(19)日本国特許庁（ＪＰ）　(12)特　許　公　報（Ｂ１）　(11)特許番号

第2938866号

(45)発行日　平成11年(1999)８月25日　(24)登録日　平成11年(1999)６月11日

(51)Int.Cl.[6]	識別記号	ＦＩ	
Ｇ１０Ｌ　3/00	５３７	Ｇ１０Ｌ　3/00	５３７Ｊ
			５３７Ｄ

請求項の数５(全 15 頁)

(21)出願番号　特願平10－243024

(22)出願日　平成10年(1998)８月28日

審査請求日　平成10年(1998)８月28日

特許法第30条第１項適用申請有り　日本音響学会平成10年度春季研究発表会講演論文集（平成10年３月17日）41－42頁に発表

特許法第30条第１項適用申請有り　平成10年３月17日に慶応義塾大学において開催された日本音響学会平成10年度春季研究発表会において発表

(73)特許権者　593118597
株式会社エイ・ティ・アール音声翻訳通信研究所
京都府相楽郡精華町大字乾谷小字三平谷５番地

(72)発明者　政瀧　浩和
京都府相楽郡精華町大字乾谷小字三平谷５番地　株式会社エイ・ティ・アール音声翻訳通信研究所内

(72)発明者　匂坂　芳典
京都府相楽郡精華町大字乾谷小字三平谷５番地　株式会社エイ・ティ・アール音声翻訳通信研究所内

(74)代理人　弁理士　青山　葆　(外２名)

審査官　涌井　智則

最終頁に続く

(54)【発明の名称】　統計的言語モデル生成装置及び音声認識装置

(57)【特許請求の範囲】

【請求項１】　複数のクラスタの統計的言語モデルを記憶する記憶手段と、

所定の複数の発声音声文を含む学習用テキストデータに基づいて各発声音声文に対する統計的言語モデルを生成して、各発声音声文が各クラスタに対応するように、上記生成した統計的言語モデルを上記記憶手段に記憶する初期化手段と、

上記学習用テキストデータの各発声音声文について、各クラスタにおける統計的言語モデルの文生成確率を計算して最大の文生成確率を有するクラスタを選択してその発声音声文を所属させるように上記記憶手段に記憶するクラスタ選択手段と、

発声音声文が属するクラスタが変化したときに、各クラスタ毎に、上記クラスタ選択手段によって選択された発

声音声文を用いて上記記憶手段に記憶された各統計的言語モデルを更新して、各クラスタに対応した統計的言語モデルを生成するモデル変更手段と、

上記複数の発声音声文に属するクラスタが１文も変化しなくなるまで、上記クラスタ選択手段の処理と、上記モデル変更手段の処理を繰り返す制御手段とを備えたことを特徴とする統計的言語モデル生成装置。

【請求項２】　請求項１記載の統計的言語モデル生成装置において、上記統計的言語モデル生成装置はさらに、

上記記憶手段に記憶された各クラスタ毎のテキストデータに基づいて、最尤推定法を用いて各クラスタ毎に単語のＮ－ｇｒａｍ（Ｎは２以上の自然数である。）の遷移確率を演算する第１の演算手段と、

上記第１の演算手段によって演算された各クラスタ毎の単語のＮ－ｇｒａｍの遷移確率の出現分布を事前知識の

所定の確率分布と仮定し、各クラスタ毎の確率分布の加重平均及び加重分散を演算した後、演算された加重平均と加重分散に基づいて事前知識の確率分布のパラメータを演算する第２の演算手段と、

上記第２の演算手段によって演算された事前知識の確率分布のパラメータと、上記学習用テキストデータうちの特定クラスタのテキストデータの事後知識における処理対象の単語列の直前の単語列の出現回数と、処理対象の単語列の出現回数とに基づいて、各クラスタ毎の単語のＮ－ｇｒａｍの遷移確率を計算することにより、各クラスタ毎の単語のＮ－ｇｒａｍの遷移確率を含む統計的言語モデルを生成する第３の演算手段とを備えたことを特徴とする統計的言語モデル生成装置。

【請求項３】　請求項２記載の統計的言語モデル生成装置において、上記統計的言語モデル生成装置はさらに、上記第３の演算手段によって演算された各クラスタ毎の単語のＮ－ｇｒａｍの遷移確率に基づいて，所定の平滑化処理を実行し、処理後の各クラスタ毎の単語のＮ－ｇｒａｍの遷移確率を含む統計的言語モデルを生成する第１の生成手段を備えたことを特徴とする統計的言語モデル生成装置。

【請求項４】　請求項３記載の統計的言語モデル生成装置において、上記統計的言語モデル生成装置はさらに、上記学習用テキストデータに基づいて、最尤推定法を用いて単語のＮ－ｇｒａｍ（Ｎは２以上の自然数である。）の遷移確率を演算して、上記単語のＮ－ｇｒａｍの遷移確率を含む別の統計的言語モデルを生成する第２の生成手段を備えたことを特徴とする統計的言語モデル生成装置。

【請求項５】　入力される発声音声文の音声信号に基づいて、所定の統計的言語モデルを用いて音声認識する音声認識装置において、

請求項４記載の統計的言語モデル生成装置と、

上記第２の生成手段によって生成された別の統計的言語モデルを用いて、入力される発声音声文の音声信号を音声認識して第１の認識仮説を出力する第１の音声認識手段と、

上記第１の音声認識手段から出力される第１の認識仮説に応答して、上記第１の生成手段によって生成された各クラスタ毎の統計的言語モデルを用いて、入力される発声音声文の音声信号を音声認識して、文生成確率が最大のクラスタの統計的言語モデル生成装置を選択するモデル選択手段と、

上記モデル選択手段によって選択されたクラスタの統計的言語モデルを用いて、上記第１の音声認識手段から出力される第１の認識仮説に対して絞込処理を行って第２の認識仮説を生成して認識結果として出力する第２の音声認識手段とを備えたことを特徴とする音声認識装置。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、学習用テキストデータに基づいて統計的言語モデルを生成する統計的言語モデル生成装置、及び上記統計的言語モデルを用いて、入力される発声音声文の音声信号を音声認識する音声認識装置に関する。

【０００２】

【従来の技術】近年、連続音声認識装置において、その性能を高めるために言語モデルを用いる方法が研究されている。これは、言語モデルを用いて、次単語を予測し探索空間を削減することにより、認識率の向上及び計算時間の削減の効果を狙ったものである。最近盛んに用いられている言語モデルとしてＮ－ｇｒａｍ（Ｎ－グラム；ここで、Ｎは２以上の自然数である。）がある。これは、大規模なテキストデータを学習し、直前のＮ－１個の単語から次の単語への遷移確率を統計的に与えるものである。複数Ｌ個の単語列$w_1^L = w_1, w_2, \cdots, w_L$の生成確率Ｐ（$w_1^L$）は次式で表される。

【０００３】

【数１】

$$P(w_1^L) = \prod_{t=1}^{L} P(w_t \mid w_{t+1-N}^{t-1})$$

【０００４】ここで、w_tは単語列w_1^Lのうちｔ番目の１つの単語を表し、w_i^jはｉ番目からｊ番目の単語列を表わす。上記数１において、確率Ｐ（$w_t \mid w_{t+1-N}^{t-1}$）は、Ｎ個の単語からなる単語列w_{t+1-N}^{t-1}が発声された後に単語w_tが発声される確率であり、以下同様に、確率Ｐ（Ａ｜Ｂ）は単語又は単語列Ｂが発声された後に単語Ａが発声される確率を意味する。また、数１における「Π」はｔ＝１からＬまでの確率Ｐ（$w_t \mid w_{t+1-N}^{t-1}$）の積を意味し、以下同様である。

【０００５】ところで、近年、上記統計的言語モデルＮ－ｇｒａｍを用いて連続音声認識の性能を向上させる手法が盛んに提案されている（例えば、従来技術文献１「Ｌ．Ｒ．Ｂａｈｌ　ｅｔ　ａｌ．，“Ａ Ｍａｘｉｍｕｍ ＬｉｋｅｌｉｈｏｏｄＡｐｐｒｏａｃｈ ｔｏ Ｃｏｎｔｉｎｕｏｕｓ Ｓｐｅｅｃｈ Ｒｅｃｏｇｎｉｔｉｏｎ”，ＩＥＥＥ Ｔｒａｎｓａｃｔｉｏｎｓ ｏｎ ＰａｔｔｅｒｎＡｎａｌｙｓｉｓ ａｎｄ Ｍａｃｈｉｎｅ Ｉｎｔｅｌｌｉｇｅｎｃｅ，ｐｐ．１７９－１９０，１９８３年」及び従来技術文献２「清水ほか，“単語グラフを用いた自由発話音声認識”，電子情報通信学会技術報告，ＳＰ９５－８８，ｐｐ．４９－５４，平成７年」参照。）。

