

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6127778号
(P6127778)

(45) 発行日 平成29年5月17日(2017.5.17)

(24) 登録日 平成29年4月21日(2017.4.21)

(51) Int.Cl.

F I

G 0 6 N 99/00 (2010.01)

G 0 6 N 99/00 1 5 3

請求項の数 4 (全 36 頁)

(21) 出願番号 特願2013-136647 (P2013-136647)
 (22) 出願日 平成25年6月28日(2013.6.28)
 (65) 公開番号 特開2015-11535 (P2015-11535A)
 (43) 公開日 平成27年1月19日(2015.1.19)
 審査請求日 平成28年3月10日(2016.3.10)

(73) 特許権者 000005223
 富士通株式会社
 神奈川県川崎市中原区上小田中4丁目1番
 1号
 (74) 代理人 100103528
 弁理士 原田 一男
 (72) 発明者 岩倉 友哉
 神奈川県川崎市中原区上小田中4丁目1番
 1号 富士通株式会社内
 審査官 多胡 滋

最終頁に続く

(54) 【発明の名称】 モデル学習方法、モデル学習プログラム及びモデル学習装置

(57) 【特許請求の範囲】

【請求項 1】

コンピュータが、

各学習事例の重みに基づいて、前記重みに対応する各学習事例から夫々の正解構造を抽出する際の手掛かりとなる素性の各々に対応する第1スコアを更新することによって、当該第1スコアの各々により規定されるモデルを算出する第1ステップと、

前記各学習事例について、当該学習事例から導かれる複数の候補構造のうち前記正解構造以外の前記候補構造の各々に対して、前記モデルにより付与される第2スコアのうち最大スコアを特定し、当該学習事例の前記正解構造に対して前記モデルにより付与される第2スコアと前記最大スコアとの差を求めると共に、前記各学習事例における前記差により定義されるエラーの上限値に基づいて、前記モデルの確信度を算出する第2ステップと、

前記確信度と前記各学習事例における前記差とに基づいて、前記各学習事例の前記重みを更新する第3ステップと、

前記第1ステップと、前記第2ステップと、前記第3ステップとを繰り返すステップと、

を実行するモデル学習方法。

【請求項 2】

前記第1ステップにおいて、前記各学習事例の前記重みの各々を、当該学習事例の学習率に設定して、前記第1スコアを更新する

請求項1記載のモデル学習方法。

【請求項 3】

各学習事例の重みに基づいて、前記重みに対応する各学習事例から夫々の正解構造を抽出する際の手掛かりとなる素性の各々に対応する第 1 スコアを更新することによって、当該第 1 スコアの各々により規定されるモデルを算出する第 1 ステップと、

前記各学習事例について、当該学習事例から導かれる複数の候補構造のうち前記正解構造以外の前記候補構造の各々に対して、前記モデルにより付与される第 2 スコアのうち最大スコアを特定し、当該学習事例の前記正解構造に対して前記モデルにより付与される第 2 スコアと前記最大スコアとの差を求めると共に、前記各学習事例における前記差により定義されるエラーの上限値に基づいて、前記モデルの確信度を算出する第 2 ステップと、

前記確信度と前記各学習事例における前記差とに基づいて、前記各学習事例の前記重みを更新する第 3 ステップと、

前記第 1 ステップと、前記第 2 ステップと、前記第 3 ステップとを繰り返すステップと、

を含み、コンピュータに実行させるためのモデル学習プログラム。

【請求項 4】

重みデータ格納部に格納されている各学習事例の重みに基づいて、学習データ格納部に格納されている前記重みに対応する各学習事例から夫々の正解構造を抽出する際の手掛かりとなる素性の各々に対応する第 1 スコアを更新することによって、当該第 1 スコアの各々により規定されるモデルを算出し、当該モデルをモデルデータ格納部に格納する第 1 算出手段と、

前記学習データ格納部に格納されている前記各学習事例について、当該学習事例から導かれる複数の候補構造のうち前記正解構造以外の前記候補構造の各々に対して、前記モデルデータ格納部に格納されている前記モデルにより付与される第 2 スコアのうち最大スコアを特定し、当該学習事例の前記正解構造に対して前記モデルにより付与される第 2 スコアと前記最大スコアとの差を求め、当該差を差データ格納部に格納すると共に、前記各学習事例における前記差により定義されるエラーの上限値に基づいて、前記モデルの確信度を算出し、当該確信度を確信度データ格納部に格納する第 2 算出手段と、

前記確信度データ格納部に格納されている前記確信度と前記差データ格納部に格納されている前記各学習事例における前記差とに基づいて、前記重みデータ格納部に格納されている前記各学習事例の前記重みを更新する更新手段と、

を有し、

前記第 1 算出手段による処理と前記第 2 算出手段による処理と前記更新手段による処理とを繰り返させるモデル学習装置。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、機械学習の技術に関する。

【背景技術】

【0002】

最初に、構造予測を行う自然言語処理のタスクの例として、単語列から品詞列へのマッピング、単語列からフレーズ列へのマッピング及び単語列から固有名詞列へのマッピングについて、順次説明する。

【0003】

図 1 に、単語列から品詞列へのマッピングの様子を示す。この例における単語列 11 は、「T a r o」、「Y a m a d a」及び「s i n g s」の単語を順に含んでいる。そして、この図は、これらの単語に名詞（図中、「N」と記す）あるいは動詞（図中、「V」と記す）のような品詞が対応付けられる様子を示している。この例では、単語列 11 に含まれる各単語に対して、品詞列 13 に含まれる「名詞」、「名詞」及び「動詞」が対応付けられている。

【0004】

図2に、単語列からフレーズ列へのマッピングの様子を示す。単語列11は、図1の場合と同様である。そして、この図は、これらの単語列からフレーズ群を抽出し、それらのフレーズ群に名詞句（図中、「NP」と記す）あるいは動詞句（図中、「VP」と記す）のような品詞句が対応付けられる様子を示している。この例では、フレーズ列21中の「Taro Yamada」の第1フレーズと「sings」の第2フレーズとが抽出され、それぞれのフレーズに「名詞句」と「動詞句」とが対応付けられている。

【0005】

図3に、単語列から固有名詞列へのマッピングの様子を示す。単語列11は、図1の場合と同様である。そして、この図は、これらの単語列に含まれる単語あるいはフレーズに、人名（図中、「P」と記す）あるいはそれ以外（図中、「O」と記す）が対応付けられる様子を示している。この例では、固有名詞列31に示すように「Taro Yamada」のフレーズが人名であると判定され、「sings」が人名以外であると判定されている。

【0006】

これらのタスクの実現方法として、教師有り学習手法が適用されてきた。教師有り学習手法では、上述のような単語列とその正解構造（たとえば、ラベル列）とが学習データとして与えられ、この学習データを用いて単語列から構造を正しくマッピングできるように学習する。例えば、分類器を組み合わせる方法の場合には、各単語にラベルを付与することで、最終的な出力を決定できるようにする。

【0007】

図4に、分類器によるマッピングの様子を示す。単語列11は、図1の場合と同様である。この例では、学習によって、単語に対してラベルを付与する分類器の組み合わせを得て、その分類器の組み合わせによって、判定対象である単語列に含まれる各単語にラベルを対応付ける。

【0008】

この例では、「名詞句の先頭」（図中、「B-NP」と記す）、「名詞句の先頭以外」（図中、「I-NP」と記す）、「動詞句の先頭」（図中、「B-VP」と記す）及び「動詞句の先頭以外」（図中、「I-VP」と記す）の4種類のラベルを用いる。

【0009】

「名詞句の先頭」の後に、「名詞句の先頭」又は「動詞句の先頭」が続く場合には、先の「名詞句の先頭」に対応する単語が、単独で名詞句に相当することを意味する。

【0010】

「名詞句の先頭」の後に、一つあるいは複数の「名詞句の先頭以外」が続く場合には、その「名詞句の先頭」に対応する単語から、最後尾の「名詞句の先頭以外」に対応する単語までの単語列からなるフレーズが、名詞句に相当することを意味する。

【0011】

「動詞句の先頭」の後に、「名詞句の先頭」又は「動詞句の先頭」が続く場合には、先の「動詞句の先頭」に対応する単語が、単独で動詞句に相当することを意味する。

【0012】

「動詞句の先頭」の後に、一つあるいは複数の「動詞句の先頭以外」が続く場合には、その「動詞句の先頭」に対応する単語から、最後尾の「動詞句の先頭以外」に対応する単語までの単語列からなるフレーズが、動詞句に相当することを意味する。

【0013】

この例では、ラベル列41に示すように、「Taro」に「名詞句の先頭」が付与され、「Yamada」に「名詞句の先頭以外」が付与され、「sings」に「動詞句の先頭」が付与されている。その結果、フレーズ列43に示すように、「Taro Yamada」が名詞句であると判定され、「sings」が動詞句であると判定される。

【0014】

また近年は、直接的に構造を予測する構造学習手法も用いられている。図5に、構造学習手法によるマッピングの様子を示す。この例では、各単語に対する4種類のラベルを選

10

20

30

40

50

択的に組み合わせたラベル列の候補群 5 1 の中から、正しいラベル列を直接選択する仕組みを学習する。ラベル列の選択は、図示するように各単語に対するラベル同士を結ぶパス群から、正しいパスを選択することに相当する。

【 0 0 1 5 】

この例では、「T a r o」に対する「名詞句の先頭」と、「Y a m a d a」に対する「名詞句の先頭以外」と、「s i n g s」に対する「動詞句の先頭」とからなるラベル列が選択され、その結果、フレーズ列 5 3 に示すように、「T a r o Y a m a d a」が名詞句であると判定され、「s i n g s」が動詞句であると判定される。

【 0 0 1 6 】

また、図 6 に、別の構造学習手法によるマッピングの様子を示す。単語列 1 1 は、図 1 の場合と同様である。この例は、単語の塊であるチャンクを単位としている。単語列を構成するチャンクに対する 4 種類のラベルを選択的に組み合わせたラベル列の候補群 6 1 の中から、正しいラベル列を直接選択する仕組みを学習する。つまり、ラベル列の選択は、図示するように各チャンクに対するラベル同士を結ぶパス群から、正しいパスを選択することに相当する。また、単語列が 1 つのチャンクとなることを想定して、そのチャンクのラベルを選択することもある。

【 0 0 1 7 】

この例では、フレーズ列 6 3 に示すように、「T a r o Y a m a d a」に対する名詞句と、「s i n g s」に対する動詞句とからなるラベル列が選択される。

【 0 0 1 8 】

これらの学習手法に加えて、更に判定精度を高めることを目的として、アンサンブル学習手法を採用している例もある。アンサンブル学習手法の一種であるブースティング手法では、複数のモデル（規則ともいう。）を学習し、それらのモデルを組み合わせることで精度の高い統合モデルを生成する。複数のモデルを学習するための学習器は、弱学習器と呼ばれ、その弱学習器で学習されたモデルは、弱仮説と呼ばれる。

【 0 0 1 9 】

ブースティング手法では、学習データに含まれる各学習事例に対して重みを設定する。そして、弱仮説によって正しい予測結果が得られる簡単な学習事例に対しては、その重みを小さくし、弱仮説によって正しい予測結果が得られない難しい学習事例に対しては、その重みを大きくするように調整する。このように各学習事例の重みを調整することによって、多様な学習事例へ適応した統合モデルが得られることが期待される。

【 0 0 2 0 】

構造予測にブースティング手法を採用するある例によれば、弱学習器として、前述したようにラベルを付与する分類器が用いられている。

【先行技術文献】

【特許文献】

【 0 0 2 1 】

【特許文献 1】特開 2 0 1 0 - 3 3 2 1 3 号公報

【非特許文献】

【 0 0 2 2 】

【非特許文献 1】Schapire, R. E. and Singer, Y.: BoosTexter: A boosting-based system for text categorization, Machine Learning, Vol. 39(2/3), pp. 135-168 (2000)

【非特許文献 2】Nagata, M.: A Stochastic Japanese Morphological Analyzer Using a Forward-DP Backward-A* N-Best Search Algorithm, COLING, pp. 201-207 (1994)

【非特許文献 3】Schapire, R. E. and Singer, Y.: Improved Boosting Algorithms Using Confidence-rated Predictions, Machine Learning, Vol. 37, No. 3, pp. 297-336 (1999)

【非特許文献 4】Cohen, W. W. and Sarawagi, S.: Exploiting dictionaries in named entity extraction: combining semi-Markov extraction processes and data integration methods, Proc. of KDD '04, pp. 89-98 (2004)

10

20

30

40

50

【非特許文献5】Sarawagi, S. and Cohen, W. W.: Semi-Markov Conditional Random Fields for Information Extraction, Proc. of NIPS '04 (2004)

【発明の概要】

【発明が解決しようとする課題】

【0023】

本発明の目的は、一側面では、構造学習の精度を高めることである。

【課題を解決するための手段】

【0024】

一態様に係るモデル学習方法は、(A)各学習事例の重みに基づいて、重みに対応する各学習事例から夫々の正解構造を抽出する際の手掛かりとなる素性の各々に対応する第1スコアを更新することによって、当該第1スコアの各々により規定されるモデルを算出する第1ステップと、(B)各学習事例について、当該学習事例から導かれる複数の候補構造のうち正解構造以外の候補構造の各々に対して、モデルにより付与される第2スコアのうち最大スコアを特定し、当該学習事例の正解構造に対してモデルにより付与される第2スコアと最大スコアとの差を求めると共に、各学習事例における当該差により定義されるエラーの上限値に基づいて、モデルの確信度を算出する第2ステップと、(C)確信度と各学習事例における上記差とに基づいて、各学習事例の重みを更新する第3ステップと、(D)第1ステップと、第2ステップと、第3ステップとを繰り返すステップとを含む。

