

(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2010-60809

(P2010-60809A)

(43) 公開日 平成22年3月18日(2010.3.18)

(51) Int.Cl. F 1 テーマコード (参考)
G 1 O L 15/10 (2006.01) G 1 O L 15/10 3 O O F 5 D O 1 5
G 1 O L 15/06 (2006.01) G 1 O L 15/06 3 O O C

審査請求 未請求 請求項の数 10 O L (全 18 頁)

(21) 出願番号 特願2008-225998 (P2008-225998) (22) 出願日 平成20年9月3日 (2008.9.3)	(71) 出願人 000004226 日本電信電話株式会社 東京都千代田区大手町二丁目3番1号 (74) 代理人 100121706 弁理士 中尾 直樹 (74) 代理人 100066153 弁理士 草野 卓 (74) 代理人 100128705 弁理士 中村 幸雄 (72) 発明者 マクダモット エリック 東京都千代田区大手町二丁目3番1号 日 本電信電話株式会社内 (72) 発明者 中村 篤 東京都千代田区大手町二丁目3番1号 日 本電信電話株式会社内 Fターム(参考) 5D015 GG04
---	---

(54) 【発明の名称】 誤り数別識別スコア・事後確率計算方法と、その方法を用いた誤り数重み付き識別学習装置と
 その方法と、その装置を用いた音声認識装置と、プログラムと記録媒体

(57) 【要約】

【課題】 誤り数別に識別スコアと事後確率を計算する。

【解決手段】 この発明の誤り数別識別スコア・事後確率計算方法は、局所スコア・局所誤り計算過程と、誤り数別識別スコア計算過程と、誤り数別事後確率計算過程とを含む。誤り数別識別スコア計算過程は、上記局所スコアと、上記局所誤り数と、上記ラティスとを入力として上記ラティスの誤り数別の識別スコアと、誤り数別前向き累積スコアと、誤り数別後ろ向き累積スコアとを計算する。誤り数別事後確率計算過程は、上記ラティスの誤り数別の識別スコアと、上記誤り数別前向き累積スコアと、上記誤り数別後ろ向き累積スコアと、上記局所スコアとを入力として上記正解シンボル系列が上記ラティスに含まれる各々の有向弧を含むことの事後確率を誤り数別に計算する。

【選択図】 図1

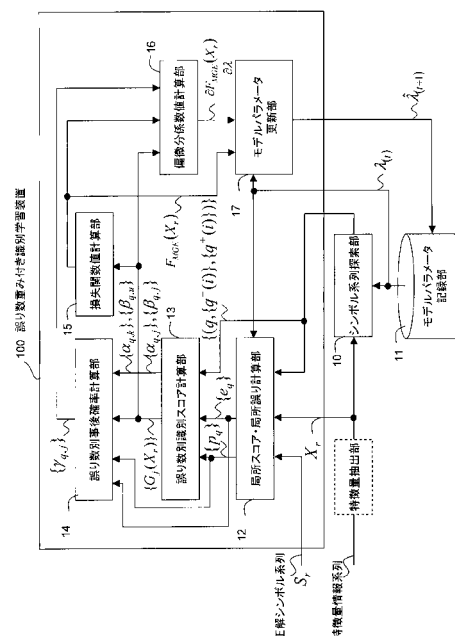


図1

【特許請求の範囲】

【請求項 1】

局所スコア・局所誤り計算部が、特徴量情報系列と、上記特徴量情報系列に対応した正解シンボル系列と、上記特徴量情報系列を複数の認識シンボル系列で表現したラティスと、モデルパラメータとを入力として、上記ラティスに含まれる各々の有向弧に対する局所スコアと、有向弧に含まれる局所誤り数とを計算する局所スコア・局所誤り計算過程と、

誤り数別識別スコア計算部が、上記局所スコアと、上記局所誤り数と、上記ラティスとを入力として、上記ラティスの誤り数別の識別スコアと誤り数別前向き累積スコアと誤り数別後ろ向き累積スコアとを計算する誤り数別識別スコア計算過程と、

誤り数別事後確率計算部が、上記ラティスの誤り数別の識別スコアと、上記誤り数別前向き累積スコアと、上記誤り数別後ろ向き累積スコアと、上記局所スコアとを入力として、上記正解シンボル系列が上記ラティスに含まれる各々の有向弧を含むことの事後確率を誤り数別に計算する誤り数別事後確率計算過程と、

を含む誤り数別識別スコア・事後確率計算方法。

【請求項 2】

請求項 1 に記載した誤り数別識別スコア・事後確率計算方法において、

上記誤り数別識別スコア計算過程は、

上記誤り数別前向き累積スコアを、上記有向弧の先行有向弧の局所スコアと上記先行有向弧までの誤り数別前向き累積スコアの誤り数の和毎に累積した前向き確率総和として求める前向き確率総和算出ステップと、

上記誤り数別後ろ向き累積スコアを、上記有向弧の後続有向弧の局所スコアと上記後続有効弧までの誤り数別後ろ向き累積スコアの誤り数の和毎に累積した後ろ向き確率総和として求める後ろ向き確率総和算出ステップと、

上記ラティスの誤り数別の識別スコアを、上記有向弧の局所スコアと上記前向き確率総和と上記後ろ向き確率総和との誤り数の和毎の事後確率として求める事後確率算出ステップと、

を含むことを特徴とする誤り数別識別スコア・事後確率計算方法。

【請求項 3】

請求項 1 又は 2 に記載した誤り数別識別スコア・事後確率計算方法と、

パターン認識部が、特徴量情報系列を入力として複数のシンボル系列を表現するラティスを上記モデルパラメータ記録部から探索して出力するパターン認識過程と、

損失関数値計算部が、上記ラティスの誤り数別の識別スコアを入力として損失関数値を計算する損失関数値計算過程と、

偏微分係数値計算部が、上記ラティスの誤り数別の識別スコアと、上記損失関数値と、上記誤り数別事後確率とを入力として有向弧での偏微分係数値を計算する偏微分係数値計算過程と、

モデルパラメータ更新部が、上記偏微分係数値と、上記損失関数と、上記モデルパラメータとを入力として上記モデルパラメータを更新するモデルパラメータ更新過程と、

を含む誤り数重み付き識別学習方法。

【請求項 4】

請求項 3 に記載した誤り数重み付き識別学習方法において、

上記損失関数値計算過程は、上記ラティスの誤り数別の識別スコアを制御係数でべき乗した値を、全ての有向弧の識別スコアを上記制御係数でべき乗した値の累積値で除し、更にその値に上記誤り数を乗算した値を上記誤り数で累積して上記損失関数とする過程であることを特徴とする誤り数重み付き識別学習方法。

【請求項 5】

モデルパラメータを記録するモデルパラメータ記録部と、

特徴量情報系列を入力として複数のシンボル系列を表現するラティスを上記モデルパラメータ記録部から探索して出力するパターン認識部と、

上記特徴量情報系列と、上記特徴量情報系列に対応した正解シンボル系列と、上記特徴量情報系列を複数の認識シンボル系列で表現したラティスと、モデルパラメータとを入力として、上記ラティスに含まれる各々の有向弧に対する局所スコアと、有向弧に含まれる局所誤り数とを計算する局所スコア・局所誤り計算部と、

上記局所スコアと、上記局所誤り数と、上記ラティスとを入力として上記ラティスの誤り数別の識別スコアと、誤り数別前向き累積スコアと、誤り数別後ろ向き累積スコアとを計算する誤り数別識別スコア計算部と、

