 3 1 2 に、符号化データを入力する（ステップ S 6 2）。また、可視素子推定部 1 3 1 1 は、L S T B M のモデルを使って可視素子（復号化データ）を推定する（ステップ S 6 3）。

このようにして、符号化とは逆の流れで、符号化データの復号化が可能となる。

【 0 0 6 7 】

[5 . 実験例]

次に、本実施の形態例によるLSTBMのモデルの有効性を検証するために実験した例について説明する。

ここでは、本実施の形態例によるLSTBMのモデルの有効性を確認するため、系列データである音声データの符号化を行い、その符号化音声の品質評価実験を行った。具体的には、評価実験用に用意された音声データベースから選んだ音声（女性アナウンサーの音声）を用いて再構築音声の品質評価実験を行った。ここでは、約4.2分の音声について、サンプリングレート20kHzを16kHzにダウンサンプリングし、窓幅256、64サンプルオーバーラップの短時間フーリエ変換を施して得られる振幅スペクトル（ $I = 129$ 次元、全フレーム数64438）を入力特徴量として、LSTBMのモデルで学習を行った。

【0068】

隠れ素子数 J については、 $J = 100$ 、 200 、 400 の3種類で比較した。モデルの学習では、フレーム数 $T = 1000$ のバッチをランダムに選択し、学習率 0.001 、減衰率 $\gamma_1 = 0.9$ 、 $\gamma_2 = 0.999$ で、250回のパラメータ更新を繰り返した。また、従来のRBMのモデルと、TRBMのモデルについても、同様の条件で学習および符号化を行った結果を、表1に示す。

【0069】

【表1】

	Train	Test
LSTBM(100)	3.86	3.81
LSTBM(200)	4.03	3.95
LSTBM(400)	4.11	4.02
RBM(100)	2.74	2.69
RBM(200)	3.29	3.20
RBM(400)	3.72	3.63
TRBM(100)	3.23	3.22
TRBM(200)	3.74	3.75
TRBM(400)	4.00	3.94

【0070】

表1において、LSTBM(100)、LSTBM(200)、LSTBM(400)は、それぞれLSTBMのモデルで、隠れ素子数100、200、400の例を示す。同様に、RBM(100)、RBM(200)、RBM(400)、TRBM(100)、TRBM(200)、TRBM(400)は、それぞれRBMのモデル又はTRBMのモデルで、隠れ素子数100、200、400の例を示す。

また、表1の「Train」は学習データで学習した際の評価値、「Test」は学習後のテスト用音声データによる符号化時の評価値を示す。

【0071】

モデルの評価には、学習用音声およびテスト用音声について、各手法から得られる復元音声に対し、PESQ(perceptual evaluation of speech quality)の手法を用いた客観品質評価を行った。ここでの評価値は、5が最も高い値で、1が最も低い値である。すなわち、5に近い程、高い評価が得られたことを示す。

ここでの復元音声とは、具体的には与えられた音声から潜在変数の期待値を計算（符号化）し、逆に潜在変数から音声の期待値を計算（復号化）して得られる振幅スペクトルと

、オリジナルの位相スペクトルから復元される複素スペクトルに逆フーリエ変換を施し、overlap add 法によって復元した音声信号である。

表 1 から分かるように、L S T B M のモデルは、いずれの隠れ素子数の場合でも、R B M や T R B M よりも高い評価値が得られている。

【 0 0 7 2 】

図 1 1 は、いずれも隠れ素子数 4 0 0 の L S T B M のモデル、R B M のモデル、T R B M のモデルの M S E (Mean Squared Error : 平均二乗誤差) を比較したものである。図 1 1 において、縦軸は M S E、横軸は繰り返し回数 (図 6 のフローチャートのステップ S 3 9 でのループ回数) を示す。

この図 1 1 から、本実施の形態例による L S T B M のモデルが、収束時の M S E が最も低くなっていることが分かる。

10

【 0 0 7 3 】

図 1 2 は、符号化前のオリジナルの音声信号のスペクトル (図 1 2 A) と、本実施の形態例の L S T B M のモデルで符号化した音声信号を復号したスペクトル (図 1 2 B) とを対比して示す波形図である。図 1 2 の縦軸は周波数、横軸は時間 (秒) である。L S T B M のモデルを適用した場合の隠れ素子数は 4 0 0 である。

図 1 2 A と図 1 2 B を比較すると分かるように、図 1 2 A のオリジナルの音声信号のスペクトルと、図 1 2 B の符号化した音声信号のスペクトルはほぼ等しく、L S T B M のモデルによる符号化では、音声の調波構造やフォルマントが正しく表現できている。

