

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局



(10) 国際公開番号

WO 2010/116554 A1

PCT

(43) 国際公開日

2010年10月14日(14.10.2010)

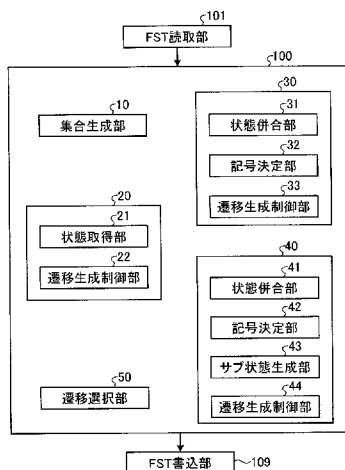
- (51) 国際特許分類:
G06F 7/00 (2006.01) G10L 15/18 (2006.01)
- (21) 国際出願番号: PCT/JP2009/068016
- (22) 国際出願日: 2009年10月19日(19.10.2009)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願 2009-082866 2009年3月30日(30.03.2009) JP
- (71) 出願人 (米国を除く全ての指定国について): 株式会社東芝 (Kabushiki Kaisha Toshiba) [JP/JP]; 〒1058001 東京都港区芝浦一丁目1番1号 Tokyo (JP).
- (72) 発明者; および
- (75) 発明者/出願人 (米国についてのみ): 永尾 学 (NAGAO, Manabu) [JP/JP]; 〒1058001 東京都港区芝浦一丁目1番1号 株式会社東芝 知的財産部内 Tokyo (JP).
- (74) 代理人: 酒井 宏明, 外 (SAKAI, Hiroaki et al.); 〒1006020 東京都千代田区霞が関三丁目2番5号 霞が関ビルディング 酒井国際特許事務所 Tokyo (JP).
- (81) 指定国 (表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL,

[続葉有]

(54) Title: FINITE STATE TRANSDUCER DETERMINATION DEVICE, AND FINITE STATE TRANSDUCER DETERMINATION METHOD

(54) 発明の名称: 有限状態トランスデューサ決定化装置及び有限状態トランスデューサ決定化方法

[図20]



(57) Abstract: A non-single-valued NFST can be determined without pretreatment. Provided are a state merging unit which extracts, from among states of the finite state transducer, a plurality of states which transit in response to a same input signal, and merges the states; and a signal determination unit wherein an identification signal corresponding to each transition from each of the plurality of pre-merged states to the next state is applied as an input signal to the next state to perform the determination; wherein the identification signal is different from an input signal assigned to the transition of the finite state transducer.

(57) 要約: 事前処理なしに単値でないNFSTを決定化する。有限状態トランスデューサが有する状態のうち、同一の入力記号により遷移する複数の状態を抽出して併合する状態併合部と、併合前の前記複数の状態のそれぞれから次の状態への遷移毎に対応する識別記号を、前記次の状態への入力記号として付与することにより決定化する記号決定部と、を有し、前記識別記号は、前記有限状態トランスデューサが有する遷移に割り当てられている入力記号とは異なる記号であることを特徴とする。

- 10 SET GENERATION UNIT
21 STATE ACQUISITION UNIT
22 TRANSITION GENERATION CONTROL UNIT
31 STATE MERGING UNIT
32 SIGNAL DETERMINATION UNIT
33 TRANSITION GENERATION CONTROL UNIT
41 STATE MERGING UNIT
42 SIGNAL DETERMINATION UNIT
43 SUB-STATE GENERATION UNIT
44 TRANSITION GENERATION CONTROL UNIT
50 TRANSITION SELECTION UNIT
101 FST READING UNIT
109 FST WRITING UNIT

WO 2010/116554 A1

WO 2010/116554 A1



NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, 添付公開書類:

CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, — 国際調査報告 (条約第 21 条(3))
TD, TG).

明 細 書

発明の名称：

有限状態トランスデューサ決定化装置及び有限状態トランスデューサ決定化方法

技術分野

[0001] 本発明は、有限状態トランスデューサ決定化装置及び有限状態トランスデューサ決定化方法に関する。

背景技術

[0002] 有限状態オートマトン (Finite State Automaton, FSA) の一種である有限状態トランスデューサ (Finite State Transducer, FST) は自然言語処理や音声認識等さまざまな分野で利用されている。形式的にはFSTは6つ組 $(Q, \Sigma, \Delta, E, I, F)$ で表される。ここで、 Q は状態の集合、 Σ は入力記号の集合、 E は遷移の集合、 I は初期状態の集合、 F は終了状態の集合、 Δ は出力記号の集合を表す。FSTにおける遷移 E は $E \subseteq Q \times \Sigma \times \Delta^* \times Q$ である。 Δ^* は Δ のクリーネ閉包 (Kleene closure) であり、空記号列 ε と Δ の組み合わせで作成できる全ての記号列から成る集合である。FSTから出力記号を取り除くとFSAとなる。またFSTは一般順序機械 (generalized sequential machine) と呼ばれることがある。

[0003] FSTが非決定的 (nondeterministic) であるとは、複数の経路に同一の入力記号列が割り当てられていることをいう。形式的には、集合 X のサイズ (要素の数) を $|X|$ と表すこととし、状態 $q \in Q$ から入力記号列 $w \in \Sigma^*$ によって遷移することができる経路の集合を $\Pi(q, w) \subseteq E^*$ とするとき、 $|\Pi(q, w)| > 1$ であるような q が存在することをいう。このようなFSTを非決定性有限状態トランスデューサ (Nondeterministic Finite State Transducer,

NFST) という。一方、FSTが決定的 (d e t e r m i n i s t i c) であるとは、ある入力記号列が与えられたとき取りうる経路が最大でも1つだけであることをいう。形式的には、全ての $q \in Q$, $w \in \Sigma^*$ に対して、 $|\Pi(q, w)| \leq 1$ であることをいう。このようなFSTを決定的有限状態トランスデューサ (D e t e r m i n i s t i c F i n i t e S t a t e T r a n s d u c e r, DFST) という。NFSTをDFSTへと変換することを決定化 (d e t e r m i n i z e) という。この決定化は例えば特許文献1や非特許文献1に記載の方法を用いることで行える。

先行技術文献

特許文献

[0004] 特許文献1：米国特許第6, 456, 971号明細書

非特許文献

[0005] 非特許文献1：M. Mohri, On Some Applications of Finite-State Automata Theory to Natural Language Processing, Natural Language Engineering, 1996, vol. 2, issue 1, Pages 61-80

非特許文献2：Jean Berstel, Transductions and Context-Free Languages, 1979, Chapter IV, Section 6

非特許文献3：C. Allauzen, M. Mohri, An Optimal Pre-determinization Algorithm for Weighted Transducers, Theoretical Computer Science, 2004, Volume 328, Issue 1-2, Implementation and application of automata, Pages 3-18

非特許文献4：M. Mohri, M. Riley, D. Hindle, A. Ljolje, F. C. N. Pereira, Full Expansion

of Context-dependent Networks in Large Vocabulary Speech Recognition, 1998, In Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP '98), Seattle, Washington

非特許文献5: C. Allauzen, M. Mohri, Generalized Optimization Algorithm for Speech Recognition Transducers, 2003, In Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP' 03)

発明の概要

発明が解決しようとする課題

[0006] しかしながら、従来の方法で単値でないNFSTを決定化するには、単値でない原因となる部分を発見するステップと、発見した部分を単値にすることで、単値なNFSTへと変換するステップから構成される事前処理が必要であった（「単値」については後述する）。

[0007] 本発明は、上記の点に鑑みて、これらの問題を解消するために発明されたものであり、事前処理なしに単値でないNFSTを決定化することにある。

課題を解決するための手段

[0008] 上述した課題を解決し、目的を達成するために、本発明は、有限状態トランスデューサが有する遷移に割り当てられている入力記号とは異なる記号である識別記号を生成する記号決定部と、前記有限状態トランスデューサが有する状態のうち、同一の入力記号による遷移先にある一または複数の状態を抽出し、抽出した前記状態をサブ状態として有する状態を生成する状態併合部と、前記状態併合部が生成する状態間の遷移の入力記号として、有限状態トランスデューサが有する遷移に割り当てられている入力記号または前記識

別記号を付与することにより決定化する単値化処理部と、を有することを特徴とする。

- [0009] 上述した課題を解決し、目的を達成するために、本発明は、また、有限状態トランスデューサが有する遷移に割り当てられている入力記号とは異なる記号である識別記号を生成する記号決定ステップと、前記有限状態トランスデューサが有する状態のうち、同一の入力による遷移先にある一または複数の状態を抽出し、抽出した前記状態をサブ状態として有する状態を生成する状態併合ステップと、前記状態併合ステップが生成する状態間の遷移の入力記号として、有限状態トランスデューサが有する遷移に割り当てられている入力記号または前記識別記号を付与することにより決定化する単値化ステップと、を有することを特徴とする。

発明の効果

- [0010] 本発明の有限状態トランスデューサ決定化装置及び有限状態トランスデューサ決定化方法によれば、事前処理なしに単値でないNFSTを決定化することが可能になる。

図面の簡単な説明

- [0011] [図1]図1は、`twins property`を持つFSTの例を示す図である。
- [図2]図2は、決定化後に出力記号が消えない例を説明する図である。
- [図3-1]図3-1は、本実施の形態の有限状態トランスデューサ決定方法の詳細を示す擬似コード（その1）である。
- [図3-2]図3-2は、本実施の形態の有限状態トランスデューサ決定方法の詳細を示す擬似コード（その2）である。
- [図4]図4は、補助入力記号を割り当てる処理の擬似コードである。
- [図5]図5は、例1における決定化前のFST T_1 を示す図である。
- [図6]図6は、例1の決定化後のFSTの初期状態を示す図である。
- [図7]図7は、例1の決定化処理の過程におけるFSTを示す図（その1）である。

[図8]図8は、例1の決定化処理の過程におけるF S Tを示す図（その2）である。

[図9]図9は、例1の決定化処理の過程におけるF S Tを示す図（その3）である。

[図10]図10は、例1の決定化後のF S Tを示す図である。

[図11]図11は、例2における決定化前のF S T T_1 を示す図である。

[図12]図12は、例2の決定化後のF S Tの初期状態を示す図である。

[図13]図13は、例2の決定化処理の過程におけるF S Tを示す図（その1）である。

[図14]図14は、例2の決定化処理の過程におけるF S Tを示す図（その2）である。

[図15]図15は、例2の決定化処理の過程におけるF S Tを示す図（その3）である。

[図16]図16は、例2の決定化後のF S Tを示す図である。

[図17]図17は、終了出力がないF S Tを処理する例の擬似コードである。

[図18]図18は、終了出力がN個あってもよいF S Tを処理する例の擬似コードである。

[図19]図19は、ハードウェアの構成の例を示す図である。

[図20]図20は、本実施の形態に係る有限状態トランスデューサ決定化装置の機能構成の例を示す図である。

[図21]図21は、音声認識におけるF S Tの例を示す図である。

[図22]図22は、翻訳におけるF S Tの例を示す図である。

[図23]図23は、検索における決定化前のF S Tの例を示す図である。

[図24]図24は、検索における決定化後のF S Tの例を示す図である。

[図25]図25は、重みが加えられたF S Tに補助入力記号を付加する処理の擬似コードである。

[図26]図26は、決定化前のW F S Tを示す図である。

[図27]図27は、決定化後のW F S Tの初期状態を示す図である。

[図28] 図 28 は、決定化処理の過程における W F S T を示す図（その 1）である。

[図29] 図 29 は、決定化処理の過程における W F S T を示す図（その 2）である。

[図30] 図 30 は、決定化後の W F S T を示す図である。

[図31] 図 31 は、音声認識における W F S T の例を示す図である。

発明を実施するための形態

[0012] 以下、本実施の形態を図面に基づき説明する。

[0013] 〔実施の形態 1〕

N F S T には単値（single-valued）であるものとそうでないものがある。N F S T が単値であるとは、どのような入力記号列を N F S T に入力しても得られる出力が最大でも 1 つであることをいう。形式的には、経路 $\pi \in \Pi(q, w)$ の出力記号列を $out(\pi)$ とするとき、 $|\{out(\pi) \mid \pi \in \Pi(q, w), q \in Q, w \in \Sigma^*\}| \leq 1$ であることをいう。同様に、単値でないとは $|\{out(\pi) \mid \pi \in \Pi(q, w), q \in Q, w \in \Sigma^*\}| > 1$ であることをいう。

[0014] 単値でない N F S T の決定化を特許文献 1 に記載の方法で行うと、決定化前には存在していた出力記号が消えることがある。非特許文献 1 に記載の方法では決定化が終了しないことがある。このような状況は、F S T が次に説明する *twins property* を持っていないときに発生する（非特許文献 2, 3）。

[0015] 状態 q から入力記号 w により状態 q' へ到達する経路の集合を $\Pi(q, w, q')$ 、経路 π の遷移先の状態を $n(\pi)$ とする。さらに $(\times)$ を記号列の結合演算とし、記号列を x とするとき、 x^{-1} は $x^{-1}(\times) x = \varepsilon$ を満たすような記号列であるとする。例えば $A(\times) B = AB$, $A^{-1}(\times) AB = B$ である。

[0016] （条件）

$$out(\pi_1)^{-1}(\times) out(\pi_2) = out(\pi_1 \pi'_1)^{-1}(\times) out$$

$(\pi_2 \pi'_2)$

を満たすとき、状態 $q_1 = n(\pi_1)$, $q_2 = n(\pi_2)$ は *twins* であるという。ここで、初期状態 $i_1, i_2 \in I$ から入力記号列 $x \in \Sigma^*$ による経路を $\pi_1 \in \Pi(i_1, x)$, $\pi_2 \in \Pi(i_2, x)$ としている。また、入力記号列 $y \in \Sigma^*$ に対して $\pi'_1 \in \Pi(q_1, y, q_1)$, $\pi'_2 \in \Pi(q_2, y, q_2)$ である。経路 π_1 の後ろに経路 π'_1 をつなげた経路を $\pi_1 \pi'_1$ と表現している。全ての q_1, q_2 の組み合わせが *twins* であるとき、その FST は *twins property* を持つという。循環経路 (*cyclic path*) がない FST の場合は $\Pi(q_1, y, q_1)$ や $\Pi(q_2, y, q_2)$ が空集合であるため、そもそも *twins* でない状態が存在しない。したがって、そのような FST も *twins property* を持つ。

[0017] *twins* でない状態を見つけ、補助入力記号 (*Auxiliary input symbol*) を該当箇所に導入することで決定化可能にする方法、言い換えると単値な FST へと変換する方法は従来より提案されている (非特許文献 3)。補助入力記号とは、元の FST には含まれていない入力記号であり、単値な FST にするために新たに導入する入力記号のことである。このような方法により、事前に単値でない FST を単値な FST へと変換することで、単値でない FST を決定化することができる。

[0018] *twins property* を持っていない出力記号が消える場合がある。図 1 のような場合である。ここで、太線の丸は初期状態、二重線の丸は終了状態を表す。「:」の左側が入力記号、右側が出力記号列である。

[0019] このような場合、終了状態から n 個の出る遷移を認めるようにし、そこに出力記号を割り当てれば (図 2)、決定化後も出力記号が消えることはなくなる。このような FST については特許文献 1、非特許文献 2 に記載がある。

[0020] 各遷移に重みが付与された FST は重み付き有限状態トランスデューサ (*Weighted Finite State Transducer, WFST*) と呼ばれる。FST の場合と同様に、WFST についても決定化を

行うことが可能であり、例えば特許文献 1 の方法によって決定化を行うことができる。WFST の場合においても、単値でない WFST が存在する。単値でない WFST を決定化するには、単値な WFST へと変換する必要があるが、この変換方法については非特許文献 3 に記載がある。この方法を用いて、まず単値でない WFST を単値な WFST へと変換し、その後、決定化するという方法を行うことで、単値でない WFST を決定化することができる。

[0021] すなわち、単値でない原因となる部分を発見するステップと、発見した部分を単値にすることで、単値な NFST へと変換するステップから構成される事前処理が必要となる。以下の実施の形態では、この事前処理を行うことなく、単値でない NFST を決定化することについて、説明する。