【０００６】しかしながら、Ｎ－ｇｒａｍはパラメータ数が多く、それぞれの値を正確に求めるためには、莫大な量のテキストデータが必要とされる。この問題を解決する方法として、学習用テキストデータに出現しない単語遷移に対しても遷移確率を与える平滑化の手法（例えば、従来技術文献３「Ｆ．Ｊｅｌｉｎｅｋ　ｅｔ　ａ

l.,“Interpolated estimation of Markov Source Parameters from Sparse Data”, Proceedings of Workshop Pattern Recognition in Practice, pp. 381−387, 1980年」、従来技術文献4「S. M. Katz et al.,“Estimation of Probabilities from Sparse Data for the Language model Component of a Speech Recognizer”, IEEE Transactions on Acoustics, Speech, and SignalProcessing, pp. 400−401, 1987年」及び従来技術文献5「川端ほか,“二項事後分布に基づくN−gram言語モデルのBack−off平滑化”, 電子情報通信学会技術報告、SP95−93, pp1−6, 平成7年」参照。）や、クラス分類、可変長N−gram等パラメータの数を減少させる手法（例えば、従来技術文献6「P. F. Brown et al.,“Class−Based n−gram models of naturallanguage”, Computational Linguistics, Vol. 18, No. 4, pp. 467−479, 1992年」、従来技術文献7「T. R. Niesler et al.,“A Variable−Length Category−Based N−gram Language Model”, Proceedings of ICASSP ’96, Vol. 1, pp. 164−167, 1996年」及び従来技術文献8「政瀧ほか,“連続音声認識のための可変長連鎖統計言語モデル”, 電子情報通信学会技術報告, SP95−73, pp. 1−6, 平成7年」参照。）等が数多く提案されている。しかしながら、これらの手法を用いても、精度の良い言語モデルを構築するためには、相当量のデータを用いる必要があると考えられる。

【0007】現在、実用化に向けて研究が行われている連続音声認識システムは、ホテル予約・スケジューリング等、タスクを限定し、システムのパフォーマンスの向上させている物が多い。しかしながら、タスク毎に大量の言語データを集めるのは困難である。特に、日本語の場合は、英語等のように単語の区切りが明確ではなく、通常人間が手作業で単語の切り出し・形態素解析の作業を行うため、大量のデータを集めるのはさらに困難である。しかしながら、データ量を増やすために、他のタスクのデータを用いるた場合、言語的特徴はタスク毎に異なるため、単純にデータを混合しても目的のタスク特有の言語特徴を効果的に表現することはできないと考えられる。

【0008】これらの問題を解決する手段として、言語モデルのタスク適応を考えられている。すなわち、目的のタスク以外のデータも含めた大量のデータを学習することによりデータ量の問題を解決し、得られたモデルの言語特徴を目的のタスクに適応させる方法である。タスク適応の手法として、従来提案されているものには大量のデータで作成したN−gramと、目的タスクのデータで作成したN−gramとを重みづけにより混合する方法（例えば、従来技術文献9「伊藤ほか,“小量テキストによる言語モデルのタスク適応の検討”, 日本音響学会講演論文集, 1−3−19, pp. 37−38, 平成8年9月」及び従来技術文献10,「坂本ほか,“音声認識における統計的言語モデルの選択の効果”, 日本音響学会講演論文集, 1−Q−24, pp. 157−158, 平成6年3月」参照。）がある。

【0009】例えば、従来技術文献9の手法を用いた第1の従来例のタスク適応化処理は、小量テキストに重みをかけて大量テキストと混合するものであり、次の手順によって言語モデルを作成する。

（a）小量テキストを、重み付きで大量テキストに加える。重み係数をwとすると、小量テキスト中でm回出現した単語は、大量テキスト中で（w・m）回出現した単語と同等に扱われることになる。

（b）このようにしてできたテキストの中で、出現頻度が一定回数未満の単語を、未知語を表す記号に置き換える。すなわち、語彙の制限を行う。

（c）テキストから統計を取り, N−gramモデルを構築する。

【0010】しかしながら、第1の従来例のタスク適応化処理においては、重み係数wは1つのみしか使用していないので、言語モデルにおける遷移確率の予測精度はいまだ比較的低く、当該言語モデルを用いて音声認識をしたときの音声認識率は比較的低いという問題点があった。

【0011】この第1の従来例の問題点を解決するために、本発明者は、特開平10−198395号の特許出願（以下、第2の従来例という。）において、特定のタスクのN−gram言語モデルの精度を向上させるため、MAP推定（最大事後確率推定）によるタスク適応の手法を提案している。

【0012】

【発明が解決しようとする課題】しかしながら、第2の従来例の手法は単語列毎にタスク間のN−gram遷移確率の平均、及び分散を用いてパラメータ推定を行うため、テキスト全体があらかじめ複数のタスクに分割されている必要があり、単一のタスクのデータには適用できないという問題点があった。また、データ量が比較的多いタスクにおいては、タスク適応の効果が小さいという問題があった。

【0013】本発明の目的は以上の問題点を解決し、各タスクに対する適応効果が大きく、従来例に比較して遷

移確率の予測精度及び信頼性を改善することができる統計的言語モデルを生成することができる統計的言語モデル生成装置、及び、当該統計的言語モデルを用いて従来例に比較して高い音声認識率で音声認識することができる音声認識装置を提供することにある。

【００１４】

【課題を解決するための手段】本発明に係る請求項１記載の統計的言語モデル生成装置は、複数のクラスタの統計的言語モデルを記憶する記憶手段と、所定の複数の発声音声文を含む学習用テキストデータに基づいて各発声音声文に対する統計的言語モデルを生成して、各発声音声文が各クラスタに対応するように、上記生成した統計的言語モデルを上記記憶手段に記憶する初期化手段と、上記学習用テキストデータの各発声音声文について、各クラスタにおける統計的言語モデルの文生成確率を計算して最大の文生成確率を有するクラスタを選択してその発声音声文を所属させるように上記記憶手段に記憶するクラスタ選択手段と、発声音声文が属するクラスタが変化したときに、各クラスタ毎に、上記クラスタ選択手段によって選択された発声音声文を用いて上記記憶手段に記憶された各統計的言語モデルを更新して、各クラスタに対応した統計的言語モデルを生成するモデル変更手段と、上記複数の発声音声文に属するクラスタが１文も変化しなくなるまで、上記クラスタ選択手段の処理と、上記モデル変更手段の処理を繰り返す制御手段とを備えたことを特徴とする。

【００１５】また、請求項２記載の統計的言語モデル生成装置は、請求項１記載の統計的言語モデル生成装置において、さらに、上記記憶手段に記憶された各クラスタ毎のテキストデータに基づいて、最尤推定法を用いて各クラスタ毎に単語のＮ－ｇｒａｍ（Ｎは２以上の自然数である。）の遷移確率を演算する第１の演算手段と、上記第１の演算手段によって演算された各クラスタ毎の単語のＮ－ｇｒａｍの遷移確率の出現分布を事前知識の所定の確率分布と仮定し、各クラスタ毎の確率分布の加重平均及び加重分散を演算した後、演算された加重平均と加重分散に基づいて事前知識の確率分布のパラメータを演算する第２の演算手段と、上記第２の演算手段によって演算された事前知識の確率分布のパラメータと、上記学習用テキストデータうちの特定クラスタのテキストデータの事後知識における処理対象の単語列の直前の単語列の出現回数と、処理対象の単語列の出現回数とに基づいて、各クラスタ毎の単語のＮ－ｇｒａｍの遷移確率を計算することにより、各クラスタ毎の単語のＮ－ｇｒａｍの遷移確率を含む統計的言語モデルを生成する第３の演算手段とを備えたことを特徴とする。

【００１６】さらに、請求項３記載の統計的言語モデル生成装置は、請求項２記載の統計的言語モデル生成装置において、さらに、上記第３の演算手段によって演算された各クラスタ毎の単語のＮ－ｇｒａｍの遷移確率に基づいて，所定の平滑化処理を実行し、処理後の各クラスタ毎の単語のＮ－ｇｒａｍの遷移確率を含む統計的言語モデルを生成する第１の生成手段を備えたことを特徴とする。

【００１７】またさらに、請求項４記載の統計的言語モデル生成装置は、請求項３記載の統計的言語モデル生成装置において、さらに、上記学習用テキストデータに基づいて、最尤推定法を用いて単語のＮ－ｇｒａｍ（Ｎは２以上の自然数である。）の遷移確率を演算して、上記単語のＮ－ｇｒａｍの遷移確率を含む別の統計的言語モデルを生成する第２の生成手段を備えたことを特徴とする。

【００１８】本発明に係る請求項５記載の音声認識装置は、入力される発声音声文の音声信号に基づいて、所定の統計的言語モデルを用いて音声認識する音声認識装置において、請求項４記載の統計的言語モデル生成装置と、上記第２の生成手段によって生成された別の統計的言語モデルを用いて、入力される発声音声文の音声信号を音声認識して第１の認識仮説を出力する第１の音声認識手段と、上記第１の音声認識手段から出力される第１の認識仮説に応答して、上記第１の生成手段によって生成された各クラスタ毎の統計的言語モデルを用いて、入力される発声音声文の音声信号を音声認識して、文生成確率が最大のクラスタの統計的言語モデル生成装置を選択するモデル選択手段と、上記モデル選択手段によって選択されたクラスタの統計的言語モデルを用いて、上記第１の音声認識手段から出力される第１の認識仮説に対して絞込処理を行って第２の認識仮説を生成して認識結果として出力する第２の音声認識手段とを備えたことを特徴とする。

【００１９】

【発明の実施の形態】以下、図面を参照して本発明に係る実施形態について説明する。

【００２０】図１に本発明に係る一実施形態の連続音声認識装置のブロック図を示す。本実施形態の連続音声認識装置は、図１において、特に、クラスタリング処理部４０と、２つの言語モデル生成部４１，４２と、２つの単語仮説絞込部６ａ，６ｂを備えたことを特徴としている。本実施形態では、上述した第２の従来例の問題点を解決するためには、同一タスクの文でも、さまざまな内容の文が存在し、タスクという尺度よりも、文毎の内容で分類した方が言語的特徴がより明確になる考えられる。本実施形態では、これらの問題を解決し、さらに精度の高いＮ－ｇｒａｍ言語モデルを得るため、学習用テキストデータメモリ２０内のテキストデータのコーパス全体をクラスタリング処理部４０によって自動的にクラスタリングし、言語モデル生成部４２において、ＭＡＰ推定を用いてそれぞれのクラスタ毎にＮ－ｇｒａｍ言語モデルを構築する。また、精度を向上させるため、従来の単語Ｎ－ｇｒａｍに代り、可変長Ｎ－ｇｒａｍ（品詞

