

【公報種別】特許法第17条の2の規定による補正の掲載

【部門区分】第6部門第2区分

【発行日】平成17年5月19日(2005.5.19)

【公開番号】特開2004-198597(P2004-198597A)

【公開日】平成16年7月15日(2004.7.15)

【年通号数】公開・登録公報2004-027

【出願番号】特願2002-365074(P2002-365074)

【国際特許分類第7版】

G 10 L 15/18

G 10 L 15/06

G 10 L 15/28

【F I】

G 10 L 3/00 5 3 7 F

G 10 L 3/00 5 2 1 W

G 10 L 3/00 5 7 1 C

G 10 L 3/00 5 3 7 D

【手続補正書】

【提出日】平成16年7月7日(2004.7.7)

【手続補正1】

【補正対象書類名】明細書

【補正対象項目名】特許請求の範囲

【補正方法】変更

【補正の内容】

【特許請求の範囲】

【請求項1】

ツリー構造の各ノードに対応する言語モデルの集合を含む階層的言語モデルを用いた音声認識装置としてコンピュータを動作させるコンピュータプログラムであって、

前記音声認識装置は、

前記ツリー構造のルートノードに対応する言語モデルを用いて入力に対する音声認識を行ない、音声認識の信頼度とともに認識結果を出力するための第1の音声認識手段と、

前記第1の音声認識手段の出力する認識結果と、前記階層的言語モデル中の前記ツリー構造の葉ノードの各々との間で計算される所定の指標に基づき、前記階層的言語モデル中の前記ツリー構造の葉ノードの内のいずれかを選択するための第1のノード選択手段と、

前記第1のノード選択手段により選択された葉ノードを特定する情報を記憶するための記憶手段と、

前記記憶手段に記憶された情報により特定される葉ノードに対応する言語モデルを用いて入力に対する音声認識を行ない、音声認識の信頼度とともに認識結果を出力するための第2の音声認識手段と、

前記第1の音声認識手段から与えられる信頼度と、前記第2の音声認識手段から与えられる信頼度とに基づいて、前記第1の音声認識手段の音声認識結果と前記第2の音声認識手段による音声認識結果とのうち、より高い信頼度に対応するものを選択するための手段とを含む、コンピュータプログラム。

【請求項2】

前記装置はさらに、

前記第1のノード選択手段により選択された葉ノードと、前記記憶手段により記憶された情報により特定される葉ノードとを比較するための比較手段と、

前記比較手段により不一致が検出された事に応答して、前記第1のノード選択手段により選択された葉ノードに対応する言語モデルを用いて、前記第1のノード選択手段により

一旦音声認識がされた入力に対し再度音声認識を行ない、音声認識の信頼度とともに認識結果を出力するための第3の音声認識手段と、

前記第1の音声認識手段より与えられる信頼度と、前記第3の音声認識手段から与えられる信頼度とに基づき、前記第1の音声認識手段の音声認識結果と前記第2の音声認識手段による音声認識結果とのうち、より高い信頼度に対応するものを選択するための手段と、

前記比較手段により不一致が検出された事に応答して、前記第1のノード選択手段が選択した葉ノードを特定する情報で前記記憶手段の記憶内容を更新するための手段とを含む、請求項1に記載のコンピュータプログラム。

【請求項3】

前記階層的言語モデルを構成する言語モデルはN-グラム言語モデルからなり、

前記第1のノード選択手段は、前記第1の音声認識手段の認識結果の尤度を、前記ツリー構造中の葉ノードの各々に対応する言語モデルに対して算出し、最も高い尤度が得られた葉ノードを選択する、請求項1に記載のコンピュータプログラム。

【請求項4】

前記階層的言語モデルを構成する言語モデルはユニグラム言語モデルである、請求項3に記載のコンピュータプログラム。

【請求項5】

各文は、所定の特徴ベクトルで表され、

前記階層的言語モデルを構成する各言語モデルは、各々特徴ベクトルのベクトル空間を二つの空間に分離する超平面で表され、

前記超平面と前記ベクトル空間内の各点との間では、各点が、前記超平面により分離された二つの空間のうち、予め定められる一方内に位置するか否かに関する信頼度を表す所定の指標の計算方法が定義されており、

