

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第6597978号
(P6597978)

(45) 発行日 令和1年10月30日(2019.10.30)

(24) 登録日 令和1年10月11日(2019.10.11)

(51) Int.Cl.		F I			
G06F	17/27	(2006.01)	G06F	17/27	610
G06N	5/04	(2006.01)	G06N	5/04	
			G06F	17/27	615

請求項の数 5 (全 15 頁)

(21) 出願番号	特願2016-223547 (P2016-223547)	(73) 特許権者	000004226
(22) 出願日	平成28年11月16日(2016.11.16)		日本電信電話株式会社
(65) 公開番号	特開2018-81505 (P2018-81505A)		東京都千代田区大手町一丁目5番1号
(43) 公開日	平成30年5月24日(2018.5.24)	(73) 特許権者	504132272
審査請求日	平成31年1月31日(2019.1.31)		国立大学法人京都大学
			京都府京都市左京区吉田本町36番地1
		(74) 代理人	110001519
			特許業務法人太陽国際特許事務所
		(72) 発明者	西野 正彬
			東京都千代田区大手町一丁目5番1号 日
			本電信電話株式会社内
		(72) 発明者	山本 章博
			京都府京都市左京区吉田本町36番地1
			国立大学法人京都大学内

最終頁に続く

(54) 【発明の名称】 構文解析装置、方法、及びプログラム

(57) 【特許請求の範囲】

【請求項1】

終端記号の系列に対して、確率文脈自由文法に基づいた構文解析を行う構文解析装置であって、

入力された、確率文脈自由文法における非終端記号を書き換えるための生成規則に関する制約条件である論理制約を表す論理関数に対応する二分決定グラフを構築するBDD構築部と、

予め与えられた生成規則の集合と、前記終端記号の系列とに基づいて、予め定められた開始記号から前記終端記号の系列を得るまでの、前記生成規則の列を導出とし、前記終端記号の系列に対して可能な導出の集合を求め、

前記求められた前記可能な導出の集合に基づいて、前記導出において各生成規則が利用されることを、命題変数を用いて表現し、

前記命題変数を用いて前記終端記号の系列に対する正しい導出であるか否かを表現する論理関数を定義し、

前記定義された論理関数に対応する二分決定グラフを構築し、

前記BDD構築部によって構築された前記二分決定グラフと、前記定義された論理関数に対応する前記二分決定グラフとの論理積となる前記二分決定グラフを用いて、前記制約条件を満たし、かつ、確率が最大となる導出を求める構文解析部と、

を含む構文解析装置。

【請求項2】

前記構文解析部は、前記終端記号の系列に含まれる前記終端記号の各々について、非終端記号を前記終端記号に書き換える生成規則が、前記導出において利用されることを表す命題変数と、前記終端記号の系列に含まれる部分系列の各々について、前記部分系列に対応して前記生成規則が、前記導出において利用されることを表す命題変数とを用いて、前記論理関数を定義する請求項 1 に記載の構文解析装置。

【請求項 3】

終端記号の系列に対して、確率文脈自由文法に基づいた構文解析を行う構文解析装置における構文解析方法であって、

BDD 構築部が、入力された、確率文脈自由文法における非終端記号を書き換えるための生成規則に関する制約条件である論理制約を表す論理関数に対応する二分決定グラフを構築するステップと、

構文解析部が、予め与えられた生成規則の集合と、前記終端記号の系列とに基づいて、予め定められた開始記号から前記終端記号の系列を得るまでの、前記生成規則の列を導出とし、前記終端記号の系列に対して可能な導出の集合を求めるステップと、

前記構文解析部が、前記求められた前記可能な導出の集合に基づいて、前記導出において各生成規則が利用されることを、命題変数を用いて表現するステップと、

前記構文解析部が、前記命題変数を用いて前記終端記号の系列に対する正しい導出であるか否かを表現する論理関数を定義するステップと、

前記構文解析部が、前記定義された論理関数に対応する二分決定グラフを構築するステップと、

前記構文解析部が、前記 BDD 構築部によって構築された前記二分決定グラフと、前記定義された論理関数に対応する前記二分決定グラフとの論理積となる前記二分決定グラフを用いて、前記制約条件を満たし、かつ、確率が最大となる導出を求めるステップと、

