(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第6849621号
(P6849621)
(45)発行日　令和3年3月24日(2021.3.24)　(24)登録日　令和3年3月8日(2021.3.8)

(51)Int.Cl.　F I
G10L 15/16 (2006.01)　G10L 15/16

請求項の数 7　(全 21 頁)

(21)出願番号　特願2018-17224(P2018-17224)
(22)出願日　平成30年2月2日(2018.2.2)
(65)公開番号　特開2019-133084(P2019-133084A)
(43)公開日　令和1年8月8日(2019.8.8)
審査請求日　令和1年12月18日(2019.12.18)

(73)特許権者　000004226
日本電信電話株式会社
東京都千代田区大手町一丁目５番１号
(74)代理人　110002147
特許業務法人酒井国際特許事務所
(72)発明者　小川　厚徳
東京都千代田区大手町一丁目５番１号　日本電信電話株式会社内
(72)発明者　デルクロア　マーク
東京都千代田区大手町一丁目５番１号　日本電信電話株式会社内
(72)発明者　苅田　成樹
東京都千代田区大手町一丁目５番１号　日本電信電話株式会社内

最終頁に続く

(54)【発明の名称】学習装置、学習方法及び学習プログラム

(57)【特許請求の範囲】
【請求項１】
精度が既知である学習用の複数の系列の入力を受け付ける入力部と、
前記複数の系列のうちの二つの系列の特徴量がそれぞれ与えられたとき、それら二つの系列の精度の高低を判定できるような、ニューラルネットワークで表されるモデルを学習する学習部と、
を有し、
前記モデルは、二つの系列を、再帰的ニューラルネットワークを用いて隠れ状態ベクトルに変換し、ニューラルネットワークを用いて、前記隠れ状態ベクトルを基に二つの系列の精度の高低の並びが正しいことを示す第１の事後確率及び二つの系列の精度の高低の並びが誤りであることを示す第２の事後確率を出力することを特徴とする学習装置。
【請求項２】
前記入力部は、音声認識精度が既知である学習用のＮベスト仮説の入力を受け付け、
前記学習部は、前記Ｎベスト仮説の二つの仮説のうち音声認識精度がより高い仮説に他方の仮説よりも高い順位が付与されている場合に正解ラベルを付与して前記モデルに学習させ、前記二つの仮説のうち音声認識精度がより高い仮説に他方の仮説よりも低い順位が付与されている場合に誤りラベルを付与して前記モデルに学習させることを特徴とする請求項１に記載の学習装置。
【請求項３】
前記二つの仮説のうちの一方の仮説は、最も音声認識精度が高いオラクル仮説であるこ

とを特徴とする請求項２に記載の学習装置。
【請求項４】
前記二つの仮説のうちの他方の仮説は、前記オラクル仮説の次に高い音声認識精度を持つ第１の仮説、Ｎベスト仮説における音声認識スコアが最も高い第２の仮説、最も低い音声認識精度を持つ第３の仮説、及び、Ｎベスト仮説における音声認識スコアが最も低い第４の仮説の少なくともいずれかを含むことを特徴とする請求項３に記載の学習装置。
【請求項５】
前記二つの仮説のうちの他方の仮説は、Ｎベスト仮説から、前記オラクル仮説、前記第１の仮説、前記第２の仮説、前記第３の仮説及び前記第４の仮説を除いた仮説から所定のルールにしたがって抽出した所定数の仮説及び前記第１から第４の仮説であることを特徴とする請求項４に記載の学習装置。
【請求項６】
学習装置が実行する学習方法であって、
精度が既知である学習用の複数の系列の入力を受け付ける工程と、
前記複数の系列のうちの二つの系列の特徴量がそれぞれ与えられたとき、それら二つの系列の精度の高低を判定できるような、ニューラルネットワークで表されるモデルを学習する工程と、
を含み、
前記モデルは、二つの系列を、再帰的ニューラルネットワークを用いて隠れ状態ベクトルに変換し、ニューラルネットワークを用いて、前記隠れ状態ベクトルを基に二つの系列の精度の高低の並びが正しいことを示す第１の事後確率及び二つの系列の精度の高低の並びが誤りであることを示す第２の事後確率を出力することを特徴とする学習方法。
【請求項７】
精度が既知である学習用の複数の系列の入力を受け付けるステップと、
前記複数の系列のうちの二つの系列の特徴量がそれぞれ与えられたとき、それら二つの系列の精度の高低を判定できるような、ニューラルネットワークで表されるモデルを学習するステップと、
をコンピュータに実行させ、
前記モデルは、二つの系列を、再帰的ニューラルネットワークを用いて隠れ状態ベクトルに変換し、ニューラルネットワークを用いて、前記隠れ状態ベクトルを基に二つの系列の精度の高低の並びが正しいことを示す第１の事後確率及び二つの系列の精度の高低の並びが誤りであることを示す第２の事後確率を出力する学習プログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、学習装置、学習方法及び学習プログラムに関する。
【背景技術】
【０００２】
音声認識は、人間が発した音声（発話）を計算機により単語列（テキスト）に変換する技術である。通常、音声認識システムは、入力された一つの発話に対して、音声認識スコアの最も高い仮説（音声認識結果）である一つの単語列（１ベスト仮説）を出力する。ただし、音声認識装置による音声認識の精度は、１００％ではない。このため、一つの入力発話に対して、１ベスト仮説のみを出力するのではなく、Ｎ（≧２）個の仮説を出力して、Ｎベストリスコアリング装置を用いて、そのＮ個仮説の中から音声認識精度が最も高いと推定される仮説を最終的な音声認識結果として出力する、Ｎベストリスコアリングと呼ばれる手法がある。なお、Ｎベストリスコアリング（モデル）とＮベストリランキング（モデル）とは同義として扱われている。
【先行技術文献】
【非特許文献】
【０００３】

【非特許文献１】T. Oba, T. Hori, A. Nakamura, and A. Ito, “Round-Robin Duel Discriminative Language Models”, IEEE Transactions on Audio, Speech, and Language Processing, vol. 20, no.4, pp.1244－1255, May 2012.
【非特許文献２】A. Ogawa and T. Hori, “Error detection and accuracy estimation in automatic speech recognition using deep bidirectional recurrent neural networks”, Speech Communication, vol. 89, pp.70－83, May 2017.

【発明の概要】
【発明が解決しようとする課題】
【０００４】
図８は、Ｎベストリスコアリングの処理手順を示す図である。このＮベストリスコアリングでは、まず、１発話の入力を受け付けると（ステップＳ３１）、音声認識を行い（ステップＳ３２）、音声認識結果であるＮ個の仮説を、これらの各スコアを基に降順でソートして出力する（ステップＳ３３）。出力した仮説は、Ｎベスト仮説である。このＮベスト仮説を用いて、Ｎベストリスコアリング処理が実行される（ステップＳ３４）。Ｎベストリスコアリング処理では、後処理として、モデルを用いて、そのＮベスト仮説に対して再度スコア付けを行う。モデルは、例えば、Ｎベストリスコアリングモデルを用いる。
【０００５】
そして、Ｎベストリスコアリング処理では、それらの再度付与されたスコアを基にＮベスト仮説を降順にソートするリランキングを行う。Ｎベストリスコアリング処理では、ソートの結果、Ｎベスト仮説の最上位にランキングされた仮説を抽出し、抽出した仮説を、最終的な音声認識結果として出力する。
【０００６】
Ｎの値としては、通常、１００～１０００程度が設定されることが多い。なお、Ｎを２以上に設定しても、一つの仮説しか得られない場合もある。その場合は、Ｎベストリスコアリングを行う意味はない。
【０００７】
図９は、Ｎベスト仮説の具体例を示す図である。図９では、Ｎを５以上に設定して音声認識を行い、５位までの五つの仮説が得られている。図９において、「[]」は、本来そこには単語がないことを示す。仮説は、音声認識スコアを基準として降順にソートされている。図９の例では、３位仮説が最も音声認識精度が高い（最も誤りが少ない）オラクル仮説である。Ｎベストリスコアリングの処理によって、この３位仮説が１位にリランキングされることが期待される。
【０００８】
ここで、Ｎベストリスコアリングモデルとして、音声認識仮説である単語列の言語としての正しさに着目して、単語のつながり易さを確率統計的に表現する（評価する）言語モデルが採用されることが多い。単語列の言語としての正しさは、単語列の自然さ、単語同士のつながりの正しさである。
【０００９】
従来、例えば、Ｎベスト仮説中の各仮説に含まれる単語のｎ連鎖（ｎは通常１～３程度）を素性として、ログリニアモデルに基づき、認識精度がより高い仮説により高いスコアを与えるような識別的言語モデルが盛んに研究されていた。
【００１０】
また、近年では、ニューラルネットワーク（Neural Network：ＮＮ）の発展に基づき、ＮＮに基づくＮＮ言語モデルのうち、再帰的ニューラルネットワーク（Recurrent Neural Network：ＲＮＮ）に基づくＲＮＮ言語モデルがＮベストリスコアリングモデルとして盛んに利用されている。
【００１１】
ここで、識別的言語モデルは、Ｎベスト仮説を用いて学習されるため、音声認識誤りを考慮したＮベストリスコアリングを行うためのモデルであるものの、最新のＮＮに基づく