【発明の効果】

【0025】

一側面としては、構造学習の精度を高めることができる。

【図面の簡単な説明】

【0026】

【図1】図1は、単語列から品詞列へのマッピングの様子を示す図である。

【図2】図2は、単語列からフレーズ列へのマッピングの様子を示す図である。

【図3】図3は、単語列から固有名詞列へのマッピングの様子を示す図である。

【図4】図4は、分類器によるマッピングの様子を示す図である。

【図5】図5は、構造学習手法によるマッピングの様子を示す図である。

【図6】図6は、構造学習手法によるマッピングの様子を示す図である。

【図7】図7は、公知技術のメイン処理フローを示す図である。

【図8】図8は、公知技術における学習データと規則の例を示す図である。

【図9】図9は、学習事例の例を示す図である。

【図10】図10は、正解構造の例を示す図である。

【図11】図11は、初期状態における学習事例の重みの例を示す図である。

【図12】図12は、初期状態における構造予測モデルの例を示す図である。

【図13】図13は、構造学習の様子を示す図である。

【図14】図14は、構造予測モデルを更新する様子を示す図である。

【図15】図15は、構造学習の様子を示す図である。

【図16】図16は、構造予測モデルを更新する様子を示す図である。

【図17】図17は、構造学習の様子を示す図である。

【図18】図18は、構造予測モデルを更新する様子を示す図である。

【図19】図19は、第2スコア算出の様子を示す図である。

【図20】図20は、第2スコア算出の様子を示す図である。

【図21】図21は、第2スコア算出の様子を示す図である。

【図22】図22は、学習事例の重みの例を示す図である。

【図23】図23は、構造学習の様子を示す図である。

【図24】図24は、構造予測モデルを更新する様子を示す図である。

【図25】図25は、構造学習の様子を示す図である。

【図26】図26は、構造予測モデルを更新する様子を示す図である。

【図27】図27は、構造学習の様子を示す図である。

10

20

30

40

50

【図 28】図 28 は、構造予測モデルを更新する様子を示す図である。

【図 29】図 29 は、モデル学習装置の機能ブロックを示す図である。

【図 30】図 30 は、モデル学習部の構成を示す図である。

【図 31】図 31 は、本発明のメイン処理フローを示す図である。

【図 32】図 32 は、初期状態における結合モデルの例を示す図である。

【図 33】図 33 は、構造学習処理フローを示す図である。

【図 34】図 34 は、構造予測モデル更新処理フローを示す図である。

【図 35】図 35 は、確信度算出処理フローを示す図である。

【図 36】図 36 は、コンピュータの機能ブロック図である。

【発明を実施するための形態】

10

【0027】

まず、本稿で用いるブースティング手法である Ada Boost (非特許文献 1) について説明する。学習の目的は、与えられた学習事例 X からラベル集合 Y へのマッピング F 、

【数 1】

$$F: X \rightarrow Y$$

を導出することである。ここでのラベルは、 $\{-1, +1\}$ である。

【0028】

図 7 に、Ada Boost のメイン処理フローを示す。学習システムは、学習データ S と、ブースティング繰り返し数 T とを受け付ける (S1)。学習データ S は、学習事例とラベルとのセットを n 個含み、 $\{(x_1, y_1), \dots, (x_n, y_n)\}$ と表される。

20

【0029】

【数 2】

$$x_i \in X$$

は、 i 番目のセットに係る学習事例を表し、

【数 3】

$$y_i \in Y$$

30

は、 i 番目のセットに係るラベルを表している。

【0030】

ブースティング繰り返し数 T は、ブースティングにおける繰り返しの回数である。

【0031】

学習システムは、学習事例の重み ($w_{1,1}, \dots, w_{1,n}$) を初期化する (S3)。学習事例の重み $w_{1,i}$ は、学習事例 x_i に対応している。各学習事例の重み $w_{1,i}$ ($1 \leq i \leq n$) に、初期値 $1/n$ が設定される。つまり、各初期の学習事例の重みは、均等とする。

【0032】

そして、カウンタ t を 1 に初期化する (S5)。

【0033】

40

学習システムは、弱学習器に弱仮説 h_t を求めさせる (S7)。弱学習器は、上述の学習データ S と学習事例の重み ($w_{1,1}, \dots, w_{1,n}$) から弱仮説 h_t を学習する。 h_t は、 t 回目のブースティングラウンドで求めた弱仮説を表している。

【0034】

次に、学習システムは、求めた弱仮説 h_t に対する確信度 β_t を算出する (S9)。

【0035】

更に、学習システムは、次の式に従って、学習事例の重みを更新する (S11)。

【数 4】

$$w_{t+1,i} = w_{t,i} \frac{e^{-y_i \alpha_t h_t(x_i)}}{Z_t(\alpha_t)} \quad (1)$$

【0036】

式中の $h_t(x_i)$ は、学習事例 x_i に対する h_t による分類結果であり、 e は、ネイピア数である。また、

【数 5】

$$Z_t(\alpha_t) = \sum_{i=1}^n w_{t,i} e^{-y_i \alpha_t h_t(x_i)} \quad (2)$$

10

は、

【数 6】

$$\sum_{i=1}^n w_{t+1,i} = 1$$

のための正規化項である。

【0037】

学習システムは、カウンタ t に 1 を加え (S13)、カウンタ t がブースティング繰り返し数 T を越えたか否かを判定する (S15)。カウンタ t がブースティング繰り返し数 T を越えていないと判定した場合には、学習システムは、S7 に戻り、上述した一連の処理を繰り返す。

20

【0038】

カウンタ t がブースティング繰り返し数 T を越えたと判定した場合には、学習システムは、最終仮説 F を求める (S17)。学習システムは、上述のループ処理により求めた T 種類の弱仮説を、以下の式に従って結合することによって、最終仮説 F を求める。

【数 7】

$$F(x) = \text{sign} \left(\sum_{t=1}^T \alpha_t h_t(x) \right)$$

30

【0039】

式中の sign は、入力値が正の場合には 1 を返し、それ以外の場合には -1 を返す関数である。

【0040】

図 8 に、AdaBoost における学習データと学習される規則との例を示す。図 8 の上側のテーブルは、AdaBoost に与えられる学習データの例である。この例で、ラベルが「+1」の場合には、名詞句の一部であることを示し、ラベルが「-1」の場合には、名詞句以外であることを示している。ここでの素性とは、規則の候補である。第 1 レコードは、素性「Yamada」が名詞句の一部であることに重み「1/3」が与えられたことを示している。第 2 レコードは、素性「sings」が名詞句以外であることに重み「1/3」が与えられたことを示している。第 3 レコードは、素性「Taro」が名詞句の一部であることに重み「1/3」が与えられたことを示している。このような学習データから図 8 の下側のテーブルに示すような規則が学習される。ここで、正の値は名詞句であることを意味し、負の値は名詞句でないことを意味している。

40

【0041】

尚、AdaBoost では、 T 種類の弱仮説で構成される最終仮説 F において、以下に示す学習エラーの上限値を持つことが証明されている。

【数 8】

$$\frac{1}{n} \sum_{i=1}^n \left[\left[F(x_i) \neq y_i \right] \right] \leq \prod_{t=1}^T Z_t(\alpha_t) \quad (3)$$

【0042】

[[]] は、ある命題 が成り立つ場合に 1 となり、ある命題 が成り立たない場合に 0 となることを意味している。

【0043】

式(1) と式(3) によって、各ブースティングラウンド t における h_t に対して、以下の式の条件を満たす確信度 α_t を選択すれば、AdaBoost による学習は収束する

10

ことが導かれる。

$Z_t(\alpha_t) < 1$

【0044】

[実施の形態の具体的内容]

本実施の形態では、構造予測モデルを規定するスコアと、構造を評価するスコアとの 2 種類のスコアを用いる。以下、構造予測モデルを規定するスコアを第 1 スコアと呼び、構造を評価するスコアを第 2 スコアと呼び、両者を区別する。

【0045】

上述したブースティング手法では、正例と負例が与えられることを想定しているが、構造学習手法においては明示的に負例が与えられるわけではない。従って、ブースティング

20

手法にそのまま構造学習手法を適用するわけには行かない。

【0046】

そこで、本実施の形態では、学習事例に対する正解構造における第 2 スコアと、正解構造以外の構造群における最高の第 2 スコアとの差を利用する。更に、上述の第 2 スコアの差で定義されるエラーの上限値に基づき、学習事例の重みを更新する。具体的には、正解構造の第 2 スコアが最高の第 2 スコアより低い場合は、不適を意味するので、その学習事例の重みを増加させる。正解構造の第 2 スコアが最高の第 2 スコアより高く、あるいは等しい場合は、好適を意味するので、その学習事例の重みを減少させる。

【0047】

以下で処理の概要を説明する。まず、学習データについて述べる。図 9 に、学習データに含まれる学習事例の例を示す。この例における学習事例は、単語列である。ここでは説明を簡単にするために、第 3 単語までの単語列の例を示しているが、単語数はもっと多くてもよい。

30

【0048】

また、学習事例の数についても、説明を簡単にするために 3 つとしているが、実際に学習を行う場合には、もっと多くの学習事例を用意する。

【0049】

第 1 学習事例： X_1 は、第 1 単語「Yamada」と第 2 単語「sings」とからなる単語列である。第 2 学習事例： X_2 は、第 1 単語「Taro」と第 2 単語「sings」とからなる単語列である。第 3 学習事例： X_3 は、第 1 単語「Taro」と第 2 単語「Yamada」と第 3 単語「sings」とからなる単語列である。

40

【0050】

学習データは、学習事例である単語列に対する正解構造も有している。この例で、図 6 に例示したように、正解構造は、チャンクとそのチャンクに対するラベルの組み合わせの列である。

【0051】

図 10 に示した正解構造の例では、図 4 で説明した「名詞句の先頭」(B - NP)、「名詞句の先頭以外」(I - NP)、「動詞句の先頭」(B - VP) 及び「動詞句の先頭以外」(I - VP) の 4 種類のラベルを用いて、チャンクと品詞句とを特定する。但し、チャンクと品詞句との組み合わせを特定する構造であれば、図 10 以外の形式であってもよ

50

い。

【 0 0 5 2 】

本実施の形態では、各学習事例に重みが付与される。図 1 1 に、初期状態における学習事例の例を示す。初期状態では、重みの合計が 1 であり、また各重みが均等になるようにする。この例では、学習事例の数が 3 であるので、各重みは $1 / 3$ とする。

【 0 0 5 3 】

次に、本実施の形態に係る構造予測モデルについて説明する。構造学習手法として、非特許文献 4 に示した手法あるいは非特許文献 5 に示した手法などを利用してよい。ここでは、非特許文献 4 における構造学習手法に対してブースティングを適用する例について説明する。図 1 2 に、初期状態における構造予測モデルの例を示す。構造予測モデルは、正解構造を予測する各手掛かりに対応する第 1 スコアを有している。この例における手掛かりは、素性に相当する。初期状態において、各第 1 スコアには初期値 0 が設定される。

【 0 0 5 4 】

この例で、第 1 レコードは、チャンク「T a r o」のラベルが名詞句 (N P) であるという手掛かりに対応する第 1 スコアに 0 が設定されていることを示している。第 2 レコードは、チャンク「Y a m a d a」のラベルが名詞句 (N P) であるという手掛かりに対応する第 1 スコアに 0 が設定されていることを示している。第 3 レコードは、チャンク「T a r o」のラベルが動詞句 (V P) であるという手掛かりに対応する第 1 スコアに 0 が設定されていることを示している。第 4 レコードは、チャンク「Y a m a d a」のラベルが動詞句 (V P) であるという手掛かりに対応する第 1 スコアに 0 が設定されていることを示している。第 5 レコードは、チャンク「s i n g s」のラベルが動詞句 (V P) であるという手掛かりに対応する第 1 スコアに 0 が設定されていることを示している。第 6 レコードは、チャンク「Y a m a d a s i n g s」のラベルが名詞句 (N P) であるという手掛かりに対応する第 1 スコアに 0 が設定されていることを示している。第 7 レコードは、チャンク「T a r o Y a m a d a」のラベルが名詞句 (N P) であるという手掛かりに対応する第 1 スコアに 0 が設定されていることを示している。手掛かりを特定する条件は、この例に限定されない。例えば、当該チャンクの前の単語あるいは後の単語に対する条件を含んでいてもよい。

【 0 0 5 5 】

図 1 3 に、構造学習の様子を示す。この図は、第 1 学習事例 X_1 の単語列 1 3 0 1 に対する候補構造 Y 群 1 3 0 3 から、図 1 2 に示した構造予測モデルに従って、最も第 2 スコアの大きい構造 Y^*_1 1 3 0 5 を抽出する様子を示している。図 1 2 に示した構造予測モデルでは、第 1 スコアがすべて 0 であるので、各候補構造 Y における第 2 スコアは等しく 0 となる。その場合には、いずれの候補構造 Y が抽出されてもよいが、この例ではチャンク「Y a m a d a s i n g s」とそのラベルの名詞句 (N P) との組み合わせである構造 Y が選択される。

【 0 0 5 6 】

第 1 学習事例 X_1 の正解構造 Y_1 1 3 0 7 は、第 1 チャンク「Y a m a d a」とそのラベルの名詞句 (N P) との組み合わせと、第 2 チャンク「s i n g s」とそのラベルの動詞句 (V P) との組み合わせの列であるので、ここで抽出された最大スコア構造 Y^*_1 1 3 0 5 は、正解構造 Y_1 1 3 0 7 と一致しないと判定される。

【 0 0 5 7 】

この図で、最大スコア構造 Y^*_1 1 3 0 5 に含まれるチャンクとそのラベルとの組み合わせを太い実線の枠で囲み、正解構造 Y_1 1 3 0 7 に含まれるチャンクとそのラベルとの組み合わせを太い破線の枠で囲んでいる。

【 0 0 5 8 】

図 1 4 に、図 1 3 の判定結果に基づいて、構造予測モデルを更新する様子を示す。本実施の形態では、最大スコア構造が正解構造と一致しないと判定された場合に、最大スコア構造に対応する各手掛かりの第 1 スコアから学習事例の重みだけ減算する。更に、正解構造に対応する各手掛かりの第 1 スコアに学習事例の重みを加算する。

【0059】

このように減算する値と加算する値とは、学習率と呼ばれる。つまり、学習事例の重みを学習率に適用して、最大スコア構造に対応する各手掛かりの第1スコアに対する減算を行うとともに、正解構造に対応する各手掛かりの第1スコアに対する加算を行う。

【0060】

最大スコア構造 $Y_{1,1}^* 1305$ に含まれるチャンク「Yamada sings」とそのラベルである名詞句(NP)との組み合わせは、構造予測モデルの第6レコードの手掛かりに合致するので、第6レコードの第1スコアを減算する。このとき、図11に示した第1学習事例の重み $W_{1,1}$ に相当する「1/3」を引くようにする。

【0061】

更に、正解構造 $Y_{1,1} 1307$ に含まれる第1チャンク「Yamada」とそのラベルの名詞句(NP)との組み合わせは、構造予測モデルの第2レコードの手掛かりに合致するので、第2レコードの第1スコアを加算する。同じく正解構造 $Y_{1,1} 1307$ に含まれる第2チャンク「sings」とそのラベルの動詞句(VP)との組み合わせは、構造予測モデルの第5レコードの手掛かりに合致するので、第5レコードの第1スコアを加算する。このとき、いずれも、図11に示した第1学習事例の重み $W_{1,1}$ に相当する「1/3」を加えるようにする。以上で、第1学習事例に対する1回目の構造学習を終える。