前記第1のノード選択手段は、前記第1の音声認識結果の特徴ベクトルと、前記ツリー構造の葉ノードに対応する言語モデルに対応する超平面の各々との間で前記指標を算出し、指標の値が最も高い信頼度を表す言語モデルに対応する葉ノードを選択する、請求項1に記載のコンピュータプログラム。

【請求項6】

前記第1の音声認識手段は、予め定められる複数個の認識結果の候補をそれぞれの信頼度とともに出力し、

前記第1の音声認識手段の認識結果の特徴ベクトルが、前記複数個の認識結果の候補を総合したものに対して算出される、請求項5に記載のコンピュータプログラム。

【請求項7】

前記装置はさらに、

前記記憶手段により選択された葉ノードと前記ルートノードの間に位置する中位のノードを特定するための手段と、

前記中位のノードに対応する言語モデルを用いて入力に対する音声認識を行ない、音声認識の信頼度とともに認識結果を出力するための第3の音声認識手段と、

前記第2の音声認識手段から与えられる信頼度と、前記第3の音声認識手段から得られる信頼度とに基づいて、前記第2の音声認識手段の音声認識結果と前記第3の音声認識手段による認識結果とのうち、より高い信頼度に対応するものを選択するための手段を含む、請求項1に記載のコンピュータプログラム。

【請求項8】

複数個のトピックに分類された文を含むコーパスから、階層化された言語モデルを作成する方法を実現する様にコンピュータを動作させるコンピュータプログラムであって、

前記方法は、

ルートノードと、前記複数個のトピックに対応する複数個の葉ノードとからなるツリー構造を準備するステップと、

前記複数個の葉ノードのうちの第1のノードを選択するステップと、

前記第1のノードに対応する第1のトピックに対し、予め定められる方法により対応するトピック間で計算される距離が最も小さくなる第2のトピックを求めるステップと、

前記第1のトピックと、前記第2のトピックとの間で計算された距離が予め定められたしきい値より小さいか否かを判定するステップと、

前記判定するステップで、前記第1のトピックと第2のトピックとの間の距離が予め定められたしきい値より小さいと判定された事に基づいて、前記第1のノードと、前記第2のトピックに対応する第2のノードとに共通する親ノードを前記ツリー構造に追加するステップと、

前記第1のノードを以後のノードの探索の対象から除くステップと、

前記選択するステップと、前記求めるステップと、前記判定するステップと、前記追加するステップと、前記除くステップとを、予め定められた条件が成立するまで繰り返し実行するステップとを含む、コンピュータプログラム。

【請求項9】

前記距離は、以下の式により算出される $D_i s t_{UNI}(C_i, C_j)$ により表され、

【数1】

$$Dist_{UNI}(C_i, C_j) = \frac{PP(T_{C_i}, M_{C_j})}{PP(T_{C_i}, M_{C_i})} + \frac{PP(T_{C_j}, M_{C_i})}{PP(T_{C_j}, M_{C_j})}$$

ただし、 C_i, C_j はそれぞれ i 番目のトピッククラスおよび j 番目のトピッククラス、 T_{C_k} は k 番目のトピックに属する文の集合、 M_{C_k} は k 番目のトピックに属する文の集合から得られた言語モデル、 $PP(T_{C_i}, M_{C_j})$ は i 番目のトピックに属する文の集合内の各文と j 番目の言語モデル M_{C_j} との間で計算されるパープレキシティである、請求項8に記載のコンピュータプログラム。