モデルではない。
【００１２】
一方、ＲＮＮ言語モデルは、最新のＮＮに基づくモデルであるものの、その学習は誤りを含まない正しい単語列を用いて行われるため、音声認識誤りを考慮できない。また、ＲＮＮ言語モデルは高いＮベストリスコアリング精度を示すものの、本来は単語列が与えられたときに、その単語列の次にどの単語が生起しやすいかを推定するモデルである。すなわち、ＲＮＮ言語モデルは、厳密には、Ｎベストリスコアリングを行うためのモデルではない。言い換えると、ＲＮＮ言語モデルの次単語を推定する機能は、Ｎベストリスコアリングを行う上で必要な機能以上の機能であると言える。
【００１３】
このように、ある一つの入力に対する解の候補として挙げられた複数の系列に対して、最も精度が高い（最も誤りが少ない）候補を判定するために、複数の系列に対して、最も精度が高い候補を判定するうえで最適なモデルを、識別的言語モデルやＲＮＮ言語モデルではなく、最新のＮＮに基づき実現することが期待されている。
【００１４】
本発明は、上記に鑑みてなされたものであって、ある一つの入力に対する解の候補として挙げられた複数の系列に対し、最も精度が高い候補を判定する上で最適なモデルを実現する学習装置、学習方法及び学習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】
【００１５】
上述した課題を解決し、目的を達成するために、本発明に係る学習装置は、精度が既知である学習用の複数の系列の入力を受け付ける入力部と、複数の系列のうちの二つの系列の特徴量がそれぞれ与えられたとき、それら二つの系列の精度の高低を判定できるような、ニューラルネットワークで表されるモデルを学習する学習部と、を有することを特徴とする。
【発明の効果】
【００１６】
本発明によれば、ある一つの入力に対する解の候補として挙げられた複数の系列に対し、最も精度が高い候補を判定する上で最適なモデルを実現する。
【図面の簡単な説明】
【００１７】
【図１】図１は、実施の形態に係るリランキング装置の機能構成の一例を示す図である。
【図２】図２は、Ｎベストリスコアリングモデルの構築例を示す図である。
【図３】図３は、図１に示すリランキング装置が実行するリランキング処理の処理手順を示すフローチャートである。
【図４】図４は、実施の形態に係る学習装置の機能構成の一例を示す図である。
【図５】図５は、図４に示す学習装置が実行する学習処理の処理手順を示すフローチャートである。
【図６】図６は、Ｎベストリランキングの評価結果を示す図である。
【図７】図７は、プログラムが実行されることにより、リランキング装置及び学習装置が実現されるコンピュータの一例を示す図である。
【図８】図８は、Ｎベストリスコアリングの処理手順を示す図である。
【図９】図９は、Ｎベスト仮説の具体例を示す図である。
【発明を実施するための形態】
【００１８】
以下、図面を参照して、本発明の一実施形態を詳細に説明する。なお、この実施の形態により本発明が限定されるものではない。また、図面の記載において、同一部分には同一の符号を付して示している。本実施の形態では、ある正解に対する候補として挙げられた複数の系列として、音声認識結果であるＮ（Ｎ≧２）ベスト仮説を例として説明する。そして、本実施の形態では、Ｎベスト仮説のうち、最終的な音声認識結果である最も音声認

識精度が高い仮説（単語列）を得るためのＮベストリランキングモデルを用いたリランキング装置、及び、Ｎベストリランキングモデルを実現する学習装置について説明する。なお、本実施の形態については、Ｎベストリスコアリング（モデル）ではなく、Ｎベストリランキング（モデル）と表現を統一して説明する。

【００１９】

まず、本実施の形態に係るリランキング装置がＮベスト仮説のリランキングを行う上で、Ｎベストリランキングモデルが有すべき必要最低限な機能について述べる。従来の方法では、Ｎベスト仮説は、リスコアリングの結果、スコアが降順になるようにソートされる。しかしながら、Ｎベストリスコアリングの主な目的は、Ｎベスト仮説から最も音声認識精度が高い仮説（オラクル仮説）を、最終的な音声認識結果として見つけ出すことである。このため、リスコアリング後のＮベスト仮説は、必ずしもソートされている必要はない。本実施の形態ではこの点に着目した。

【００２０】

すなわち、本実施の形態では、Ｎベスト仮説の中からオラクル仮説をリランキングにより見つけ出すためにＮベストリランキングモデルに必要最低限な機能は、Ｎベスト仮説中の二つの仮説に着目したときに、どちらの仮説の方がより高い音声認識精度を有しているかを判定できることである点に着目した。言い換えると、Ｎベストリランキングモデルに必要最低限な機能は、Ｎベスト仮説中の二つの仮説を対象に、一対一の仮説比較を行うことができることである。

【００２１】

そこで、本実施の形態に係るリランキング装置は、ＮＮで表され、一対一の二つの仮説の比較を行う機能を持つＮベストリランキングモデルを用いることによって、二つの仮説のうち音声認識精度がより高い仮説を判定する機能を持たせた。そして、本実施の形態に係るリランキング装置は、音声認識精度がより高い仮説を次の判定対象の一方の仮説として残し、未判定の仮説から他方の仮説を選択して、Ｎベストリランキングモデルを用いた比較を行う。本実施の形態に係るリランキング装置は、前回の判定で音声認識精度がより高いと判定された仮説を判定対象の一方の仮説として選択し、未判定の仮説のいずれかを他方の仮説として選択し、Ｎベストリランキングモデルによる二つの仮説に対する比較処理を繰り返す。これによって、本実施の形態では、Ｎベスト仮説の中からオラクル仮説を見つけ出すことを可能にした。

【００２２】

［実施の形態］

［リランキング装置］

次に、実施の形態に係るリランキング装置について説明する。このリランキング装置は、音声認識結果であるＮベスト仮説のうちの二つの仮説に対する、ＮＮで表されるＮベストリランキングモデルを用いた音声認識精度の高低の判定を繰り返し実行して、最も音声認識精度の高い仮説を最終的な音声認識結果として出力する。

【００２３】

図１は、実施の形態に係るリランキング装置の機能構成の一例を示す図である。実施の形態１に係るリランキング装置１０は、例えば、ＲＯＭ（Read Only Memory）、ＲＡＭ（Random Access Memory）、ＣＰＵ（Central Processing Unit）等を含むコンピュータ等に所定のプログラムが読み込まれて、ＣＰＵが所定のプログラムを実行することで実現される。

【００２４】

リランキング装置１０は、音声認識装置２から出力されたＮベスト仮説の入力を受け付ける。そして、リランキング装置１０は、このＮベスト仮説のうち、二つの仮説に対する音声認識精度の高低についての判定を、全Ｎベスト仮説について実行し、音声認識精度が高い仮説として残った仮説を、最終的な音声認識結果として出力する。なお、音声認識装置２は、１発話が入力されると、例えば、音声認識用のモデルを用いて音声認識を行い、音声認識結果としてＮベスト仮説を出力する。音声認識用のモデルは、学習用の複数の発

話と、各発話に対応する書き起こし（正解単語列）を学習データとして用いて学習（モデルパラメータが最適化）されている。

【００２５】

リランキング装置１０は、Ｎベストリランキングモデル記憶部１１、仮説入力部１２、仮説選択部１３、特徴量抽出部１４、判定部１５、実行制御部１６及び出力部１７を有する。

【００２６】

Ｎベストリランキングモデル記憶部１１は、Ｎベストリランキングモデルを記憶する。Ｎベストリランキングモデルは、ＮＮで表されるモデルである。Ｎベストリランキングモデルは、音声認識精度が既知である学習用のＮベスト仮説を用いて予め学習される。Ｎベストリランキングモデルは、学習用のＮベスト仮説のうち二つの系列の複数の組み合わせについて、二つの系列の特徴量が与えられたときに、その二つの系列の音声認識精度の高低を判定できるように学習される。Ｎベストリランキングモデルは、二つの仮説を、ＲＮＮを用いて隠れ状態ベクトルに変換する。そして、Ｎベストリランキングモデルは、ＮＮを用いて、隠れ状態ベクトルを基に二つの仮説の精度の高低の並びが正しいことを示す第１の事後確率及び二つの仮説の精度の高低の並びが誤りであることを示す第２の事後確率を出力する。言い換えると、Ｎベストリランキングモデルでは、ＲＮＮの後段に２クラス分類ＦＦＮＮが接続される。この２クラス分類ＦＦＮＮは、ＲＮＮが変換した隠れ状態ベクトルを基に、二つの仮説のＮベスト仮説における順位の上下関係が正しいことを示す第１の事後確率及び二つの仮説のＮベスト仮説における順位の上下関係が誤りであることを示す第２の事後確率を出力する。

【００２７】

仮説入力部１２は、Ｎベスト仮説の入力を受け付ける。Ｎベスト仮説は、音声認識装置２が出力する。或いは、他の装置が、ネットワーク等を介して、Ｎベスト仮説をリランキング装置１０に入力してもよい。

【００２８】

仮説選択部１３は、入力を受け付けたＮベスト仮説のうち、一対一の比較対象である二つの仮説を選択する。仮説選択部１３は、一定のルールに従い、Ｎベスト仮説の中から、任意の二つの仮説を１組とし選択する。具体的には、仮説選択部１３は、二つの仮説の一方の仮説として、比較対象時に最高の精度を持つと推定される仮説を選択する。仮説選択部１３は、二つの仮説の他方の仮説として、前回比較対象となった仮説の順位の次の順位の仮説を選択する。このように、仮説選択部１３は、全Ｎベスト仮説について一対一の比較が実行されるように、Ｎベスト仮説から、比較対象の二つの仮説を選択する。

【００２９】

特徴量抽出部１４は、一対一の比較対象である二つの仮説について、それぞれの特徴量を抽出する。特徴量抽出部１４は、一対一の比較対象であるＮベスト仮説中のｕ位の仮説（単語列）と、Ｎベスト仮説中のｖ位（ｕ＜ｖ≦Ｎ）の仮説とについて、それぞれの特徴量を抽出する。特徴量抽出部１４は、仮説中の各単語単位で特徴量ベクトルを抽出する。各単語の特徴量ベクトルは、例えば、離散値である単語ＩＤをＮＮによる単語の埋め込み処理により連続値のベクトルとして表現した単語ベクトルに、音声認識処理により得られる単語単位の音響スコア（対数尤度）や言語スコア（対数確率）などを補助特徴量として、単語ベクトルに連結したものである。