上記ラティスの誤り数別の識別スコアと、上記誤り数別前向き累積スコアと、上記誤り数別後ろ向き累積スコアと、上記局所スコアとを入力として上記正解シンボル系列が上記ラティスに含まれる各々の有向弧を含むことの事後確率を誤り数別に計算する誤り数別事後確率計算部と、

上記ラティスの誤り数別の識別スコアを入力として損失関数値を計算する損失関数値計算部と、

上記ラティスの誤り数別の識別スコアと、上記損失関数値と、上記誤り数別事後確率とを入力として有向弧での偏微分係数値を計算する偏微分係数値計算部と、

上記偏微分係数値と、上記損失関数と、上記モデルパラメータとを入力として上記モデルパラメータを更新するモデルパラメータ更新部と、

を具備する誤り数重み付き識別学習装置。

【請求項 6】

請求項 5 に記載した誤り数重み付き識別学習装置において、

上記損失関数値計算部は、上記ラティスの誤り数別の識別スコアを制御係数でべき乗した値を、上記ラティスの誤り数別の識別スコアの累積値を上記制御係数でべき乗した値で除し、更に上記誤り数を乗算した値を、誤り数で累積して上記損失関数とするものであることを特徴とする誤り数重み付き識別学習装置。

【請求項 7】

請求項 5 又は 6 に記載した誤り数重み付き識別学習装置と、

音声情報列を入力として音声情報列の音声特徴量情報系列を算出する音声特徴量抽出部と、

上記誤り数重み付き識別学習装置で学習した音響モデルを記録する音響モデル記録部と、

上記音声特徴量情報系列に応じて、上記音響モデルパラメータ記録部内に記録された特徴量情報系列を探索して複数のシンボル系列で表現される単語ラティスを出力する単語列探索部と、

を具備する音声認識装置。

【請求項 8】

請求項 5 又は 6 に記載した誤り数重み付き識別学習装置としてコンピュータを機能させるための装置プログラム。

【請求項 9】

請求項 7 に記載した音声認識装置としてコンピュータを機能させるための装置プログラム。

【請求項 10】

請求項 8 又は 9 に記載した装置プログラムを記録したコンピュータで読み取り可能な記録媒体。

【発明の詳細な説明】

【技術分野】

【0001】

この発明は、音声、静止画像、動画像等の時間軸上や空間軸上の概念情報を表現する信号の特徴量情報系列を用いて、予め定められた離散値で表現したシンボル系列としてパターン認識するためのモデルパラメータ学習に用いる誤り数別識別スコア・事後確率計算方法と、その方法を用いた誤り数重み付き識別学習装置とその方法と、その装置を用いた音

10

20

30

40

50

声認識装置と、プログラムと記録媒体に関する。

【背景技術】

【0002】

パターン認識の誤りを少なくするためには、誤り数をモデルパラメータの関数（損失関数）として推定し、この損失関数値が小さくなるようにモデルパラメータの学習を行うことが有効である。このような学習を行う代表的な従来技術として、MPE/MWE（Minimum Phone Error/Minimum Word Error）学習方法が知られている（非特許文献1）。この学習方法を、パターン認識の代表的な例である連続音声認識に適用した場合を例に説明する。

【0003】

図7に音声認識装置800の機能構成例を示す。音声認識装置800は、MPE/MWE学習方法を用いた音響モデル学習装置700と、音声特徴量抽出部80と、単語系列探索部81と、音響モデルパラメータ記録部82とを備える。音響モデル学習装置700は、局所スコア・局所誤り計算部70、誤り数平均値計算部71、前向き・後ろ向きスコア計算部72、事後確率計算部73、損失関数値計算部74、偏微分係数値計算部75、モデルパラメータ更新部76を備える。

【0004】

音声特徴量抽出部80は、離散値で形成される音声情報列を入力として音声情報列の音声特徴量情報系列 X_r を算出する。音声特徴量情報系列 X_r は、例えばメル周波数ケプストラム係数（MFCC）分析によって抽出される。単語系列探索部81は、その音声特徴量情報系列 X_r に応じて、音響モデルパラメータ記録部82内に記録された特徴量情報系列を探索して複数のシンボル系列で表現される単語ラティスを出力する。図8に「今日の北海道」に対応した単語ラティスの一例を示す。単語ラティスは単語グラフとも呼ばれ、複数の節点を、単語あるいは部分単語系列に対応する有向弧 q で結んだ構造を持っている。

【0005】

また、有向弧 q は、音声特徴量情報系列 X_r に含まれる部分特徴量系列にも対応している。ラティスに含まれる有向弧 q の各々と単語あるいは部分単語系列との対応、及びラティスに含まれる有向弧 q の各々と部分特徴量系列との対応は、単語列探索の結果として得られるものである。

【0006】

局所スコア・局所誤り計算部70は、正解系列単語列 S_r と、音声特徴量情報系列 X_r と、モデルパラメータ $\lambda(t)$ と、単語ラティスとを入力として、有向弧 q の局所的なスコアである局所スコア p_q と、局所的な誤り数である局所誤り数 e_q を計算する。局所スコア p_q は、式(1)で計算できる。局所スコア p_q と局所誤り数 e_q は、ラティスに含まれる有効弧 q の各々について計算されるので、有効弧 q についての集合を成している。これを図7では $\{p_q\}$ 、 $\{e_q\}$ と表記している。他の信号についても同様である。

【0007】

【数1】

$$p_q = P_{AG}(s_q)^{\eta\psi} p_{\Lambda_A}(x_q|s_q)^{\psi} \quad (1)$$

【0008】

ここで $P_{AG}(s_q)$ は、有向弧 q に割り付けられた局所言語スコアである。 $p_{\Lambda_A}(x_q|s_q)$ は、有向弧 q に割り付けられた局所音響スコアである。局所言語スコアと局所音響スコアの積が局所スコアとなる。 η と ψ は制御係数であり、 η が大きいほど局所スコアにおける局所言語スコア $P_{AG}(s_q)$ の寄与率が大きくなる係数である。 ψ が局所スコア p_q を計算する際の部分単語系列 s_q 毎のバラツキを抑制（ $\psi \rightarrow$ 小）、又は強調（ $\psi \rightarrow$ 大）する係数である。また、局所スコア・局所誤り計算部70は、局所誤り数 e_q を計数する。局所誤り数 e_q の計数例を図9に示す。図9(a)は単語単位の誤り数の例を示す。図9(b)は音素単位の誤り数、図9(c)は時間フレーム単位の誤り数の例を示す

10

20

30

40

50

。全て、正解文を「今日の北海道の天気は晴れです」としたときのそれぞれの単位毎の誤り数の例である。例えば、単語単位では誤り数が2、音素単位では誤り数が5、フレーム単位では誤り数が35である。この局所誤り数 e_q は、正解単語系列 S_r と、音声特徴量情報系列 X_r と、単語ラティス Q とを対比させることで計数する。ここで Q とは、有向弧 q の接続関係を含む集合である。

【0009】

誤り数平均値計算部71は、有向弧 q を通る全ての単語系列について式(2)に示す誤り数の平均値 $c(q)$ を計算する。

【0010】

【数2】

$$c(q) = \frac{\sum_{S|q \in S} E(S, S_r) P_{\Lambda}(S|X_r)}{\sum_{S|q \in S} P_{\Lambda}(S|X_r)} \quad (2)$$