【0062】

続いて、図15に、第2学習事例に対する1回目の構造学習の様子を示す。この図は、第2学習事例 X_2 の単語列1501に対する候補構造Y群1503から、図14に示した構造予測モデルに従って、最も第2スコアの大きい構造 $Y_{2,1}^* 1505$ を抽出する様子を示している。ここでは、チャンク「Taro」とそのラベルの動詞句(VP)との組み合わせと、チャンク「sings」とそのラベルの動詞句(VP)との組み合わせとの列が選択される。

【0063】

第2学習事例 X_2 については、ここで抽出された最大スコア構造 $Y_{2,1}^* 1505$ は、正解構造 $Y_{2,1} 1507$ と一致しないと判定される。

【0064】

この図でも、最大スコア構造 $Y_{2,1}^* 1505$ に含まれるチャンクとそのラベルとの組み合わせを太い実線の枠で囲み、正解構造 $Y_{2,1} 1507$ に含まれるチャンクとそのラベルとの組み合わせを太い破線の枠で囲んでいる。

【0065】

図16に、図15の判定結果に基づいて、構造予測モデルを更新する様子を示す。最大スコア構造 $Y_{2,1}^* 1505$ に含まれるチャンク「Taro」とそのラベルの動詞句(VP)との組み合わせは、構造予測モデルの第3レコードの手掛かりに合致するので、第3レコードの第1スコアを減算する。同じく最大スコア構造 $Y_{2,1}^* 1505$ に含まれるチャンク「sings」とそのラベルの動詞句(VP)との組み合わせは、構造予測モデルの第5レコードの手掛かりに合致するので、第5レコードの第1スコアを減算する。このとき、いずれも、図11に示した第2学習事例の重み $W_{1,2}$ に相当する「1/3」を引くようにする。

【0066】

更に、正解構造 $Y_{2,1} 1507$ に含まれる第1チャンク「Taro」とそのラベルの名詞句(NP)との組み合わせは、構造予測モデルの第1レコードの手掛かりに合致するので、第1レコードの第1スコアを加算する。同じく正解構造 $Y_{2,1} 1507$ に含まれる第2チャンク「sings」とそのラベルの動詞句(VP)との組み合わせは、構造予測モデルの第5レコードの手掛かりに合致するので、第5レコードの第1スコアを加算する。このとき、いずれも、図11に示した第1学習事例の重み $W_{1,2}$ に相当する「1/3」を加えるようにする。第5レコードの第1スコアは、減算と加算とが重なるため、結果的に値が変動しない。以上で、第2学習事例に対する1回目の構造学習を終える。

【0067】

10

20

30

40

50

続いて、図 17 に、第 3 学習事例に対する 1 回目の構造学習の様子を示す。この図は、第 3 学習事例 X_3 の単語列 1 7 0 1 に対する候補構造 Y 群 1 7 0 3 から、図 16 に示した構造予測モデルに従って、最も第 2 スコアの大きい構造 Y^*_3 1 7 0 5 を抽出する様子を示している。ここでは、チャンク「T a r o」とそのラベルの名詞句 (NP) との組み合わせと、チャンク「Y a m a d a」とそのラベルの名詞句 (NP) との組み合わせと、チャンク「s i n g s」とそのラベルの動詞句 (VP) との組み合わせとの列が選択される。

【0068】

第 3 学習事例 X_3 の正解構造 Y_3 1 7 0 7 は、第 1 チャンク「T a r o Y a m a d a」とそのラベルの名詞句 (NP) との組み合わせと、第 2 チャンク「s i n g s」とそのラベルの動詞句 (VP) との組み合わせの列であるので、ここで抽出された最大スコア構造 Y^*_3 1 7 0 5 は、正解構造 Y_3 1 7 0 7 と一致しないと判定される。

10

【0069】

この図でも、最大スコア構造 Y^*_3 1 7 0 5 に含まれるチャンクとそのラベルとの組み合わせを太い実線の枠で囲み、正解構造 Y_3 1 7 0 7 に含まれるチャンクとそのラベルとの組み合わせを太い破線の枠で囲んでいる。

【0070】

図 18 に、図 17 の判定結果に基づいて、構造予測モデルを更新する様子を示す。最大スコア構造 Y^*_3 1 7 0 5 に含まれるチャンク「T a r o」とそのラベルである名詞句 (NP) との組み合わせは、構造予測モデルの第 1 レコードの手掛かりに合致するので、第 1 レコードの第 1 スコアを減算する。同じく最大スコア構造 Y^*_3 1 7 0 5 に含まれるチャンク「Y a m a d a」とそのラベル名詞句 (NP) との組み合わせは、構造予測モデルの第 2 レコードの手掛かりに合致するので、第 2 レコードの第 1 スコアを減算する。同じく最大スコア構造 Y^*_3 1 7 0 5 に含まれるチャンク「s i n g s」とそのラベル動詞句 (VP) との組み合わせは、構造予測モデルの第 5 レコードの手掛かりに合致するので、第 5 レコードの第 1 スコアを減算する。このとき、いずれも、図 11 に示した第 3 学習事例の重み $W_{1,3}$ に相当する「1 / 3」を引くようにする。

20

【0071】

更に、正解構造 Y_3 1 7 0 7 に含まれる第 1 チャンク「T a r o Y a m a d a」とそのラベルの名詞句 (NP) との組み合わせは、構造予測モデルの第 7 レコードの手掛かりに合致するので、第 7 レコードの第 1 スコアを加算する。同じく正解構造 Y_3 1 7 0 7 に含まれる第 2 チャンク「s i n g s」とそのラベルの動詞句 (VP) との組み合わせは、構造予測モデルの第 5 レコードの手掛かりに合致するので、第 5 レコードの第 1 スコアを加算する。このとき、いずれも、図 11 に示した第 1 学習事例の重み $W_{1,3}$ に相当する「1 / 3」を加えるようにする。第 5 レコードの第 1 スコアは、減算と加算とが重なるため、結果的に値が変動しない。以上で、第 3 学習事例に対する 1 回目の構造学習を終える。

30

【0072】

通常は、上述した構造学習を複数回繰り返すが、ここでは説明を簡単にするために、繰り返しは行わずに、構造学習を一旦終了する。

【0073】

このようにして、構造予測モデルが得られると、次に構造予測モデルに対する確信度を算出する。本実施の形態では、正解構造を除く候補構造群に対する第 2 スコアのうち最大スコアと、正解構造に対する第 2 スコアである正解スコアとの差に基づいて、確信度を算出する。尚、第 2 スコアは、その時点における構造予測モデルに基づいて算出される。

40

【0074】

図 19 に、第 1 学習事例に対する 1 回目の第 2 スコア算出の様子を示す。この図は、第 1 学習事例 X_1 の単語列 1 3 0 1 に対する候補構造 Y 群 1 3 0 3 から、図 18 に示した構造予測モデルに従って、最も第 2 スコアの大きい構造 Y^*_1 1 9 0 1 (正解構造 Y_1 1 3 0 7 を除く) を抽出する様子を示している。ここではチャンク「Y a m a d a」とそのラベルの動詞句 (VP) との組み合わせと、チャンク「s i n g s」とそのラベルの動詞句 (VP) との組み合わせとの列が選択される。

50

【0075】

最も第2スコアの大きい構造 $Y^t_1 1901$ に対する第2スコア $s_1(X_1, Y^t_1)$ は、チャンク「Yamada」とそのラベルの動詞句(VP)との組み合わせに対するスコアと、チャンク「sings」とそのラベルの動詞句(VP)との組み合わせに対するスコアとの合計となる。チャンク「Yamada」とそのラベルの動詞句(VP)との組み合わせに対するスコアの算出には、図18に示した構造予測モデルでこの組み合わせに合致する手掛かりを示す第4レコードに設定されている第1スコアを用いる。つまり、0となる。チャンク「sings」とそのラベルの動詞句(VP)との組み合わせに対するスコアの算出には、図18に示した構造予測モデルでこの組み合わせに合致する手掛かりを示す第5レコードに設定されている第1スコアを用いる。つまり、1/3となる。従って、最も第2スコアの大きい構造 $Y^t_1 1901$ に対する最大スコア $s_1(X_1, Y^t_1)$ は、1/3となる。

10

【0076】

更に、正解構造 $Y_1 1307$ の第2スコアである正解スコアも算出する。正解構造 $Y_1 1307$ に対する第2スコア $s_1(X_1, Y_1)$ は、第1チャンク「Yamada」とそのラベルの名詞句(NP)との組み合わせに対するスコアと、第2チャンク「sings」とそのラベルの動詞句(VP)との組み合わせに対するスコアとの合計となる。第1チャンク「Yamada」とそのラベルの名詞句(NP)との組み合わせに対するスコアの算出には、図18に示した構造予測モデルでこの組み合わせに合致する手掛かりを示す第2レコードに設定されている第1スコアを用いる。つまり、0となる。第2チャンク「sings」とそのラベルの動詞句(VP)との組み合わせに対するスコアの算出には、図18に示した構造予測モデルでこの組み合わせに合致する手掛かりを示す第5レコードに設定されている第1スコアを用いる。つまり、1/3となる。従って、正解構造 $Y_1 1307$ の第2スコアである正解スコア $s_1(X_1, Y_1)$ は、1/3となる。

20

【0077】

そのため、第1学習事例 X_1 については、正解スコアと最大スコアとの差は、0となる。

【0078】

図20に、第2学習事例に対する1回目の第2スコア算出の様子を示す。この図は、第2学習事例 X_2 の単語列1501に対する候補構造Y群1503から、図18に示した構造予測モデルに従って、最も第2スコアの大きい構造 $Y^t_2 2001$ (正解構造 $Y_2 1507$ を除く)を抽出する様子を示している。ここではチャンク「Taro」とそのラベルの動詞句(VP)との組み合わせと、チャンク「sings」とそのラベルの動詞句(VP)との組み合わせとの列が選択される。

30

【0079】

最も第2スコアの大きい構造 $Y^t_2 2001$ に対する第2スコア $s_1(X_2, Y^t_2)$ は、チャンク「Taro」とそのラベルの動詞句(VP)との組み合わせに対するスコアと、チャンク「sings」とそのラベルの動詞句(VP)との組み合わせに対するスコアとの合計となる。チャンク「Taro」とそのラベルの動詞句(VP)との組み合わせに対するスコアの算出には、図18に示した構造予測モデルでこの組み合わせに合致する手掛かりを示す第3レコードに設定されている第1スコアを用いる。つまり、-1/3となる。チャンク「sings」とそのラベルの動詞句(VP)との組み合わせに対するスコアの算出には、図18に示した構造予測モデルでこの組み合わせに合致する手掛かりを示す第5レコードに設定されている第1スコアを用いる。つまり、1/3となる。従って、最も第2スコアの大きい構造 $Y^t_2 2001$ に対する最大スコア $s_1(X_2, Y^t_2)$ は、0となる。

40

【0080】

更に、正解構造 $Y_2 1507$ の第2スコアである正解スコアも算出する。正解構造 $Y_2 1507$ に対する第2スコア $s_1(X_2, Y_2)$ は、第1チャンク「Taro」とそのラベルの名詞句(NP)との組み合わせに対するスコアと、第2チャンク「sings」とその

50

ラベルの動詞句 (VP) との組み合わせに対するスコアとの合計となる。第1チャンク「Taro」とそのラベルの名詞句 (NP) との組み合わせに対するスコアの算出には、図18に示した構造予測モデルでこの組み合わせに合致する手掛かりを示す第1レコードに設定されている第1スコアを用いる。つまり、0となる。第2チャンク「sings」とそのラベルの動詞句 (VP) との組み合わせに対するスコアの算出には、図18に示した構造予測モデルでこの組み合わせに合致する手掛かりを示す第5レコードに設定されている第1スコアを用いる。つまり、1/3となる。従って、正解構造 $Y_2 1507$ の第2スコアである正解スコア $s_1(X_2, Y_2)$ は、1/3となる。

【0081】

そのため、第2学習事例 X_2 について、正解スコアと最大スコアとの差は、1/3となる。

10

【0082】

図21に、第3学習事例に対する1回目のスコア算出の様子を示す。この図は、第3学習事例 X_3 の単語列 1701 に対する候補構造 Y 群 1703 から、図18に示した構造予測モデルに従って、最も第2スコアの大きい構造 $Y^t_3 2101$ (正解構造 $Y_3 1707$ を除く) を抽出する様子を示している。ここではチャンク「Taro」とそのラベルの名詞句 (NP) との組み合わせと、チャンク「Yamada」とそのラベルの名詞句 (NP) との組み合わせと、チャンク「sings」とそのラベルの動詞句 (VP) との組み合わせとの列が選択される。

【0083】

20

最も第2スコアの大きい構造 $Y^t_3 2101$ に対する第2スコア $s_1(X_3, Y^t_3)$ は、チャンク「Taro」とそのラベルの名詞句 (NP) との組み合わせに対するスコアと、チャンク「Yamada」とそのラベルの名詞句 (NP) との組み合わせに対するスコアと、チャンク「sings」とそのラベルの動詞句 (VP) との組み合わせに対するスコアとの合計となる。チャンク「Taro」とそのラベルの名詞句 (NP) との組み合わせに対するスコアの算出には、図18に示した構造予測モデルでこの組み合わせに合致する手掛かりを示す第1レコードに設定されている第1スコアを用いる。つまり、0となる。チャンク「Yamada」とそのラベルの名詞句 (NP) との組み合わせに対するスコアの算出には、図18に示した構造予測モデルでこの組み合わせに合致する手掛かりを示す第2レコードに設定されている第1スコアを用いる。つまり、0となる。チャンク「sings」とそのラベルの動詞句 (VP) との組み合わせに対するスコアの算出には、図18に示した構造予測モデルでこの組み合わせに合致する手掛かりを示す第5レコードに設定されている第1スコアを用いる。つまり、1/3となる。従って、最も第2スコアの大きい構造 $Y^t_3 2101$ に対する最大スコア $s_1(X_3, Y^t_3)$ は、1/3となる。

30

【0084】

更に、正解構造 $Y_3 1707$ の第2スコアである正解スコアも算出する。正解構造 $Y_3 1707$ に対する第2スコア $s_1(X_3, Y_3)$ は、第1チャンク「Taro Yamada」とそのラベルの名詞句 (NP) との組み合わせに対するスコアと、第2チャンク「sings」とそのラベルの動詞句 (VP) との組み合わせに対するスコアとの合計となる。第1チャンク「Taro Yamada」とそのラベルの名詞句 (NP) との組み合わせに対するスコアの算出には、図18に示した構造予測モデルでこの組み合わせに合致する手掛かりを示す第7レコードに設定されている第1スコアを用いる。つまり、1/3となる。第2チャンク「sings」とそのラベルの動詞句 (VP) との組み合わせに対するスコアの算出には、図18に示した構造予測モデルでこの組み合わせに合致する手掛かりを示す第5レコードに設定されている第1スコアを用いる。つまり、1/3となる。従って、正解構造 $Y_3 1707$ の第2スコアである正解スコア $s_1(X_3, Y_3)$ は、2/3となる。

