(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7192995号
(P7192995)

(45)発行日　令和4年12月20日(2022.12.20)

(24)登録日　令和4年12月12日(2022.12.12)

(51)国際特許分類

Ｇ１０Ｌ　15/08　(2006.01)	Ｇ１０Ｌ	15/08	２００Ｚ
Ｇ１０Ｌ　15/16　(2006.01)	Ｇ１０Ｌ	15/16	

ＦＩ

請求項の数　10　（全26頁）

(21)出願番号　特願2021-537548(P2021-537548)
(86)(22)出願日　令和1年8月8日(2019.8.8)
(86)国際出願番号　PCT/JP2019/031517
(87)国際公開番号　WO2021/024491
(87)国際公開日　令和3年2月11日(2021.2.11)
審査請求日　令和3年11月19日(2021.11.19)

(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
弁理士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
(72)発明者　中谷 智広
最終頁に続く

(54)【発明の名称】　判定装置、学習装置、判定方法及び判定プログラム

(57)【特許請求の範囲】

【請求項１】

音声認識精度のスコアが対応付けられたＮベスト仮説の入力を受け付ける入力部と、

入力を受け付けた前記Ｎベスト仮説のうち、判定対象である二つの仮説を選択する選択部と、

選択された二つの仮説が与えられたとき、前記二つの仮説を隠れ状態ベクトルに変換し、前記二つの仮説の隠れ状態ベクトルを基に前記二つの仮説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複数の補助モデルと、前記複数の補助モデルでそれぞれ変換された前記二つの仮説の隠れ状態ベクトルを基に、前記二つの仮説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メインモデルとを用いて、前記二つの仮説の精度の高低を判定する判定部と、

を有することを特徴とする判定装置。

【請求項２】

前記選択部は、前記Ｎベスト仮説のスコアの昇順に前記二つの仮説を選択することを特徴とする請求項１に記載の判定装置。

【請求項３】

前記判定部は、前記メインモデルから出力された判定情報、または、各補助モデルから出力された判定情報と前記メインモデルから出力された判定情報とに対して計算した重み付け和の値、に基づいて前記二つの仮説の精度の高低を判定することを特徴とする請求項１または２に記載の判定装置。

【請求項４】
各補助モデルは、前記二つの仮説を、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事後確率を出力し、
前記メインモデルは、ニューラルネットワークを用いて、前記複数の補助モデルでそれぞれ変換された前記二つの仮説の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事後確率を出力する
ことを特徴とする請求項１～３のいずれか一つに記載の判定装置。
【請求項５】
音声認識精度が既知である学習用の二つの仮説の入力を受け付ける入力部と、
前記二つの仮説が与えられたとき、前記二つの仮説を隠れ状態ベクトルに変換し、前記二つの仮説の隠れ状態ベクトルを基に前記二つの仮説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複数の補助モデルと、前記複数の補助モデルでそれぞれ変換された前記二つの仮説の隠れ状態ベクトルを基に、前記二つの仮説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メインモデルとに対し、各ニューラルネットワークが前記二つの仮説の精度の高低を判定するタスクを個別に行うとみなしたマルチタスク学習を行わせる学習部と、
を有することを特徴とする学習装置。
【請求項６】
前記学習部は、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複数の補助モデル及び前記メインモデルに学習させ、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複数の補助モデル及び前記メインモデルに学習させることを特徴とする請求項５に記載の学習装置。
【請求項７】
前記学習部は、各ニューラルネットワークによって実行された各タスクについて所定の損失をそれぞれ計算し、各損失の重み付け和に基づいて、各ニューラルネットワークのパラメータの値を更新することを特徴とする請求項５または６に記載の学習装置。
【請求項８】
各補助モデルは、前記二つの仮説を、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事後確率を出力し、
前記メインモデルは、ニューラルネットワークを用いて、前記複数の補助モデルでそれぞれ変換された前記二つの仮説の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事後確率を出力する
ことを特徴とする請求項５～７のいずれか一つに記載の学習装置。
【請求項９】
判定装置が実行する判定方法であって、
音声認識精度のスコアが対応付けられたＮベスト仮説の入力を受け付ける工程と、
入力を受け付けた前記Ｎベスト仮説のうち、判定対象である二つの仮説を選択する工程と、
選択された二つの仮説が与えられたとき、前記二つの仮説を隠れ状態ベクトルに変換し、前記二つの仮説の隠れ状態ベクトルを基に前記二つの仮説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複数の補助モデルと、前記複数の補助モデルでそれぞれ変換された前記二つの仮説の隠れ状態ベクトルを基に、前記二つの仮説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メインモデルとを用いて、前記二つの仮説の精度の高低を判定する工程と、
を含んだことを特徴とする判定方法。
【請求項１０】
音声認識精度のスコアが対応付けられたＮベスト仮説の入力を受け付けるステップと、

入力を受け付けた前記Ｎベスト仮説のうち、判定対象である二つの仮説を選択するステップと、
選択された二つの仮説が与えられたとき、前記二つの仮説を隠れ状態ベクトルに変換し、前記二つの仮説の隠れ状態ベクトルを基に前記二つの仮説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複数の補助モデルと、前記複数の補助モデルでそれぞれ変換された前記二つの仮説の隠れ状態ベクトルを基に、前記二つの仮説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メインモデルとを用いて、前記二つの仮説の精度の高低を判定するステップと、
をコンピュータに実行させるための判定プログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、判定装置、学習装置、判定方法及び判定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とＮベストリランキングとは同義として扱われている。
【０００３】
Ｎベストリスコアリング方法では、音声認識結果である仮説の中からスコアの高い所定数（Ｎ個）の仮説を出力する。そして、Ｎベストリスコアリング方法では、この中から尤もらしい仮説を音声認識結果として出力する。ここで、スコアが最大となる仮説が必ずしもベストな仮説とは限らない。このため、二つの仮説のうち尤もらしい仮説（正解に近い仮説）を選択する二択問題をトーナメント方式で繰り返し適用することで、尤もらしい仮説を選択するリランキング装置が提案されている（例えば、非特許文献１参照）。
【先行技術文献】
【非特許文献】
【０００４】
【文献】Atsunori Ogawa, Marc Delcroix, Shigeki Karita, Tomohiro Nakatani, “RESCORING N-BEST SPEECH RECOGNITION LIST BASED ON ONE-ON-ONE HYPOTHESIS COMPARISON USING ENCODER-CLASSIFIER MODEL”, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) , pp. 6099-6103, 2018.
【発明の概要】
【発明が解決しようとする課題】
【０００５】
非特許文献１に記載のリランキング方法では、Ｎ仮説をスコアの降順に並べ、先頭の仮説（スコアが最も高い仮説）から順に二つの仮説を選択し、学習済みの二択問題を解くニューラルネットワーク（ＮＮ）にこれらの仮説を入力することによって、いずれかの仮説を選択する処理を繰り返し行い、最終的に選択された仮説を音声認識結果として出力することが記載されている。非特許文献１に記載のリランキング方法では、一定の精度で音声認識結果を出力するが、さらに、近年では、音声認識結果の出力に対して、精度の安定化が要求されている。
【０００６】

本発明は、上記に鑑みてなされたものであって、ある音声信号に対する解の候補として挙げられた複数の仮説に対し、最も精度が高い仮説を安定した精度で判定することができる判定装置、学習装置、判定方法及び判定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】