ここで、 S は単語系列の全体を意味し、 E は単語系列全体の誤り数を意味する。

【0011】

誤り数の平均値 $c(q)$ は、前向き・後ろ向きアルゴリズムと、平均誤り数伝播アルゴリズムとの組み合わせによって効率的に計算することができる。つまり、式(3)に示すように局所誤り数 e_q に、前向き平均誤り数 α'_q と後ろ向き平均誤り数 β'_q とを加えた値で求めることができる。

【0012】

【数3】

$$c(q) = \alpha'_q + e_q + \beta'_q \quad (3)$$

【0013】

前向き平均誤り数 α'_q の求め方の概念図を10に示す。前向き平均誤り数 α'_q を求める前に有向弧 q の前向き累積確率 α_q を式(4)の計算により求める。有向弧 q の前向き累積確率 α_q は、有向弧 q の始端に終端を接続している先行有向弧 $q^-(i)$ ($i = 1, \dots, N_q$) のそれぞれの前向き向き累積確率 $\alpha_{q^-(i)}$ に、それぞれの局所確率 $p_{q^-(i)}$ を乗じた値の累積である。

【0014】

【数4】

$$\alpha_q = \sum_{i=1, N_q} \alpha_{q^-(i)} \cdot p_{q^-(i)} \quad (4)$$

【0015】

有向弧 q の前向き累積確率 α_q と後述する後ろ向き累積確率 β_q は、前向き・後ろ向きスコア計算72で、局所スコア p_q を入力として計算される。前向き累積確率 α_q と後述する後ろ向き累積確率 β_q は、誤り数平均値計算部71と事後確率計算部73に出力される。

【0016】

誤り数平均値計算部71は、局所スコア p_q と、局所誤り数 e_q と、単語ラティス q を入力として式(3)の計算によって有向弧 q を通る全ての単語系列についての誤り数の平均値 $c(q)$ を計算する。式(3)の計算に当たって、必要な前向き平均誤り数 α'_q は式(5)で計算される。

【0017】

10

20

30

40

【数 5】

$$\alpha'_q = \frac{\sum_{i=1, N_q^-} \alpha_{q^-(i)} p_{q^-(i)} (\alpha'_{q^-(i)} + e_{q^-(i)})}{\sum_{i=1, N_q^-} \alpha_{q^-(i)} p_{q^-(i)}} \quad (5)$$

$e_{q^-(i)}$ は先行有向弧 $q^-(i)$ の局所誤り数である。

【0018】

同様に後ろ向き累積確率 β_q と、後ろ向き平均誤り数 β'_q は、式(6)と式(7)で計算できる。

【0019】

10

【数 6】

$$\beta_q = \sum_{i=1, N_q^+} \beta_{q^+(i)} \cdot p_{q^+(i)} \quad (6)$$

$$\beta'_q = \frac{\sum_{i=1, N_q^+} \beta_{q^+(i)} p_{q^+(i)} (e_{q^+(i)} + \beta'_{q^+(i)})}{\sum_{i=1, N_q^+} \beta_{q^+(i)} p_{q^+(i)}} \quad (7)$$

【0020】

誤り数の平均値 $c(q)$ の求め方の概念図を図11に示す。有向弧 q に割り付けられた部分単語系列 s_q が例えば「東海道」であり、その有向弧の誤り数の平均値 $c(q)$ は式(3)で与えられる。

20

【0021】

単語ラティスの始節点を始端とする有向弧から順に漸化的に式(4)、式(5)を適用し、さらに単語ラティスの終節点を終端とする有向弧から順に漸化的に式(6)、式(7)を適用すれば全ての有向弧 q について、式(3)の計算により誤り数の平均値 $c(q)$ を求めることができる。有向弧 q は $((q, \{q^-(i)\}, \{q^+(i)\}) \in Q)$ である。以降、簡単に $q \in Q$ と表記する。

【0022】

事後確率計算部73は、前向き累積確率 α_q と後ろ向き累積確率 β_q とを入力として、認識候補単語系列が有向弧 q を通ることの事後確率 $P_\Lambda(q | X_r)$ を式(8)で計算する。

30

【0023】

【数 7】

$$P_\Lambda(q | X_r) = \frac{\sum_{S|q \in S} P_\Lambda(S | X_r)}{\sum_{q \in Q} \sum_{S|q \in S} P_\Lambda(S | X_r)} \quad (8)$$

【0024】

損失関数値計算部74は、誤り数の平均値 $c(q)$ と、事後確率 $P_\Lambda(q | X_r)$ とを入力として式(9)の計算を行って単語ラティスに対する損失関数 $F_{MPE}(X_r)$ を求める。

40

【0025】

【数 8】

$$F_{MPE}(X_r) = \sum_{q \in Q} c(q) P_\Lambda(q | X_r) \quad (9)$$

【0026】

損失関数 $F_{MPE}(X_r)$ を、最適化手法を用いてモデルパラメータ λ について最小化することにより、認識誤りを少なくする。具体的な最適化手法としては、確率的降下 (PD

50

: Probabilistic Descent) 法、Quickprop法、Rprop法、拡張Baum-Welch (EBW: Extended Baum-Welch) 法等を用いることができる。何れの最適化手法においてもモデルパラメータ λ ($\in \Lambda$) についての $F_{MPE}(X_r)$ の偏微分係数値 $\partial F_{MPE}(X_r) / \partial \lambda$ の計算が必要となる。偏微分係数値 $\partial F_{MPE}(X_r) / \partial \lambda$ は、 $\log p_{\Lambda}(X_r, q) - \log(P_{\Lambda G}(q) - \eta \Psi p_{\Lambda A}(X_r, q) \Psi)$ と置いたとき、式 (10) に示すように分解できる。

【0027】

【数9】

$$\frac{\partial F_{MPE}(X_r)}{\partial \lambda} = \frac{\partial F_{MPE}(X_r)}{\partial \log p_{\Lambda}(X_r, q)} \frac{\partial \log p_{\Lambda}(X_r, q)}{\partial \lambda} \quad (10)$$

【0028】

式 (10) の中で特に、 X_r が有向弧 q に対応する部分単語系列を含む単語列を意図して発声された音声の特徴量情報系列であることの尤もらしさを表す識別スコア $\log p_{\Lambda}(X_r, q)$ についての偏微分係数値が重要である。

【0029】

偏微分係数値計算部 75 は、誤り数の平均値 $c(q)$ と、事後確率 $P_{\Lambda}(q | X_r)$ と、損失関数 $F_{MPE}(X_r)$ とを入力として、有向弧 q での偏微分係数値を式 (11) の計算で求める。

【0030】

【数10】

$$\frac{\partial F_{MPE}(X_r)}{\partial \log p_{\Lambda}(X_r, q)} = P_{\Lambda}(q | X_r) (c(q) - F_{MPE}(X_r)) \quad (11)$$

【0031】

つまり、単語ラティスに適用される MPE/MWE 学習法では、有向弧 q の誤り数の平均値 $c(q)$ と、認識誤り数の平均値 $F_{MPE}(X_r)$ との差、つまり認識誤り数の平均値 $F_{MPE}(X_r)$ を基準にして学習を行う。関数 $F_{MPE}(X_r)$ よりも小さな誤り数の平均値 $c(q)$ を持つ場合は、識別スコア $\log p_{\Lambda}(X_r, q)$ を高くするように学習を進める。逆に有向弧 q が $F_{MPE}(X_r)$ よりも大きな誤り数の平均値 $c(q)$ を持つ場合は、識別スコア $\log p_{\Lambda}(X_r, q)$ を低くするように学習を進める。これを繰り返すことで、損失関数 $F_{MPE}(X_r)$ を最小化することができる。モデルパラメータ更新部 76 は損失関数が低くなるように音響モデルパラメータ記録部 82 に記録されたモデルパラメータ $\lambda_{(t)}$ を、モデルパラメータ $\lambda_{(t+1)}$ に更新する。

【非特許文献1】D.Povey and P.Woodland, "Minimum Phone Error and I-smoothing for improved discriminative training," in Proc.ICASSP02, pp.105-108, 2002.