を含む構文解析方法。

【請求項 4】

前記構文解析部が定義するステップは、前記終端記号の系列に含まれる前記終端記号の各々について、非終端記号を前記終端記号に書き換える生成規則が、前記導出において利用されることを表す命題変数と、前記終端記号の系列に含まれる部分系列の各々について、前記部分系列に対応して前記生成規則が、前記導出において利用されることを表す命題変数とを用いて、前記論理関数を定義する請求項 3 に記載の構文解析方法。

【請求項 5】

コンピュータを、請求項 1 又は請求項 2 に記載の構文解析装置の各部として機能させるためのプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、構文解析装置、方法、及びプログラムに係り、特に、終端記号の系列に対して、構文解析を行うための構文解析装置、方法、及びプログラムに関する。

【背景技術】

【0002】

確率文脈自由文法は自然言語文の構文解析に用いられる基礎的な技術である。確率文脈自由文法 G は、 $G = (M, T, R, S, P)$ からなる 5 つ組として定義される。ここで M は非終端記号の集合、 T は終端記号の集合、 R は生成規則の集合、 S は開始記号、 P は生成規則に付与された確率値の集合である。確率文脈自由文法がチョムスキー標準形であるとすれば、生成規則は以下 (1) ~ (3) 式のいずれかである。

【0003】

【数 1】

$$A \rightarrow BC$$

10

20

30

40

50

・・・ (1)

$A \rightarrow \alpha$

・・・ (2)

$S \rightarrow \epsilon$

・・・ (3)

【 0 0 0 4 】

ここでA、B、及びCは非終端記号、 α は終端記号、Sは開始記号、 ϵ は空列を表すものとする。

【 0 0 0 5 】

各規則は、規則の左辺（以下では頭部とよぶ）の非終端記号Aを右辺の記号列に書き換えることを表している。ある規則 $A \rightarrow BC$ を適用するとは、非終端記号Aを含む系列において、Aを非終端記号の列BCに書き換えることを表す。

【 0 0 0 6 】

各規則に対してはその規則が適用される確率が付与されている。与えられた確率文脈自由文法において、規則の頭部が非終端記号Aであるような規則の集合を r_1, \dots, r_m とすると、各規則 r_i には出現確率 w_i が与えられる。ここで確率 w_i は

【 0 0 0 7 】

【数2】

$$0 < w_i < 1, \sum_{i=1}^m w_i = 1$$

【 0 0 0 8 】

を満たす。各確率は、非終端記号Aが与えられたときに確率 w_i で生成規則 r_i を適用することを表している。

【 0 0 0 9 】

ある確率文脈自由文法Gが与えられると、それをを用いた構文解析が可能となる。すなわち、与えられた終端記号の列 $\alpha_1, \alpha_2, \dots, \alpha_n$ に対して、開始記号Sから当該終端記号の列を導出するための生成規則の適用による非終端記号の書き換えの列を得ることができる。ある構文解析は、開始記号に対して適用した生成規則の列として表現することができる。この列を導出ともよぶ。また、生成規則の列に含まれる各規則に付与された確率の積として、その解析が得られる確率が定義される。

【 0 0 1 0 】

確率文脈自由文法を用いることで自然言語文のように曖昧さのある対象に対する構文解析を行うことができるため、自然言語文の解析や、RNA（リボ核酸）の二次構造予測などに用いられてきた。また、確率文脈自由文法が与えられたときに、ある系列に対する構文解析のうち、確率が最大となるものを発見する問題は、系列の長さをNとすると $O(N^3)$ の動的計画法を用いることで解くことができる。

【先行技術文献】

【非特許文献】

【 0 0 1 1 】

【非特許文献1】Bryant, Randal E. "Graph-based algorithms for boolean function manipulation." Computers, IEEE Transactions on 100.8 (1986): 677-691.

【非特許文献2】Knuth, D. E. (2009). The Art of Computer Programming: Bitwise Tricks & Techniques; Binary Decision Diagrams, volume 4, fascicle 1.