上述した課題を解決し、目的を達成するために、本発明に係る判定装置は、音声認識精度のスコアが対応付けられたＮベスト仮説の入力を受け付ける入力部と、入力を受け付けたＮベスト仮説のうち、判定対象である二つの仮説を選択する選択部と、選択された二つの仮説が与えられたとき、二つの仮説を隠れ状態ベクトルに変換し、二つの仮説の隠れ状態ベクトルを基に二つの仮説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複数の補助モデルと、複数の補助モデルでそれぞれ変換された二つの仮説の隠れ状態ベクトルを基に、二つの仮説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メインモデルとを用いて、二つの仮説の精度の高低を判定する判定部と、を有することを特徴とする。

【０００８】

また、本発明に係る学習装置は、音声認識精度が既知である学習用の二つの仮説の入力を受け付ける入力部と、二つの仮説が与えられたとき、二つの仮説を隠れ状態ベクトルに変換し、二つの仮説の隠れ状態ベクトルを基に二つの仮説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複数の補助モデルと、複数の補助モデルでそれぞれ変換された二つの仮説の隠れ状態ベクトルを基に、二つの仮説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メインモデルとに対し、各ニューラルネットワークが二つの仮説の精度の高低を判定するタスクを個別に行うとみなしたマルチタスク学習を行わせる学習部と、を有することを特徴とする。

【０００９】

また、本発明に係る判定方法は、判定装置が実行する判定方法であって、音声認識精度のスコアが対応付けられたＮベスト仮説の入力を受け付ける工程と、入力を受け付けたＮベスト仮説のうち、判定対象である二つの仮説を選択する工程と、選択された二つの仮説が与えられたとき、二つの仮説を隠れ状態ベクトルに変換し、二つの仮説の隠れ状態ベクトルを基に二つの仮説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複数の補助モデルと、複数の補助モデルでそれぞれ変換された二つの仮説の隠れ状態ベクトルを基に、二つの仮説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メインモデルとを用いて、二つの仮説の精度の高低を判定する工程と、を含んだことを特徴とする。

【００１０】

また、本発明に係る判定プログラムは、音声認識精度のスコアが対応付けられたＮベスト仮説の入力を受け付けるステップと、入力を受け付けたＮベスト仮説のうち、判定対象である二つの仮説を選択するステップと、選択された二つの仮説が与えられたとき、二つの仮説を隠れ状態ベクトルに変換し、二つの仮説の隠れ状態ベクトルを基に二つの仮説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複数の補助モデルと、複数の補助モデルでそれぞれ変換された二つの仮説の隠れ状態ベクトルを基に、二つの仮説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メインモデルとを用いて、二つの仮説の精度の高低を判定するステップと、をコンピュータに実行させる。

【発明の効果】

【００１１】

本発明によれば、ある音声信号に対する解の候補として挙げられた複数の仮説に対し、最も精度が高い仮説を、安定した精度で判定することができる。

【図面の簡単な説明】

【００１２】

【図１】図１は、実施の形態１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、第１補助モデル～第Ｍ補助モデル及びメインモデルの構成を説明する図である。
【図３】図３は、第１補助モデルの構築例を示す図である。
【図４】図４は、実施の形態１に係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５】図５は、図１に示すリランキング装置が、Ｎベスト仮説に対して実行するリランキング処理を説明する図である。
【図６】図６は、実施の形態２に係る学習装置の機能構成の一例を示す図である。
【図７】図７は、図６に示す入替部の処理を説明する図である。
【図８】図８は、実施の形態２に係る学習処理の処理手順を示すフローチャートである。
【図９】図９は、実施の形態３に係るリランキング装置の要部構成を示す図である。
【図１０】図１０は、実施の形態３に係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１１】図１１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。
【発明を実施するための形態】
【００１３】
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。
【００１４】
本実施の形態では、音声認識結果であるＮ（Ｎ≧２）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）を得るためのモデルを用いたリランキング装置、及び、Ｎベストのリランキング処理に用いる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ではなく、Ｎベストリランキングと表現を統一して説明する。
【００１５】
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、本実施の形態におけるモデルが有すべき必要最低限な機能について述べる。本実施の形態では、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が目的である。
【００１６】
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、本実施の形態における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。
【００１７】
そこで、本実施の形態に係るリランキング装置は、一対一の二つの仮説の比較を行う機能を持つ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。さらに、本実施の形態では、モデルとして、ニューラルネットワーク（ＮＮ）で表されるメインモデルと、ＮＮで表される複数の補助モデルとを用いる。各補助モデルは、二つの仮説が与えられたとき、二つの仮説を隠れ状態ベクトルに変換し、二つの仮説の隠れ状態ベクトルを基に二つの仮説の精度の高低を判定するモデルである。メインモデルは、複数の補助モデルでそれぞれ変換された二つの仮説の隠れ状態ベクトルを基に、二つの仮説の精度の高低を判定するモデルである。
【００１８】
そして、本実施の形態に係るリランキング装置は、Ｎベスト仮説のスコアの昇順に二つの仮説を選択し、選択した二つの仮説のうち、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、未判定の仮説から昇順に他方の仮説を選択して、複数の補助

モデル及びメイン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、未判定の仮説のうち最も順位の低い仮説を他方の仮説として選択し、複数の補助モデル及びメイン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、安定した精度で、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。

【００１９】

［実施の形態１］

［リランキング装置］

まず、実施の形態１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して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。

【００２０】

図１は、実施の形態１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。

【００２１】

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、スコアと対応付けて、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。

【００２２】

リランキング装置１０は、モデル記憶部１１、仮説入力部１２、仮説選択部１３（選択部）、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。

【００２３】

モデル記憶部１１は、補助モデル及びメインモデル１１０を記憶する。図１の例では、モデル記憶部１１は、補助モデルとして、第１補助モデル１１１～第Ｍ補助モデル１１Ｍを記憶する。第１補助モデル１１１～第Ｍ補助モデル１１Ｍ及びメインモデル１１０は、ＮＮで表されるモデルである。第１補助モデル１１１～第Ｍ補助モデル１１Ｍ及びメインモデル１１０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。

【００２４】

第１補助モデル補助１１１～第Ｍ補助モデル１１Ｍは、選択された二つの仮説が与えられたとき、二つの仮説を隠れ状態ベクトルに変換し、二つの仮説の隠れ状態ベクトルを基に二つの仮説の精度の高低を判定できるような、ＮＮで表される。第１補助モデル１１１～第Ｍ補助モデル１１Ｍは、学習用のＮベスト仮説のうち二つの仮説が与えられたときに、二つの仮説について、その二つの仮説の音声認識精度の高低を判定できるように学習される。第１補助モデル１１１～第Ｍ補助モデル１１Ｍは、二つの仮説を、それぞれＲＮＮを用いて隠れ状態ベクトルに変換する。そして、第１補助モデル１１１～第Ｍ補助モデル１１Ｍ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事後確率をそれぞれ生成する。

【００２５】

メインモデル１１０は、第１補助モデル１１１～第Ｍ補助モデル１１Ｍにおいてそれぞ

れ変換された二つの仮説の隠れ状態ベクトルを基に、二つの仮説の精度の高低を判定できるような、ＮＮで表される。メインモデル１１０は、第１補助モデル１１１～第Ｍ補助モデル１１Ｍにおいてそれぞれ変換された学習用の二つの仮説の隠れ状態ベクトルを基に、二つの仮説の精度の高低を判定できるように学習される。メインモデル１１０は、ＮＮを用いて、第１補助モデル１１１～第Ｍ補助モデル１１Ｍにおいてそれぞれ変換された学習用の二つの仮説の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事後確率を生成する。