【発明の開示】

【発明が解決しようとする課題】

【0032】

従来 of MPE/MWE 学習法では、対象とするシンボル系列集合内での認識誤り数の平均値と、ラティス全体としての認識誤り数の平均値との差と、対象とする単語系列集合の事後確率の大きさを基準にしてモデルパラメータの学習を行っている。そのため、次のような問題点が生じる。

【0033】

学習データの量が十分でない場合、或いは、学習データと認識対象とする特徴量情報系列とが異なる統計的性質を有している場合には、認識対象とする特徴量情報系列の誤り数を十分削減する学習効果が得られない。特に認識対象とする特徴量情報系列を初期モデル

で認識した結果と、学習データを認識した結果とで、誤り数毎のシンボル系列数の分布が大きく異なる場合には学習効果が得られない。

【0034】

また、探索部の認識結果出力特性によるが、探索部から出力されるラティスは、同じシンボルの並びでもシンボル境界の時刻のみが異なるシンボル系列が多数含まれることがある。このような場合には、誤り数毎のシンボル系列数の分布の偏りが大きくなり学習の効果は限定的になる。

【0035】

この発明は、このような問題点に鑑みてなされたものであり、学習データの偏りやシンボル系列の出現傾向の偏りによる悪影響が小さく、より高い認識精度が得られるパラメータの学習に用いる誤り数別識別スコア・事後確率計算方法とその方法を用いた誤り数重み付き識別学習装置とその方法と、その装置を用いた音声認識装置と、そのプログラムと記録媒体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】

【0036】

この発明の誤り数別識別スコア・事後確率計算方法は、局所スコア・局所誤り計算過程と、誤り数別識別スコア計算過程と、誤り数別事後確率計算過程とを含む。局所スコア・局所誤り計算過程は、特徴量情報系列と、上記特徴量情報系列に対応した正解シンボル系列と、上記特徴量情報系列を複数の認識シンボル系列で表現したラティスと、モデルパラメータとを入力として、上記ラティスに含まれる各々の有向弧に対する局所スコアと有向弧に含まれる局所誤り数とを計算する。誤り数別識別スコア計算過程は、上記局所スコアと、上記局所誤り数と、上記ラティスとを入力として、上記ラティスの誤り数別の識別スコアと誤り数別前向き累積スコアと誤り数別後ろ向き累積スコアとを計算する。誤り数別事後確率計算過程は、上記ラティスの誤り数別の識別スコアと、上記誤り数別前向き累積スコアと、上記誤り数別後ろ向き累積スコアと、上記局所スコアとを入力として、上記正解シンボル系列が上記ラティスに含まれる各々の有向弧を含むことの事後確率を誤り数別に計算する。

【0037】

また、この発明の誤り数重み付き識別学習方法は、この発明の誤り数別識別スコア・事後確率計算方法を用いる。また、この発明の音声認識装置は、この発明の誤り数重み付き識別学習方法を用いる。

【発明の効果】

【0038】

この発明の誤り数別識別スコア・事後確率計算方法は、誤り数別識別スコア計算過程でラティスの誤り数別の識別スコアと、誤り数別前向き累積スコアと、誤り数別後ろ向き累積スコアとを計算し、誤り数別事後確率計算過程で正解シンボル系列がラティスに含まれる各々の有向弧を含むことの事後確率を誤り数別に計算する。つまり、誤り数別に識別スコアと事後確率を計算するので、学習データの偏りや認識シンボル系列の出現傾向の偏りの影響を受け難いモデルパラメータの学習装置に用いることができる誤り数別識別スコア・事後確率計算方法とすることができる。

【0039】

また、この発明の誤り数別識別スコア・事後確率計算方法を用いた誤り数重み付き識別学習方法は、認識性能を向上させる学習方法とすることができる。更に、その誤り数重み付き識別学習方法を用いた音声認識装置は、認識率を向上させたものにすることができる。

【発明を実施するための最良の形態】

【0040】

以下、この発明の実施の形態を図面を参照して説明する。複数の図面中同一のものには同じ参照符号を付し、説明は繰り返さない。

【実施例1】

【0041】

図1にこの発明の誤り数重み付き識別学習装置100の機能構成例を示す。その動作フローを図2に示す。誤り数重み付き識別学習装置100は、局所スコア・局所誤り計算部12と、誤り数別識別スコア計算部13と、誤り数別事後確率計算部14と、損失関数値計算部15と、偏微分係数値計算部16と、モデルパラメータ更新部17とを備える。誤り数重み付き識別学習装置100の外部に設けられたシンボル系列探索部10とモデルパラメータ記録部11とは、従来技術で説明した音響モデル学習装置700の外部に設けられた単語系列探索部81と音響モデルパラメータ記録部82と、基本的に同じものである。また、誤り数重み付き識別学習装置100内部の局所スコア・局所誤り計算部12とモデルパラメータ更新部17とは、従来技術の音響モデル学習装置700の局所スコア・局所誤り計算部70とモデルパラメータ更新部76と、基本的に同じものである。つまり、これらの部分については、モデルパラメータが音声認識用であるか否かの違いしかない。

10

【0042】

以降の説明では異なる部分のみを説明する。図2のステップS10とステップS12までの動作は、従来の音響モデル学習装置700と同じである。なお、誤り数重み付き識別学習装置100、及び音響モデル学習装置700は、例えばROM、RAM、CPU等で構成されるコンピュータに所定のプログラムが読み込まれて、CPUがそのプログラムを実行することで実現される。

20

【0043】

誤り数別識別スコア計算部13は、局所スコア・局所誤り計算部12で計算された局所スコア p_q と、局所誤り数 e_q と、特徴量情報系列を複数の認識シンボル系列で表現したラティスとを入力として、式(12)に示すラティスの誤り数別の識別スコア $G_j(X_r)$ (以降、識別スコア $G_j(X_r)$ と略す)を求める(ステップS13)。

【0044】

【数11】

$$G_j(X_r) = \sum_{S|E(S,S_r)=j} P_{\Lambda G}(S)^{\eta\psi} p_{\Lambda_A}(X_r|S)^{\psi} \quad (12)$$

【0045】

識別スコア $G_j(X_r)$ は、誤り数別の前向き・後ろ向きアルゴリズムによって効率的に計算することができる。今、有向弧 q での局所的なスコアを局所スコア p_q 、有向弧 q での局所的な誤り数を局所誤り数 e_q とする。連続パターン認識の結果として有向弧 q に割り付けられた部分特徴量情報系列を x_q 、部分単語系列を s_q とすると、有向弧 q の局所スコア p_q は従来技術と同じように式(1)で計算することができる。有向弧 q の局所スコアと、先行有向弧の全て $q^-(i)$ ($i=1, \dots, N_{q^-}$)の前向き累積スコア $\alpha_{q^-(i)}$ とで、前向き確率の総和 $\alpha_{q,j}$ を式(13)で計算する。

30

【0046】

【数12】

$$\alpha_{q,j} = \sum_{\forall(i,k)|k+e_{q^-(i)}=j} \alpha_{q^-(i),k} \cdot p_{q^-(i)} \quad (13)$$