40

【0085】

そのため、第3学習事例 X_3 について、正解スコアと最大スコアとの差は、1/3となる。

50

【0086】

これらの差に基づいて、後述する式(4)を満たす確信度を特定する。この例では、後述する式(5)により、

$$\begin{aligned} Z^{-1}(\cdot) & (\cdot \text{は、} Z \text{ の上のハットを意味する。}) \\ & = 1/3 \exp(-\cdot * (0)) + 1/3 \exp(-\cdot * (1/3)) + 1/3 \exp(-\cdot * (1/3)) \\ & = 1/3 + 2/3 \exp(-\cdot / 3) < 1 \end{aligned}$$

となる \hat{z}_1 を見つける。

【0087】

この例では、 $0 < \text{確信度 } \hat{z}_1$ であれば足りるので、確信度 \hat{z}_1 を1とする。従って、

$$Z^{-1}(1) = 0.81102$$

となる。

【0088】

そして、後述する式(6)により、以下の通り第2学習事例の重みを算出する。

$$\begin{aligned} W_{2,1} & = W_{1,1} \exp(-\cdot (0)) / Z^{-1}(1) \\ & = (1/3) / Z^{-1}(1) = 0.411 \end{aligned}$$

$$\begin{aligned} W_{2,2} & = W_{1,2} \exp(-\cdot (1/3)) / Z^{-1}(1) \\ & = (1/3) \exp(-\cdot (1/3)) / Z^{-1}(1) = 0.29449 \end{aligned}$$

$$\begin{aligned} W_{2,3} & = W_{1,3} \exp(-\cdot (1/3)) / Z^{-1}(1) \\ & = (1/3) \exp(-\cdot (1/3)) / Z^{-1}(1) = 0.29449 \end{aligned}$$

【0089】

図22に、更新された学習事例の重みの例を示す。このように、学習事例の重みを更新した上で、更に構造学習を繰り返す。

【0090】

図23に、第1学習事例に対する2回目の構造学習の様子を示す。2回目の構造学習において、構造予測モデルは初期化され、図12の状態に戻る。この図は、第1学習事例 X_1 の単語列1301に対する候補構造 Y 群1303から、図12に示した構造予測モデルに従って、最も第2スコアの大きい構造 $Y^*_{1,2301}$ を抽出する様子を示している。ここでは、チャンク「Yamada sings」とそのラベルの名詞句(NP)との組み合わせが選択される。

【0091】

第1学習事例 X_1 の正解構造 $Y_1 1307$ は、第1チャンク「Yamada」とそのラベルの名詞句(NP)との組み合わせと、第2チャンク「sings」とそのラベルの動詞句(VP)との組み合わせの列であるので、ここで抽出された最大スコア構造 $Y^*_{1,2301}$ は、正解構造 $Y_1 1307$ と一致しないと判定される。

【0092】

この図でも、最大スコア構造 $Y^*_{1,2301}$ に含まれるチャンクとそのラベルとの組み合わせを太い実線の枠で囲み、正解構造 $Y_1 1307$ に含まれるチャンクとそのラベルとの組み合わせを太い破線の枠で囲んでいる。

【0093】

図24に、図23の判定結果に基づいて、構造予測モデルを更新する様子を示す。最大スコア構造 $Y^*_{1,2301}$ に含まれるチャンク「Yamada sings」とそのラベルの名詞句(NP)との組み合わせは、構造予測モデルの第6レコードの手掛かりに合致するので、第6レコードの第1スコアを減算する。このとき、図22に示した第1学習事例の重み $W_{2,1}$ に相当する「0.411」を引くようにする。1回目に比べて、重みが増している。

【0094】

更に、正解構造 $Y_1 1307$ に含まれる第1チャンク「Yamada」とそのラベルの名詞句(NP)との組み合わせは、構造予測モデルの第2レコードの手掛かりに合致するので、第2レコードの第1スコアを加算する。同じく正解構造 $Y_1 1307$ に含まれる第

10

20

30

40

50

2 チャンク「sing s」とそのラベルの動詞句 (V P) との組み合わせは、構造予測モデルの第 5 レコードの手掛かりに合致するので、第 5 レコードの第 1 スコアを加算する。このとき、いずれも、図 2 2 に示した第 1 学習事例の重み $W_{2,1}$ に相当する「0.411」を加えるようにする。1 回目に比べて、重みが増している。以上で、第 1 学習事例に対する 2 回目の構造学習を終える。

【0095】

続いて、図 2 5 に、第 2 学習事例に対する 2 回目の構造学習の様子を示す。この図は、第 2 学習事例 X_2 の単語列 1 5 0 1 に対する候補構造 Y 群 1 5 0 3 から、図 2 4 に示した構造予測モデルに従って、最も第 2 スコアの大きい構造 $Y^*_2 2 5 0 1$ を抽出する様子を示している。ここでは、チャンク「Tar o」とそのラベルの動詞句 (V P) との組み合わせと、チャンク「sing s」とそのラベルの動詞句 (V P) との組み合わせとの列が選択される。

10

【0096】

第 2 学習事例 X_2 の正解構造 $Y_2 1 5 0 7$ は、第 1 チャンク「Tar o」とそのラベルの名詞句 (N P) との組み合わせと、第 2 チャンク「sing s」とそのラベルの動詞句 (V P) との組み合わせの列であるので、ここで抽出された最大スコア構造 $Y^*_2 2 5 0 1$ は、正解構造 $Y_2 1 5 0 7$ と一致しないと判定される。

【0097】

この図でも、最大スコア構造 $Y^*_2 2 5 0 1$ に含まれるチャンクとそのラベルとの組み合わせを太い実線の枠で囲み、正解構造 $Y_2 1 5 0 7$ に含まれるチャンクとそのラベルとの組み合わせを太い破線の枠で囲んでいる。

20

【0098】

図 2 6 に、図 2 5 の判定結果に基づいて、構造予測モデルを更新する様子を示す。最大スコア構造 $Y^*_2 2 5 0 1$ に含まれるチャンク「Tar o」とそのラベルの動詞句 (V P) との組み合わせは、構造予測モデルの第 3 レコードの手掛かりに合致するので、第 3 レコードの第 1 スコアを減算する。同じく最大スコア構造 $Y^*_2 2 5 0 1$ に含まれるチャンク「sing s」とそのラベルの動詞句 (V P) との組み合わせは、構造予測モデルの第 5 レコードの手掛かりに合致するので、第 5 レコードの第 1 スコアを減算する。このとき、いずれも、図 2 2 に示した第 2 学習事例の重み $W_{2,2}$ に相当する「0.29449」を引くようにする。1 回目に比べて、重みが減っている。

30

【0099】

更に、正解構造 $Y_2 1 5 0 7$ に含まれる第 1 チャンク「Tar o」とそのラベルの名詞句 (N P) との組み合わせは、構造予測モデルの第 1 レコードの手掛かりに合致するので、第 1 レコードの第 1 スコアを加算する。同じく正解構造 $Y_2 1 5 0 7$ に含まれる第 2 チャンク「sing s」とそのラベルの動詞句 (V P) との組み合わせは、構造予測モデルの第 5 レコードの手掛かりに合致するので、第 5 レコードの第 1 スコアを加算する。このとき、いずれも、図 2 2 に示した第 2 学習事例の重み $W_{2,2}$ に相当する「0.29449」を加えるようにする。1 回目に比べて、重みが減っている。第 5 レコードの第 1 スコアは、減算と加算とが重なるため、結果的に値が変動しない。以上で、第 2 学習事例に対する 2 回目の構造学習を終える。

40

【0100】

続いて、図 2 7 に、第 3 学習事例に対する 2 回目の構造学習の様子を示す。この図は、第 3 学習事例 X_3 の単語列 1 7 0 1 に対する候補構造 Y 群 1 7 0 3 から、図 2 6 に示した構造予測モデルに従って、最も第 2 スコアの大きい構造 $Y^*_3 2 7 0 1$ を抽出する様子を示している。ここでは、チャンク「Tar o」とそのラベルの名詞句 (N P) との組み合わせと、チャンク「Yamada」とそのラベルの名詞句 (N P) との組み合わせと、チャンク「sing s」とそのラベルの動詞句 (V P) との組み合わせとの列が選択される。

【0101】

第 3 学習事例 X_3 の正解構造 $Y_3 1 7 0 7$ は、第 1 チャンク「Tar o Yamada」とそのラベルの名詞句 (N P) との組み合わせと、第 2 チャンク「sing s」とそのラ

50

ベルの動詞句 (VP) との組み合わせの列であるので、ここで抽出された最大スコア構造 $Y^*_3 2701$ は、正解構造 $Y_3 1707$ と一致しないと判定される。

【0102】

この図でも、最大スコア構造 $Y^*_3 2701$ に含まれるチャンクとそのラベルとの組み合わせを太い実線の枠で囲み、正解構造 $Y_3 1707$ に含まれるチャンクとそのラベルとの組み合わせを太い破線の枠で囲んでいる。

【0103】

図28に、図27の判定結果に基づいて、構造予測モデルを更新する様子を示す。最大スコア構造 $Y^*_3 2701$ に含まれるチャンク「Taro」とそのラベルの名詞句 (NP) との組み合わせは、構造予測モデルの第1レコードの手掛かりに合致するので、第1レコードの第1スコアを減算する。同じく最大スコア構造 $Y^*_3 2701$ に含まれるチャンク「Yamada」とそのラベルの名詞句 (NP) との組み合わせは、構造予測モデルの第2レコードの手掛かりに合致するので、第2レコードの第1スコアを減算する。同じく最大スコア構造 $Y^*_3 2701$ に含まれるチャンク「sings」とそのラベルの動詞句 (VP) との組み合わせは、構造予測モデルの第5レコードの手掛かりに合致するので、第5レコードの第1スコアを減算する。このとき、いずれも、図22に示した第3学習事例の重み $W_{2,3}$ に相当する「0.29449」を引くようにする。1回目に比べて、重みが減っている。

【0104】

更に、正解構造 $Y_3 1707$ に含まれる第1チャンク「Taro Yamada」とそのラベルの名詞句 (NP) との組み合わせは、構造予測モデルの第7レコードの手掛かりに合致するので、第7レコードの第1スコアを加算する。同じく正解構造 $Y_3 1707$ に含まれる第2チャンク「sings」とそのラベルの動詞句 (VP) との組み合わせは、構造予測モデルの第5レコードの手掛かりに合致するので、第5レコードの第1スコアを加算する。このとき、いずれも、図22に示した第3学習事例の重み $W_{2,3}$ に相当する「0.29449」を加えるようにする。1回目に比べて、重みが減っている。第5レコードの第1スコアは、減算と加算とが重なるため、結果的に値が変動しない。以上で、第3学習事例に対する2回目の構造学習を終える。

【0105】

そして、2回目の構造予測モデルについても確信度を算出する。図28に示した2回目の構造予測モデルは、図18に示した1回目の構造予測モデルよりも構造を予測する精度が向上している。

【0106】

例えば、チャンク「Taro」とそのラベルの名詞句 (NP) との組み合わせに合致する手掛りに対応する第1レコードの第1スコア「0」は、チャンク「Taro」とそのラベルの動詞句 (VP) との組み合わせに合致する手掛りに対応する第3レコードの第1スコア「-0.29449」よりも大きい。従って、チャンク「Taro」について、正しいラベルである名詞句 (NP) を導くようになっている。

【0107】

例えば、チャンク「Yamada」とそのラベルの名詞句 (NP) との組み合わせに合致する手掛りに対応する第2レコードの第1スコア「0.11651」は、チャンク「Yamada」とそのラベルの動詞句 (VP) との組み合わせに合致する手掛りに対応する第4レコードの第1スコア「0」よりも大きい。従って、チャンク「Yamada」についても、正しいラベルである名詞句 (NP) を導くようになっている。

【0108】

例えば、チャンク「Taro Yamada」とそのラベルの名詞句 (NP) との組み合わせに合致する手掛りに対応する第7レコードの第1スコア「0.29449」は、チャンク「Taro」とそのラベルの名詞句 (NP) との組み合わせに合致する手掛りに対応する第1レコードの第1スコア「0」と、チャンク「Yamada」とそのラベルの名詞句 (NP) との組み合わせとの列に合致する手掛りに対応する第2レコードの第1スコ

10

20

30

40

50

ア「0.11651」との合計「0.11651」よりも大きい。従って、チャンク「Taro Yamada」とそのラベルの名詞句（NP）との組み合わせを正しく導くようになっている。

【0109】

尚、後述するように、それぞれの構造予測モデルに基づいて、結合モデルを求める。以上で、処理の概要についての説明を終える。

【0110】

以下、具体的なシステムの構成及び動作について説明する。図29に、モデル学習装置の機能ブロックを示す。本システムは、学習データなどの入力を行うための学習データ入力部101と、学習データ入力部101によって入力された学習データなどを格納する学習データ格納部103と、学習データ格納部103に格納されているデータを用いて本実施の形態における主要な処理を実施するモデル学習部105とを有する。

10

【0111】

更に、本システムは、モデル学習部105で算出する第2スコアの差データを格納する差データ格納部106と、モデル学習部105で算出する重みデータを格納する重みデータ格納部107と、モデル学習部105で算出する確信度データを格納する確信度格納部108と、モデル学習部105の処理結果であるモデルデータ（例えば、構造予測モデル及び統合モデル）を格納するモデルデータ格納部109とを有する。

【0112】

更に、本システムは、生成されたモデルを適用して分類を行う対象である分類データを入力するための分類データ入力部111と、分類データ入力部111によって入力された分類データを格納する分類データ格納部113と、モデルデータ格納部109に格納されているモデルデータを、分類データ格納部113に格納されている処理対象データに適用して、分類を行うモデル適用部115と、モデル適用部115の処理結果を格納する処理結果格納部117とを有する。

20

【0113】

図30に、モデル学習部105の構成を示す。モデル学習部105は、ブースティング部201と構造学習部203とを有する。ブースティング部201は、ブースティング処理を行う。構造学習部203は、構造学習処理を行う。

【0114】

ブースティング部201は、初期化部205、確信度算出部207、結合モデル更新部209及び重み更新部211を有する。初期化部205は、データの初期化を行う。確信度算出部207は、構造予測モデルに対する確信度を算出する。結合モデル更新部209は、結合モデルを更新する。重み更新部211は、学習事例の重みを更新する。