[0022] 次に、本実施の形態やそれ以降の実施の形態に必要な記号、用語等について説明する。

[0023] 空集合は $\emptyset$ または $\{\}$ で表す。空記号列は ε で表す。空記号列は、空列や空文字列などと呼ばれることがある。遷移 $e \in E$ の遷移元状態を $p(e)$ 、遷移先状態を $n(e)$ 、入力記号を $i_n(e)$ 、出力記号列を $out(e)$ とする。同様に、経路 $\pi \in E^*$ の始端状態を $p(\pi)$ 、終端状態を $n(\pi)$ 、入力記号列を $i_n(\pi)$ 、出力記号列を $out(\pi)$ とする。経路 π 上にある遷移の総数を $|\pi|$ とする。 π 上の j 個目の遷移先の状態を $(\pi)_j$ とする。 $p(\pi) = (\pi)_0$ 、 $n(\pi) = (\pi)_{|\pi|}$ である。補助入力記号は $\#_n$ (n は整数) で表すこととする。複数の補助入力記号を区別できればよいので、ここでは $\#_n$ の n は整数であるとする。

[0024] プログラム等で集合を表現するときは、集合の要素を格納でき、集合内に重複した要素が発生しなければどのように表現してもよく、例えば、配列、リンクリスト、2 分木、ハッシュテーブルなどが挙げられる。

[0025] 本実施の形態での方法は、`twins property` を持たない FST 全てについて決定化できるというわけではない。本実施の形態での方法によって決定化でできる NFST は次の条件を満たすもののみである。

- [0026] 初期状態から状態 q_1 , q_2 までの経路をそれぞれ π_1 , π_2 とし, 状態 q_1 から q_1 までの経路を π'_1 , 状態 q_2 から q_2 までの経路を π'_2 とし, さらに $\text{in}(\pi_1 \pi'_1) = \text{in}(\pi_2 \pi'_2)$ であるとする。このような状態 q_1 と q_2 の全てに対して twins であるか, 次の条件を満たすとき, 本実施の形態の方法によって単値へと変換しながら決定化することができる。
- [0027] その条件とは, $(\pi_1)_j = (\pi_2)_j$ であるような j が存在し, 状態 $(\pi_1)_j$ と $(\pi_2)_j$ までの π_1 と π_2 上にある遷移の出力記号列を全て ε とすると, 状態 q_1 と q_2 が twins となることである。
- [0028] 以上の条件を満たさない NFST は単値にできないため, 本実施の形態で述べる方法による決定化処理が完了することはない。
- [0029] 本実施の形態の方法の詳細は図 3-1 及び図 3-2 (以下、「図 3」という。) の擬似コードの通りである。なお、図 3-2 に示す擬似コードは、図 3-1 に続く擬似コードであり、両者を併せて 1 つの関数 determine となる。この擬似コードは、 FST に初期出力と終了出力を加えた FST の決定化を行うことができる。ただし、終了出力は出力記号列 1 つのみとしている。初期出力がない FST , 終了出力が複数ある FST , 終了出力がない FST に対する処理方法は後に説明する。この擬似コードによって決定化を行う対象の FST を $T_1 = (Q_1, \Sigma_1, \Delta, E_1, I_1, F_1, \lambda_1, \rho_1)$, 決定化後の FST を $T_2 = (Q_2, \Sigma_2, \Delta, E_2, i_2, F_2, \lambda_2, \rho_2)$ とする。 Q_1, Q_2 は状態の集合, Σ_1, Σ_2 は入力記号の集合, Δ は出力記号の集合, E_1, E_2 は遷移の集合, I_1 は初期状態の集合, i_2 は初期状態, λ_1 は初期出力関数, λ_2 は初期出力, ρ_1, ρ_2 は終了出力関数を表す。初期状態 $i \in I$ に対し, $\lambda_1(i) \in \Delta^*$ であり, $\lambda_2 \in \Delta^*$ である。終了状態 $q \in F_1$ に対して $\rho_1(q) \in \Delta^*$ であり, 同様に $q \in F_2$ に, 対して $\rho_2(q) \in \Delta^*$ である。
- [0030] 出力記号の演算は一般化のため $(+)$ と $(\times)$ で表す。図中では $(+)$ は $+$ を丸で囲った記号, $(\times)$ は $\times$ を丸で囲った記号で表す。本実施の形態では, $(+)$ は前方最長一致 ($\text{longest common prefix}$

）演算を、 $(\times)$ は結合演算を表す。例えば $a b c (+) a d c = a$, $a (\times) b = a b$ となる。さらに特別な値である $1\#$ を導入する。本実施の形態では、 $1\#$ を空記号列 ε として扱う。言い換えると、本実施の形態に限っては図 3 の $1\#$ を ε と置き換えても意味は同じであるということである。記号列を x とするとき、 x^{-1} は $x^{-1} (\times) x = 1\# = \varepsilon$ を満たすような記号列であるとする。例えば、 $a^{-1} (\times) a b = b$ となる。 $a u x (q, v)$ は出力記号 $v \in \Delta^*$ と状態 $q \in Q_2$ に対応する補助入力記号を表している。 $a u x (q, v)$ の詳細については後述する。

[0031] T_2 の状態の集合 Q_2 の要素である状態を q_2 とするとき、 $(T_1$ の状態、入力記号または出力記号列) の 2 つ組の集合で q_2 は表される。形式的には、 $q_2 \subseteq \{ (q_1, w) \mid q_1 \in Q_1, w \in (\Sigma_1 \cup \Delta^*) \}$ である。コンピュータ上で実際に動作させる際には、図 3 の擬似コードと同じく、 T_2 の各状態を区別するために 2 つ組の集合を用いてもよいし、 T_2 の各状態に番号を割り当てておき、さらにその番号と 2 つ組の集合とを相互に関連付けて記憶するようにしておいてもよい。また、決定化が完了すれば、2 つ組の集合は不要なので、各状態には区別するための情報を含めておくだけでよい。例えば、 T_2 の各状態に互いに異なる番号を割り当てればよい。

[0032] 図 3 で使われている $\gamma (q_2, \alpha)$ と $\xi (q_2, \alpha, q_n)$ は

$$\gamma (q_2, \alpha) = \{ n(e) \mid e \in E_1, (q, \cdot) \in q_2, p(e) = q, i n(e) = \alpha \}$$

$$\xi (q_2, \alpha, q_n) = \{ (w, e) \mid (q, w) \in q_2, e \in E_1, p(e) = q, n(e) = q_n, i n(e) = \alpha \}$$

のように定義される。 $\gamma (q_2, \alpha)$ は決定化後の状態 q_2 の要素である決定化前の状態のそれぞれからの入力記号 α による遷移先の状態の集合を表している。 $\xi (q_2, \alpha, q_n)$ は、決定化後の状態 q_2 の要素である決定化前の状態が遷移元状態であり、決定化前の状態 $q_n \in Q_1$ が遷移先であり、さらに入力記号 α が割り当てられている遷移の集合を表している。

[0033] 図 3 の擬似コードを順に説明する。なお、本実施の形態では、擬似コード

の 1 行に対応する処理毎に、一のステップが実行される。1 行目では FST T_2 の状態の集合 Q_2 、遷移の集合 E_2 、終了状態の集合 F_2 を空集合にすることで初期化する。入力記号の集合 Σ_2 は FST T_1 の入力記号の集合 Σ_1 を代入する。

[0034] 2 行目では FST T_1 の初期状態に割り当てられている出力記号列全てについて最長前方一致演算を行った結果を FST T_2 の初期出力記号列 λ_2 へと代入する。

[0035] 3 行目では T_2 の初期状態 i_2 を作成する。 T_1 の初期状態と出力記号列の 2 つ組を要素とする集合として初期状態は表される。 i_2 に含まれる要素を (i, w) とするとき、 w は T_1 の初期状態 i に対応する初期出力記号列 $\lambda_1(i)$ の前方から順に λ_2 と同じ出力記号列を取り除いたものとなる。例えば、 $\lambda_1(i)$ が abc で、 λ_2 が ab なら $w = (ab)^{-1}(\times) abc = c$ となる。

[0036] 4 行目ではスタック S に T_2 の初期状態を設定する。この時点では S に含まれる要素は i_2 のみである。なお、スタック S は値を 1 つずつ追加・取り出しができることができればどのようなものでもかまわないので、スタックではなく例えばキューとして実装してもよい。さらに Q_2 に i_2 を追加する。

[0037] 5 行目ではスタック S が空か否かを判断する。空なら決定化が完了したことを意味するので、処理を終了する。空でない場合は 6 行目以降を処理していく。

6 行目ではスタック S から要素を 1 つ取り出し、 q_2 に代入する。取り出した要素はスタック S から除去される。

[0038] 7 行目では q_2 に含まれる要素である 2 つ組（状態、記号列）のうち、少なくとも 1 つの記号列が Σ_1 に含まれていれば 8 行目に処理を進める。そうでなければ 14 行目に処理を進める。

[0039] 8 行目では q_2 に含まれている要素である 2 つ組を (q, w) とする。なお、必ず $|q_2| = 1$ であるので q, w は一意に決まる。

[0040] 9 行目では 10 行目で作成する遷移の遷移先状態 q_{n2} を作成する。状態 q_n

$_2$ は状態 q と出力記号列 $1^\#$ から成る 2 つ組を要素とする集合となる。

[0041] 1 0 行目では遷移元状態が q_2 , 入力記号が w , 出力記号が $1^\#$, 遷移先状態が q_{n_2} である遷移を E_2 へ追加する。

[0042] 1 1 行目では q_{n_2} が既に Q_2 に含まれているか否かを判断し, もし含まれていなければ, 1 2 行目へと処理を進める。1 2 行目ではスタック S に q_{n_2} を追加し, さらに Q_2 にも q_{n_2} を追加する。

[0043] 1 4 行目から 2 3 行目では T_2 の終了状態と終了出力を作成する処理である。

1 4 行目では q_2 の要素である 2 つ組に含まれている状態 $q \in Q_1$ のうち終了状態 F_1 に含まれているものが少なくとも 1 つあるか否かを判断する。少なくとも 1 つあれば 1 5 行目の処理に進む。

[0044] 1 5 行目では q_2 に含まれる 2 つ組 (q, w) のうち, $q \in F_1$ であるような 2 つ組 (q, w) の全てに対して, $w(x) \rho_1(q)$ を計算し, その計算で得られた出力記号列すべてを含む集合を W とする。

[0045] 1 6 行目では出力記号列の集合 W に含まれる出力記号列全てに対して $(+)$ による演算を行い, その結果を w_2 に代入する。例えば $W = \{x_1, x_2, x_3\}$ であれば, $x_1(+)x_2(+)x_3$ という計算をすることになる。

[0046] 1 7 行目では 1 6 行目で計算した w_2 と同じ出力記号列が少なくとも 1 つ W に含まれていれば 1 8 行目に処理を進める。そうでなければ 1 9 行目に処理を進める。

[0047] 1 8 行目では w_2 を q_2 の終了出力として $\rho_2(q_2)$ に代入し, さらに終了状態として F_2 に q_2 を追加する。

[0048] 1 9 行目は W に含まれる出力記号列 w のうち, w_2 と異なるものについて, 2 0 行目から 2 3 行目を処理することを表している。

2 0 行目では状態 q_{n_2} を空集合にする。

2 1 行目では遷移元状態を q_2 , 入力記号を (q_{n_2}, w) , 出力記号列を w , 遷移先状態を q_{n_2} とするような遷移を E_2 へ追加する。また, 追加した遷移の入力記号を Σ_2 に追加する。

- [0049] 22行目では T_2 の状態の集合 Q_2 に q_{n2} が含まれているか否かを判断する。
含まれていなければ23行目へと処理を進める。
23行目では q_{n2} の終了出力 $\rho_2(q_{n2})$ を $1^\#$ とし、 T_2 の状態の集合 Q_2 と終了状態の集合 $F_2 \rightarrow q_{n2}$ を追加する。
- [0050] 24行目は、 Σ_1 に含まれる入力記号 α についてそれぞれ25行目から37行目の処理を行うことを表している。2つ組(T_1 の状態, 出力記号列)として q_2 に含まれている T_1 の状態が遷移元状態である遷移に割り当てられている全ての入力記号について25行目から37行目までを処理できればよいので、それらに含まれていない入力記号については25行目から37行目までを処理する必要はない。ただし、処理したとしても26行目や32行目の条件を満たすことがないので、結果は処理しなかった場合と同じである。
- [0051] 25行目では $\gamma(q_2, \alpha)$ の各要素 q_n に対して次の処理を行い、 q_n と得られた結果で2つ組(q_n , 出力記号列の集合)を作り、全ての q_n に対応する2つ組(q_n , 出力記号列の集合)の集合を θ に代入する。その処理とは、 w を出力記号列、 e を遷移とするとき、 $\xi(q_2, \alpha, q_n)$ の各要素(w, e)に対して $w(x) \circ ut(e)$ を計算した結果を要素とする集合を作成する処理である。
- [0052] 26行目では25行目で作成した θ の要素である2つ組(q_n, W)のうち、 W のサイズが2以上である要素について27行目から31行目の処理を行う。
27行目では2つ組(q_n, α)を q_{n2} へ代入する。
28行目は、 W に含まれる要素 w についてそれぞれ29行目を処理すること
を表している。
- [0053] 29行目では遷移元状態を q_2 、入力記号を $aux(q_{n2}, w)$ 、出力記号列を w 、遷移先状態を q_{n2} とする遷移を E_2 へ追加する。さらに、追加した遷移の入力記号を Σ_2 へ追加する。
30行目では q_{n2} が Q_2 に含まれているか否かを判断し、含まれていなければ31行目へと処理を進める。

- [0054] 3 1 行目ではスタック S に q_{n_2} を追加し、さらに Q_2 に q_{n_2} を追加する。
 3 2 行目では θ の要素である 2 つ組 (q_n, W) のうち、 W のサイズが 1 の 2 つ組が少なくとも 1 つあれば 3 3 行目から 3 7 行目の処理を行う。
- [0055] 3 3 行目では θ の要素である 2 つ組 (q_n, W) のうち、 W のサイズが 1 であるもの全てについて、出力記号列 $w \in W$ を $(+)$ で計算し、得られた結果を w_2 へ代入する。例えば $r_1, r_2, r_3 \in Q_1$ として $\theta = \{ (r_1, \{a b\}), (r_2, \{a c\}), (r_3, \{b, c d\}) \}$ であるとするとき、 W のサイズが 1 である 2 つ組は $(r_1, \{a b\}), (r_2, \{a c\})$ であり、これらの出力記号列を $(+)$ で計算すると、 $a b (+) a c = a$ となるので、 $w_2 = a$ となる。
- [0056] 3 4 行目は 3 3 行目と同様に θ の W のサイズが 1 である 2 つ組 (q_n, W) に対し、 q_n と W の要素 w の前方から w_2 に一致する部分を除いた出力記号列とから成る 2 つ組を作成し、それらの集合を q_{n_2} へ代入する。3 3 行目の例を用いて説明すると、3 3 行目と同様に W のサイズが 1 である 2 つ組は $(r_1, \{a b\}), (r_2, \{a c\})$ であるので、 r_1 に対しては $a^{-1} (x) a b = b$ 、 r_2 に対しては $a^{-1} (x) a c = c$ となる。その結果、この例の場合は $q_{n_2} = \{ (r_1, b), (r_2, c) \}$ となる。
- [0057] 3 5 行目では遷移元状態を q_2 、入力記号を α 、出力記号を w_2 、遷移先状態を q_{n_2} とする遷移を E_2 へ追加する。
 3 6 行目では q_{n_2} が Q_2 に含まれているか否かを判断し、含まれていなければ 3 7 行目の処理を行う。
- [0058] 3 7 行目ではスタック S に q_{n_2} を追加し、さらに Q_2 に q_{n_2} を追加する。
 決定化が完了すると、各状態に関連付けられている 2 つ組 (状態, 記号列) の集合を記憶しておく必要はなくなるので、各状態に互いに異なる番号を割り当て直すという処理を行うこともできる。また、決定化中に随時番号を割り当てておき、その番号に 2 つ組 (状態, 記号列) の集合を関連付けておくようにすることもできる。
- [0059] 決定化後の状態 $q \in Q_2$ と出力記号列 $v \in \Delta^*$ を引数として取り、戻り値と

して補助入力記号を返すアルゴリズム aux は、 q と v とから成る 2 つ組のそれぞれに対して、互いに異なる補助入力記号を割り当てることができればよい。

補助入力記号を y として、3 つ組 (q, v, y) を要素とする集合を Y とし、決定化を行う前に空集合に初期化されているものとする。すると、アルゴリズム aux は図 4 のような手順によって実現できる。

[0060] ここで、 $make_aux$ は引数として番号 n が渡されると、補助入力記号 $\#_n$ を返すという処理を行う。

[0061] 図 4 の擬似コードの動作例を次に示す。説明を簡単にするため、決定化後の FST の状態は単に番号で表すものとする。 Y が空集合のときに $aux(1, B)$ が呼び出されると、 Y が空集合なので 1 行目の条件が満たされることはない。したがって、3 行目へと処理を進める。3 行目では、 $make_aux(1)$ が呼び出される。 $|Y| = 0$ であるからである。すると、補助入力記号 $\#_1$ が得られるので y に代入し、4 行目で Y に $(1, B, \#_1)$ を追加し、5 行目で得られた $\#_1$ を返す。