【００３０】

判定部１５は、一対一の比較対象の二つの仮説に対し、Ｎベストリランキングモデルを用いて、いずれの仮説がより高い音声認識精度を有しているかを判定する。具体的には、一対一の比較対象であるｕ位の仮説と、ｖ（ｕ＜ｖ≦Ｎ）位の仮説との特徴量をＮベストリランキングモデルに入力し、Ｎベストリランキングモデルによる出力結果を用いて、どちらの仮説が高い音声認識精度を有しているかを判定する。ｕ位及びｖ位で表す仮説の順位は、Ｎベスト仮説において既に付与されているものである。リランキング装置１０では、順位の再設定を行わない。

【００３１】
ここで、Ｎベストリランキングモデルは、ｕ位の仮説の特徴量及びｖ位の仮説の特徴量が入力されると、ｕ位の仮説がｖ位の仮説よりも音声認識精度が高いことを示す第１の事後確率と、ｖ位の仮説がｕ位の仮説よりも音声認識精度が高いことを示す第２の事後確率とを出力する。判定部１５は、第１の事後確率が第２の事後確率よりも高い場合には、ｕ位の仮説がｖ位の仮説よりも音声認識精度が高いと判定する。また、判定部１５は、第１の事後確率が第２の事後確率よりも低い場合には、ｖ位の仮説よりもｕ位の仮説よりも音声認識精度が高いと判定する。
【００３２】
なお、リランキング装置１０では、特徴量抽出部１４の機能を、Ｎベストリランキングモデルが有してもよい。この場合、判定部１５は、比較対象である二つの仮説をＮベストリランキングモデルに入力する。
【００３３】
そして、判定部１５は、比較対象の二つの系列のうち、より精度が高いと判定した仮説を次の判定時における比較対象として残し、他方の仮説を以降比較対象から外す。仮説選択部１３は、判定部１５によって精度が高いと判定された仮説を二つの系列の一方の仮説として選択し、Ｎベスト仮説のうち、判定部１５による判定が行われていない仮説のいずれかを他方の仮説として選択する。具体的には、前述したように、仮説選択部１３は、判定部１５が残した仮説を二つの仮説の一方の仮説として選択し、Ｎベスト仮説のうち、前回比較対象となった仮説の順位の次の順位の仮説を二つの仮説の他方の仮説として選択する。
【００３４】
実行制御部１６は、判定部１５による判定処理と仮説選択部１４による選択処理とを、所定条件に達するまで繰り返す制御を行う。この場合、実行制御部１６は、全Ｎベスト仮説について一対一の比較が実行されるように、仮説選択部１３における比較対象の二つの仮説の選択処理、特徴量抽出部１４における特徴量抽出処理、及び、判定部１５における判定処理を繰り返す制御を行う。具体的に、実行制御部１６は、比較対象である仮説の順位がＮになるまで、仮説の選択処理、特徴量抽出処理及び判定処理を繰り返す制御を行う。
【００３５】
出力部１７は、仮説の選択処理、特徴量抽出処理、判定処理及び順位の設定処理が繰り返された結果、Ｎベスト仮説のうち、所定条件に達した場合、比較対象として残っている仮説を、最も音声認識精度が高い仮説、すなわち、最終的な音声認識結果として出力する。出力部１７は、最後の判定処理で精度が高いと判定された仮説を最終的な音声認識結果として出力する。
【００３６】
次に、Ｎベストリランキングモデルに必要最低限な機能要件を数式で定義する。$W^{(u)}=w_1^{(u)}, w_2^{(u)}, \cdots, w_{L(W^{(u)})}^{(u)}$を、Ｎベスト仮説中のｕ位の仮説（単語列）と定義する。また、$L(W^{(u)})$を、$W^{(u)}$の長さ（単語数）と定義する。
【００３７】
また、$A^{(u)}=a_1^{(u)}, a_2^{(u)}, \cdots, a_{L(W^{(u)})}^{(u)}$を$W^{(u)}$に対応する補助特徴量ベクトル列と定義する。$W^{(u)}$中のｉ番目の単語$w_i^{(u)}$の補助特徴量ベクトル$a_i^{(u)}$は、例えば、音声認識装置による音声認識処理の結果として得られる音響スコア（対数尤度）や言語スコア（対数確率）などである（詳細は、例えば、非特許文献２を参照）。
【００３８】
また、$X^{(u)}=x_1^{(u)}, x_2^{(u)}, \cdots, x_{L(W^{(u)})}^{(u)}$を$W^{(u)}$に対応する特徴量ベクトル列と定義する。$W^{(u)}$中のｉ番目の単語$w_i^{(u)}$の特徴量ベクトル$x_i^{(u)}$は、$x_i^{(u)}=\mathrm{concat}(\mathrm{embed}(w_i^{(u)}), a_i$

（ｕ）で得られる。ここで、ｃｏｎｃａｔ（・）は、ベクトルの連結処理を表す。また、ｅｍｂｅｄ（・）は、ＮＮによる単語の埋め込み処理（離散値の単語ＩＤを連続値のベクトルで表現する処理）（詳細は、例えば、坪井祐太，海野裕也，鈴木潤，深層学習による自然言語処理，MLP機械学習プロフェッショナルシリーズ，講談社，2017．（以降、参考文献１とする。）を参照）を表す。なお、ｅｍｂｅｄ（・）を行うＮＮもＮベストリスコアリングモデルの一部であり、そのパラメータは、後述のエンコーダＲＮＮ及び２クラス分類ＦＦＮＮのパラメータと同時に学習（最適化）される。

【００３９】

Ｎベスト仮説中のｕ番目の仮説Ｗ（ｕ）とｖ番目の仮説Ｗ（ｖ）（ｕ＜ｖ≦Ｎ）の特徴量ベクトル列Ｘ（ｕ），Ｘ（ｖ）が与えられたとき、リランキング装置１０におけるＮベストリランキングモデルは、２クラスの記号ｙ＝｛０，１｝の事後確率Ｐを出力する。ｙ＝０は、Ｗ（ｕ）及び仮説Ｗ（ｖ）の順位の上下関係が正しいことを示す。また、ｙ＝１は、Ｗ（ｕ）及び仮説Ｗ（ｖ）の順位の上下関係が誤りであることを示す。Ｐ（０｜Ｘ（ｕ），Ｘ（ｖ））は、ｕ位の仮説とｖ位の仮説との順位の上下関係が正しさを確率的に表現する第１の事後確率である。Ｐ（１｜Ｘ（ｕ），Ｘ（ｖ））は、ｕ位の仮説とｖ位の仮説との順位の上下関係が誤りであることを確率的に表現する第２の事後確率である。

【００４０】

判定部１５は、Ｎベストリランキングモデルから出力された第１の事後確率Ｐ（０｜Ｘ（ｕ），Ｘ（ｖ））及び第２の事後確率Ｐ（１｜Ｘ（ｕ），Ｘ（ｖ））を取得し、取得した二つの事後確率の大小を比較して、ｕ位の仮説及びｖ位の仮説のいずれがより音声認識精度が高いかを判定する。判定部１５は、第１の事後確率Ｐ（０｜Ｘ（ｕ），Ｘ（ｖ））が第２の事後確率Ｐ（１｜Ｘ（ｕ），Ｘ（ｖ））よりも高い場合には、ｕ位の仮説がｖ位の仮説よりも音声認識精度が高いと判定する。また、判定部１５は、第１の事後確率Ｐ（０｜Ｘ（ｕ），Ｘ（ｖ））が第２の事後確率Ｐ（１｜Ｘ（ｕ），Ｘ（ｖ））よりも低い場合には、ｖ位の仮説がｕ位の仮説よりも音声認識精度が高いと判定する。

【００４１】

すなわち、判定部１５は、以下の（１－１）式及び（１－２）式に示すように、ｕ位の仮説及びｖ位の仮説のいずれがより音声認識精度が高いかを判定する。

【００４２】

Ｐ（０｜Ｘ（ｕ），Ｘ（ｖ））≧Ｐ（１｜Ｘ（ｕ），Ｘ（ｖ））
ｉｆ　ａｃｃ（Ｗ（ｕ））≧ａｃｃ（Ｗ（ｖ））　　・・・（１－１）
Ｐ（０｜Ｘ（ｕ），Ｘ（ｖ））＜Ｐ（１｜Ｘ（ｕ），Ｘ（ｖ））
ｏｔｈｅｒｗｉｓｅ　　・・・（１－２）

【００４３】

ここで、ａｃｃ（・）は、与えられた仮説（単語列）の音声認識精度を返す関数Σ_yＰ（ｙ｜Ｘ（ｕ），Ｘ（ｖ））＝１である。（１－１）式の１段目に示す不等式が満足される場合、判定部１５は、仮説Ｗ（ｕ）は仮説Ｗ（ｖ）以上の音声認識精度を持つと判定する。また、（１－２）式の不等式が満足される場合、判定部１５は、Ｗ（ｕ）はＷ（ｖ）よりも低い音声認識精度を持つと判定する。