30

【0115】

学習データ格納部103、差データ格納部106、重みデータ格納部107、確信度格納部108、モデルデータ格納部109、分類データ格納部113及び処理結果格納部117は、例えば後述する図36に示されたメモリ2501、HDD（Hard Disk Drive）2505あるいはドライブ装置2513を用いる。

【0116】

学習データ入力部101、モデル学習部105、分類データ入力部111、モデル適用部115、ブースティング部201、構造学習部203、初期化部205、確信度算出部207、結合モデル更新部209及び重み更新部211は、その処理の一部又は全部を、例えば後述する図36に示されたメモリ2501に格納されたプログラムに含まれる命令をCPU（Central Processing Unit）2503で読み込み、実行することによって実現されるようにしてもよい。

40

【0117】

次に、図31乃至図35を用いて、本実施の形態におけるモデル学習部105等の処理内容について説明する。

【0118】

50

図 3 1 に、本発明のメイン処理フローを示す。まず、学習データ入力部 1 0 1 は、例えばユーザからの指示に従って、学習データ $S = \{ (X_1, Y_1), \dots, (X_n, Y_n) \}$ と、ブースティング繰り返し数 T についての入力を受け付ける (S 1 0 1)。

【0 1 1 9】

学習データ入力部 1 0 1 は、例えば、図 9 に示した学習事例と、図 1 0 に示した正解構造とを受け付ける。ブースティング繰り返し数 T は、ブースティングにおける繰り返しの回数である。

【0 1 2 0】

モデル学習部 1 0 5 は、モデルデータ格納部 1 0 9 に初期の結合モデル M_0 を設定する (S 1 0 3)。具体的には、ブースティング部 2 0 1 の初期化部 2 0 5 がこの処理を行う。図 3 2 に、初期状態における結合モデルの例を示す。初期化部 2 0 5 は、各手掛かりに対する第 1 スコアに 0 を設定する。

【0 1 2 1】

モデル学習部 1 0 5 は、重みデータ格納部 1 0 7 に格納されている学習事例の重み ($w_{1,1}, \dots, w_{1,n}$) を初期化する (S 1 0 5)。具体的には、ブースティング部 2 0 1 の初期化部 2 0 5 がこの処理を行う。各重みに、等しく $1/n$ の値が設定される。

【0 1 2 2】

モデル学習部 1 0 5 のブースティング部 2 0 1 は、パラメータ t を 1 に初期化する (S 1 0 7)。

【0 1 2 3】

モデル学習部 1 0 5 は、構造学習処理を実施する (S 1 0 9)。具体的には、構造学習部 2 0 3 がこの処理を行う。

【0 1 2 4】

図 3 3 に、構造学習処理フローを示す。構造学習部 2 0 3 は、以下の式に示すように、学習事例の重み ($w_{t,1}, \dots, w_{t,n}$) を学習事例の学習率 ($\alpha_1, \dots, \alpha_n$) にセットする (S 2 0 1)。学習事例の学習率 ($\alpha_1, \dots, \alpha_n$) は、構造学習処理における内部パラメータである。

【数 9】

$$\{\varepsilon_i\}_{i=1}^n = \{w_{t,i}\}_{i=1}^n$$

ここでは、非特許文献 4 に示した手法を利用する場合の例を示しているが、非特許文献 5 に示した手法を利用する場合にも、同様に学習事例の重みを学習事例の学習率にセットするようにしてもよい。

【0 1 2 5】

構造学習部 2 0 3 は、モデルデータ格納部 1 0 9 に格納されている構造予測モデル m を初期化する (S 2 0 3)。このとき、構造学習部 2 0 3 は、モデルデータ格納部 1 0 9 に格納されている構造予測モデル m に含まれる各スコアを 0 にする。これにより、構造予測モデル m は、図 1 2 に示した状態となる。

【0 1 2 6】

構造学習部 2 0 3 は、パラメータ p を 1 に初期化する (S 2 0 5)。

【0 1 2 7】

構造学習部 2 0 3 は、パラメータ i を 1 に初期化する (S 2 0 7)。

【0 1 2 8】

構造学習部 2 0 3 は、以下の式に従って、学習事例 X_i についての候補構造 Y 群のうち、最大スコアを得る構造 Y_i^* を特定する (S 2 0 9)。以下で説明する構造は、例えば前述したチャンクとラベルとの組み合わせの列に相当する。

【数 1 0】

$$Y_i^* = \underset{Y \in \mathcal{Y}(X_i)}{\operatorname{argmax}} m \cdot \Phi(X, Y)$$

10

20

30

40

50

【 0 1 2 9 】

(X , Y) は、学習事例 X の単語列 と構造 Y とを、 D 次元ベクトルへマッピングする関数であり、以下の式のように定義される。

【 数 1 1 】

$$\Phi(X, Y) = \sum_{d=1}^D \sum_{j=1}^{|Y|} \phi_d(X, Y_{(j)})$$

【 0 1 3 0 】

D は、構造予測モデル m における手掛かりの数に相当する。 | Y | は、構造 Y に含まれるチャンクの数である。 Y (j) は、構造 Y に含まれる j 番目のチャンクとそのラベル l (Y (j)) との組み合わせである。

10

【 0 1 3 1 】

また、 $\phi_d(X, Y_{(j)})$ は、学習事例である単語列 X と組み合わせ Y (j) を受け取り、 D 次元のベクトルを返す関数である。単語列 X と組み合わせ Y (j) とが、 d 番目の手掛かりに相当する素性を満たす場合には、 d 次元目の値に 1 が設定され、他の値に 0 が設定されたベクトルが返される。単語列 X と組み合わせ Y (j) とが、 d 番目の手掛かりに相当する素性を満たさない場合には、 d 次元目の値に 0 が設定され、更に他の値にも 0 が設定されたベクトルが返される。

【 0 1 3 2 】

20

素性には、例えば「 Y (j) のチャンクが『 M r . 』であり、且つそのラベルが名詞句である」、「 Y (j) のチャンクにおける先頭の単語の前に出現する単語が『 M r . 』であり、且つそのチャンクのラベル l (Y (j)) が動詞句である」、あるいは「 Y (j) のチャンクにおける最後の単語の後に出現する単語が『 i n 』であり、且つそのチャンクのラベル l (Y (j)) が名詞句である」のようなものを用いてもよい。

【 0 1 3 3 】

繰り返しによって、構造予測モデル m は随時更新されるが、構造学習部 2 0 3 は、 S 2 0 9 の処理時点における構造予測モデル m を用いる。 $w_d(X, Y)$ は、学習事例 X と構造 Y の組み合わせに与えられるスコアである。

【 0 1 3 4 】

30

最大スコア構造 Y_i^* が、正解構造 Y_i と異なる場合には、以下の式に従って、構造予測モデル m は更新される。

【 数 1 2 】

$$m = m + \varepsilon_i \left(\Phi(X_i, Y_i) - \Phi(X_i, Y_i^*) \right)$$

【 0 1 3 5 】

そのため、構造学習部 2 0 3 は、最大スコア構造 Y_i^* は正解構造 Y_i と一致するか否かを判定する。(S 2 1 1)。最大スコア構造 Y_i^* は正解構造 Y_i と一致すると判定した場合には、そのまま S 2 1 5 の処理へ移る。

40

【 0 1 3 6 】

他方、最大スコア構造 Y_i^* は正解構造 Y_i と一致しないと判定した場合には、構造学習部 2 0 3 は、構造予測モデル更新処理を実施する (S 2 1 3)。

【 0 1 3 7 】

図 3 4 に、構造予測モデル更新処理フローを示す。構造学習部 2 0 3 は、最大スコア構造 Y_i^* に対応する各手掛かりの第 1 スコアを、学習率 η_i に従って減算する (S 3 0 1)。構造学習部 2 0 3 は、図 1 4、図 1 6、図 1 8、図 2 4、図 2 6 及び図 2 8 における太い実線枠の例のように、最大スコア構造 Y_i^* に含まれる各チャンクとそのラベルの組み合わせと合致する各手掛かりを特定し、その手掛かりの第 1 スコアに学習率 η_i の値を減算する。

50

【 0 1 3 8 】

更に、構造学習部 2 0 3 は、正解構造 Y_i に対応する各手掛かりの第 1 スコアを、学習率 α_i に従って加算する (S 3 0 3)。構造学習部 2 0 3 は、図 1 4、図 1 6、図 1 8、図 2 4、図 2 6 及び図 2 8 における太い破線枠の例のように、正解構造 Y_i に含まれる各チャンクとそのラベルの組み合わせと合致する各手掛かりを特定し、その手掛かりの第 1 スコアに学習率 α_i の値を加算する。そして、図 3 3 の S 2 1 5 の処理に戻る。

【 0 1 3 9 】

図 3 3 の処理に戻って、構造学習部 2 0 3 は、パラメータ i に 1 を加え (S 2 1 5)、パラメータ i が学習事例の数 n を越えたか否かを判定する (S 2 1 7)。パラメータ i が学習事例の数 n を越えていないと判定した場合には、S 2 0 9 に戻って一連の処理を繰り返す。パラメータ i が学習事例の数 n を越えたと判定した場合には、S 2 1 9 に移って、構造学習部 2 0 3 は、パラメータ p に 1 を加え (S 2 1 9)、パラメータ p が構造学習の繰り返し数 P を越えたか否かを判定する (S 2 1 7)。パラメータ p が構造学習の繰り返し数 P を越えていないと判定した場合には、S 2 0 7 に戻って一連の処理を繰り返す。パラメータ p が構造学習の繰り返し数 P を越えたと判定した場合には、構造学習処理を終え、図 3 1 の S 1 1 1 に戻る。

【 0 1 4 0 】

図 3 1 の説明に戻る。モデル学習部 1 0 5 は、確信度算出処理を実施する (S 1 1 1)。具体的には、ブースティング部 2 0 1 の確信度算出部 2 0 7 がこの処理を行う。

【 0 1 4 1 】

図 3 5 に、確信度算出処理フローを示す。確信度算出部 2 0 7 は、パラメータ i に 1 を設定する (S 4 0 1)。パラメータ i は、S 4 0 3 から S 4 0 7 までのルーチンを計数し、学習事例を特定する。

【 0 1 4 2 】

確信度算出部 2 0 7 は、候補構造 Y 群のうち、正解構造 Y_i を除き最大スコアを得る構造 Y^t_i を特定する (S 4 0 3)。最大スコアを得る構造 Y^t_i を特定するための式を、以下に示す。

【 数 1 3 】

$$Y_i^t = \underset{Y \in \mathcal{Y}(X_i) \wedge Y \neq Y_i}{\operatorname{argmax}} s_t(X_i, Y)$$

【 0 1 4 3 】

上記の $s_t(X_i, Y)$ は、以下の式に示すように、 t 番目のブースティングラウンドにおける学習事例 X_i に対する構造 Y のスコアを意味する。

【 数 1 4 】

$$s_t(X, Y) = m_t \cdot \Phi(X, Y)$$

【 0 1 4 4 】

m_t は、 t 番目のブースティングラウンドにおける構造予測モデルである。

【 0 1 4 5 】

上述の通り正解構造 Y_i は除かれるので、候補構造 Y 群のスコアうち、正解構造 Y_i のスコアが最も大きい場合には、2 番目に大きいスコアを得た構造を Y^t_i とする。このとき、例えば forward-DP backward-A* N-best search algorithm (非特許文献 2) の技術を用いるようにしてもよい。

【 0 1 4 6 】

確信度算出部 2 0 7 は、以下の式に従って、正解構造 Y_i に対する第 2 スコアから最大スコア構造 Y^t_i に対する第 2 スコアを引いて差 $d_i(X_i)$ を求める (S 4 0 5)。確信度算出部 2 0 7 は、求めた差を差データ格納部 1 0 6 に格納する。

【数 1 5】

$$d_t(X_i) = s_t(X_i, Y_i) - s_t(X_i, Y_i^t)$$

【0 1 4 7】

この差を用いることによって、誤ってチャンキングされた学習事例、あるいは誤ってラベル付けされた学習事例に対する重みが高くなるように調整される。

【0 1 4 8】

確信度算出部 2 0 7 は、パラメータ i に 1 を加え (S 4 0 7)、パラメータ i が学習事例の数 n を越えたか否かを判定する (S 4 0 9)。パラメータ i が学習事例の数 n を越えていないと判定した場合には、S 4 0 3 へ戻り一連の処理を繰り返す。

10

【0 1 4 9】

パラメータ i が学習事例の数 n を越えたと判定した場合には、確信度算出部 2 0 7 は、学習モデル M_t に対する確信度 α_t を算出する。具体的には、確信度算出部 2 0 7 は、差データ格納部 1 0 6 に格納されている各学習事例における差に基づき、以下の式で示される収束条件を満たす確信度 α_t を特定する (S 4 1 1)。

【数 1 6】

$$\tilde{Z}_t(\alpha_t) < 1 \quad (4)$$

【0 1 5 0】

このとき、

20

【数 1 7】

$$\tilde{Z}_t(\alpha_t) = \sum_{i=1}^n w_{t,i} e^{-\alpha_t d_t(X_i)} \quad (5)$$

である。

【0 1 5 1】

この式によれば、構造学習処理 (S 1 0 9) によって各学習事例に与えられる実数値であるスコアを指数として扱うことになる。従って、収束条件を満たす確信度 α_t を解析的に計算することは、難しい場合もある。

【0 1 5 2】

30

確信度 α_t を計算する方法として、Bisection Search を利用するようにしてもよい。 $\{-1, +1\}$ の二値を予測する弱仮説 h_t を用いた場合の確信度の計算 (AdaBoost: 非特許文献 1) に合わせるために、本実施の形態では、弱仮説 h_t を $\text{sign}(d_t(X_i))$ と定義する。そして、式 (2) における $Z_t(\alpha_t)$ を最小化するときの値 α_t を求め、それを確信度とする。そのため、以下の式に示すように、 Z_t に対する α_t による微分を行う。

【数 1 8】

$$Z'_t(\alpha_t) = \sum_{i=1}^n -h_t(X_i) w_{t,i} e^{-\alpha_t h_t(X_i)}$$

40

【0 1 5 3】

そして、上記式が 0 となる条件を解くことで、以下の式に示すように確信度を求める。

【数 1 9】

$$\tilde{\alpha}_t = \frac{1}{2} \log \left(\frac{\sum_{i=1}^n w_{t,i} \left[\left[h_t(X_i) = 1 \right] \right]}{\sum_{i=1}^n w_{t,i} \left[\left[h_t(X_i) = -1 \right] \right]} \right)$$

