

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2017-16384
(P2017-16384A)

(43) 公開日 平成29年1月19日(2017.1.19)

(51) Int.Cl. F I テーマコード(参考)
 G06N 3/08 (2006.01) G06N 3/08 5B091
 G06F 17/27 (2006.01) G06F 17/27 615

審査請求 未請求 請求項の数 5 O L (全 18 頁)

(21) 出願番号 特願2015-132347(P2015-132347)
 (22) 出願日 平成27年7月1日(2015.7.1)

(71) 出願人 000004352
 日本放送協会
 東京都渋谷区神南2丁目2番1号
 (74) 代理人 110001807
 特許業務法人磯野国際特許商標事務所
 (72) 発明者 熊野 正
 東京都世田谷区砧一丁目10番11号 日
 本放送協会放送技術研究所内
 Fターム(参考) 5B091 AA15 CA01 EA01

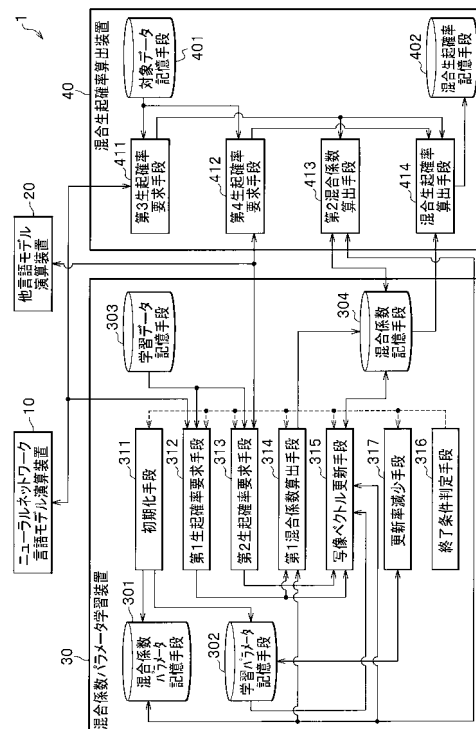
(54) 【発明の名称】 混合係数パラメータ学習装置、混合生起確率算出装置、及び、これらのプログラム

(57) 【要約】 (修正有)

【課題】 混合生起確率の正確性を向上させる混合係数パラメータ学習装置を提供する。

【解決手段】 混合係数パラメータ学習装置30は、ニューラルネットワーク言語モデル演算装置10に隠れ層ベクトル及び生起確率を要求する第1生起確率要求手段312と、他言語モデル演算装置20に生起確率を要求する第2生起確率入力手段313と、隠れ層ベクトルから混合係数を算出する第1混合係数算出手段314と、確率の勾配降下法により、写像ベクトルを更新する写像ベクトル更新手段315と、更新率を減少させる更新率減少手段317と、終了条件を満たすまで、写像ベクトル更新手段315に写像ベクトルを更新させる終了条件判定手段316とを備える。

【選択図】 図2



【特許請求の範囲】**【請求項 1】**

ニューラルネットワーク確率モデルと前記ニューラルネットワーク確率モデル以外の他の確率モデルとのそれぞれで求めた前要素系列に対する次要素の生起確率を混合するときの混合係数の算出に必要なパラメータを学習する混合係数パラメータ学習装置であって、

前記ニューラルネットワーク確率モデルの隠れ層ベクトルと、前記ニューラルネットワーク確率モデルで求めた生起確率とが入力される第 1 生起確率入力手段と、

前記他の確率モデルで求めた生起確率が入力される第 2 生起確率入力手段と、

予め設定された写像ベクトルにより前記隠れ層ベクトルを実数値のスカラに線形写像し、前記実数値のスカラをシグモイド関数で非線形変換することで、前記混合係数を算出する第 1 混合係数算出手段と、

前記ニューラルネットワーク確率モデルと前記他の確率モデルとのそれぞれで求めた生起確率、前記混合係数、及び、予め設定された更新率を用いた確率的勾配降下法により、前記パラメータとしての前記写像ベクトルを更新する写像ベクトル更新手段と、

予め設定された更新率減少規則に従って前記更新率を減少させる更新率減少手段と、

予め設定された終了条件を満たすか否かを判定し、前記終了条件を満たすまで、減少させた前記更新率で前記写像ベクトル更新手段に前記写像ベクトルを更新させる終了条件判定手段と、

を備えることを特徴とする混合係数パラメータ学習装置。

【請求項 2】

前記写像ベクトル更新手段は、前記写像ベクトルを更新するときに正則化を行うことを特徴とする請求項 1 に記載の混合係数パラメータ学習装置。

【請求項 3】

ニューラルネットワーク確率モデルと前記ニューラルネットワーク確率モデル以外の他の確率モデルとのそれぞれで求めた前要素系列に対する次要素の生起確率を混合した混合生起確率を算出する混合生起確率算出装置であって、

前記ニューラルネットワークの隠れ層ベクトルと、前記ニューラルネットワーク確率モデルで求めた生起確率とが入力される第 3 生起確率入力手段と、

前記他の確率モデルで求めた生起確率が入力される第 4 生起確率入力手段と、

請求項 1 に記載の混合係数パラメータ学習装置が学習した写像ベクトルで前記隠れ層ベクトルを実数値のスカラに線形写像し、前記実数値のスカラをシグモイド関数で非線形変換することで、混合係数を算出する第 2 混合係数算出手段と、

前記混合係数を用いて、前記ニューラルネットワーク確率モデルと前記他の確率モデルとで求めた前記次要素の生起確率を混合することで、前記混合生起確率を算出する混合生起確率算出手段と、

を備えることを特徴とする混合生起確率算出装置。

【請求項 4】

コンピュータを、請求項 1 又は請求項 2 に記載の混合係数パラメータ学習装置として機能させるための混合係数パラメータ学習プログラム。

【請求項 5】

コンピュータを、請求項 3 に記載の混合生起確率算出装置として機能させるための混合生起確率算出プログラム。

【発明の詳細な説明】**【技術分野】****【0001】**

本願発明は、混合係数の算出に必要なパラメータを学習する混合係数パラメータ学習装置、ニューラルネットワーク確率モデルと他の確率モデルとの混合生起確率を算出する混合生起確率算出装置、及び、これらのプログラムに関する。

【背景技術】**【0002】**

10

20

30

40

50

統計的言語モデル（以後、「言語モデル」とは、ある言語又はドメインにおいて、単語系列 $w_1 w_2 \dots w_n$ が生起する確率 $p(w_1 w_2 \dots w_n)$ を計算する手段、及び、その手段によって計算するのに必要な各種統計量の一覧として定義される。言語モデルによって言語の生起を確率モデル化することは、統計的自然言語処理の最も基本的な技術の一つであり、音声認識、機械翻訳をはじめとする各種自然言語処理技術に用いられている。

【0003】

なお、言語モデルとは、ある言語、又は、その言語の特定分野における表現（単語系列）生起の確率モデルのことであり、一般的には予め与えられた当該言語又はその言語の当該分野のコーパスから学習する。

また、コーパスとは、ある言語又はその言語の特定分野で観測された単語系列の実例である。

また、 w_1, w_2, w_n は、単語を表す。

【0004】

単語系列の生起確率 $p(w_1 w_2 \dots w_n)$ は、一般的には、系列の各単語がそれ以前の単語系を前文脈として生起する確率の積、すなわち、 $p(w_1) \times p(w_2 | w_1) \times p(w_3 | w_1 w_2) \times \dots \times p(w_n | w_1 w_2 \dots w_{n-1})$ としてモデル化される。つまり、言語モデルは、前文脈が与えられた条件下での次単語生起の予測モデルであると言える。