【請求項10】

前記言語モデルはユニグラムの言語モデルである、請求項9に記載のプログラム。

【請求項11】

前記コーパス内の各文は所定の特徴ベクトルで表され、

前記モデルは前記特徴ベクトルのベクトル空間内においてサポートベクトルマシン (SVM) により求められる超平面により表され、

前記距離は、以下の式 $D_i s t_{SVM}(C_i, C_j)$ により算出され、

【数2】

$$Dist_{SVM}(C_i, C_j) = \left\| dist_{avg}(C_i, \bar{X}c_j) - dist_{avg}(C_j, \bar{X}c_j) \right\| + \left\| dist_{avg}(C_j, \bar{X}c_i) - dist_{avg}(C_i, \bar{X}c_i) \right\|$$

ただし、

$$dist_{avg}(C_i, \bar{X}c_j) = \frac{1}{n} \sum_{k=0}^{n_i} dist(x_k, \bar{X}c_j),$$

$\bar{X}c_i$ はトピッククラス C_i に対する SVM 超平面、

$dist(x_k, \bar{X}c_i)$ は SVM 超平面 $\bar{X}c_i$ から点 x_k までの距離、

$dist_{avg}(C_i, \bar{X}c_j)$ は SVM 超平面 $\bar{X}c_j$ からトピッククラス C_i 内の全ての点までの平均距離

(クラス内の場合に正、それ以外の場合に負)

n_i はトピッククラス C_i のトレーニングセットのサイズ

である、請求項8に記載のコンピュータプログラム。

【請求項 1 2】

複数個のトピックに分類された文を含むコーパスの各文を、音声認識のために再度複数個のトピックに分類するための文分類装置としてコンピュータを動作させるコンピュータプログラムであって、

前記文分類装置は、

前記コーパスの前記複数個のトピックに分類された文の集合に対して予め定められた方法により、前記複数個のトピックにそれぞれ対応した複数個の言語モデルを作成するための言語モデル作成手段と、

前記コーパス内の各文と、前記言語モデル作成手段により作成された前記複数個の言語モデルとの間で予め定義される距離を算出するための距離算出手段と、

前記コーパス内の各文を、前記距離算出手段により算出された距離が最も小さい言語モデルに対応するトピックに分類するための分類手段と、

予め定められた条件が満足されるまで、前記言語モデル作成手段、前記距離算出手段、および前記分類手段が繰り返し動作する様に制御するための繰り返し制御手段とを含む、コンピュータプログラム。

【請求項 1 3】

前記制御手段は、前記分類手段による分類結果が収束するまで、前記言語モデル作成手段、前記距離算出手段、および前記分類手段が繰り返し動作する様に制御するための手段を含む、請求項 1 2 に記載のコンピュータプログラム。

【請求項 1 4】

前記制御手段は、所定の回数だけ繰り返しが行なわれるまで、前記言語モデル作成手段、前記距離算出手段、および前記分類手段が繰り返し動作する様に制御するための手段を含む、請求項 1 2 に記載のコンピュータプログラム。

【請求項 1 5】

前記距離算出手段は、文 t と言語モデル M_i との距離 $D_{i s t}(M_i, t)$ を、文 t と言語モデル M_i との間で計算されるパープレキシティ $PP(M_i, t)$ により算出する、請求項 1 2 に記載のコンピュータプログラム。

【手続補正 2】

【補正対象書類名】明細書

【補正対象項目名】0 0 0 7

【補正方法】変更

【補正の内容】

【0 0 0 7】

【非特許文献 1】

S . S n e f f 、 R . L a u 、 J . P o l i f r o n i 著、「Galaxy - II 会話システムの編成、通信、および制御」、ユーロスピーチ予稿集、1999年 (S . S n e f f , R . L a u , J . P o l i f r o n i , "O r g a n i z a t i o n , C o m m u n i c a t i o n , a n d C o n t r o l i n t h e G a l a x y - I I C o n v e r s a t i o n a l S y s t e m " , P r o c . E u r o s p e e c h , 1 9 9 9 .)

【非特許文献 2】

F . W e s s e l 、 A . B a a d e r 著、「1点排除推定を用いたロバストな対話状態依存の言語モデリング」、ICASSP 予稿集第2巻、741 - 744 頁、1999年 (F . W e s s e l , A . B a a d e r , " R o b u s t D i a l o g u e - S t a t e D e p e n d e n t L a n g u a g e M o d e l i n g u s i n g L e a v i n g - O n e - O u t " , P r o c . I C A S S P V o l . 2 , p p . 7 4 1 - 7 4 4 , 1 9 9 9 .)