と可変長単語列の複合Ｎ－ｇｒａｍ）を用いる。

【００２１】すなわち、本実施形態の連続音声認識装置は、言語モデルの学習に用いるテキストコーパスをクラスタリングし、クラスタ毎の言語特徴を明確にさせ、言語モデルの精度を向上させる。しかしながら、入力された音声の発話文が属するクラスタをあらかじめ知ることは不可能である。このため、まず最初に、入力音声をコーパス全体で作成した言語モデルで認識を行い、次に、認識結果からクラスタ別の言語モデルを１つのみ選択し再度認識を行う、という２段階の認識を行う。

【００２２】中間認識結果Ｗからクラスタ別の言語モデルＬＭｃの選択は、次式のように確率が最も高いものを選択することにより行う。

【００２３】

【数２】

$$LMc = \underset{LMk}{\mathrm{argmax}} P(LMk \mid W)$$

【００２４】上記式は、Ｂａｙｅｓ則を用いると、次式のように表される。

【００２５】

【数３】

$$LMc = \underset{LMk}{\mathrm{argmax}} P(W \mid LMk)\, P(LMk)$$

【００２６】ここで、Ｐ（ＬＭｋ）は、言語モデルＬＭｋの事前確率である。この確率は前発話の内容等より求めることができるが、本実施形態ではこの確率に関しては考慮しない。従って、次の式によりクラスタ言語モデルを選択する。

【００２７】

【数４】

$$LMc = \underset{LMk}{\mathrm{argmax}} P(W \mid LMk)$$

【００２８】すなわち、Ｋ個のクラスタのそれぞれのモデルＬＭｋで認識結果文Ｗに対する生成確率Ｐ（Ｗ｜ＬＭｋ）（１≦ｋ≦Ｋ）を求め、確率の最も高いクラスタモデルＬＭｃを選択する。この選択処理は、言語モデル選択部８により行われる。

【００２９】次いで、クラスタリング処理部４０によるコーパスのクラスタリングについて説明する。コーパスを自動クラスタリングするために、公知のＫ－ｍｅａｎｓ法に類似した方法を用いた。Ｋ－ｍｅａｎｓ法は、サンプルを距離が最も近いクラスタ中心に所属させる形でクラスタリングを行う手法である。この手法を文のクラスタリングに適用するため、次の２点で通常の方法と異なる。

（１）クラスタ中心ベクトルをそのクラスタに属する文で生成される言語モデルとする。

（２）距離尺度に文の生成確率Ｐ（Ｗ｜ＬＭｋ）を用いる。

【００３０】これらは、前述した認識結果からクラスタモデルの選択で用いる手法と同一であり、妥当な方法であると考える。以下に、クラスタリングの手順を示す。

＜ステップＳＳ１＞クラスタモデルの初期化：クラスタ数をＫとし、適当な手法によりコーパスからＫ文を選択して全クラスタに１文ずつ配置し、クラスタ毎の言語モデルＬＭ１、ＬＭ２、…、ＬＭＫを作成する。

＜ステップＳＳ２＞クラスタの選択：コーパスの全文について、各クラスタにおける言語モデルの文生成確率を求め、最も確率の高いクラスタを選択し、その文を所属させる。

＜ステップＳＳ３＞クラスタモデルの変更：各クラスタ毎に、ステップＳＳ２で選択した文を用いて言語モデルＬＭ１、ＬＭ２、…、ＬＭＫを更新する。

＜ステップＳＳ４＞終了条件：文の属するクラスタが１文も変化しない場合、クラスタリングを終了する。それ以外の場合は、ステップＳＳ２及びＳＳ３の処理を繰り返す。ただし、ある程度の回数を繰り返してもクラスタリングが収束しない場合は強制終了させる。

【００３１】次いで、ＭＡＰ推定によるＮ－ｇｒａｍの適応について説明する。クラスタリングを行うことにより、クラスタ毎の言語的特徴は明確になるものの、クラスタ毎のデータ量は減少するため、Ｎ－ｇｒａｍのパラメータ推定の精度が低下することが考えられる。このため、第２の従来例で開示されたＭＡＰ推定を用いた適応の手法を用い、パラメータ推定の精度を向上させる。

【００３２】まず、ＭＡＰ推定法による遷移確率の算出について述べる。通常、Ｎ－ｇｒａｍの遷移確率は、ＭＬ（Ｍａｘｉｍｕｍ　Ｌｉｋｅｌｉｈｏｏｄ；最尤）推定法により算出される。最尤推定法は、観測したサンプル値（すなわち、テキストデータの単語）ｘに対して、遷移確率ｐが決まったときに単語ｘとなる尤度関数ｆ（ｘ｜ｐ）を最大にさせる値として、Ｎ－ｇｒａｍ遷移確率p_{ML}が次式で定められる。

【００３３】

【数５】

$$p_{ML} = \underset{p}{\mathrm{argmax}}\ f(x \mid p)$$

【００３４】ここで、関数ａｒｇｍａｘは、ｐを変化したときに関数の引数が最大となるときのｐの値を示す関数である。議論を簡単にするため、単語w_kからw_lへのｂｉｇｒａｍの遷移確率ｐ（w_l｜w_k）について考え、また、以下のような変数の定義を行う。

（ａ）Ｎ：学習用テキストデータ中の単語w_kの出現回数。

（ｂ）ｎ：学習用テキストデータ中の単語列w_k w_lの出現回数。

（ｃ）ｐ：単語w_kからw_lへの遷移確率。

【００３５】このとき尤度関数ｆ（ｘ｜ｐ）は、単語w_kがＮ回観測され、次に単語w_lに続く回数がｎ回で、それ以外の単語に続く回数が（Ｎ－ｎ）回であるから、次

式を得ることができる。

【００３６】

【数６】$f(p)=p^{n}(1-p)^{N-n}$

【００３７】ｆ（ｐ）の最大化条件ｄｌｏｇｆ（ｐ）／ｄｐ＝０を解くことにより、Ｎ－ｇｒａｍの遷移確率は次式のように計算される。

【００３８】

【数７】$p_{ML}=n/N$

【００３９】従って、もし単語列$w_k\ w_1$が観測データ上で出現しない場合、ｎ＝０であるから、遷移確率は０と推定されてしまう。これに対して、ＭＡＰ（ＭａｘｉｍｕｍＡ－ｐｏｓｔｅｒｉｏｒｉ　Ｐｒｏｂａｂｉｌｉｔｙ；最大事後確率）推定においては、最尤推定法を用いて、観測したサンプル値ｘに対して、遷移確率ｐが決定される事後確率関数ｈ（ｐ｜ｘ）を最大化する値として、Ｎ－ｇｒａｍの遷移確率が次式で求められる。

【００４０】

【数８】

$$p_{MAP}=\underset{p}{\mathrm{argmax}}\ h(p\mid x)$$

【００４１】ここで、Ｂａｙｅｓ則を用いると、上記数８は次式のように変形される。

【００４２】

【数９】

$$p_{MAP}=\underset{p}{\mathrm{argmax}}\ f(x\mid p)\,g(p)$$

【００４３】ここで、ｇ（ｐ）は、各クラスタのテキストデータに基づいて予め決定される、Ｎ－ｇｒａｍの遷移確率ｐの事前分布である。すなわち、ＭＡＰ推定法を用いると、Ｎ－ｇｒａｍの遷移確率はある事前知識より＊得られる分布ｇ（ｐ）に従う変数とし、この事前分布と実際に観測されたサンプル値とを用いて、実際の遷移確率が推定される。このため、観測データで出現しない単語遷移に対しても、事前知識により０でない遷移確率を与えることができる。

【００４４】次いで、ｂｉｇｒａｍを例にとり、ＭＡＰ推定法によりＮ－ｇｒａｍの遷移確率を求める方法を示す。ただし、変数の定義は上述と同じものを用いる。まず、遷移確率ｐの事前分布としてベータ分布（$ap^{\alpha-1}(1-p)^{\beta-1}$，ここで、α及びβはベータ分布の正のパラメータであり、ａは正規化のための正の定数である。）を用いる。なお、$0\leqq p\leqq 1$である。ベータ分布を用いる理由は次の２点である。

（ａ）ベータ分布は２項分布の自然共役事前分布で、ＭＡＰ推定によるパラメータの解が求まりやすい。

（ｂ）ベータ分布のパラメータα，βを変化させることにより、様々な形状の分布を表すことができる。

ここで、ベータ分布は、連続変数の確率分布の一種であり、ガンマ関数をもとにして構成されるベータ関数が表示に含まれる。なお、本実施形態においては、ベータ分布を用いるが、本発明はこれに限らず、ベータ分布に代えて、ディリクレ分布を用いてもよい。ディリクレ分布は、ベータ分布を多変量分布に拡張したものであり、多項分布の自然共役事前分布である。