【0005】

言語モデルの最も一般的な実現手法は、 n -gram 言語モデルである。この n -gram 言語モデルは、前記条件となる前文脈を直近の $n-1$ 単語に制限し（但し、 n は 1 以上の整数）、学習コーパスから、 $n-1$ 単語の列である前文脈の異なり毎に次単語生起頻度を収集した結果に基づき、各前文脈条件下の次単語生起確率を推定するものである。

【0006】

n -gram 言語モデルでは、精度よく次単語の生起確率を推定するために長い前文脈を参照する（大きな値の n を用いる）必要がある。また、 n -gram 言語モデルでは、各前文脈に対して十分な実例を集める必要があるが、長い前文脈を用いるほど前文脈の異なりが増加するため、正確性を向上させるために非常に大きな学習コーパスを用意する必要がある。

【0007】

近年、この n -gram 言語モデルに対して、ニューラルネットワークを用いた言語モデル実現手法が提案されている。この手法は、ニューラルネットワークを用いて、各単語を表すものとして、固定次元で各次元が実数値である単語表現ベクトルへの写像を学習し、前文脈として単語列の各単語に対応する単語表現ベクトルの組み合わせを用いるものである。

【0008】

例えば、非特許文献 1 に記載の NNLM (Neural Network Language Model) は、図 5 のようなニューラルネットワークを構築する。以後、言語モデルについて、有限個 $|V|$ 種類の単語のみを扱うものとし、各単語は $1 \sim |V|$ の数値として表すことにする。このとき、 $|V|$ 種類の単語の中には、必ず文頭を表す特殊な単語を含むものとする。ここで、各単語 w に対応する予め定めた固定次元数 m の単語表現ベクトルを $C(w)$ とする。また、単語系列 $w_1 w_2 \dots w_t$ の生起に関して、単語 w_t の $n-1$ 個の前文脈を表す $n-1$ 個の単語表現ベクトルを連結した $n \times m$ 次元の入力ベクトル $x(t) = [C(w_{t-n+1}), \dots, C(w_{t-2}), C(w_{t-1})]$ から予め定めた固有次元数 h のベクトルへの線形写像を $Hx(t)$ とする。

なお、前文脈の長さが $n-1$ 未満である（すなわち $t < n$ である）場合には、単語 w_1 の前に $n-t$ 個の文頭を表す単語を補うことで入力ベクトル $x(t)$ を作成するものとする。

【0009】

また、線形写像 $Hx(t)$ の各次元を非線形関数 f （例えば、双曲線正接関数 \tanh

10

20

30

40

50

)で変換した隠れ層ベクトル $z(t)$ から $|V|$ 次元ベクトル $y(t)$ への線形写像を $Uz(t)$ とする。

また、 $y(t)$ の各次元を式(1)に示す関数で変換した $|V|$ 次元ベクトルを出力ベクトル $p(t)$ とする。この場合、次単語が w_t である確率を以下の式(1)~式(3)のように定義する(但し、 y_i は y の i 次元の値)。

また、入力ベクトル $x(t)$ の (t) は、前文脈 $w_1 w_2 \dots w_{t-1}$ に後続する次単語 w_t の生起確率に関わる入力ベクトル x を意味する(他のベクトルも同様)。

また、図5の「」はベクトルの要素を表す。

【0010】

【数1】

$$p(w_t | w_{t-n+1} \dots w_{t-2} w_{t-1}) = \frac{e^{y w_t(t)}}{\sum_i e^{y_i(t)}} \quad \dots \text{式(1)}$$

10

【0011】

【数2】

$$y(t) = Uz(t) \quad \dots \text{式(2)}$$

【0012】

【数3】

$$z(t) = f(Hx(t)) \quad \dots \text{式(3)}$$

20

【0013】

n, m, h を予め設定し、学習コーパスの各単語 w_t に対して、前文脈 $w_{t-n+1}, \dots, w_{t-2}, w_{t-1}$ をニューラルネットワークに入力して次単語生起の確率分布を出力(順方向伝搬)し、出力ベクトルと正解ベクトルとの交差エントロピー誤差をニューラルネットワークに逆方向伝搬させながら、以下の式(4)~式(6)のように単語表現ベクトル C 、入力層から隠れ層への重み H 、隠れ層から出力層への重み U を確率的勾配降下法により更新する(但し、 ϵ は更新率)。これを学習コーパス全体で何回か繰り返すことによって学習を実現する。

なお、正解ベクトルとは、単語 w_t の生起確率を1とし、それ以外の単語の生起確率を0としたベクトルである。

30

【0014】

【数4】

$$U \leftarrow U + \epsilon \frac{\partial \log p(w_t | w_{t-n+1} \dots w_{t-2} w_{t-1})}{\partial U} \quad \dots \text{式(4)}$$

【0015】

【数5】

$$H \leftarrow H + \epsilon \frac{\partial \log p(w_t | w_{t-n+1} \dots w_{t-2} w_{t-1})}{\partial H} \quad \dots \text{式(5)}$$

40

【0016】

【数6】

$$C \leftarrow C + \epsilon \frac{\partial \log p(w_t | w_{t-n+1} \dots w_{t-2} w_{t-1})}{\partial C} \quad \dots \text{式(6)}$$

【0017】

単語表現ベクトル C の学習の結果、類似した単語が近い単語表現ベクトルに写像され、 H の学習の結果、類似した単語が近い隠れ層ベクトルに写像されるので、小規模な学習コーパスから学習した場合でも高い正確性を得ることができる。

50

【 0 0 1 8 】

また、NNLMとは異なる手法として、非特許文献2に記載のRNNLM (Recurrent Neural Network Language Model) が提案されている。前記したように、NNLMでは、単語 w_t に対して予め定めた $n - 1$ 個の単語 $w_{t-n+1}, \dots, w_{t-2}, w_{t-1}$ から前文脈を表す隠れ層ベクトル $z(t)$ を計算する。一方、このRNNLMは、図6のように、隠れ層ベクトル $z(t)$ を、1つ前の単語 w_{t-1} 及びこの単語 w_{t-1} に対する前文脈を表す隠れ層ベクトル $z(t-1)$ から計算する。これにより、RNNLMは、明示的な前文脈長 n を与えることなく、長い前文脈を反映した次単語生起の予測を行うことを可能にする。

【 0 0 1 9 】

10

一般的には、これらニューラルネットワーク言語モデルは、他の言語モデル (例えば、 n -gram言語モデル) と組み合わせて用いられる。具体的には、ニューラルネットワーク言語モデルによる生起確率を p_N 、他の言語モデルによる生起確率を p_0 、混合比率を λ とする。この場合、以下の式(7)のように、 λ : $1 - \lambda$ の比率で両言語モデルの生起確率 p_N 、 p_0 を混合したものを混合生起確率 p として算出する。

【 0 0 2 0 】

【数7】

$$p(w_t | \dots w_{t-1}) = \lambda p_N(w_t | \dots w_{t-1}) + (1 - \lambda) p_0(w_t | \dots w_{t-1}) \quad \dots \text{式 (7)}$$

【 0 0 2 1 】

20

なお、式(7)では、 λ が混合係数を表す。通常、混合係数 λ は、学習済みの両言語モデルを用意した上で、別途用意したテストコーパスに対して最も正確性が高くなる値を決定し、決定した値を固定的に用いる。

【 0 0 2 2 】

このように生起確率を混合するには、以下のような理由がある。

1) ニューラルネットワーク言語モデルでは学習コーパスに表れなかった単語 (未知語) の生起確率を推定する一般的手法がないが、 n -gram言語モデルでは未知語に適切な生起確率を割り当てることができる。

2) ニューラルネットワーク言語モデルは学習のための計算量が n -gram言語モデルと比較して非常に大きいため、ニューラルネットワーク言語モデルで小規模のドメインに特化した学習コーパスを用いて学習し、より広範囲の大規模な学習コーパスから学習した n -gram言語モデルと組み合わせることが現実的である。