【００２６】

第１補助モデル１１１～第Ｍ補助モデル１１Ｍ及びメインモデル１１０は、各ニューラルネットワークが二つの仮説の精度の高低を判定するタスクを個別に行うとみなしたマルチタスク学習によって学習が実行される。

【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、Ｎベスト仮説のスコアの昇順に選択する。仮説選択部１３は、最初の判定においては、Ｎベスト仮説のうち、スコアが最下位である仮説と、最下位の仮説より１つ順位が高い仮説とを判定対象として選択する。仮説選択部１３は、以降の判定においては、二つの仮説の一方の仮説として、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選択する。そして、仮説選択部１３は、二つの仮説の他方の仮説として、未判定の仮説のうち、最もスコアの順位が低い仮説を選択する。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、昇順に、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ｖ位の仮説とＮベスト仮説中のｕ（ｕ＜ｖ≦Ｎ）位の仮説（単語列）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、第１補助モデル１１１～第Ｍ補助モデル１１Ｍ及びメインモデル１１０を用いて、二つの仮説の精度の高低を判定する。判定部１５は、一対一の比較対象であるｖ位の仮説とｕ位の仮説とを第１補助モデル１１１～第Ｍ補助モデル１１Ｍにそれぞれ入力し、メインモデル１１０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、第１補助モデル１１１～第Ｍ補助モデル１１Ｍ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事後確率を出力する。メインモデル１１０は、第１補助モデル１１１～第Ｍ補助モデル１１Ｍにおいてそれぞれ変換された二つの仮説の隠れ状態ベクトルが入力されると、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事後確率を出力する。判定部１５は、メインモデル１１０による事後確率が０．５以上である場合には、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、メインモデ

ル１１０による事後確率が０．５未満である場合には、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。

【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、第１補助モデル１１１～第Ｍ補助モデル１１Ｍ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第１補助モデル１１１～第Ｍ補助モデル１１Ｍにそれぞれ入力する。

【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、他方の仮説を以降の比較対象から外す。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、判定部１５による判定が行われていない仮説のうち最もスコアの順位が低い仮説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１つ上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１３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、１位の仮説に対して比較処理が行われるまで、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、所定条件に達した場合、Ｎベスト仮説のうち、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】

［定義］

まず、リランキング装置１０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、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.（以降、参考文献１とする。）を参照）。

【００３８】

補助特徴量ベクトル$a_i^{(u)}$は、１７次元の基本の補助特徴量ベクトルを含む。さらに、補助特徴量ベクトル$a_i^{(u)}$では、前向き（forword）ＬＳＴＭＬＭの単語予測スコアを１８次元の補助特徴量として用いてもよい。ＬＳＴＭＬＭは、長短期記憶メモリ（long short-term memory：ＬＳＴＭ）ユニットを用いた再帰的ニューラルネットワーク（Recurrent Neural Network：ＲＮＮ）言語モデルであり、後述するように、本実施の形態１～３における各補助モデルを構成するモデルである。そして、補助特徴量ベクトルａ

$_i^{(u)}$では、後向き（backword）ＬＳＴＭＬＭの単語予測スコアを１９次元の補助特徴量として用いてもよい。後向きＬＳＴＭＬＭは、未来の単語列から現在の単語の生起確率を予測するものであるであり、前向きＬＳＴＭＬＭと相補的な単語予測能力を持つことから各補助モデルが出力する判定情報の精度向上が期待できる。

【００３９】

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、$x_i^{(u)}=concat(embed(w_i^{(u)}),a_i^{(u)})$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２とする。）を参照）を表す。なお、ｅｍｂｅｄ（・）を行うＮＮも第１補助モデル１１１～第Ｍ補助モデル１１Ｍ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。

【００４０】

そして、$P(0\mid X^{(u)},X^{(v)})$を、二つの仮説$W^{(u)}$，$W^{(v)}$の精度の高低の並びが正しいことを示す事後確率と定義する。$P(0\mid X^{(u)},X^{(v)})$は、第１補助モデル１１１～第Ｍ補助モデル１１Ｍ及びメインモデル１１０のそれぞれにおいて、生成される。

【００４１】

[補助モデル及びメインモデルの構成]

第１補助モデル１１１～第Ｍ補助モデル１１Ｍ、メインモデル及び構成について説明する。図２は、第１補助モデル１１１～第Ｍ補助モデル１１Ｍ及びメインモデル１１０の構成を説明する図である。図２では、処理の流れを説明するため、判定部１５も記載される。

【００４２】

図２に示すように、各補助モデルは、それぞれ、二つの第１変換部、第１結合部及び判定情報生成部を有する。具体的に、第１補助モデル１１１を例に説明する。第１補助モデル１１１は、二つの第１変換部１１１－１ｕ，１１１－１ｖ、第１結合部１１１－２及び判定情報生成部１１１－３を有する。

【００４３】

第１変換部１１１－１ｕは、比較対象の二つの仮説$W^{(u)}$，$W^{(v)}$のうち、仮説$W^{(u)}$の特徴量$X^{(u)}$の入力を受け付け、隠れ状態ベクトルに変換する。第１変換部１１１－１ｖは、比較対象の二つの仮説$W^{(u)}$，$W^{(v)}$のうち、仮説$W^{(v)}$の特徴量$X^{(v)}$の入力を受け付け、隠れ状態ベクトルに変換する。

【００４４】

第１結合部１１１－２は、第１変換部１１１－１ｕ，第１変換部１１１－１ｖが変換した二つの隠れ状態ベクトルを結合する。判定情報生成部１１１－３は、二つの仮説$W^{(u)}$，$W^{(v)}$の精度の高低の並びが正しいことを示す事後確率$P(0\mid X^{(u)},X^{(v)})$を、判定情報として生成する。他の補助モデルも、第１補助モデル１１１と同じ構成であり、与えられた二つの仮説$W^{(u)}$，$W^{(v)}$に対し、それぞれ、隠れ状態ベクトルの変換、隠れ状態ベクトルの結合、及び、判定情報の生成を含むタスクをそれぞれ実行できるようにしている。なお、各補助モデルは、学習時におけるランダム初期化時における初期値がそれぞれ異なる。

【００４５】

メインモデル１１０は、メイン結合部１１０－１と、判定情報生成部１１０－２とを有する。メイン結合部１１０－１は、第１補助モデル１１１～第Ｍ補助モデル１１Ｍでそれぞれ変換された二つの仮説の隠れ状態ベクトルを結合する。判定情報生成部１１１－３は、二つの仮説$W^{(u)}$，$W^{(v)}$の精度の高低の並びが正しいことを示す事後確率$P(0\mid X^{(u)},X^{(v)})$を、判定情報として生成する。