【 0 0 7 4 】

以上説明したように、長期記憶と短期記憶の構造を持つボルツマンマシンベースの確率モデルである L S T B M を適用することで、従来よりも高品質な符号化や復号化が行えるようになる。

20

【 0 0 7 5 】

[6 . 変形例]

なお、上述した実施の形態例では、実験例として音声データに適用した場合を説明したが、本発明による L S T B M は、他の様々な時系列の信号の符号化および復号化に適用が可能である。例えば画像データの符号化および復号化に本発明による L S T B M を適用してもよい。さらに、本発明による L S T B M は、音声データや画像データ以外の時系列データの符号化および復号化に適用してもよい。

30

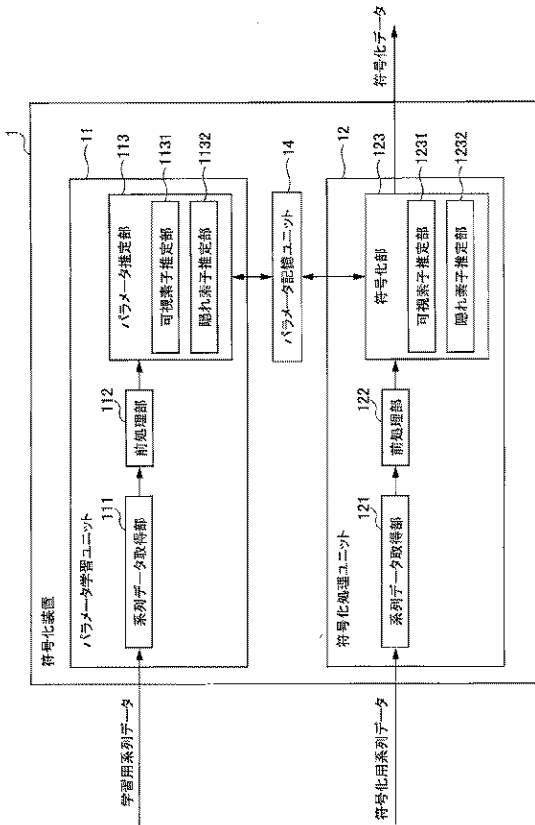
【 符号の説明 】

【 0 0 7 6 】

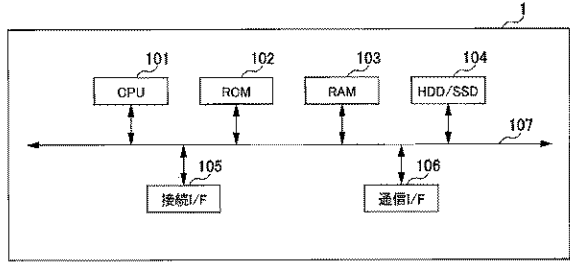
1 . . . 符号化装置、2 . . . 復号化装置、1 1 . . . パラメータ学習ユニット、1 1 a . . . 入力ゲート、1 1 b . . . 可視素子セル、1 1 c . . . 記憶セル、1 1 d , 1 1 g , 1 1 i . . . 演算器、1 1 e . . . 短期記憶セル、1 1 f . . . 忘却ゲート、1 1 h . . . 出力ゲート、1 1 j . . . 隠れ素子セル、1 2 . . . 符号化処理ユニット、1 3 . . . 復号化処理ユニット、1 4 . . . パラメータ記憶ユニット、1 0 1 . . . C P U (中央制御ユニット)、1 0 2 . . . R O M、1 0 3 . . . R A M、1 0 4 . . . H D D / S D D、1 0 5 . . . 接続 I / F、1 0 6 . . . 通信 I / F、1 1 1 , 1 2 1 . . . 系列データ取得部、1 1 2 , 1 2 2 . . . 前処理部、1 1 3 . . . パラメータ推定部、1 2 3 . . . 符号化部、1 3 1 . . . 復号化部、1 3 2 . . . 後処理部、1 3 3 . . . 出力部、1 1 3 1 , 1 2 3 1 , 1 3 1 1 . . . 可視素子推定部、1 1 3 2 , 1 2 3 2 , 1 3 1 2 . . . 隠れ素子推定部