[0062] 再度 $aux(1, B)$ が呼び出されると、 $Y = \{(1, B, \#_1)\}$ であることから 1 行目の条件を満たすので、2 行目で、条件を満たした Y の要素の 3 番目の要素 y_1 である $\#_1$ を結果として返す。

[0063] 次に $aux(2, B)$ が呼び出されると、1 行目の条件を満たさない。 $q_1 = 2$ 、 $v_1 = B$ であるような要素は Y には今は含まれていないからである。そこで、3 行目で新しい補助入力記号を作成する。今、 $|Y| = 1$ であるので、 $make_aux(2)$ を実行することになり、 $\#_2$ が得られ、 y に代入する。これを使って Y に $(2, B, \#_2)$ を追加すると、 $Y = \{(1, B, \#_1), (2, B, \#_2)\}$ となる。そして、 $\#_2$ を結果として返す。

[0064] 以上のようにしていくことで、状態 q と出力記号列 v の組み合わせに対して、それぞれ互いに異なる補助入力記号を割り当てていくことができる。また、必要な補助入力記号のみを作成することができる。

[0065] なお、決定化後の状態から出て行く遷移の作成が状態ごとに完了するたび

に、 Y を空集合に初期化することもできる。具体的には図 3 の 6 行目から 3 7 行目の処理が終わるたびに Y を空集合にすればよい。このようにすると、生成される補助入力記号の種類を少なく抑えることができる。処理する決定化後の状態ごとに、必要に応じて $\#_1$ から順に生成されるためである。その代わり、状態 q と出力記号列 v が同じであっても異なる補助入力記号が割り当てられることがある。

[0066] (例 1)

本実施の形態における決定化方法がどのように動作するのかについて、図 5 の FST を決定前の FST T_1 として説明する。図 3 と同様に決定化後の FST は T_2 とする。なおこの例ではスタック S はスタックとして、つまり先入れ後出し方式で動作するものとする。図 5 より $Q_1 = \{0, 1, 2, 3, 4, 5, 6\}$, $I_1 = \{0\}$, $F_1 = \{6\}$, $\Sigma_1 = \{a, b, c, d\}$, $\Delta_1 = \{A, B, C\}$ である。太線の円が初期状態であり、二重線の円が終了状態を表している。各遷移のラベルのうち、「:」の左側が入力記号、右側が出力記号列を表している。

[0067] 図 3 の 1 行目でまず決定化後の FST の初期化が行われる。

次に 2 行目で初期出力を計算する。この例の FST には初期出力はない。言い換えると $\lambda_1(0) = 1^\# = \varepsilon$, つまり空記号列を出力することである。したがって、 $\lambda_2 = \varepsilon$ となる。

[0068] 3 行目では決定化後の FST の初期状態を作成する。 $i_2 = \{(0, \varepsilon)\}$ となる (図 6)。

[0069] 4 行目では 3 行目で作成した状態 i_2 をスタック S と Q_2 に追加し、その結果、 $S = \{\{(0, \varepsilon)\}\}$, $Q_2 = \{\{(0, \varepsilon)\}\}$ となる。

[0070] $|S| = 1$ であるので、5 行目の条件を満たし、6 行目に処理を進める。6 行目でスタック S から値を取り出し、 q_2 に代入すると、 $S = \phi$, $q_2 = \{(0, \varepsilon)\}$ となる。

[0071] 7 行目では、判断しなければならない w は $w = \varepsilon$ の場合のみである。 w は Σ_1 には含まれていないので、7 行目の条件は満たされない。したがって、1

4 行目に処理を進める。

1 4 行目では、 q_2 に含まれている 2 つ組（状態、記号列）のうちの状態が F_1 に含まれているか否かを判断する。いま、 q_2 に含まれる状態は状態 0 のみであり、これは F_1 には含まれていない。したがって、処理を 2 4 行目に進める。

[0072] $\Sigma_1 = \{a, b, c\}$ ではあるが、状態 0 から出て行く遷移の入力記号は a のみであるので、 $\alpha = a$ の場合のみを処理すればよい。したがって、2 4 行目では $\alpha = a$ として 2 5 行目以降の処理を行う。

[0073] 2 5 行目では、 $\theta = \{ (1, \{\varepsilon(x) A\}), (2, \{\varepsilon(x) B\}), (3, \{\varepsilon(x) C\}) \} = \{ (1, \{A\}), (2, \{B\}), (3, \{C\}) \}$ となる。

2 6 行目の $|W| > 1$ の条件を満たす θ の要素はないので、3 2 行目へと処理を進める。

$|W| = 1$ を満たす θ の要素は存在するので 3 2 行目の条件を満たし、3 3 行目へと処理を進める。

[0074] 3 3 行目の処理で、 $w_2 = A (+) B (+) C = \varepsilon = 1^\#$ となる。

3 4 行目の処理で、 $q_{n_2} = \{ (1, A), (2, B), (3, C) \}$ が得られる。 $x^{-1}(x)x = \varepsilon$ であるので、 $\varepsilon^{-1} = \varepsilon$ である。 $x = \varepsilon$ とすると $\varepsilon^{-1}(x)\varepsilon = \varepsilon$ なので両辺に右側から ε を (x) 演算でつなげると $\varepsilon^{-1}(x)\varepsilon(x)\varepsilon = \varepsilon(x)\varepsilon$ となり、 $\varepsilon^{-1}(x)\varepsilon = \varepsilon(x)\varepsilon$ となり、両辺の右側の ε を取り除くと $\varepsilon^{-1} = \varepsilon$ と変形できるからである。したがって、例えば、 q_{n_2} の最初の要素である $(1, A)$ は $w_2(x)A = \varepsilon^{-1}(x)A = \varepsilon(x)A = A$ のような計算から得られる。

[0075] 3 5 行目で状態元状態が $\{ (0, \varepsilon) \}$ で入力記号が a 、出力記号が ε 、遷移先状態が $\{ (1, A), (2, B), (3, C) \}$ である遷移を作成し E_2 に加える。その結果、図 7 のうち初期状態と状態 $S_1 O_1$ とそれらの間の遷移とから構成される FST になる。

[0076] 今 q_{n_2} は Q_2 には含まれていないので 3 6 行目を満たし、3 7 行目が処理さ

れる。

37行目では、 q_{n2} をSと Q_2 に追加するので、 $S = \{ \{ (1, A), (2, B), (3, C) \} \}$ 、 $q_2 = \{ \{ (0, \varepsilon) \}, \{ (1, A), (2, B), (3, C) \} \}$ となる。

[0077] 5行目に処理が戻り、スタックSは空集合ではないので、6行目に処理を進む。

6行目で、 $q_2 = \{ (1, A), (2, B), (3, C) \}$ となり、 $S = \phi$ となる。

AもBもCも Σ_1 には含まれていないので、7行目の条件は満たされず、14行目に処理が進む。

[0078] 状態1, 2, 3は F_1 には含まれていない、つまり終了状態ではないので、14行目の条件は満たさない。したがって24行目に処理が進む。

[0079] 状態1, 2, 3から出て行く遷移の入力記号はbのみであるので、25行目から37行目の処理は $\alpha = b$ の場合の処理のみを行えばよい。

25行目では、 $\theta = \{ (4, \{ A(x)\varepsilon, B(x)\varepsilon \}), (5, \{ C(x)\varepsilon \}) \} = \{ (4, \{ A, B \}), (5, \{ C \}) \}$ となる。

[0080] 26行目では θ の要素のうち、 $(4, \{ A, B \})$ が $|W| > 1$ の条件を満たすので、これについて27行目から31行目の処理を行う。

27行目で、 $q_{n2} = \{ (q_n, \alpha) \} = \{ (4, b) \}$ となり、28行目へと処理が進む。

[0081] いま、 $W = \{ A, B \}$ であるので、まず、 $w = A$ について29行目の処理を行う。ここで $aux(q_{n2}, w) = aux(\{ (4, b) \}, A) = \#_1$ とする。そうすると、29行目では、遷移元状態が $\{ (1, A), (2, B), (3, C) \}$ 、入力記号が $\#_1$ 、出力記号列がA、遷移先状態が $\{ (4, b) \}$ である遷移が E_2 に追加され、さらに Σ_2 に $\#_1$ が追加される。同様に $w = B$ についても処理を行うが、このとき $aux(\{ (4, b) \}, B) = \#_2$ とすると、遷移元状態が $\{ (1, A), (2, B), (3, C) \}$ 、入力記号が $\#_2$ 、出力記号列がB、遷移先状態が $\{ (4, b) \}$ である遷移が E_2 に追

加され、さらに Σ_2 に $\#_2$ が追加される。その結果、図7のうち、初期状態と状態S 1 0 1と状態S 1 0 2とそれらの状態間の遷移とから構成されるF S Tになる。

[0082] 3 0 行目で、 Q_2 に $\{(4, b)\}$ が含まれていないと判定され、3 1 行目でSと Q_2 に $\{(4, b)\}$ が追加される。その結果、 $S = \{ \{(4, b)\} \}$ 、 $Q_2 = \{ \{(0, \varepsilon)\}, \{(1, A), (2, B), (3, C)\}, \{(4, b)\} \}$ となる。

[0083] 3 2 行目の処理に進むが、 $|W| = 1$ を満たす要素が θ に含まれているので、3 2 行目の条件を満たし、3 3 行目以降の処理が行われる。ここで、条件を満たすのは、 $(5, \{C\})$ である。

[0084] 3 3 行目で、 $w_2 = C$ となる。

3 4 行目で、 $q_{n_2} = \{(5, C^{-1}(x)C)\} = \{(5, \varepsilon)\}$ となる。

3 5 行目で、遷移元状態が $\{(1, A), (2, B), (3, C)\}$ 、入力記号がb、出力記号列がC、遷移先状態が $\{(5, \varepsilon)\}$ である遷移が追加される。その結果、図7のうち、初期状態と状態S 1 0 1と状態S 1 0 2と状態S 1 0 3とそれらの状態間の遷移とから構成されるF S Tになる。

[0085] $q_{n_2} = \{(5, \varepsilon)\}$ なので3 6 行目の条件を満たし、3 7 行目に処理が進む。

3 7 行目で、Sと Q_2 に $\{(5, \varepsilon)\}$ が追加される。その結果、 $S = \{ \{(4, b)\}, \{(5, \varepsilon)\} \}$ 、 $Q_2 = \{ \{(0, \varepsilon)\}, \{(1, A), (2, B), (3, C)\}, \{(4, b)\}, \{(5, \varepsilon)\} \}$ となる。

[0086] 5 行目に処理が戻るが、Sは空集合ではないので6 行目に処理を進める。この例ではSはスタックであるとしているので、6 行目で $q_2 = \{(5, \varepsilon)\}$ となる。また、 $S = \{ \{(4, b)\} \}$ となる。

[0087] q_2 に含まれる記号列は ε のみであるので、7 行目の条件は満たさない。そのため、1 4 行目に処理が進む。

[0088] 状態5は終了状態ではないので、1 4 行目の条件は満たさず、2 4 行目の処理へ進む。

25行目以降で処理すべき α は c だけなので、 $\alpha = c$ とする。

25行目で、 $\theta = \{ (6, \{ \varepsilon (x) \varepsilon \}) \} = \{ (6, \{ \varepsilon \}) \}$ となる。

26行目を満たす θ の要素はないので、32行目へと処理を進める。

θ の要素は1つであり、その要素に関して $W = \{ \varepsilon \}$ であるので、 $|W| = 1$ となり、32行目の条件を満たす。したがって、33行目以降の処理に進む。

[0089] 33行目で、 $w_2 = \varepsilon$ になる。

34行目で、 $q_{n2} = \{ (6, \varepsilon) \}$ となる。

[0090] 35行目で、遷移元状態が $\{ (5, \varepsilon) \}$ 、入力記号が c 、出力記号が ε 、遷移先状態が $\{ (6, \varepsilon) \}$ の遷移が E_2 に追加される。その結果、状態 S_{104} を含む図7のようなFSTになる。

$q_{n2} = \{ (6, \varepsilon) \}$ なので36行目の条件を満たし、37行目に処理が進む。37行目で、 S と Q_2 に $\{ (6, \varepsilon) \}$ が追加される。その結果、 $S = \{ \{ (4, b) \}, \{ (6, \varepsilon) \} \}$ 、 $Q_2 = \{ \{ (0, \varepsilon) \}, \{ (1, A), (2, B), (3, C) \}, \{ (4, b) \}, \{ (5, \varepsilon) \} \}$ 、 $\{ (6, \varepsilon) \}$ となる。

[0091] 5行目に処理が戻るが、 S は空集合ではないので6行目に処理を進める。

6行目で $q_2 = \{ (6, \varepsilon) \}$ となる。また、 $S = \{ \{ (4, b) \} \}$ となる。

q_2 に含まれる記号列は ε のみであるので、7行目の条件は満たさない。そのため、14行目に処理が進む。

[0092] 状態6は終了状態なので、14行目の条件を満たし、15行目の処理へ進む。

$\rho_1(6) = \varepsilon$ であるので、15行目で $W = \{ \varepsilon (x) \varepsilon \} = \{ \varepsilon \}$ となる。

16行目で、 $w_2 = \varepsilon$ となる。

W の要素は1つで、それは ε であるので、17行目の条件を満たし、18行目へ処理を進める。

[0093] 18行目で、 $\rho_2(\{(6, \varepsilon)\}) = \varepsilon$ 、 $F_2 = \{\{(6, \varepsilon)\}\}$ となる。その結果、図8のようになる。

19行目を満たす W の要素はないので、20行目から23行目は処理されず、24行目の処理へ進む。

[0094] 25行目以降で処理すべき α は d だけなので、 $\alpha = d$ とする。

25行目で、 $\theta = \{(0, \{\varepsilon(\times)\varepsilon\})\} = \{(0, \{\varepsilon\})\}$ となる。

26行目を満たす θ の要素はないので、32行目へと処理を進める。

θ の要素は1つであり、その要素に関して $W = \{\varepsilon\}$ であるので、 $|W| = 1$ となり、32行目の条件を満たす。したがって、33行目以降の処理に進む。

[0095] 33行目で、 $w_2 = \varepsilon$ になる。

34行目で、 $q_{n2} = \{(0, \varepsilon)\}$ となる。

35行目で、遷移元状態が $\{(6, \varepsilon)\}$ 、入力記号が d 、出力記号が ε 、遷移先状態が $\{(0, \varepsilon)\}$ の遷移が E_2 に追加される。その結果、図9のうち、状態 $S105$ とその状態への遷移とその状態からの遷移を除いた部分から構成される FST になる。

$q_{n2} = \{(0, \varepsilon)\}$ なので36行目の条件を満たさないため、37行目は処理されない。

[0096] 5行目に処理が戻るが、 S は空ではないので6行目に処理を進める。

6行目で $q_2 = \{(4, b)\}$ となる。また、 $S = \phi$ となる。

q_2 に含まれる記号列は b であるので、7行目の条件を満たし、8行目へ処理を進める。

8行目で、 $q = 4$ 、 $w = b$ となる。

9行目で、 $q_{n2} = \{(4, 1^\#)\} = \{(4, \varepsilon)\}$ となる。

[0097] 10行目で、遷移元状態が $\{(4, b)\}$ 、入力記号が b 、出力記号が ε 、遷移先状態が $\{(4, \varepsilon)\}$ の遷移が E_2 に追加される。その結果、図9のうち、状態 $S105$ から出て行く遷移を除いた部分から構成される FST に

なる。

$q_{n2} = \{ (4, \varepsilon) \}$ であるので、11行目の条件を満たし、12行目へ処理を進める。

12行目で、 S と Q_2 に $\{ (4, \varepsilon) \}$ が追加される。その結果、 $S = \{ \{ (4, \varepsilon) \} \}$ 、 $Q_2 = \{ \{ (0, \varepsilon) \}, \{ (1, A), (2, B), (3, C) \}, \{ (4, b) \}, \{ (5, \varepsilon) \} \}$ 、 $\{ (6, \varepsilon) \}$ 、 $\{ (4, \varepsilon) \}$ となる。

[0098] 5行目に処理が戻るが、 S は空ではないので6行目に処理を進める。

6行目で $q_2 = \{ (4, \varepsilon) \}$ となる。また、 $S = \phi$ となる。

q_2 に含まれる記号列は ε のみであるので、7行目の条件は満たさない。そのため、14行目に処理が進む。

[0099] 状態4は終了状態ではないので、14行目の条件は満たさず、24行目の処理へ進む。

25行目以降で処理すべき α は c だけなので、 $\alpha = c$ とする。

25行目で、 $\theta = \{ (6, \{ \varepsilon (\times) \varepsilon \}) \} = \{ (6, \{ \varepsilon \}) \}$ となる。

