

明 細 書

発明の名称：

単語アライメントスコア算出装置、単語アライメント装置、及びコンピュータプログラム

技術分野

[0001] 本発明は、対訳文対内で各単語の対応関係（単語アライメント）を解析する単語アライメント装置及びそのための単語アライメントスコア算出装置に関する。

背景技術

[0002] 単語アライメントは、統計的機械翻訳（Statistical Machine Translation：SMT）に欠かせない重要な処理の一つである。図1に単語アライメントの例を示す。

[0003] 図1を参照して、日本語文と英語との対訳文対20を考える。対訳文対20の各文は予め単語ごとに区切られている。対訳文対20は、日本語文30（「私 | が | 利用 | 料金 | を | 払う」（「|」は単語の区切りを示す。））と、英語文32（「I pay usage fees.」）とを含む。単語アライメントは、例えば日本語文30を構成する各単語が英語文32中のどの単語（又は単語群）に翻訳されるか、すなわち日本語文30を構成する各単語が英語文32中のどの単語（又は単語群）に対応するかを推定する処理である。図1には日本語から英語への単語アライメントを示したが、英語から日本語への単語アライメントも同様に行なわれる。

[0004] SMTでは、このような単語アライメントが非常に重要な役割を果たす。SMTでは、上記したような対訳を多数含む対訳コーパスを準備する。各対訳について単語アライメントを行なう。この単語アライメントに基づいて、統計的な処理により翻訳モデルが作成される。この過程を翻訳モデルの学習と呼ぶ。翻訳モデルは、端的に言えば、一方の言語のある単語が他方の言語のどの単語に翻訳されるかを確率の形で示すものである。SMTでは、原言

語の文が与えられたときに、翻訳先の言語（目的言語）の文の候補を多数準備し、その原言語の文が目的言語の各文の候補から生成された確率を求め、目的言語の文のうちでその確率が最も高い文を原言語の文に対する翻訳文と推定する。この過程で、上記した翻訳モデルが使用される。

[0005] 明らかに、SMTの性能を改善するためには翻訳モデルの精度を高める必要がある。そのためには翻訳モデルの学習に使用される対訳コーパスの単語アライメントの精度を高めなければならない。したがって、SMTの性能を改善するために、対訳に対して単語アライメントを行なう単語アライメント装置の性能向上が求められている。

[0006] 広く使われている単語アライメントの方法として、IBMモデル（後掲の非特許文献1を参照）、及びHMMモデル（非特許文献2を参照）がある。これらは、単語アライメントが、ある確率分布により生成されることを想定し、実際に観測された単語アライメントからその確率分布を推定（学習）するモデル（生成モデル）である。原言語の文を $f_1^J = f_1, \dots, f_J$ 、目的言語の文を $e_1^I = e_1, \dots, e_I$ とすると、原言語の文 f_1^J は単語アライメント a_1^J にしたがって目的言語の文 e_1^I から生成され、その生成確率は以下の式（1）で計算する。式（1）中の各 a_j は、原言語の単語 f_j が目的言語の単語 e_{a_j} に対応付けられることを示す隠れ変数である。なお、以下のテキストでは、ある下付き文字にさらに下付き文字が付いていることをアンダースコア「 」で示し、下付き文字の範囲をアンダースコアの後ろの「 」で示す。すなわち、「 $e_{a_{\{j\}}}$ 」という表現は、「 e 」に添えられた下付き文字が通常の表現では「 a_j 」であることを示し、「 $e_{a_{\{j-1\}}}$ 」という表現は e の下付き文字が a_{j-1} であることを示し、「 $e_{a_{\{j-1\}}}$ 」という表現は e の下付き文字が a_{j-1} であることを示す。

[0007] [数1]

$$p(f_1^J | e_1^I) = \sum_{a_1^J} p(f_1^J, a_1^J | e_1^I) \cdots (1)$$

[0008]

[数2]

$$p(f_1^J, a_1^J | e_1^I) = \prod_{j=1}^J p_a(a_j | a_{j-1}, j) \cdot p_t(f_j | e_{a_j}) \cdots (2)$$

式(2)において p_a はアライメント確率、 p_t は語彙翻訳確率である。

[0009] これらのモデルは、対訳文対 (f_1^J, e_1^I) に対し、以下の式(3)を満たす最適なアライメント $\hat{a}$ (記号「 $\hat{\cdot}$ 」は、本来は直後の文字の直上に記載されるべきものである。)をforward-backwardアルゴリズムを使う等して特定する。最適なアライメント $\hat{a}$ はビタビアライメントと呼ばれる。

[0010] [数3]

$$\hat{a} = \arg \max_{a_1^J} p(f_1^J, a_1^J | e_1^I) \cdots (3)$$

[0011] 非特許文献3は、フィード・フォワード型ニューラル・ネットワーク(feed forward neural networks (FFNN))の一種であるContext-Dependent Deep Neural Network for HMMを非特許文献2のHMMモデルに適用し、アライメント確率に対応するアライメントスコア、語彙選択確率に対応する語彙スコアをFFNNで計算するアライメント手法を提案した。すなわち、対訳文対 (f_1^J, e_1^I) に対するアライメント a_1^J のスコア $s_{NN}(a_1^J | f_1^J, e_1^I)$ は以下の式(4)により表される。

[0012] [数4]

$$\begin{aligned} s_{NN}(a_1^J | f_1^J, e_1^I) \\ = \prod_{j=1}^J t_a(a_j - a_{j-1} | c(e_{a_{j-1}})) \cdot t_t(f_j, e_{a_j} | c(f_j), c(e_{a_{j-1}})) \cdots (4) \end{aligned}$$

[0013] 非特許文献3の手法では、全単語における正規化の計算量が膨大となるために、確率のかわりにスコアを用いている。ここで、 t_a 及び t_t はそれぞれ、

式(2)の p_a 及び p_t に対応する。 s_{mn} はアライメント a_j^d のスコアであり、 $c(w)$ は、単語 w の文脈を表す。ビタビアライメントは、このモデルでもHMMモデルと同様、forward-backwardアルゴリズムにより決定する。

[0014] 式(4)における語彙翻訳スコア $t_t(f_j, e_{a_{\{j\}}} | c(f_j), c(e_{a_{\{j\}-1}))$ を計算するためのニューラル・ネットワークのネットワーク構造(語彙翻訳モデル)を図3に示す。図3に示すニューラル・ネットワーク60は、原言語の単語 f_{j-1} 、 f_j 、及び f_{j+1} 、並びに目的言語の単語 $e_{a_{\{j\}-1}}$ 、 $e_{a_{\{j\}}}$ 、及び $e_{a_{\{j\}+1}}$ を受け、これらをベクトル z_0 に変換する入力層(Lookup層)70と、ベクトル z_0 を受けて式(5)による出力ベクトル z_1 を出力する隠れ層72と、ベクトル z_1 を受けて式(6)により語彙翻訳スコア76を計算し出力する出力層74とを含む。各層はそれぞれ、重み行列 L 、 $\{H, B_H\}$ 、及び $\{O, B_O\}$ を持つ。ここでは隠れ層が1層の場合を説明するが、複数個の隠れ層を用いてもよい。

[0015] 重み行列 L はembedding行列であり、各単語のword embeddingを管理する。word embeddingとは、単語を特徴付ける低次元の実ベクトルであり、単語の統語的、意味的特性を表す。原言語の単語集合を V_f 、目的言語の単語集合を V_e 、word embeddingの長さを M とすると、重み行列 L は $M \times (|V_f| + |V_e|)$ 行列である。ただし、 V_f と V_e には、それぞれ、未知語を表す〈unk〉と単語「null」を表す〈null〉を追加する。

