

分散型問題解決システムの一考察

山崎 晴明

沖電気工業(株) 総合システム研究所

本稿では、分散型問題解決システムのプロトタイプとして開発した SD³(system for Distributed Database with Deductive Search mechanism) システムについて報告する。本システムは、リレーショナルデータベース上の演繹検索を分散型で実行するものである。さらに本稿では SD³ システムにおける演繹検索動作の概要を述べるとともに、通信/処理の時間比が与えられた場合の全解探索問題における最適知識割り当て法、一般の分散型本探索システムにおける知識の編成タイプとその探索アルゴリズム等についても論ずる。

A PROTOTYPE SYSTEM FOR DISTRIBUTED PROBLEM SOLVING

HARUAKI YAMAZAKI

OKI electric Industry Co., Ltd, Systems Lab. 10-12, Shibaura 4-chome Minatoku, Tokyo 108, Japan.

The system called SD³ (System for Distributed Data with Deductive Search mechanism) has been implemented as a prototype for a distributed problem solving system. This system executes the deductive search on the relational databases in a distributed manner. In this paper, we describe the outline of this system, and discuss the optimal knowledge allocation for the exhaustive search problem provided that the communication/processing time ratio is given. Furthermore, we also discuss the variation of organizational types of distributed search systems, and the difference of the search algorithm for each type.

1. はじめに

関係データベース上に演繹推論機構を載せ、データベース機能の強化をはかるアプローチが、近年注目されている。筆者等は、こうした機能の実現を目指して、まず単体での演繹推論機構を持った関係データベースシステムの開発を行った。

次にシステムの応答性、信頼性等を向上させるためそれらを LAN で結合し、分散型で演繹検索を行なわせる SD³ (System for Distributed Database with Deductive Search mechanism) と呼ぶシステムを開発した。これは、一般の問題解決や意志決定支援といった広汎な応用が期待される分散型問題解決システムのプロトタイプを目指したものである。以下に SD³ システムの概要を示す。

2. 開発システムの概要³⁾

システムの全体構成を図1に示す。図1においてAPは具体的問題に依存した手続き部分である。各ノードは演繹探索を実行する DDS とデータベースの検索を実行する DBS とから成る。各ノードの DBS を結合することによって分散データベースシステムが構成され、さらに DDS を統合させることにより、分散型演繹データベースシステムが構成される。なお、これらのノードを結合するローカルネットワーク BANET は、同報通信をベースに構成された独特のネットワークアーキテクチャであり、コミットメント制御機能をはじめ分散データ処理を行ううえで極めて重要な機能をネットワーク中に取り込んでいる^{2,2)}

図2は、各ノードにおけるモジュールの論理構成を示したものである。図2において、ICM は、分散型問題解決の protocols を実行し、ADN を用いて

他ノードからの受信情報が自ノードに関連するか否かの決定、自身、導出した情報の他ノードへの送出等を行う。

EVL は、探索戦略を実行する。その際、評価関数に沿った探索、全解探索のいずれも実行可能であり、また探索の深さも自由に設定できる。また自ノードで導出した情報が自ノードでは処理できない場合は、それを ICM に渡す。

IDB には、ホーン節で記述したルール群、関数定義等が格納されており、HCI が EVL より与えられるゴール節をもとに演繹探索を実行する。なお探索の戦略 (e.g. depth/breadth first, 評価値による枝刈り等) および探索の深さは EVL が DBM を介して TDB に書き込むことで、HCI に指示がなされる。