【００４５】上記数９のＭＡＰ推定法の定義に従うと、遷移確率p_{MAP}は、尤度関数ｆ（ｐ）と事前分布ｇ（ｐ）とを用いて次式のように求められる。

【００４６】

【数１０】

$$\begin{aligned}p_{MAP}&=\underset{p}{\mathrm{argmax}}\ \{p^{n}(1-p)^{N-n}\,ap^{\alpha-1}(1-p)^{\beta-1}\}\\&\equiv\underset{p}{\mathrm{argmax}}\ L(p)\end{aligned}$$

【００４７】ここで、関数Ｌ（ｐ）が最大となるための条件ｄｌｏｇＬ（ｐ）／ｄｐ＝０をｐについて解くと、単語のｂｉｇｒａｍの遷移確率p_{MAP}は次式のように求まる。

【００４８】

【数１１】$p_{MAP}=(n+\alpha-1)/(N+\alpha+\beta-2)$

【００４９】ここで、パラメータα及びβは、事前分布であるベータ分布のパラメータであるが、これらは、次式のように求めることができる。なお、ベータ分布の平均μ及び分散σ^2は以下の式となることが知られている（例えば、従来技術文献５参照。）。

【００５０】

【数１２】$\mu=\alpha/(\alpha+\beta)$

$\sigma^2=(\alpha\beta)/\{(\alpha+\beta)^2(\alpha+\beta+1)\}$

【００５１】これらの式をα，α＋βについて解くと、次式が得られる。

【００５２】

【数１３】$\alpha=\{\mu^2(1-\mu)\}/\sigma^2-\mu$

$\alpha+\beta=\{\mu(1-\mu)\}/\sigma^2-1$

【００５３】以上より、観測テキストデータから頻度を計算することにより得られるパラメータＮ，ｎ、及び事前分布の平均μ及び分散σ^2により、上記数１１及び数１３を用いて、単語のｂｉｇｒａｍの遷移確率を求めることができる。

【００５４】これまでの議論は、単語のｂｉｇｒａｍについてのみの議論であったが、一般に、ＭＡＰ推定法によるＮ－ｇｒａｍの遷移確率$p(w_n\mid w_1^{n-1})$は、直前の単語w_kを直前の単語列w_1^{n-1}と置き換え、パラメータＮ及びｎを次のように定義すれば、同じ議論が通用

することは明らかである。

（ａ）Ｎ：学習用テキストデータ中の単語列w_1^{n-1}の出現回数（ｃ（w_1^{n-1}））、すなわち、処理対象の単語列の直前の単語列の出現回数である。

（ｂ）ｎ：学習用テキストデータ中の単語列w_1^{n}の出現回数（ｃ（w_1^{n}））、すなわち、処理対象の単語列の出現回数である。

【００５５】次いで、ＭＡＰ推定法を用いたクラスタ適応化処理について述べる。上述のＭＡＰ推定法によるＮ－ｇｒａｍをクラスタ適応化に応用するために、図３に示すように、複数のクラスタより構成される大量のテキストデータから構成される不特定のタスクのテキストデータに基づくＮ－ｇｒａｍを事前知識とし、目的の特定クラスタｉのテキストデータ２１－ｉを事後知識とみな＊す。不特定のクラスタのＮ－ｇｒａｍを事前知識とみなしたとき、その事前分布は、各クラスタにおけるＮ－ｇｒａｍ遷移確率の分布と考えることができる。ただし、各クラスタにおけるＮ－ｇｒａｍ遷移確率は最尤推定法により求められる。この事前分布をベータ分布と仮定してＭＡＰ推定法の事前分布として用いる。このとき、事前分布の加重平均μ、及び加重分散σ^2は次式で求められる。

【００５６】

【数１４】

$$\mu=\sum_i c_i(w_n)\,p_i(w_n\mid w_1^{n-1})\ /\ \sum_i c_i(w_1^{n-1})$$

【数１５】

$$\sigma^2=\sum_i c_i(w_1^{n-1})\,p_i(w_n\mid w_1^{n-1})^2\ /\ \sum_i c_i(w_1^{n-1})-\mu^2$$

【００５７】ここで、c_i（w_1^{n-1}）はクラスタｉにおいて単語列w_1^{n-1}の出現頻度であり、p_i（w_n｜w_1^{n-1}）はクラスタｉにおける単語列w_1^{n-1}からw_nへの遷移確率である。また、事後知識を目的のクラスタのテキストデータとすると、前述のパラメータＮ及びｎは次のように表される。

（ａ）Ｎ：目的の特定クラスタｉのテキストデータ２１－ｉ中の単語列w_1^{n-1}の出現頻度、すなわち、処理対象の単語列の直前の単語列の出現回数である。

（ｂ）ｎ：目的の特定クラスタｉのテキストデータ２１－ｉ中の単語列w_1^{n}の出現頻度、すなわち、処理対象の単語列の出現回数である。

以上の加重平均μ，加重分散σ^2，パラメータｎ及びＮを上述の数１０及び数１２に代入することにより、ＭＡＰ推定法によるタスク適応後のＮ－ｇｒａｍ遷移確率が得られる。

【００５８】さらに、Ｂａｃｋ－ｏｆｆ平滑化法による遷移確率の平滑化について述べる。上記でＭＡＰ推定法によるタスク適応の基本原理を述べたが、実際に言語モデルとして使用するには、２つの問題がある。１つは、平滑化の問題である。不特定タスクの大量のテキストデータを用いても、出現しない単語列が存在し、ＭＡＰ推定法を用いても、Ｎ－ｇｒａｍの遷移確率が０となってしまう。従って、平滑化処理によりテキストに出現しな※い単語組に対しても、０でない遷移確率を与える必要がある。もう１つの問題は、本発明に係るタスク適応化処理は、全ての遷移確率を独立に求める手法であるため、遷移確率の和が１になるとは限らない。連続音声認識等に適用する際は、問題とはならないが、パープレキシティで評価する際は、１に正規化されていないと、正しい評価ができない。従って、近年盛んに用いられているＢａｃｋ－Ｏｆｆ平滑化法（例えば、従来技術文献４参照。）を拡張して、これらの問題を解決する方法を述べる。

【００５９】単語列w_1^{n}が不特定のクラスタのテキストデータ２１－ｋに含まれる場合は、上記のタスク適応化処理により、遷移確率p_{MAP}（w_n｜w_1^{n-1}）を求め、チューリング（Ｔｕｒｉｎｇ）推定法により、確率p_{MAP}（w_n｜w_1^{n-1}）を軽減する。ただし、軽減係数は不特定のクラスタのテキストデータ３１の頻度（c_I（w_1^{n}））を用いて計算する。当該軽減により生じた確率の余剰分をw_1^{n}が不特定のクラスタのテキストデータ３１に含まれない単語連鎖に対して、（ｎ－１）－ｇｒａｍの遷移確率に比例して配分する。以上をまとめると、クラスタ適応化されたＮ－ｇｒａｍの平滑化後の遷移確率Ｐｓ（w_n｜w_1^{n-1}）は次式で表される。

【００６０】

【数１６】

$$\begin{aligned}
&Ps(w_n\mid w_1^{n-1})\\
&=Ph(w_n\mid w_1^{n-1}),\ c_I(w_1^{n-1})>0\text{のとき}\\
&=\alpha(w_1^{n-1})\,Ps(w_n\mid w_2^{n-1}),\ c_I(w_1^{n-1})=0,\ c_I(w_2^{n-1})>0\text{のとき}\\
&=Ps(w_n\mid w_2^{n-1}),\ c_I(w_1^{n-1})=0,\ c_I(w_2^{n-1})=0\text{のとき}
\end{aligned}$$

【００６１】上記の数１６において、Ｐｈはクラスタ適応化により得られる確率に軽減係数をかけたものであり、次式で与えられる。

【００６２】

【数１７】$Ph(w_n\mid w_1^{n-1})=\{c_I(w_1^{n})+1\}/\{c_I(w_1^{n})\}\times\{n\,c_I(w_1^{n})+1\}/\{n\ c_{ci}(w_1^{n})\}\cdot p_{MAP}(w_1^{n})$

【００６３】ここで、n_cは、不特定のクラスタのテキストデータ３１中にｃ回出現する単語列の種類数（異なり）であり、また、数１６で、α（w_1^{n-1}）は正規化のための係数であり、次のように求められる。

【００６４】

【数１８】$\alpha(w_1^{n-1})=Aa/Ab$

ここで、

$$Aa \equiv 1-\sum_{w_n : c_I(w_1^n)>0} Ph(w_n \mid w_1^{n-1})$$

$$Ab \equiv 1-\sum_{w_n : c_I(w_2^n)>0} Ph(w_n \mid w_1^{n-1})$$

【００６５】以上のＢａｃｋ－ｏｆｆ平滑化法を応用した手法を用いることにより、学習データ上に出現しない単語連鎖に対しても確率値を与えることができる。また、遷移確率p_{MAP}が正規化されていなくても、上記数１８においてαを求める際に正規化の計算を行うため、遷移確率の和は自動的に１になる。

【００６６】従って、本実施形態で用いるＭＡＰ推定法によるＮ－ｇｒａｍの適応手法について要約すると、以下の通りとなる。ＭＡＰ推定法による単語列ｈから次単語ｗへの単語Ｎ－ｇｒａｍの遷移確率Ｐ（ｗ｜ｈ）は次式により与えられる。

【００６７】

【数１９】Ｐ（ｗ｜ｈ）＝｛Ｎ（ｈ，ｗ）＋α－１｝／（Ｎ（ｈ）＋α＋β－２）

【００６８】ここで、Ｎ（＃）はそのクラスタでの単語（列）＃の出現頻度である。また、α及びβは事前分布として用いるベータ分布（$ap\alpha^{-1}(1-p)\beta^{-1}$，ａ）のパラメータであり、次式により求められる。