40

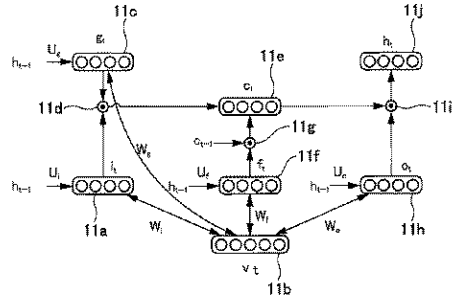
【図1】



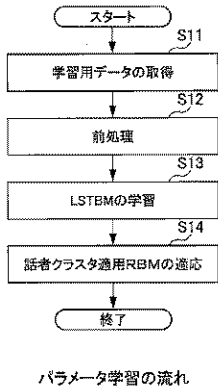
【図2】



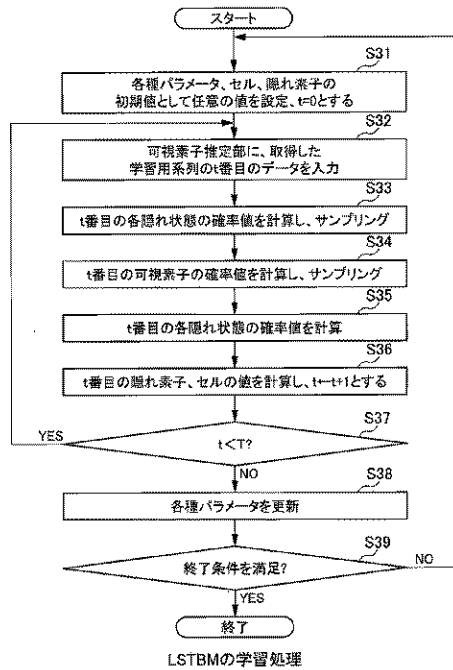
【図3】



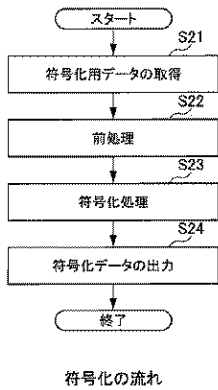
【図4】



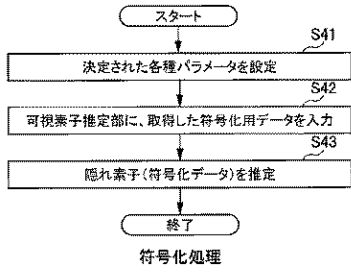
【図6】



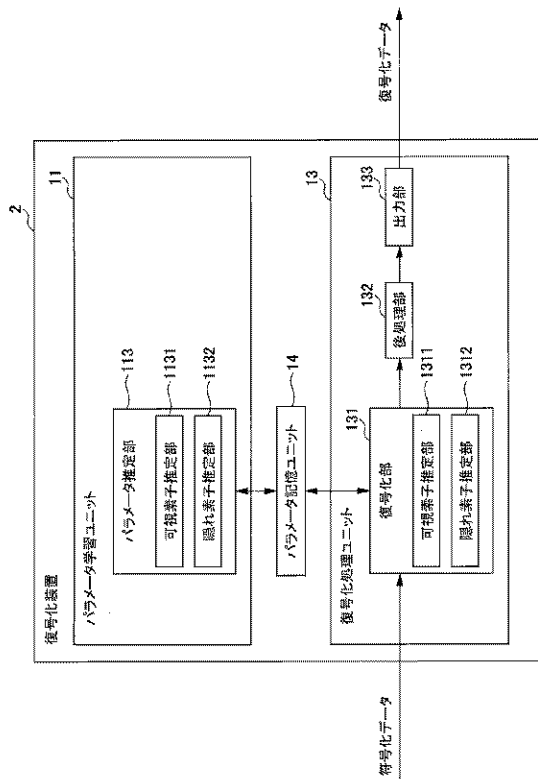
【図5】



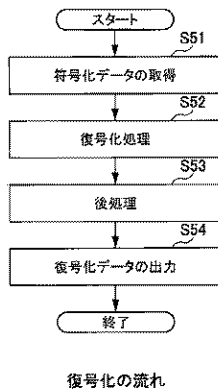
【図7】



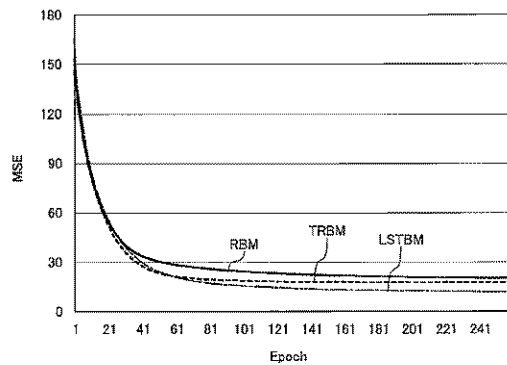
【図8】



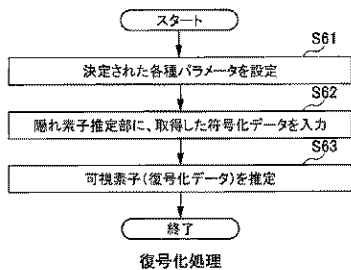
【図9】



【図11】



【図10】



【図12】