【００４６】
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$X^{(u)}$，$X^{(v)}$が各補助モデルに与えられたとき、メインモデル１１０は、記号ｙ＝｛０｝の事後確率Ｐ（０｜$X^{(u)}$，$X^{(v)}$）を出力する。
【００４７】
判定部１５は、メインモデル１１０が出力した事後確率Ｐ（０｜$X^{(u)}$，$X^{(v)}$）を受け取り、判定を行う。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事後確率である。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事後確率Ｐ（０｜$X^{(u)}$，$X^{(v)}$）を取得し、取得した事後確率を所定の閾値と比較して、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。
【００４８】
具体的には、判定部１５は、事後確率Ｐ（０｜$X^{(u)}$，$X^{(v)}$）が０．５以上である場合には、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し、ｙ＝０を出力する。また、判定部１５は、事後確率Ｐ（０｜$X^{(u)}$，$X^{(v)}$）が０．５未満である場合には、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し、ｙ＝１を出力する。
【００４９】
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。
【００５０】
Ｐ（０｜$X^{(u)}$，$X^{(v)}$）≧０．５
if WER (Word error rate) of $W^{(u)}$≦WER of $W^{(v)}$・・・（１－１）
Ｐ（０｜$X^{(u)}$，$X^{(v)}$）＜０．５
otherwise ・・・（１－２）
【００５１】
ここで、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であるため、（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。
【００５２】
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は、$W^{(v)}$として使用する。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。
【００５３】
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、$W^{(v)}$として引き続き使用する。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。
【００５４】
［補助モデルの構築例］
第１補助モデル１１１～第Ｍ補助モデル１１Ｍの構築例について説明する。第１補助モデル１１１～第Ｍ補助モデル１１Ｍは、同じ構成であるため、図３を参照し、第１補助モデル１１１の構築例を説明する。図３は、第１補助モデル１１１の構築例を示す図である

。なお、図３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。

【００５５】

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、第１補助モデル１１１は、二つの仮説の特徴量を、ＲＮＮを用いて隠れ状態ベクトルに変換する。具体的には、第１補助モデル１１１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２参照）のエンコーダＲＮＮ１１１－１ａを第１変換部１１１－１ｕ，１１１－１ｖとして有する。

【００５６】

第１補助モデル１１１は、エンコーダＲＮＮ１１１－１ａ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。そして、第１補助モデル１１１～第Ｍ補助モデル１１Ｍ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。

【００５７】

エンコーダＲＮＮ１１１－１ａの処理について説明する。エンコーダＲＮＮ１１１－１ａ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２参照）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。

【００５８】

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

【００５９】

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ＲＮＮ１１１－１ａ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。

【００６０】

エンコーダＲＮＮ１１１－１ａ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３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。

【００６１】

第１補助モデル１１１は、以上で得た二つの隠れ状態ベクトル$h_{L(W(u))}^{(u)}$，$h_{L(W(v))}^{(v)}$を、第１結合部１１１－２で連結した隠れ状態ベクトル$h^{\{(u,v)\}}$をエンコーダＲＮＮ１１１－ａの出力として以下の（３）式のように得る。

【００６２】

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

【００６３】

そして、第１補助モデル１１１は、エンコーダＲＮＮ１１１－１ａ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、第１補助モデル１１１は、１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）１１１－３ａ（詳細は、例えば、参考文献２を参照）を、判定情報生成部１１３として用いる。

エンコーダＲＮＮ１１１－１ａの出力として得た隠れ状態ベクトル$h^{\{(u,v)\}}$が、１層の１クラス分類ＦＦＮＮ１１１－３ａに入力され、最終的に、１クラスのｙ＝｛０｝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。

【００６４】

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

$$P(y \mid X^{(u)}, X^{(v)}) = \mathrm{sigmoid}(z^{\{(u,v)\}})_y \quad \cdots (5)$$

【００６５】

ここで、ｌｉｎｅａｒ（・）は、線形変換処理（詳細は、例えば、参考文献２を参照）を表す。ｓｉｇｍｏｉｄ（・）は、シグモイド処理を表す。

【００６６】

また、メインモデル１１０では、メイン結合部１１０－１は、第１結合部１１１－２と同様のベクトル連結処理を行う。また、メインモデル１１０では、判定情報生成部１１０－３は、判定情報生成部１１１－３の１層の１クラス分類ＦＦＮＮ１１１－３ａと同様の構成の１クラス分類ＦＦＮＮによって構成される。

【００６７】

［補助モデル及びメインモデルの他の構築例１］

なお、第１補助モデル１１１及びメインモデル１１０は、１クラス分類ＦＦＮＮにおけるシグモイド処理に代えて、ソフトマックス処理を行ってもよい。この場合、エンコーダＲＮＮ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（６），（７）式のように得ることができる。なお、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。

【００６８】

$$z^{\{(u,v)\}} = \mathrm{linear}(h^{\{(u,v)\}}) \quad \cdots (6)$$

$$P(y \mid X^{(u)}, X^{(v)}) = \mathrm{softmax}(z^{\{(u,v)\}})_y \quad \cdots (7)$$

【００６９】

ここで、ｓｏｆｔｍａｘ（・）は、ソフトマックス処理を表す。また、$\mathrm{softmax}(\cdot)_y$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。

【００７０】

この場合、判定部１５は、メインモデル１１０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。

【００７１】

［補助モデルの他の構築例２］

なお、図３に示すエンコーダＲＮＮ１１１－１ａ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。

【００７２】

［補助モデルの他の構築例３］

また、ＬＳＴＭユニットの代わりに、単純な（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。

【００７３】

［補助モデル及びメインモデルの他の構築例４］

さらに、補助モデル及びメインモデル１１０は、図３の構築例では、１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、補助モデル及びメインモデル１１０の他の構築例１～４の用語の詳細については、例えば、参考文献２を参照いただきたい。

【００７４】

［補助モデルの他の構築例５］

また、補助モデルは、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。

【００７５】

［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４は、実施の形態１に係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。

【００７６】

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、スコアの昇順に、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（$u < v \leqq N$）。まず、仮説選択部１３は、$u = N - 1$、$v = N$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（$X^{(u)}$，$X^{(v)}$）を各補助モデル（第１補助モデル１１１～第Ｍ補助モデル１１Ｍ）に入力する（ステップＳ５）。

【００７７】

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、事後確率$P(0 \mid X^{(u)}, X^{(v)})$を取得する。

【００７８】

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、$P(0 \mid X^{(u)}, X^{(v)}) \geqq 0.5$であるか否かを判定する（ステップＳ７）。$P(0 \mid X^{(u)}, X^{(v)}) \geqq 0.5$である場合（ステップＳ７：Ｙｅｓ）、判定部１５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し、実行制御部１６は、ｋについて$k = u$と設定する（ステップＳ８）。ｋは、比較処理後の仮説のうち、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、$P(0 \mid X^{(u)}, X^{(v)}) \geqq 0.5$でない場合（ステップＳ７：Ｎｏ）、判定部１５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し、実行制御部１６は、$k = v$と設定する（ステップＳ９）。

【００７９】

続いて、実行制御部１６は、$u = 1$であるか否かを判定する（ステップＳ１０）。$u = 1$でない場合（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、実行制御部１６は、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、$u = u - 1$、$v = k$に設定し（ステップＳ

１１）、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。

【００８０】

また、ｕ＝１である場合（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、実行制御部１６は、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。

【００８１】

このように、実施の形態１に係るリランキング装置１０は、一対一の二つの仮説の比較を行う機能を持つ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。さらに、リランキング装置１０では、モデルとして、ニューラルネットワーク（ＮＮ）で表されるメインモデル１１０と、ＮＮで表される複数の補助モデルとを用いる。