[0016] この語彙翻訳モデルは、入力として、計算対象である原言語の単語 f_j と目的言語の単語 $e_{a_{\{j\}}}$ とともにそれらの文脈単語を受付ける。文脈単語とは、予め定めたサイズの窓内に存在する単語である。ここでは、図3に示すように窓幅が3の場合を想定している。入力層70は、原言語の入力部80と、目的言語の入力部82とを含む。原言語の入力部80は、計算対象である原言語の単語 f_j とともに、その前後の2つの単語 f_{j-1} 及び f_{j+1} を受付け、embedding行列(L)から対応する列を見つけ、word embeddingベクトルのうち、原言語部分を出力する。目的言語の入力部82は、計算対象である目的言語の単語 $e_{a_{\{j\}}}$ とともに、その前後の2つの単語 $e_{a_{\{j\}-1}}$ 及び $e_{a_{\{j\}+1}}$ を受付け、embe

ding行列 (L) から対応する列を見つけ、word embeddingベクトルのうち、目的言語の部分を出力する。原言語の入力部 80 の出力及び目的言語の入力部 82 の出力は連結されて実ベクトル z_0 を形成し隠れ層 72 の入力に与えられる。次に隠れ層 72 が実ベクトル z_0 の非線形な特徴を捉えてベクトル z_1 を出力する。最後に、出力層 74 が隠れ層 72 の出力するベクトル z_1 を受け取り、以下の式で表される語彙翻訳スコア 76 を計算して出力する。

[0017] [数5]

$$t_t(f_j, e_{a_j} | f_{j-1}^{j+1}, e_{a_{j-1}}^{a_{j+1}})$$

[0018] 隠れ層 72 及び出力層 74 が行なう具体的な計算は以下の通りである。

[0019] [数6]

$$z_1 = f(H \times z_0 + B_H), \quad (5)$$

$$t_t = O \times z_1 + B_O \quad (6)$$

ここで、 H 、 B_H 、 O 、 B_O はそれぞれ、 $|z_1| \times |z_0|$ 、 $|z_1| \times 1$ 、 $1 \times |z_1|$ 、 1×1 行列である。 $f(x)$ は非線形活性化関数であり、ここでは $\text{htanh}(x)$ を用いている。 $\text{htanh}(x)$ は次のような関数である。

[0020] [数7]

$$\text{htanh}(x) = \begin{cases} -1 & (x < -1) \\ 1 & (x > 1) \\ x & (\text{上記以外}) \end{cases}$$

[0021] アライメントスコア $t_a(a_j - a_{j-1} | c(e_{a_{\{j\}-1}}))$ を計算するアライメントモデルもこれと同様に構成できる。

[0022] 各モデルの学習では、次式 (7) のランキング損失を最小化するように、各層の重み行列を確率的勾配降下法 (Stochastic Gradient Descent: SGD) により学習する。各重みの勾配はバックプロパゲーションで計算する。

[0023] [数8]

$$\text{loss}(\theta) = \sum_{(f,e) \in T} \max\{0, 1 - s_\theta(a^+ | f, e) + s_\theta(a^- | f, e)\}. \quad (7)$$

ここで、 θ は最適化するパラメータ（重み行列の重み）、 T は学習データ、 s_{θ} はパラメータ θ のモデルによる a_{j-1} のスコア（式（４）参照）、 a^+ は正解アライメント、 a^- はパラメータ θ のモデルでスコアが最高の不正解アライメントを示す。

先行技術文献

非特許文献

[0024] 非特許文献1: Peter F. Brown, Stephen A. Della Pietra, Vincent J. Della Pietra, and Robert L. Mercer. 1993. The Mathematics of Statistical Machine Translation: Parameter Estimation. *Computational Linguistics*, 19(2):263-311.

非特許文献2: Stephan Vogel, Hermann Ney, and Christoph Tillmann. 1996. HMM-based Word Alignment in Statistical Translation. In *Proceedings of the 16th International Conference on Computational Linguistics*, pages 836-841.

非特許文献3: Nan Yang, Shujie Liu, Mu Li, Ming Zhou, and Nenghai Yu. 2013. Word Alignment Modeling with Context Dependent Deep Neural Network. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 166-175.

発明の概要

発明が解決しようとする課題

[0025] 式（２）及び式（４）のいずれにおいても、各単語のアライメント a_j は直前のアライメント a_{j-1} に依存している。図２に示すように、日本語文３０の先頭の単語「私」のアライメント４０を行なった後、次の単語「が」のアライメント４２を行なう際には、アライメント４０の結果が参照される。すなわち、最適なアライメント $\hat{a}$ を求める際、各単語のアライメントは、直前のアライメント関係に基づいて（手がかりとして）決定される。

[0026] しかし、直前のアライメント関係のみでアライメントの手がかりとして十分かどうかは定かではない。単語アライメントの精度を高めるためには他の手法も考え、より精度の高い方法があればそれを特定する必要がある。

[0027] したがって本発明の1つの目的は、従来の手法より精度が高い単語アライメントが行なえる単語アライメント装置及びそのための単語アライメントスコア算出装置、ならびにそれらのためのコンピュータプログラムを提供することである。

課題を解決するための手段

[0028] 本発明の第1の局面に係る単語アライメントスコア算出装置は、第1の言語と第2の言語との対訳文対の単語アライメントのスコアを算出する。この装置は、対訳文対と、当該対訳文対に対する単語アライメントとを受けて、所定の順序にしたがって、対訳文対の第1の言語の文の単語を順番に選択する選択手段と、選択手段により選択された単語と、単語アライメントにより当該単語と対応付けられた第2の言語の単語とからなる単語対が正しい単語対である可能性を示すスコアを対訳文対の第1の言語の文の全ての単語について算出し、当該スコアに基づいて、単語アライメントのスコアを算出するスコア算出手段とを含む。スコア算出手段は、ある単語対のスコアを算出するにあたって、当該ある単語対を構成する第1の言語の単語より前に選択手段により選択された単語のアライメントの全てに基づいて当該ある単語対のスコアを算出する。

[0029] 好ましくは、選択手段は、対訳文対の内、第1の言語の文の単語を第1の言語の文の先頭から順番に選択する手段を含む。

[0030] より好ましくは、スコア算出手段は、選択手段により選択された単語と、対訳文対の第2の言語の文のうちで単語アライメントにより当該単語と対応付けられた単語とからなる単語対が正しい単語対である可能性を示すスコアを算出する第1の算出手段と、対訳文対の第1の言語の文の全ての単語について第1のスコア算出手段により算出されたスコアに基づいて、単語アライメントのスコアを算出する第2の算出手段とを含む。