30

【先行技術文献】

【非特許文献】

【 0 0 2 3 】

【非特許文献1】A Neural Probabilistic Language Model, Yoshua Bengio et.al, Journal of Machine Learning Research 3, (2003), 1137-1155

【非特許文献2】Static Language Model based on Neural Network, Tomas Mikolov

【発明の概要】

【発明が解決しようとする課題】

40

【 0 0 2 4 】

しかし、ニューラルネットワーク言語モデルでは、前文脈に関係なく固定的な混合係数を用いているため、混合生起確率の正確性が低くなるという問題がある。例えば、前文脈「私は」の次に表れる単語を正確に予測するためには巨大な学習コーパスで学習する必要がある。一方、前文脈「私は山にいきまし」の次に表れる単語は「た」「て」くらいしかあり得ず、小規模な学習コーパスで学習しても正確に予測可能である。すなわち、ニューラルネットワーク言語モデルでは、前文脈に応じて異なる混合係数を用いることで、混合生起確率の正確性を向上させることができる。

【 0 0 2 5 】

本願発明は、前記した課題に鑑みて、混合生起確率の正確性を向上させる混合係数パラ

50

メータ学習装置、混合生起確率算出装置、及び、これらのプログラムを提供することを課題とする。

【課題を解決するための手段】

【0026】

前記した課題に鑑みて、本願発明に係る混合係数パラメータ学習装置は、ニューラルネットワーク確率モデルと前記ニューラルネットワーク確率モデル以外の他の確率モデルとのそれぞれで求めた前要素系列に対する次要素の生起確率を混合するときの混合係数の算出に必要なパラメータを学習する混合係数パラメータ学習装置であって、第1生起確率入力手段と、第2生起確率入力手段と、第1混合係数算出手段と、写像ベクトル更新手段と、更新率減少手段と、終了条件判定手段とを備える構成とした。

10

【0027】

かかる構成によれば、混合係数パラメータ学習装置は、第1生起確率入力手段によって、前記ニューラルネットワーク確率モデルの隠れ層ベクトルと、前記ニューラルネットワーク確率モデルで求めた生起確率とが入力される。

【0028】

すなわち、学習済みのニューラルネットワーク確率モデルに前要素系列を入力すると、ニューラルネットワーク確率モデルの隠れ層ベクトルとして、汎化された前要素の表現が得られる。従って、学習済みのニューラルネットワーク確率モデルの隠れ層ベクトルから混合係数への写像ベクトルを学習すれば、前要素系列に応じた混合係数を求めることが可能となる。

20

【0029】

混合係数パラメータ学習装置は、第2生起確率入力手段によって、前記他の確率モデルで求めた生起確率が入力される。

混合係数パラメータ学習装置は、第1混合係数算出手段によって、予め設定された写像ベクトルにより前記隠れ層ベクトルを実数値のスカラに線形写像し、前記実数値のスカラをシグモイド関数で非線形変換することで、前記混合係数を算出する。

【0030】

混合係数パラメータ学習装置は、写像ベクトル更新手段によって、前記ニューラルネットワーク確率モデルと前記他の確率モデルとのそれぞれで求めた生起確率、前記混合係数、及び、予め設定された更新率を用いた確率的勾配降下法により、前記パラメータとしての前記写像ベクトルを更新する。

30

【0031】

混合係数パラメータ学習装置は、更新率減少手段によって、予め設定された更新率減少規則に従って前記更新率を減少させる。

混合係数パラメータ学習装置は、終了条件判定手段によって、予め設定された終了条件を満たすか否かを判定し、前記終了条件を満たすまで、減少させた前記更新率で前記写像ベクトル更新手段に前記写像ベクトルを更新させる。例えば、この終了条件は、更新率を減少させても、生起確率が変化しないという条件である。

【0032】

また、前記した課題に鑑みて、本願発明に係る混合生起確率算出装置は、ニューラルネットワーク確率モデルと前記ニューラルネットワーク確率モデル以外の他の確率モデルとのそれぞれで求めた前要素系列に対する次要素の生起確率を混合した混合生起確率を算出する混合生起確率算出装置であって、第3生起確率入力手段と、第4生起確率入力手段と、第2混合係数算出手段と、混合生起確率算出手段とを備える構成とした。

40

【0033】

かかる構成によれば、混合生起確率算出装置は、第3生起確率入力手段によって、前記ニューラルネットワークの隠れ層ベクトルと、前記ニューラルネットワーク確率モデルで求めた生起確率とが入力される。

混合生起確率算出装置は、第4生起確率入力手段によって、前記他の確率モデルで求めた生起確率が入力される。

50

【 0 0 3 4 】

混合生起確率算出装置は、第2混合係数算出手段によって、本願発明に係る混合係数パラメータ学習装置が学習した写像ベクトルで前記隠れ層ベクトルを実数値のスカラに線形写像し、前記実数値のスカラをシグモイド関数で非線形変換することで、前要素系列に応じた混合係数を算出する。

【 0 0 3 5 】

混合生起確率算出装置は、混合生起確率算出手段によって、前要素系列に応じた前記混合係数を用いて、前記ニューラルネットワーク確率モデルと前記他の確率モデルとで求めた前記次要素の生起確率を混合することで、前記混合生起確率を算出する。

【 発明の効果 】

10

【 0 0 3 6 】

本願発明は、以下のような優れた効果を奏する。

本願発明によれば、学習済みのニューラルネットワーク確率モデルの隠れ層ベクトルから混合係数への写像ベクトルを学習する。これにより、前要素系列に応じた混合係数が求められるので、混合生起確率の正確性を向上させることができる。

【 図面の簡単な説明 】

【 0 0 3 7 】

【 図 1 】本願発明における写像ベクトルの学習手順を説明する説明図である。

【 図 2 】本願発明の実施形態に係る混合生起確率算出システムの構成を示すブロック図である。

20

【 図 3 】図 2 の混合係数パラメータ算出装置の動作を示すフローチャートである。

【 図 4 】図 2 の混合生起確率算出装置の動作を示すフローチャートである。

【 図 5 】従来の N N L M の処理概要を説明する説明図である。

【 図 6 】従来の R N N L M の処理概要を説明する説明図である。

【 発明を実施するための形態 】

【 0 0 3 8 】

以下、本願発明の実施形態に係る混合生起確率算出システム 1 について説明する。

最初に、図 1 を参照し、写像ベクトルの学習手順と、混合生起確率の算出手順とを説明する。その後、混合生起確率算出システム 1 の構成について説明する。

【 0 0 3 9 】

30

ここで、前文脈（前要素系列）を表現した h 次元の隠れ層ベクトルを持ち、この隠れ層ベクトルからの写像として次単語の生起確率を演算するニューラルネットワーク言語モデル演算装置 10 があることとする。このニューラルネットワーク言語モデル演算装置 10 は、生起確率の計算に必要な各統計量が学習コーパス等により学習済みであることとする。

【 0 0 4 0 】

また、 n -gram 言語モデル等の他の言語モデルで次単語の生起確率を推定する他言語モデル演算装置 20（図 2）があることとする。この他言語モデル演算装置 20 では、生起確率の計算に必要な各統計量が学習コーパス等により学習済みであることとする。

なお、他言語モデル演算装置 20 の各統計量を学習するために用いる学習コーパス等は、前記のニューラルネットワーク言語モデル演算装置 10 の各統計量を学習するために用いた学習コーパスと同一である必要はない。

40

【 0 0 4 1 】

< 写像ベクトルの学習手順 >

ニューラルネットワーク言語モデル演算装置 10 に前文脈を入力すると、隠れ層ベクトルとして、汎化された前文脈の表現が得られる。そこで、本願発明は、図 1 のように、ニューラルネットワーク言語モデル演算装置 10 に、隠れ層ベクトル z から混合係数 w への写像を行う処理を追加し、この写像ベクトル s を学習することとする。