26行目を満たす θ の要素はないので、32行目へと処理を進める。

[0100] θ の要素は1つであり、その要素に関して $W = \{ \varepsilon \}$ であるので、 $|W| = 1$ となり、32行目の条件を満たす。したがって、33行目以降の処理に進む。

33行目で、 $w_2 = \varepsilon$ になる。

34行目で、 $q_{n2} = \{ (6, \varepsilon) \}$ となる。

35行目で、遷移元状態が $\{ (4, \varepsilon) \}$ 、入力記号が c 、出力記号が ε 、遷移先状態が $\{ (6, \varepsilon) \}$ の遷移が E_2 に追加される。その結果、図9のようになる。

[0101] $q_{n2} = \{ (6, \varepsilon) \}$ なので36行目の条件を満たさず、37行目は処理されない。

5行目に処理が戻るが、いま S は空集合なので、5行目の条件を満たさない

。したがって、決定化処理が完了したことになる。

各状態に関連付けられている2つ組の集合を番号に置き換えて、図10のよ
うにすることもできる。

[0102] (例2)

初期状態や終了状態が複数ある場合に本実施の形態における決定化方法が
どのように動作するのかについて、図11のFSTを決定前のFST T_1 と
して説明する。

図3と同様に決定化後のFSTは T_2 とする。なおこの例でもスタックSはス
タックとして、つまり先入れ後出し方式で動作するものとする。図11より
 $Q_1 = \{0, 1, 2, 3, 4, 5\}$, $I = \{0, 1\}$, $F = \{3, 4, 5\}$,
 $\Sigma_1 = \{a, b\}$, $\Delta_1 = \{A, B, C, D\}$, $\lambda_1(0) = A$, $\lambda_1(1) =$
 AB , $\rho_1(3) = D$, $\rho_1(4) = A$, $\rho_1(5) = \varepsilon$ である。初期状態と終
了状態に記載されている／の左側が状態番号で、／の右側がそれぞれ初期出
力記号と終了出力記号である。

[0103] 図3の1行目でまず決定化後のFSTの初期化が行われる。

次に2行目で初期出力を計算する。 $\lambda_2 = \lambda_1(0) (+) \lambda_1(1) = A (+)$
 $AB = A$ となる。

3行目では決定化後のFSTの初期状態を作成する。 $i_2 = \{(0, A^{-1}(\times)$
 $)A), (1, A^{-1}(\times)AB)\} = \{(0, \varepsilon), (1, B)\}$ となる(図12)。

[0104] 4行目では3行目で作成した状態 i_2 をスタックSと Q_2 に追加し、その結
果、 $S = \{ \{(0, \varepsilon), (1, B)\} \}$, $Q_2 = \{ \{(0, \varepsilon), (1, B)$
 $\} \}$ となる。

$|S| = 1$ であるので、5行目の条件を満たし、6行目に処理を進める。

6行目でスタックSから値を取り出し、 q_2 に代入すると、 $S = \phi$, $q_2 = \{$
 $(0, \varepsilon), (1, B)\}$ となる。

[0105] 7行目では、判断しなければならないwに該当する記号列は ε とBである
。どちらも Σ_1 には含まれていないので、7行目の条件は満たされない。した

がって、14行目に処理を進める。

[0106] 14行目では、 q_2 に含まれている2つ組（状態、記号列）のうちの状態が F_1 に含まれているか否かを判断する。いま、 q_2 に含まれる状態は状態0と状態1であるが、これらは F_1 には含まれていない。したがって、処理を24行目に進める。

$\Sigma_1 = \{a, b\}$ ではあるが、状態0と状態1から出て行く遷移の入力記号は a のみであるので、 $\alpha = a$ の場合のみを処理すればよい。したがって、24行目では $\alpha = a$ として25行目以降の処理を行う。

[0107] 25行目では、 $\theta = \{ (2, \{ \varepsilon (x) \varepsilon, B(x) C \}) \} = \{ (2, \{ \varepsilon, BC \}) \}$ となる。

26行目では θ の要素のうち、 $(2, \{ \varepsilon, BC \})$ が $|W| > 1$ の条件を満たすので、これについて27行目から31行目の処理を行う。

27行目で、 $q_{n2} = \{ (q_n, \alpha) \} = \{ (2, a) \}$ となり、28行目へと処理が進む。

[0108] いま、 $W = \{ \varepsilon, BC \}$ であるので、まず、 $w = \varepsilon$ について29行目の処理を行う。ここで $a \cup x (q_{n2}, w) = a \cup x (\{ (2, a) \}, \varepsilon) = \#_1$ とする。そうすると、29行目では、遷移元状態が $\{ (0, \varepsilon), (1, B) \}$ 、入力記号が $\#_1$ 、出力記号列が ε 、遷移先状態が $\{ (2, a) \}$ である遷移が E_2 に追加され、さらに Σ_2 に $\#_1$ が追加される。同様に $w = BC$ についても処理を行うが、このとき $a \cup x (\{ (2, a) \}, BC) = \#_2$ とすると、遷移元状態が $\{ (0, \varepsilon), (1, B) \}$ 、入力記号が $\#_2$ 、出力記号列が BC 、遷移先状態が $\{ (2, a) \}$ である遷移が E_2 に追加され、さらに Σ_2 に $\#_2$ が追加される。その結果、図13のうち、初期状態と状態S201とそれらの状態間の遷移から構成されるFSTになる。

[0109] 30行目で、 Q_2 に $\{ (2, a) \}$ が含まれていないと判定され、31行目でSと Q_2 に $\{ (2, a) \}$ が追加される。その結果、 $S = \{ \{ (2, a) \} \}$ 、 $Q_2 = \{ \{ (0, \varepsilon), (1, B) \}, \{ (2, a) \} \}$ となる。

32行目の処理に進むが、 $|W| = 1$ を満たす要素が θ には含まれていない

ので、32行目の条件を満たさない。したがって、33行目以降の処理は行われず、5行目に処理が戻る。

[0110] Sは空集合ではないので6行目に処理を進める。

6行目で $q_2 = \{(2, a)\}$ となる。また、 $S = \phi$ となる。

q_2 に含まれる記号列はaであるので、7行目の条件を満たし、8行目へ処理を進める。

[0111] 8行目で、 $q = 2$ 、 $w = a$ となる。

9行目で、 $q_{n2} = \{(2, 1^\#)\} = \{(2, \varepsilon)\}$ となる。

10行目で、遷移元状態が $\{(2, a)\}$ 、入力記号がa、出力記号列が ε 、遷移先状態が $\{(2, \varepsilon)\}$ の遷移が E_2 に追加される。その結果、図13のうち、初期状態と状態S201と状態S202とそれらの状態間の遷移から構成されるFSTになる。

[0112] $q_{n2} = \{(2, \varepsilon)\}$ であるので、11行目の条件を満たし、12行目へ処理を進める。

12行目で、Sと Q_2 に $\{(2, \varepsilon)\}$ が追加される。その結果、 $S = \{\{(2, \varepsilon)\}\}$ 、 $Q_2 = \{\{(0, \varepsilon), (1, B)\}, \{(2, a)\}, \{(2, \varepsilon)\}\}$ となる。

5行目に処理が戻るが、Sは空ではないので6行目に処理を進める。

6行目で $q_2 = \{(2, \varepsilon)\}$ となる。また、 $S = \phi$ となる。

[0113] q_2 に含まれる記号列は ε のみであるので、7行目の条件は満たさない。そのため、14行目に処理が進む。

状態2は終了状態ではないので、14行目の条件は満たさず、24行目の処理へ進む。

[0114] 25行目以降で処理すべき α はbだけなので、 $\alpha = b$ とする。

25行目で、 $\theta = \{(3, \{\varepsilon(x)\varepsilon\}), (4, \{\varepsilon(x)D\}), (5, \{\varepsilon(x)DB\})\} = \{(3, \{\varepsilon\}), (4, \{D\}), (5, \{DB\})\}$ となる。

26行目を満たす θ の要素はないので、32行目へと処理を進める。

θ の要素は3つあり、それらの要素に関して W は $\{\varepsilon\}$, $\{D\}$, $\{DB\}$ であるので、それら全てについて $|W| = 1$ となる。したがって、32行目の条件を満たし、33行目以降の処理に進む。

[0115] 33行目で、 $w_2 = \varepsilon$ になる。

34行目で、 $q_{n2} = \{(3, \varepsilon), (4, D), (5, DB)\}$ となる。

35行目で、遷移元状態が $\{(2, \varepsilon)\}$ 、入力記号が b 、出力記号が ε 、遷移先状態が $\{(3, \varepsilon), (4, D), (5, DB)\}$ の遷移が E_2 に追加される。その結果、図13のFSTになる。

[0116] $q_{n2} = \{(3, \varepsilon), (4, D), (5, DB)\}$ なので36行目の条件を満たし、37行目に処理が進む。37行目で、 S と Q_2 に $\{(3, \varepsilon), (4, D), (5, DB)\}$ が追加される。その結果、 $S = \{\{(3, \varepsilon), (4, D), (5, DB)\}\}$ 、 $Q_2 = \{\{(0, \varepsilon), (1, B)\}, \{(2, a)\}, \{(2, \varepsilon)\}, \{(3, \varepsilon), (4, D), (5, DB)\}\}$ となる。

[0117] 5行目に処理が戻るが、 S は空集合ではないので6行目に処理を進める。

6行目で $q_2 = \{(3, \varepsilon), (4, D), (5, DB)\}$ となる。また、 $S = \phi$ となる。

q_2 に含まれる記号列は ε と D と DB であるので、7行目の条件は満たさない。そのため、14行目に処理が進む。

[0118] 状態3, 4, 5は終了状態なので、14行目の条件を満たし、15行目の処理へ進む。

$\rho_1(3) = D$, $\rho_1(4) = A$, $\rho_1(5) = \varepsilon$ であるので、15行目で $W = \{\varepsilon(\times)D, D(\times)A, DB(\times)\varepsilon\} = \{D, DA, DB\}$ となる。

16行目で、 $w_2 = D(+)DA(+)DB = D$ となる。

W の要素には D が含まれているので、17行目の条件を満たし、18行目へ処理を進める。

[0119] 18行目で、 $\rho_2(\{(3, \varepsilon), (4, D), (5, DB)\}) = D$, $F_2 = \{\{(3, \varepsilon), (4, D), (5, DB)\}\}$ となる。その結果、図1

4のようになる。

[0120] 19行目を満たす W の要素があるので、まずは $w=DA$ について20行目から23行目の処理を行う。

20行目で $q_{n2}=\phi$ とする。

[0121] $aux(q_{n2}, w) = aux(\phi, DA) = \#_3$ とすると、21行目で、遷移元状態が $\{(3, \varepsilon), (4, D), (5, DB)\}$ 、入力記号が $\#_3$ 、出力記号が DA 、遷移先状態が ϕ の遷移が E_2 に追加される。

ϕ は Q_2 に含まれていないので、22行目を満たし、23行目へ進む。

23行目で、 $\rho_2(q_{n2}) = \rho_2(\phi) = 1^\# = \varepsilon$ 、 $Q_2 = \{ \{(0, \varepsilon), (1, B)\}, \{(2, a)\}, \{(2, \varepsilon)\}, \{(3, \varepsilon), (4, D), (5, DB)\}, \phi \}$ 、 $F_2 = \{ \{(3, \varepsilon), (4, D), (5, DB)\}, \phi \}$ となる。その結果、図14に図15の状態 S_{204} と遷移 t_{204a} が追加される。つまり図15のうち、遷移 t_{204b} を除く部分で構成される FST となる。

[0122] 次に、 $w=DB$ について20行目から23行目の処理を行う。

20行目で $q_{n2}=\phi$ とする。

$aux(q_{n2}, w) = aux(\phi, DB) = \#_4$ とすると、21行目で、遷移元状態が $\{(3, \varepsilon), (4, D), (5, DB)\}$ 、入力記号が $\#_4$ 、出力記号が DB 、遷移先状態が ϕ の遷移が E_2 に追加される。図15でいうと遷移 t_{204b} が追加される。その結果、図15の FST となる。

[0123] いま、 ϕ は Q_2 に含まれているので、22行目を満たさない。

w_2 でない W の要素の全てについて処理を終えたので、24行目に処理を進める。

状態3, 4, 5から出て行く遷移は存在しないので、24行目以降の処理を行っても何も起きない。

[0124] 5行目に処理が戻るが、いま S は空集合なので、5行目の条件を満たさない。したがって、決定化処理が完了したことになる。

前述したように、補助入力記号の番号($\#_n$ の n の部分)を状態ごとに1から

振るようにすることもできる。その場合、決定化後のF S Tは図16のようになる。

[0125] (初期出力がない場合のF S Tの処理について)

初期出力がないF S Tを用いる場合、決定化したいN F S Tの初期出力は ε であると考えられる。したがって、図3の2行目で、 $\lambda_2 = \varepsilon$ となる。つまり、決定化後のD F S Tも初期出力がないD F S Tになる。したがって、決定化前の初期状態に割り当てられている初期出力が全て ε であるとするだけで、初期出力のないN F S Tから初期出力のないD F S Tへと決定化する処理を図3の方法を用いることで行うことができる。

[0126] (終了出力がないF S Tの処理について)

終了出力がないF S T、つまり、 ρ が定義されないF S Tに関する決定化を行う場合には、図3の方法の15行目から23行目を図17のように変更すればよい。

決定化前のF S Tの終了出力がないので、 W は q_2 に含まれる2つ組(状態、記号列)のうち、 F_1 に含まれる状態と組になっている記号列から成る。ただし、記号列が $1^\#$ であるものは含めない(図17の1行目)。

[0127] W が空集合であれば(図17の2行目)、3行目で q_2 を F_2 に追加する。

W の全ての要素に対して図17の5行目から8行目の処理を行うが、5行目から7行目の動作は図3の20行目から22行目と同じである。図17の8行目では、 q_{n_2} を Q_2 と F_2 に追加する処理を行う。

[0128] 以上をまとめると、終了出力が必要ないときは単に q_2 を終了状態にする。そうでないときは新たに終了状態を作成し、その状態を遷移先状態、 q_2 を遷移元状態とする遷移を作成し、その遷移の出力記号列に w を割り当てるということである。

[0129] (終了出力が N 個あってもよいF S Tの処理について)

各終了状態に終了出力を最大で N 個($N \geq 0$)割り当てることができるF S Tの場合、状態 q に割り当てられる終了出力 $\rho(q)$ は出力記号列の集合となる。ただし、 $|\rho(q)| \leq N$ を満たすとする。 $\rho(q) = \phi$ の場合、

状態 q の終了出力はないということになる。このような F S T に関する決定化を行う場合には、図 3 の方法の 15 行目から 23 行目を図 18 のように変更すればよい。

[0130] 図 18 の 1 行目で q_2 を F_2 に追加する。

2 行目では終了出力の一覧を作成する。 q_2 に含まれる 2 つ組 (状態 q , 記号列 w) のうち、状態 q が決定化前の F S T の終了状態であるような全ての 2 つ組に対し、その記号列 w と状態 q に割り当てられている終了出力 y から $x = w(x)y$ を計算し、そのうち、 x が 1# でないような x からなる記号列の集合を W とする。

[0131] 3 行目では W のサイズが N より大きいかな否かを判断し、大きい場合には 4 行目へと処理を進め、そうでない場合には 11 行目へと処理を進める。 W のサイズが N より大きい場合は、 W に含まれる終了出力の全てを q_2 の終了出力に割り当てることができないからである。

4 行目では、図 3 の 20 行目と同様に遷移先状態 q_{n2} を作成する。

5 行目では、作成した q_{n2} が Q_2 に含まれているかな否かを判断し、含まれていなければ 6 行目の処理を行う。

[0132] 6 行目では、 q_{n2} の終了出力 $\rho_2(q_{n2})$ を空集合に設定し、 q_{n2} を Q_2 と F_2 に追加する。

7 行目では W のサイズが N より大きいかなを判断し、大きい場合には 8 行目から 10 行目を繰り返し処理する。

8 行目では、 W から出力記号を 1 つ取り出し、それを w とする。取り出し方は例えば W に含まれる出力記号列のうち長い出力記号列から取り出すようにすればよい。短い出力記号列から取り出すようにすることもできる。