【００４４】

したがって、（１－１）式の１段目に示す不等式が満足される場合、Ｗ（ｕ）及びＷ（ｖ）のランキングの上下関係（ｕ＜ｖ）が正しいと推定される。このため、判定部１５は、Ｗ（ｕ）を、Ｗ（ｖ）との一対一の仮説比較においてＷ（ｖ）よりも音声認識精度が高い仮説として残し、次の一対一の仮説比較でもＷ（ｕ）として引き続き使用する。なお、判定部１５は、Ｗ（ｖ）を、Ｗ（ｕ）よりも音声認識精度が低い仮説として扱い、最も音声認識精度が高い仮説の候補、すなわち、最終的な音声認識結果の候補から除外する。

【００４５】

そして、（１－２）式の１段目不等式が満足される場合は、Ｗ（ｕ）及びＷ（ｖ）のランキングの上下関係は、誤りであると推定される。すなわち、Ｗ（ｕ）及びＷ（ｖ）のランキングの上下関係は逆であると推定される。このため、判定部１５は、Ｗ（ｖ）を、Ｗ

$^{(u)}$との一対一の仮説比較においてW$^{(u)}$よりも音声認識精度が高い仮説として残し、次の一対一の仮説比較ではW$^{(u)}$として使用する。なお、判定部１５は、元のW$^{(u)}$を、元のW$^{(v)}$よりも音声認識精度が低い仮説として扱い、最も音声認識精度が高い仮説の候補、すなわち、最終的な音声認識結果の候補から除外する。なお、Ｎベストリランキングモデルは、第１の事後確率Ｐ（０｜Ｘ$^{(u)}$，Ｘ$^{(v)}$）及び第２の事後確率Ｐ（１｜Ｘ$^{(u)}$，Ｘ$^{(v)}$）の事後確率の大小を比較して、ｕ位の仮説及びｖ位の仮説のいずれがより音声認識精度が高いかを判定し、仮説の残存の判定までを推定してもよい。

【００４６】

［Ｎベストリランキングモデルの構築例］

図２は、Ｎベストリランキングモデルの構築例を示す図である。なお、図２では、簡単のため、単語の埋め込み処理ｅｍｂｅｄ（・）を行うＮＮは省略されている。以下、その詳細について説明する。

【００４７】

比較対象の仮説W$^{(u)}$の長さ（単語数）Ｌ（W$^{(u)}$）と仮説W$^{(v)}$（ｕ＜ｖ≦Ｎ）の長さＬ（W$^{(v)}$）とが異なる可能性がある。この長さの違いを吸収するため、Ｎベストランキングモデルは、二つの仮説を、ＲＮＮを用いて隠れ状態ベクトルに変換する。具体的には、Ｎベストリランキングモデルは、この処理を行うために、エンコーダ－デコーダモデル（詳細は、例えば、参考文献１参照）のエンコーダＲＮＮ１１１を有する。

【００４８】

Ｎベストリランキングモデルは、エンコーダＲＮＮ１１１を用いてW$^{(u)}$とW$^{(v)}$を固定長の隠れ状態ベクトルで表現することができる。そして、Ｎベストリランキングモデルは、これらの隠れ状態ベクトルを用いることによって、W$^{(u)}$とW$^{(v)}$とを公平に比較することが可能になる。

【００４９】

エンコーダＲＮＮ１１１の処理について説明する。エンコーダＲＮＮ１１１は、ＲＮＮの一種である長短期記憶メモリ（ｌｏｎｇ　ｓｈｏｒｔ－ｔｅｒｍ　ｍｅｍｏｒｙ：ＬＳＴＭ）ユニット（詳細は、例えば、参考文献１参照）を有する。ＬＳＴＭユニットは、W$^{(u)}$のｉ番目の単語$w_i^{(u)}$の特徴量ベクトル$x_i^{(u)}$と、ｉ－１番目の隠れ状態ベクトル$h_{\{i-1\}}^{(u)}$が与えられたとき、ｉ番目の隠れ状態ベクトル$h_i^{(u)}$を以下の（２）式のように与える。

【００５０】

$$h_i^{(u)} = lstm(x_i^{(u)},\ h_{\{i-1\}}^{(u)}) \quad \cdots (2)$$

【００５１】

ここで、ｌｓｔｍ（・）は、１層単方向（ｕｎｉｄｉｒｅｃｔｉｏｎａｌ）のＬＳＴＭユニットの処理を示す。また、$h_i^{(u)}=0$(ゼロベクトル)である。$h_i^{(u)}$は、単語列$w_1^{(u)}$，$w_2^{(u)}$，・・・，$w_i^{(u)}$の特徴量ベクトル列$x_1^{(u)}$，$x_2^{(u)}$，・・・，$x_i^{(u)}$をエンコード（符号化）したものである。エンコーダＲＮＮ１１１は、この処理を、特徴量ベクトル列Ｘ$^{(u)}$中の各特徴量ベクトル$x_i^{(u)}$に対して繰り返すことで、Ｘ$^{(u)}$をエンコードした隠れ状態ベクトル$h_{L(W(u))}^{(u)}$を得ることができる。

【００５２】

エンコーダＲＮＮ１１１は、同様の処理を特徴量ベクトル列Ｘ$^{(v)}$に対しても行い、Ｘ$^{(v)}$をエンコードした隠れ状態ベクトル$h_{L(W(v))}^{(v)}$を得る。なお、Ｘ$^{(u)}$に対して処理を行うＬＳＴＭユニットと、Ｘ$^{(v)}$に対して処理を行うＬＳＴＭユニットは同じもの、すなわち、パラメータが共有されていてもよいし、別のＬＳＴＭユニットであってもよい。また、図２では、$x_{L(W(u))}^{(u)}$，$x_{L(W(v))}^{(v)}$，$h_{L(W(u))}^{(u)}$，$h_{L(W(v))}^{(v)}$の下付き部分Ｌ（Ｗ（ｕ））は、Ｌ（W$^{(u)}$）と示している。

【００５３】

Ｎベストリランキングモデルは、以上で得た二つの隠れ状態ベクトル$h_{L(W(u))}$

$h_{L(W(u))}^{(u)}$, $h_{L(W(v))}^{(v)}$を連結した隠れ状態ベクトル$h^{\{(u,v)\}}$をエンコーダＲＮＮ１１の出力として以下の（３）式のように得る。

【００５４】

$$h^{\{(u,v)\}}=concat(h_{L(W(u))}^{(u)},h_{L(W(v))}^{(v)}) \quad \cdots(3)$$

【００５５】

そして、Ｎベストリランキングモデルは、エンコーダＲＮＮ１１１の後段に、クラス分類（ｙ＝０　ｏｒ　１）を行うためのＮＮを連結する。例えば、Ｎベストリランキングモデルは、２クラス分類のためのＮＮとして、１層のフィードフォワード型ＮＮ（ＦＦＮＮ）１１２（詳細は、例えば、参考文献１を参照）を用いる。エンコーダＲＮＮ１１１の出力として得た隠れ状態ベクトル$h^{\{(u,v)\}}$が、１層の２クラス分類ＦＦＮＮに入力され、最終的に、２クラスの記号ｙ＝｛０，１｝の事後確率Ｐ（ｙ｜$X^{(u)}$，$X^{(v)}$）を以下の（４），（５）式のように得ることができる。

【００５６】

$$z^{\{(u,v)\}}=linear(h^{\{(u,v)\}}) \quad \cdots(4)$$

$$P(y|X^{(u)},X^{(v)})=softmax(z^{\{(u,v)\}})_y \quad \cdots(5)$$

【００５７】

ここで、ｌｉｎｅａｒ（・）は、線形変換処理（詳細は、例えば、参考文献１を参照）を表す。ｓｏｆｔｍａｘ（・）は、ソフトマックス処理を表す。また、$softmax(\cdot)_y$は、ソフトマックス処理の結果として得られる事後確率ベクトルのｙ番目の要素（確率値）を表す。

【００５８】

［Ｎベストリランキングモデルの他の構築例１］

なお、図２に示すエンコーダＲＮＮ１１１のＬＳＴＭユニットは、１層単方向のＬＳＴＭユニットとしたが、複数層または双方向（ｂｉｄｉｒｅｃｔｉｏｎａｌ）のＬＳＴＭユニットであってもよい。

【００５９】

［Ｎベストリランキングモデルの他の構築例２］

また、ＬＳＴＭユニットの代わりに、単純な（下記のｓｉｇｍｏｉｄ関数等を活性化関数として持つ。）ＲＮＮや、Ｇａｔｅｄ　Ｒｅｃｕｒｒｅｎｔ　Ｕｎｉｔ（ＧＲＵ）を用いてもよい。

【００６０】

［Ｎベストリランキングモデルの他の構築例３］

さらに、Ｎベストリランキングモデルは、図２の構築例では、２クラス分類ＮＮとして、１層のフィードフォワード型ＮＮを用いたが、複数層のフィードフォワード型ＮＮを用いてもよい。Ｎベストリランキングモデルは、複数層のフィードフォワード型ＮＮを用いる場合、活性化関数として、ｓｉｇｍｏｉｄ関数、ｔａｎｈ関数、Ｒｅｃｔｉｆｉｅｄ　Ｌｉｎｅａｒ　Ｕｎｉｔ（ＲｅＬＵ）関数、Ｐａｒａｍｅｔｒｉｃ　ＲｅＬＵ（ＰＲｅＬＵ）関数などを用いることができる。なお、Ｎベストリランキングモデルの他の構築例１～３の用語の詳細については、例えば、参考文献１を参照いただきたい。

【００６１】

［Ｎベストリランキングモデルの他の構築例４］

また、Ｎベストリランキングモデルは、従来のＮベストリスコアリングモデル（例えばＲＮＮ言語モデル）により計算されたスコアを、特徴量ベクトルにおける新たな次元として追加して利用することも可能である。