【 0 0 4 2 】

具体的には、ニューラルネットワーク言語モデルによる処理に以下の式（8）及び式（

50

9) の処理を加えることで、前文脈... w_{t-1} が与えられたときの次単語 w_t の生起確率の算出に必要な混合係数 $\lambda(t)$ を計算するようにする。

【0043】

【数8】

$$\lambda(t) = \frac{1}{1+e^{-s(t)}} \quad \dots \text{式 (8)}$$

【0044】

【数9】

$$s(t) = b + Sz(t) \quad \dots \text{式 (9)}$$

10

【0045】

なお、式(8)は、実数値のスカラー $s(t)$ から混合係数 $\lambda(t)$ へのシグモイド関数による非線形変換を表している。

また、式(9)は、隠れ層ベクトル $z(t)$ から実数値のスカラー $s(t)$ への線形写像 $Sz(t)$ を表している。また、式(9)では、 b がバイアス値を表している。

【0046】

写像ベクトル S 及びバイアス値 b の学習は、何らかの学習コーパス中の各単語 w_t について、生起確率 $p_N(w_t | \dots w_{t-1})$ と生起確率 $p_O(w_t | \dots w_{t-1})$ とを式(8)で定義される混合係数 $\lambda(t)$ を用いて、以下の式(10)に従って混合した混合生起確率 $p(w_t | \dots w_{t-1})$ が最大となるように、以下の手順1~手順3で行う。

20

【0047】

【数10】

$$p(w_t | \dots w_{t-1}) = \lambda(t)p_N(w_t | \dots w_{t-1}) + (1 - \lambda(t))p_O(w_t | \dots w_{t-1})$$

…式(10)

【0048】

なお、学習コーパスは、ニューラルネットワーク言語モデルや別の言語モデルの学習に用いたコーパスと同じもの、又は、そのコーパスと異なるものでもよい。

また、生起確率 $p_N(w_t | \dots w_{t-1})$ は、前文脈... w_{t-1} をニューラルネットワーク言語モデルに与えて得られる次単語 w_t の生起確率である。

30

また、生起確率 $p_O(w_t | \dots w_{t-1})$ は、前文脈... w_{t-1} を他の言語モデルに与えて得られる次単語 w_t の生起確率である。

【0049】

手順1. 更新率 α を予め設定する。

手順2. 学習コーパス中の各単語 w_t に対して以下の(a)~(c)の処理を実行する。

(a) ニューラルネットワーク言語モデル演算装置10に適宜(NNMLのように前文脈が固定されている場合、その長さで区切った)前文脈... w_{t-1} を入力して順方向伝搬を行うことで、隠れ層ベクトル $z(t)$ 及び次単語 w_t の生起確率 $p_N(w_t | \dots w_{t-1})$ を求める。同様に、他の言語モデルの生起確率 $p_O(w_t | \dots w_{t-1})$ を求める。

40

【0050】

(b) 隠れ層から順方向伝搬を行うことで、混合係数 $\lambda(t)$ を求める。すなわち、式(8)及び式(9)を用いて、混合係数 $\lambda(t)$ を求める。

【0051】

(c) 確率的勾配降下法により写像ベクトル S を更新する。すなわち、 h 次元の写像ベクトル S の各次元 S_i を、以下の式(11)及び式(12)のように混合係数 $\lambda(t)$ が反映された確率的勾配降下法により更新する。

【0052】

【数 1 1】

$$S_i \leftarrow S_i + \varepsilon \frac{\partial \log p(w_t | \dots w_{t-1})}{\partial S_i} \quad \dots \text{式 (1 1)}$$

【0 0 5 3】

【数 1 2】

$$\frac{\partial \log p(w_t | \dots w_{t-1})}{\partial S_i} = z_i(t) \frac{p_N(w_t | \dots w_{t-1}) - p_O(w_t | \dots w_{t-1})}{p(w_t | \dots w_{t-1})} \lambda(t)(1 - \lambda(t))$$

…式 (1 2)

10

【0 0 5 4】

さらに、式 (9) のバイアス値 b も学習の対象となる。このため、以下の式 (1 3) 及び式 (1 4) のようにバイアス値 b も更新する。

【0 0 5 5】

【数 1 3】

$$b \leftarrow b + \varepsilon \frac{\partial \log p(w_t | \dots w_{t-1})}{\partial b} \quad \dots \text{式 (1 3)}$$

【0 0 5 6】

【数 1 4】

$$\frac{\partial \log p(w_t | \dots w_{t-1})}{\partial b} = \frac{p_N(w_t | \dots w_{t-1}) - p_O(w_t | \dots w_{t-1})}{p(w_t | \dots w_{t-1})} \lambda(t)(1 - \lambda(t))$$

…式 (1 4)

20

【0 0 5 7】

手順 2 (c) において、写像ベクトル S を更新する際、ニューラルネットワーク言語モデルによる過学習を防止するため、一例として式 (1 5) のように、正則化を行ってもよい。さらに、バイアス値 b についても、写像ベクトル S と同様、正則化を行ってもよい。

なお、式 (1 5) では、 β が正則化係数を表す。例えば、正則化係数 β は、更新率 ε より小さな値とする。

30

【0 0 5 8】

【数 1 5】

$$S_i \leftarrow S_i + \varepsilon \frac{\partial \log p(w_t | \dots w_{t-1})}{\partial S_i} - \beta S_i \quad \dots \text{式 (1 5)}$$

【0 0 5 9】

手順 3 . 所定の終了条件に合致するまで、手順 2 に戻って処理を繰り返す。このとき、所定の更新率減少規則に従って、更新率 ε を減少させる。

なお、終了条件及び更新率減少規則の詳細は、後記する。

【0 0 6 0】

40

< 混合生起確率の算出手順 >

次単語 w_t の生起確率の計算は、前記した学習結果を用いて、以下の手順 4 ~ 手順 6 で行う。

【0 0 6 1】

手順 4 . ニューラルネットワーク言語モデル演算装置に適宜 (NNML のように前文脈が固定されている場合、その長さで区切った) 前文脈 $\dots w_{t-1}$ を入力して順方向伝搬を行うことで、隠れ層ベクトル $z(t)$ 及び次単語 w_t の生起確率 $p_N(w_t | \dots w_{t-1})$ を求める。同様に、他の言語モデルの生起確率 $p_O(w_t | \dots w_{t-1})$ を求める。

なお、この手順 4 は、写像ベクトル S の学習手順 2 (a) と同じ処理である。

【0 0 6 2】

50

手順 5 . 隠れ層から順方向伝搬を行うことで、混合係数 $\lambda(t)$ を求める。すなわち、学習した写像ベクトル S 及びバイアス値 b を式 (9) に代入して、混合係数 $\lambda(t)$ を求める。なお、この手順 5 は、写像ベクトル S の学習手順 2 (b) と同じ処理である。

手順 6 . 式 (16) を用いて、混合生起確率 $p(w_t | \dots w_{t-1})$ を求める。

【0063】

【数16】

$$p(w_t | \dots w_{t-1}) = \lambda(t)p_N(w_t | \dots w_{t-1}) + (1 - \lambda(t))p_O(w_t | \dots w_{t-1})$$

…式 (16)

【0064】

図 2 を参照し、本願発明の実施形態に係る混合生起確率算出システム 1 の構成について説明する。

【0065】

混合生起確率算出システム 1 は、ニューラルネットワーク言語モデルで求めた生起確率 p_N と、他の言語モデルで求めた生起確率 p_O とを混合した混合生起確率 P を算出するものである。図 2 のように、混合生起確率算出システム 1 は、ニューラルネットワーク言語モデル演算装置 10 と、他言語モデル演算装置 20 と、混合係数パラメータ学習装置 30 と、混合生起確率算出装置 40 とを備える。