- [0031] さらに好ましくは、第2の算出手段は、対訳文対の第1の言語の文の全ての単語について第1の算出手段により算出されたスコアを互いに乗算することにより、単語アライメントのスコアを算出する手段を含む。
- [0032] スコア算出手段は、選択手段により選択された単語を受ける第1の入力と、単語アライメントにより当該単語に対応付けられた第2の言語の単語を受ける第2の入力とを持つリカレント型のニューラル・ネットワークと、選択手段により選択された単語と、単語アライメントにより当該単語に対応付けられた単語とを、第1の入力及び第2の入力にそれぞれ与える入力制御手段とを含んでもよい。リカレント型のニューラル・ネットワークは、第1の入力及び第2の入力を持ち、第1の入力及び第2の入力にそれぞれ与えられた単語からword embeddingベクトルを算出し出力する入力層と、入力層の出力を受け、予め定められた非線形演算により、入力層からの2つの出力の関係を表すベクトルを生成する隠れ層と、隠れ層の出力に基づいてスコアを算出し出力する出力層とを含む。隠れ層の出力は、単語アライメントスコア算出装置に次の単語対が入力として与えられたときに、隠れ層に入力として与えられる。
- [0033] 本発明の第2の局面に係る単語アライメント装置は、第1の言語と第2の言語との対訳文対の単語アライメントを推定する単語アライメント装置である。この装置は、上記したいずれかの単語アライメントスコア算出装置と、対訳文対に対して複数個の単語アライメント候補を生成する単語アライメント候補生成手段と、単語アライメント候補生成手段により生成された複数個の単語アライメント候補の各々について、単語アライメントスコア算出装置を用いて対訳文対に対する単語アライメントスコアを算出する算出手段と、算出手段により複数個の単語アライメント候補に対して算出された単語アライメントスコアのうち、最も高いスコアに対応する単語アライメント候補を、対訳文対の単語アライメントとして決定し出力する単語アライメント決定手段とを含む。
- [0034] 本発明の第3の局面に係るコンピュータプログラムは、コンピュータによ

り実行されると、当該コンピュータを、上記したいずれかの装置の各手段として機能させる。

図面の簡単な説明

[0035] [図1]単語アライメントを説明する模式図である。

[図2]従来の単語アライメントの方法を説明する模式図である。

[図3]非特許文献3による語彙翻訳モデルを実現するニューラル・ネットワークの構成を模式的に示す図である。

[図4]本発明の1実施の形態におけるRNNによる単語アライメントモデルの構成を示す模式図である。

[図5]対訳コーパスを用いて図4に示す単語アライメントモデルのための学習を行なうためのコンピュータプログラムの制御構造を示すフローチャートである。

[図6]図4に示す単語アライメントモデルを用いてコーパス中の対訳の単語アライメントを行なうコンピュータプログラムの制御構造を示すフローチャートである。

[図7]対訳コーパスを用いて図4に示す単語アライメントモデルの片方向の学習を行なうためのコンピュータプログラムの制御構造を示すフローチャートである。

[図8]実験に用いたデータセットのサイズを示す表形式の図である。

[図9]本発明の1実施の形態による単語アライメント性能を表形式で示す図である。

[図10]本発明の1実施の形態による単語アライメント結果を用いて学習した統計的機械翻訳装置の翻訳性能を表形式で示す図である。

[図11]単語アライメントの例を示す図である。

[図12]単語アライメントの他の例を示す図である。

[図13]様々なサイズの学習データにより学習した際のBTECコーパスに対する単語アライメント性能を表形式で示す図である。

[図14]本発明の1実施の形態に係る単語アライメント装置を実現するコンピ

ユータの外観を示す図である。

[図15]図14に示すコンピュータの内部構成を示すブロック図である。

発明を実施するための形態

[0036] 以下の説明及び図面では、同一の部品には同一の参照番号を付してある。したがって、それらについての詳細な説明は繰返さない。

[0037] [基本的考え方]

本実施の形態では、最適なアライメント $\hat{a}$ を求める際、各単語のアライメントは、文頭から直前までの全てのアライメント関係に基づいて決定する。例えば、アライメント系列 $a_1^J = a_1, \dots, a_J$ のスコアを次の式(8)のように以前のアライメント関係全てに依存したスコアにより求める。スコアは確率でもよい。

[0038] [数9]

$$s_{NN}(a_1^J | f_1^J, e_1^J) = \prod_{j=1}^J t_{RNN}(a_j | a_1^{j-1}, f_j, e_{a_j}) \cdots (8)$$

[0039] [第1の実施の形態]

このために本実施の形態では、RNN (Recurrent Neural Network) によるアライメントモデルを採用する。このモデルは、RNNを用いてアライメント a_1^J のスコア s_{NN} を式(8)を用いて計算する。式(8)によれば、第 j 番目のアライメント a_j の予測は、それに先立つ全てのアライメント a_1^{j-1} に依存する。この例では、従来のFFNN型のものと同様、確率ではなくスコアを用いている点に注意する必要がある。

[0040] 図4に、このRNNによるモデル100の構成を示す。図4を参照して、このモデル100は、入力層(Lookup層)110と、隠れ層112と、出力層114とを含む。これらはそれぞれ、重み行列 L , $\{H^d, R^d, B_{H^d}\}$ 、及び $\{O, B_o\}$ を持つ。隠れ層112の行列 (H^d, R^d, B_{H^d}) はアライメントに依存する。ここで d はアライメント a_{j-1} からアライメント a_j までのジャンプ距離を示す： $d = a_j - a_{j-1}$ 。なお、後述する実験では、ジャンプ距離が8

以上のもの及び -8 以下のものはそれぞれ一まとめで「 ≥ 8 」距離及び「 ≤ -8 」距離とした。つまり、隠れ層 1 1 2 は重み行列 $\{H^{\leq -8}, H^{-7}, \dots, H^7, H^{\geq 8}, R^{\leq -8}, R^{-7}, \dots, R^7, R^{\geq 8}, B_H^{\leq -8}, B_H^{-7}, \dots, B_H^7, B_H^{\geq 8}\}$ を持つ。

[0041] 図 4 を参照して、入力層 1 1 0 は、原言語の入力部 1 3 0 と、目的言語の入力部 1 3 2 とを含む。原言語の入力部 1 3 0 は、計算対象である原言語の単語 f_j を受け、embedding 行列 (L) から対応する列を見つけ、word embedding ベクトルのうち、原言語部分を出力する。目的言語の入力部 1 3 2 は、計算対象である目的言語の単語 $e_{a_{\{j\}}}$ を受け、embedding 行列 (L) から対応する列を見つけ、word embedding ベクトルのうち、目的言語の部分を出力する。原言語の入力部 1 3 0 の出力及び目的言語の入力部 1 3 2 の出力は連結されて実ベクトル x_j を形成し隠れ層 1 1 2 の入力に与えられる。隠れ層 1 1 2 の前回の出力は一時記憶されており、循環接続 1 1 8 により隠れ層 1 1 2 の入力に前回の出力ベクトル y_{j-1} として与えられる。次に隠れ層 1 1 2 が実ベクトル x_j 及び出力ベクトル y_{j-1} の非線形な特徴を捉えてベクトル y_j を出力する。このベクトルは一時記憶され、原言語の次の単語 f_{j+1} の処理時に循環接続 1 1 8 を介して隠れ層 1 1 2 の入力に与えられる。なお、 $j = 0$ のときにはベクトル y_{j-1} は全て 0 に初期化されている。隠れ層 1 1 2 における計算では、特定のジャンプ距離 d に依存した重み行列 $\{H^d, R^d, B_H^d\}$ が使われていることに注意が必要である。最後に、出力層 1 1 4 が隠れ層 1 1 2 の出力するベクトル y_j を受け取り、次式で表されるスコア 1 0 2、つまり