【００８２】

すなわち、リランキング装置１０では、複数の補助モデルを設け、入力された二つの仮説に対して、各補助モデルにタスクを実行させている。各補助モデルの構造は同じであっても、学習時においてパラメータのランダム初期化を行うので、同じ入力仮説に対しても異なる隠れ状態ベクトルを出力する。これにより、ある二つの入力仮説に対して、ある補助モデルが出力する隠れ状態ベクトルが適切なものでなかったとしても、別の補助モデルが適切な隠れ状態ベクトルを出力できる可能性が高まる。つまり、正確な仮説の判定結果を生成するのに適した隠れ状態ベクトルが、いずれかの補助ネットワークから出力される可能性が高くなる。この結果、リランキング装置１０のメインモデル１１０には、適切な二つの仮説に対応する隠れ状態ベクトルが安定して入力されるため、メインモデル１１０の出力値の精度も安定する。このように、実施の形態１に係るリランキングモデルは、安定した精度で、Ｎベスト仮説の中からオラクル仮説を見つけ出すことができる。

【００８３】

また、リランキング装置１０は、Ｎベスト仮説のスコアの昇順に二つの仮説を選択する。言い換えると、リランキング装置１０は、Ｎベスト仮説のうち、スコアが最も低い仮説から順に仮説ペアを選択する。図５は、図１に示すリランキング装置１０が、Ｎベスト仮説に対して実行するリランキング処理を説明する図である。

【００８４】

一般には、スコアが高い仮説の方が、尤もらしい仮説である可能性が高い。スコアの高い順に仮説を選択していくと、最もスコアの高い仮説は、Ｎ－１回の判定処理に勝ち抜かなければ、最終的な出力仮説として選択されず、尤もらしい仮説として選ばれにくくなってしまう。

【００８５】

そこで、図５に示すように、リランキング装置１０は、最終的に出力仮説として選ばれる可能性の高い仮説について、少ない判定回数で済むように、Ｎベスト仮説のうち、スコアが最も低い仮説から順に仮説ペアを選択する。言い換えると、リランキング装置１０は、図５に示すように、スコアの最も高い仮説については、シード権を与え、Ｎベスト仮説全体に対する比較処理の後の方の処理で比較処理が行われるようにし、尤もらしい仮説として選ばれやすくしている。このように、リランキング装置１０は、最終的に出力仮説として選ばれる可能性の高い仮説が、尤もらしい仮説として選ばれやすいため、安定した精度で、Ｎベスト仮説の中からオラクル仮説を見つけ出すことができる。

【００８６】

［実施の形態２］

［学習装置］

次に、実施の形態２として、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６は、実施の形態２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２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６に示すように、モデル記憶部２１、学習装置２０は、仮説入力部２２及び学習部２３を有する。

【００８７】

モデル記憶部２１は、学習対象の第１補助モデル１１１～第Ｍ補助モデル１１Ｍ及びメインモデル１１０を記憶する。第１補助モデル１１１～第Ｍ補助モデル１１Ｍ及びメインモデル１１０は、選択された二つの仮説が与えられたとき、二つの仮説を隠れ状態ベクトルに変換し、二つの仮説の隠れ状態ベクトルを基に二つの仮説の精度の高低を判定できるような、ＮＮで表される。第１補助モデル１１１～第Ｍ補助モデル１１Ｍ及びは、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、第１補助モデル１１１～第Ｍ補助モデル１１Ｍ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事後確率を判定情報として生成する。

【００８８】

メインモデル１１０は、第１補助モデル１１１～第Ｍ補助モデル１１Ｍにおいてそれぞれ変換された二つの仮説の隠れ状態ベクトルを基に、二つの仮説の精度の高低を判定できるような、ＮＮで表される。メインモデル１１０は、ＮＮを用いて、第１補助モデル１１１～第Ｍ補助モデル１１Ｍにおいてそれぞれ変換された学習用の二つの仮説の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事後確率を生成する。

【００８９】

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。

【００９０】

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、第１補助モデル１１１～第Ｍ補助モデル１１Ｍ及びメインモデル１１０に対し、各ＮＮが二つの仮説の精度の高低を判定するタスクを個別に行うとみなしたマルチタスク学習を行わせる。学習部２３は、各ＮＮによって実行された各タスクについて所定の損失をそれぞれ計算し、各損失の重み付け和を全体の損失関数とする。そして、学習部２３は、この全体の損失関数に基づいて、各ＮＮのパラメータの値を更新する。

【００９１】

なお、学習部２３は、各損失に対し、等重みで重み付けをしてもよい。また、メインモデル１１０が出力する判定情報が判定部１５における判定に使用されるため、学習部２３は、メインモデル１１０に、他の補助モデルよりも多めの重みを付けてもよい。

【００９２】

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、第１補助モデル１１１～第Ｍ補助モデル１１Ｍ及びメインモデル１１０に与える。これによって、学習部２３は、第１補助モデル１１１～第Ｍ補助モデル１１Ｍ及びメインモデル１１０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、第１補助モデル１１１～第Ｍ補助モデル１１Ｍ及びメインモデル１１０の学習（パラメータの最適化）を行う。

【００９３】

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第１補助モデル１１１～第Ｍ補助モデル１１Ｍに入力し、第１補助モデル１１１～第Ｍ補助モデル１１Ｍ及びメインモデル１１０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、第１補助モデル１１１～第Ｍ補助モデル１１Ｍ及びメ

インモデル１１０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。

【００９４】

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、第１補助モデル１１１～第Ｍ補助モデル１１Ｍ及びメインモデル１１０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、第１補助モデル１１１～第Ｍ補助モデル１１Ｍ及びメインモデル１１０に学習させる。

【００９５】

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、第１補助モデル１１１～第Ｍ補助モデル１１Ｍ及びメインモデル１１０の学習を行う。図７は、図６に示す入替部２３２の処理を説明する図である。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（図７の（１）参照）、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（図７の（２）参照）。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。

【００９６】

［学習処理の処理手順］

次に、図６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８は、実施の形態２に係る学習処理の処理手順を示すフローチャートである。図８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、$W^{(u)}$の精度は、$W^{(v)}$の精度よりも高いときの学習処理の処理手順を示す。

【００９７】

図８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$X^{(u)}$，$X^{(v)}$を第１補助モデル１１１～第Ｍ補助モデル１１Ｍに入力する（ステップＳ２２）。そして、学習部２３は、第１補助モデル１１１～第Ｍ補助モデル１１Ｍ及びメインモデル１１０にマルチタスク学習を行わせて、第１補助モデル１１１～第Ｍ補助モデル１１Ｍ及びメインモデル１１０のモデルパラメータを更新させる（ステップＳ２３）。

【００９８】

すなわち、この二つの仮説の$W^{(u)}$，$W^{(v)}$の特徴量ベクトル$X^{(u)}$，$X^{(v)}$を第１補助モデル１１１～第Ｍ補助モデル１１Ｍに入力した場合、第１補助モデル１１１～第Ｍ補助モデル１１Ｍ及びメインモデル１１０は、理想的には、$P(0 \mid X^{(u)}, X^{(v)})=1$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、第１補助モデル１１１～第Ｍ補助モデル１１Ｍ及びメインモデル１１０のモデルパラメータ（エンコーダＲＮＮ（ＬＳＴＭユニット）、１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。

【００９９】

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、入替部２３２は、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、$W^{(u)}$の精度は、$W^{(v)}$の精度よりも低い。よって、この二つの仮説$W^{(u)}$，$W^{(v)}$の特徴量ベクトル$X^{(u)}$，$X^{(v)}$を第１補助モデル１１１～第Ｍ補助モデル１１Ｍ及びメインモデル１１０に入力した場合、第１補助モデル１１１～第Ｍ補助モデル１１Ｍ及びメインモデル１１０は、理想的には，$P(0 \mid X^{(u)}, X^{(v)})=0$の事後確率を出力すべきである。