【００６２】

［リランキング処理の処理手順］

次に、図１に示すリランキング装置１０が実行するリランキング処理の処理手順について説明する。図３は、図１に示すリランキング装置１０が実行するリランキング処理の処

理手順を示すフローチャートである。

【００６３】

まず、仮説入力部１２が、リランキング対象のＮベスト仮説の入力を受け付けると（ステップＳ１）、仮説選択部１３は、入力を受け付けたＮベスト仮説のうち、順次、一対一の比較対象であるｕ位及びｖ位の二つの仮説を選択する（ｕ＜ｖ≦Ｎ）。まず、仮説選択部１３は、ｕ＝１、ｖ＝２に設定する（ステップＳ２）。そして、仮説選択部１３は、入力を受け付けたＮベスト仮説から、ｕ位及びｖ位の二つの仮説$W^{(u)}$，$W^{(v)}$をＮベスト仮説から選択する（ステップＳ３）。続いて、特徴量抽出部１４は、仮説$W^{(u)}$，$W^{(v)}$の特徴量を抽出する（ステップＳ４）。判定部１５は、仮説$W^{(u)}$，$W^{(v)}$の特徴量（$X^{(u)}$，$X^{(v)}$）をＮベストリランキングモデルに入力する（ステップＳ５）。

【００６４】

判定部１５は、Ｎベストリランキングモデルからの出力結果を取得する（ステップＳ６）。具体的には、判定部１５は、第１の事後確率Ｐ（０｜$X^{(u)}$，$X^{(v)}$）及び第２の事後確率Ｐ（１｜$X^{(u)}$，$X^{(v)}$）を取得する。

【００６５】

そして、（１－１）式及び（１－２）式において説明したように、判定部１５は、ａｃｃ（$W^{(u)}$）≧ａｃｃ（$W^{(v)}$）であるか否かを判定する（ステップＳ７）。判定部１５は、Ｐ（０｜$X^{(u)}$，$X^{(v)}$）≧Ｐ（１｜$X^{(u)}$，$X^{(v)}$）の場合、ａｃｃ（$W^{(u)}$）≧ａｃｃ（$W^{(v)}$）であると判定する。一方、判定部１５は、Ｐ（０｜$X^{(u)}$，$X^{(v)}$）＜Ｐ（１｜$X^{(u)}$，$X^{(v)}$）の場合、ａｃｃ（$W^{(u)}$）≧ａｃｃ（$W^{(v)}$）でないと判定する。

【００６６】

判定部１５がａｃｃ（$W^{(u)}$）≧ａｃｃ（$W^{(v)}$）であると判定した場合（ステップＳ７：Ｙｅｓ）、順位設定部１６は、ｋについてｋ＝ｕと設定する（ステップＳ８）。ｋは、最も音声認識精度が高い仮説のＮベスト仮説における順位（ランキング）である。一方、判定部１５がａｃｃ（$W^{(u)}$）≧ａｃｃ（$W^{(v)}$）でないと判定した場合（ステップＳ７：Ｎｏ）、順位設定部１６は、ｋ＝ｖと設定する（ステップＳ９）。

【００６７】

続いて、実行制御部１６は、ｖ＝Ｎであるか否かを判定する（ステップＳ１０）。実行制御部１６は、ｖ＝Ｎでないと判定した場合（ステップＳ１０：Ｎｏ）、必要な一対一の仮説比較処理がまだ全ては終了していないため、仮説選択部１３に対し、比較対象の次の仮説の選択を行わせる。具体的には、仮説選択部１３は、ｕ＝ｋ、ｖ＝ｖ＋１に設定し（ステップＳ１１）、ステップＳ３に戻り、次の判定対象のＮベスト仮説$W^{(u)}$，$W^{(v)}$を選択する。そして、リランキング装置１０は、このＮベスト仮説$W^{(u)}$，$W^{(v)}$に対して、ステップＳ４～ステップＳ１０の処理を実行する。

【００６８】

また、実行制御部１６は、ｖ＝Ｎであると判定した場合（ステップＳ１０：Ｙｅｓ）、必要な一対一の比較処理が全て終了したため、ｋ位の$W^{(k)}$を最も音声認識精度が高いと推定される仮説、すなわち、最終的な音声認識結果として出力し（ステップＳ１２）、処理を終了する。このように、リランキング装置１０では、任意の二つの仮説を１組とし、複数の組についてそれぞれ音声認識精度の高低の判定を繰り返すことで、最も音声認識精度が高いと推定される仮説を、最終的な音声認識結果として出力することができる。

【００６９】

［学習装置］

次に、リランキング装置１０が用いるＮベストリランキングモデルを学習する学習装置について説明する。図４は、実施の形態に係る学習装置の機能構成の一例を示す図である。実施の形態１に係る学習装置２０は、例えば、ＲＯＭ、ＲＡＭ、ＣＰＵ等を含むコンピュータ等に所定のプログラムが読み込まれて、ＣＰＵが所定のプログラムを実行することで実現される。図４に示すように、Ｎベストリランキングモデル記憶部２１、学習装置２

０は、仮説入力部２２及び学習部２３を有する。

【００７０】

Ｎベストリランキングモデル記憶部２１は、学習対象のＮベストリランキングモデルを記憶する。Ｎベストリランキングモデルは、ＮＮで表される。Ｎベストリランキングモデルは、Ｎベスト仮説のうちの二つの仮説を、ＲＮＮを用いて隠れ状態ベクトルに変換する。そして、Ｎベストリランキングモデルは、ＮＮを用いて、隠れ状態ベクトルを基に二つの仮説の精度の高低の並びが正しいことを示す第１の事後確率及び二つの仮説の精度の高低の並びが誤りであることを示す第２の事後確率を出力する。

【００７１】

仮説入力部２２は、音声認識精度が既知である学習用のＮベスト仮説の入力を受け付ける。学習用のＮベスト仮説として、学習データ中の各発話に対して音声認識が行われ、各発話のＮベスト仮説が得られているものとする。また学習データであるので、全ての仮説の音声認識精度は、既知である。また、Ｎベスト仮説中の全ての仮説に対して、前述のように、特徴量ベクトル列が抽出されているものとする。

【００７２】

学習部２３は、学習用のＮベスト仮説のうちの二つの仮説の特徴量がそれぞれ与えられたときに、それら二つの仮説の精度の高低が判定できるような、Ｎベストリランキングモデルを学習する。学習部２３では、学習用のＮベスト仮説のうちの二つの仮説の特徴量ベクトル列と、これらに対応する教師ラベル（後述）とを、Ｎベストリランキングモデルに与える。これによって、学習部２３は、Ｎベストリランキングモデルがこれら二つの仮説の音声認識精度の高低を正しく判定できるように、Ｎベストリランキングモデルの学習（パラメータの最適化）を行う。具体的には、学習部２３は、特徴量ベクトル列と、対応する教師ラベルとをＮベストリランキングモデルに入力し、Ｎベストリランキングモデルがこれらの特徴量ベクトルを与えられたときに対応する教師ラベルを正しく出力できるように、Ｎベストリランキングモデルの学習を行う。学習部２３は、教師ラベル付与部２３１及び入替部２３２を有する。

【００７３】

教師ラベル付与部２３１は、二つの仮説のうち音声認識精度がより高い仮説に他方の仮説よりも高い順位が付与されている場合に正解を表す教師ラベル（ｙ＝０）を付与して、Ｎベストリランキングモデルに学習させる。また、教師ラベル付与部２３１は、二つの仮説のうち音声認識精度がより高い仮説に他方の仮説よりも低い順位が付与されている場合に誤りを表す教師ラベル（ｙ＝１）を付与し、Ｎベストリランキングモデルに学習させる。

【００７４】

入替部２３２は、学習用のＮベスト仮説のうちの二つの仮説の順位を入れ換え、対応する教師ラベルも入れ換えて、Ｎベストリランキングモデルの学習を行う。例えば、教師ラベルとしてｙ＝０が付与されている二つの仮説については、二つの仮説の順位を入れ換え、教師ラベルｙを１に変える。一方、教師ラベルとしてｙ＝１が付与されている二つの仮説については、二つの仮説の順位を入れ換え、教師ラベルｙを０に変える。

【００７５】

［学習処理の処理手順］

次に、図４に示す学習装置２０が実行する学習処理の処理手順について説明する。図５は、図４に示す学習装置が実行する学習処理の処理手順を示すフローチャートである。図５では、Ｎベスト仮説から二つの仮説としてＷ（ｕ），Ｗ（ｖ）（ｕ＜ｖ≦Ｎ）が与えられ、かつ、ａｃｃ(Ｗ（ｕ）)≧ａｃｃ(Ｗ（ｖ）)であるときの学習処理の処理手順を示す。

【００７６】

図５に示すように、教師ラベル付与部２３１が、教師ラベルｙ＝０を付与し（ステップＳ２１）、Ｗ（ｕ），Ｗ（ｖ）の特徴量Ｘ（ｕ），Ｘ（ｖ）をＮベストリランキングモデルに入力し（ステップＳ２２）、Ｎベストリランキングモデルを学習させ、Ｎベストリラ

ンキングモデルのモデルパラメータを更新させる（ステップＳ２３）。すなわち、（１－１）式に従うと、この二つの仮説のW（ｕ），W（ｖ）の特徴量ベクトルX（ｕ），X（ｖ）をＮベストリランキングモデルに入力した場合、モデルは、理想的には、Ｐ（０｜X（ｕ），X（ｖ））＝１との事後確率を出力すべきである。このため、教師ラベル付与部２３１は、教師ラベルとして、ｙ＝０を与える。以上の入力を基に、学習部２３は、モデルパラメータ（エンコーダＲＮＮ（ＬＳＴＭユニット）及び２クラス分類ＦＦＮＮ及び単語の埋め込み処理ｅｍｂｅｄ（・）を行うＮＮのパラメータを同時に）を更新させる。