【非特許文献 3】

G . R i c c a r d i 、 A . G o r i n 、 A . L j o l i e 、 M . R i l e y 著、「自動電話受付のための話し言葉システム」、ICASSP 予稿集、第2巻、1143 - 1146 頁、1997年 (G . R i c c a r d i , A . G o r i n , A . L j o l i e , M . R i l e y ,

"A Spoken Language System for Automated Call Routing", Proc. ICASSP, Vol. 2, pp. 1143 - 1146, 1997)

【非特許文献4】

T. タケザワ、E. スミタ、F. スガヤ、H. ヤマモト、S. ヤマモト著、「実世界における旅行会話の音声翻訳のための大規模バイリンガルコーパスに向けて」、LREC 2002、147-152頁、2002年(T. Takezawa et al. "Towards a Broad-coverage Bilingual Corpus for Speech Translation of Travel Conversation in the Real World", LREC 2002, pp. 147-152, 2002.)

【発明が解決しようとする課題】

しかし、京都大学で開発された方式では、ドメインの数に応じた数の音声認識装置を並列に動作させる必要がある。そのため、ドメインの数が増えるたびに、システムのリソースを追加する必要があり計算コストが増えるという問題がある。また、こうしたシステムでは、ユーザの最初の発話からそれ以後の対話のトピックを同定する必要がある。そのため、できるだけトピック同定の精度を高めるとともに、万が一正確にトピックを同定できないおそれがある場合にも、できるだけ大きな誤りを生じない様なロバスト性が望まれる。特にドメインの切替が生じた場合、こうした問題が生じ易い。

【手續補正3】

【補正対象書類名】明細書

【補正対象項目名】 0018

【補正方法】変更

【補正の内容】

〔 0 0 1 8 〕

この発明の第2の局面に係るコンピュータプログラムは、複数個のトピックに分類された文を含むコーパスから、階層化された言語モデルを作成する方法を実現する様にコンピュータを動作させるコンピュータプログラムである。この方法は、ルートノードと、複数個のトピックに対応する複数個の葉ノードとからなるツリー構造を準備するステップと、複数個の葉ノードのうちの第1のノードを選択するステップと、第1のノードに対応する第1のトピックに対し、予め定められる方法により対応するトピック間で計算される距離が最も小さくなる第2のトピックを求めるステップと、第1のトピックと、第2のトピックとの間で計算された距離が予め定められたしきい値より小さいか否かを判定するステップと、判定するステップで、第1のトピックと第2のトピックとの間の距離が予め定められたしきい値より小さいと判定された事に基づいて、第1のノードと、第2のトピックに対応する第2のノードとに共通する親ノードをツリー構造に追加するステップと、第1のノードを以後のノードの探索の対象から除くステップと、選択するステップと、求めるステップと、判定するステップと、追加するステップと、除くステップとを、予め定められた条件が成立するまで繰り返し実行するステップとを含む。

【手續補正4】

【補正対象書類名】明細書

【補正対象項目名】 0 0 3 4

【補正方法】変更

【補正の内容】

【 0 0 3 4 】

最下層の言語モデル 64 に含まれるトピック毎の言語モデル 100 ~ 122 は各々、トピック分類済コーパス 30 に含まれる文のうち、特定のトピックのタグが付された文の集合から作成された言語モデルである。たとえばトピック毎の言語モデル 100 は「両替」というトピックのタグが付された文の集合から作成された言語モデルであり、トピック毎の言語モデル 102 は「ショッピング」というトピックが付された文の集合から作成され

た言語モデルであり、以下同様である。ただし、後述する様に本実施の形態では、トピック分類済コーパス30の各文にはトピックの再割当処理が行なわれ、再割当がされたトピックに基づいて最下層の言語モデル64内の各言語モデル100～122が作成される。従って、トピック分類済コーパス30で各文に付されたトピックのタグが最下層の言語モデル64に含まれるトピック毎の言語モデル100～122の作成にそのまま使用されるわけではない。

