

プロダクションシステムにおける最大・最小値の取扱いと リアルタイム機能への応用

竹中 一起、横井 玉雄

住友金属工業㈱ システムエンジニアリング本部

プロダクションシステムの枠組の中で、対象データの最大値とか最小値を扱う場合、高速のパターンマッチング手法であるReteアルゴリズムを採用したとしても、データの選出にはかなりの計算時間を要していた。本稿では、最大・最小値の取扱いを効率化するために、Reteネットワーク内でWMEを順序列として管理する方式を提案し、Reteネットワーク上での実現方法を示すとともにその効果について述べる。さらに本方式を拡張すれば、サンプリングや時間概念を有するデータについても、Reteネットワーク内で簡潔にかつ高速に処理できることを示す。

Effective Approach to Maximum/Minimum Data in Production System
and Application to Realtime Environment (in Japanese)

Kazuki TAKENAKA and Tamao YOKOI

Sumitomo Metal Industries, Ltd.

1-3, Nishinagasu-hondori, Amagasaki, Hyogo, 660, Japan

In the production system, it takes so much time to deal with the maximum or minimum data of objects, even if Rete match algorithm is adopted. In this paper, we propose a new method by which WMEs are managed orderly in Rete network, in order to process those data efficiently. Then we show how to implement the method on Rete network, and discuss the effect. Finally we show that the method can be successfully applied to sampling or time-based data.

I. 緒言

手続き型言語によるコンピュータシステムの限界を打ち破るものとして、AI手法、特にエキスパートシステムの枠組が注目されている。我々もプロセス制御システムのより一層の高度化・知的化を図るべく、早くからこの技術に着目し、プロセス制御用計算機上で稼働するエキスパートシステム構築ツールMARKS-II[1]を独自に開発し、社内応用システムへの適用を進めてきた。

MARKS-IIでは、推論方式としてプロダクション型と分類型の2種類の推論エンジンを備えている。プロダクション型についてはReteアルゴリズム[2]を採用して、パターンマッチング処理を効率化し、さらに数々の独自の高速化手法[3]を駆使することにより、実時間制御に適用するに十分な高速性を実現している。

ところで、特に設計や計画問題に関する知識を整理してみると、例えば「重量比が最小の材料に関して、含有成分が最大の元素が◇◇であって、その含有量が△△であれば、――。」というように、ある要素の値が最大とか最小とかを扱う知識が非常に多いのに気付く。ところがReteアルゴリズムでは推論データであるワーキングメモリ・エレメント(WME: MARKS-IIでは「知識データオブジェクト」と呼ぶ)を順序列として管理していないため、最大データや最小データの選出にはかなりの計算時間を要していた。そこで最大・最小の取扱いを効率化することを目的として、Reteネットワーク内でWMEを順序列として管理する方式を提案する。

本稿では、まず今回の提案のベースとなるReteアルゴリズムについて簡単にレビューした後、最大・最小の取扱いについてReteアルゴリズムに内在する問題点を指摘する。次に今回提案する新しいWMEの管理方式とマッチング・アルゴリズムについて述べ、Reteネットワーク上での実現方法を示すとともにその効果について評価する。そして最

後にリアルタイム機能への応用として、本方式を拡張すれば、サンプリングや時間概念を有するデータについても、Reteネットワーク上で簡潔にかつ高速に処理できることを示す。

II. Reteアルゴリズム

Reteアルゴリズムとは、1回のルール実行によって変化するWMEの数は通常極く少数であること、従って条件部の成立可否をチェックし直す必要のあるルール数はそう多くないことに着目し、変化したWMEに関係するルールだけにマッチング処理を限定することにより推論を効率化することを狙ったものである。さらに複数のルール内に現れる同一固定値のチェックを融合し、また前回までのルール実行で部分的にマッチングしている情報を保持しておくことによってマッチング処理の回数を激減させている。

具体的には知識ベースの全ルールを、図1に示すようなネットワークに展開し、WMEを「トークン」という形でネットワークに流すことによりルールのマッチング処理を行っている。このときトークンの頭に、データ生成のときは+、消去の時は-の符号をつけてネットワークに流す。