【０１００】

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$X^{(u)}$，$X^{(v)}$を第１補助モデル１１１～第Ｍ補

助モデル１１Ｍ及びメインモデル１１０に入力する（ステップＳ２６）。学習部２３は、以上の入力を基に、第１補助モデル１１１～第Ｍ補助モデル１１Ｍ及びメインモデル１１０にマルチタスク学習を行わせて、第１補助モデル１１１～第Ｍ補助モデル１１Ｍ及びメインモデル１１０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。

【０１０１】

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２参照）と同様に行うことができる。

【０１０２】

［実施の形態２の効果］

このように、実施の形態２に係る学習装置２０は、第１補助モデル１１１～第Ｍ補助モデル１１Ｍ及びメインモデル１１０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マルチタスク学習を行わせている。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第１補助モデル１１１～第Ｍ補助モデル１１Ｍ及びメインモデル１１０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第１補助モデル１１１～第Ｍ補助モデル１１Ｍ及びメインモデル１１０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、安定した制度でオラクル仮説を抽出することができる。

【０１０３】

［学習処理の効率化例１］

図８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、$E \times M \times N \times 2 \times {}_{N}C_{2}$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。

【０１０４】

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、$E \times M \times N \times 2 \times (N-1)$に削減することができる。

【０１０５】

［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。

【０１０６】

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。

【０１０７】

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。

【０１０８】

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。

【０１０９】

［評価］

実際に、実施の形態１におけるＮベストリランキングと、非特許文献１記載のＮベストリランキングとの比較評価を行った。表１は、ＣＳＪ音声コーパスを用いて、非特許文献１記載のＮベストリランキングとの比較評価する１００（＝Ｎ）ベストリランキング評価を行った結果を示す表である。表の数値は、ＷＥＲ（Word error rate）率［%］であり、Ｄｅｖ（Development）、Ｅｖａｌ（Evaluation）を示す。

【０１１０】

【表１】

(表1)

No.	Model	Dev	Eval
1	Single-encoder DDM	16.4	13.7
2	Eight-encoder DDM	16.1	13.4
3	2 with fwd & bwd LSTMLM scores	15.2	12.6
4	Oracle	11.6	9.7

【０１１１】

表１の通番「１」は、非特許文献１記載のＮベストリランキング結果である。表１の通番「２」は、実施の形態１に係るリランキング装置１０であって８個の補助モデルを有する場合の結果である。通番「３」は、通番「２」の条件に加え、前向き及び後ろ向きＬＳＴＭＬＭの単語予測スコアを１８時限目及び１９次元目の補助特徴として用いている。通番「４」は、参考のために示されたオラクルである。

【０１１２】

表１に示すように、通番「１」の非特許文献１記載のリランキング方法でも、十分にＷＥＲを削減できるが、通番「２」の８個の補助モデルを有するリランキング装置１０では、さらにＷＥＲ削減が実現できる。また、通番「３」の評価結果に示すように、両方向のＬＳＴＭＬＭの単語予測スコアと８個の補助モデルを用いることで、相補的なＷＥＲ削減効果が得られることが確認できた。また、通番「２」以外にも、補助モデルの個数を、２または４とした構成でも評価を行っており、補助モデルの個数を増やすほどＷＥＲを削減できる傾向が確認できた。

【０１１３】

この評価結果から、本実施の形態１に係るリランキング装置１０は、非特許文献１記載のリランキング方法と比して、安定したＷＥＲ削減が実現できる。

【０１１４】
［実施の形態３］
なお、実施の形態１のリランキング装置１０は、メインモデル１１０の出力を用いて判定を行ったが、メインモデル１１０の出力の他に各補助モデルの出力を用いて判定を行ってもよい。図９は、実施の形態３に係るリランキング装置の要部構成を示す図である。
【０１１５】
図３に示すように、リランキング装置は、判定部１５の前段に、重み付け部１８を有する。重み付け部１８は、第１補助モデル１１１～第Ｍ補助モデル１１Ｍ及びメインモデル１１０から出力された全ての判定情報を取得し、各判定情報に対して重み付け和を計算する。
【０１１６】
なお、各判定情報に対応する重みは予め設定されている。重み付け部１８は、各判定情報に対し、全補助モデル及びメインモデル１１０に対して等重みで重み付けをしてもよい。また、重み付け部１８は、メインモデル１１０に、他の補助モデルよりも多めの重みを付けてもよい。また、重み付け部１８は、予め各判定情報に対する重みを学習した１層の線形ＮＮを有し、各判定情報が入力されると各判定対象に対する重みを求めてもよい。
【０１１７】
判定部１５は、重み付け部１８が計算した重み付け和の値に基づいて二つの仮説の精度の高低を判定する。例えば、判定部１５は、判定情報のそれぞれが、仮説Ｗ（ｕ）が選択される確率を示すものとして、判定情報の重みづけ和を０～１の範囲に収まるように正規化した値が０．５以上であれば仮説Ｗ（ｕ）を選択し、そうでなければ仮説Ｗ（ｖ）を選択する。
【０１１８】
［判定処理の処理手順］
図１０は、実施の形態３に係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。
【０１１９】
図１０に示すステップＳ３１～ステップＳ３６は、図４に示すステップＳ１～ステップＳ６と同じ処理である。重み付け部１８は、第１補助モデル１１１～第Ｍ補助モデル１１Ｍ及びメインモデル１１０から出力された全ての判定情報を取得し、各判定情報に対して重み付け和を計算する重み付け処理を行う（ステップＳ３７）。そして、重み付け部１８が計算した重み付け和の値に基づいて二つの仮説の精度の高低を判定する。ステップＳ３８～ステップＳ４３は、図４に示すステップＳ７～ステップＳ１２と同じ処理である。
【０１２０】
［実施の形態３の効果］
この実施の形態３に示すように、メインモデル１１０による判定情報に加え、全補助モデルによる判定情報を用いて、判定を行うことも可能である。この際、実施の形態３では、各補助モデル或いはメインモデル１１０に応じて、各判定情報に対する重み付けを行い、重み付け和の値に基づいて二つの仮説の精度の高低を判定するため、オラクル仮説を抽出精度を保持することができる。
【０１２１】
なお、本実施の形態では、全ての仮説に対して、比較処理を行うため、Ｎベスト仮説のソートも可能である。
【０１２２】
また、本実施の形態１～３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１に例示する第１補助モデル１１１～第Ｍ補助モデル１１Ｍ及びメインモデル１１０について説明した。ただし、本実施の形態１～３の第１補助モデル１１１～第Ｍ補助モデル１１Ｍ及びメインモデル１１０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。
【０１２３】
このため、本実施の形態１～３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１～３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１～３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１～３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。
【０１２４】
また、この場合には、本実施の形態１～３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で表される第１補助モデル１１１～第Ｍ補助モデル１１Ｍ及びメインモデル１１０にマルチタスク学習を行わせる。そして、本実施の形態１～３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第１補助モデル１１１～第Ｍ補助モデル１１Ｍ及びメインモデル１１０に学習させる。そして、本実施の形態１～３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第１補助モデル１１１～第Ｍ補助モデル１１Ｍ及びメインモデル１１０に学習させる。本実施の形態１～３では、この第１補助モデル１１１～第Ｍ補助モデル１１Ｍ及びメインモデル１１０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。
【０１２５】
［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。
【０１２６】
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。
【０１２７】
［プログラム］
図１１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。