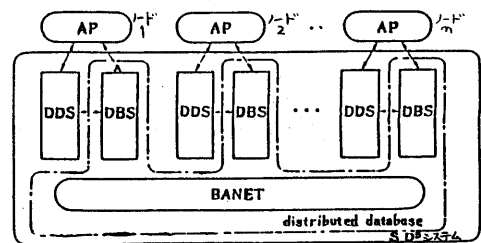


図1. SD³システムの全体構成

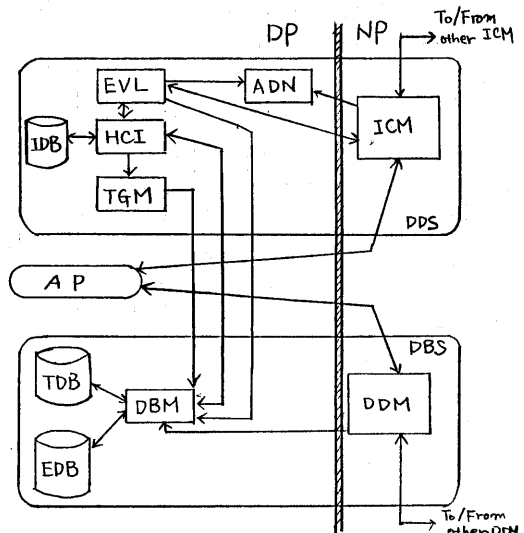


図2. ノード内モジュール構成

また EDB には探索を実行するために必要なソフト集合が、TDB には導出された中間結果および EVL により指示された探索戦略が格納される。これらを管理するのが DBM であり、さらに DDM は、分散データの検索、更新の同期等の処理を実行する。TGM は探索の途中で導出された中間結果を TDB に書き込むためのトリガ機構である。

なお各ノードは、図 2 に示すようにデータベースおよび演繹検索を行う DP (Data Processor) と、ノード間の通信処理を実行する NP (Network Processor) の 2 のプロセッサより成っている。

3. 分散型木探索システムとその評価

SD³システムの各ノードは、分散形で 1 つの組織的な木探索を行っている。本節では、こうした分散型の探索方式を形式的に扱い、評価を行うため、まずいくつかの用語を以下に定義しておく。

(1) システム内のあるノードに 1 つの中間結果が与えられたとする。その直接の後続中間結果が少なくとも 1 つそのノードで導出できるとき、与えられた中間結果はそのノードにとって 展開可能 であるという。さらにすべての後続中間結果をそのノードで導出できるとき、その中間結果はそのノードにとって 全展開可能 と呼ぶ。

(2) あるノードが展開可能な中間結果の集合をそのノードの 関連域 と呼び、全展開可能な中間結果の集合をそのノードの 達成域 と呼ぶ。

3.1 分散型木探索システムの類別

関連域と達成域という概念を用いて

分散型木探索システムの編成形態を類別することができる。

1 つは、任意の中間結果に対し、その中間結果を達成域にもつようなノードが常にシステム内に少なくとも 1 つ存在する場合で、ここではこれを 全編成 されたシステムと呼ぶことにする。

これに対し、与えられた中間結果を自分の関連域にもつすべてのノードがその展開を実行して直接の後続中間結果を導出したのち、その和集合をとれば与えられた中間結果の直接の後続中間結果をすべて導出したことと等価となるシステムもある。このようなシステムを 半編成 システムと呼ぶ。

次に分散データベースにおけるジョイン演算の例を用いてこれらの差異を論ずる。

たとえば、 P および Q をバイナリリレーションとする。いま P のオ 1 カラムをある値によって制限し、 P のオ 2 カラムと Q のオ 1 カラムとのジョインを行うキュアリ (e.g. $P(a, y) \wedge Q(y, x)$) を考える。これは深さ 1 の木の全探索とみなすことができる。

いま、 P と Q とをそれぞれ 3 つの部分 $P_1, P_2, P_3, Q_1, Q_2, Q_3$ に分割し、 (P_i, Q_i) をノード N_i に格納したとする (図 3)。このように配置したシステムでは、格納データの移動を伴わないうで解を求めることはできない。それ故これは 未編成 システムである。

一方、 P はそのまま Q を重複させ全ノードに割り当てたとする。この場合、各ノードのローカルなジョイン結果をあわせれば解が得られるから半編成システムである (図 4)。

次に P をオ 1 カラムでソートし、オ 1 カラムが同一値を持つタプルは同一ノードに收容されるよう P を再分割したとする (図 5)。これは、1 つのノードでのジョイン演算の実行により必ず解が得られるから全編成システムで