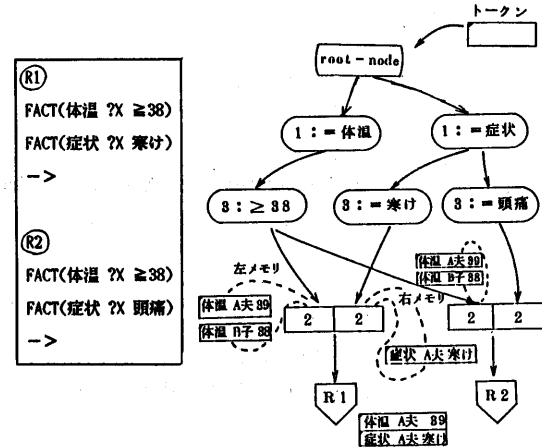


図1. Reteネットワーク

表1. Reteネットワークのノード

名称	ルールとの対応	主な処理
ルート・ノード □	ルールの集合 (知識エント)	ルールの実行部を「トークン」という型にして、全ての子ノードに流す。
イントラ・ノード ○	パターン内の 固定データのチェック	流れてきたトークンが 条件を満たすとき 下流にトークンを流す。 条件を満たさないとき トークンを消す。
インター・ノード □	パターン間のチェック ・変数のバインド ・パターン間での一致チェック (ルールのパターン毎に1個)	①流れてきたトークンを、その方向に応じて右メモリ or 左メモリに格納する。 ②反対側のメモリ内(右から流れてきた場合は左メモリを探す)に格納されているトークンと条件のチェックを行い、満足していれば2つのトークンを合成して下流に流す。
プロダクション・ノード ▽	条件部の成立 (ルール毎に1個)	本ノードに達したトークンは条件を成立させるWMEの組であるので、競合集合に登録する。

ネットワークのノードには4種類があり、各ノードでの処理を表1にまとめる。

Ⅲ. 従来方式の問題点

(1) 知識ベース記述上の問題点

パターンマッチング処理にReteアルゴリズムを採用し、OPS5[4]並みの記述力をもつツールでは、膨大な量のWMEからあるパターンの条件を満たすWMEを選び出し、あるいは複数のパターンを共通に満足するWMEを選別したりする方法としては、基本的にはWMEのデータ部の一致とか比較・範囲指定（イントラノード）、複数WME間の指定要素の一致や比較（インタノード）、あるいはそれらのAND/OR結合ぐらいいしか提供されていない。従ってある条件を満足するデータのなかで指定要素の値が最大というのを表現するには、まずある条件を満足するWMEを任意の一つを選び出し、次に同一条件を満足する他の全てのWMEと比較して、それより大きい値を持つものはないと記述せねばならない。すなわちMARKS-IIの文法に沿って記述すれば、

```
FACT (優先度 ?材料 ?優先度) ..... a
~FACT (優先度 ?材料 > ?優先度) ..... b
```

[~はそれに続くデータが存在しないことを意味する]となる。これは利用者の立場からは知識ベースが理解しづらく、内部処理的には多大の計算量を要するという問題点がある。いま「優先度」という名のWMEがN個あったとする。上記bの比較処理は、aを満足するデータ1個毎に最大N回ずつ行われる為、結局Nの2乗のオーダーの比較演算が必要となる。

(2) Reteアルゴリズムに内在する問題点

前節では、最大値や最小値を持つWMEを選出するのにN（指定の条件を満たすWMEの総数）の2乗のオーダーの計算量を要すると述べたが、Reteアルゴリズムでは部分的なマッチングの状態がインタノードに保存されていることより、実際には新規に追加あるいは削除されたWMEについてのみ、マッチング処理をやり直せば良い。Reteアルゴリズムでは、WMEの任意の要素に設定された固定データによるチェック（データの一致や範囲指定を調べるもので、イントラノードで処理される）を通過したトークンは、インタノードに蓄積された後、パターン間のマッチング処理の対象となる。このときトークンは、初めて流れ込んだインタノードの右メモリに、またトークン間でマッチングが確立した場合はそれらの合成トークンが直後のインタノー

ドの左メモリに格納される。Reteアルゴリズムではこれらのトークンの右/左メモリへの登録順については、何ら言及しておらず、通常は競合解消戦略の関係から到着順に蓄積される。このため新規に追加/削除されたWMEについてのみマッチング処理をやり直せば良いとは言っても、データ群が値の大小順に順序列を構成していないため無駄な処理を多々含むことになる。

以上についてさらに詳しく検討を加える。

*RULE 搬送ルール

```
FACT (搬送指示 ?材料 ?F置場 ?T置場 ..... ) a
FACT (優先度 ?材料 ?優先度) ..... b
~FACT (優先度 ?材料 > ?優先度) ..... c
->
```