【発明の概要】

【発明が解決しようとする課題】

10

20

30

40

50

【 0 0 1 2 】

確率文脈自由文法を用いた構文解析には動的計画法を用いた効率的なアルゴリズムが存在し、系列の要素間の関係性を考慮した解析が可能であることから、様々な問題に利用される。一方で、既存の動的計画法に基づく構文解析アルゴリズムでは、適用される規則の間の制約条件を考慮することができなかった。ここで制約条件とは、例えばある規則 r R が導出中に出現する回数に制限を加えたり、規則 r , s R が導出中で同時に出現することを禁止したりするものである。

【 0 0 1 3 】

本発明は、上記事情を鑑みて成されたものであり、生成規則に関する任意の制約条件を考慮した構文解析を行うことができる構文解析装置、方法、及びプログラムを提供することを目的とする。

10

【課題を解決するための手段】

【 0 0 1 4 】

上記目的を達成するために、第1の発明に係る構文解析装置は、終端記号の系列に対して、確率文脈自由文法に基づいた構文解析を行う構文解析装置であって、入力された、確率文脈自由文法における非終端記号を書き換えるための生成規則に関する制約条件である論理制約を表す論理関数に対応する二分決定グラフを構築するBDD構築部と、予め与えられた生成規則の集合と、前記終端記号の系列とに基づいて、予め定められた開始記号から前記終端記号の系列を得るまでの、前記生成規則の列を導出とし、前記終端記号の系列に対して可能な導出の集合を求め、前記求められた前記可能な導出の集合に基づいて、前記導出において各生成規則が利用されることを、命題変数を用いて表現し、前記命題変数を用いて前記終端記号の系列に対する正しい導出であるか否かを表現する論理関数を定義し、前記定義された論理関数に対応する二分決定グラフを構築し、前記BDD構築部によって構築された前記二分決定グラフと、前記定義された論理関数に対応する前記二分決定グラフとの論理積となる前記二分決定グラフを用いて、前記制約条件を満たし、かつ、確率が最大となる導出を求める構文解析部と、を含んで構成されている。

20

【 0 0 1 5 】

また、第1の発明に係る構文解析装置において、前記構文解析部は、前記終端記号の系列に含まれる前記終端記号の各々について、非終端記号を前記終端記号に書き換える生成規則が、前記導出において利用されることを表す命題変数と、前記終端記号の系列に含まれる部分系列の各々について、前記部分系列に対応して前記生成規則が、前記導出において利用されることを表す命題変数とを用いて、前記論理関数を定義するようにしてもよい。

30

【 0 0 1 6 】

第2の発明に係る構文解析方法は、終端記号の系列に対して、確率文脈自由文法に基づいた構文解析を行う構文解析装置における構文解析方法であって、BDD構築部が、入力された、確率文脈自由文法における非終端記号を書き換えるための生成規則に関する制約条件である論理制約を表す論理関数に対応する二分決定グラフを構築するステップと、構文解析部が、予め与えられた生成規則の集合と、前記終端記号の系列とに基づいて、予め定められた開始記号から前記終端記号の系列を得るまでの、前記生成規則の列を導出とし、前記終端記号の系列に対して可能な導出の集合を求めるステップと、前記構文解析部が、前記求められた前記可能な導出の集合に基づいて、前記導出において各生成規則が利用されることを、命題変数を用いて表現するステップと、前記構文解析部が、前記命題変数を用いて前記終端記号の系列に対する正しい導出であるか否かを表現する論理関数を定義するステップと、前記構文解析部が、前記定義された論理関数に対応する二分決定グラフを構築するステップと、前記構文解析部が、前記BDD構築部によって構築された前記二分決定グラフと、前記定義された論理関数に対応する前記二分決定グラフとの論理積となる前記二分決定グラフを用いて、前記制約条件を満たし、かつ、確率が最大となる導出を求めるステップと、を含んで実行することを特徴とする。

40

【 0 0 1 7 】

50

また、第2の発明に係る構文解析方法において、前記構文解析部が定義するステップは、前記終端記号の系列に含まれる前記終端記号の各々について、非終端記号を前記終端記号に書き換える生成規則が、前記導出において利用されることを表す命題変数と、前記終端記号の系列に含まれる部分系列の各々について、前記部分系列に対応して前記生成規則が、前記導出において利用されることを表す命題変数とを用いて、前記論理関数を定義するようにしてもよい。

【0018】

また、第3の発明に係るプログラムは、コンピュータを、第1の発明に係る構文解析装置の各部として機能させるためのプログラムである。

【発明の効果】

10

【0019】

本発明の構文解析装置、方法、及びプログラムによれば、生成規則に関する制約条件である論理制約を表す論理関数に対応する二分決定グラフを構築し、予め与えられた生成規則の集合と、終端記号の系列とに基づいて、予め定められた開始記号から終端記号の系列を得るまでの、生成規則の列を導出とし、終端記号の系列に対して可能な導出の集合を求め、求められた可能な導出の集合に基づいて、導出において各生成規則が利用されることを、命題変数を用いて表現し、命題変数を用いて終端記号の系列に対する正しい導出であるか否かを表現する論理関数を定義し、定義された論理関数に対応する二分決定グラフを構築し、構築された二分決定グラフと、定義された論理関数に対応する二分決定グラフとの論理積となる二分決定グラフを用いて、制約条件を満たし、かつ、確率が最大となる導出を求めることにより、生成規則に関する任意の制約条件を考慮した構文解析を行うことができる、という効果が得られる。