【００７７】

そして、入替部２３２は、仮説W（ｕ），W（ｖ）の順位を入れ替える（ステップＳ２４）。すなわち、元々、W（ｖ）であった仮説をW（ｕ）とし、元々、W（ｕ）であった仮説をW（ｖ）とする。この場合には、ａｃｃ（W（ｕ））≧ａｃｃ（W（ｖ））ではない。よって、（１－２）式に従えば、この二つの仮説W（ｕ），W（ｖ）の特徴量ベクトルX（ｕ），X（ｖ）をＮベストリランキングモデルに入力した場合、モデルは理想的には，Ｐ（１｜X（ｕ），X（ｖ））＝１との事後確率を出力すべきである。このため、教師ラベル付与部２３１は、教師ラベルとして、ｙ＝１を付与し（ステップＳ２５）、W（ｕ），W（ｖ）の特徴量X（ｕ），X（ｖ）をＮベストリランキングモデルに入力する（ステップＳ２６）。学習部２３は、以上の入力を基に、Ｎベストリランキングモデルを学習させ、Ｎベストリランキングモデルのモデルパラメータを更新させて（ステップＳ２７）、二つの仮説W（ｕ），W（ｖ）に対する学習処理を終了する。

【００７８】

学習装置２０は、上記の手順を、学習データ中の各発話のＮベスト仮説について繰り返し、更にはその繰り返し自体を何度か（何エポックか）繰り返す。学習部２３は、学習の更なる具体的な手順については、従来のＮＮの学習（詳細は、例えば、参考文献１参照）と同様に行うことができる。

【００７９】

［学習処理の効率化例１］

図５に示す学習処理の処理手順は、計算コストが高い。例えば、Ｅをエポック数、Ｍを学習データ中の発話数とすると、上記の学習手順におけるモデルパラメータの更新回数は、最大で、Ｅ×Ｍ×Ｎ×２×${}_{N}C_{2}$になる。通常、Ｅは数十程度、Ｍは少なくとも数万、Ｎは上記の通り１００～１０００程度であるので、モデルパラメータの更新回数は、膨大な数に達する。このため、本実施の形態では、学習の効率化を図ることが好ましい。そこで、以下に、学習の効率化例１について述べる。

【００８０】

上述したように、Ｎベストリスコアリングの主な目的は、Ｎベスト仮説からオラクル仮説を最終的な音声認識結果として見つけ出すことである。言い換えれば、オラクル仮説をその他のＮ－１個の仮説から精度よく区別できればよい。これを実現するために、学習の際に、Ｎベストリランキングモデルに入力する二つの仮説のうちの一方をオラクル仮説とする。これにより、モデルパラメータの更新回数を、Ｅ×Ｍ×Ｎ×２×（Ｎ－１）に削減することができる。

【００８１】

［学習処理の効率化例２］

次に、学習の効率化例２について説明する。学習の効率化例１では、Ｎベスト仮説が与えられたとき、その中に含まれるオラクル仮説とその他のＮ－１個の仮説とを比較していた。学習処理の効率化例２では、オラクル仮説と比較するその他の仮説の個数を絞り込む。

【００８２】

例えば、まず、下の典型的な四つの仮説を選択する。

仮説１は、オラクル仮説の次に高い音声認識精度を持つ仮説である。

仮説２は、音声認識スコアが最も高い仮説である。

仮説３は、最も低い音声認識精度を持つ仮説である。

仮説４は、音声認識スコアが最も低い仮説である。

【００８３】

仮説１と仮説２とは、音声認識精度が高い（または高いと推定される）仮説で、オラクル仮説との区別が難しい仮説である。一方、仮説３と仮説４とは、音声認識精度が低い（または低いと推定される）仮説で、オラクル仮説との区別が容易な（確実に区別しないといけない）仮説である。その他の仮説をこの四つのみに絞り込む場合は、モデルパラメータの更新回数は、Ｅ×Ｍ×Ｎ×２×４にまで削減することができる。

【００８４】

ただし、上記の四つの仮説のみではオラクル仮説の対立仮説としての多様性が十分に確保できないと考えられる場合、Ｎベスト仮説から、オラクル仮説とこれらの四つの仮説を除いた、残りのＮ－５個の仮説から、所定のルールにしたがって抽出した所定数の仮説を選択して前記四つの仮説と共に対立仮説として用いてもよい。例えば、二つの仮説のうちの他方の仮説として、オラクル仮説とこれらの四つの仮説を除いた、残りのＮ－５個の仮説から、等間隔に、或いは、はランダムに、Ｑ個の仮説を選択して四つの仮説と共に他方の仮説として用いる。このとき、モデルパラメータの更新回数は、Ｅ×Ｍ×Ｎ×２×（４＋Ｑ）となる。例えば、Ｑは、５～５０である。

【００８５】

［評価］

実際に、本実施の形態におけるＮベストリスコアリングと、従来のＲＮＮ言語モデルとの比較評価を行った。Ｎベストリランキングモデルを使用する際（評価時）、モデルは、（１－１）式及び（１－２）式にしたがい、２クラスの事後確率Ｐ（ｙ｜Ｘ$^{(u)}$，Ｘ$^{(v)}$），ｙ＝｛０，１｝を推定する。リランキング装置１０は、これらの事後確率をそのまま用いてＮベストリランキングを行ってもよい。また、リランキング装置１０は、従来のＮベストリランキングモデルと同様に、（６）式を用いて、元々の音声認識スコアとＮベストリランキングモデルによるスコア（事後確率の対数値）とを重み付け加算し、その値を基に、Ｎベストリランキングを行ってもよい。

【００８６】

スコア＝（１－λ）×音声認識スコア＋λ×Ｎベストリランキングモデルによるスコア
・・・（６）

【００８７】

なお、（６）式において、λは、Ｎベストリランキングモデルの重みであり、０≦λ≦１である。リランキング装置１０は、λ＝１に設定した場合は、音声認識スコアを用いず、Ｎベストリランキングモデルによるスコアのみを用いて、Ｎベストリランキングを行う。

【００８８】

［評価結果］

図６は、Ｎベストリランキングの評価結果を示す図である。図６では、Ｎベストリランキングの評価結果の例として、日本語話し言葉コーパスを用いて、従来のＲＮＮ言語モデル、本実施の形態のリランキング装置１０が用いるＮベストリランキングモデル、及び、ＲＮＮ言語モデルのスコアを特徴量ベクトルの一次元として加えたＮベストリランキングモデル（他の構築例４）を比較評価した結果を示す。評価は、（６）式に従い、音声認識スコアとＮベストリランキングモデルによるスコアとを重み加算したスコアを用いてＮベストリランキングを行っている。

【００８９】

図６に示すように、従来のＲＮＮ言語モデルよりも、本実施の形態のＮベストリランキングモデルの方が着実に音声認識精度を改善できることが分かる。また、構築例４のように、ＲＮＮ言語モデルのスコアを特徴量ベクトルの一次元として加えることによって、Ｎベストリランキングモデルの音声認識精度をさらに改善できることが分かる。

【００９０】

さらに、図６より、従来のＲＮＮ言語モデルでは、音声認識スコアを使用する必要があ

り、かつ、重みλに比較的狭い最適値があることが分かる。本評価では、従来のＲＮＮ言語モデルは、λ＝０．８付近である。一方、本実施の形態のＮベストリランキングモデルを用いた場合、従来のＲＮＮ言語モデルの場合と比較して、λの最適値の範囲が広いことが分かる。すなわち、本実施の形態のＮベストリランキングモデルは、λの値に頑健である。或いは、本実施の形態のＮベストリランキングモデルを用いた場合、λ＝１において最高か最高に近い音声認識精度が得られているので、音声認識スコアを使用しなくてもよいことが分かる。

【００９１】

［実施の形態の効果］

本実施の形態に係るリランキング装置１０では、音声認識結果であるＮベスト仮説の入力を受け付け、Ｎベスト仮説中の二つの仮説に対し、ＮＮで表されるＮベストリランキングモデルを用いて、いずれの仮説がより高い音声認識精度を有しているかを判定する。

【００９２】

前述したように、リランキング装置１０がＮベスト仮説のリランキングを行う上で、Ｎベストリランキングモデルが有すべき必要最低限の機能は、Ｎベスト仮説から最も高精度な仮説（オラクル仮説）を、最終的な音声認識結果として見つけ出すことである。このため、リスコアリング後のＮベスト仮説は、必ずしもソートされている必要はない。

【００９３】

そこで、本実施の形態では、Ｎベスト仮説の中からオラクル仮説をリランキングにより見つけ出すために、Ｎベストリランキングモデルに、Ｎベスト仮説中の二つの仮説のうちどちらの仮説の方がより高い音声認識精度を有しているかを判定できる機能を持たせた。言い換えると、本実施の形態では、Ｎベストリランキングモデルに、Ｎベスト仮説中の二つの仮説を対象に、一対一の仮説比較を実行できる機能を持たせた。

【００９４】

具体的には、リランキング装置１０は、ＮＮで表され、一対一の二つの仮説の比較を行う機能を持つＮベストリランキングモデルを用い、Ｎベストリランキングモデルを用いた一対一の二つの仮説に対する比較処理を繰り返すことによって、Ｎベスト仮説の中からオラクル仮説を見つけ出すことを可能にしている。