50

【 0 1 5 4 】

最後に、以下の範囲で、式 (5) を最小化する値を *B i s e c t i o n S e a r c h* によって選択し、それを確信度とする。

【 数 2 0 】

$$0 \sim 2 \times \tilde{\alpha}_t$$

【 0 1 5 5 】

このように探索の範囲を広くすることによって、より正確な確信度が得られると期待される。確信度算出部 2 0 7 は、算出した確信度を確信度格納部 1 0 8 に格納する。以上で、確信度算出処理についての説明を終える。

10

【 0 1 5 6 】

図 3 1 の説明に戻る。モデル学習部 1 0 5 は、モデルデータ格納部 1 0 9 に格納されている構造予測モデルに基づき、モデルデータ格納部 1 0 9 に格納されている結合モデルを更新する (S 1 1 3)。具体的には、ブースティング部 2 0 1 の結合モデル更新部 2 0 9 が、以下の式に従って、前回のブースティングラウンド $t - 1$ における結合モデル M_{t-1} 、今回のブースティングラウンド t で求めた構造予測モデル m_t 及び確信度 α_t に基づいて、今回のブースティングラウンド t における結合モデル M_t を算出する。このとき、結合モデル更新部 2 0 9 は、確信度格納部 1 0 8 から確信度 α_t を読み取る。

【 数 2 1 】

$$M_t = M_{t-1} + \alpha_t m_t$$

20

【 0 1 5 7 】

続いて、モデル学習部 1 0 5 は、重みデータ格納部 1 0 7 に格納されている学習事例の重みを更新する (S 1 1 5)。具体的には、ブースティング部 2 0 1 の重み更新部 2 1 1 が、以下の式に従って、今回のブースティングラウンド t における学習事例の重み ($w_{t,1}, \dots, w_{t,n}$) から、次回のブースティングラウンド $t+1$ における学習事例の重み ($w_{t+1,1}, \dots, w_{t+1,n}$) を求める。

【 数 2 2 】

$$w_{t+1,i} = \frac{w_{t,i} e^{-\alpha_t d_t(X_i)}}{\tilde{Z}_t(\alpha_t)} \quad (6)$$

30

【 0 1 5 8 】

正解構造 Y_i のスコア $s_t(X_i, Y_i)$ が、最大スコア $s_t(X_i, Y^{t,i})$ より大きい場合、つまり $0 < d_t(X_i)$ の場合には、正しく解析されたことを意味する。従って、次回に用いる学習事例の重みを減少させる。

【 0 1 5 9 】

他方、正解構造 Y_i のスコア $s_t(X_i, Y_i)$ が、最大スコア $s_t(X_i, Y^{t,i})$ より小さい場合、つまり $0 > d_t(X_i)$ の場合には、正しく解析されなかったことを意味する。従って、次回に用いる学習事例の重みを増加させる。

40

【 0 1 6 0 】

モデル学習部 1 0 5 のブースティング部 2 0 1 は、パラメータ t を 1 インクリメントする (S 1 1 7)。

【 0 1 6 1 】

続いて、モデル学習部 1 0 5 のブースティング部 2 0 1 は、パラメータ t がブースティング繰り返し数 T を越えたか否かを判定する (S 1 1 9)。パラメータ t がブースティング繰り返し数 T を越えていないと判定した場合には、S 1 0 9 に戻って一連の処理を繰り返す。パラメータ t がブースティング繰り返し数 T を越えたと判定した場合には、図 3 1 の処理を終了する。

【 0 1 6 2 】

50

最後に、モデル適用部 115 におけるモデル適用処理について説明する。モデル適用部 115 は、従来の構造学習におけるマッピングを実施する。モデル適用部 115 の処理内容は、モデルデータ格納部 109 に格納されるモデルデータ（この例では、結合データ）と、分類データ入力部 111 によって入力され分類データ格納部 113 に格納されている分類データ（この例では、単語列）とを用いて、従来通りに実行される。この例では、以下の式に従って、分類データの単語列 X における候補構造 Y 群のうち、最終の結合モデル M_T に基づいて算出したスコアが最大となる構造 Y^* を特定する。

【数 2 3】

$$Y^* = \underset{Y \in \mathcal{Y}(X)}{\operatorname{argmax}} M_T \cdot \Phi(X, Y)$$

10

【0163】

そして、処理結果（この例では、構造 Y^* ）が、処理結果格納部 117 に格納される。

【0164】

[付録]

本実施の形態に係る学習方法は、各ブースティングラウンド t で、 $0 < \alpha_t$ を満たし、且つ式（4）を満たす確信度 α_t を計算できれば、本発明の前提と同様に収束することを示す。

【0165】

本実施の形態に係る学習方法では、以下に示す学習エラーの上限値が得られることを証明する。 20

【数 2 4】

$$\frac{1}{n} \sum_{i=1}^n \left[\left[Y_i^* \neq Y_i \right] \right] \leq \prod_{t=1}^T \tilde{Z}_t(\alpha_t)$$

【0166】

尚、

【数 2 5】

$$Y_i^* = \underset{Y \in \mathcal{Y}(X_i)}{\operatorname{argmax}} M_T \cdot \Phi(X_i, Y)$$

30

である。

【0167】

まず、式（6）を展開することにより、次の式が得られる。

【数 2 6】

$$\begin{aligned}
 w_{T+1,i} &= \frac{w_{T,i} e^{-\alpha_T d_T(X_i)}}{\tilde{Z}_T(\alpha_T)} \\
 &= \frac{e^{-\sum_{t=1}^T \alpha_t d_t(X_i)}}{n \prod_{t=1}^T \tilde{Z}_t(\alpha_t)} \\
 &= \frac{e^{-\sum_{t=1}^T \alpha_t m_t(\Phi(X_i, Y_i) - \Phi(X_i, Y_i^t))}}{n \prod_{t=1}^T \tilde{Z}_t(\alpha_t)}
 \end{aligned}$$

10

【0 1 6 8】

従って、 $Y_i^* = Y_i$ の場合には、

【数 2 7】

$$\frac{e^{-\sum_{t=1}^T \alpha_t m_t(\Phi(X_i, Y_i) - \Phi(X_i, Y_i^*))}}{n \prod_{t=1}^T \tilde{Z}_t(\alpha_t)} \leq w_{T+1,i}$$

20

となる。

【0 1 6 9】

これは、1 から T までの範囲の t に対して、以下の式を満たすからである。

【数 2 8】

$$m_t \Phi(X_i, Y_i^*) \leq m_t \Phi(X_i, Y_i^t)$$

30

【0 1 7 0】

更に、 $Y_i^* = Y_i$ の場合には、次の式が満たされる。

【数 2 9】

$$\begin{aligned}
 1 &\leq e^{-\sum_{t=1}^T \alpha_t m_t(\Phi(X_i, Y_i) - \Phi(X_i, Y_i^*))} \\
 &\leq e^{-\sum_{t=1}^T \alpha_t m_t(\Phi(X_i, Y_i) - \Phi(X_i, Y_i^t))} \\
 &= e^{-\sum_{t=1}^T \alpha_t d_t(X_i)}
 \end{aligned}$$

40

【0 1 7 1】

よって、

【数 3 0】

$$\left[\left[Y_i^* \neq Y_i \right] \right] \leq e^{-\sum_{t=1}^T \alpha_t d_t(X_i)}$$

となる。

【 0 1 7 2】

これらの結果から、以下に示すように、上述した学習エラーの上限値が得られる。

【数 3 1】

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \left[\left[Y_i^* \neq Y_i \right] \right] &\leq \frac{\sum_{i=1}^n e^{-\sum_{t=1}^T \alpha_t d_t(X_i)}}{n} \\ &= \sum_{i=1}^n \left(\prod_{t=1}^T \tilde{Z}_t(\alpha_t) \right) w_{T+1,i} \\ &= \prod_{t=1}^T \tilde{Z}_t(\alpha_t) \end{aligned}$$

10

20

【 0 1 7 3】

以上のように、従来のブースティング手法と同様、本実施の形態も収束性を備える。

【 0 1 7 4】

以上本発明の一実施の形態を説明したが、本発明はこれに限定されるものではない。例えば、上述の機能ブロック構成は実際のプログラムモジュール構成に一致しない場合もある。

【 0 1 7 5】

また、上で説明した各記憶領域の構成は一例であって、上記のような構成でなければならないわけではない。さらに、処理フローにおいても、処理結果が変わらなければ処理の順番を入れ替えることも可能である。さらに、並列に実行させるようにしても良い。

30

【 0 1 7 6】

なお、上で述べたモデル学習装置は、コンピュータ装置であって、図 3 6 に示すように、メモリ 2 5 0 1 と C P U (Central Processing Unit) 2 5 0 3 とハードディスク・ドライブ (H D D : Hard Disk Drive) 2 5 0 5 と表示装置 2 5 0 9 に接続される表示制御部 2 5 0 7 とリムーバブル・ディスク 2 5 1 1 用のドライブ装置 2 5 1 3 と入力装置 2 5 1 5 とネットワークに接続するための通信制御部 2 5 1 7 とがバス 2 5 1 9 で接続されている。オペレーティング・システム (O S : Operating System) 及び本実施例における処理を実施するためのアプリケーション・プログラムは、H D D 2 5 0 5 に格納されており、C P U 2 5 0 3 により実行される際には H D D 2 5 0 5 からメモリ 2 5 0 1 に読み出される。C P U 2 5 0 3 は、アプリケーション・プログラムの処理内容に応じて表示制御部 2 5 0 7、通信制御部 2 5 1 7、ドライブ装置 2 5 1 3 を制御して、所定の動作を行わせる。また、処理途中のデータについては、主としてメモリ 2 5 0 1 に格納されるが、H D D 2 5 0 5 に格納されるようにしてもよい。本発明の実施例では、上で述べた処理を実施するためのアプリケーション・プログラムはコンピュータ読み取り可能なリムーバブル・ディスク 2 5 1 1 に格納されて頒布され、ドライブ装置 2 5 1 3 から H D D 2 5 0 5 にインストールされる。インターネットなどのネットワーク及び通信制御部 2 5 1 7 を経由して、H D D 2 5 0 5 にインストールされる場合もある。このようなコンピュータ装置は、上で述べた C P U 2 5 0 3、メモリ 2 5 0 1 などのハードウェアと O S 及びアプリケーション・プログラムなどのプログラムとが有機的に協働することにより、上で述べたような

40

50

各種機能を実現する。

【0177】

以上述べた本発明の実施の形態をまとめると、以下のようになる。

【0178】

本実施の形態に係るモデル学習方法は、(A)各学習事例の重みに基づいて、重みに対応する各学習事例から夫々の正解構造を抽出する際の手掛かりとなる素性の各々に対応する第1スコアを更新することによって、当該第1スコアの各々により規定されるモデルを算出する第1ステップと、(B)各学習事例について、当該学習事例から導かれる複数の候補構造のうち正解構造以外の候補構造の各々に対して、モデルにより付与される第2スコアのうち最大スコアを特定し、当該学習事例の正解構造に対してモデルにより付与される第2スコアと最大スコアとの差を求めると共に、各学習事例における当該差により定義されるエラーの上限値に基づいて、モデルの確信度を算出する第2ステップと、(C)確信度と各学習事例における上記差とに基づいて、各学習事例の重みを更新する第3ステップと、(D)第1ステップと、第2ステップと、第3ステップとを繰り返すステップとを含む。

10

【0179】

このようにすれば、構造学習で得られたモデルの確信度を正解構造の第2スコアと他の候補構造の最大スコアとの差で表すことによって、次の構造学習に用いる学習事例の重みを更新できるようになる。更に、エラーの上限値に基づいて構造学習におけるトレーニングエラーを収束させることができる。例えば、学習事例の重みに基づくブースティング手法を構造学習に適用して、構造学習の精度を高めることができる。

20

【0180】

更に、上記第1ステップにおいて、重みデータ格納部に格納されている各学習事例の重みの各々を、学習事例の学習率に設定して、上記第1スコアを更新するようにしてもよい。

【0181】

このようにすれば、学習事例の重みを直接的に構造学習によるモデルに反映できる。

【0182】

なお、上記方法による処理をコンピュータに行わせるためのプログラムを作成することができ、当該プログラムは、例えばフレキシブルディスク、CD-ROM、光磁気ディスク、半導体メモリ、ハードディスク等のコンピュータ読み取り可能な記憶媒体又は記憶装置に格納されるようにしてもよい。尚、中間的な処理結果は、一般的にメインメモリ等の記憶装置に一時保管される。

30

【0183】

以上の実施例を含む実施形態に関し、さらに以下の付記を開示する。

【0184】

(付記1)

コンピュータが、

各学習事例の重みに基づいて、前記重みに対応する各学習事例から夫々の正解構造を抽出する際の手掛かりとなる素性の各々に対応する第1スコアを更新することによって、当該第1スコアの各々により規定されるモデルを算出する第1ステップと、

40

前記各学習事例について、当該学習事例から導かれる複数の候補構造のうち前記正解構造以外の前記候補構造の各々に対して、前記モデルにより付与される第2スコアのうち最大スコアを特定し、当該学習事例の前記正解構造に対して前記モデルにより付与される第2スコアと前記最大スコアとの差を求めると共に、前記各学習事例における前記差により定義されるエラーの上限値に基づいて、前記モデルの確信度を算出する第2ステップと、

前記確信度と前記各学習事例における前記差とに基づいて、前記各学習事例の前記重みを更新する第3ステップと、

前記第1ステップと、前記第2ステップと、前記第3ステップとを繰り返すステップと

50

を実行するモデル学習方法。

【0185】

(付記2)

前記第1ステップにおいて、前記各学習事例の前記重みの各々を、当該学習事例の学習率に設定して、前記第1スコアを更新する

付記1記載のモデル学習方法。

【0186】

(付記3)

各学習事例の重みに基づいて、前記重みに対応する各学習事例から夫々の正解構造を抽出する際の手掛かりとなる素性の各々に対応する第1スコアを更新することによって、当該第1スコアの各々により規定されるモデルを算出する第1ステップと、

前記各学習事例について、当該学習事例から導かれる複数の候補構造のうち前記正解構造以外の前記候補構造の各々に対して、前記モデルにより付与される第2スコアのうち最大スコアを特定し、当該学習事例の前記正解構造に対して前記モデルにより付与される第2スコアと前記最大スコアとの差を求めると共に、前記各学習事例における前記差により定義されるエラーの上限値に基づいて、前記モデルの確信度を算出する第2ステップと、

前記確信度と前記各学習事例における前記差とに基づいて、前記各学習事例の前記重みを更新する第3ステップと、

前記第1ステップと、前記第2ステップと、前記第3ステップとを繰り返すステップと、

を含み、コンピュータに実行させるためのモデル学習プログラム。

【0187】

(付記4)