20

【図面の簡単な説明】

【0020】

【図1】二分決定グラフ(BDD)の一例を示す図である。

【図2】本発明の実施の形態に係る構文解析装置の構成を示すブロック図である。

【図3】本発明の実施の形態に係る構文解析装置における構文解析処理ルーチンを示すフローチャートである。

【発明を実施するための形態】

【0021】

30

以下、図面を参照して本発明の実施の形態を詳細に説明する。

【0022】

<本発明の実施の形態に係る概要>

【0023】

まず、本発明の実施の形態における概要を説明する。

【0024】

本実施の形態で提案する、論理制約が存在するもとの確率文脈自由文法に基づく構文解析の特徴は、まず確率文脈自由文法における各生成規則を命題変数に置き換えて、可能な導出の集合を、論理関数を用いて表現し、論理関数を二分決定グラフ(BDD)とよばれるデータ構造を用いて表現することにある。二分決定グラフを用いて論理関数を表現することによって、可能な導出の集合のうち制約条件を満たし、かつ確率最大となる導出の検索を行うことができるようになる。

40

【0025】

<本発明の実施の形態に係る原理>

【0026】

次に、本発明の実施の形態における二分決定グラフ(BDD)の原理を説明する。

【0027】

二分決定グラフ(BDD)は論理関数を有向非巡回グラフとして表現するデータ構造である。N引数の論理関数

【0028】

50

【数3】

$$f : \{0, 1\}^N \mapsto \{0, 1\}$$

【0029】

は、 N 個の命題変数 x_1, \dots, x_N を引数として受け取り、0または1を出力する関数である。論理関数は命題変数と論理演算子を用いて記述することができる。例えば x_1, x_2 を引数として受け取り、 $x_1 = 1$ かつ $x_2 = 1$ のときに1、それ以外の場合に0を返すような論理関数は、

【0030】

【数4】

$$f(x_1, x_2) = x_1 \wedge x_2$$

【0031】

として表現することができる。3引数の論理関数

【0032】

【数5】

$$f(x, y, z) = (x \wedge y) \vee (x \wedge z) \vee (y \wedge z)$$

【0033】

を表現するBDDを図1に示す。BDDは2種類のノードからなる。

【0034】

そのノードを親とするようなアークが存在しないノードを終端ノードとよび、図中では四角で表現する。それ以外のノードを分岐ノードとよび、図中では丸で表現する。BDDは必ず記号

【0035】

【数6】

\perp, \top

【0036】

が付与された2種類の終端ノードを含む。また、分岐ノードは必ずそのノードを親とする2つのアークを持つ。それぞれのアークは区別され、0枝、1枝とよばれる。図1中では0枝は破線、1枝は実線で表現される。各分岐ノードは論理関数のいずれかの引数である命題変数に対応しており、対応している変数は図1中では分岐ノードの丸の中に書かれた記号で表現されている。対応している変数をBDDの分岐ノードのラベルとよぶ。また、最上位の分岐ノードをBDDの根とよぶ。

【0037】

BDD上で最上位のノードから、入力に応じた枝を辿ることによって、ある引数の値に対する論理関数の出力を得ることができる。例で挙げた論理関数に入力として $x = 1, y = 1, z = 1$ が与えられたときに図1のBDDを用いて関数を評価する方法を示す。まず、根のラベルは x であり、入力では $x = 1$ であることから、根から1枝を辿って図中の直下にある分岐ノードに遷移する。辿った先のノードのラベルは y であり、入力は $y = 1$ であることから、ノードから1枝を辿って

【数7】

\top

10

20

30

40

50

終端ノードに遷移する。

【数 8】

┆

終端ノードに遷移した場合、関数は与えられた入力に対して 1 を返すことを示すため、入力 $x = 1$, $y = 1$, $z = 1$ に対する出力は 1 であることが分かる。同様に、入力として $x = 1$, $y = 0$, $z = 0$ が与えられたとすると、BDD を根から順番に辿るとこの場合は