[0042] [数10]

$$t_{RNN}(a_j | a_1^{j-1}, f_j, e_{a_j})$$

を計算して出力する。

[0043] 隠れ層 1 1 2 の出力するベクトル y_j を受けた出力層 1 1 4 は、ベクトル y_j に応答してアライメント a_j のスコア 1 0 2 (t_{RNN}) を式 (10) により算出し出力する。従来の F F N N モデル (図 3) は語彙翻訳スコアとアライメントスコアとの 2 つの要素からなっていたが、この実施の形態の R N N による

モデルは、単に1つのスコア102しか出力しない点に注意が必要である。

[0044] このモデルの隠れ層112及び出力層114における計算は以下の式により行なわれる。

[0045] [数11]

$$y_j = f(H^d \times x_j + R^d \times y_{j-1} + B_H^d) \quad (9)$$

$$t_{RNN} = O \times y_j + B_O \quad (10)$$

ただし H^d 、 R^d 、 B_H^d 、 O 、及び B_O はそれぞれ、 $|y_j| \times |x_j|$ 、 $|y_j| \times |y_{j-1}|$ 、 $|y_j| \times 1$ 、 $1 \times |y_j|$ 、及び 1×1 行列である。ここでは $|y_j| = |y_{j-1}|$ であることに注意。 $f(x)$ は非線形活性化関数であり、この実施の形態では $\text{htanh}(x)$ である。

[0046] この実施の形態でビタビアライメントはforward-backwardアルゴリズムにより決定するが、厳密にはダイナミック・プログラミングによるforward-backwardアルゴリズムは使用できない。 y_j に対するアライメントの履歴が長いためである。したがってここでは、ヒューリスティックなビームサーチによりビタビアライメントを近似的に算出する。具体的には、forwardアルゴリズムの際、各 f_j に対して、予め指定するビーム幅分の状態のみを保持し、それ以外は切り捨てる。

[0047] このように、RNNによるモデルは、循環接続された隠れ層を持つ。この循環接続により、これまでのアライメントの履歴を隠れ層112によってコンパクトにエンコードして伝播できる。そのため、その隠れ層112の設定にしたがってスコアを求めることにより、以前のアライメント関係全体を考慮したスコアを計算できる。

[0048] 〈学習〉

学習では、各層の重み行列の最適化を、バッチサイズDのミニバッチSGDにより行なった。この手法は、単純なSGD ($D=1$) よりも収束が速く安定している。勾配は、通時的逆伝播 (Back propagation through time : BPTT) アルゴリズムにより計算した。BPTTでは、ネットワークを時間

(j) 方向に展開し、時間ステップごとに勾配を計算する。さらに目的関数にl2正規化項を加えることによりモデルの過学習を防止する。

[0049] RNNによるモデルは、FFNNによるモデルと同様に教師付き学習による学習ができる。この場合、学習は式(7)により定義されるランキング損失に基づいて行なわれる。また、この学習方法以外に、アライメントの両方向の整合性をとる制約を組込んだり、教師なし学習をしたりすることで、更なる性能改善ができる。教師付き学習では、教師データ(理想的なアライメント)が必要となる。こうした問題を解決するため、本実施の形態では、ラベルなしの学習データから学習を行なえる、NCE(Noise-Contrastive Estimation)を用いた教師なし学習を用いる。

[0050] 〈教師なし学習〉

Dyerらは、contrastive estimation(CE)に基づく教師なし学習によるアライメントモデルを提案している(Chris Dyer, Jonathan Clark, Alon Lavie, and Noah A. Smith. 2011. Unsupervised Word Alignment with Arbitrary Features. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies Volume 1, pages 409-419.)。CEは、観測されたデータに対し、その周辺のデータを擬似的なネガティブサンプルとみなし、観測データとその周辺データを識別するモデルを学習する。Dyerらは、学習データTとして与えられた対訳文について考えられる限りで可能な単語アライメントを観測データ、翻訳の全探索空間 Ω をその周辺データとみなした。

[0051] 本実施の形態ではこのアイデアをマージン付きランキング損失に導入した。

[0052] [数12]

$$\begin{aligned} & \text{loss}(\theta) \\ &= \max \left\{ 0, 1 - \sum_{(f^+, e^+) \in T} E_{\Phi} [s_{\theta}(a | f^+, e^+)] + \sum_{(f^+, e^-) \in \Omega} E_{\Phi} [s_{\theta}(a | f^+, e^-)] \right\} \quad (11) \end{aligned}$$

ここで Φ は (f, e) が与えられたときに可能な全てのアライメントの集合を指し、 $E_{\Phi}[s_{\theta}]$ は Φ 上でのスコア s_{θ} の期待値を表し、 e^+ は学習データ中の目的言語の文を表し、 e^- は擬似的な目的言語の文を表す。期待値の第1項は観測データに対するものであり、第2項は周辺データに関するものである。

[0053] 全探索空間 Ω に対する計算量は膨大となる。計算量を削減するためにNCEを採用した。NCEは、全探索空間 Ω 中からランダムに選んだサンプル文を e^- とする。また、ビーム幅 W のビームサーチにより期待値を算出し、スコアの低いアライメントを捨てることで計算量をさらに削減する。そして、上記式(11)をオンライン処理に適した形式に変形する。

[0054] [数13]

$loss(\theta)$

$$= \sum_{(f^+, e^+) \in T} \max \left\{ 0, 1 - E_{GEN}[s_{\theta}(a|f^+, e^+)] + \frac{1}{N} \sum_{e^-} E_{GEN}[s_{\theta}(a|f^+, e^-)] \right\} \quad (12)$$

ただし、 e^+ は学習データ中で f^+ の対訳となっている目的言語の文、すなわち $(f^+, e^+) \in T$ であり、 e^- はランダムサンプリングにより生成された、長さが $|e^+|$ である擬似的な目的言語の文であり、 N は f^+ に対して生成する擬似的な目的言語の文の数を表す。 GEN は可能な全てのアライメントの集合 Φ の、ビームサーチにより生成されたサブセットを示す。

[0055] e^- を生成する単純な方法では、目的言語の単語集合 V_e からランダムに $|e^+|$ 回目的言語の単語をサンプリングし、それらを並べることで e^- を生成する。モデルの学習により効果的なネガティブサンプルを生成するためには、原言語の単語 $f_i \in f^+$ と学習データの対訳文内で共起する目的言語の単語の内、 l_0 priorを組込んだIBMモデル1においてしきい値 C より大きな確率を持つ単語（翻訳候補の単語）と判定された単語集合からサンプリングする。 l_0 priorを組込んだIBMモデル1は、標準的なIBMモデル1と比較してより疎なアライメントを生成するため、翻訳候補の数を削減する上では適している。

[0056] 〈合意制約〉

F F N NによるモデルもR N NによるモデルもHMMアライメントモデルを基礎にしており、したがって非対称である。すなわち、目的言語側から見たときに1対多のアライメントを表し得る。このような非対称モデルの学習は、2つのアライメント方向で行なわれることが通常である。しかし、このような方向性を持つモデルについて、両者が一致するような学習を行なうとアライメント性能が向上することが分かった。このような制約を合意制約と呼ぶ。以下、上記したモデルに合意制約を課して学習を行なう方法を説明する。

[0057] 具体的には、合意制約は、双方向におけるword embeddingを一致させるという条件を課す。この実施の形態では、双方向のモデルについて、以下に示すような、word embeddingの相違を表現するペナルティ項を導入した目的関数にしたがって学習を行なう。