重みデータ格納部に格納されている各学習事例の重みに基づいて、学習データ格納部に格納されている前記重みに対応する各学習事例から夫々の正解構造を抽出する際の手掛かりとなる素性の各々に対応する第1スコアを更新することによって、当該第1スコアの各々により規定されるモデルを算出し、当該モデルをモデルデータ格納部に格納する第1算出手段と、

前記学習データ格納部に格納されている前記各学習事例について、当該学習事例から導かれる複数の候補構造のうち前記正解構造以外の前記候補構造の各々に対して、前記モデルデータ格納部に格納されている前記モデルにより付与される第2スコアのうち最大スコアを特定し、当該学習事例の前記正解構造に対して前記モデルにより付与される第2スコアと前記最大スコアとの差を求め、当該差を差データ格納部に格納すると共に、前記各学習事例における前記差により定義されるエラーの上限値に基づいて、前記モデルの確信度を算出し、当該確信度を確信度データ格納部に格納する第2算出手段と、

前記確信度データ格納部に格納されている前記確信度と前記差データ格納部に格納されている前記各学習事例における前記差とに基づいて、前記重みデータ格納部に格納されている前記各学習事例の前記重みを更新する更新手段と、

を有し、

前記第1算出手段による処理と前記第2算出手段による処理と前記更新手段による処理とを繰り返させるモデル学習装置。

【符号の説明】

【0188】

101	学習データ入力部	103	学習データ格納部
105	モデル学習部	106	差データ格納部
107	重みデータ格納部	108	確信度格納部
109	モデルデータ格納部	111	分類データ入力部
113	分類データ格納部	115	モデル適用部
117	処理結果格納部	201	ブースティング部
203	構造学習部	205	初期化部

10

20

30

40

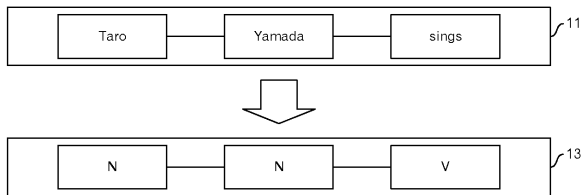
50

2 0 7 確信度算出部

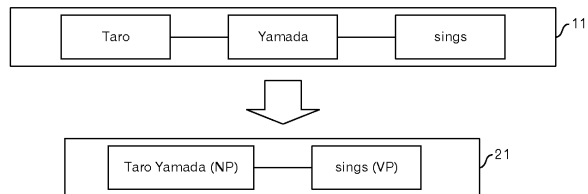
2 0 9 結合モデル更新部

2 1 1 重み更新部

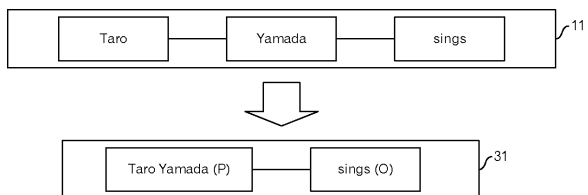
【図 1】



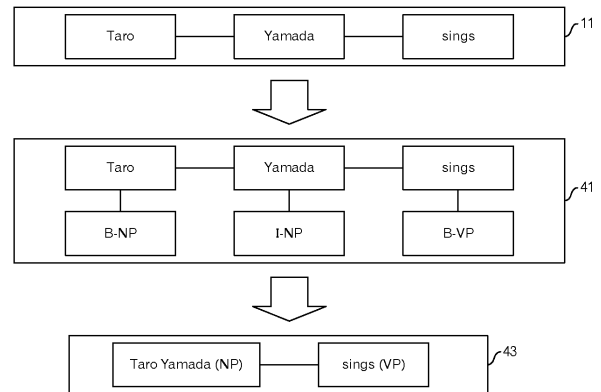
【図 2】



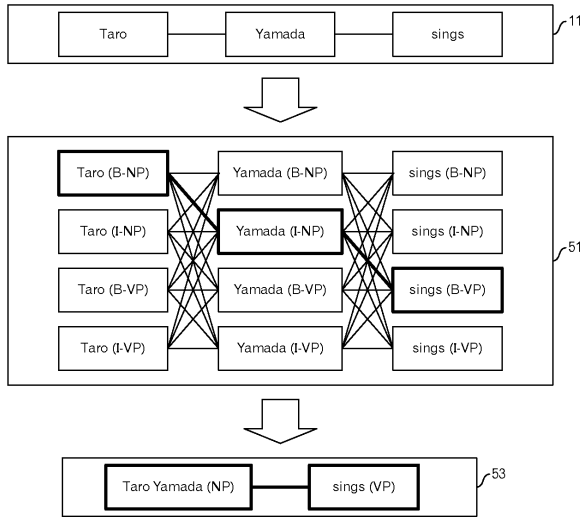
【図 3】



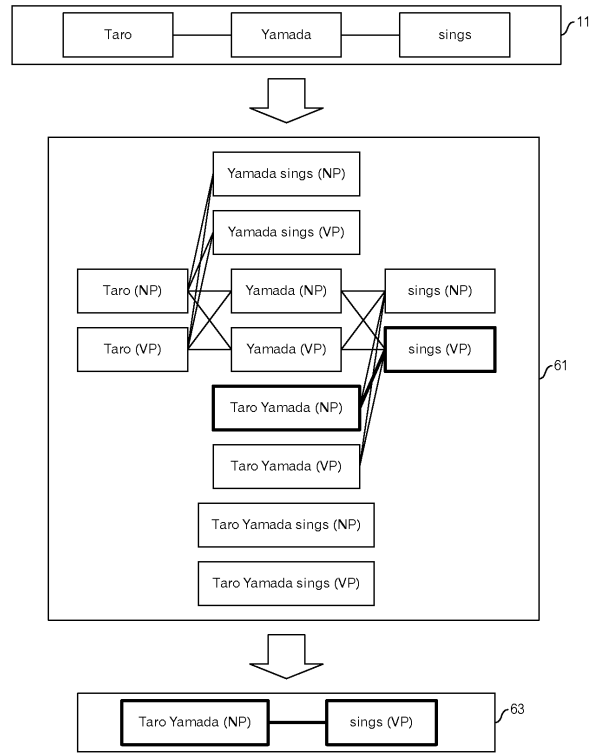
【図 4】



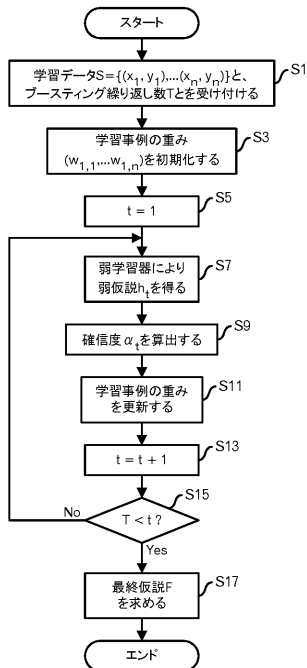
【図 5】



【図 6】



【図 7】



【図 8】

ラベル	重み	素性
+1	1/3	Yamada
-1	1/3	sings
+1	1/3	Taro



規則	スコア
sings	-0.69954
Yamada	0.29845
Taro	0.29845

【図 9】

	第1単語	第2単語	第3単語
第1学習事例X1	Yamada	sings	-
第2学習事例X2	Taro	sings	-
第3学習事例X3	Taro	Yamada	sings

【図 10】

	第1単語	第1ラベル	第2単語	第2ラベル	第3単語	第3ラベル
第1正解構造 Y_1	Yamada	B-NP	sings	B-VP	-	-
第2正解構造 Y_2	Taro	B-NP	signs	B-VP	-	-
第3正解構造 Y_3	Taro	B-NP	Yamada	I-NP	signs	B-VP

【図 11】

第1学習事例の重み $w_{1,1}$	1/3
第2学習事例の重み $w_{1,2}$	1/3
第3学習事例の重み $w_{1,3}$	1/3

【図 12】

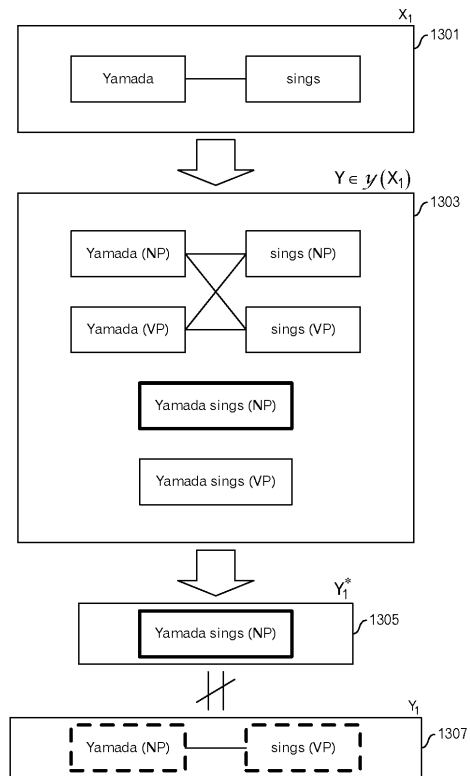
手掛かり	第1スコア
Taro (NP)	0
Yamada (NP)	0
Taro (VP)	0
Yamada (VP)	0
sings (VP)	0
Yamada sings (NP)	0
Taro Yamada (NP)	0
\vdots	\vdots

【図 14】

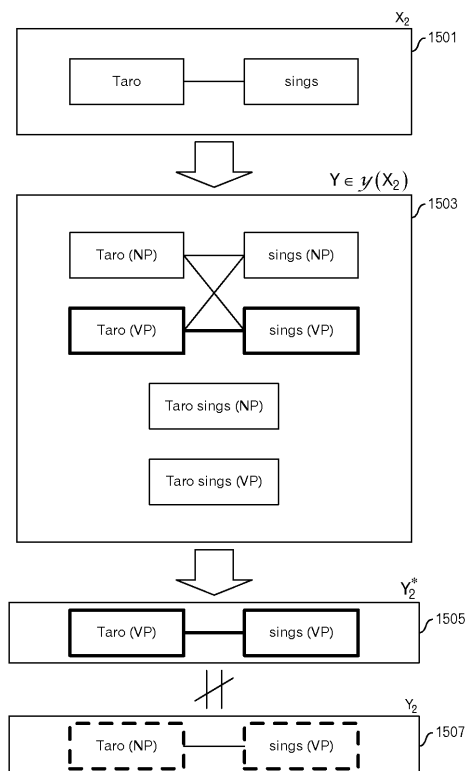
手掛かり	第1スコア
Taro (NP)	0
Yamada (NP)	1/3
Taro (VP)	0
Yamada (VP)	0
sings (VP)	1/3
Yamada sings (NP)	-1/3
Taro Yamada (NP)	0
\vdots	\vdots