これはある材料搬送システムにおいて最大の優先度を持つ搬送指示情報を選び出すルールである。いま下図のように搬送指示FACTがM個、優先度FACTがN個あるとし、部分的なマッチング状況がインタノードに保存されているものとする。また簡単のために優先度FACTの優先度は全て異なるものとする。

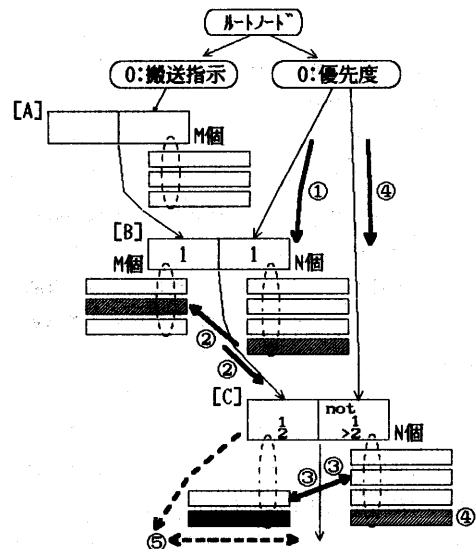


図2. 最大値のマッチング処理

ここで優先度FACTが新規に登録された場合を考える。この場合は、+の符号を持つトークンがルートノードより流れ、以下のように処理が進行する。

- ①bのパターンには固定値チェックがないため、FACTのワイルド外名のチェックの後、[B]のインタノードに達し、まず右メモリに格納される。
- ②次に反対側の左メモリに格納されているトークンとパターンマッチングを行う。一致するものがあればそれらの合成トークンを作り下流に流す。
- ③②で作成された合成トークンは次の[C]の左メモリに格納された後、反対側の右メモリと第一要素の一致と第二要素の比較処理を実施する。
- ④b、cについては同一種類のWMEに関する条件であることから、①のワイルド外名チェックのインタノードが分岐し、今回新規に作成された優先度FACTは、[B]と共に[C]にも流される。[C]では、[B]と同様まず右メモリにトークンが格納される。
- ⑤cは「存在しない」ことを意味するNOTパターンであることから、この場合は反対側の左メモリを調べ、右メモリ側に一致と比較の条件を満たすFACTが存在しないことによりNOT条件を満足している合成トークンをサーチする。条件を満たすトークンが存在すれば、今回の新規FACTの追加によりこの条件の成立が侵害されていないかを調べる。条件を満足できなくなった場合には、下流のノードに対して、その旨を知らせるトークンを順次流す。

次に計算量について考察する。オリジナルのReteでは、通常インタノードへの到着順で右/左メモリのトークンの管理を行っている為、①については右メモリの末尾にトークンをつなぐだけで処理が終了する。②については左メモリの全てのトークンとマッチング演算をする必要がある為、M回の演算が行われる。③の処理は②で一致したものについてのみ行われるが、一致したものの1個につき最高N回のマッチング演算が必要となる。⑤ではN回のサーチ演算が行われる。結局、M+Nのオーダの計算量となる。

今度はある優先度FACTが削除された場合を考える。この場合は-の符号を持つトークンがネットワークに流される。

- ①WMEの削除を意味する-トークンは、FACTのワイルド外名のチェックの後、[B]のインタノードに達する。今回は-トークンの為、右メモリ内に同一内容のトークンがあるかを調べ、存在する場合はそのトークンを消去すると同時に反対側の左メモリにマッチングしているトークンがないか調べる。ここでは右メモリに関して高々N回と左メモリに関してM回のマッチング演

算が行なわれる。

- ②①の左メモリにマッチングするものがある場合は、-の符号を持つ合成トークンが[C]に流され、左メモリから同一内容を持つトークンを消すと同時に、それが反対側の右メモリとマッチングしていなければ、さらに下流に対して合成トークンを流す。
- ④-トークンは、[C]にも流される。[C]の右メモリに関しては、①と同様、同一内容を持つトークンが消去されるが、[C]ノードは、NOT処理を行うノードの為、左メモリの処理については壊滅的な事態が発生する。即ち今回の-トークンにより、条件成立の障害となっていたトークンがなくなったかも知れないため、全ての左要素に対して、右要素とのマッチングをやり直す必要がある。いま優先度FACTがN個あるとすると、Nの2乗のオーダの再計算が必要となる。