[0133] 9 行目では遷移元状態を q_2 , 入力記号を $aux(q_{n2}, w)$, 出力記号列を w , 遷移先状態を q_{n2} とする遷移を E_2 へ追加する。さらに、追加した遷移の入力記号を Σ_2 へ追加する。

10 行目では、 W から w を取り除く。

11 行目では、 W に残っている出力記号を q_2 の終了出力 $\rho_2(q_2)$ へと割り

当てる。

[0134] (ハードウェア構成)

図19は本実施の形態に記載した方法を実施するために必要な決定化装置2700のハードウェア構成を示したブロック図である。図19に示されているように決定化装置2700は、CPU (Central Processing Unit) 1、操作部2、表示部3、ROM (Read Only Memory) 4、RAM (Random Access Memory) 5、記憶部6、バス7等を備えている。各部はバス7により接続されている。

[0135] CPU 1は、RAM 5の所定領域を作業領域として、ROM 4又は記憶部6に予め記憶された各種制御プログラムとの協働により各種処理を実行し、決定化装置2700を構成する各部の動作を統括的に制御する。またCPU 1は、ROM 4又は記憶部6に予め記憶された所定のプログラムとの協働により後述する各機能部の機能を実現させる。

[0136] 操作部2は、マウスやキーボード等の入力デバイスであって、ユーザから操作入力された情報を指示信号として受け付け、その指示信号をCPU 1に出力する。表示部3は、LCD (Liquid Crystal Display) 等の表示装置により構成され、CPU 1からの表示信号に基づいて、各種情報を表示する。

[0137] ROM 4は、決定化装置2700の制御にかかるプログラムや各種設定情報等を書き換え不可能に記憶する。RAM 5は、SDRAM等の揮発性の記憶媒体であって、CPU 1の作業エリアとして機能する。具体的には、決定化処理時において生成される各種変数やパラメータの値等を一時記憶するバッファ等の役割を果たす。

[0138] 記憶部6は、フラッシュメモリ等の半導体による記憶媒体や、磁氣的又は光学的に記録可能な記憶媒体を有し、決定化装置2700の制御にかかるプログラムや各種設定情報等を書き換え可能に記憶する。また記憶部6は、FSTに係る各種の情報を予め記憶したり、決定化処理の処理結果であるFS

Tを記録したりする。

[0139] (処理部構成)

図20は、本実施形態に係る有限状態トランスデューサの決定化方法を実行する、決定化装置の機能構成を示す図である。図20の決定化装置100は、FST読取部101からFSTが入力され、FST書込部109に対して決定化したFSTを出力する。

[0140] 決定化装置100は、集合生成部10、単値状態処理部20、終了状態処理部30、単値化処理部40、及び、遷移選択部50を有する。

[0141] 集合生成部10は、FST読取部101から入力されるFSTから、状態と、状態毎に対応する遷移に割り当てられている入力記号、出力記号、及び、遷移先の状態と、を含む集合をそれぞれ生成する。決定化後のFSTに対応する集合についても空集合であるものを含め生成する。また、その他のワーク領域として使用する集合等についても生成する。生成された集合は、単値状態処理部20、終了状態処理部30、単値化処理部40、及び、遷移選択部50によって処理され、FSTの決定化が行われる。集合生成部10は、図3の1行目ないし4行目の処理も行う。

[0142] 単値状態処理部20は、状態取得部21、及び、遷移生成制御部22を有する。状態取得部21は、集合生成部10により生成された集合の中から、決定化後の状態のうち、その状態から出て行く遷移の生成処理がまだ行われていない状態を取得する。この状態を決定化処理中状態と呼ぶこととし、図3の q_2 に対応する。遷移生成制御部22は、決定化処理中状態から出て行く遷移とその遷移先の状態を生成し、生成された状態を決定化後のFSTに含ませる。

[0143] なお、単値状態処理部20は、図3の6行目から12行目の処理を行う。状態取得部21は、図3の6行目の処理を行う。遷移生成制御部22は、図3の7行目ないし12行目の処理を行う。

[0144] 終了状態処理部30は、FSTの終了状態を生成する。終了状態処理部30は、状態併合部31、記号決定部32、及び、遷移生成制御部33を有す

る。状態併合部 31 は、決定化処理中状態に対応する一または複数の終了出力を取得する。

[0145] 記号決定部 32 は、状態併合部 31 が取得した終了出力が複数あるときに、必要に応じて遷移生成制御部 33 によって生成される遷移に対応する識別記号を決定する。ここで決定される識別記号は、FST 読取部 101 が読み取った FST に含まれているどの入力記号とも異なる記号となる。遷移生成制御部 33 は、遷移の生成や終了状態を生成する処理の全体を制御する。

[0146] なお、終了状態処理部 30 は、図 3 の 13 行目ないし 23 行目の処理を行う。状態併合部 31 は、図 3 の 15 行目の処理を行う。記号決定部 32 は、図 3 の 21 行目の処理のうち、関数 aux による処理を行う。遷移生成制御部 33 は、図 3 の 13 行目ないし 23 行目の処理のうち、状態併合部 31 及び記号決定部 32 による処理ではない処理を行う。

[0147] 単値化処理部 40 は、同一の入力記号が割り当てられている複数の遷移を併合して単値化すると共に決定化する処理を行う。単値化処理部 40 は、状態併合部 41、記号決定部 42、サブ状態生成部 43、遷移生成制御部 44 を有する。

[0148] 状態併合部 41 は、集合生成部 10 により生成された集合の中から、同一の入力記号による遷移先にある複数の状態を取得し、それらをサブ状態として有する状態を生成する。同一の入力記号による遷移先の状態が一つだけの場合にも複数ある場合と同様にサブ状態を有するようにしておくと、サブ状態に関する処理が容易になる。

[0149] 記号決定部 42 は、状態併合部 41 により生成された一の状態に対し、入力記号として扱う識別記号を一または複数生成する。生成された識別記号のそれぞれは、一の状態からの出て行く一または複数の遷移に割り当てられる出力記号のそれぞれに対応する。ここで決定される識別記号は、記号決定部 32 と同様に、FST 読取部 101 が読み取った FST に含まれているどの入力記号とも異なる記号となる。サブ状態生成部 43 は、状態併合部 41 が生成する状態に含まれるサブ状態を生成する。なお、記号決定部 42 が生成

する識別記号は、状態併合部 4 1 が生成した状態毎に対応する。

[0150] 遷移生成制御部 4 4 は、単値化処理部 4 0 が行う処理の全体の制御を行う。

[0151] なお、状態併合部 4 1 は、図 3 の 2 5 行目の処理を行う。記号決定部 4 2 は、図 3 の 2 9 行目の処理のうち、関数 aux による処理を行う。サブ状態生成部 4 3 は、図 3 の 2 7 行目と 3 4 行目の処理を行う。遷移生成制御部 4 4 は、図 3 の 2 4 行目ないし 3 7 行目の処理のうち、単値化処理部 4 0 が有する他の処理部が行う処理ではない処理を行う。

[0152] 遷移選択部 5 0 は、入力される F S T が重み付き F S T の場合に、重みに基づいて、最良の経路となる遷移を選択する。

[0153] (音声認識での F S T の具体例)

F S T は音声認識や自然言語処理の分野などで利用されている。ひとつの例として、音素列から単語へと変換する機能を N F S T として実現する方法が挙げられる。一般的にはこのような N F S T には決定化処理が施される。各状態からの遷移先の状態が入力記号を与えると一意に決まり、扱い易いからである。このとき、同音異表記語が存在すると、同じ発音列に対して異なる単語が割り当てられることになる。つまり単値でない N F S T が作成されることになる。具体的には図 2 1 のようになる。この例では、入力記号を発音記号として、出力記号列を単語としている。発音列を数単語にわたって変換したいときは、この図の場合、入力記号と出力記号が共に ε である状態 5 から状態 0 へと遷移を作成すればよい。

[0154] (翻訳での F S T の具体例)

発音から単語へと変換できる F S T を実現できるので、同様に、ある言語の文字を入力記号に、別の言語の文字列を出力記号に割り当てるとする。例えば、日本語で「会議」という文字列は英語では `meeting`, `conference`, `council` などに翻訳できるので、それを F S T として表すと図 2 2 のようになる。この例の場合も単値な F S T でないので、終了出力を割り当てない F S T を用いる場合には、従来の方法によって決定化す

ることができない。1単語だけでなく複数の単語を連続して他の言語に翻訳するには、状態1, 2, 3から状態4への遷移の遷移先を状態0に変更することで実現できる。このときは終了出力を複数割り当てられるFSTを用いても、補助入力記号を追加する事前処理なしでは従来の決定化の方法では決定化することができない。本実施の形態の方法であれば、事前処理なしで決定化を行うことができる。

[0155] (検索でのFSTの具体例)

検索したい対象の文章をFSTで表現すると、文書内の検索を容易に行うことができるようになる。

[0156] 例えば、「Good morning」という文章があったとする。文章の先頭「G」を1文字目、次の「o」を2文字目のようにする。各文字が現れる文章内での位置を出力記号、文章内の文字を入力記号として、図23のようなFSTを作成する。〈s p〉はスペースを表すものとする。例えばこの文章内の文字列「r n i」が出現する位置を検索したい場合、このFSTに入力記号列「r n i」を入力することになる。そうすると、状態0、状態8、状態9、状態10と遷移し、出力記号列は「8」となる。つまり、「r n i」は8文字目に出現することがわかる。入力記号に「o」が割り当てられている状態0からの遷移を見ると、「o : 2」、「o : 3」、「o : 7」の3つがある。これらの遷移の遷移先は全て終了状態であるので、入力記号列として「o」を入力すると、2, 3, 7の3種類を出力することができる。つまり、このFSTは単値でないということである。

[0157] FSTを決定化しておけば、検索したい文字列、つまり入力記号列に対するFSTでの経路が一意に定まるので、検索時の計算量を決定化しない場合に比べて減らすことができる。一方、このFSTは単値ではないため従来の決定化の手法では決定化できない。そこで本実施の形態における決定化を用いて単値なFSTへと変形しながら決定化を行う。そして、決定化したFSTを用いて検索するときには、補助入力記号を持つ遷移を全て ε 遷移とみなせばよい。 ε 遷移とは、入力記号を消費せずに次の状態へと遷移ができる遷

移のことである。

- [0158] 図23のFSTを本実施の形態における方法によって決定化したFSTを図24に示す。決定化後のFSTへ入力記号列「o」を入力すると、初期状態である状態0から状態13へと遷移する。補助入力記号である $\#_1$, $\#_2$, $\#_3$ は ε であるとみなすので、FSTに入力される入力記号列を消費せずに遷移できる先は状態15のみである。それらの遷移の出力記号列は2, 3, 7なので、文章中の「o」が現れる位置は、2文字目, 3文字目, 7文字目ということになる。FSTに入力される入力記号列が「or」であれば、状態0から状態13を経由して状態8へと遷移する経路が1つのみあり、その経路の出力記号は7である。つまり、「or」が現れる位置は7文字目ということになる。

- [0159] (効果)

以上のように本実施の形態によると、事前に単値に変形する処理を行うことなく、単値でないFSTを単値でないFSTへと変形しながら決定化することができる。

- [0160] (実施の形態2)

(準備)

本実施の形態では、FSTの各遷移に割り当てられている入力記号と出力記号列に加えて、重みが加えられた、重み付き有限状態トランスデューサ (Weighted Finite State Transducer, WFST) の決定化方法のうち、単値でないWFSTを単値にしながら決定化する方法について述べる。ここでいう重みとは、何らかの確率値やスコア、ペナルティなどであり、入力記号を受理する経路に沿って何らかの規則 (足し算, 掛け算, 最小値, 最大値などなんでも良い) によって演算されるものである。

- [0161] 重みがとりうる値の集合を R_w とする。 R_w は例えば整数であったり実数であったりする。また、FSTの場合には Δ としていた出力記号の集合を R_o とする。そうすると、WFSTは9つ組 $(Q, \Sigma, R_o, R_w, E, I, F, \lambda$

, ρ)で表すことができる。 $Q, \Sigma, I, F, \lambda, \rho$ はFSTの場合と同様である。遷移の集合 E に関しては、 $E \subseteq Q \times \Sigma \times \Delta \times Q$ となる。ここで Δ を出力とし、 $\Delta = R_o^* \times R_w$ とした。遷移 $e \in E$ に割り当てられている出力記号列と重みはFSTの場合と同じように、 $out(e) \in \Delta$ とする。つまり、 $out(e)$ は遷移 e に割り当てられている2つ組(出力記号列, 重み)を表すということである。初期出力関数 λ は初期状態を $i \in I$ とするとき $\lambda(i) \in \Delta$ である。同様に終了状態を $q \in F$ とするとき、終了出力関数は $\rho(q) \in \Delta$ である。

[0162] (方法)

出力 Δ の計算方法を次のように定義すると、単値でないWFSTの決定化に実施の形態1の決定化方法をそのまま用いることができるようになる。

出力記号列 R_o の演算 $(+)_o$ と $(\times)_o$ と特別な値 $1^\#_o$ を実施の形態1と同じように定義する。つまり、 $(+)_o$ を前方最長一致(longest common prefix)演算であると定義し、 $(\times)_o$ を結合演算であると定義し、 $1^\#_o$ を空記号列 ε であると定義する。

[0163] 重み R_w の演算 $(+)_w$ と $(\times)_w$ と特別な値 $1^\#_w$ は例えば次のように定義する。値 $x, y \in R_w$ に対して、 $x(+)_w y = \min(x, y)$, $x(\times)_w y = x + y$ と定義し、 $1^\#_w = 0$ と定義する。これは非特許文献5でtropical semiringと呼ばれている演算方法である。他にも、 $x(+)_w y = x + y$, $x(\times)_w y = x \times y$ と定義し、 $1^\#_w = 1$ と定義することができる。非特許文献5でlog semiringと呼ばれている演算方法を用いることも可能である。

[0164] 以上のように出力記号列と重みの演算方法を定義すると、出力記号列と重みの2つ組である出力 Δ の演算 $(+)$ と $(\times)$ と特別な値 $1^\#$ は、 $x_1, x_2 \in R_o$, $y_1, y_2 \in R_w$ であるとして、それぞれ $(x_1, y_1)(+)(x_2, y_2) = (x_1(+)_o x_2, y_1(+)_w y_2)$, $(x_1, y_1)(\times)(x_2, y_2) = (x_1(\times)_o x_2, y_1(\times)_w y_2)$, $1^\# = (1^\#_o, 1^\#_w)$ と定義できる。さらに出力 Δ に関する和集合の演算 $\cup$ を次式のように定義する。

[0165] [数1]

$$\{(x_1, y_1)\} \cup \{(x_2, y_2)\} = \begin{cases} \{(x_1, y_1 \oplus_w y_2)\} & \text{if } x_1 = x_2 \\ \{(x_1, y_1), (x_2, y_2)\} & \text{if } x_1 \neq x_2 \end{cases} \quad \dots (1)$$

[0166] 決定化後の状態 $q \in Q_2$ と出力 $v \in \Delta$ を引数にとり、戻り値として補助入力記号返すアルゴリズム $a u x$ を図 25 のようにする。1 行目で出力 Δ の出力記号列と重みをそれぞれ x, y とし、2 行目で状態 q と出力記号列 x を引数として $a u x_0$ を呼び出し、その戻り値を返すという処理を行うということである。

ここで、 $a u x_0$ は実施の形態 1 で使用したアルゴリズム $a u x$ に相当し、例えば図 4 のような方法である。

[0167] 図 3 で使用されている $(+)$, $(\times)$, $1^\#$, U , $a u x$ を以上のように定義し直すと、単値でない WFS T に関しても、図 3 の方法により単値に変形しながら決定化を行うことができる。

[0168] (例)

本実施の形態における WFS T の決定化方法がどのように動作するのかについて、図 26 の WFS T を決定化前の WFS T T_1 とし、決定化後の WFS T を T_2 として説明する。初期出力と終了出力は $1^\#$ とし、図からは省略している。

[0169] 図 26 より $Q_1 = \{0, 1, 2, 3\}$, $I_1 = \{0\}$, $F_1 = \{3\}$, $\Sigma_1 = \{a, b\}$, $R_0 = \{A, B\}$ である。 R_w は整数とする。実施の形態 1 における F S T の図と同様に、太線の円が初期状態であり、二重線の円が終了状態を表している。各遷移のラベルのうち、「:」の左側が入力記号、「:」と「/」の間が出力記号列、「/」の右側が重みを表している。なお、この例では重み $x, y \in R_w$ を $x (+)_w y = \min(x, y)$, $x (\times)_w y = x + y$, $1^\#_w = 0$ で演算するものとする。そうすると、 $x^{-1} (\times) x = 1^\#_w$ より $x^{-1} + x = 0$ であるので、 $x^{-1} = -x$ となる。