【００９５】

さらに、学習装置２０は、Ｎベストリランキングモデルに、音声認識精度が既知である学習用のＮベスト仮説のうちの二つの仮説を１組として、複数の組についてそれぞれ音声認識精度の高低を判定できるように予め学習させている。したがって、学習装置２０は、Ｎベストリランキングを行う上で最適なモデルを、最新のＮＮに基づき実現することができる。そして、リランキング装置１０は、学習装置２０において学習されたＮベストリランキングモデルを使用することによって、一対一の二つの仮説の比較を精度よく行うことができ、このリランキング装置１０によるオラクル仮説の抽出を高精度にできる。

【００９６】

このように、本実施の形態によれば、Ｎベスト仮説の中から最終的な音声認識結果を、精度よく得ることができる、ＮＮで表されたＮベストリランキングモデルを実現することができる。そして、本実施によれば、Ｎベストリランキングモデルを用いることによって、最終的な音声認識結果を精度よく得ることができる。

【００９７】

なお、本実施の形態では、一対一の仮説比較をオラクル仮説（最も精度が高い仮説）と推定される仮説が見つかった時点で処理を終了していたが、オラクル仮説と推定される仮説を除いたＮ－１個の仮説に対して，オラクル仮説を見つけるのと同様の処理を行うことで、二番目に精度が高いと推定される仮説を見つけ出すことができる。以降、この処理を繰り返すことによって、Ｎベスト仮説のソートも可能である。

【００９８】

また、本実施の形態では、音声認識のＮベスト仮説をリランキングするためのモデルとして、図２に例示するＮベストリランキングモデルについて説明した。ただし、本実施の

形態のモデルは、音声認識のＮベスト仮説への適用にとどまらず、Ｎベスト仮説を採用しているあらゆるタスクに適用可能である。例えば、機械翻訳や文章要約などにも本実施の形態を適用することが可能である。また、文字列に限らず、数字やアルファベットを含む複数の系列にも本実施の系列を適用することが可能である。

【００９９】

このため、本実施の形態は、ある一つの入力に対する解の候補として挙げられた複数の系列であれば、このうちの二つの系列に対し、ＮＮで表されるモデルを用いて、二つの系列のうちより精度が高い（誤りが少ない）系列を判定できる。そして、本実施の形態では、二つの系列のうち、より精度が高いと判定した系列を比較対象として残し、他方の系列を比較対象から外し、精度が高いと判定した系列を二つの系列の一方の仮説として選択し、複数の系列のうち、判定が行われていない系列のいずれかを他方の仮説として選択する。そして、本実施の形態では、判定処理と選択処理とを、所定条件に達するまで順次実行させせる。これによって、本実施の形態によれば、所定条件に達した場合に比較対象として残っている系列を、最も精度が高い系列、すなわち、最終的な出力として出力することができる。

【０１００】

また、この場合には、本実施の形態では、精度が既知である学習用の複数の系列のうちの二つの系列の特徴量が与えられたとき、それら二つの系列の精度の高低が判定できるような、ＮＮで表されるモデルを学習する。そして、本実施の形態では、二つの系列のうち精度がより高い（誤りがより少ない）系列に他方の系列よりも高い順位が付与されている場合に正解を示す教師ラベルを付与してモデルに学習させる。そして、本実施の形態では、二つの系列のうち精度がより高い（誤りがより少ない）系列に他方の系列よりも低い順位が付与されている場に誤りを示す教師ラベルを付与してモデルに学習させる。本実施の形態では、このモデルによって、一対一の二つの系列の比較が高精度で行うことができ、この結果、最も精度の高い系列を精度よく得ることができる。

【０１０１】

［システム構成等］

図示した各装置の各構成要素は機能概念的なものであり、必ずしも物理的に図示の如く構成されていることを要しない。すなわち、各装置の分散・統合の具体的形態は図示のものに限られず、その全部又は一部を、各種の負荷や使用状況等に応じて、任意の単位で機能的又は物理的に分散・統合して構成することができる。例えば、リランキング装置１０及び学習装置２０は、一体の装置であってもよい。さらに、各装置にて行なわれる各処理機能は、その全部又は任意の一部が、ＣＰＵ及び当該ＣＰＵにて解析実行されるプログラムにて実現され、あるいは、ワイヤードロジックによるハードウェアとして実現され得る。

【０１０２】

また、本実施形態において説明した各処理のうち、自動的に行われるものとして説明した処理の全部又は一部を手動的におこなうこともでき、あるいは、手動的におこなわれるものとして説明した処理の全部又は一部を公知の方法で自動的におこなうこともできる。また、本実施形態において説明した各処理は、記載の順にしたが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にあるいは個別に実行されてもよい。この他、上記文書中や図面中で示した処理手順、制御手順、具体的名称、各種のデータやパラメータを含む情報については、特記する場合を除いて任意に変更することができる。

【０１０３】

［プログラム］

図７は、プログラムが実行されることにより、リランキング装置１０或いは学習装置２０が実現されるコンピュータの一例を示す図である。コンピュータ１０００は、例えば、メモリ１０１０、ＣＰＵ１０２０を有する。また、コンピュータ１０００は、ハードディスクドライブインタフェース１０３０、ディスクドライブインタフェース１０４０、シリ

アルポートインタフェース１０５０、ビデオアダプタ１０６０、ネットワークインタフェース１０７０を有する。これらの各部は、バス１０８０によって接続される。

【０１０４】

メモリ１０１０は、ＲＯＭ１０１１及びＲＡＭ１０１２を含む。ＲＯＭ１０１１は、例えば、ＢＩＯＳ（Basic Input Output System）等のブートプログラムを記憶する。ハードディスクドライブインタフェース１０３０は、ハードディスクドライブ１０３１に接続される。ディスクドライブインタフェース１０４０は、ディスクドライブ１０４１に接続される。例えば磁気ディスクや光ディスク等の着脱可能な記憶媒体が、ディスクドライブ１０４１に挿入される。シリアルポートインタフェース１０５０は、例えばマウス１１１０、キーボード１１２０に接続される。ビデオアダプタ１０６０は、例えばディスプレイ１１３０に接続される。

【０１０５】

ハードディスクドライブ１０３１は、例えば、ＯＳ１０９１、アプリケーションプログラム１０９２、プログラムモジュール１０９３、プログラムデータ１０９４を記憶する。すなわち、リランキング装置１０或いは学習装置２０の各処理を規定するプログラムは、コンピュータ１０００により実行可能なコードが記述されたプログラムモジュール１０９３として実装される。プログラムモジュール１０９３は、例えばハードディスクドライブ１０３１に記憶される。例えば、リランキング装置１０或いは学習装置２０における機能構成と同様の処理を実行するためのプログラムモジュール１０９３が、ハードディスクドライブ１０３１に記憶される。なお、ハードディスクドライブ１０３１は、ＳＳＤ（Solid State Drive）により代替されてもよい。

【０１０６】

また、上述した実施形態の処理で用いられる設定データは、プログラムデータ１０９４として、例えばメモリ１０１０やハードディスクドライブ１０３１に記憶される。そして、ＣＰＵ１０２０が、メモリ１０１０やハードディスクドライブ１０３１に記憶されたプログラムモジュール１０９３やプログラムデータ１０９４を必要に応じてＲＡＭ１０１２に読み出して実行する。

【０１０７】

なお、プログラムモジュール１０９３やプログラムデータ１０９４は、ハードディスクドライブ１０３１に記憶される場合に限らず、例えば着脱可能な記憶媒体に記憶され、ディスクドライブ１０４１等を介してＣＰＵ１０２０によって読み出されてもよい。あるいは、プログラムモジュール１０９３及びプログラムデータ１０９４は、ネットワーク（ＬＡＮ（Local Area Network）、ＷＡＮ（Wide Area Network）等）を介して接続された他のコンピュータに記憶されてもよい。そして、プログラムモジュール１０９３及びプログラムデータ１０９４は、他のコンピュータから、ネットワークインタフェース１０７０を介してＣＰＵ１０２０によって読み出されてもよい。

【０１０８】

以上、本発明者によってなされた発明を適用した実施形態について説明したが、本実施形態による本発明の開示の一部をなす記述及び図面により本発明は限定されることはない。すなわち、本実施形態に基づいて当業者等によりなされる他の実施形態、実施例及び運用技術等は全て本発明の範疇に含まれる。

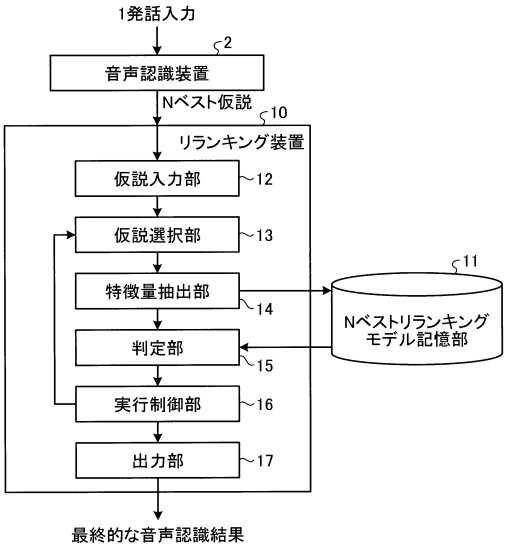
【符号の説明】

【０１０９】

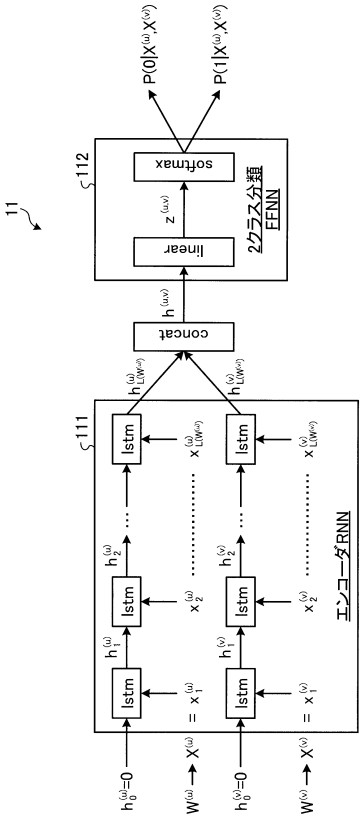
２　音声認識装置
１０　リランキング装置
１１，２１　Ｎベストリランキングモデル記憶部
１２　仮説入力部
１３　仮説選択部
１４　特徴量抽出部
１５　判定部