【００６９】

【数２０】$\alpha=\{\mu^2(1-\mu)/\sigma^2\}-\mu$

【数２１】$\alpha+\beta=\{\mu(1-\mu)/\sigma^2\}-1$

【００７０】上式のμ及びσ^2は、クラスタ毎の遷移確率Ｐ（ｗ｜ｈ）の分布の平均及び分散である。また、本実施形態で用いる可変長Ｎ－ｇｒａｍは、クラスＮ－ｇｒａｍを基本としたモデルであり、遷移確率はＰ（ｗｓ｜ｃ（ｗｓ））・Ｐ（ｃ（ｗｓ）｜ｃ（ｈ））として与えられる。ただし、ｗｓは可変長の単語列で、ｃ（＃）は単語（列）＃の属するクラスである。Ｐ（ｃ（ｗｓ）｜ｃ（ｈ））はクラス間の遷移確率であり、上記数１９と同様に与えることができる。また、Ｐ（ｗｓ｜ｃ（ｗｓ））はその単語の属するクラスから単語の出現確率であり、ＭＡＰ推定により次式で与えられる。

【００７１】

【数２２】Ｐ（ｗｓ｜ｃ（ｗｓ））＝｛Ｎ（ｗｓ）＋α－１｝／｛Ｎ（ｃ（ｗｓ））＋α＋β－２｝

【００７２】また、公知のＢａｃｋ－ｏｆｆ平滑化法を用い、コーパス上に出現しなかった単語遷移に対して確率を与えるとともに、遷移確率の和が１になるよう確率の正規化を行う。

【００７３】図４は、図１のクラスタリング処理部４０によって実行されるクラスタリング処理を示すフローチャートである。図４において、まず、ステップＳ１において学習用テキストデータメモリ２０からＫ個の発声音声文をランダムに選択し、ステップＳ２において、上記

選択した発声音声文を、学習用テキストデータメモリ２１の各メモリ２１－１乃至２１－Ｋに、クラスタ１からクラスタＭへの順番に１文ずつ書き込む。次いで、ステップＳ３において学習用テキストデータメモリ２１の各クラスタのテキストデータを読み出し、ステップＳ４において上記読み出した各クラスタのテキストデータから、各クラスタ毎に統計的言語モデルを生成する。ここで、生成された統計的言語モデルはクラスタリング処理部４０の内部メモリ又は統計的言語モデルメモリ３２に記憶される。

【００７４】さらに、ステップＳ５において学習用テキストデータメモリ２０から１文ずつ読み出し、ステップＳ６においてステップＳ４で生成した各統計的言語モデルに対して、ステップＳ５で読み出した文の生成確率を計算し、確率の最も高いクラスタＣを選択し、ステップＳ７においてステップＳ５で読み出した文を、学習用テキストデータメモリ２１のメモリ２１－ＣにクラスタＣとして書き込む。そして、ステップＳ８においてステップＳ５で読み出した文は最後の文か否かが判断され、ＮＯであるときは、次の文を処理するために、ステップＳ５に戻る。一方、ステップＳ８でＹＥＳのときは、ステップＳ９において、ステップＳ６で選択されたクラスタＣが１文でも変化したか否かが判断され、ＹＥＳのときは再度クラスタリング処理を実行するために、ステップＳ３に戻る。ステップＳ９でＮＯであるときは、当該クラスタリング処理を終了する。

【００７５】図５は、図１の言語モデル生成部４１によって実行される言語モデル生成処理を示すフローチャートである。図５において、まず、ステップＳ１１において学習用テキストデータメモリ２０からコーパスのテキストデータを読み出し、ステップＳ１２において読み出したテキストデータに基づいて最尤推定法を用いて単語ｂｉｇｒａｍの遷移確率を数６を用いて計算する。次いで、ステップＳ１３において計算された単語ｂｉｇｒａｍの遷移確率を含む統計的言語モデルを生成して、統計的言語モデルメモリ３１に記憶して当該言語モデル生成処理を終了する。

【００７６】図６は、図１の言語モデル生成部４２によって実行される言語モデル生成処理を示すフローチャートである。図６において、まず、ステップＳ２１において学習用テキストデータメモリ２０から各クラスタｋのテキストデータ２１－ｋ（ｋ＝１，２，…，Ｋ）を読み出す。次いで、ステップＳ２２において、読み出した各クラスタｋのテキストデータ２１－ｋに基づいて最尤推定法を用いて各クラスタｋ毎に単語ｂｉｇｒａｍの遷移確率を数６を用いて計算し、ステップＳ２３において各クラスタｋの単語ｂｉｇｒａｍの遷移確率の出現頻度分布をベータ分布と仮定し、ベータ分布の加重平均μ及び加重分散σ^2を数１３及び数１４を用いて計算した後これらに基づいて数１２を用いてパラメータα及びβを計

算する。さらに、ステップＳ２４において事前知識のパラメータα及びβと、特定クラスタのテキストデータ２１－ｉの事後知識のパラメータN及びnとに基づいて数１０を用いて各クラスタｋ毎の単語ｂｉｇｒａｍの遷移確率ｐを計算する。ここで、上記ステップＳ２１からＳ２４までの処理は、すべてのクラスタｋ＝１，２，…，Kについて実行される。さらに、ステップＳ２５において各クラスタｋ毎の単語ｂｉｇｒａｍの遷移確率ｐに基づいてＢａｃｋ－ｏｆｆ平滑化処理を実行し、処理後の各クラスタｋ毎の単語ｂｉｇｒａｍの遷移確率を含む統計的言語モデルを生成して、クラスタ適応化された統計的言語モデルメモリ３２に記憶して、当該言語モデル生成処理を終了する。

【００７７】次いで、図１に示す連続音声認識装置の構成及び動作について説明する。図１において、単語照合部４に接続された音素隠れマルコフモデル（以下、隠れマルコフモデルをHMMという。）メモリ１１内の音素HMMは、各状態を含んで表され、各状態はそれぞれ以下の情報を有する。

（ａ）状態番号、（ｂ）受理可能なコンテキストクラス、（ｃ）先行状態、及び後続状態のリスト、（ｄ）出力確率密度分布のパラメータ、及び（ｅ）自己遷移確率及び後続状態への遷移確率。

なお、本実施形態において用いる音素HMMは、各分布がどの話者に由来するかを特定する必要があるため、所定の話者混合HMMを変換して生成する。ここで、出力確率密度関数は３４次元の対角共分散行列をもつ混合ガウス分布である。また、単語照合部４に接続された単語辞書メモリ１２内の単語辞書は、音素HMMメモリ１１内の音素HMMの各単語毎にシンボルで表した読みを示すシンボル列を格納する。

【００７８】図１において、話者の発声音声はマイクロホン１に入力されて音声信号に変換された後、特徴抽出部２に入力される。特徴抽出部２は、入力された音声信号をA／D変換した後、例えばＬＰＣ分析を実行し、対数パワー、１６次ケプストラム係数、Δ対数パワー及び１６次Δケプストラム係数を含む３４次元の特徴パラメータを抽出する。抽出された特徴パラメータの時系列はバッファメモリ３を介して単語照合部４に入力される。

【００７９】単語照合部４は、ワンーパス・ビタビ復号化法を用いて、バッファメモリ３を介して入力される特徴パラメータのデータに基づいて、音素HMM１１と単語辞書１２とを用いて単語仮説を検出し尤度を計算して出力する。ここで、単語照合部４は、各時刻の各HMMの状態毎に、単語内の尤度と発声開始からの尤度を計算する。尤度は、単語の識別番号、単語の開始時刻、先行単語の違い毎に個別にもつ。また、計算処理量の削減のために、音素HMM１１及び単語辞書１２とに基づいて計算される総尤度のうちの低い尤度のグリッド仮説を削減する。単語照合部４は、その結果の単語仮説と尤度の情報を発声開始時刻からの時間情報（具体的には、例えばフレーム番号）とともにバッファメモリ５を介して単語仮説絞込部６ａに出力する。

【００８０】単語仮説絞込部６ａは、単語照合部４からバッファメモリ５を介して出力される単語仮説に基づいて、統計的言語モデルメモリ３２内の統計的言語モデルを参照して、終了時刻が等しく開始時刻が異なる同一の単語の単語仮説に対して、当該単語の先頭音素環境毎に、発声開始時刻から当該単語の終了時刻に至る計算された総尤度のうちの最も高い尤度を有する１つの単語仮説で代表させるように単語仮説の絞り込みを行った後、絞り込み後のすべての単語仮説の単語列のうち、最大の総尤度を有する仮説の単語列を認識結果としてバッファメモリ７を介して言語モデル選択部８に出力する。本実施形態においては、好ましくは、処理すべき当該単語の先頭音素環境とは、当該単語より先行する単語仮説の最終音素と、当該単語の単語仮説の最初の２つの音素とを含む３つの音素並びをいう。