[0170] 図 3 の 1 行目でまず決定化後の WFS T の初期化が行われる。次に 2 行目で初期出力を計算する。この例の WFS T には初期出力はない。言い換える

と $\lambda_1(0) = 1^\# = (\varepsilon, 0)$, つまり空記号列と重み 0 を出力するという
ことである。したがって, $\lambda_2 = (\varepsilon, 0)$ となる。

[0171] 3 行目では決定化後の WFS T の初期状態を作成する。 $i_2 = \{ (0, (\varepsilon, 0)) \}$ となる (図 27)。

[0172] 4 行目では 3 行目で作成した状態 i_2 をスタック S と Q_2 に追加し, その結果, $S = \{ \{ (0, (\varepsilon, 0)) \} \}$, $Q_2 = \{ \{ (0, (\varepsilon, 0)) \} \}$ となる。

$|S| = 1$ であるので, 5 行目の条件を満たし, 6 行目に処理を進める。

6 行目でスタック S から値を取り出し, q_2 に代入すると, $S = \phi$, $Q_2 = \{ (0, (\varepsilon, 0)) \}$ となる。

[0173] 7 行目では, 判断しなければならない w は $w = (\varepsilon, 0)$ の場合のみである。 w は Σ_1 には含まれていないので, 7 行目の条件は満たされない。したがって, 14 行目に処理を進める。

[0174] 14 行目では, q_2 に含まれている 2 つ組 (状態, 出力) のうちの状態が F_1 に含まれているか否かを判断する。いま, q_2 に含まれる状態は状態 0 のみであり, これは F_1 には含まれていない。したがって, 処理を 24 行目に進める。

[0175] $\Sigma_1 = \{a, b\}$ ではあるが, 状態 0 から出て行く遷移の入力記号は a のみであるので, $\alpha = a$ の場合のみを処理すればよい。したがって, 24 行目では $\alpha = a$ として 25 行目以降の処理を行う。

[0176] 25 行目では, $\theta = \{ (1, \{ (\varepsilon, 0) (x) (A, 1) \}), (2, \{ (\varepsilon, 0) (x) (A, 3) \}) \} = \{ (1, \{ (\varepsilon (x) \circ A, 0 (x)_w 1) \}), (2, \{ (\varepsilon (x) \circ A, 0 (x)_w 3) \}) \} = \{ (1, \{ (A, 0+1) \}), (2, \{ (A, 0+3) \}) \} = \{ (1, \{ (A, 1) \}), (2, \{ (A, 3) \}) \}$ となる。

26 行目の $|W| > 1$ の条件を満たす θ の要素はないので, 32 行目へと処理を進める。

$|W| = 1$ を満たす θ の要素は存在するので 32 行目の条件を満たし, 33

行目へと処理を進める。

[0177] 3 3 行目の処理で, $w_2 = (A, 1) (+) (A, 3) = (A (+) \circ A, 1 (+) \circ w_3) = (A, \min(1, 3)) = (A, 1)$ となる。

3 4 行目の処理で, $q_{n_2} = \{ (1, (A, 1)^{-1} (\times) (A, 1)), (2, (A, 1)^{-1} (\times) (A, 3)) \} = \{ (1, (A^{-1} (\times) \circ A, 1^{-1} (\times) \circ w_1)), (2, (A^{-1} (\times) \circ A, 1^{-1} (\times) \circ w_3)) \} = \{ (1, (\varepsilon, -1+1)), (2, (\varepsilon, -1+3)) \} = \{ (1, (\varepsilon, 0)), (2, (\varepsilon, 2)) \}$ が得られる。

3 5 行目で状態元状態が $\{ (0, (\varepsilon, 0)) \}$ で入力記号が a , 出力記号が A , 重みが 1, 遷移先状態が $\{ (1, (\varepsilon, 0)), (2, (\varepsilon, 2)) \}$ である遷移を作成し E_2 に加える。その結果, 図 2 8 の初期状態と状態 $S_3 0 1$ とそれらの状態間の遷移から構成される F S T になる。

[0178] 今 q_{n_2} は Q_2 には含まれていないので 3 6 行目を満たし, 3 7 行目が処理される。

3 7 行目では, q_{n_2} を S と Q_2 に追加するので, $S = \{ (1, (\varepsilon, 0)), (2, (\varepsilon, 2)) \}$, $Q_2 = \{ (0, (\varepsilon, 0)), (1, (\varepsilon, 0)), (2, (\varepsilon, 2)) \}$ となる。

[0179] 5 行目に処理が戻り, スタック S は空集合ではないので, 6 行目に処理を進む。

6 行目で, $q_2 = \{ (1, (\varepsilon, 0)), (2, (\varepsilon, 2)) \}$ となり, $S = \phi$ となる。

$(\varepsilon, 0)$ も $(\varepsilon, 2)$ も Σ_1 には含まれていないので, 7 行目の条件は満たされず, 1 4 行目に処理が進む。

[0180] 状態 1, 2 は F_1 には含まれていない, つまり終了状態ではないので, 1 4 行目の条件は満たさない。したがって 2 4 行目に処理が進む。

状態 1, 2 から出て行く遷移の入力記号は b のみであるので, 2 5 行目から 3 7 行目の処理は $\alpha = b$ の場合の処理のみを行えばよい。

[0181] 2 5 行目では, $\theta = \{ (3, \{ (\varepsilon, 0) (\times) (A, 2) \} \cup \{ (\varepsilon,$

$0) (x) (B, 1) \} \cup \{ (\varepsilon, 2) (x) (A, 5) \} \cup \{ (\varepsilon, 2) (x) (A, 2) \}) \} = \{ (3, \{ (A, 2) \} \cup \{ (B, 1) \} \cup \{ (A, 7) \} \cup \{ (A, 4) \}) \} = \{ (3, \{ (A, 2 (+)_w 7) \} \cup \{ (B, 1 (+)_w 4) \}) \} = \{ (3, \{ (A, \min(2, 7)) \} \cup \{ (B, \min(1, 4)) \}) \} = \{ (3, \{ (A, 2) \} \cup \{ (B, 1) \}) \} = \{ (3, \{ (A, 2), (B, 1) \}) \}$ となる。

[0182] 26行目では θ の要素のうち、 $(3, \{ (A, 2), (B, 1) \})$ が $|W| > 1$ の条件を満たすので、これについて27行目から31行目の処理を行う。

27行目で、 $q_{n2} = \{ (q_n, \alpha) \} = \{ (3, b) \}$ となり、28行目へと処理が進む。

[0183] いま、 $W = \{ (A, 2), (B, 1) \}$ であるので、まず、 $w = (A, 2)$ について29行目の処理を行う。ここで $aux(q_{n2}, w) = aux(\{ (3, b) \}, (A, 2)) = \#_1$ とする。そうすると、29行目では、遷移元状態が $\{ (1, (\varepsilon, 0)), (2, (\varepsilon, 2)) \}$ 、入力記号が $\#_1$ 、出力が $(A, 2)$ 、遷移先状態が $\{ (3, b) \}$ である遷移が E_2 に追加され、さらに Σ_2 に $\#_1$ が追加される。同様に $w = (B, 1)$ についても処理を行うが、このとき $aux(\{ (3, b) \}, (B, 1)) = \#_2$ とすると、遷移元状態が $\{ (1, (\varepsilon, 0)), (2, (\varepsilon, 2)) \}$ 、入力記号が $\#_2$ 、出力が $(B, 1)$ 、遷移先状態が $\{ (3, b) \}$ である遷移が E_2 に追加され、さらに Σ_2 に $\#_2$ が追加される。その結果、図28の初期状態と状態S301と状態S302と、それらの状態間の遷移から構成されるFSTになる。

[0184] 30行目で、 Q_2 に $\{ (3, b) \}$ が含まれていないと判定され、31行目でSと Q_2 に $\{ (3, b) \}$ が追加される。その結果、 $S = \{ \{ (3, b) \} \}$ 、 $Q_2 = \{ \{ (0, (\varepsilon, 0)) \}, \{ (1, (\varepsilon, 0)), (2, (\varepsilon, 2)) \}, \{ (3, b) \} \}$ となる。

[0185] 32行目の処理に進むが、 $|W| = 1$ を満たす要素が θ に含まれていないので、32行目の条件を満たさない。したがって、33行目以降の処理は行

われず、5行目に処理が戻る。

[0186] Sは空集合ではないので6行目に処理を進める。

6行目で $q_2 = \{ (3, b) \}$ となる。また、 $S = \phi$ となる。

q_2 に含まれる記号列はbであるので、7行目の条件を満たし、8行目へ処理を進める。

[0187] 8行目で、 $q = 3$ 、 $w = b$ となる。

9行目で、 $q_{n2} = \{ (3, 1^\#) \} = \{ (3, (\varepsilon, 0)) \}$ となる。

10行目で、遷移元状態が $\{ (3, b) \}$ 、入力記号がb、出力記号列が ε 、重みが0、遷移先状態が $\{ (3, (\varepsilon, 0)) \}$ の遷移が E_2 に追加される。その結果、図28の状態S303を含む全ての状態と遷移から構成されるFSTになる。

[0188] $q_{n2} = \{ (3, \varepsilon) \}$ であるので、11行目の条件を満たし、12行目へ処理を進める。

12行目で、Sと Q_2 に $\{ (3, (\varepsilon, 0)) \}$ が追加される。その結果、 $S = \{ \{ (3, (\varepsilon, 0)) \} \}$ 、 $Q_2 = \{ \{ (0, (\varepsilon, 0)) \}, \{ (1, (\varepsilon, 0)) \}, \{ (2, (\varepsilon, 2)) \}, \{ (3, b) \}, \{ (3, (\varepsilon, 0)) \} \}$ となる。

[0189] 5行目に処理が戻るが、Sは空ではないので6行目に処理を進める。

6行目で $q_2 = \{ (3, (\varepsilon, 0)) \}$ となる。また、 $S = \phi$ となる。

q_2 に含まれる記号列は $(\varepsilon, 0)$ のみであるので、7行目の条件を満たさない。そのため、14行目に処理が進む。

[0190] 状態3は終了状態なので、14行目の条件を満たし、15行目の処理へ進む。

$\rho_1(3) = (\varepsilon, 0)$ であるので、15行目で $W = \{ (\varepsilon, 0) (\times) (\varepsilon, 0) \} = \{ (\varepsilon, 0) \}$ となる。

16行目で、 $w_2 = \{ (\varepsilon, 0) \}$ となる。

Wの要素は1つで、それは $(\varepsilon, 0)$ であるので、17行目の条件を満たし、18行目へ処理を進める。

[0191] 18行目で、 $\rho_2(\{(3, (\varepsilon, 0))\}) = (\varepsilon, 0)$ 、 $F_2 = \{\{(3, (\varepsilon, 0))\}\}$ となる。その結果、図29のようになる。

[0192] 19行目を満たすWの要素はないので、20行目から23行目は処理されず、24行目の処理へ進む。

状態3から出て行く遷移は存在しないので、24行目以降の処理を行っても何も起きない。

[0193] 5行目に処理が戻るが、いまSは空集合なので、5行目の条件を満たさない。したがって、決定化処理が完了したことになる。

F S Tの場合と同様に、各状態に関連付けられている2つ組の集合を番号に置き換えて、図30のようにすることもできる。

[0194] (音声認識での具体例)

W F S Tを用いた具体例として、音声認識の単語辞書に単語ごとに挿入ペナルティを重みとして設定したW F S Tを図31に挙げる。この場合、s e eを出力する場合は挿入ペナルティが10、s e aを出力するときは挿入ペナルティが20となる。音声認識における他の使用例については非特許文献4に記述がある。

[0195] 以上のように、実施の形態2の方法によると、単値でないW F S Tを単値なW F S Tへと変形しながら決定化することができる。

[0196] なお、本発明は、上記実施の形態そのままに限定されるものではなく、実施段階ではその要旨を逸脱しない範囲で構成要素を変形して具体化することができる。また、上記実施の形態に開示されている複数の構成要素の適宜な組み合わせにより、種々の発明を形成することができる。例えば、実施の形態に示される全構成要素からいくつかの構成要素を削除してもよい。さらに、異なる実施の形態にわたる構成要素を適宜組み合わせても良い。

[0197] 以上のように、本発明にかかる決定化装置は、音声認識技術や翻訳技術、要約技術に適している。

符号の説明

[0198] 10 集合生成部

- 2 0 単値状態処理部
- 2 1 状態取得部
- 2 2 遷移生成制御部
- 3 0 終了状態処理部
- 3 1 状態併合部
- 3 2 記号決定部
- 3 3 遷移生成制御部
- 4 0 単値化処理部
- 4 1 状態併合部
- 4 2 記号決定部
- 4 3 サブ状態生成部
- 4 4 遷移生成制御部
- 5 0 遷移選択部
- 1 0 0 決定化装置
- 1 0 1 F S T 読取部
- 1 0 9 F S T 書込部
- 2 7 0 0 決定化装置

請求の範囲

- [請求項1] 有限状態トランスデューサが有する遷移に割り当てられている入力記号とは異なる記号である識別記号を生成する記号決定部と、
- 前記有限状態トランスデューサが有する状態のうち、同一の入力記号による遷移先にある一または複数の状態を抽出し、抽出した前記状態をサブ状態として有する状態を生成する状態併合部と、
- 前記状態併合部が生成する状態間の遷移の入力記号として、有限状態トランスデューサが有する遷移に割り当てられている入力記号または前記識別記号を付与することにより決定化する単値化処理部と、
- を有することを特徴とする有限状態トランスデューサ決定化装置。
- [請求項2] 前記識別記号が付与された入力による遷移先の状態を、前記サブ状態のそれぞれの遷移先の状態とする遷移生成制御部を有することを特徴とする請求項1記載の有限状態トランスデューサ決定化装置。
- [請求項3] 前記遷移生成制御部は、前記サブ状態の次の状態から遷移する際の入力記号を、決定化前の前記次の状態に対応する入力記号とすることを特徴とする請求項2記載の有限状態トランスデューサ決定化装置。
- [請求項4] 前記サブ状態から次の状態への出力記号は、空を含むことを特徴とする請求項2記載の有限状態トランスデューサ決定化装置。
- [請求項5] 前記状態併合部が生成した状態が有する一または複数のサブ状態からの遷移のうち、同一の入力記号により遷移する複数の遷移があり、かつ、前記複数の遷移による出力記号が複数あり、前記複数の遷移による遷移先の状態が同一の場合に、
- 前記状態併合部は、前記複数の遷移の遷移先の状態を生成し、
- 前記記号決定部は、前記複数の遷移毎に対応する識別記号を付与することを特徴とする請求項1記載の有限状態トランスデューサ決定化装置。
- [請求項6] 前記記号決定部は、前記状態併合部が生成した状態と該状態への遷移の出力記号とを、該状態への遷移の識別記号に対応づけることを特

徴とする請求項 1 記載の有限状態トランスデューサ決定化装置。

[請求項7] 前記記号決定部は、さらに、一の状態からの遷移毎に、前記識別記号を一意にすることを特徴とする請求項 6 記載の有限状態トランスデューサ決定化装置。

[請求項8] 前記記号決定部は、前記識別記号に対応付けられる状態毎、及び、出力記号毎、に、前記識別記号を一意にすることを特徴とする請求項 6 記載の有限状態トランスデューサ決定化装置。

[請求項9] 前記有限状態トランスデューサが、重み付き有限状態トランスデューサの場合に、

前記識別記号に対応する遷移のうち、重みに基づいて、最良の経路となる遷移を選択する遷移選択部を有することを特徴とする請求項 1 記載の有限状態トランスデューサ決定化装置。

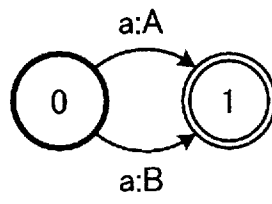
[請求項10] 有限状態トランスデューサが有する遷移に割り当てられている入力記号とは異なる記号である識別記号を生成する記号決定ステップと、

前記有限状態トランスデューサが有する状態のうち、同一の入力による遷移先にある一または複数の状態を抽出し、抽出した前記状態をサブ状態として有する状態を生成する状態併合ステップと、

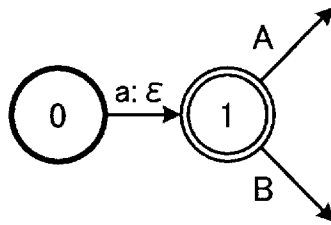
前記状態併合ステップが生成する状態間の遷移の入力記号として、有限状態トランスデューサが有する遷移に割り当てられている入力記号または前記識別記号を付与することにより決定化する単値化ステップと、