【0038】

【数 9】

10

┆

【0039】

終端ノードに到達する。

【0040】

【数 10】

┆

20

【0041】

終端ノードに到達したときには関数は 0 を返すことを示すため、入力 $x = 1$, $y = 0$, $z = 0$ に対する出力は 0 であることが分かる。

【0042】

BDD を用いて論理関数を表現することで、上述した関数の評価のほかにも論理関数間の演算を行うことができたり、関数のモデル（関数が 1 を返すような引数の値の組合せの個数）を数えるなどの演算を高速に実行することが可能となる。

【0043】

< 本発明の実施の形態に係る構文解析装置の構成 >

【0044】

30

次に、本発明の実施の形態に係る構文解析装置の構成について説明する。図 2 に示すように、本発明の実施の形態に係る構文解析装置 100 は、CPU と、RAM と、後述する構文解析処理ルーチンを実行するためのプログラムや各種データを記憶した ROM と、を含むコンピュータで構成することが出来る。この構文解析装置 100 は、機能的には図 2 に示すように文入力部 10 と、論理制約入力部 12 と、生成規則入力部 14 と、演算部 20 と、出力部 50 とを備えている。

【0045】

文入力部 10 は、解析の対象となる終端記号の系列（例えば、自然言語文）の入力を受け付ける。

【0046】

40

論理制約入力部 12 は、確率文脈自由文法における非終端記号を書き換えるための生成規則に関する制約条件である論理制約を表す論理関数を受け付ける。論理関数の引数については後述する。

【0047】

生成規則入力部 14 は、チョムスキー標準形で与えられる生成規則の集合を入力として受け取る。

【0048】

演算部 20 は、生成規則記憶部 28 と、BDD 構築部 30 と、構文解析部 32 とを含んで構成されている。

【0049】

50

生成規則記憶部 28 は、生成規則入力部 14 で受け付けた生成規則の集合を記憶している。

【0050】

BDD構築部 30 は、論理制約入力部 12 で受け付けた論理関数に対応する二分決定グラフを構築する。

【0051】

構文解析部 32 は、生成規則記憶部 28 に記憶されている生成規則の集合と、文入力部 10 で受け付けた終端記号の系列とに基づいて、予め定められた開始記号 S から当該終端記号の系列を得るまでに適用される生成規則の列を導出とし、終端記号の系列に対して可能な導出の集合を求め、求められた可能な導出の集合に基づいて、導出において各生成規則が利用されることを、命題変数を用いて表現し、命題変数を用いて終端記号の系列に対する正しい導出であるか否かを表現する論理関数を定義する。構文解析部 32 は、定義された論理関数に対応する二分決定グラフを構築し、BDD構築部 30 によって構築された二分決定グラフと、定義された論理関数に対応する二分決定グラフとの論理積となる二分決定グラフを用いて、可能な導出の集合のうち、論理制約入力部 12 で受け付けた論理関数が表す制約条件を満たし、かつ、確率が最大となる導出を求める。ここで、論理関数を定義する際には、終端記号の系列に含まれる終端記号 t_i の各々について、非終端記号を終端記号 t_i に書き換える生成規則 r_j が、導出において利用されることを表す命題変数 d_{ij} と、終端記号の系列に含まれる部分系列 $t_{i:j}$ の各々について、部分系列 $t_{i:j}$ に対応して生成規則 r_1 が、導出において利用されることを表す命題変数 b_{ij1} とを用いる。

【0052】

<本発明の実施の形態に係る構文解析装置の作用>

【0053】

次に、本発明の実施の形態に係る構文解析装置 100 の作用について説明する。生成規則入力部 14 が生成規則の集合を受け付けると、生成規則記憶部 28 に生成規則の集合が記憶される。そして、構文解析装置 100 は、図 3 に示す構文解析処理ルーチンを実行する。

【0054】

まず、ステップ S100 では、文入力部 10 が終端記号の系列を受け付け、論理制約入力部 12 が論理制約を表す論理関数を受け付ける。受け付けた長さ N の終端記号の列を t_1, t_2, \dots, t_N とする。 T を終端記号の集合とすると、各 t_i は T であるとする。また、生成規則記憶部 28 に記憶されている生成規則の集合を R とする。 R に含まれる生成規則のうち、非終端記号を終端記号に置き換える規則、すなわち $A \rightarrow B$ の形をしている規則を q_1, \dots, q_K とし、非終端記号を非終端記号のペアに置き換える規則、すなわち $A \rightarrow BC$ の形をしている規則を r_1, \dots, r_M とする。論理制約は、後述する命題変数 d_i, b_i を引数とする任意の論理関数として表現されるものとする。