【００８１】単語仮説絞込部６ａの処理においては、例えば、図２に示すように、（ｉ－１）番目の単語W_{i-1}の次に、音素列a_1，a_2，…，a_nからなるｉ番目の単語W_iがくるときに、単語W_{i-1}の単語仮説として６つの仮説Ｗａ，Ｗｂ，Ｗｃ，Ｗｄ，Ｗｅ，Ｗｆが存在している。ここで、前者３つの単語仮説Ｗａ，Ｗｂ，Ｗｃの最終音素は／ｘ／であるとし、後者３つの単語仮説Ｗｄ，Ｗｅ，Ｗｆの最終音素は／ｙ／であるとする。終了時刻t_eと先頭音素環境が等しい仮説（図２では先頭音素環境が“ｘ／a_1／a_2”である上から３つの単語仮説）のうち総尤度が最も高い仮説（例えば、図２において１番上の仮説）以外を削除する。なお、上から４番めの仮説は先頭音素環境が違うため、すなわち、先行する単語仮説の最終音素がｘではなくｙであるので、上から４番めの仮説を削除しない。すなわち、先行する単語仮説の最終音素毎に１つのみ仮説を残す。図２の例では、最終音素／ｘ／に対して１つの仮説を残し、最終音素／ｙ／に対して１つの仮説を残す。

【００８２】次いで、言語モデル選択部８は、上述のように、数４に従ってクラスタの統計的言語モデルを統計的言語モデルメモリ３２から選択し、すなわち、K個のクラスタのそれぞれのモデルＬＭｋで認識結果文Wに対する生成確率Ｐ（Ｗ｜ＬＭｋ）（１≦ｋ≦K）を求め、確率の最も高いクラスタモデルＬＭｃを選択して、その選択情報を単語仮説絞込部６ｂに出力する。これに応答して、単語仮説絞込部６ｂは、単語仮説絞込部６ａによって絞り込まれた単語仮説に対して、再度、統計的言語モデルメモリ３２で選択された統計的言語モデルを用いて、単語仮説絞込部６ａと同様の処理を実行して、単語仮説の絞込処理を実行して、処理後の例えば最尤の絞り込んだ単語仮説（ここで、ｎ－ｂｅｓｔでもよい。）を認識結果として出力する。

【００８３】以上の実施形態においては、当該単語の先頭音素環境とは、当該単語より先行する単語仮説の最終音素と、当該単語の単語仮説の最初の２つの音素とを含む３つの音素並びとして定義されているが、本発明はこれに限らず、先行する単語仮説の最終音素と、最終音素と連続する先行する単語仮説の少なくとも１つの音素とを含む先行単語仮説の音素列と、当該単語の単語仮説の最初の音素を含む音素列とを含む音素並びとしてもよい。

【００８４】以上の実施形態において、特徴抽出部２と、単語照合部４と、単語仮説絞込部６ａ，６ｂと、クラスタリング処理部４０と、言語モデル生成部４１，４２とは、例えば、デジタル電子計算機などのコンピュータで構成され、バッファメモリ３，５と、音素HMMメモリ１１と、単語辞書メモリ１２と、学習用テキストデータメモリ２０，２１と、統計的言語モデルメモリ３１，３２とは、例えばハードデイスクメモリなどの記憶装置で構成される。

【００８５】以上実施形態においては、単語照合部４と単語仮説絞込部６ａ，６ｂとを用いて音声認識を行っているが、本発明はこれに限らず、例えば、音素HMM１１を参照する音素照合部と、例えばＯｎｅ　Ｐａｓｓ　ＤＰアルゴリズムを用いて統計的言語モデルを参照して単語の音声認識を行う音声認識部とで構成してもよい。ただし、本実施形態の場合、統計的言語モデルメモリ３１を参照して音声認識する第１の音声認識部と、統計的言語モデルメモリ３２内で言語モデル選択部８によって＊選択された１つのクラスタの統計的言語モデルを参照して音声認識する第２の音声認識部とを備えることになる。

【００８６】

【実施例】本発明者は、本実施形態で用いるタスク適応化された統計的言語モデルの性能を確認するため、評価実験を行った。実験で用いたデータは、本特許出願人が所有する自然発話データベース（例えば、従来技術文献１１「Ｔ．Ｍｏｒｉｍｏｔｏ　ｅｔ　ａｌ．，“Ａ Ｓｐｅｅｃｈ　ａｎｄ　Ｌａｎｇｕａｇｅ　Ｄａｔａｂａｓｅ　ｆｏｒ　Ｓｐｅｅｃｈ　Ｔｒａｎｓｌａｔｉｏｎ　Ｒｅｓｅａｒｃｈ”，ＩＣＳＬＰ，ｐｐ．１７９１－１７９４，１９９４年」参照。）であり、本データベースのサイズは、１，３３２対話、３２，０７４文、５９７，６２６単語で、語彙は7、２２１語である。このうち評価用として「ホテルの部屋の予約」タスクから４０対話、１１６６文、１８、３８１単語を選択し、残りのデータを言語モデルの学習に使用した。

【００８７】最初にテストセットパープレキシティにより評価を行った。可変長Ｎ－ｇｒａｍは活用形及び活用型を含む１５８品詞による初期クラスから、５００クラス分離を行ったモデルを使用した。クラスタ数4、8、１６、３２、６４の時のクラスタモデルと、データベース全体で作成したモデル（クラスタ数１）とのパープレキシティの比較を表１に示す。

【００８８】

【表1】

パープレキシティによる比較

全体モデル	クラスタモデル（クラスタ数）				
	4	8	16	32	64
14.21	13.00	12.33	11.44	10.44	9.72

【００８９】ここで、パープレキシティとは以下のように定義される。例えば、複数ｎ個の単語からなる長い単語列$w_1^n = w_1 w_2 \cdots w_n$があるときのエントロピーＨ（ｎ）は次式で表される。

【００９０】

【数２３】

$$H(n) = -(1/n) \cdot \log_2 P(w_1^n)$$

【００９１】ここで、$P(w_1^n)$は単語列w_1^nの生成確率であり、パープレキシティＰＰ（ｎ）は次式で表される。

【００９２】

【数２４】

$$PP(n) = 2^{H(n)}$$

【００９３】上記表１より、クラスタ数に比例してパープレキシティが減少しており、クラスタ毎の言語的特徴がよりできたと考えられる。クラスタ数が６４の時は、全体モデルよりもパープレキシティが約３２％減少した。また、評価に用いた「ホテルの部屋の予約」タスクのデータは、データ量が多いために第２の従来例では、タスク適応の効果は、単語ｂｉｇｒａｍで５％程度と小さかったが、本実施形態に係る装置では、文の内容毎に適応モデルを作成するため、大きな精度向上が得られたと考えられる。計算量の都合のため、クラスタ数は最大６４としたが、さらにクラスタ数を増加させることにより、パープレキシティは減少すると考えられる。ただし、クラスタ数を多くしすぎると各クラスタのデータ量が少なくなりすぎ、パラメータ推定が困難になるため、限界はあると考えられる。

【００９４】次に、連続音声認識に適用した際の認識率によって評価を行った。音響モデルにはＭＬ－ＳＳＳ法（従来技術文献１２「Ｍ．Ｏｓｔｅｎｄｏｒｆ　ｅｔ　ａｌ．，”ＨＭＭ　ｔｏｐｏｌｏｇｙ　ｄｅｓｉｇｎ　ｕｓｉｎｇ　ｍａｘｉｍｕｍｌｉｋｅｌｉｈｏｏｄ　ｓ

uccessive state splitting”, Computer Speech and language, No. 11, pp. 17－41, 1997年」参照。）によるHMM網（801状態5混合分布）の不特定話者モデルを用い、単語グラフサーチ法（従来技術文献13「清水ほか, ”単語グラフを用いた自由発話音声認識”, 電子情報通信学会研究報告, SP95－88, pp. 49－54, 1995年12月」参＊照。）により認識解の探索を行った。言語モデルは、コーパス全体で作成したモデルとクラスタ数4、16、64のクラスタモデルとを比較した。表2に単語認識率（Accuracy）（%）及び文認識率（%）を示す。

【0095】

【表2】

連続音声認識における性能比較

認識率の種類	全体モデル	クラスタモデル（クラスタ数）		
		4	16	64
単語認識率	77.66	78.69	79.06	78.54
文認識率	33.43	35.82	36.12	37.31

【0096】上記表2より、単語認識率はクラスタ数16の時に全体モデルより約1．4%向上（改善率約6%）し、文認識率はクラスタ数64の時に最大約3．9%向上（改善率約6%）し、連続音声認識における有効性を確認した。クラスタ数64の時の単語認識率はクラスタ数4、16の時よりも低下しているが、これは、誤認識が生じた際にクラスタモデルの選択が正しく行われないことが原因と考えられる。

【0097】以上説明したように、本実施形態によれば、コーパスの各文をクラスタリングし、それぞれのクラスタ毎にMAP推定によるN－gram型の言語モデルを作成することにより言語特徴をより効果的に表現できる手法を開示している。実験の結果、パープレキシティは最大約32%減少し、また、連続音声認識に適用した際、単語認識率及び文認識率共に最大約6%改善し、本手法の有効性を確認した。すなわち、本実施形態によれば、少量のテキストデータを用いて、従来例に比較して、より高い遷移確率の予測精度及び信頼性を有する統計的言語モデルを生成することができるとともに、タスク選択を自動的に行うことができ、選択された統計的言語モデルを用いて音声認識することにより、従来例に比較してより高い音声認識率で連続的に音声認識することができる。