【0066】

[ニューラルネットワーク言語モデル演算装置の構成]

ニューラルネットワーク言語モデル演算装置 10 は、ニューラルネットワーク言語モデルにより、生起確率 p_N を演算するものである。例えば、ニューラルネットワーク言語モデル演算装置 10 は、隠れ層を用いるニューラルネットワーク (例えば、NNLM、RNNLM) を用いることができる。

【0067】

具体的には、ニューラルネットワーク言語モデル演算装置 10 は、前文脈 w_1, w_2, \dots, w_{t-1} が入力されると、当該前文脈に後続する単語 w_t の生起確率 $p_N(w_t | \dots w_{t-1})$ を演算する。また、ニューラルネットワーク言語モデル演算装置 10 は、ニューラルネットワークの出力層ベクトル $p(t)$ を演算する際、ニューラルネットワークの入力層ベクトル $x(t)$ から算出した隠れ層ベクトル $z(t)$ を記憶し、記憶した隠れ層ベクトル $z(t)$ を混合係数パラメータ学習装置 30 又は混合生起確率算出装置 40 に出力する。

【0068】

NNLM の場合、ニューラルネットワーク言語モデル演算装置 10 は、参照可能な前文脈の長さが、前文脈の末尾から所定の単語数 $n-1$ までに限定される (n は 1 以上の整数)。

例えば、前文脈が w_1, w_2, \dots, w_{t-1} の場合、参照可能な前文脈が $w_{t-n+1}, w_{t-n+1}, \dots, w_{t-1}$ となる。

ニューラルネットワーク言語モデル演算装置 10 は、入力された前文脈の各単語に対応した単語表現ベクトル $C(w)$ を記憶し、長さ $n-1$ の前文脈 $w_{t-n+1}, w_{t-n+1}, \dots, w_{t-1}$ が入力されると、その各単語に応じた単語表現ベクトル $C(w)$ を連結してニューラルネットワークの入力層ベクトル $x(t)$ に設定する。そして、ニューラルネットワーク言語モデル演算装置 10 は、順方向伝搬を行い、ニューラルネットワークの隠れ層ベクトル $z(t)$ 及び出力層ベクトル $p(t)$ を算出する。

出力層ベクトル $p(t)$ は、単語の異なり数の次元を持つベクトルであり、ベクトルの各次元の値がその次元に対応した単語の生起確率を表す。なお、隠れ層ベクトル $z(t)$ を「前文脈 w_1, w_2, \dots, w_{t-1} の隠れ層表現」と呼ぶ。

【0069】

RNNLM の場合、ニューラルネットワーク言語モデル演算装置 10 は、内部にこれまで順に入力された単語系列 w_1, w_2, \dots を前文脈とする隠れ層ベクトル z を記憶してい

10

20

30

40

50

る。初期状態において、隠れ層ベクトル z は、ニューラルネットワーク言語モデル演算装置 10 に固有の初期値に設定される。

ニューラルネットワーク言語モデル演算装置 10 は、 i 番目の単語 w_i を入力すると、当該単語 w_i に対応した次元のみが 1 で、他のすべての次元が 0 であるベクトルを入力層 $x(i)$ に設定する。そして、ニューラルネットワーク言語モデル演算装置 10 は、入力層 $x(i)$ 及び記憶している前入力 of 隠れ層ベクトル $z(i)$ から順方向伝搬を行い、ニューラルネットワークの隠れ層ベクトル $z(i+1)$ 及び出力層ベクトル $p(i+1)$ を算出する。単語 w_1, w_2, \dots, w_{t-1} までの入力及び順方向伝搬が終了したとき、隠れ層ベクトル $z(t)$ は、前記した NNLM を用いた場合の「前文脈 w_1, w_2, \dots, w_{t-1} の隠れ層表現」と同様のものになる。すなわち、ニューラルネットワーク言語モデル演算装置 10 は、隠れ層ベクトル $z(t)$ を用いた順方向伝搬により出力層ベクトル $p(t)$ を算出し、次単語の生起確率 p_N を求める。

【0070】

なお、ニューラルネットワーク言語モデル演算装置 10 は、学習済みであり（学習データにより順方向伝搬のための写像行列が適切な値に設定済みであり）、その学習結果が記憶されていることとする。

また、ニューラルネットワーク言語モデル演算装置 10 は、一般的な構成のため、これ以上の説明を省略する。

【0071】

[他言語モデル演算装置の構成]

他言語モデル演算装置 20 は、ニューラルネットワーク言語モデル以外の他の言語モデル（例えば、 n -gram 言語モデル）により、生起確率 p_0 を演算するものである。具体的には、他言語モデル演算装置 20 は、前文脈 w_1, w_2, \dots, w_{t-1} が入力されると、当該前文脈に後続する任意の単語 w_t の生起確率 $p_0(w_t | \dots w_{t-1})$ を演算して出力する。

【0072】

なお、他言語モデル演算装置 20 は、確率値の計算に必要な各種パラメータが予め設定されていることとする。

また、他言語モデル演算装置 20 は、一般的な構成のため、これ以上の説明を省略する。

【0073】

[混合係数パラメータ学習装置の構成]

混合係数パラメータ学習装置 30 は、ニューラルネットワーク言語モデルと他の言語モデルとのそれぞれで求めた生起確率 p_N, p_0 を混合するときの混合係数の算出に必要なパラメータを学習するものである。

【0074】

図 2 のように、混合係数パラメータ学習装置 30 は、混合係数パラメータ記憶手段 301 と、学習パラメータ記憶手段 302 と、学習データ記憶手段 303 と、混合係数記憶手段 304 と、初期化手段 311 と、第 1 生起確率要求手段（第 1 生起確率入力手段）312 と、第 2 生起確率要求手段（第 2 生起確率入力手段）313 と、第 1 混合係数算出手段 314 と、写像ベクトル更新手段 315 と、終了条件判定手段 316 と、更新率減少手段 317 とを備える。

【0075】

混合係数パラメータ記憶手段 301 は、混合係数の算出に必要な混合係数パラメータを記憶するメモリ、ハードディスク等の記憶手段である。具体的には、混合係数パラメータ記憶手段 301 は、写像ベクトル S 、バイアス値 b 等の混合係数パラメータを記憶する。この写像ベクトル S は、ニューラルネットワークの隠れ層ベクトル z の次元数 h と同一次元数である。

【0076】

学習パラメータ記憶手段 302 は、写像ベクトル S の学習に必要なパラメータを記憶す

10

20

30

40

50

るメモリ、ハードディスク等の記憶手段である。具体的には、学習パラメータ記憶手段 302 は、更新率、正則化係数等の学習パラメータを記憶する。

【0077】

学習データ記憶手段 303 は、写像ベクトル S の学習に必要な学習データである単語列を記憶するメモリ、ハードディスク等の記憶手段である。この学習データは、ニューラルネットワーク言語モデル演算装置 10 及び他言語モデル演算装置 20 で学習に用いたものと同一でなくともよい。

混合係数記憶手段 304 は、混合係数を記憶するメモリ、ハードディスク等の記憶手段である。

【0078】

初期化手段 311 は、混合係数パラメータ及び学習パラメータの初期化を行うものである。具体的には、初期化手段 311 は、混合係数パラメータ記憶手段 301 の写像ベクトル S の各次元の値、および、バイアス値 b を乱数で初期化する。また、初期化手段 311 は、学習パラメータ記憶手段 302 の更新率及び正則化係数を予め設定した値で初期化する。