【0055】

次に、ステップ S102 では、ステップ S100 で受け付けた論理制約を表す論理関数に対応する二分決定グラフを構築する。二分決定グラフの構築には、例えば非特許文献 1 に記載されているような二分決定グラフ構築のための既存手法を用いる。

【0056】

ステップ S104 では、ステップ S100 で受け付けた生成規則の集合と、終端記号の系列とに基づいて、予め定められた開始記号 X から終端記号の系列を得るまでの、生成規則の列を導出とし、終端記号の系列に対して可能な導出の集合を求める。ここでは、Cocke-Younger-Kasami (CYK) アルゴリズムを実行して、生成規則の集合による可能な導出の集合を求める。CYK アルゴリズムは動的計画法によって文脈自由文法の構文解析を行うアルゴリズムである。終端記号の系列の i 文字目から j 文字目までからなる部分系列を $t_{i:j}$ とすると、CYK アルゴリズムは任意の非終端記号を開始記号とする各部分系列への導出を再帰的に求める。すなわち、 $S_{i:j}(X)$ を、部分系列 $t_{i:j}$ に対する、非終端記号 X を

10

20

30

40

50

開始記号とする導出において、最初に利用された生成規則のインデックスの集合であるとすると、 $(i < j)$ であるときの $S_{ij}^{(X)}$ は以下(4)式のように再帰的に求められる。

【0057】

【数11】

$$S_{ij}^{(X)} = \bigcup_{k=i}^{j-1} \{i \mid r_i = (X \rightarrow YZ), S_{ik}^{(Y)} \neq \emptyset, S_{k+1,j}^{(Z)} \neq \emptyset\}$$

10

・・・(4)

【0058】

また、 $S_{ii}^{(X)}$ は、導出において利用される、非終端記号 X を終端記号 a_i に書き換える生成規則のインデックスの集合であり、以下(5)式のように定義される。

【0059】

【数12】

$$S_{ii}^{(X)} = \{j \mid q_j = (X \rightarrow \alpha), \alpha = a_i\}$$

20

・・・(5)

【0060】

すべての $S_{ij}^{(X)}$ の計算は $O(N^3)$ 時間で実行することができる。

【0061】

ステップS106では、求められた可能な導出の集合に基づいて、導出において各生成規則が利用されることを、命題変数を用いて表現する。命題変数は、終端記号の系列に含まれる終端記号 a_i の各々について、非終端記号を終端記号 a_i に書き換える生成規則 r_j が、導出において利用されることを表す命題変数 d_{ij} と、終端記号の系列に含まれる部分系列 a_{ij} の各々について、部分系列 a_{ij} に対応して生成規則 r_{j1} が、導出において利用されることを表す命題変数 b_{ij1} とを用いる。すなわち、命題変数 d_{ij} は、終端記号 a_i の各々について、

30

【0062】

【数13】

$$S_{ii}^{(X)} \neq \emptyset$$

【0063】

であるような生成規則のインデックス

40

【0064】

【数14】

$$j \in S_{ii}^{(X)}$$

【0065】

に対して定義される。

【0066】

また、命題変数 b_{ij1} は、終端記号 a_i から a_j について、 $i = j$ であり各

【0067】

50

【数 15】

$$S_{ij}^{(X)} \neq \emptyset$$

【0068】

である場合において、生成規則のインデックス $l = S_{ij}^{(X)}$ に対応する生成規則 $r_l = (X \rightarrow YZ)$ と、ある $k : i \leq k \leq j - 1$ について、

【0069】

【数 16】

$$S_{ik}^{(Y)} \neq \emptyset \text{ かつ } S_{k+1,j}^{(Z)} \neq \emptyset$$

10

【0070】

である生成規則のインデックス l に対して定義される。

【0071】

ステップ S 108 では、ステップ S 106 で得られた各命題変数 d_{ij} 、 b_{ijl} を用いて、終端記号の系列に対する正しい導出であるか否かを表現する論理関数を定義する。ここでは、論理式間を満たすべき制約の条件をそれぞれ以下 (6) ~ (11) のように設定することで、正しい導出を表現する論理関数を構成する。まず、各命題変数 d_{i*} のうち、二つ以上が同時に真とならない制約を以下 (6) 式とする。

20

【0072】

【数 17】

$$G_i = \bigwedge_{d_{ij}, d_{ik}: j \neq k} \neg d_{ij} \vee \neg d_{ik}$$

... (6)