40

【0047】

つまり、有向弧 $q(i)$ と先行有向弧以前の累積誤り数 k の全ての組み合わせにおいて、先行有向弧以前の累積誤り数 k と、先行有向弧 $q^-(i)$ の局所誤り数 $e_{q^-(i)}$ との和である j 毎に、前向き確率の総和を求める。

【0048】

単語ラティスの始節点を始端とする有向弧から順に、先行有向弧を漏れなく用いて漸化的に式(13)を適用すれば、全ての有向弧 $q \in Q$ について前向き累積誤り数別の前向き確率 $\alpha_{q,j}$ (以降、誤り数別前向き累積スコアと称する)を求めることができる。なお

50

、先頭の有向弧については、始節点を終端とした $p_{q^-(1)} = 1$ 、 $\alpha_{q^-(1), j} = 1$ （全ての j について）、 $e_{q^-} = 0$ の仮想の先行有向弧 $q^-(1) = q^{\text{start}}$ を考える。

【0049】

同様に、単語ラティスの終節点を終端とする有向弧から順に後続有向弧を漏れなく用いて漸化的に式（14）を適用すれば、全ての有向弧 $q \in Q$ について後ろ向き累積誤り数別の後ろ向き確率の総和 $\beta_{q, j}$ （以降、誤り数別後ろ向き累積スコアと称する）を求めることができる。

【0050】

【数13】

$$\beta_{q,j} = \sum_{\forall(i,u) \mid u+e_{q^+(i)}=j} \beta_{q^-(i),u} \cdot p_{q^-(i)} \quad (14)$$

【0051】

全ての有向弧 $q \in Q$ について、誤り数別前向き累積スコア $\alpha_{q, j}$ 、又は誤り数別後ろ向き累積スコア $\beta_{q, j}$ を計算すれば、識別スコア $G_j(X_r)$ は式（15）で得られる。

【0052】

【数14】

$$G_j(X_r) = \alpha_j^{\text{final}} = \beta_j^{\text{start}} \quad (15)$$

【0053】

つまり識別スコア $G_j(X_r)$ は、終端の有向弧の誤り数別前向き累積スコア $\alpha_j^{\text{final}} = G_j(X_r)$ 又は、始端の有向弧の誤り数別後ろ向き累積スコア $\beta_j^{\text{start}} = G_j(X_r)$ で表すことができる。

【0054】

誤り数別事後確率計算部14は、識別スコア $G_j(X_r)$ と、誤り数別前向き累積スコア $\alpha_{q, j}$ と、誤り数別後ろ向き累積スコア $\beta_{q, j}$ と、有向弧 q の局所スコア p_q とを入力として、誤り数 j の認識シンボル系列が有向弧 q を通ることの事後確率 $\gamma_{q, j}$ を式（16）で計算する（ステップS14）。

【0055】

【数15】

$$\gamma_{q,j} = \frac{\sum_{\forall(k,u) \mid k+e_q+u=j} \alpha_{q,k} \cdot p_q \cdot \beta_{q,u}}{G_j(X_r)} \quad (16)$$

【0056】

式（13）～式（16）の動作の概念図を図3に示す。局所誤り $e_q = 2$ の有向弧 q の前向き/後ろ向き累積スコア統合 $\zeta_{q, j}$ （以降、は累積スコア統合 $\zeta_{q, j}$ と略す）式（17）で計算できる。つまり、式（16）の分子である。

【0057】

【数16】

$$\zeta_{q,j} = \sum_{\forall(k,u) \mid k+e_q+u=j} \alpha_{q,k} \cdot p_q \cdot \beta_{q,u} \quad (17)$$

【0058】

累積スコア統合 $\zeta_{q, j}$ は、累積誤り数別の先行有向弧以前の累積スコア $\alpha_{q, k}$ と、後続有向弧までの累積スコア $\beta_{q, u}$ と、事後確率を求めたい有向弧 q の局所スコア p_q との積の累積である。この累積スコア統合 $\zeta_{q, j}$ を識別スコア $G_j(X_r)$ で除した値が、誤り数 j の認識シンボル系列が有向弧 q を通ることの事後確率 $\gamma_{q, j}$ である。 α_q

10

20

30

40

50

$\gamma_{q,j}$ 、 $\beta_{q,j}$ 、 $\gamma_{q,j}$ は、有向弧 q と誤り数 j 毎のそれぞれの値である。

【0059】

損失関数値計算部 15 は、誤り数別の識別スコア $G_j(X_r)$ を入力として、例えば式 (18) に示す損失関数値を計算する (ステップ S15)。

【0060】

【数17】

$$F_{MGE1}(X_r) = \sum_j \frac{G_j(X_r)^\phi}{\sum_i G_i(X_r)^\phi} \quad (18)$$

10

【0061】

ここで ϕ は、誤り数別の識別スコア $G_j(X_r)$ の誤り数 j 毎のバラツキを抑制 ($\phi \rightarrow$ 小)、又は強調 ($\phi \rightarrow$ 大) するための制御係数である。

【0062】

偏微分係数値計算部 16 は、モデルパラメータについての式 (18) の損失関数を最小化する偏微分係数値 (式 (19)) を計算する (ステップ S16)。

【0063】

【数18】

$$\frac{\partial F_{MGE1}(X_r)}{\partial \lambda} = \frac{\partial F_{MGE1}(X_r)}{\partial \log p_\Lambda(X_r, q)} \frac{\partial \log p_\Lambda(X_r, q)}{\partial \lambda} \quad (19)$$

20

【0064】

式 (19) の右辺第一項の値が重要であって、偏微分係数値計算部 16 は、その右辺第一項を、誤り数別の識別スコア $G_j(X_r)$ と、損失関数値 $F_{MGE1}(X_r)$ と、事後確率 $\gamma_{q,j}$ とを入力として式 (20) で計算する。

【0065】

【数19】

$$\begin{aligned} \frac{\partial F_{MGE1}(X_r)}{\partial \log p_\Lambda(X_r, q)} &= \sum_j \frac{\partial F_{MGE1}(X_r)}{\partial \log G_j(X_r)} \frac{\partial \log G_j(X_r)}{\partial \log p_\Lambda(X_r, q)} \\ &= \sum_j \left(\phi \frac{G_j(X_r)^\phi}{\sum_i G_i(X_r)^\phi} (j - F_{MGE1}(X_r)) \gamma_{q,j} \right) \end{aligned} \quad (20)$$

30

【0066】

式 (18) と式 (19) とを用いて最適化手法を適用することで、損失関数値を最小化することができる。最適化の収束を判定するのに特徴量情報系列の総損失 $\Gamma_{MGE1}(Z)$ (式

40

21) とその偏微分係数値 (式 (22)) を用いる。Z は様々な部分特徴量情報系列 $X_{r1},$

$\dots, X_{rm}$ を含む全体の特徴量情報系列 ($Z \in \{X_{r(m)} \mid m = 1, \dots, M\}$) である。

$X_{r(m)}$

は式中の表記が正しい。

【0067】

【数 2 0】

$$\Gamma_{MGE1}(Z) = \sum_{X_{r_m} \in Z} F_{MGE1}(X_{r_m}) \quad (21)$$

$$\frac{\partial \Gamma_{MGE1}(Z)}{\partial \lambda} = \sum_{X_{r_m} \in Z} \frac{\partial F_{MGE1}(X_{r_m})}{\partial \lambda} \quad (22)$$

【0068】

モデルパラメータ更新部 17 は、式 (18) の偏微分係数値と、損失関数値 $F_{MGE1}(X_r)$ と、モデルパラメータ記録部 11 に記録されたモデルパラメータ $\lambda_{(t)}$ とを入力として、最適化手法を用いて損失関数を最小化するモデルパラメータ $\lambda_{(t+1)}$ に更新する (ステップ S17)。