【手続補正5】

【補正対象書類名】明細書

【補正対象項目名】0043

【補正方法】変更

【補正の内容】

【0043】

次に、この第2次のトピック再割当後のコーパス168に含まれる各トピック別の文の集合 $T_{1 \sim N}$ から、言語モデル $M_{1 \sim N}$ からなる第3次の言語モデル群170を作成する事ができる。以下同様にして、引き続き各文にトピックを再割当し、それぞれのトピックに対応した言語モデルを作成する処理を行なう事ができる。こうした処理を、たとえば分類結果が収束するまで繰り返し行なうか、または予め定められた回数だけ行ない、その結果得られたものをトピック再分類済言語モデル142(図3)とする。本実施の形態では、分類結果が収束するまで上記した繰り返しを行なう。

【手続補正6】

【補正対象書類名】明細書

【補正対象項目名】0051

【補正方法】変更

【補正の内容】

【0051】

【数6】

$$Dist_{UNI}(C_i, C_j) = \frac{PP(T_{Ci}, M_{Cj})}{PP(T_{Ci}, M_{Ci})} + \frac{PP(T_{Cj}, M_{Ci})}{PP(T_{Cj}, M_{Cj})}$$

ただし、 T_{Ck} はk番目のトピックに属する文の集合であり、 M_{Ck} はk番目のトピックに属する文の集合から得られたユニグラムモデルであり、 $PP(T_{Ci}, M_{Cj})$ はi番目のトピックに属する文の集合内の各文とj番目のモデル M_{Cj} との間で計算されるパープレキシティである

【手続補正7】

【補正対象書類名】明細書

【補正対象項目名】0066

【補正方法】変更

【補正の内容】

【0066】

-階層的言語モデル作成部32の動作-

階層的言語モデル作成部32は以下の様に動作する。図3を参照して、まずトピック再分類部140がトピック分類済コーパス30内の各文にトピックの再割当を行ない、トピック再分類済言語モデル142を作成する。その詳細は図4および図5に示した。

【手続補正8】

【補正対象書類名】明細書

【補正対象項目名】0072

【補正方法】変更

【補正の内容】

【0072】

続いて図5のステップ186で、第3次の言語モデル群170が収束したかどうか、す

なわち第3次の言語モデル群170が前回の処理で得られた第2次の言語モデル群166と一致しているか否かが判定される。判定結果がNOであれば制御はステップ182に戻り、以下同様の処理が繰り返される。判定結果がYESであればこの処理を終了する。以上の処理で、図3に示すトピック再分類済言語モデル142が得られる。

【手続補正9】

【補正対象書類名】明細書

【補正対象項目名】0073

【補正方法】変更

【補正の内容】

【0073】

続いて、図3に示すトピッククラスタリング部144が、トピック再分類済言語モデル142に含まれる言語モデルをクラスタリングし、階層的言語モデル34を作成する。すなわち、図6を参照して、まず繰り返し変数iに0を代入する（ステップ190）。続いて変数iに1を加算する（ステップ192）。ステップ194で変数iの値（=1）が最大値（本実施の形態では「N」）と等しいか否かが判定される。本実施の形態では図2に示される様にN=12とすれば、繰り返しの1回目ではこの判定結果はNOとなる。従つて、制御はステップ196に進む。

【手続補正10】

【補正対象書類名】明細書

【補正対象項目名】0120

【補正方法】変更

【補正の内容】

【0120】

-トピック検出の精度-

次に、トピックの検出精度について調べた。この精度は、自動音声認識の結果を、元の音声からの書き起こし文に基づいて行なわれた、精度100%の判定結果と比較する事により行なった。ユニグラムおよびSVM方式によるトピック検出精度を表3に示す。

【手続補正11】

【補正対象書類名】図面

【補正対象項目名】図6

【補正方法】変更

【補正の内容】

【図6】