【０１２８】

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。

【０１２９】

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。

【０１３０】

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。

【０１３１】

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。

【０１３２】

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】

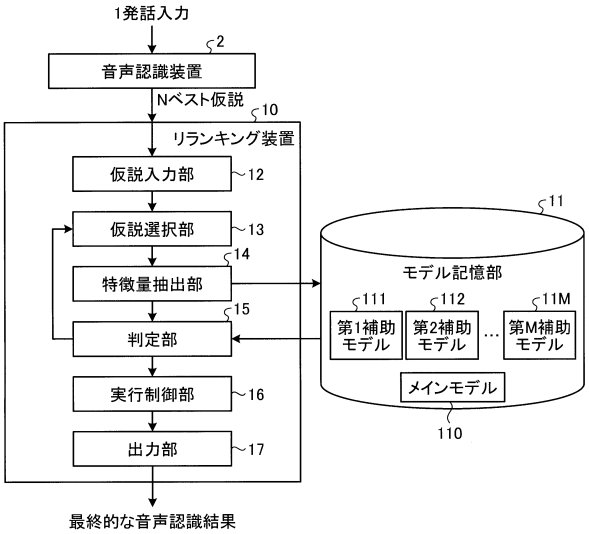
【０１３３】

２　音声認識装置
１０　リランキング装置
１１，２１　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

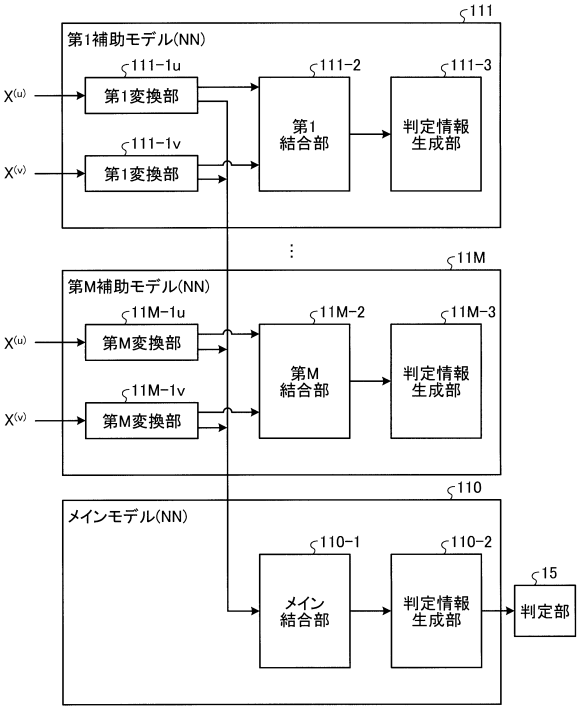
１６　実行制御部
１７　出力部
１８　重み付け部
２０　学習装置
２２　仮説入力部
２３　学習部
１１０　メインモデル
１１１～１１Ｍ　第１補助モデル～第Ｍ補助モデル
２３１　教師ラベル付与部
２３２　入替部

【図面】

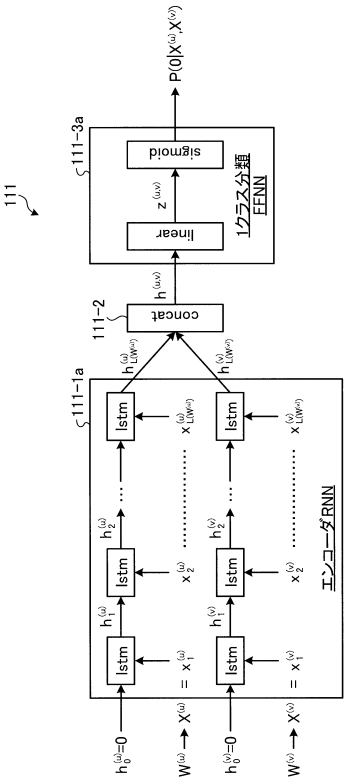
【図１】



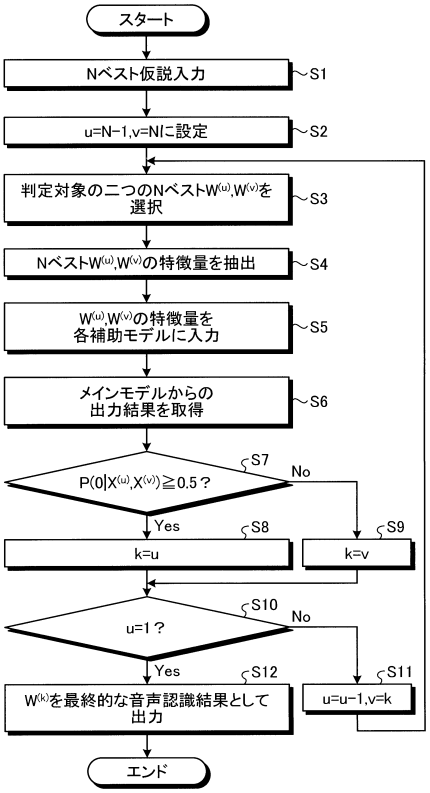
【図２】



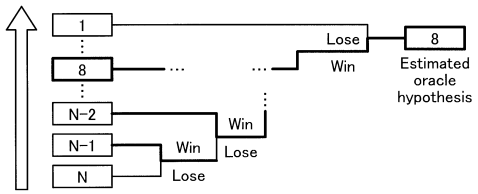
【図３】


111
111-3a
1クラス分類FFNN
linear
sigmoid
111-2
concat
111-1a
lstm
エンコーダRNN

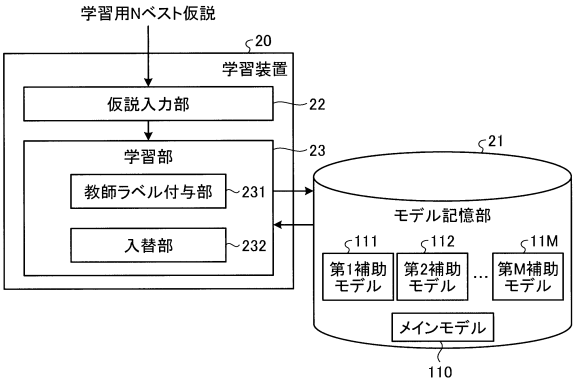
【図４】


スタート
Nベスト仮説入力
S1
u=N-1,v=N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)の特徴量を各補助モデルに入力
S5
メインモデルからの出力結果を取得
S6
P(0|X(u),X(v))≧0.5？
S7
No
Yes
k=u
S8
k=v
S9
u=1？
S10
No
Yes
W(k)を最終的な音声認識結果として出力
S12
u=u-1,v=k
S11
エンド