[0058] [数14]

$$\arg \min_{\theta_{FE}} \{ \text{loss}(\theta_{FE}) + \alpha \| \theta_{LEF} - \theta_{LFE} \| \} \quad (13)$$

$$\arg \min_{\theta_{EF}} \{ \text{loss}(\theta_{EF}) + \alpha \| \theta_{LFE} - \theta_{LEF} \| \} \quad (14)$$

ただし θ_{FE} (θ_{EF}) は原言語から目的言語 (目的言語から原言語) へのアライメントモデルにおける各層の重みを表し、 θ_L は入力層 (Lookup層)、すなわちword embeddingの重みを表し、 α は合意制約の強さを制御するパラメータを示す。「 $\| \theta \|$ 」はノルムを表す。後述する実験では2-ノルムを使用した。式(13)及び式(14)はいずれも教師付き学習及び教師なし学習の双方に利用できる。それぞれ式(7)及び式(12)を式(13)、式(14)の $\text{loss}(\theta)$ に代入すればよい。

[0059] 図5を参照して、本実施の形態における学習を実現するプログラムは、以下のような制御構造を有する。なお、後述するようにこの実施の形態に係る単語アライメントモデル学習装置、及び単語アライメント装置は、コンピュ

ータハードウェアと、そのコンピュータハードウェアにより実行されるコンピュータプログラムとにより実現されることが一般的でかつ現実的である。しかし、これらはいずれも専用のハードウェアにより実現してもよいし、プログラマブルロジックデバイスにより実現してもよい。

[0060] このプログラムは、まず、予めパラメータを記入した設定ファイルを読み、バッチサイズ D 、 N 、 C 、 W 、 α を設定する（ステップ150）。次に、 θ_{FE}^1 、 θ_{EF}^1 と、最大繰返し回数を表す定数 $MaxIter$ との入力を受け、あわせて学習データ T と $IBM1$ を読み込む（ステップ152）。ここで $IBM1$ は、原言語及び目的言語の各単語に対して、 $l0$ priorを組み込んだ IBM モデル1によって見出された翻訳候補の単語を並べたリストである。このプログラムはさらに、上記ステップに続いて、 $1 \leq t \leq MaxIter$ を満たす全ての t に対して以下の処理を繰返すステップ154と、ステップ154が完了したときに得られる値 $\theta_{EF}^{MaxIter+1}$ 及び $\theta_{FE}^{MaxIter+1}$ を出力して処理を終了するステップ156とを含む。

[0061] ステップ154で各 t に対して繰返される処理は、学習データ T から D 個の対訳文対 $(f^+, e^+)^D$ をサンプリングするステップ170と、 $l0$ priorを組み込んだ IBM モデル1によって見出された f^+ 中の各単語の翻訳候補（ $IBM1$ ）に基づいて、各 f^+ に対して N 個の擬似的ネガティブサンプルを生成するステップ172と、同様の処理により、 e^+ 中の各単語の翻訳候補に基づいて、各 e^+ に対して N 個の擬似的ネガティブサンプルを生成するステップ174とを含む。さらに、ステップ176及び178により、前述した目的関数にしたがってニューラル・ネットワークの各層の重みが更新される。 θ_{EF}^t 及び θ_{FE}^t は各繰返しで同時に更新され、 θ_{FE}^t と θ_{EF}^t を更新する際には、それぞれ、word embeddingを相互に合意（一致）させるために θ_{EF}^{t-1} と θ_{FE}^{t-1} の値が用いられる。

[0062] 図6に、このRNNによるモデルを用いて対訳コーパスの単語アライメントを行なうコンピュータプログラムの制御構造を示す。これにより、例えば、統計的機械翻訳装置の学習に用いる学習データ（対訳コーパス）に単語ア

ライメントを付与できる。図6を参照して、このプログラムは、RNNによるモデルの学習済パラメータを読み、RNNを初期化するステップ190と、対訳コーパスを開くステップ192と、全ての対訳 (f_1^J, e_1^I) に対して、単語アライメントを実行するステップ194と、ステップ194により全ての対訳に対して単語アライメントが終了した学習データファイルを閉じて処理を終了するステップ196とを含む。

[0063] ステップ194は、処理中の対訳 (f_1^J, e_1^I) について考えられる全てのアライメント ($a_{1^J,k}$) ($k=1\sim K$) に対して後述するスコア s_k を算出し記憶するステップ200と、全てのアライメント ($a_{1^J,k}$) ($k=1\sim K$) のうち、ステップ200により記憶されたスコア s_k が最大となるアライメントを処理中の対訳に対するアライメントとして選択し、対訳にラベルを付して学習データを更新するステップ202とを含む。

[0064] ステップ200は、スコア s_k を1に初期化するステップ210と、変数 $j=1\sim J$ の順序で順番に変化させて原言語の単語 f_j を選択しながら、各 j に対し、単語 f_j と、その単語 f_j に対してアラインメント a_j によりアラインされる単語 $e_{a_{\{j\}}}$ について、それまでに得られたアライメント結果を全て ($a_{1^{j-1}}$) 考慮してスコア s_k を更新していき、最終的なスコア s_k を算出するステップ212と、ステップ212で算出されたスコア s_k を k 番目のアライメントに対するスコアとして記憶するステップ214とを含む。

[0065] ステップ212は、スコア $t_{RNN}(a_j | a_{1^{j-1}}, f_j, e_{a_{\{j\}}})$ を算出するステップ220と、スコア s_k にスコア t_{RNN} を乗じてスコア s_k を更新するステップ222とを含む。

[0066] [第2の実施の形態]

第1の実施の形態は、RNNによるモデルに対して双方向の合意制約を課したものである。しかし本発明はそのような合意制約を課したものには限定されない。合意制約のない一方向のRNNモデルを用いるものでもよい。

[0067] 図7に、この第2の実施の形態で使用するモデルの学習を実現するコンピュータプログラムの制御構造を示す。図7に示すプログラムは、図5に示

すステップ154及びステップ156に代えて、ステップ240及び242を含んでいる。ステップ240は、実質的には、図5のステップ154からステップ174及びステップ178を取り除き、2つの変数のうち、一方の変数のみを計算するようにしたものである。ステップ242は、図5のステップ156で出力される2つの変数のうち、ステップ240で計算される一方の変数の値のみを出力する点のみがステップ156と異なっている。

[0068] 図5におけるステップ176及び178は、それぞれ、式(13)、式(14)を目的関数として各重みを更新するのに対し、図7のステップ176では、式(7)又は(12)を目的関数とする点が異なる。図7において、その他個々のステップで行なわれる処理はそれぞれ図5に示すもので実行されるものと同じである。

[0069] [実験]

上記実施の形態に記載した単語アライメント方法の性能を調べるために実験を行なった。実験では、コーパスBTEC (Toshiyuki Takezawa, Eiichiro Sumita, Fumiaki Sugaya, Hirofumi Yamamoto, and Seiichi Yamamoto. 2002. Toward a Broad-coverage Bilingual Corpus for Speech Translation of Travel Conversations in the Real World. In Proceedings of the 3rd International Conference on Language Resources and Evaluation, pages 147-152.)での日本語と英語との単語アライメントと、2003 NAACL 共通タスク (Rada Mihalcea and Ted Pedersen. 2003. An Evaluation Exercise for Word Alignment. In Proceedings of the HLT-NAACL 2003 Workshop on Building and Using Parallel Texts: Data Driven Machine Translation and Beyond, pages 1-10.)で用いられたHansardsデータセット (Hansards)でのフランス語と英語とのアライメントとを行なった。さらに、FBISコーパスにおける中国語から英語への翻訳タスク、IWSLT2007における日本語から英語への翻訳タスク、及びNTCIR日本語-英語特許翻訳タスクについてもその翻訳性能を評価した。