30

【0073】

同様に、各命題変数 b_{ij*} のうち、二つ以上が同時に真とならない制約を以下 (7) 式のように表す。

【0074】

【数 18】

$$H_{ij} = \bigwedge_{b_{ijk}, b_{ijl}: k \neq l} \neg b_{ijk} \vee \neg b_{ijl}$$

... (7)

40

【0075】

次に、各集合 $S_{ij}^{(X)}$ にインデックスが含まれる生成規則のうちいずれかが真となる制約を、以下 (8) 式のように表す。

【0076】

【数 19】

$$F_{ij}^{(X)} = \bigvee_{k \in S_{ij}^{(X)}} b_{ijk}$$

・・・(8)

【0077】

同様に $S_{ii}^{(X)}$ にインデックスが含まれる生成規則のうちいずれかが真となる制約を、以下(9)式のように表す。

10

【0078】

【数 20】

$$F_{ii}^{(X)} = \bigvee_{j \in S_{ii}^{(X)}} d_{ij}$$

・・・(9)

【0079】

また、命題変数 b_{ijk} に対し、 $r_k = (X \ Y \ Z)$ であるならば、以下(10)式の制約を設定する。

20

【0080】

【数 21】

$$D_{ijk} = \neg b_{ijk} \vee \left(\bigvee_{l=i}^{j-1} F_{i,l}^{(Y)} \wedge F_{l+1,j}^{(Z)} \right)$$

・・・(10)

30

【0081】

最後に、 $i < k \leq j < l$ であるような全ての b_{ijm} 、 b_{kln} に対して、以下(11)式の制約を設定する。

【0082】

【数 22】

$$C = \bigwedge_{b_{ijm}, b_{kln}: i < k \leq j < l} \neg b_{ijm} \vee \neg b_{kln}$$

・・・(11)

40

【0083】

上記(6)～(11)式を用いると、正しい導出に対応する規則の集合が与えられたときに真を返す論理関数は、命題変数 d_{ij} 、 b_{ijk} を引数とする関数として以下(12)のように表現できる。

【0084】

【数 2 3】

$$\left(\bigwedge_{i=1}^N G_i \right) \wedge \left(\bigwedge_{i,j:1 \leq i < j \leq N} H_{ij} \right) \wedge \left(\bigwedge_{b_{ijk}} D_{ijk} \right) \wedge C$$

・・・(12)

【0085】

ステップS110では、ステップS108で定義された論理関数に対応する二分決定グラフを構築する。二分決定グラフの構築は、ステップS102と同様に非特許文献1に記載の既存手法を用いる。

10

【0086】

ステップS112では、ステップS102で構築された二分決定グラフと、ステップS110で構築された二分決定グラフとの論理積となる二分決定グラフを計算する。

【0087】

ステップS114では、ステップS112で構築された二分決定グラフを用いて、可能な導出のうち、ステップS100で受け付けた論理関数が表す制約条件を満たし、かつ、確率が最大となる導出を求める。ステップS112で計算した二分決定グラフを用いることで、導出確率を最大とする導出は二分決定グラフのサイズに比例する動的計画法で求めることができる。このとき、N引数 x_1, \dots, x_N の論理関数 $f(x_1, \dots, x_n)$ が二分決定グラフで表されているとする。また、各 x_i について重み u_i が定められているとすると、二分決定グラフを用いた、非特許文献2に記載されているアルゴリズム(非特許文献2の7ページに記載されている"AlgorithmB"を参照)を用いることで、以下(13)~(15)式を解くことができる。

20

【0088】

【数 2 4】

$$\text{Maximize } \sum_{i=1}^N u_i x_i$$

30

・・・(13)

$$\text{Subject to } f(x_1, \dots, x_N) = 1$$

・・・(14)

$$x_i \in \{0, 1\} \quad \forall i$$

40

・・・(15)

【0089】

本実施の形態の問題では、各命題変数 x_i について対応する生成規則 r_j が定まるので、その生成規則の出現確率 w_j の対数を取り、出現確率 w_j の対数を命題変数 x_i に対応する重み u_i として設定することによって、上記の定式化によって確率最大かつ制約条件を満たす導出を得ることが可能となる。

【0090】

以上説明したように、本発明の実施の形態に係る構文解析装置によれば、入力された生成規則に関する制約条件である論理制約を表す論理関数に対応する二分決定グラフを構築し、予め与えられた生成規則の集合と、終端記号の系列とに基づいて、予め定められた開