【0079】

第 1 生起確率要求手段 312 は、学習データ記憶手段 303 の前文脈をニューラルネットワーク言語モデル演算装置 10 に出力することで、隠れ層ベクトル z 及び生起確率 p_N を要求するものである。この要求に応じて、第 1 生起確率要求手段 312 は、ニューラルネットワーク言語モデル演算装置 10 から、隠れ層ベクトル z 及び生起確率 p_N が入力される。そして、第 1 生起確率要求手段 312 は、入力された隠れ層ベクトル z 及び生起確率 p_N を第 1 混合係数算出手段 314 及び写像ベクトル更新手段 315 に出力する。

【0080】

第 2 生起確率要求手段 313 は、学習データ記憶手段 303 の前文脈を他言語モデル演算装置 20 に出力することで、生起確率 p_O を要求するものである。ここで、第 2 生起確率要求手段 313 は、第 1 生起確率要求手段 312 と同一の前文脈を他言語モデル演算装置 20 に出力する。この要求に応じて、第 2 生起確率要求手段 313 は、他言語モデル演算装置 20 から、生起確率 p_O が入力される。そして、第 2 生起確率要求手段 313 は、入力された生起確率 p_O を写像ベクトル更新手段 315 に出力する。

【0081】

第 1 混合係数算出手段 314 は、式 (9) を用いて、混合係数パラメータ記憶手段 301 の写像ベクトル S により、第 1 生起確率要求手段 312 から入力された隠れ層ベクトル z を実数値のスカラ s に線形写像するものである。また、第 1 混合係数算出手段 314 は、式 (8) を用いて、実数値のスカラ s をシグモイド関数で非線形変換することで、混合係数を算出する。そして、第 1 混合係数算出手段 314 は、算出した混合係数を混合係数記憶手段 304 に記憶する。

【0082】

写像ベクトル更新手段 315 は、第 1 生起確率要求手段 312 からの生起確率 p_N 、第 2 生起確率要求手段 313 からの生起確率 p_O 、混合係数記憶手段 304 の混合係数、及び、学習パラメータ記憶手段 302 の更新率を用いた確率的勾配降下法により、混合係数記憶手段 304 の写像ベクトル S を更新するものである。つまり、写像ベクトル更新手段 315 は、式 (11) 及び式 (12) で表される確率的勾配降下法を用いて、写像ベクトル S を更新する。

【0083】

終了条件判定手段 316 は、予め設定された終了条件を満たすか否かを判定し、この終了条件を満たすまで、後記する更新率減少手段 317 が減少させた更新率で写像ベクトル更新手段 315 に写像ベクトル S を更新させるものである。例えば、終了条件判定手段 316 は、予め設定した回数だけ更新率を減少させて混合生起確率 p の値が変化しなかった場合、終了条件を満たすと判定する。

【0084】

10

20

30

40

50

ここで、終了条件を満たしていない場合、終了条件判定手段 3 1 6 は、更新率減少手段 3 1 7 に更新率 の減少を指令する。その後、終了条件判定手段 3 1 6 は、第 1 生起確率要求手段 3 1 2、第 2 生起確率要求手段 3 1 3、第 1 混合係数算出手段 3 1 4、及び、写像ベクトル更新手段 3 1 5 に処理の再実行を指令する。

一方、終了条件を満たしている場合、終了条件判定手段 3 1 6 は、処理を終了する。

なお、図 2 では、終了条件判定手段 3 1 6 からの指令信号を破線で図示した。

【0085】

更新率減少手段 3 1 7 は、予め設定された更新率減少規則に従って、必要に応じて学習パラメータ記憶手段 3 0 2 の更新率 を減少させるものである。例えば、更新率減少規則としては、更新率 の値から予め設定した値を減算するという規則があげられる。

10

【0086】

[混合生起確率算出装置の構成]

混合生起確率算出装置 4 0 は、ニューラルネットワーク言語モデルと他の確率モデルとのそれぞれで求めた生起確率 p_N 、 p_O を混合した混合生起確率 p を算出するものである。図 2 のように、混合生起確率算出装置 4 0 は、対象データ記憶手段 4 0 1 と、混合生起確率記憶手段 4 0 2 と、第 3 生起確率要求手段（第 3 生起確率入力手段）4 1 1 と、第 4 生起確率要求手段（第 4 生起確率入力手段）4 1 2 と、第 2 混合係数算出手段 4 1 3 と、混合生起確率算出手段 4 1 4 とを備える。

【0087】

対象データ記憶手段 4 0 1 は、混合生起確率 p の算出対象となる前文脈及び次単語を表す単語列を記憶するメモリ、ハードディスク等の記憶手段である。この対象データ記憶手段 4 0 1 の単語列は、学習データ記憶手段 3 0 3 の単語列と異なるものである。

20

混合生起確率記憶手段 4 0 2 は、混合生起確率 p を記憶するメモリ、ハードディスク等の記憶手段である。

【0088】

第 3 生起確率要求手段 4 1 1 は、対象データ記憶手段 4 0 1 の前文脈をニューラルネットワーク言語モデル演算装置 1 0 に出力することで、隠れ層ベクトル z 及び生起確率 p_N を要求するものである。この要求に応じて、第 3 生起確率要求手段 4 1 1 は、ニューラルネットワーク言語モデル演算装置 1 0 から、隠れ層ベクトル z 及び生起確率 p_N が入力される。そして、第 3 生起確率要求手段 4 1 1 は、入力された隠れ層ベクトル z 及び生起確率 p_N を第 2 混合係数算出手段 4 1 3 及び混合生起確率算出手段 4 1 4 に出力する。

30

【0089】

第 4 生起確率要求手段 4 1 2 は、対象データ記憶手段 4 0 1 の前文脈を他言語モデル演算装置 2 0 に出力することで、生起確率 p_O を要求するものである。ここで、第 4 生起確率要求手段 4 1 2 は、第 3 生起確率要求手段 4 1 1 と同一の前文脈を他言語モデル演算装置 2 0 に出力する。この要求に応じて、第 4 生起確率要求手段 4 1 2 は、他言語モデル演算装置 2 0 から、生起確率 p_O が入力される。そして、第 4 生起確率要求手段 4 1 2 は、入力された生起確率 p_O を混合生起確率算出手段 4 1 4 に出力する。

【0090】

第 2 混合係数算出手段 4 1 3 は、式 (9) を用いて、混合係数パラメータ記憶手段 3 0 1 の写像ベクトル S により、第 3 生起確率要求手段 4 1 1 から入力された隠れ層ベクトル z を実数値のスカラ s に線形写像するものである。また、第 2 混合係数算出手段 4 1 3 は、式 (8) を用いて、実数値のスカラ s をシグモイド関数で非線形変換することで、混合係数 を算出する。そして、第 2 混合係数算出手段 4 1 3 は、算出した混合係数を混合係数記憶手段 3 0 4 に記憶する。

40

【0091】

混合生起確率算出手段 4 1 4 は、混合係数記憶手段 3 0 4 の混合係数 を用いて、第 3 生起確率要求手段 4 1 1 から入力された生起確率 p_N と、第 4 生起確率要求手段 4 1 2 から入力された生起確率 p_O とを混合することで、混合生起確率 p を算出するものである。そして、混合生起確率算出手段 4 1 4 は、算出した混合生起確率 p を混合生起確率記憶手

50

段 4 0 2 に記憶する。

【 0 0 9 2 】

[混合係数パラメータ学習装置の動作]

図 3 を参照し、混合係数パラメータ学習装置 3 0 の動作について説明する (適宜図 2 参照) 。

【 0 0 9 3 】

混合係数パラメータ学習装置 3 0 は、初期化手段 3 1 1 によって、写像ベクトル S 、バイアス値 b 等の混合係数パラメータを初期化する (ステップ S 1) 。