IV. 新方式による解決策

(1) 最大値・最小値の取り扱い

最大・最小に関する以上の問題点は、Reteアルゴリズムでは部分的なマッチングの状況をインタノードに保存しているものの、大小関係による順序づけた管理を行っていないことに起因していた。そこで本稿では、インタノードの左あるいは右メモリにおけるトークンを、注目する要素の値の昇順あるいは降順で管理することにより、知識ベース記述の簡潔化とマッチング処理の高速化を同時に達成する方式を提案する。

『インタノードの右メモリ、あるいは左メモリごとに、各々以下の情報を持ち、最大値・最小値指定の場合は値の降順あるいは昇順で、例えば2重リンクリストのような柔軟なデータ構造によってトークンを管理する。

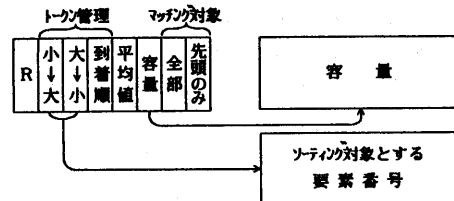


図3. ノードメモリの管理情報

また最大値・最小値指定の場合は、先頭に位置するトークンのみをマッチングの対象とする。但し、2番目以降が先頭と同一値の場合は、2番目以降であってもマッチングの対象とする』

れており、両者が一致する必要があるが、この場合 a パターンでは A, b パターンでは B とバインドされている為、結局条件部は成立しなくなる。これは、前節で示した !MAX, !MIN は、固定データによるチェックを満足した (特定のイントラノードの並びを通過した) WME 群の中で指定した要素が最大あるいは最小の値をもつ WME を選出するものであった為である。

そこで新たに #MAX と #MIN という 2 つの記述子を導入する。これは意味的には図 4 に示すように、指定したパターンまでの間でパターンマッチングをした WME の組から、ある要素に着目して最大あるいは最小の値 (即ち部分的にマッチングした部分集合内での最大・最小) を持つデータ組を選出するものである。すなわち ! はイントラノードを通過したデータを対象としているのに対し、# はインタノードに於けるマッチングが成立したデータを対象としたものである。

(3) Rete ネットワーク上でのインプリメント

! については、イントラノードを通過したデータを対象としたものであり、2 入力 1 出力の形をとるインタノードに対しイントラノードからは必ず右方向から装入するので以前に示したように、右メモリを昇順あるいは降順にソーティングして管理することにより容易に実現できた。

これに対し、# はインタノードに於けるマッチングが成立した WME の組を対象としたものである。ところで、あるインタノードでマッチングが成立した場合は、それらの合成トークンが作られ下流に流される。下流はもちろんインタノード (プロダクションノードも含む) であり、その左方向から装入される。すなわち # の場合は、ルールで記述されたパターンの次パターンに対応するインタノードの左メモリを、指定要素の昇順あるいは降順にソーティングして、管理すれば良いことになる。

以上まとめると、インプリメント上は、

!MAX, !MIN :

指定パターンに対応するインタノードの右メモリ

#MAX, #MIN :

指定パターンの次パターンに対応するインタノードの左メモリ

を、昇順あるいは降順で管理すればよい。

(4) 知識ベースの記述

MARKS-II では、FACT, フレーム, プログラム変数等、全ての知識データオブジェクトに関して、ある要素が

最大とか最小とかを表現する場合

最大 (極大) : !MAX #MAX

最小 (極小) : !MIN #MIN

という記述子を使用することにより簡潔に記述することができる。たとえば

FACT (関係 10 ? !MAX ? x)

は、第 1 要素が 10 という値を持つ「関係」という名の FACT のうち、第 3 要素の値が最大のものを意味する。ここで、! は、各々のルールパターンに関して固定データ指定による条件を満足したオブジェクトの中で指定要素が最大あるいは最小を意味し、# は各ルールの先頭から指定パターンまでの範囲でパターンマッチングしたオブジェクトの組の中で、指定した要素に注目した場合の最大 (極大) あるいは最小 (極小) を意味する。