50

始記号から終端記号の系列を得るまでの、生成規則の列を導出とし、終端記号の系列に対して可能な導出の集合を求め、求められた可能な導出の集合に基づいて、導出において各生成規則が利用されることを、命題変数を用いて表現し、命題変数を用いて終端記号の系列に対する正しい導出であるか否かを表現する論理関数を定義し、定義された論理関数に対応する二分決定グラフを構築し、構築された二分決定グラフと、定義された論理関数に対応する二分決定グラフとの論理積となる二分決定グラフを用いて、制約条件を満たし、かつ、確率が最大となる導出を求めることにより、生成規則に関する任意の制約条件を考慮した構文解析を行うことができる。

【 0 0 9 1 】

なお、本発明は、上述した実施の形態に限定されるものではなく、この発明の要旨を逸脱しない範囲内で様々な変形や応用が可能である。

10

【 0 0 9 2 】

例えば、上述した実施の形態における構文解析装置は、自然言語文に対して、構文解析を行う場合について説明したが、これに限定されるものではなく、自然言語文以外の終端記号の列を受け付けて、構文解析を行うようにしてもよい。例えば、RNAなどの二次構造予測など、何らかの構造に対して、構文解析を行うようにしてもよい。

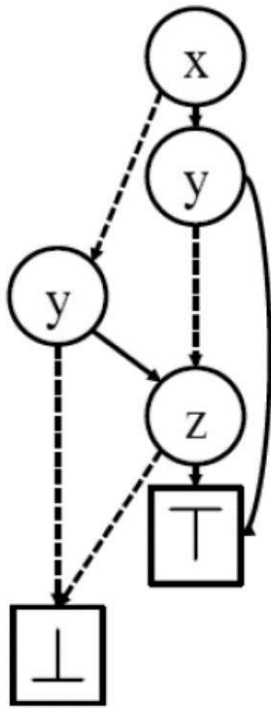
【 符号の説明 】

【 0 0 9 3 】

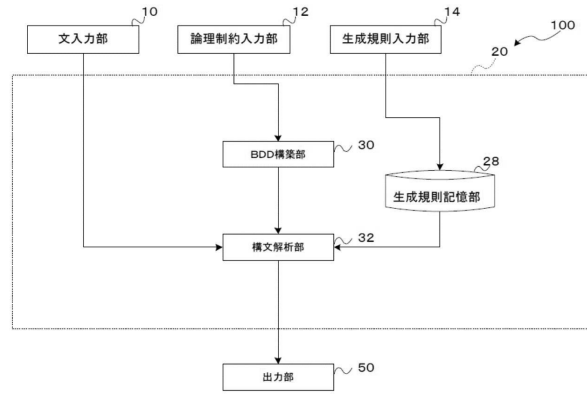
- 1 0 文入力部
- 1 2 論理制約入力部
- 1 4 生成規則入力部
- 2 0 演算部
- 2 8 生成規則記憶部
- 3 0 B D D 構築部
- 3 2 構文解析部
- 5 0 出力部
- 1 0 0 構文解析装置

20

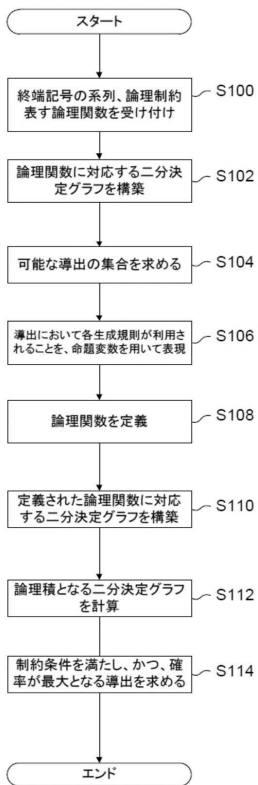
【図1】



【図2】



【図3】



フロントページの続き

審査官 萩島 豪

- (56)参考文献 特開2010-066970(JP,A)
特開平11-085748(JP,A)
特開平07-064792(JP,A)
特開2014-029622(JP,A)
佐藤 泰介, ベイジアンネットワークと離散構造処理系, 人工知能学会誌 第25巻 第6号,
日本, (社)人工知能学会, 2010年11月 1日, 第25巻, pp.796-802

(58)調査した分野(Int.Cl., DB名)

G06F 17/20 - 17/28
G06N 5/04