を有することを特徴とする有限状態トランスデューサ決定化方法。

[図1]



[図2]



[図3-1]

determine (T_1, T_2)

1 $Q_2 \leftarrow \phi, E_2 \leftarrow \phi, \Sigma_2 \leftarrow \Sigma_1, F_2 \leftarrow \phi$

2 $\lambda_2 \leftarrow \bigoplus_{i \in I_1} \lambda_1(i)$

3 $i_2 \leftarrow \{(i, \lambda_2^{-1} \otimes \lambda_1(i)) \mid i \in I_1\}$

4 $S \leftarrow \{i_2\}, Q_2 \leftarrow Q_2 \cup \{i_2\}$

5 while $S \neq \phi$

6 $q_2 \leftarrow \text{pop}(S)$

7 if $\{w \in \Sigma_1 \mid (q, w) \in q_2\} \neq \phi$

8 let $q, w : (q, w) \in q_2$

9 $q_{n2} \leftarrow \{(q, l^\#)\}$

10 $E_2 \leftarrow E_2 \cup \{(q_2, w, l^\#, q_{n2})\}$

11 if $q_{n2} \notin Q_2$

12 push $(S, q_{n2}), Q_2 \leftarrow Q_2 \cup \{q_{n2}\}$

13 else

14 if $\{q \in F_1 \mid (q, w) \in q_2\} \neq \phi$

15 $W \leftarrow \bigcup_{q \in F_1, (q, w) \in q_2} \{w \otimes \rho_1(q)\}$

16 $w_2 \leftarrow \bigoplus_{w \in W} w$

17 if $\{w \mid w \in W, w = w_2\} \neq \phi$

18 $\rho_2(q_2) \leftarrow w_2, F_2 \leftarrow F_2 \cup \{q_2\}$

19 foreach $w \in W$ such that $w \neq w_2$

20 $q_{n2} \leftarrow \phi$

21 $E_2 \leftarrow E_2 \cup \{(q_2, \text{aux}(q_{n2}, w), w, q_{n2})\}, \Sigma_2 \leftarrow \Sigma_2 \cup \{\text{aux}(q_{n2}, w)\}$

22 if $q_{n2} \notin Q_2$

23 $\rho_2(q_{n2}) \leftarrow l^\#, Q_2 \leftarrow Q_2 \cup \{q_{n2}\}, F_2 \leftarrow F_2 \cup \{q_{n2}\}$

24 foreach $\alpha \in \Sigma_1$

25 $\theta \leftarrow \bigcup_{q_n \in \gamma(q_2, \alpha)} \left\{ \left(q_n, \bigcup_{(w, e) \in \xi(q_2, \alpha, q_n)} \{w \otimes \text{out}(e)\} \right) \right\}$

[図3-2]

```

26      foreach  $(q_n, W) \in \theta$  such that  $|W| > 1$ 
27           $q_{n2} \leftarrow \{(q_n, \alpha)\}$ 
28          foreach  $w \in W$ 
29               $E_2 \leftarrow E_2 \cup \{(q_2, \text{aux}(q_{n2}, w), w, q_{n2})\}, \Sigma_2 \leftarrow \Sigma_2 \cup \{\text{aux}(q_{n2}, w)\}$ 
30          if  $q_{n2} \notin Q_2$ 
31              push( $S, q_{n2}$ ),  $Q_2 \leftarrow Q_2 \cup \{q_{n2}\}$ 
32      if  $\{(q_n, W) \in \theta \mid |W| = 1\} \neq \phi$ 
33           $w_2 \leftarrow \bigoplus_{(q_n, W) \in \theta, w \in W, |W|=1} w$ 
34           $q_{n2} \leftarrow \{(q_n, w_2^{-1} \otimes w) \mid (q_n, W) \in \theta, |W| = 1, w \in W\}$ 
35           $E_2 \leftarrow E_2 \cup \{(q_2, \alpha, w_2, q_{n2})\}$ 
36          if  $q_{n2} \notin Q_2$ 
37              push( $S, q_{n2}$ ),  $Q_2 \leftarrow Q_2 \cup \{q_{n2}\}$ 

```

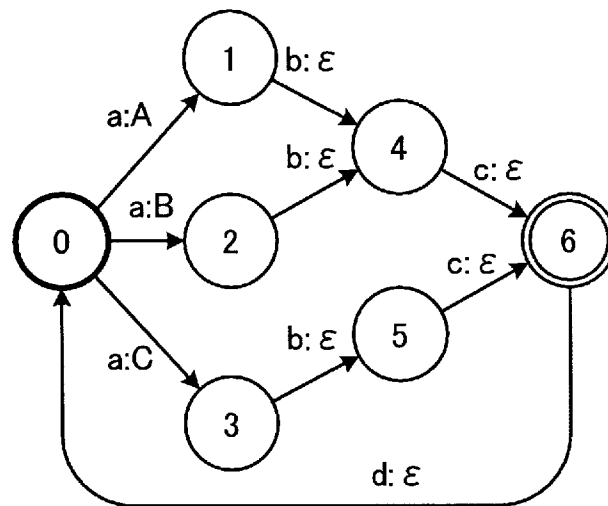
[図4]

```

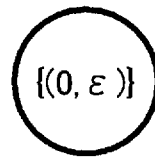
aux(q, v)
1  if  $\{(q_1, v_1, y_1) \in Y \mid q_1 = q, v_1 = v\} \neq \emptyset$ 
2    return  $y_1$ 
3   $y \leftarrow \text{make\_aux}(|Y|+1)$ 
4   $Y \leftarrow Y \cup \{(q, v, y)\}$ 
5  return  $y$ 

```

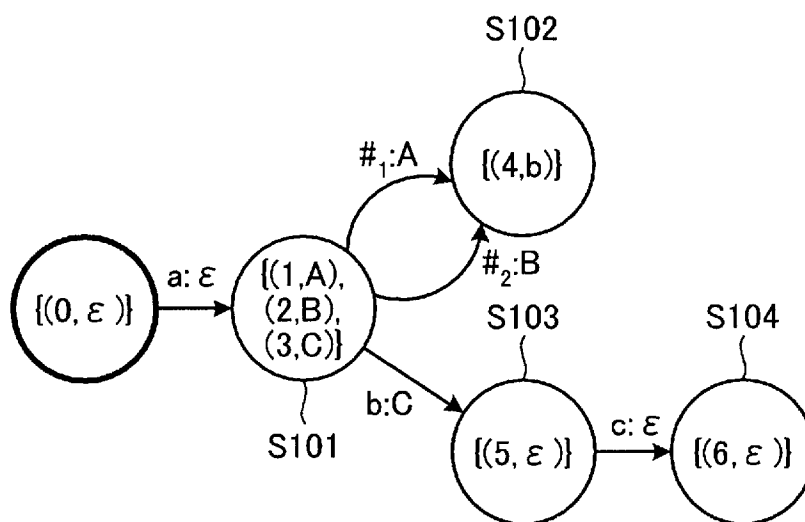
[図5]



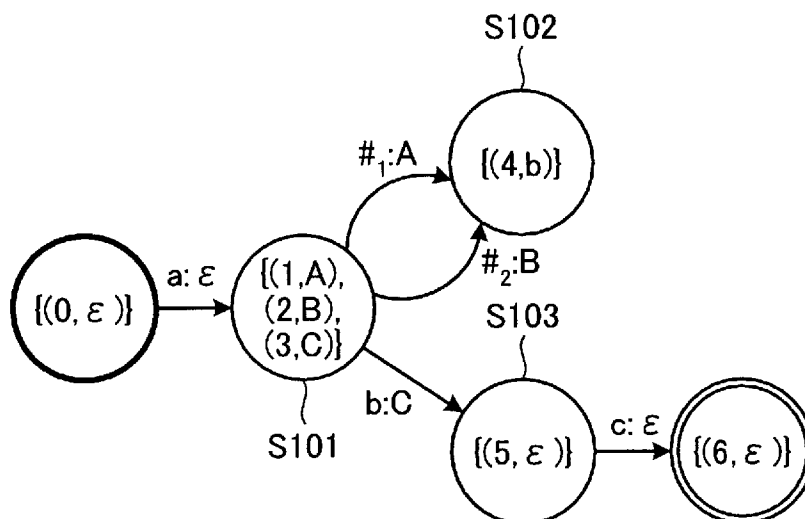
[図6]



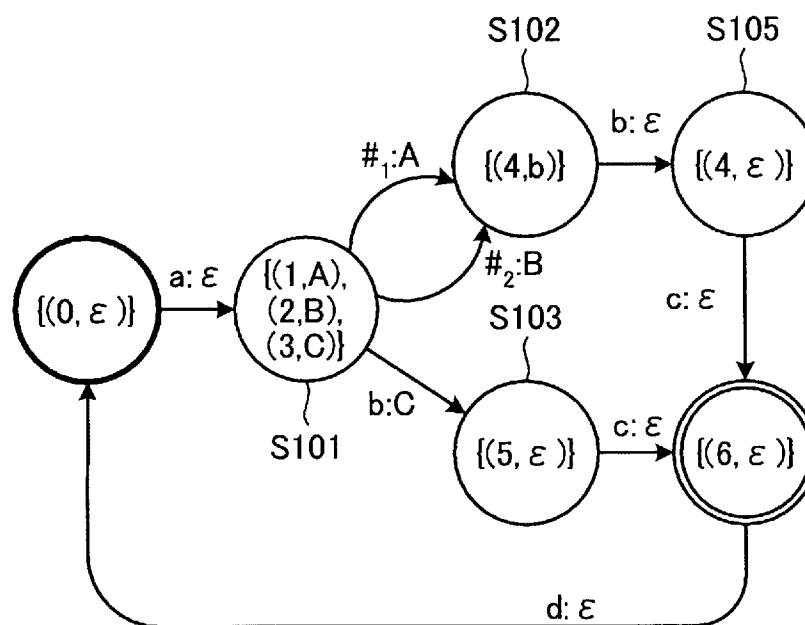
[図7]



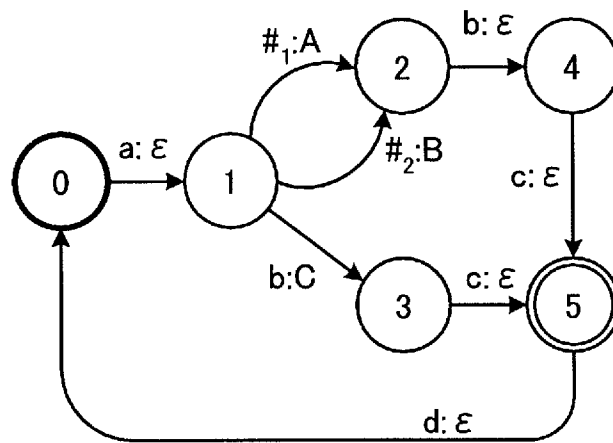
[図8]



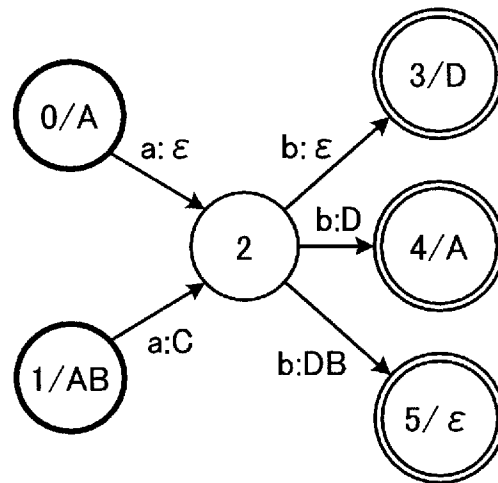
[図9]



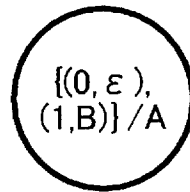
[図10]



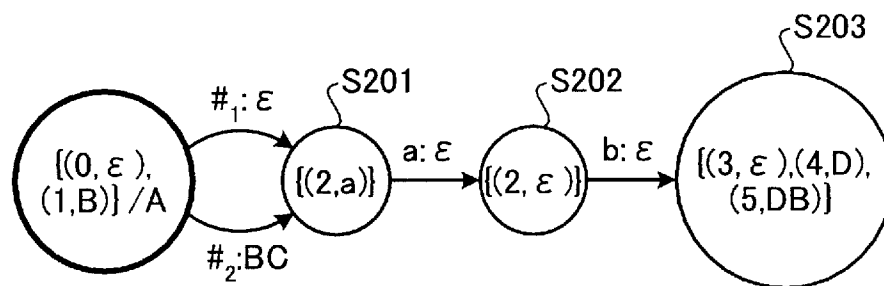
[図11]



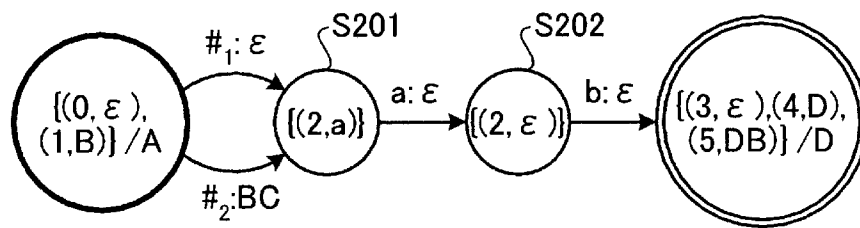
[図12]



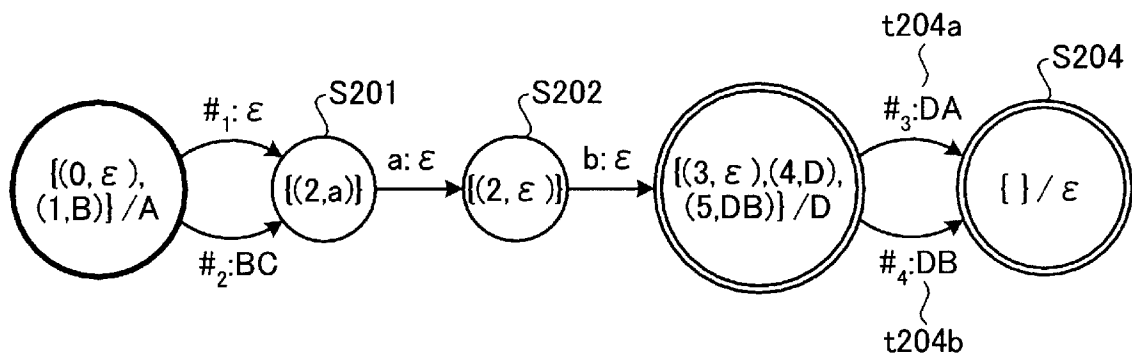
[図13]



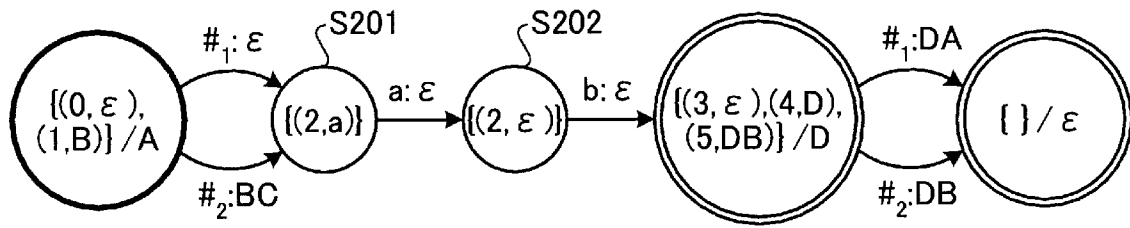
[図14]



[図15]



[図16]



[図17]

```

1   $W \leftarrow \bigcup_{q \in F_1, (q, w) \in q_2, w \neq l^\#} w$ 
2  if  $W = \phi$ 
3     $F_2 \leftarrow F_2 \cup \{q_2\}$ 
4  foreach  $w \in W$ 
5     $q_{n2} \leftarrow \phi$ 
6     $E_2 \leftarrow E_2 \cup \{(q_2, \text{aux}(q_{n2}, w), w, q_{n2})\}, \Sigma_2 \leftarrow \Sigma_2 \cup \{\text{aux}(q_{n2}, w)\}$ 
7    if  $q_{n2} \notin Q_2$ 
8       $Q_2 \leftarrow Q_2 \cup \{q_{n2}\}, F_2 \leftarrow F_2 \cup \{q_{n2}\}$ 