[0070] 図8に、使用したデータセットのサイズを表形式で示す。BTEC及びHansardsについては、予備実験によりアライメントタスクのためのハイパーパラメータが設定済だったため、開発セットは用いていない。BTECデータはIWSLTの学習データ中の最初の9,960個の対訳文対であり、Gohら(Chooi-Ling Goh, Taro Watanabe, Hirofumi Yamamoto, and Eiichiro Sumita, 2010, Constraining a Generative Word Alignment Model with Discriminative Output, IEICE Transactions, 93-D(7):1976-1983.)により正しい単語アライメントが付されている。このデータを2つに分割した。先頭の9000文対からなる学習データと、最後の960文対からなるテストデータである。BTECの全データについては単語アライメントがされているが、Hansards中の学習データはラベル付けがされていない。FBI Sについては、NIST 02評価データを開発データとして用い、NIST 03及び04の評価データをテストデータとして用いた。

[0071] 〈比較対象〉

上記実施の形態に係るRNNによるアライメントモデルについて、2つのベースラインモデルと対比した。第1はIBMモデル4であり、第2は隠れ層を1つ持つFFNNによるモデルである。IBMモデル4はOch及びNey (Franz Josef Och and Hermann Ney, 2003, A Systematic Comparison of Various Statistical Alignment Models, Computational Linguistics, 29:19-51.)により示されたモデルシーケンス ($1^5 H^5 3^5 4^5$: IBMモデル1による5回の繰返し、その後にHMMモデルによる5回の繰返し、...)により学習したものであり、これはGIZA++に対するデフォルトの設定である(IBM4)。FFNNによるモデルについては、word embedding長Mを30に、隠れ層のユニット数 $|z_1|$ を100に、窓幅を5に、それぞれ設定した。Yangら(Nan Yang, Shujie Liu, Mu Li, Ming Zhou, and Nenghai Yu, 2013, Word Alignment Modeling with Context Dependent Deep Neural Network, In Proceedings of t

he 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 166—175.) の教示にしたがい、既に説明したような教師あり学習によって F F N N によるモデルの学習を行なった。

[0072] R N N によるモデルについては、word embedding 長 M を 30 に、隠れ層のユニット数 $|y_i|$ を 100 に、それぞれ設定した。評価実験では、R N N による 4 つのモデル、すなわち $R N N_s$ 、 $R N N_{s+c}$ 、 $R N N_u$ 、及び $R N N_{u+c}$ の性能を評価した。ここで「s / u」は教師あり / 教師なし学習によるモデルであることを示し、「+ c」は合意制約が課されているか否かを示す。

[0073] I B M 4 を除くモデルに関する学習では、各層の重みを最初に初期化した。すなわち、入力層 (Lookup 層) L の重みについては原言語及び目的言語の双方について word embedding を学習データのそれぞれの側から学習し、得られた word embedding を L に設定した。これにより局所最適解に陥ってしまうことが避けられる。他の重みについては、閉区間 $[-0.1, 0.1]$ の範囲内でランダムな値に設定した。word embedding の学習では Mikolov ら (Tomas Mikolov, Martin Karafiat, Lukas Burget, Jan Cernocky, and Sanjeev Khudanpur. 2010. Recurrent Neural Network based Language Model. In Proceedings of 11th Annual Conference of the International Speech Communication Association, pages 1045-1048.) に基づく R N N L M ツールキット (<http://www.fit.vutbr.cz/~imikolov/>) をデフォルト設定で用いた。ただしここでは、5 回未満しか出現しない単語は全て <unk> という特別なトークンにまとめた。次に、重みの各々をミニバッチ S G D を用いて最適化した。ここでのバッチサイズ D は 100 に、学習率は 0.01 に、 L_2 正規化パラメータは 0.1 に、それぞれ設定した。学習は 50 世代の繰返し後に終了した。他のパラメータは以下のとおりである。教師なし学習におけるパラメータ W 、 N 及び C はそれぞれ 100, 50 及び 0.001 に設定した。合意制約の強度を示すパラメータ α は 0.1 に設定した。

[0074] 翻訳タスクでは、K o e h n ら (Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constanin, and Evan Herbst. 2007. Moses: Open Source Toolkit for Statistical Machine Translation. In Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics on Interactive Poster and Demonstration Sessions, pages 177-180.) によるフレーズベースのSMTを用いた。日本語及び中国語の文は全て、Chasen及びStanford segmenterを使用してそれぞれ単語分割を行なった。学習では、40語を超える長い文は除いた。改良型のKneser-Ney (modified Kneser-Ney) 平滑化を用いるSRILMツールキット (Stolcke, 2002) を使用して、学習データの英語側を用いてIWSLT及びNTCIRのための5グラム言語モデルの学習を行ない、English Gigaword CorpusのXinhua側を用いてFBISのための5グラム言語モデルの学習を行なった。SMTの重みパラメータは、開発データを用いてMERT (Franz Josef Och, 2003. Minimum Error Rate Training in Statistical Machine Translation. In Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics, pages 160-167.) を用いて調整した。

[0075] 〈単語アライメントの結果〉

図9に、単語アライメントの性能をF値により評価したものを示す。以下の説明では、MODEL (R) 及びMODEL (I) はそれぞれ、正解アライメント、IBMモデル4による単語アライメントを教師データとして学習したモデルを示す。Hansardsでは、学習データからランダムにサンプリングした100Kのデータにより全てのモデルの学習を行なった。単語アライメントの評価は、最初に各モデルを双方向に適用しそれぞれの方向のアライメントを解析した後、「grow-diag-final-and」ヒューリスティック (Philipp Koehn, Franz Josef Och, and Daniel Marcu, 2003. Stat

istical Phrase-Based Translation. In Proceedings of the 2003 Human Language Technology Conference: North American Chapter of the Association for Computational Linguistics, pages 48-54.) によってそれらのアライメントを組合わせたものに対して行なった。単語アライメントの性能に関する有意性検定は、有意レベル5%で符号検定により行なった。図9の「+」は、ベースライン (IBM4 及び FFNN_s (R/I)) との間に有意な差があることを示す。

[0076] 図9において、RNN_{utc}は、上記実施の形態に記載した全ての特徴を備えたものである。すなわち、RNNによるモデル、教師なし学習、及び合意制約を全て適用したものである。そしてこのRNN_{utc}が、BTEC及びHansardsの双方において最もよい性能を示した。特に、ベースラインのシステムとの間の差は統計的に有意であった。図9からはまた、BTECについてはRNN_s (R/I) がFFNN_s (R/I) に対して統計的に有意な差を示したことが分かる。これらの結果から、RNNによるモデルにより、直前だけでなくそれ以前のアライメント結果を考慮して単語アライメントを行なうことにより、性能が向上することが確認できる。

[0077] 図9によれば、RNN_{s+c} (R/I) 及びRNN_{utc}が、いずれのタスクにおいてもそれぞれRNN_s (R/I) 及びRNN_uよりよい性能を示している。この結果から、合意制約を課すとモデルの性能が向上することが分かる。