混合係数パラメータ学習装置 3 0 は、初期化手段 3 1 1 によって、更新率、正則化係数等の学習パラメータを初期化する (ステップ S 2) 。

混合係数パラメータ学習装置 3 0 は、カウンタ i の値を 1 に初期化する (ステップ S 3) 。

【 0 0 9 4 】

混合係数パラメータ学習装置 3 0 は、第 1 生起確率要求手段 3 1 2 によって、学習データ記憶手段 3 0 3 の単語列 w_1, w_2, \dots, w_N のうち、先頭から $i - 1$ 個の単語列 w_1, w_2, \dots, w_{i-1} を前文脈としてニューラルネットワーク言語モデル演算装置 1 0 に出力する。

混合係数パラメータ学習装置 3 0 は、第 1 生起確率要求手段 3 1 2 によって、ニューラルネットワーク言語モデル演算装置 1 0 から、隠れ層ベクトル $z(i)$ 及び次単語 w_i の生起確率 $p_N(w_i | w_1 w_2 \dots w_{i-1})$ が入力される (ステップ S 4) 。

【 0 0 9 5 】

混合係数パラメータ学習装置 3 0 は、第 2 生起確率要求手段 3 1 3 によって、ステップ S 4 と同一の前文脈 w_1, w_2, \dots, w_{i-1} を他言語モデル演算装置 2 0 に出力する。

混合係数パラメータ学習装置 3 0 は、第 2 生起確率要求手段 3 1 3 によって、他言語モデル演算装置 2 0 から、次単語 w_i の生起確率 $p_o(w_i | w_1 w_2 \dots w_{i-1})$ が入力される (ステップ S 5) 。

【 0 0 9 6 】

混合係数パラメータ学習装置 3 0 は、第 1 混合係数算出手段 3 1 4 によって、ステップ S 4 で入力された隠れ層のベクトル $z(i)$ 及び写像ベクトル S を用いて、式 (8) 及び式 (9) に従って混合係数 $\alpha(i)$ を算出する (ステップ S 6) 。

【 0 0 9 7 】

混合係数パラメータ学習装置 3 0 は、写像ベクトル更新手段 3 1 5 によって、ステップ S 4 で入力された生起確率 $p_N(w_i | w_1 w_2 \dots w_{i-1})$ と、ステップ S 5 で入力された生起確率 $p_o(w_i | w_1 w_2 \dots w_{i-1})$ と、ステップ S 6 で算出した混合係数 $\alpha(i)$ と、更新率 η を用いて、式 (11) 及び式 (12) で写像ベクトル S を更新する (ステップ S 7) 。

【 0 0 9 8 】

混合係数パラメータ学習装置 3 0 は、カウンタ i をインクリメントする (ステップ S 8) 。

混合係数パラメータ学習装置 3 0 は、カウンタ i が単語最大数 N 以下であるか否かを判定する (ステップ S 9) 。

カウンタ i が単語最大数 N 以下の場合 (ステップ S 9 で Yes)、混合係数パラメータ学習装置 3 0 は、ステップ S 4 の処理に戻る。

【 0 0 9 9 】

カウンタ i が単語最大数 N 以下でない場合 (ステップ S 9 で No)、混合係数パラメータ学習装置 3 0 は、終了条件判定手段 3 1 6 によって、終了条件を満たすか否かを判定する (ステップ S 10) 。

終了条件を満たす場合 (ステップ S 10 で Yes)、混合係数パラメータ学習装置 3 0 は、処理を終了する。

【 0 1 0 0 】

10

20

30

40

50

終了条件を満たさない場合（ステップS10でNo）、混合係数パラメータ学習装置30は、更新率減少手段317によって、更新率減少規則に従って、必要に応じて更新率を減少させ（ステップS11）、ステップS3の処理に戻る。

【0101】

[混合生起確率算出装置の動作]

図4を参照し、混合生起確率算出装置40の動作について説明する（適宜図1参照）。

【0102】

混合生起確率算出装置40は、第3生起確率要求手段411によって、対象データ記憶手段401の単語列 w_1, w_2, \dots, w_{t-1} を前文脈としてニューラルネットワーク言語モデル演算装置10に出力する。

混合生起確率算出装置40は、第3生起確率要求手段411によって、ニューラルネットワーク言語モデル演算装置10から、隠れ層ベクトル z 及び次単語 w_t の生起確率 $p_N(w_t | w_1 w_2 \dots w_{t-1})$ が入力される（ステップS21）。

【0103】

混合生起確率算出装置40は、第4生起確率要求手段412によって、ステップS21と同一の前文脈 w_1, w_2, \dots, w_{t-1} を他言語モデル演算装置20に出力する。

混合生起確率算出装置40は、第4生起確率要求手段412によって、他言語モデル演算装置20から、次単語 w_i の生起確率 $p_O(w_t | w_1 w_2 \dots w_{t-1})$ が入力される（ステップS22）。

【0104】

混合生起確率算出装置40は、第2混合係数算出手段413によって、ステップS21で入力された隠れ層のベクトル z 及び写像ベクトル S を用いて、式(8)及び式(9)に従って混合係数 (t) を算出する（ステップS23）

【0105】

混合生起確率算出装置40は、混合生起確率算出手段414によって、ステップ21で入力された生起確率 $p_N(w_t | w_1 w_2 \dots w_{t-1})$ とステップ22で入力された生起確率 $p_O(w_t | w_1, w_2, \dots, w_{t-1})$ との混合生起確率 $p(w_t | w_1 w_2 \dots w_{t-1})$ を、式(16)で算出する（ステップS24）。

【0106】

[作用・効果]

以上のように、混合生起確率算出システム1は、ニューラルネットワーク言語モデルにより写像ベクトル S を学習し、学習した写像ベクトル S により前文脈に応じた混合係数を求めている。これにより、混合生起確率算出システム1は、*n-gram*言語モデル等の他の言語モデルと混合して混合生起確率 p を算出する際、従来よりも混合生起確率 p の正確性を向上させることができる。

【0107】

(変形例)

以上、本願発明の各実施形態を詳述してきたが、本願発明は前記した実施形態に限られるものではなく、本願発明の要旨を逸脱しない範囲の設計変更等も含まれる。

【0108】

前記した実施形態では、混合係数パラメータ学習装置が混合係数パラメータ記憶手段及び混合係数記憶手段を備えることとして説明したが、本願発明は、これに限定されない。つまり、混合生起確率算出装置が混合係数パラメータ記憶手段及び混合係数記憶手段を備えてもよい。

【0109】

前記した実施形態では、本願発明を言語モデルに適用する例を説明したが、本願発明が適用可能な確率モデルはこれに限定されず、何らかの記号系列に後続して生起する記号の生起確率モデル一般に適用することができる。

【0110】

前記した実施形態では、正則化を行うこととして説明したが、本願発明は、正則化を行

10

20

30

40

50

わなくともよい。

前記した実施形態では、バイアス値 b を用いることとして説明したが、本願発明は、バイアス値 b を用いなくともよい。この場合、前記した式 (9) の代わりに以下の式 (1 7) を用いることになる。

【 0 1 1 1 】

【 数 1 7 】

$$s(t) = Sz(t) \quad \dots \text{式 (1 7)}$$

【 0 1 1 2 】

前記した実施形態では、混合係数パラメータ学習装置を独立したハードウェアとして説明したが、本願発明は、これに限定されない。例えば、混合係数パラメータ学習装置は、コンピュータが備える CPU、メモリ、ハードディスク等のハードウェア資源を協調動作させる混合係数パラメータ学習プログラムで実現することもできる。このプログラムは、通信回線を介して配布してもよく、CD-ROM やフラッシュメモリ等の記録媒体に書き込んで配布してもよい。