【0098】以上の実施形態において、統計的言語モデルは、N－gramの言語モデルを含むが、ここで、Nは2及び3に限らず、4以上の自然数であってもよい。

【0099】

【発明の効果】以上詳述したように本発明に係る請求項1記載の統計的言語モデル生成装置によれば、複数のクラスタの統計的言語モデルを記憶する記憶手段と、所定の複数の発声音声文を含む学習用テキストデータに基づいて各発声音声文に対する統計的言語モデルを生成して、各発声音声文が各クラスタに対応するように、上記生成した統計的言語モデルを上記記憶手段に記憶する初期化手段と、上記学習用テキストデータの各発声音声文について、各クラスタにおける統計的言語モデルの文生成確率を計算して最大の文生成確率を有するクラスタを選択してその発声音声文を所属させるように上記記憶手段に記憶するクラスタ選択手段と、発声音声文が属するクラスタが変化したときに、各クラスタ毎に、上記クラスタ選択手段によって選択された発声音声文を用いて上記記憶手段に記憶された各統計的言語モデルを更新して、各クラスタに対応した統計的言語モデルを生成するモデル変更手段と、上記複数の発声音声文に属するクラスタが1文も変化しなくなるまで、上記クラスタ選択手段の処理と、上記モデル変更手段の処理を繰り返す制御手段とを備える。従って、少量のテキストデータを用いて、従来例に比較して、より高い遷移確率の予測精度及び信頼性を有する統計的言語モデルを生成することができる。

【0100】また、請求項2記載の統計的言語モデル生成装置によれば、請求項1記載の統計的言語モデル生成装置において、さらに、上記記憶手段に記憶された各クラスタ毎のテキストデータに基づいて、最尤推定法を用いて各クラスタ毎に単語のN－gram（Nは2以上の自然数である。）の遷移確率を演算する第1の演算手段と、上記第1の演算手段によって演算された各クラスタ毎の単語のN－gramの遷移確率の出現分布を事前知識の所定の確率分布と仮定し、各クラスタ毎の確率分布の加重平均及び加重分散を演算した後、演算された加重平均と加重分散に基づいて事前知識の確率分布のパラメータを演算する第2の演算手段と、上記第2の演算手段によって演算された事前知識の確率分布のパラメータと、上記学習用テキストデータうちの特定クラスタのテキストデータの事後知識における処理対象の単語列の直前の単語列の出現回数と、処理対象の単語列の出現回数

とに基づいて、各クラスタ毎の単語のＮ－ｇｒａｍの遷移確率を計算することにより、各クラスタ毎の単語のＮ－ｇｒａｍの遷移確率を含む統計的言語モデルを生成する第３の演算手段とを備える。従って、少量のテキストデータを用いて、従来例に比較して、より高い遷移確率の予測精度及び信頼性を有する統計的言語モデルを生成することができる。

【０１０１】さらに、請求項３記載の統計的言語モデル生成装置によれば、請求項２記載の統計的言語モデル生成装置において、さらに、上記第３の演算手段によって演算された各クラスタ毎の単語のＮ－ｇｒａｍの遷移確率に基づいて，所定の平滑化処理を実行し、処理後の各クラスタ毎の単語のＮ－ｇｒａｍの遷移確率を含む統計的言語モデルを生成する第１の生成手段を備える。従って、少量のテキストデータを用いて、従来例に比較して、より高い遷移確率の予測精度及び信頼性を有する統計的言語モデルを生成することができる。

【０１０２】またさらに、請求項４記載の統計的言語モデル生成装置によれば、請求項３記載の統計的言語モデル生成装置において、さらに、上記学習用テキストデータに基づいて、最尤推定法を用いて単語のＮ－ｇｒａｍ（Ｎは２以上の自然数である。）の遷移確率を演算して、上記単語のＮ－ｇｒａｍの遷移確率を含む別の統計的言語モデルを生成する第２の生成手段を備える。従って、少量のテキストデータを用いて、従来例に比較して、より高い遷移確率の予測精度及び信頼性を有する統計的言語モデルを生成することができる。

【０１０３】本発明に係る請求項５記載の音声認識装置によれば、入力される発声音声文の音声信号に基づいて、所定の統計的言語モデルを用いて音声認識する音声認識装置において、請求項４記載の統計的言語モデル生成装置と、上記第２の生成手段によって生成された別の統計的言語モデルを用いて、入力される発声音声文の音声信号を音声認識して第１の認識仮説を出力する第１の音声認識手段と、上記第１の音声認識手段から出力される第１の認識仮説に応答して、上記第１の生成手段によって生成された各クラスタ毎の統計的言語モデルを用いて、入力される発声音声文の音声信号を音声認識して、文生成確率が最大のクラスタの統計的言語モデル生成装置を選択するモデル選択手段と、上記モデル選択手段によって選択されたクラスタの統計的言語モデルを用いて、上記第１の音声認識手段から出力される第１の認識仮説に対して絞込処理を行って第２の認識仮説を生成して認識結果として出力する第２の音声認識手段とを備える。従って、少量のテキストデータを用いて、従来例に比較して、より高い遷移確率の予測精度及び信頼性を有する統計的言語モデルを生成することができるとともに、タスク選択を自動的に行うことができ、選択された統計的言語モデルを用いて音声認識することにより、従来例に比較してより高い音声認識率で連続的に音声認識することができる。

【図面の簡単な説明】

【図１】　本発明に係る一実施形態である連続音声認識装置のブロック図である。

【図２】　図１の連続音声認識装置における単語仮説絞込部６ａ及び６ｂの処理を示すタイミングチャートである。

【図３】　図１の言語モデル生成部４２の処理を示すブロック図である。

【図４】　図１のクラスタリング処理部４０によって実行されるクラスタリング処理を示すフローチャートである。

【図５】　図１の言語モデル生成部４１によって実行される言語モデル生成処理を示すフローチャートである。

【図６】　図１の言語モデル生成部４２によって実行される言語モデル生成処理を示すフローチャートである。

【符号の説明】

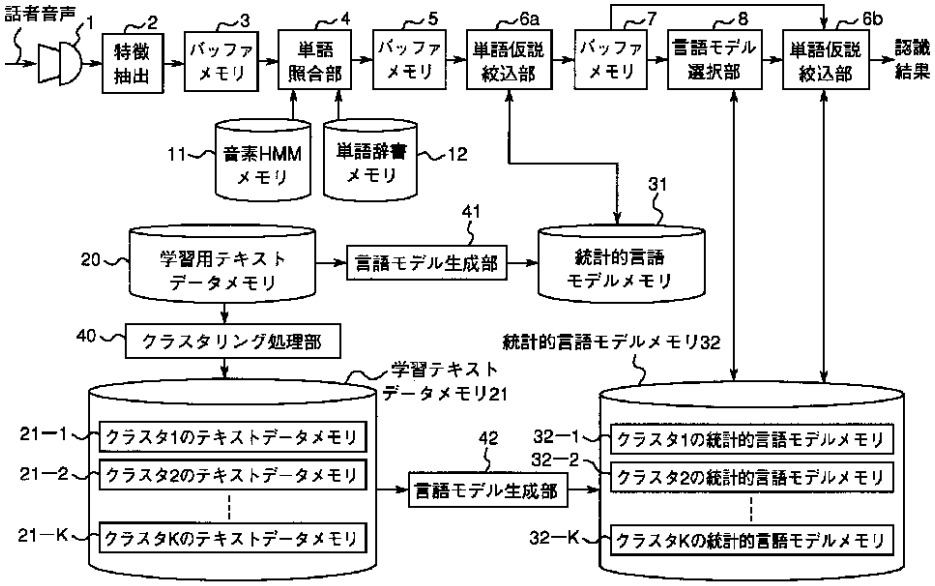
１…マイクロホン、
２…特徴抽出部、
３，５，７…バッファメモリ、
４…単語照合部、
６ａ，６ｂ…単語仮説絞込部、
８…言語モデル選択部、
１１…音素HMMメモリ、
１２…単語辞書メモリ、
２０，２１…学習用テキストデータメモリ、
２１－ｋ…各クラスタのテキストデータメモリ、
３１，３２…統計的言語モデルメモリ、
３２－ｋ…各クラスタの統計的言語モデルメモリ、
４１，４２…言語モデル生成部。

【要約】

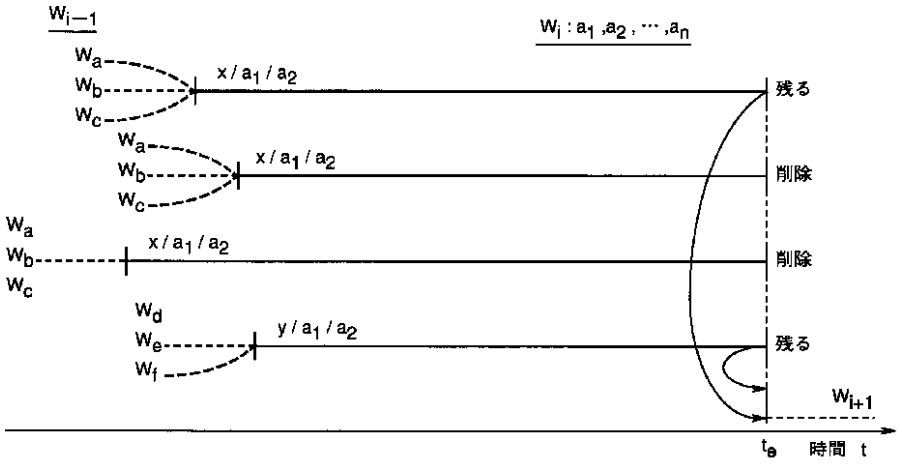
【課題】　遷移確率の予測精度及び信頼性を改善可能な統計的言語モデルを生成し、統計的言語モデルを用いてより高い音声認識率で音声認識する。