```

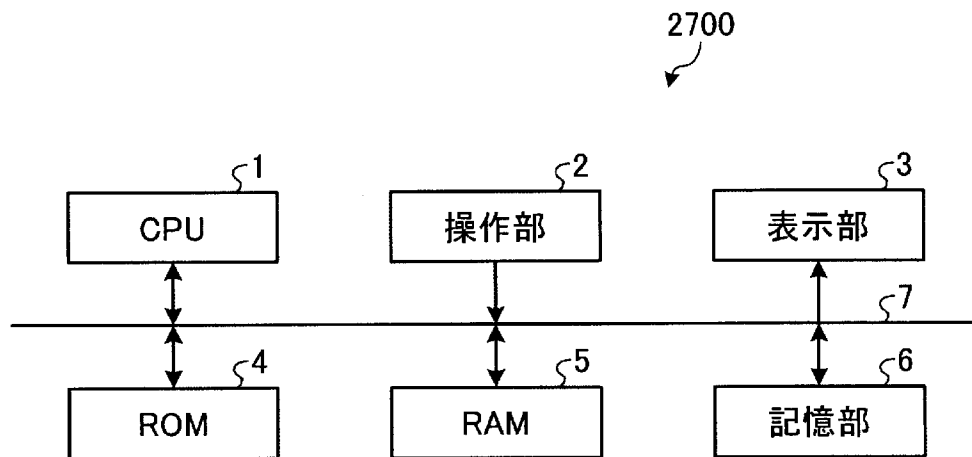
[図18]

```

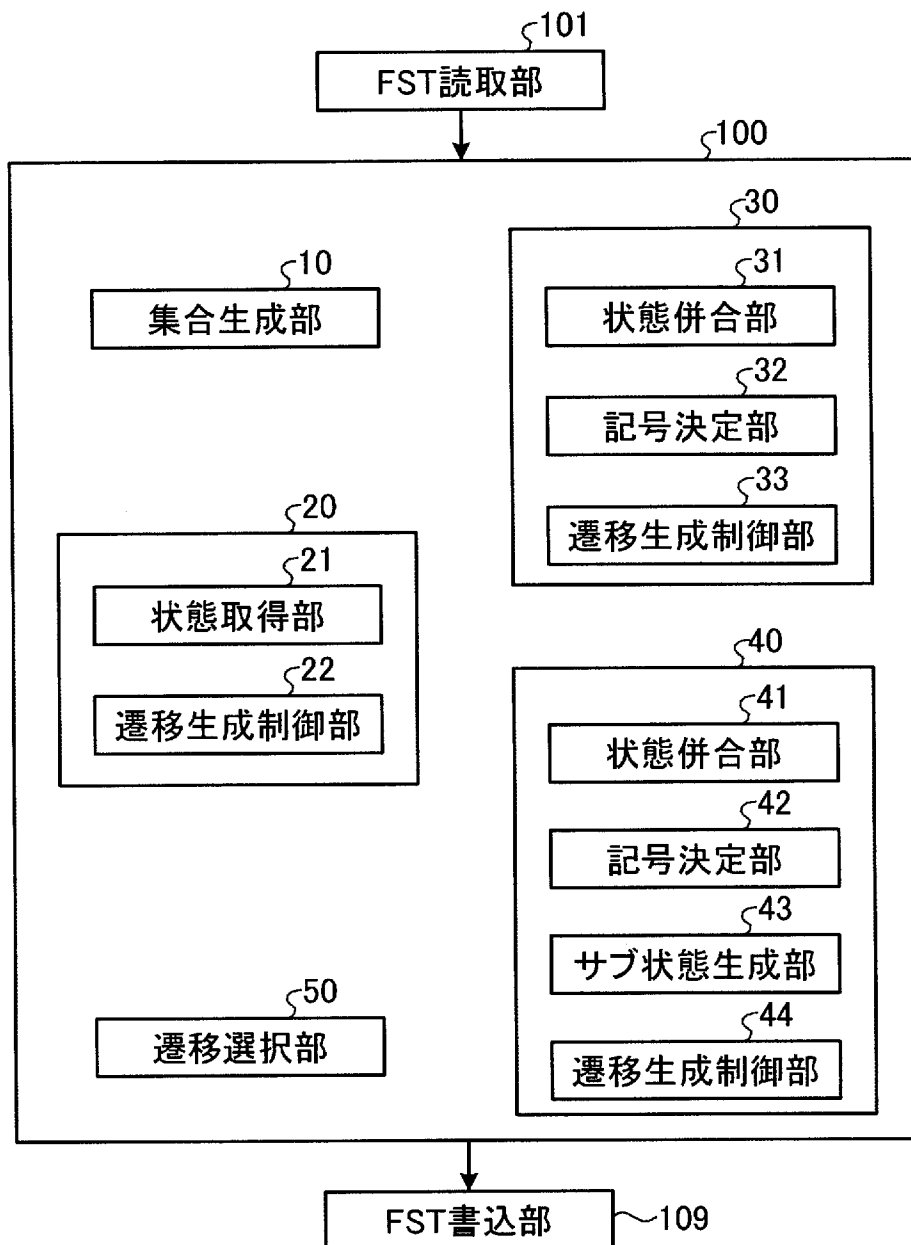
1   $F_2 \leftarrow F_2 \cup \{q_2\}$ 
2   $W \leftarrow \{x | x = w \otimes y, x \neq 1^\#, y \in \rho_1(q), (q, w) \in q_2, q \in F_1\}$ 
3  if  $|W| > N$ 
4     $q_{n2} \leftarrow \phi$ 
5    if  $q_{n2} \notin Q_2$ 
6       $\rho_2(q_{n2}) \leftarrow \phi, Q_2 \leftarrow Q_2 \cup \{q_{n2}\}, F_2 \leftarrow F_2 \cup \{q_{n2}\}$ 
7    while  $|W| > N$ 
8      let  $w : w \in W$ 
9       $E_2 \leftarrow E_2 \cup \{(q_2, \text{aux}(q_{n2}, w), w, q_{n2})\}, \Sigma_2 \leftarrow \Sigma_2 \cup \{\text{aux}(q_{n2}, w)\}$ 
10      $W \leftarrow W \setminus \{w\}$ 
11   $\rho_2(q_2) \leftarrow W$ 

```

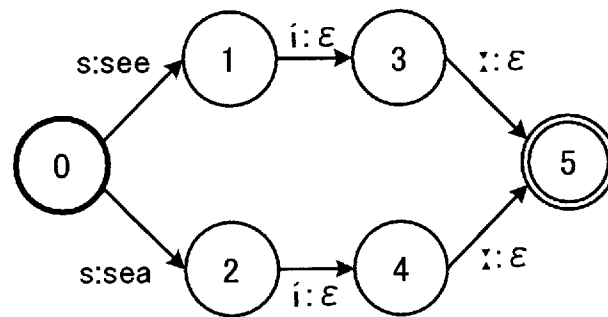
[図19]



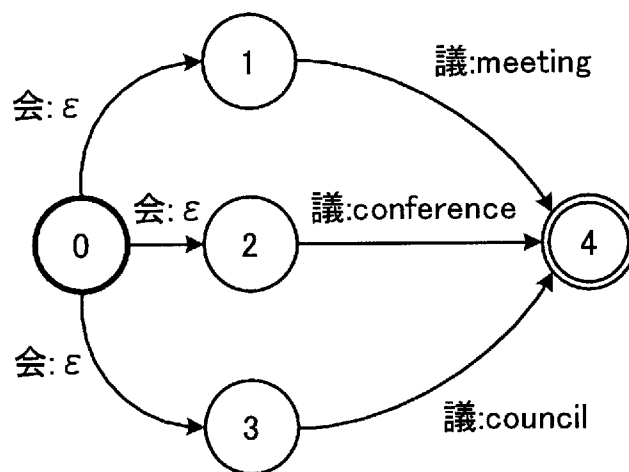
[図20]



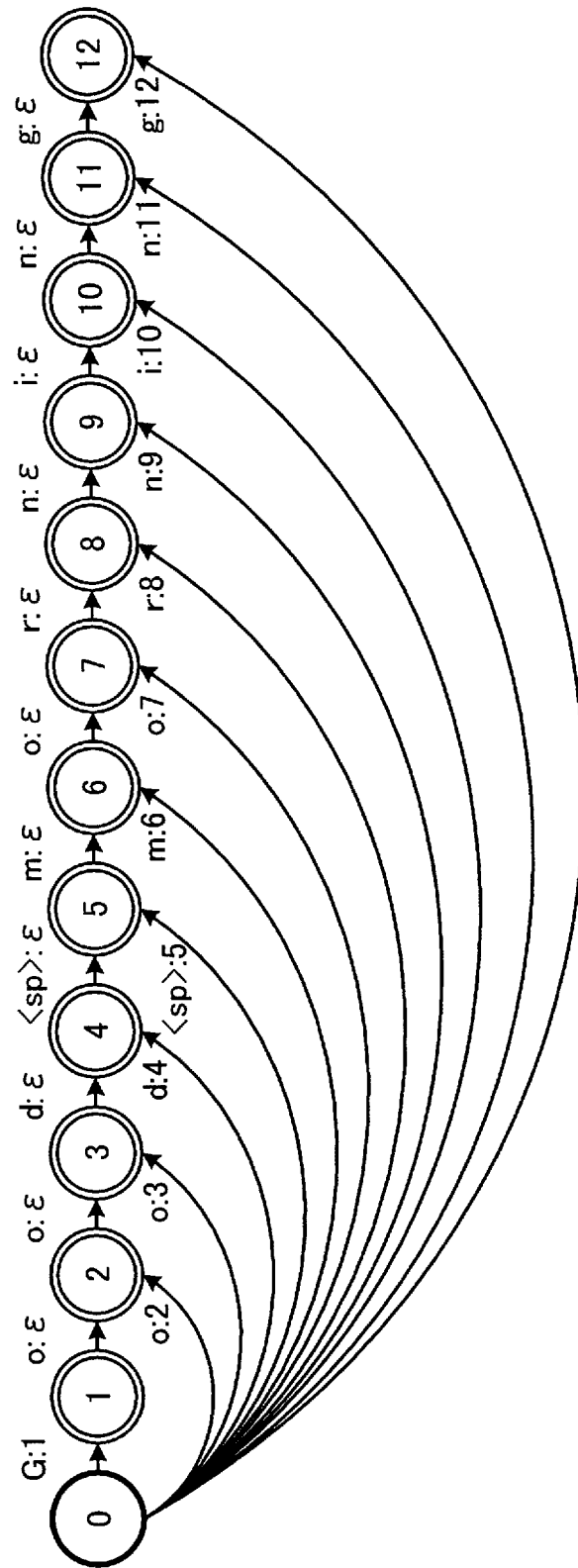
[図21]



[図22]



[図23]



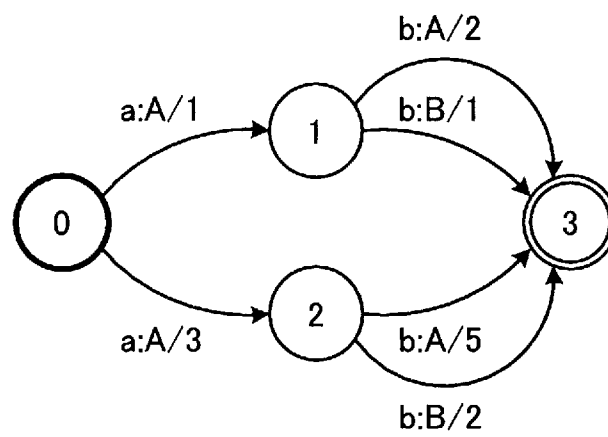
[図25]

```

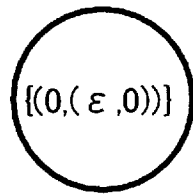
aux(q,v)
1  (x,y) ← v
2  return aux0(q,x)

```

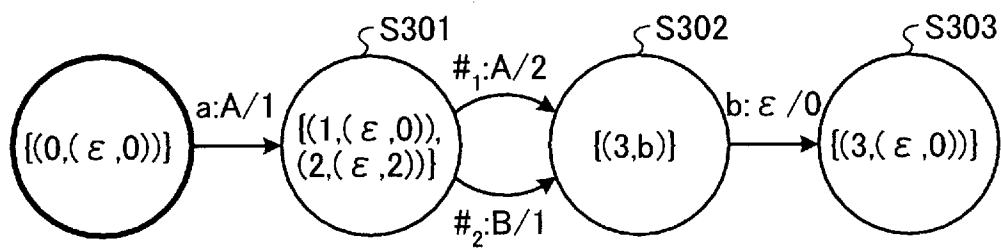
[図26]



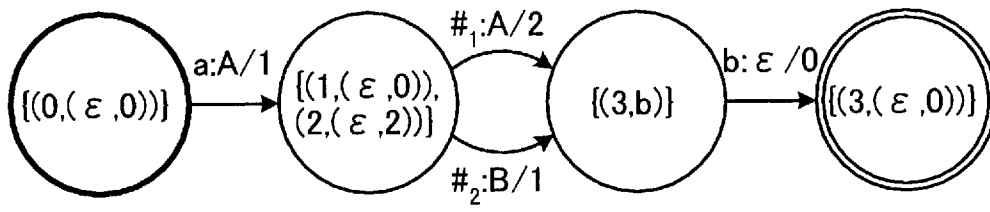
[図27]



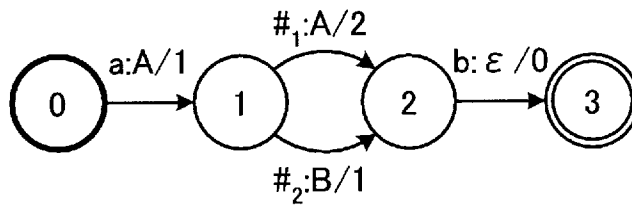
[図28]



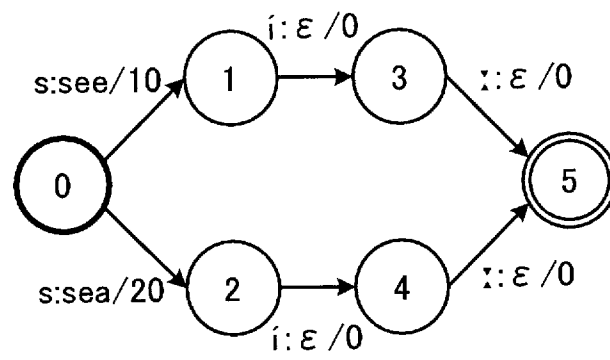
[図29]



[図30]



[図31]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2009/068016

A. CLASSIFICATION OF SUBJECT MATTER

G06F7/00(2006.01) i, G10L15/18(2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F7/00, G10L15/18

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2009
Kokai Jitsuyo Shinan Koho	1971-2009	Toroku Jitsuyo Shinan Koho	1994-2009

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2006-031403 A (Nippon Telegraph And Telephone Corp.), 02 February 2006 (02.02.2006), paragraphs [0054] to [0058]; fig. 8, 13 to 14 (Family: none)	1-10
A	JP 2009-058989 A (Toshiba Corp.), 19 March 2009 (19.03.2009), entire text; fig. 2 to 3, 5 to 6 (Family: none)	1-10

☐ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search
07 December, 2009 (07.12.09)

Date of mailing of the international search report
22 December, 2009 (22.12.09)

Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

A. 発明の属する分野の分類（国際特許分類（I P C））

Int.Cl. G06F7/00(2006.01)i, G10L15/18(2006.01)i

B. 調査を行った分野

調査を行った最小限資料（国際特許分類（I P C））

Int.Cl. G06F7/00, G10L15/18

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1 9 2 2 - 1 9 9 6 年
日本国公開実用新案公報	1 9 7 1 - 2 0 0 9 年
日本国実用新案登録公報	1 9 9 6 - 2 0 0 9 年
日本国登録実用新案公報	1 9 9 4 - 2 0 0 9 年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	JP 2006-031403 A（日本電信電話株式会社）2006.02.02, 段落【0054】－【0058】、第8図、第13-14図（ファミリーなし）	1-10
A	JP 2009-058989 A（株式会社東芝）2009.03.19, 全文、第2-3図、第5-6図（ファミリーなし）	1-10

C欄の続きにも文献が列挙されている。

パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」特に関連のある文献ではなく、一般的技術水準を示すもの
「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）
「O」口頭による開示、使用、展示等に言及する文献
「P」国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの

「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの

「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの

「&」同一パテントファミリー文献

国際調査を完了した日

0 7 . 1 2 . 2 0 0 9

国際調査報告の発送日

2 2 . 1 2 . 2 0 0 9

国際調査機関の名称及びあて先

日本国特許庁（I S A / J P）

郵便番号100-8915

東京都千代田区霞が関三丁目4番3号

特許庁審査官（権限のある職員）

緑川 隆

5 E

2 9 5 0

電話番号 0 3 - 3 5 8 1 - 1 1 0 1 内線 3 5 2 1