10

【0069】

以上のように誤り数別の識別スコアを用いてモデルパラメータを更新するので学習データの偏りや認識シンボル系列の出現傾向の偏りの影響を受け難いモデルパラメータの学習装置が実現できる。学習データの偏りや認識シンボル系列の出現傾向の偏りの影響を受け難い理由を図 4 と図 5 に示す。図 4 と図 5 は、単語シンボル系列当りの誤り数と、事後確率と偏微分係数値との関係を示したものである。横軸が単語シンボル系列当りの誤り数、図 4 (a)、図 4 (c)、図 5 (a)、図 5 (c) の縦軸が事後確率、図 4 (b)、図 4 (d)、図 5 (b)、図 5 (d) の縦軸が偏微分係数値である。

20

【0070】

部分単語系列毎のバラツキを調整するパラメータ ψ は全て 0.1 に固定されている。誤り数 j 毎のバラツキを調整するパラメータ ϕ は図 4 (a)、(b) と図 5 (a)、(b) が 1.0、図 4 (c)、(d) と図 5 (c)、(d) が $\phi = 0.25$ に設定されている。ここで $\phi = 1$ に設定すると、事後確率と偏微分係数値は、単語シンボル系列の平均値となる。これは式 (18) から明らかである。つまり、図 4 (a)、(b) と図 5 (a)、(b) とは、従来技術の特性を示している。したがって、図 4 (c)、(d) と図 5 (c)、(d) の特性が実施例 1 で得られる事後確率と偏微分係数値である。

【0071】

30

図 4 (a) と図 4 (c)、図 4 (b) と図 4 (d) とを比較すると、従来技術の特性は、単語シンボル系列当りの誤り数が 30 を超えるまで変化がない。それに対して実施例 1 の特性には変化が見られる。図 5 (a) と図 5 (c)、図 5 (b) と図 5 (d) との比較も同じである。図 5 (c)、(d) の特性は、式 (22) と式 (23) に示す損失関数と偏微分係数値を用いたもので、パラメータ $v = 0.25$ 、 $\varepsilon = 10$ としたものである。このように変化が有るということは、学習に寄与することを意味する。

【0072】

損失関数値と偏微分係数値を求める式には、上記したもの以外の式を適用することが可能である。例えば、式 (23) と式 (24) に示すように、誤り数に閾値 ε を設け、閾値 ε を境界として損失関数値に重み付けするようにしてもよい。

40

【0073】

【数 2 1】

$$F_{MGE2}(X_r) = \begin{cases} \sum_j j \left(\frac{G_j(X_r)^\phi}{\sum_{i < \varepsilon} G_i(X_r)^\phi} \right) \times \frac{\left(\sum_{i < \varepsilon} G_i(X_r)^\phi \right)^\nu}{\left(\sum_{i < \varepsilon} G_i(X_r)^\phi \right) + \left(\sum_{i \geq \varepsilon} G_i(X_r)^\phi \right)^\nu} & i < \varepsilon \\ \sum_j j \left(\frac{G_j(X_r)^\phi}{\sum_{i \geq \varepsilon} G_i(X_r)^\phi} \right) \times \frac{\left(\sum_{i \geq \varepsilon} G_i(X_r)^\phi \right)^\nu}{\left(\sum_{i < \varepsilon} G_i(X_r)^\phi \right) + \left(\sum_{i \geq \varepsilon} G_i(X_r)^\phi \right)^\nu} & i \geq \varepsilon \end{cases} \quad (23)$$

$$\frac{\partial F_{MGE2}(X_r)}{\partial \log p_\Lambda(X_r, q)} = \sum_j \left(\phi \nu \frac{G_j(X_r)^\phi}{\sum_i G_i(X_r)^\phi} (j - F_{MGE2}(X_r)) \gamma_{q,j} \right) \quad (24)$$

【0074】

ここで ν は、閾値 ε を境界としてバラツキを制御するパラメータである。また、式 (25) と式 (26) に示すような損失関数値と偏微分係数値を用いてもよい。

【0075】

【数 2 2】

$$F_{MGE3}(X_r) = \sum_j j \frac{G_j(X_r)^\phi e^{-\sigma j}}{\sum_i G_i(X_r)^\phi e^{-\sigma i}} \quad (25)$$

$$\frac{\partial F_{MGE3}(X_r)}{\partial \log p_\Lambda(X_r, q)} = \sum_j \left(\phi \frac{G_j(X_r)^\phi e^{-\sigma j}}{\sum_i G_i(X_r)^\phi e^{-\sigma i}} (j - F_{MGE3}(X_r)) \gamma_{q,j} \right) \quad (26)$$

【0076】

ここで、 σ は誤り数が大きくなるほど指数的に識別スコアを減衰させる減衰係数である。〔応用例〕

この発明の誤り数重み付き識別学習装置 100 を音声認識に応用した例を説明する。図 6 にこの発明の誤り数重み付き識別学習装置 100 を用いて音声認識装置 600 を構成した機能構成例を示す。音声認識装置 600 は、従来技術で説明した音声認識装置 800 の音声認識用学習装置 700 をこの発明の誤り数重み付き識別学習装置 100 に置き換えたものである。

10

20

30

40

50

【0077】

この発明の音声認識装置600は、音響モデルを誤り数別に学習するので、認識スコアの精度を高められる。よって、誤認識の少ない音声認識装置を実現することができる。

【0078】

〔実験結果〕

この発明の誤り数重み付き識別学習方法の効果を確認する目的で実験を行った。この発明による学習方法、損失関数値を式(22)、偏微分係数値を式(23)で求め、 $\psi = 0.04$ 、 $\phi = 0.25$ 、 $v = 0.65$ 、 $\varepsilon = 20$ の条件で、日本語の学会講演約230時間の音声进行学习した。その後、上記学習データとは別の約130分の長さの評価音声を、この発明の音声認識装置700で音声認識した結果の単語誤り率は18.8%であった。従来技術の音声認識装置800で評価音声を音声認識した結果の単語誤り率は19.3%であった。また、初期値のモデルパラメータで音声認識した結果の単語誤り率は21.6%であった。したがって、初期の誤り率を100とした相対誤り削減率はこの発明の方法が13.0%、従来法の削減率が10.6%であり、この発明の学習方法の方が優れた認識性能を示すことが確認できた。

【0079】

なお、この発明の技術思想に基づく誤り数別識別スコア・事後確率計算方法とその方法を用いた誤り数重み付き識別学習装置とその方法は、上述の実施形態に限定されるものではなく、この発明の趣旨を逸脱しない範囲で適宜変更が可能である。上記した装置及び方法において説明した処理は、記載の順に従って時系列に実行されるのみならず、処理を実行する装置の処理能力あるいは必要に応じて並列的あるいは個別に実行されるとしてもよい。

【0080】

また、上記装置における処理手段をコンピュータによって実現する場合、各装置が有すべき機能の処理内容はプログラムによって記述される。そして、このプログラムをコンピュータで実行することにより、各装置における処理手段がコンピュータ上で実現される。