但し、

① 数値型の要素またはスロットに限る。

② 同一パターン内で記述できる個数は、(!MAX, !MIN) のうちの 1 個と、(#MAX, #MIN) のうちの 1 個、計 2 個までである。

③ NOT パターン内では、!, # とも使用不可。

④ 最大値や最小値をバインドして後方で使用したい場合は、従来通り ? 変数を使用する

例) !MAX & ? x

次に知識ベース記述上の注意として、一見同じ意味をもつように見受けられるが、実は全く別の意味を持つ記述例を挙げる。

[A] FACT (成分 ?x ?y >0.08 & !MIN & ?z)

[B] FACT (成分 ?x ?y !MIN & ?z)

TEST (?z > 0.08)

[A] は 3 番目の要素が 0.08 以上のものの中での最小値を意味し、図 4 のデータの場合は ?z に 0.1 がバインドされる。これに対し、[B] は 3 番目の要素の最小値が 0.08 より大きいかわりを意味し、この場合は ?z に 0.02 がバインドする為、条件は成立しない。

(5) 新方式の効果

! 単体の定量的評価については既に述べた。実際レベルの総合効果については、パターンマッチング処理が、今回の ! や # の出現頻度、ルール数や各ルール毎のパターン数、WME の数、それにルールの記述方法などに依存し、これらの要素が複雑に絡み合うため、定量的に評価することは極めて困難である。本機能を適用した実システムから、経

験的に得られた定性的な効果について以下に列挙する。

- ①一般的には同一のワヅェット名を持つWMEの数が増大するほど効率が低下するが、今回の！や井を導入すればWME数への依存度が小さくなる。特にWMEが多い大規模システムほど効果が大きい。
- ②余分なインタノード（NOT）がなくなり、かつ無駄な部分マッチングが防止されるため、自由メモリの大幅な節約になる。
- ③記述力が向上し、知識ベースの見通しが良くなる。自然な表現となる。
- ④井よりも！の方が速度面、メモリ面とも効果が大きい。これは！がインタノードにおけるマッチング前、井がマッチング後を対象にしていることから明らかである。

V. リアルタイム機能への応用

(1) 従来型システムとの協調

MARKS-IIでは、従来型制御システムと協調した複合型のESを容易に構築できるよう、設計当初から従来型システムとエキスパートシステムとの密接なデータリンクを非常に重視して開発を進め、現在は表2に示す3種類の型態[5]を提供している。なお、これらは混在が可能で、応用システムのニーズに応じて組合せて使用できる。

表2. 従来型システムとの協調型態

タイプ	エキスパートシステム	従来型システム
Calling	MARKS-II	手続き
Polling	MARKS-II (データ変更時、推論を起動)	監視 データ領域 設定 データ収集タスク 読出し
Called	MARKS-II	起動 生成/消去 ドライバタスク

Calling型は、ルールの右辺（実行部）で、従来型システムの手続きを直接呼び出せる型のものである。

Polling型は、プロセスデータ等に割り付けられた従来型言語の変数を、ESのルールから直接参照したり更新したりできるもので、MARKS-IIの大きな特徴の一つである。表2で、データ収集タスクは、なんらかの外部要因によって割込起動され、プロセス信号入出力装置からデータを取り込み、データ領域にセットしている。推論エンジンはこのタスクとは独立に動作し、周期的にデータ領域をポーリングする。そして外部の変数値と対応する知識データオブジェクトの値に相違があれば、その変更内容を内部

に反映し、条件部を満足するルールを実行する。ルールを記述する専門家は、知識ベースの最初の部分でこれらの変数のデータタイプを宣言するだけで良く、外部データを獲得する際の複雑で微妙なタイミングを考慮しなくて済む。

Called型は、従来型システムからESを、手続き呼出しの形で呼出すことができるものであり、以下の3種類の手続きを装備している。

- 1° 推論の起動
- 2° 知識データ・オブジェクトの生成 (assert)
- 3° " 削除 (retract)

この型のリンク方法を使えば、従来型システムからESを、必要なときにいつでも起動することができる。

(2) 周期データの取扱い

外部データのハンドリングは、Polling型とCalled型の2つの型態で可能である。Polling型はデータ自体は簡素な型だが、データ収集処理が複雑な場合に、またCalled型は比較的多量の情報を含むデータを扱う場合にと使い分けられる。このうちPolling型は、外部データの変化をMARKSが常時監視し、ルールの条件部が満足された時点で推論実行を起動するタイプであるが、運用上次の問題点を含んでいる。