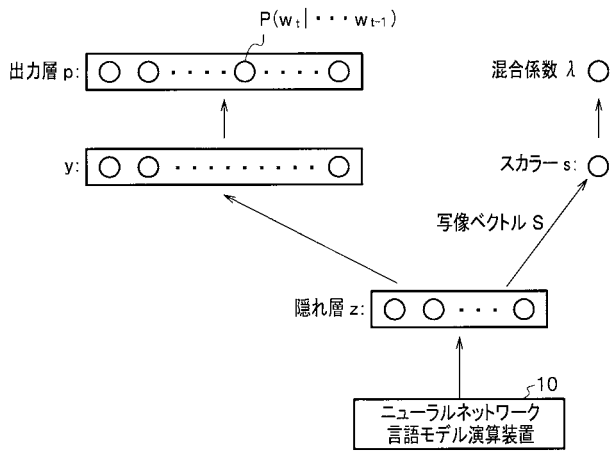
また、混合生起確率算出装置は、混合係数パラメータ学習装置と同様、混合生起確率算出プログラムで実現することもできる。

【 符号の説明 】

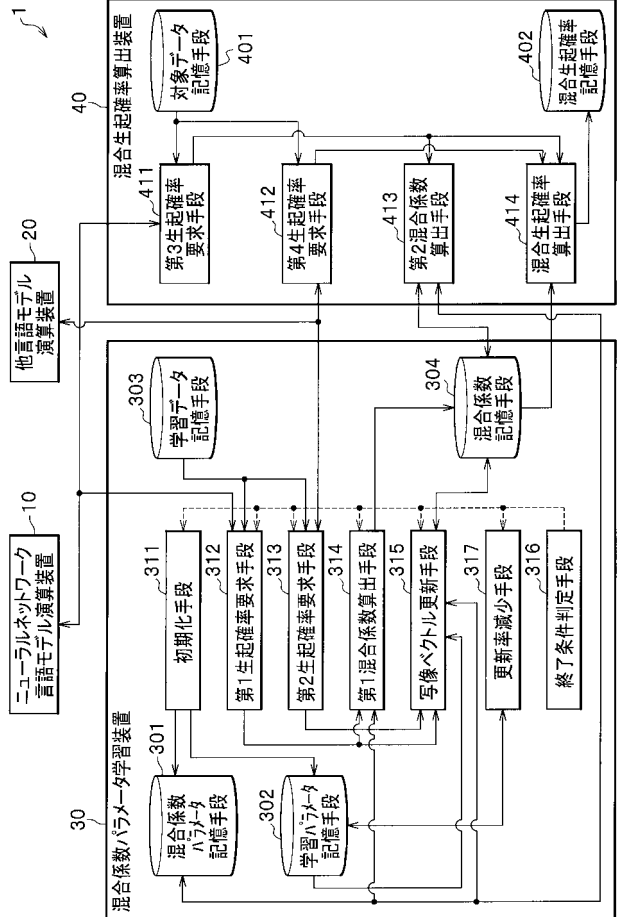
【 0 1 1 3 】

- | | | |
|-------|-------------------------------|----|
| 1 | 混合生起確率算出システム | 20 |
| 1 0 | ニューラルネットワーク言語モデル演算装置 | |
| 2 0 | 他言語モデル演算装置 | |
| 3 0 | 混合係数パラメータ学習装置 | |
| 3 0 1 | 混合係数パラメータ記憶手段 | |
| 3 0 2 | 学習パラメータ記憶手段 | |
| 3 0 3 | 学習データ記憶手段 | |
| 3 0 4 | 混合係数記憶手段 | |
| 3 1 1 | 初期化手段 | |
| 3 1 2 | 第 1 生起確率要求手段 (第 1 生起確率入力手段) | |
| 3 1 3 | 第 2 生起確率要求手段 (第 2 生起確率入力手段) | 30 |
| 3 1 4 | 第 1 混合係数算出手段 | |
| 3 1 5 | 写像ベクトル更新手段 | |
| 3 1 6 | 終了条件判定手段 | |
| 3 1 7 | 更新率減少手段 | |
| 4 0 | 混合生起確率算出装置 | |
| 4 0 1 | 対象データ記憶手段 | |
| 4 0 2 | 混合生起確率記憶手段 | |
| 4 1 1 | 第 3 生起確率要求手段 (第 3 生起確率入力手段) | |
| 4 1 2 | 第 4 生起確率要求手段 (第 4 生起確率入力手段) | |
| 4 1 3 | 第 2 混合係数算出手段 | 40 |
| 4 1 4 | 混合生起確率算出手段 | |

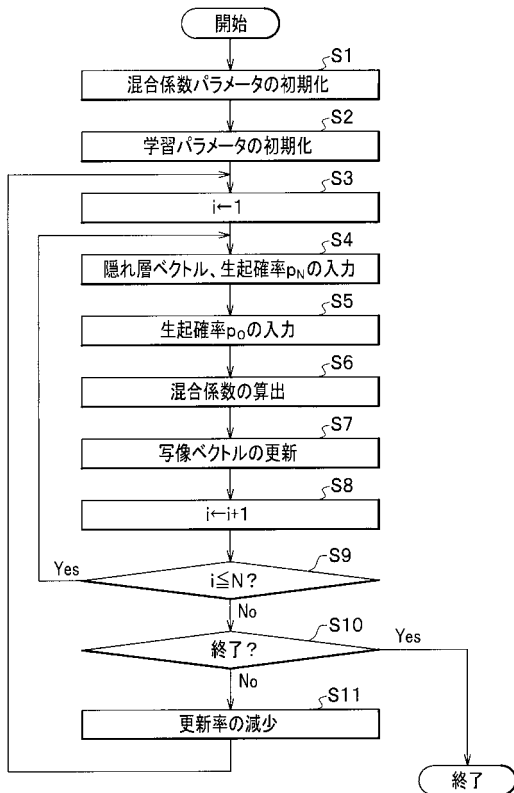
【図1】



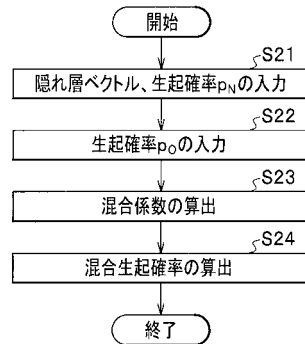
【図2】



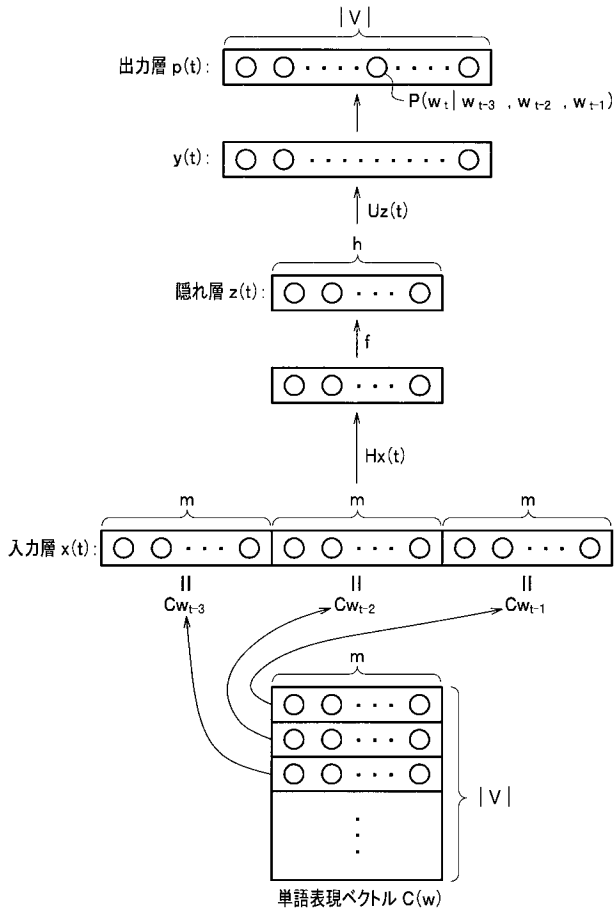
【図3】



【図4】



【 図 5 】



【 図 6 】