【0081】

この処理内容を記述したプログラムは、コンピュータで読み取り可能な記録媒体に記録しておくことができる。コンピュータで読み取り可能な記録媒体としては、例えば、磁気記録装置、光ディスク、光磁気記録媒体、半導体メモリ等のようなものでもよい。具体的には、例えば、磁気記録装置として、ハードディスク装置、フレキシブルディスク、磁気テープ等を、光ディスクとして、DVD (Digital Versatile Disc)、DVD-RAM (Random Access Memory)、CD-ROM (Compact Disc Read Only Memory)、CD-R (Recordable) / RW (ReWritable) 等を、光磁気記録媒体として、MO (Magneto Optical disc) 等を、半導体メモリとしてフラッシュメモリー等を用いることができる。

【0082】

また、このプログラムの流通は、例えば、そのプログラムを記録したDVD、CD-ROM等の可搬型記録媒体を販売、譲渡、貸与等することによって行う。さらに、このプログラムをサーバコンピュータの記録装置に格納しておき、ネットワークを介して、サーバコンピュータから他のコンピュータにそのプログラムを転送することにより、このプログラムを流通させる構成としてもよい。

【0083】

また、各手段は、コンピュータ上で所定のプログラムを実行させることにより構成することにしてもよいし、これらの処理内容の少なくとも一部をハードウェア的に実現することとしてもよい。

【図面の簡単な説明】

【0084】

【図1】この発明の誤り数重み付き識別装置100の機能構成例を示す図。

【図2】誤り数重み付き識別装置100の動作フローを示す図。

【図3】誤り数別前向き累積スコア～事後確率 $\gamma_{q,j}$ を求める動作を概念的に示す図。

【図 4】学習データの偏りや認識シンボル系列の出現傾向の偏りの影響を受け難い理由を示す図であり、(a)は従来技術の誤り数別事後確率を示す図、(b)は従来技術の偏微分係数値を示す図、(c)は実施例 1 の誤り数別事後確率を示す図、(d)は実施例 1 の誤り数別偏微分係数値を示す図である。

【図 5】学習データの偏りや認識シンボル系列の出現傾向の偏りの影響を受け難い理由を示す図であり、(a)は従来技術の誤り数別事後確率を示す図、(b)は従来技術の偏微分係数値を示す図、(c)は実施例 1 の誤り数別事後確率を示す図、(d)は実施例 1 の誤り数別偏微分係数値を示す図である。

【図 6】この発明の音声認識装置 600 の機能構成を示す図。

【図 7】従来の音響モデル学習装置の機能構成例を示す図。

【図 8】「今日の北海道」に対応した単語ラティスの一例を示す図。

【図 9】局所誤り数 e_q の計数例を示す図であり、(a)は単語単位の誤り数の例を示す図、(b)は音素単位の誤り数を示す図、(c)は時間フレーム単位の誤り数の例を示す図である。

【図 10】前向き平均誤り数 α'_q の求め方の概念図を示す図。

【図 11】誤り数の平均値 $c(q)$ の求め方の概念図を示す図。

10

【図 1】

【図 2】

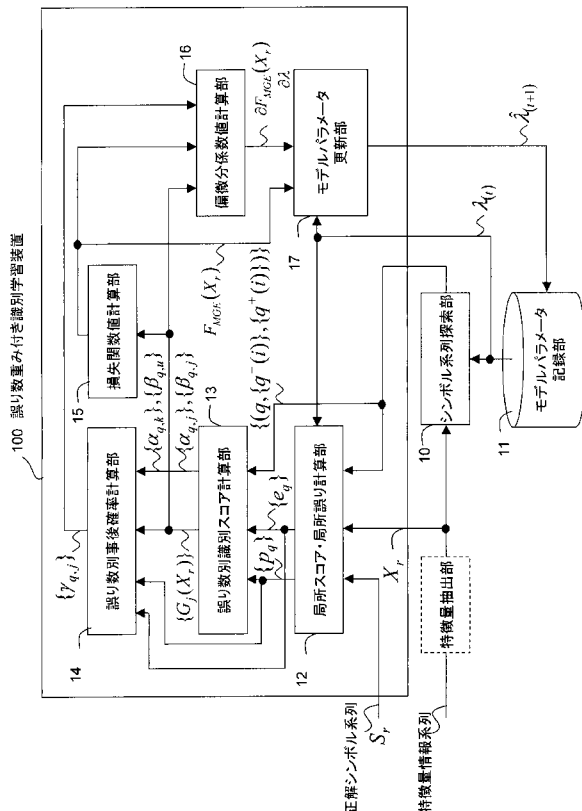


図 1

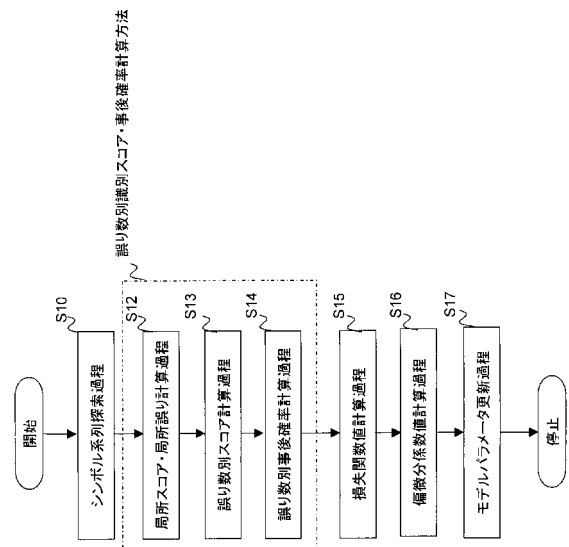
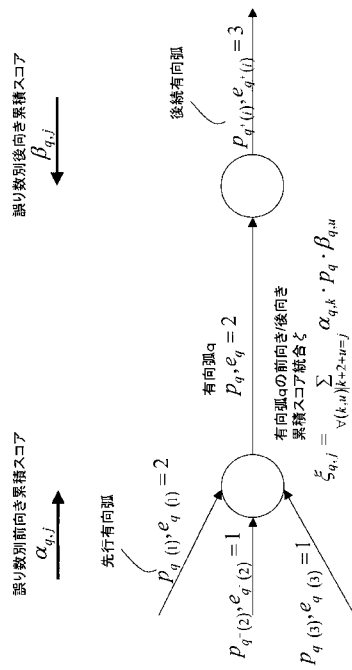
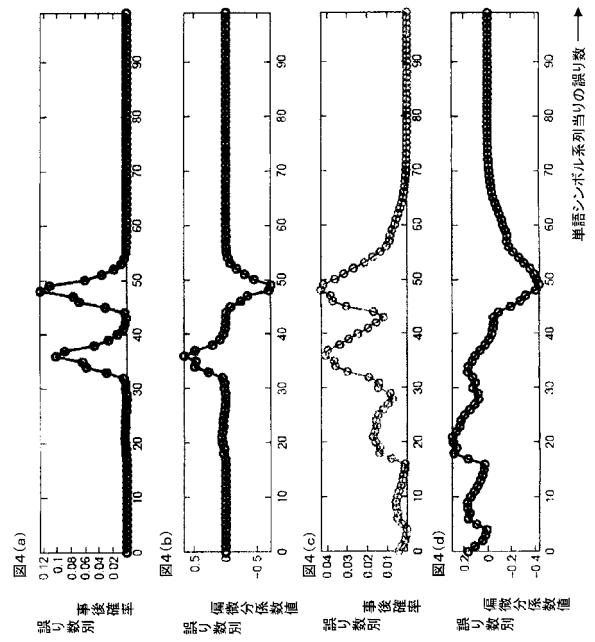