[0078] BTECでは、RNN_u及びRNN_{utc}の性能はそれぞれRNN_s (I) 及びRNN_{s+c} (I) の性能をかなり上回っているが、Hansardsではこれらの性能の間には大きな差はない。これは、学習データの自動アライメント結果を教師データとして使ったときのように、教師データの精度が高くないときには、上記実施の形態の教師なし学習が効果を持つことを示している。

[0079] 〈機械翻訳の結果〉

図10は、大文字と小文字を区別したBLEU値 (Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: a Method for Automatic Evaluation of Machine Translation. In Proce

edings of 40th Annual Meeting of the Association for Computational Linguistics, pages 311-318.) により測定した翻訳精度を示す。各値は、3回の独立のMERT処理で得られたBLEU値の算術平均である。

- [0080] NTCIR及びFBISでは、各アライメントモデルをランダムにサンプリングした100Kのデータにより学習した後に、このアライメントモデルを用いて単語アライメントされた学習データ全てを用いて翻訳モデルの学習を行なった。さらに、全ての学習データを用いて学習したIBMモデル4によるSMTシステム(IBM4_{all})の評価も行なった。翻訳性能に関する有意性検定は、有意レベル5%でbootstrapによる検定手法(Philipp Koehn, 2004, Statistical Significance Tests for Machine Translation Evaluation, In Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, pages 388-395.)により行なった。図10において「*」は、翻訳性能が2つのベースライン、IBM4及びFFNN_s(l)の双方に対して有意な差を持つことを示す。
- [0081] 図10の結果から、RNN_u及びRNN_{u+c}は全てのタスクでFFNN_s(l)及びIBM4を上回る性能を示していることが分かる。したがって、RNNによるモデルを用いて単語アライメントした学習データにより翻訳モデルの学習を行なうことで、翻訳性能が向上することが分かる。さらに、図10からは、RNNによるモデルは、NTCIR及びFBISタスクでは、学習データの一部しか用いずに学習を行なっただけであるにもかかわらず、IBM4_{all}に匹敵する性能を示すことが分かる。
- [0082] 図11に、日本語と英語の対訳文に対するRNN_s(R)による単語アライメントの結果を○で、FFNN_s(R)による単語アライメントの結果を△で、正しいアライメントを実線の□で、それぞれ示す。図11からは、RNN_s(R)によるモデルが、遠く離れた2つの単語を含む複雑なアライメントをほぼ正確に行なっていることがわかる。FFNN_s(R)では特に、遠く離れた2つの単語のアライメントがうまく行なわれていない。これは、FFNN_s

(R) が直前のアライメントの結果しか用いていないのに対し、 $RNN_s(R)$ (R) が、直前だけでなくそれ以前のアライメントの結果も用いており、ある意味ではフレーズレベルでのアライメントを行なっているといえることによる。

[0083] 図12に、フランス語と英語の対訳文に対する $RNN_s(R)$ と $FFNN_s(R)$ による単語アライメントの結果を示す。英語とフランス語とでは語順が似通っているため、直前のアライメント結果が単語アライメントの大きな手がかりになっており、 $FFNN_s(R)$ でもアライメントを精度高くできることがわかる。したがってこのような言語対では $RNN_s(R)$ を用いてもそれほど大きな効果は得られないと考えられる。

[0084] 〈学習データサイズ〉

図13に、学習データのサイズを変えて学習した各モデルを用いてBTECの単語アライメントした結果(F値)を示す。学習データは、IWSLT用の学習データ(40K)、BTEC用の学習データ(9K)、及びBTECの学習データからランダムにサンプリングした学習データ(1K)である。なお、IWSLTについてはアライメントの正解データがないため、 $RNN_{s+c}(R)$ についてはIWSLT用のデータ(40K)を用いた学習はできなかった。

[0085] 図13によれば、上記実施の形態に係る RNN によるモデルの性能は、1Kのラベル付きデータで学習した場合も、9Kのラベル付きデータで学習した場合も、40Kのラベルなしの学習データにより学習したIBMモデル4を上回る。すなわち、 RNN によるモデルは、IBMモデル4の学習データの25%にも満たないデータで学習してもIBMモデル4を上回る性能を示す。したがって、ごく少ない学習データにより学習した RNN_{s+c} モデルを用いたSMTによれば、図10に示すような学習データの全体を用いて学習したIBMモデル4によるSMTと同等以上の性能を得ることができる。

[0086] 以上のように上記実施の形態に係る RNN によるモデルを用いると、直前の単語アライメントの結果だけではなく、それ以前の単語アライメントの結

果を用いて単語のアライメントを行なえる。その結果、実験で明らかになったように、このモデルを用いた単語アライメントは従来の、直前の単語アライメントのみに依存して単語アライメントを行なうモデルよりも高い性能を示す。しかもこのモデルは教師なし学習での学習も可能で、その結果得られるモデルの性能も高い。さらに、学習に必要とされる学習データの数が少なくても従来のモデルと同等以上の性能が得られる。

[0087] [コンピュータによる実現]

上記した単語アライメントモデル学習装置及び単語アライメント装置は、実質的には、コンピュータハードウェアと、コンピュータハードウェアと協働するコンピュータプログラムとにより実現できる。

[0088] <ソフトウェア構成>

ソフトウェア構成については、図5、図6及び図7に記載したとおりである。

[0089] <ハードウェア構成>

図14は、上記した単語アライメントモデル学習装置及び単語アライメント装置を実現する、コンピュータ340を含むコンピュータシステム330の外観を示し、図15はコンピュータ340の内部構成を示す。

[0090] 図14を参照して、このコンピュータシステム330は、コンピュータ340と、このコンピュータ340に接続されるモニタ342、キーボード346、及びマウス348とを含む。コンピュータ340は、DVD362（図15を参照）が装着可能なDVDドライブ350と、リムーバブルメモリ364（図15を参照）が装着されるメモリポート352とを有する。

[0091] 図15を参照して、コンピュータ340は、上記したDVDドライブ350及びメモリポート352に加え、CPU（中央演算処理装置）356と、CPU356に接続されたバス366と、ブートアッププログラム等を記憶する、書換え可能な読出専用メモリ（ROM）358と、プログラム命令、システムプログラム、及び作業データ等を記憶するランダムアクセスメモリ（RAM）360と、コンピュータプログラム及び対訳コーパスなどを記憶

するハードディスクドライブ（HDD）354と、他の端末との通信を可能とするネットワークインターフェイス（I/F）344を含む。

[0092] コンピュータシステム330を上記した実施の形態に係る単語アライメントモデル学習装置及び単語アライメント装置の各機能部として機能させるためのコンピュータプログラムは、予めリムーバブルメモリ364に記憶され、リムーバブルメモリ364をメモリポート352に装着した後にROM358の書換えプログラムを起動することによりROM358又はHDD354に転送され記憶される。又は、プログラムを、ネットワークI/F344を介した通信によりネットワーク上の他の装置からRAM360に転送し、その後にROM358又はHDD354に書き込むようにしてもよい。プログラムは実行の際にROM358又はHDD354から読出され、RAM360にロードされ、CPU356により実行される。