【解決手段】　学習用テキストデータ全体をクラスタリング処理部４０によって自動的にクラスタリングしてクラスタ毎のテキストデータをメモリ２１に記憶し、言語モデル生成部４２によりＭＡＰ推定法を用いて各クラスタ毎の統計的言語モデルを生成してメモリ３２に記憶する。一方、学習用テキストデータ全体に対して統計的言語モデルを生成してメモリ３１に記憶する。単語照合部４による単語仮説の生成の後、単語仮説絞込部６ａはメモリ３１内の統計的言語モデルを用いて単語仮説の絞込処理を実行した後、言語モデル選択部８はメモリ３２内の各クラスタの統計的言語モデルのうちで文生成確率が最大のモデルを選択して、単語仮説絞込部６ｂは選択されたモデルを用いて再度の絞込処理を行って認識結果を出力する。

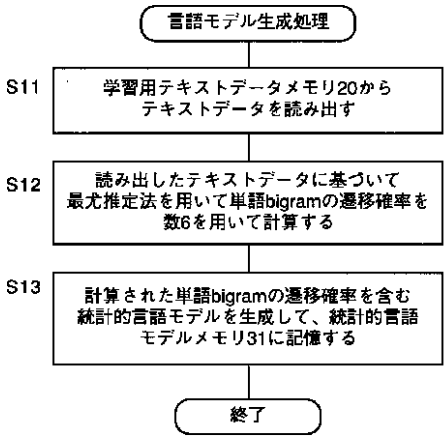
【図１】


話者音声
1
2
特徴抽出
3
バッファメモリ
4
単語照合部
5
バッファメモリ
6a
単語仮説絞込部
7
バッファメモリ
8
言語モデル選択部
6b
単語仮説絞込部
認識結果
11
音素HMMメモリ
12
単語辞書メモリ
31
41
20
学習用テキストデータメモリ
言語モデル生成部
統計的言語モデルメモリ
40
クラスタリング処理部
統計的言語モデルメモリ32
学習テキストデータメモリ21
21−1
クラスタ1のテキストデータメモリ
21−2
クラスタ2のテキストデータメモリ
21−K
クラスタKのテキストデータメモリ
42
言語モデル生成部
32−1
クラスタ1の統計的言語モデルメモリ
32−2
クラスタ2の統計的言語モデルメモリ
32−K
クラスタKの統計的言語モデルメモリ

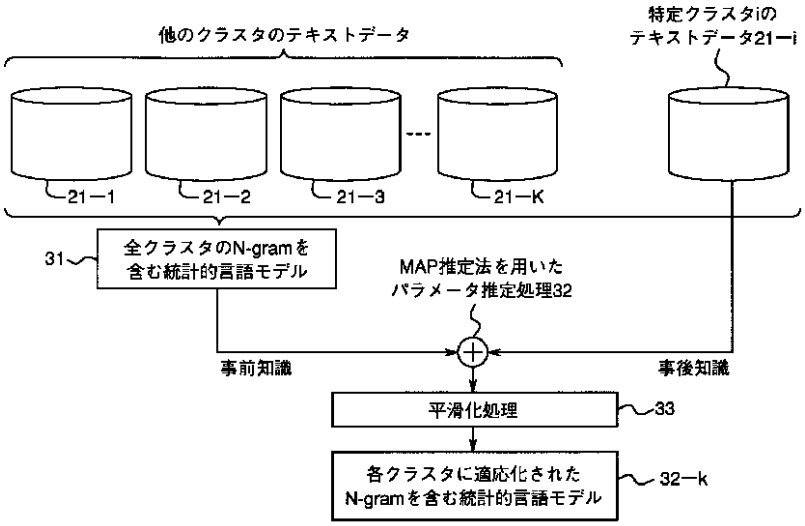
【図２】


W_{i-1}
$W_i : a_1, a_2, \cdots, a_n$
W_a
W_b
W_c
$x / a_1 / a_2$
残る
削除
削除
W_d
W_e
W_f
$y / a_1 / a_2$
残る
W_{i+1}
t_e
時間 t

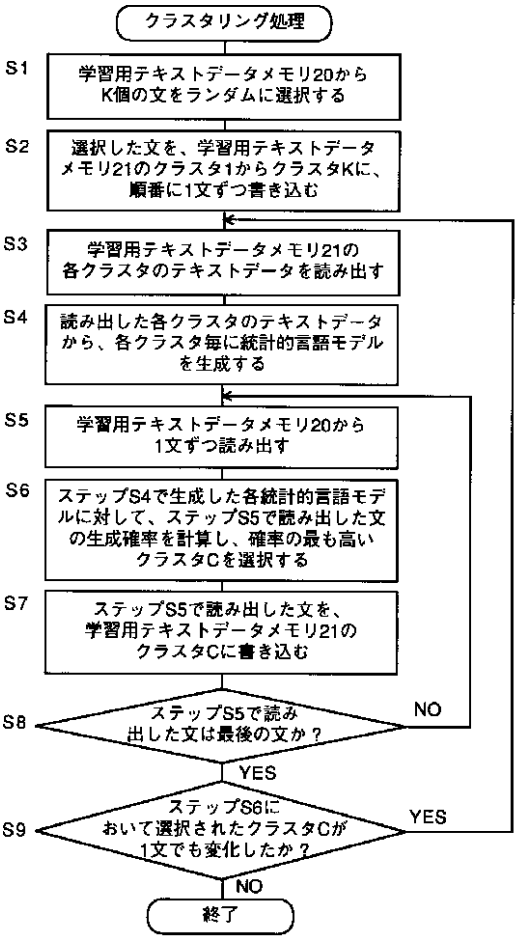
【図５】


言語モデル生成処理
S11
学習用テキストデータメモリ20からテキストデータを読み出す
S12
読み出したテキストデータに基づいて最尤推定法を用いて単語bigramの遷移確率を数6を用いて計算する
S13
計算された単語bigramの遷移確率を含む統計的言語モデルを生成して、統計的言語モデルメモリ31に記憶する
終了

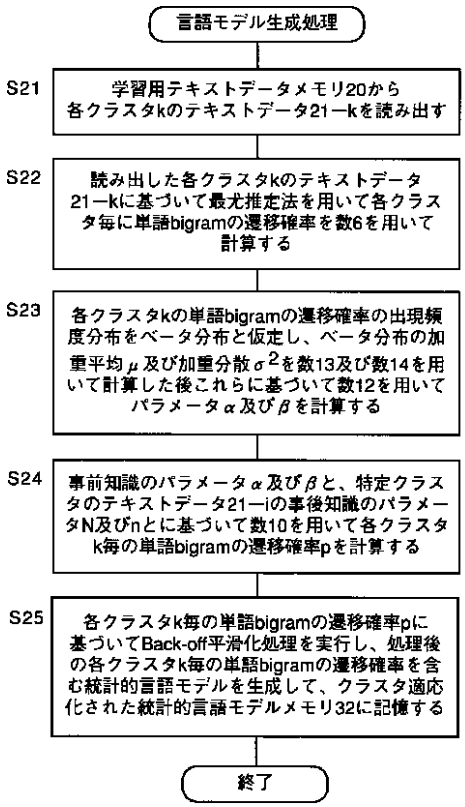
【図３】


他のクラスタのテキストデータ
特定クラスタiの
テキストデータ21−i
21−1
21−2
21−3
21−K
31
全クラスタのN-gramを
含む統計的言語モデル
MAP推定法を用いた
パラメータ推定処理32
事前知識
事後知識
平滑化処理
33
各クラスタに適応化された
N-gramを含む統計的言語モデル
32−k

【図４】


クラスタリング処理
S1
学習用テキストデータメモリ20から
K個の文をランダムに選択する
S2
選択した文を、学習用テキストデータ
メモリ21のクラスタ1からクラスタKに、
順番に1文ずつ書き込む
S3
学習用テキストデータメモリ21の
各クラスタのテキストデータを読み出す
S4
読み出した各クラスタのテキストデータ
から、各クラスタ毎に統計的言語モデル
を生成する
S5
学習用テキストデータメモリ20から
1文ずつ読み出す
S6
ステップS4で生成した各統計的言語モデ
ルに対して、ステップS5で読み出した文
の生成確率を計算し、確率の最も高い
クラスタCを選択する
S7
ステップS5で読み出した文を、
学習用テキストデータメモリ21の
クラスタCに書き込む
S8
ステップS5で読み
出した文は最後の文か？
NO
YES
S9
ステップS6に
おいて選択されたクラスタCが
1文でも変化したか？
YES
NO
終了

【図６】


言語モデル生成処理
S21
学習用テキストデータメモリ20から
各クラスタkのテキストデータ21−kを読み出す
S22
読み出した各クラスタkのテキストデータ
21−kに基づいて最尤推定法を用いて各クラス
タ毎に単語bigramの遷移確率を数6を用いて
計算する
S23
各クラスタkの単語bigramの遷移確率の出現頻
度分布をベータ分布と仮定し、ベータ分布の加
重平均μ及び加重分散σ2を数13及び数14を用
いて計算した後これらに基づいて数12を用いて
パラメータα及びβを計算する
S24
事前知識のパラメータα及びβと、特定クラス
タのテキストデータ21−iの事後知識のパラメ
タN及びnとに基づいて数10を用いて各クラスタ
k毎の単語bigramの遷移確率pを計算する
S25
各クラスタk毎の単語bigramの遷移確率pに
基づいてBack-off平滑化処理を実行し、処理後
の各クラスタk毎の単語bigramの遷移確率を含
む統計的言語モデルを生成して、クラスタ適応
化された統計的言語モデルメモリ32に記憶する
終了

フロントページの続き

(56)参考文献　　特開　平６－27985（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.[6]，ＤＢ名)
G10L　3/00 - 9/20
ＪＩＣＳＴファイル（ＪＯＩＳ）