図 2

【 図 3 】

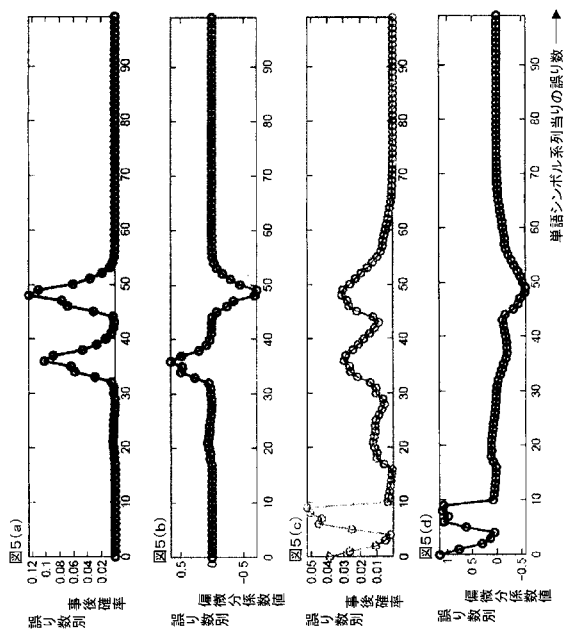


【図 4】



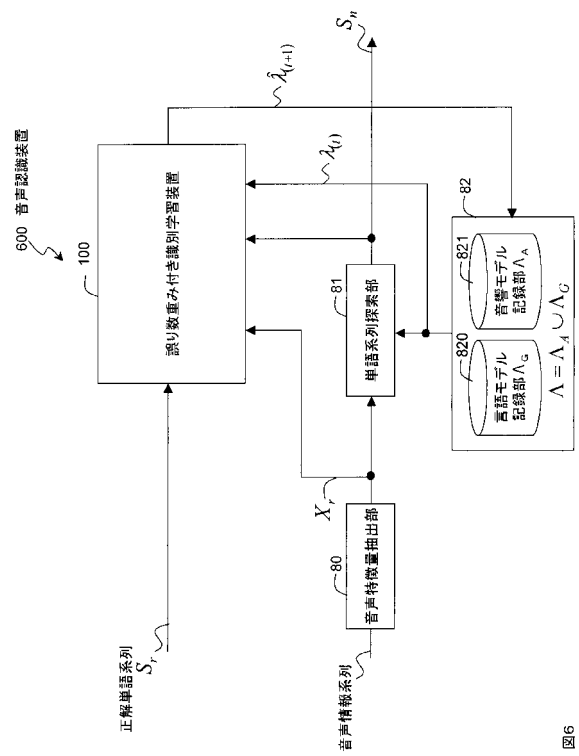
4

【図 5】



5

【図 6】



6

【図 7】

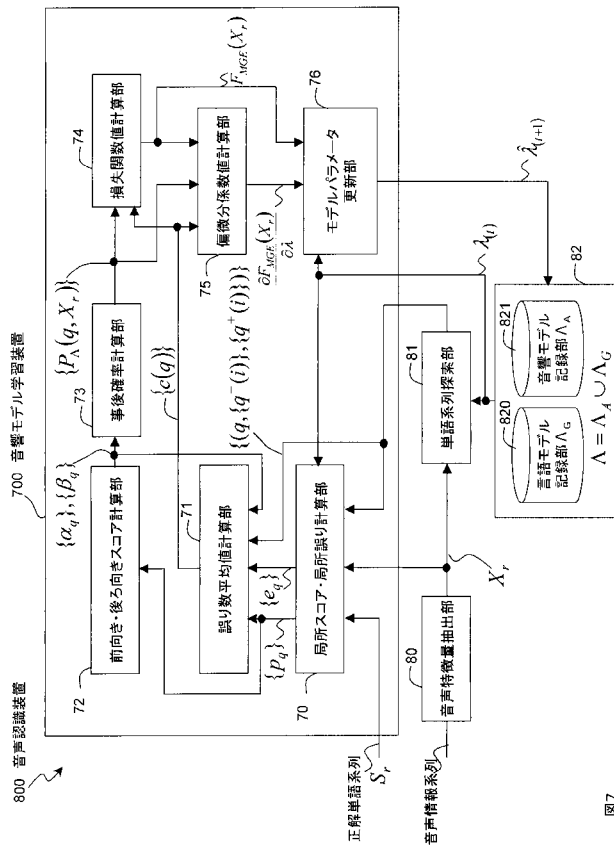


図 7

【図 8】

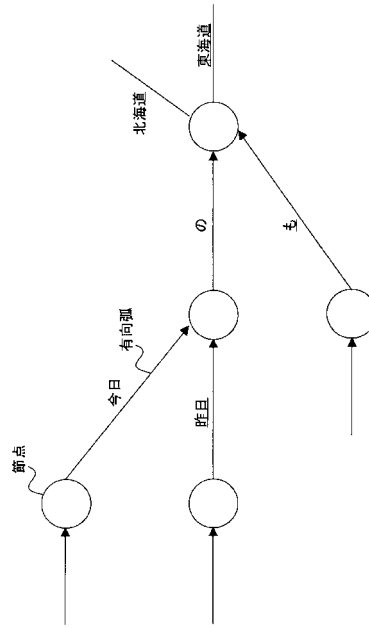


図 8

【図 9】

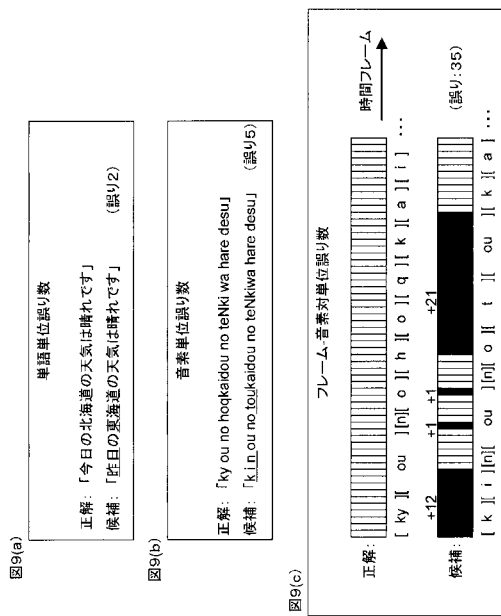


図 9

【図 10】

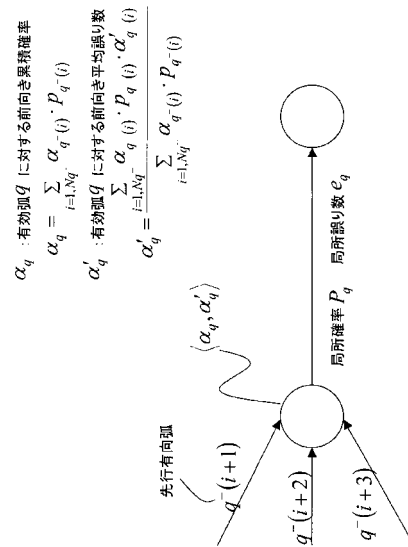


図 10

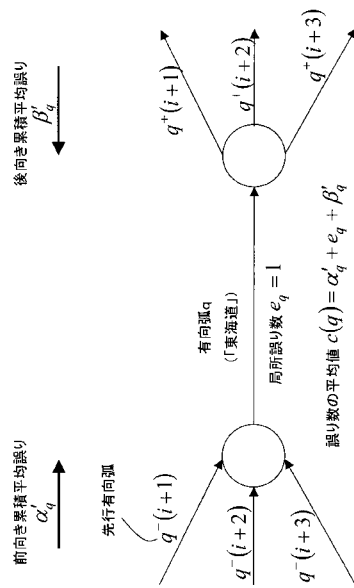


図 1 1