１６　実行制御部
１７　出力部
２０　学習装置
２２　仮説入力部
２３　学習部
２３１　教師ラベル付与部
２３２　入替部

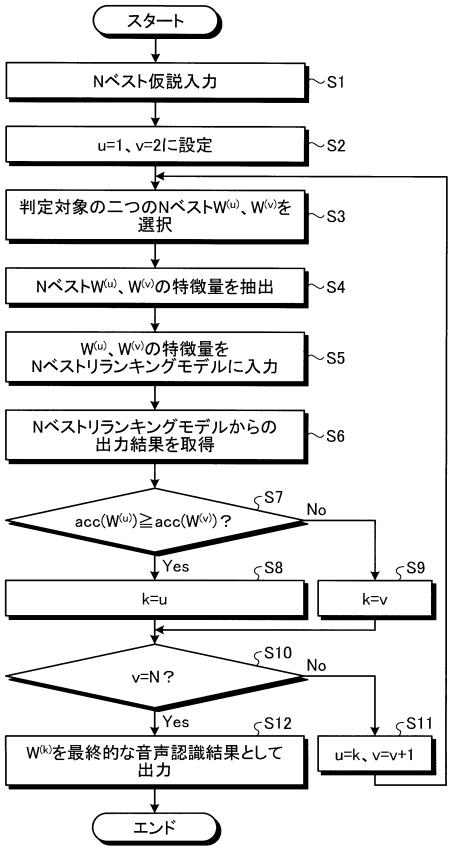
【図１】



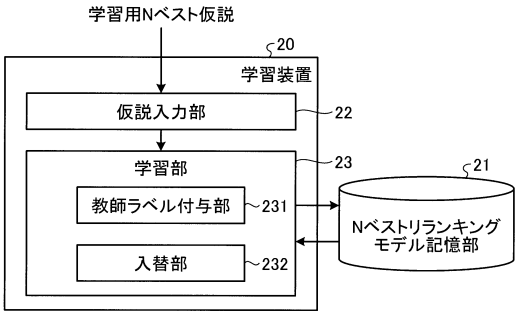
【図２】



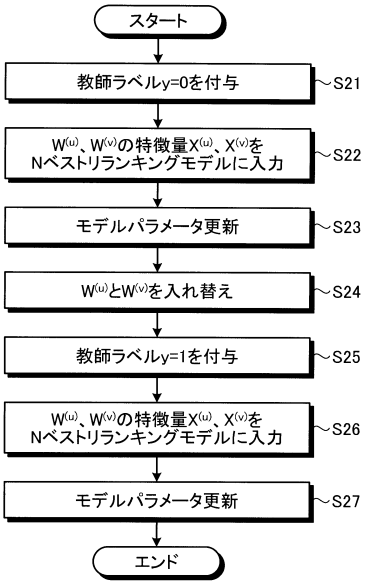
【図３】


スタート
Nベスト仮説入力
S1
u=1、v=2に設定
S2
判定対象の二つのNベストW(u)、W(v)を選択
S3
NベストW(u)、W(v)の特徴量を抽出
S4
W(u)、W(v)の特徴量をNベストリランキングモデルに入力
S5
Nベストリランキングモデルからの出力結果を取得
S6
acc(W(u))≧acc(W(v))？
S7
No
Yes
S8
k=u
S9
k=v
v=N？
S10
No
Yes
S12
W(k)を最終的な音声認識結果として出力
S11
u=k、v=v+1
エンド

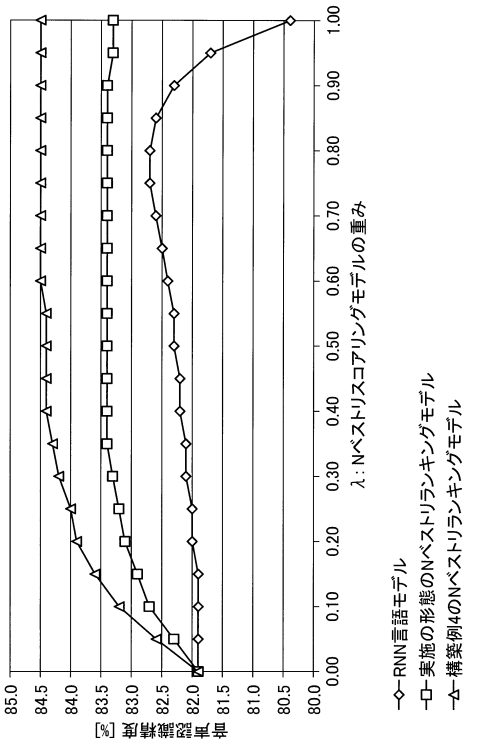
【図４】


学習用Nベスト仮説
20
学習装置
仮説入力部
22
学習部
23
21
教師ラベル付与部
231
入替部
232
Nベストリランキングモデル記憶部

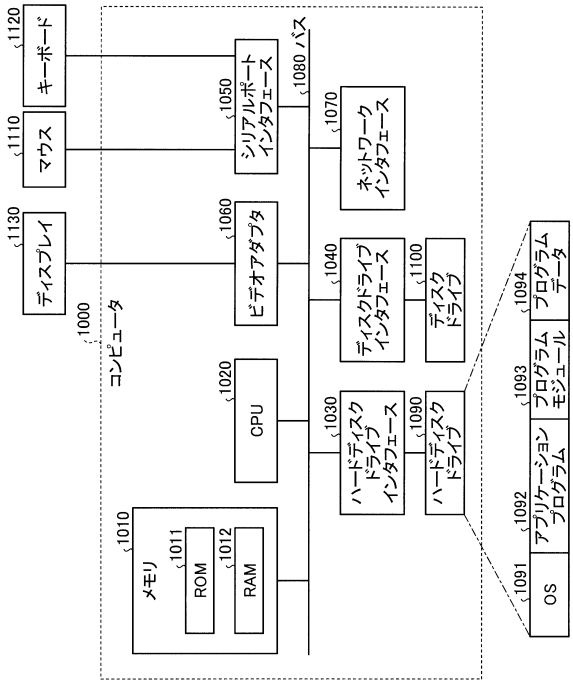
【図５】


スタート
教師ラベルy=0を付与
S21
W(u)、W(v)の特徴量X(u)、X(v)をNベストリランキングモデルに入力
S22
モデルパラメータ更新
S23
W(u)とW(v)を入れ替え
S24
教師ラベルy=1を付与
S25
W(u)、W(v)の特徴量X(u)、X(v)をNベストリランキングモデルに入力
S26
モデルパラメータ更新
S27
エンド

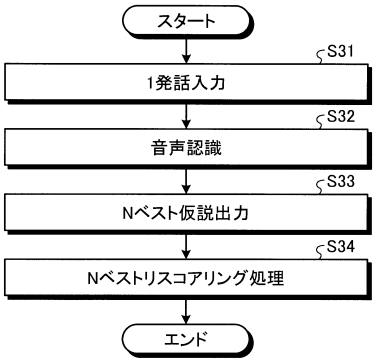
【図６】


音声認識精度 [%]
85.0
84.5
84.0
83.5
83.0
82.5
82.0
81.5
81.0
80.5
80.0
0.00
0.10
0.20
0.30
0.40
0.50
0.60
0.70
0.80
0.90
1.00
λ：Nベストリスコアリングモデルの重み
RNN言語モデル
実施の形態のNベストリランキングモデル
構築例4のNベストリランキングモデル

【図７】



【図８】



【図９】

発話(正解単語列): 私 と 友人 は 2018 年 4 月 に □ 名古屋 大学 に 入学 し ます □

1位仮説(認識結果): 私 と 求人 で 2018 念 7 月 に □ 名古屋 退学 に 入閣 し ます □ 120.3
2位仮説(認識結果): 轍 と 友人 は 2018 年 4 月 に □ 名古屋 大学 に 入閣 し ます □ 115.8
3位仮説(認識結果): 私 と 友人 は 2018 年 4 月 に □ 名古屋 大学 に 入閣 し ます □ 108.5
4位仮説(認識結果): 轍 と 友人 は 2017 年 4 月 に □ 名古屋 大学 に 入閣 し ます □ 101.4
5位仮説(認識結果): 私 と 求人 で 2018 念 7 月 に あ 名古屋 退学 に 入閣 し まし た 98.2

音声認識スコア

フロントページの続き

(72)発明者　中谷　智広
東京都千代田区大手町一丁目５番１号　日本電信電話株式会社内

審査官　中村　天真

(56)参考文献　特開２０１１－２４３１４７（ＪＰ，Ａ）
岩立将和，外２名，トーナメントモデルを用いた日本語係り受け解析，自然言語処理，２００８年１０月，第１５巻，第５号，p.169-185
島岡聖世，外２名，オートエンコーダにおける単語ベクトルの学習，言語処理学会第１９回年次大会発表論文集，２０１３年　３月，p.612-615
小川厚徳，外３名，一対一の仮説比較を行うencoder-classifierモデルを用いたNベスト音声認識仮説のリスコアリング，日本音響学会講演論文集，２０１８年　３月，p.23-24

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ１０Ｌ　　１５／００－１５／３４
Ｇ０６Ｆ　　４０／２０－４０／５８