$w_{1,1} = 1/3$ を加える
 $w_{1,1} = 1/3$ を引く

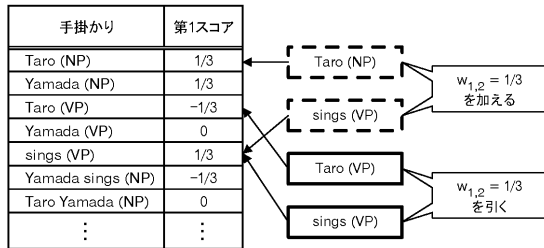
【図 13】



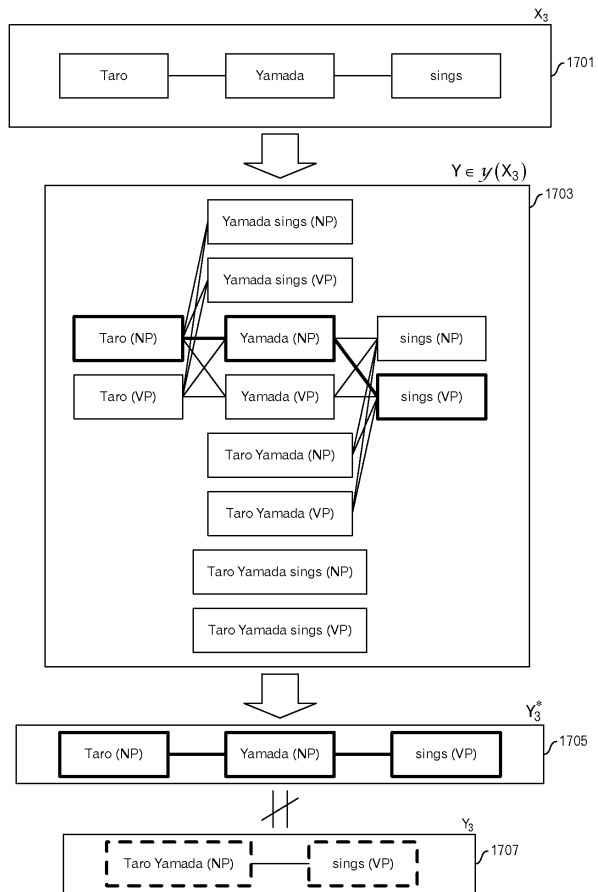
【図 15】



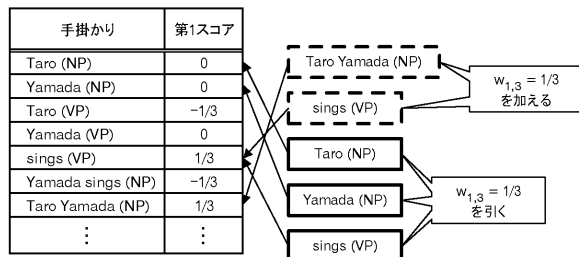
【図 16】



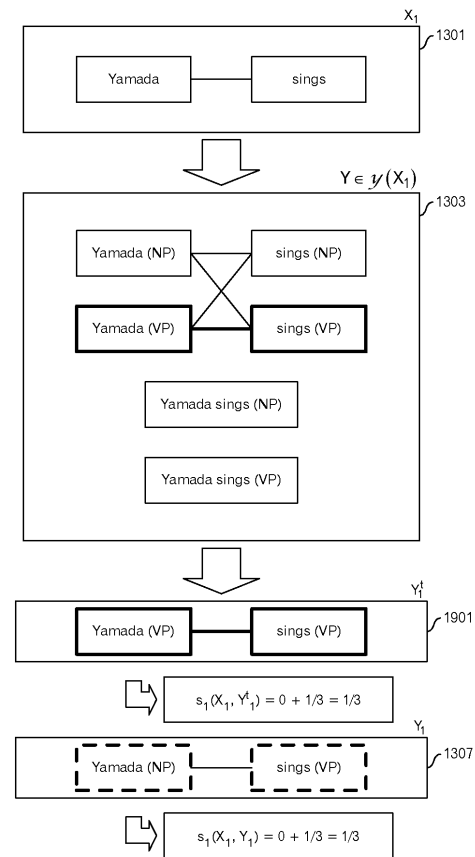
【図 17】



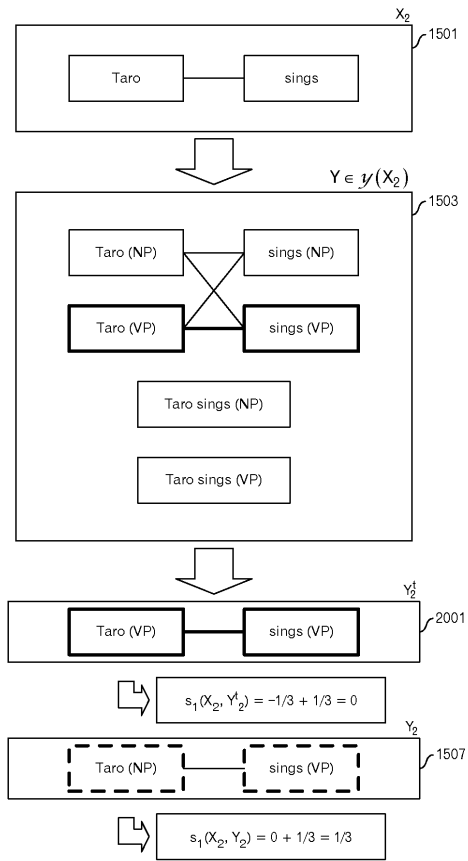
【図 18】



【図 19】



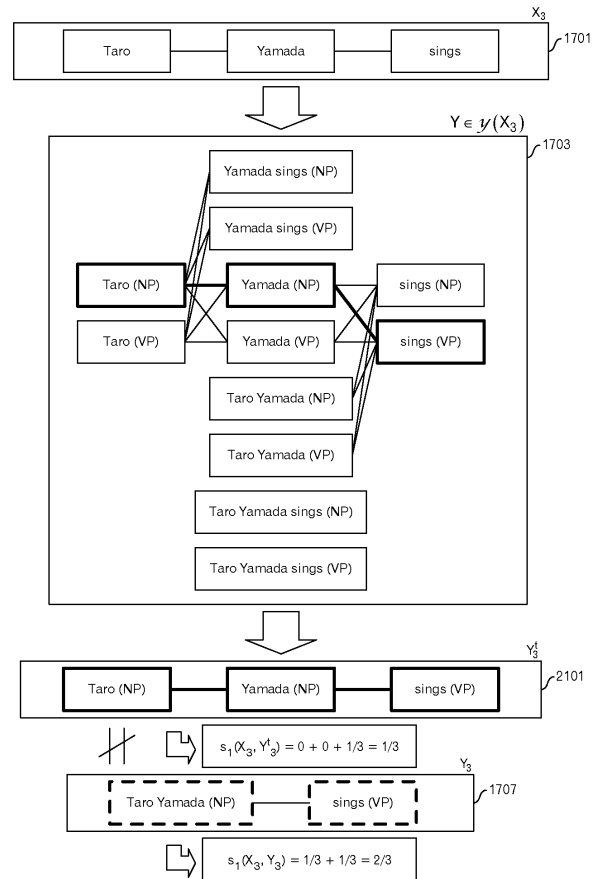
【図 20】



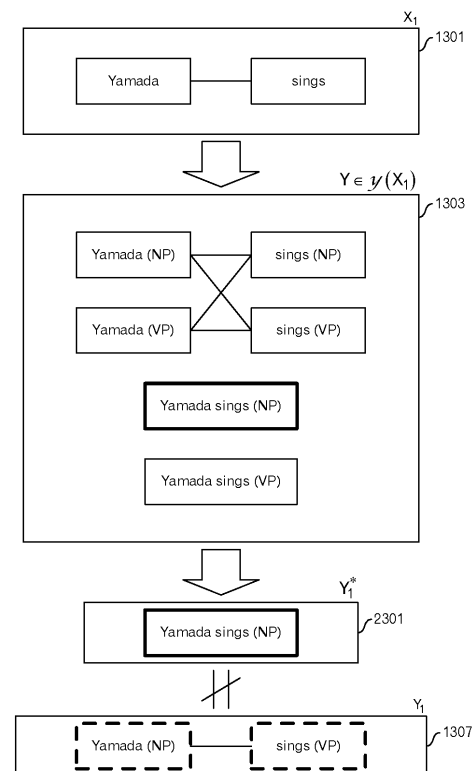
【図 22】

第1学習事例の重み $w_{2,1}$	0.411
第2学習事例の重み $w_{2,2}$	0.29449
第3学習事例の重み $w_{2,3}$	0.29449

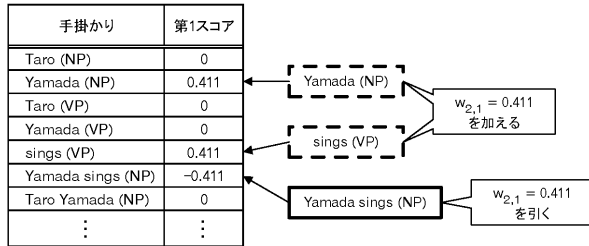
【図 21】



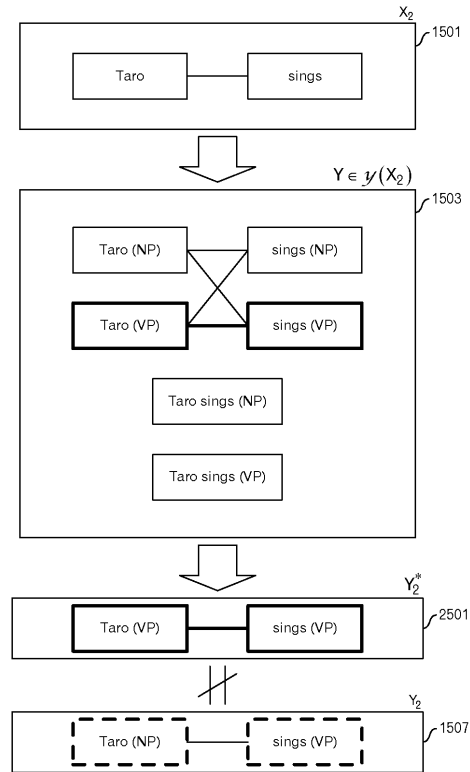
【図 23】



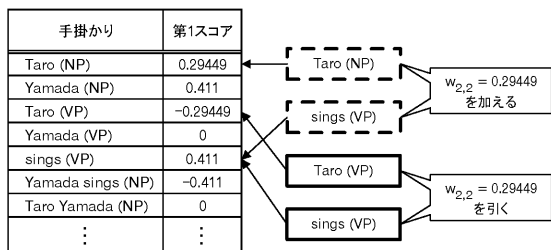
【図 24】



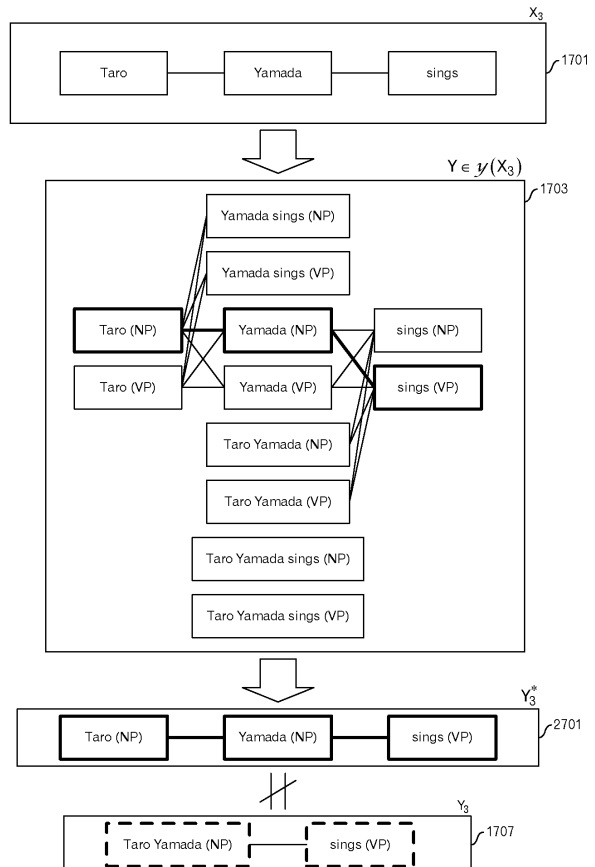
【図 25】



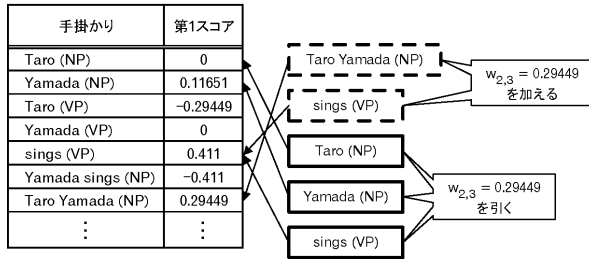
【図 26】



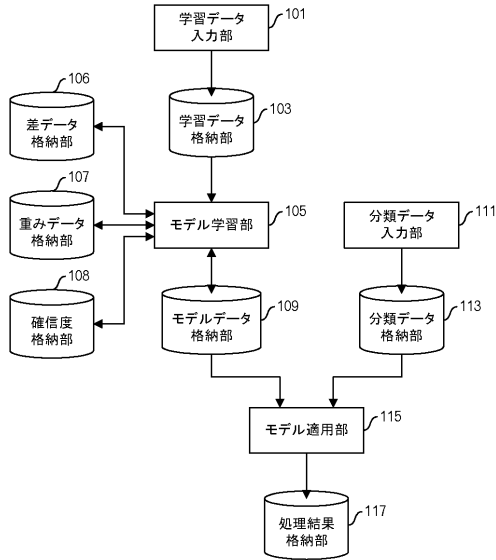
【図 27】



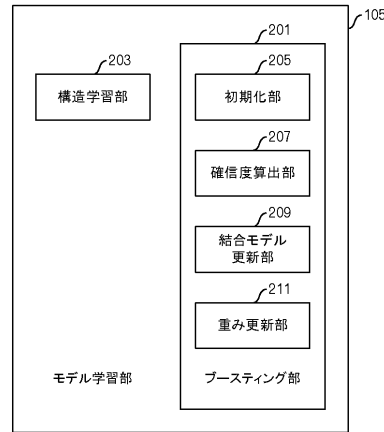
【図 28】



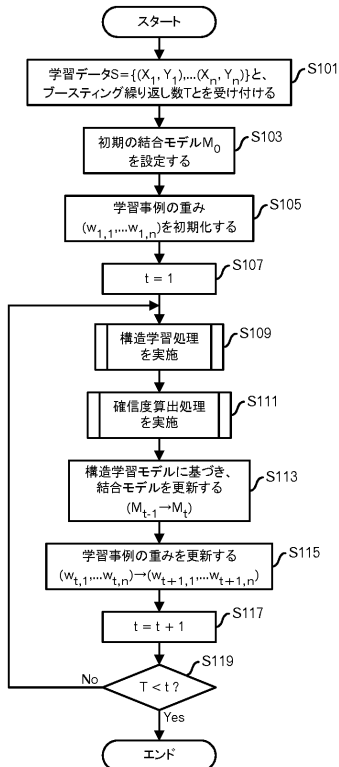
【図 29】



【図 30】



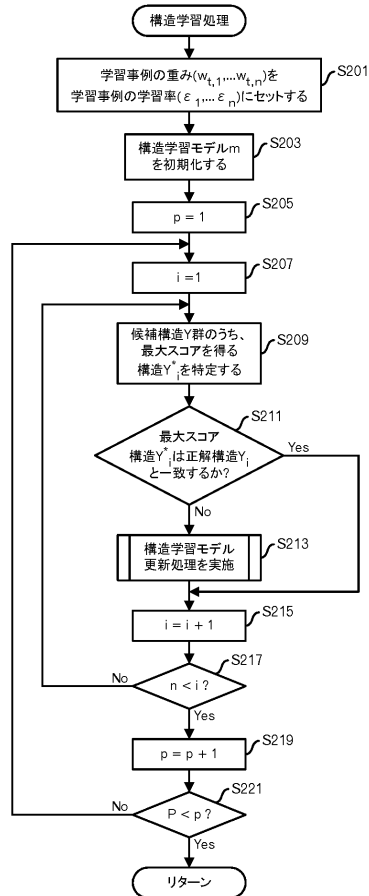
【図 31】



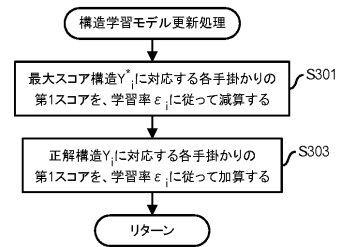
【図 32】

手掛かり	第1スコア
Taro (NP)	0
Yamada (NP)	0
Taro (VP)	0
Yamada (VP)	0
sings (VP)	0
Yamada sings (NP)	0
Taro Yamada (NP)	0
⋮	⋮

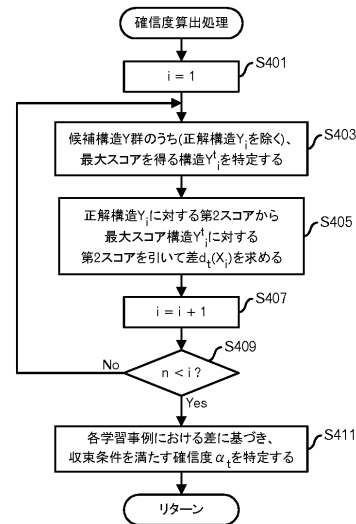
【図 3 3】



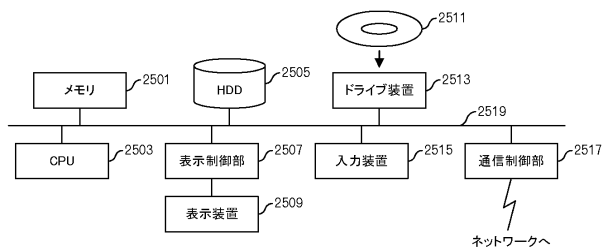
【図 3 4】



【図 3 5】



【図 3 6】



フロントページの続き

(56)参考文献 岩倉友哉，外 2 名，大規模自然言語処理学習データのための複数弱仮説を生成する弱学習器を用いる Ada Boost 手法，電気学会論文誌 C 電子・情報・システム部門誌，日本，（社）電気学会，2010 年 1 月 1 日，第130巻，第1号，pp.83-91

(58)調査した分野(Int.Cl.，DB名)
G 0 6 N 9 9 / 0 0