[0093] ROM358又はHDD354に記憶されたプログラムは、コンピュータ340を、上記実施の形態に係る単語アライメントモデル学習装置及び単語アライメント装置の各機能部として機能させるための複数の命令からなる命令列を含む。この動作を行なわせるのに必要な基本的機能のいくつかはコンピュータ340上で動作するオペレーティングシステム若しくはサードパーティのプログラム、又は、コンピュータ340にインストールされている各種プログラミングツールキット又はプログラムライブラリにより実行時に動的に提供されることがある。したがって、このプログラム自体はこの実施の形態に係る単語アライメントモデル学習装置及び単語アライメント装置を実現するのに必要な機能全てを必ずしも含まなくてよい。このプログラムは、命令のうち、所望の結果が得られるように制御されたやり方で適切な機能又はプログラミングツールキット内の適切なプログラムツールをコンピュータ340の記憶装置内から動的に呼出すことにより、上記したシステムとしての機能を実現する命令のみを含んでいればよい。もちろん、プログラムのみで必要な機能を全て提供するようにしてもよい。

[0094] コンピュータプログラムを実行する際のコンピュータシステム330の動

作は周知である。したがってここではその詳細については繰返さない。

[0095] 上記実施の形態では、まず英語文の先頭から単語を順番に選択して行き、アライメントにしたがって各アライメントのスコアを算出した。しかし本発明はそのような実施の形態には限定されない。単語を選択する順番は任意でよく、全単語を所定の順番で選択できればどのような順番でもよい。もっとも、上記実施の形態のように一方の言語の文の先頭の単語から順番に他方の言語の単語とアライメントしていくのが簡単である。

[0096] また、上記実施の形態ではリカレント型のニューラル・ネットワークの各層の関数として特定の関数を用いた。しかし本発明はそのような実施の形態には限定されない。例えば隠れ層において、二つの単語の間の非線形的関係を表現できるものであればどのような関数を用いてもよい。入力層及び出力層についても同様である。また上記した実施の形態では、ニューラル・ネットワークの出力はスコアであったが、前述したとおりこれは2つの単語が正しい対応付けになっている確率であってもよい。確率もある種のスコアと考えることができる。

[0097] 上記実施の形態では、RNN型のニューラル・ネットワークの学習と単語アライメントとを同一のコンピュータで実行するものとした。しかしもちろん本発明はそのような実施の形態には限定されない。学習により得られたニューラル・ネットワークのパラメータを他のコンピュータにコピーしてRNN型のニューラル・ネットワークをセットアップすることにより、任意のコンピュータで単語アライメントを行なうことができる。

[0098] 今回開示された実施の形態は単に例示であって、本発明が上記した実施の形態のみに制限されるわけではない。本発明の範囲は、発明の詳細な説明の記載を参酌した上で、請求の範囲の各請求項によって示され、そこに記載された文言と均等の意味及び範囲内での全ての変更を含む。

産業上の利用可能性

[0099] この発明は、2つの文、例えば異なる言語の2文の間で対応する単語を特定するために利用でき、典型的には、統計的言語翻訳のための翻訳モデルの

学習データの作成、翻訳者及び翻訳監督者のための翻訳の検証装置、2つの文書の間と比較のための装置等に利用できる。

符号の説明

- [0100] 2 0 対訳文対
3 0 日本語文
3 2 英語文
4 0, 4 2 アライメント
6 0 ニューラル・ネットワーク
7 0, 1 1 0 入力層 (Lookup層)
7 2, 1 1 2 隠れ層
7 4, 1 1 4 出力層
7 6 語彙翻訳スコア
8 0, 1 3 0 原言語の入力部
8 2, 1 3 2 目的言語の入力部
1 0 0 RNNによるモデル
1 0 2 スコア
1 1 8 循環接続

請求の範囲

- [請求項1] 第1の言語と第2の言語との対訳文対の単語アライメントのスコアを算出する単語アライメントスコア算出装置であって、
- 前記対訳文対と、当該対訳文対に対する単語アライメントとを受け、所定の順序にしたがって、前記対訳文対の前記第1の言語の文の単語を順番に選択する選択手段と、
- 前記選択手段により選択された単語と、前記単語アライメントにより当該単語と対応付けられた前記第2の言語の単語とからなる単語対が正しい単語対である可能性を示すスコアを前記対訳文対の前記第1の言語の文の全ての単語について算出し、当該スコアに基づいて、前記単語アライメントのスコアを算出するスコア算出手段とを含み、
- 前記スコア算出手段は、ある単語対のスコアを算出するにあたって、前記単語アライメントのうち、当該ある単語対を構成する前記第1の言語の単語より前に前記選択手段により選択された単語のアライメントの全てに基づいて当該ある単語対のスコアを算出する、単語アライメントスコア算出装置。
- [請求項2] 前記選択手段は、前記対訳文対の内、前記第1の言語の文の単語を前記第1の言語の文の先頭から順番に選択する手段を含む、請求項1に記載の単語アライメントスコア算出装置。
- [請求項3] 前記スコア算出手段は、
- 前記選択手段により選択された単語と、前記対訳文対の前記第2の言語の文のうちで前記単語アライメントにより当該単語と対応付けられた単語とからなる単語対が正しい単語対である可能性を示すスコアを算出する第1の算出手段と、
- 前記対訳文対の前記第1の言語の文の全ての単語について前記第1のスコア算出手段により算出されたスコアに基づいて、前記単語アライメントのスコアを算出する第2の算出手段とを含む、請求項1又は請求項2に記載の単語アライメントスコア算出装置。

[請求項4] 前記第2の算出手段は、前記対訳文対の前記第1の言語の文の全ての単語について前記第1の算出手段により算出されたスコアを互いに乗算することにより、前記単語アライメントのスコアを算出する手段を含む、請求項3に記載の単語アライメントスコア算出装置。

[請求項5] 前記スコア算出手段は、前記選択手段により選択された単語を受ける第1の入力と、前記単語アライメントにより当該単語に対応付けられた前記第2の言語の単語を受ける第2の入力とを持つリカレント型のニューラルネットワークと、

前記選択手段により選択された単語と、前記単語アライメントにより当該単語に対応付けられた単語とを、前記第1の入力及び前記第2の入力にそれぞれ与える入力制御手段とを含み、

前記リカレント型のニューラルネットワークは、

前記第1の入力及び前記第2の入力を持ち、前記第1の入力及び前記第2の入力にそれぞれ与えられた単語からword embeddingベクトルを算出し出力する入力層と、

前記入力層の出力を受け、予め定められた非線形演算により、前記入力層からの2つの出力の関係を表すベクトルを生成する隠れ層と、

前記隠れ層の出力に基づいて前記スコアを算出し出力する出力層とを含み、

前記隠れ層の出力は、前記単語アライメントスコア算出装置に次の単語対が入力として与えられたときに、前記隠れ層に入力として与えられる、請求項1～請求項4のいずれかに記載の単語アライメントスコア算出装置。

[請求項6] 第1の言語と第2の言語との対訳文対の単語アライメントを推定する単語アライメント装置であって、

請求項1～請求項5のいずれかに記載の単語アライメントスコア算出装置と、

前記対訳文対に対して複数個の単語アライメント候補を生成する単

語アライメント候補生成手段と、

前記単語アライメント候補生成手段により生成された前記複数個の単語アライメント候補の各々について、前記単語アライメントスコア算出装置を用いて前記対訳文対に対する単語アライメントスコアを算出する算出手段と、

前記算出手段により前記複数個の単語アライメント候補に対して算出された単語アライメントスコアのうち、最も高いスコアに対応する単語アライメント候補を、前記対訳文対の単語アライメントとして決定し出力する単語アライメント決定手段とを含む、単語アライメント装置。

【請求項7】 コンピュータにより実行されると、当該コンピュータを、請求項1～請求項6のいずれかに記載の各手段として機能させる、コンピュータプログラム。