①推論実行中は、Called型による推論起動時、知識ユニット切換え時、および蓄積トークンバッファ[3]満杯時に外部データをポーリングしているが、このタイミングを外部からは把握しにくい。

②推論のアイドル中は一定周期毎に、外部データをポーリングしており、CPU負荷が重い。

そこで今回、Polling型のプログラム入力変数について周期データの取扱いを新設することにする。これは、知識ベース記述のデータ定義の際に個々のプログラム入力変数ごとに独立したサンプリング周期の指定を可能とし、ルール条件部で、ある外部データの

- a n周期前のデータ
- b 過去n周期間の最大（最小）値
- c 過去n周期間の平均値

を扱えるように拡張するものである。推論実行時、推論エンジンは指定されたサンプリング周期毎にプログラム入力変数の値を取込み、タイムスタンプをつけてReteネットに流す。なお周期指定されたプログラム入力変数については、上記①②のタイミングでのデータ監視は行なわない。

表3. Reteネットワークにおける周期データの取扱い

種別	n 周期前	過去 n 周期間の最大(最小)値	過去 n 周期間の平均値
ノード型態			
トークン管理	到着順	降順 または 昇順	到着順 + ダミートークン(常時先頭)
マッチング対象	先頭のみ	先頭のみ	先頭のみ (再マッチングは合計値の 変化時のみ)
トークン受信時の 主な処理	<p>・右メモリの容量は(n+1)とし、上流からトークンが流れてくるとメモリリンクの末尾にトークンをつなぎ、同時に先頭のトークンをリンクからはずす。</p>	<p>・右メモリの容量は(n+1)とし、上流からトークンが流れてくるとリンク中の適切な場所に格納する。同時に最古のトークンをリンクからはずす。</p>	<p>・メモリリンクの先頭に、トークンの合計値を保持する為のダミートークンを設ける。 ・容量は(n+1)+1=n+2とする。 ・トークンをリンクにつなぐ場合は、ダミートークンの合計値を加算し、はずすときは減算する。</p>

(3) 周期データのインプリメント

Reteネットワークに於ける周期データの取扱いは、今回の最大・最小機能を応用すれば、インプリメントは表3に示すように比較的容易に行える。

今回は周期データの取扱いはプログラム変数のみとするが、他の型の知識データオブジェクトについても同様に拡張可能である。ところで通常のReteネットワークでは、イントラノードを通過できないトークンは消去されるが、今回の場合は指定の時間範囲をトークンの個数で管理しているため、イントラノードの条件を満足しない場合でもインタノードにダミーのトークンを流す必要がある。ダミートークンは個数の管理にのみ使用するもので、イントラノードではチェックなしに下流に流し、インタノードでは右メモリに保存するもののマッチングの対象とはしない。

VI. 結言

プロダクションシステムの枠組の中で、対象データの最大値とか最小値とかを扱う場合、高速のパターンマッチング手法であるReteアルゴリズムを採用したとしても最大データや最小データの選出にはかなりの計算時間を要していた。そこで本稿では、Reteネットワーク内でWMEを順序列として管理することにより、最大・最小の取扱いを大幅に効率化できる方式を提案し、Reteネットワーク上での実現方法を示すとともにその効果について述べた。さらにリ

アルタイム機能への応用として、本方式を拡張すれば、サンプリングや時間概念を有するデータについても、Reteネットワーク内で簡潔にかつ高速に処理できることを示した。

本稿で述べた最大・最小機能は、既にインプリメントを完了し、大規模な搬送スケジューリングシステムに適用して、その有効性を確認している。またリアルタイム機能への応用については、現在、詳細仕様の詰めを行っている。

【参考文献】

- [1]横井、竹中:「プロセス制御向エキスパートシステム構築ツール(1)-概要-」、情報処理学会第35回全国大会,1987.
- [2]Forgy,C.L.:「Rete:A Fast Algorithm for the Many Pattern/Many Object Pattern Matching Problem」, Artificial Intelligence 19,1982.
- [3]竹中、横井:「プロセス制御向エキスパートシステム構築ツール(2)-高速推論機構-」、情報処理学会第35回全国大会,1987.
- [4]Forgy,C.L.:「OPSS User's Manual」, Technical Report CS-81-135,Department of Computer Science,CMU,1981.
- [5]K.Takenaka et al.:「Expert System Building Tool for Proccss Control:MARKS-II」, International Workshop on Artificial Intelligence for Industrial Application,1988.