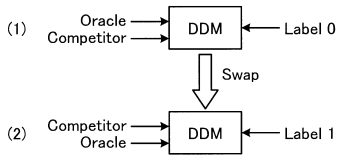
【図５】


1
8
N-2
N-1
N
Lose
Win
Estimated oracle hypothesis

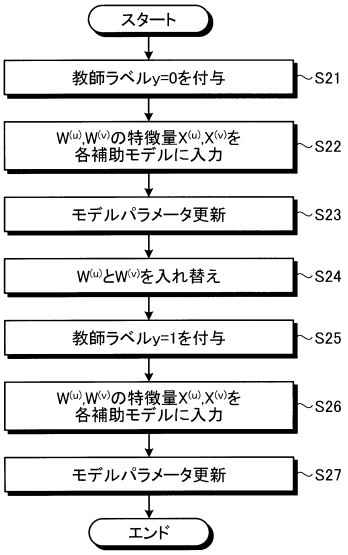
【図６】


学習用Nベスト仮説
20
学習装置
仮説入力部
22
学習部
23
教師ラベル付与部
231
入替部
232
21
モデル記憶部
111
112
11M
第1補助モデル
第2補助モデル
第M補助モデル
メインモデル
110

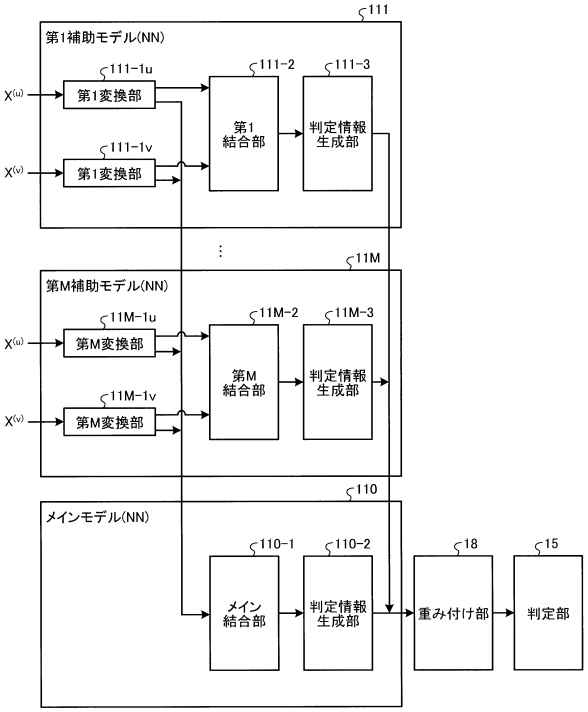
【図７】


(1)
Oracle
Competitor
DDM
Label 0
Swap
(2)
Competitor
Oracle
DDM
Label 1

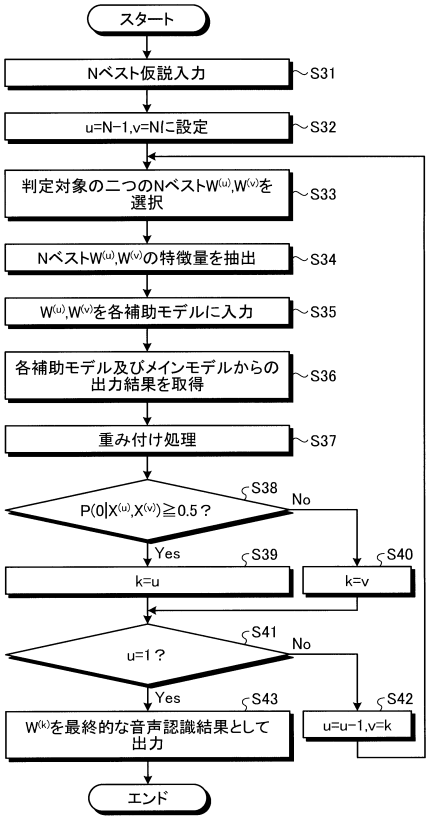
【図８】


スタート
教師ラベルy=0を付与
S21
W(ω),W(v)の特徴量X(ω),X(v)を各補助モデルに入力
S22
モデルパラメータ更新
S23
W(ω)とW(v)を入れ替え
S24
教師ラベルy=1を付与
S25
W(ω),W(v)の特徴量X(ω),X(v)を各補助モデルに入力
S26
モデルパラメータ更新
S27
エンド

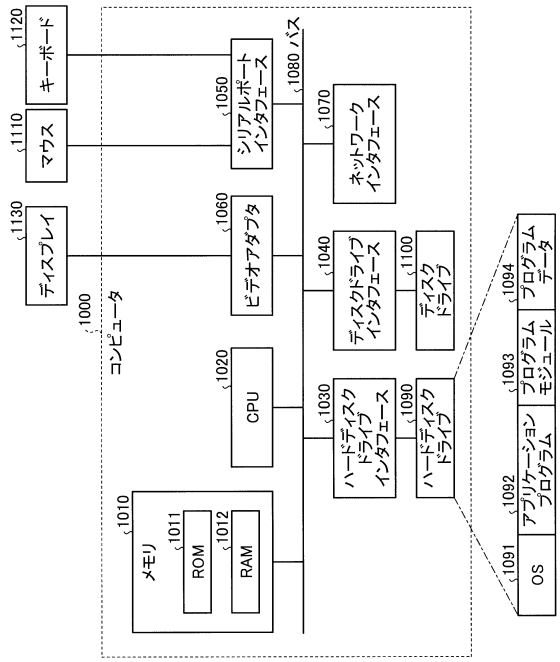
【図９】


111
第1補助モデル(NN)
X(ω)
X(v)
111-1u
第1変換部
111-1v
第1変換部
111-2
第1結合部
111-3
判定情報生成部
11M
第M補助モデル(NN)
11M-1u
第M変換部
11M-1v
第M変換部
11M-2
第M結合部
11M-3
判定情報生成部
110
メインモデル(NN)
110-1
メイン結合部
110-2
判定情報生成部
18
重み付け部
15
判定部

【図１０】


スタート
Nベスト仮説入力
S31
u=N−1,v=Nに設定
S32
判定対象の二つのNベストW(ω),W(v)を選択
S33
NベストW(ω),W(v)の特徴量を抽出
S34
W(ω),W(v)を各補助モデルに入力
S35
各補助モデル及びメインモデルからの出力結果を取得
S36
重み付け処理
S37
P(0|X(ω),X(v))≧0.5？
S38
No
Yes
k=u
S39
k=v
S40
u=1？
S41
No
Yes
W(k)を最終的な音声認識結果として出力
S43
u=u−1,v=k
S42
エンド

【図１１】


1000
コンピュータ
1010
メモリ
1011
ROM
1012
RAM
1020
CPU
1030
ハードディスク
ドライブ
インタフェース
1040
ディスクドライブ
インタフェース
1050
シリアルポート
インタフェース
1060
ビデオアダプタ
1070
ネットワーク
インタフェース
1080 バス
1090
ハードディスク
ドライブ
1091
OS
1092
アプリケーション
プログラム
1093
プログラム
モジュール
1094
プログラム
データ
1100
ディスク
ドライブ
1110
マウス
1120
キーボード
1130
ディスプレイ

フロントページの続き

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　山下 剛史

(56)参考文献　特開２０１１－２４３１４７（ＪＰ，Ａ）
特開２０１８－６００４７（ＪＰ，Ａ）
米国特許出願公開第２０１７／０２２１４７４（ＵＳ，Ａ１）
米国特許第１００３２４６３（ＵＳ，Ｂ１）
小川厚徳他，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，日本音響学会２０１８年春季研究発表会講演論文集［ＣＤ-ＲＯＭ］，2018年03月，pp.23-24
田中智大他，複数仮説を考慮したニューラル誤り訂正言語モデルの検討，電子情報通信学会技術研究報告，2018年08月，Vol.118，No.198，pp.31-36

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