ある。

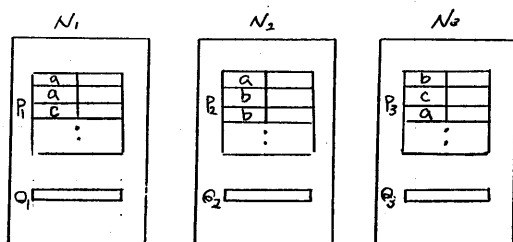


図3. 木形成システム

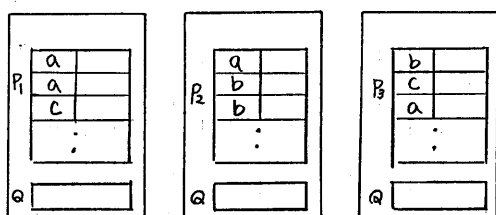


図4 半編成システム

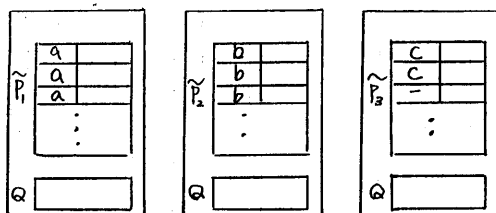


図5 全編成システム

3.2 ノード間プロトコル

分散形で木探索を実行する場合、その処理アルゴリズムおよびノード間プロトコルは、前述の編成形態に応じて異なったものと仮定することが予想される。

SD³システムにおけるノード間プロトコルは、全編成もしくは半編成のシステムを対象としており、どちらの編成に対しても同一のプロトコルの適用が可能である。以下にプロトコルの概要を述べる。

なお、ネットワークとしては同報通信を基本としたバス型のLANを想定し

ていること、またR-Q、B-Qは中間結果をリンクづけするキューであり、R-Qには展開を行っている中間結果が、B-Qには展開を他ノードに委託しようとしている中間結果がそれぞれリンクづけされることに注意されたい。

＜プロトコル概要＞

- ①あるノード N_i に問題 S_0 が与えられる。
- ② N_i は S_0 の展開を開始する。
- ③導出中間結果 S_i が N_i の達成域を越えた場合、 N_i は協調メッセージにより S_i を全ノードに同報する。但し S_i が N_i の関連域にある場合は展開を続行する。 S_i が関連域を越えてしまった場合は、それをB-Qにおく。
- ④ N_i の関連域に解以外の中間結果が存在しなくなるまで、 N_i はR-Q内にある中間結果の展開を続行する。
- ⑤ S_i を受信したノード N_j は、それが N_j の関連域にあれば優先度を付加して競合メッセージを全ノードに同報する。なお付加する優先度はシステムが全編成か半編成かによって異なる。
- ⑥全編成の場合、自ノードのR-Qにある中間結果の枚数とノード番号 N_i との和とする。半編成の場合、 S_i が達成域にあれば0、そうでなければ1とする。
- ⑦競合メッセージを送出したノードはただちに S_i をR-Qにおき、展開の対象とする。
- ⑧他ノード N_k からの S_i に対する競合メッセージを受信した N_j は、優先度を比較し、自身の方が低ければR-Qより S_i およびその後続中間結果を取り除く。優先度が同じか高ければ処理を続行する。
- ⑨他ノードからの S_i に対する競合メッセージを1つでも受信した N_i は、 S_i をB-Qより取り除く。もし S_i およびその後続中間結果がR-Qにある場合は、⑦と同様の処理となる。

図6にプロトコル概要を示す。

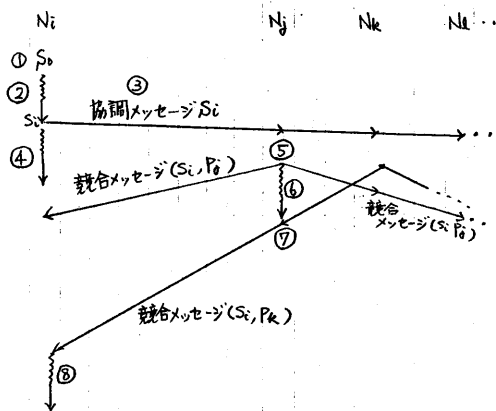


図6 ノード間プロトコル概要

3.3. 分散型木探索システムの評価

本節では評価対象として、全網成システムにおける全解探索問題をとりあげる。システムの実行効率に特に影響を与える要因として重要なものは、各ノードの保持する達成域の大きさである。達成域を大きくしすぎると並列実行による効果がほとんど得られない（問題が単一ノードに閉じてしまう）し、逆に達成域を小さくしすぎると通信オーバーヘッドが極めて大になってしまうことが予想される。この最適な達成域の大きさを、本節では展開の結果生ずる直接の後続中間結果が、もとの中間結果と同一達成域に入る確率という観点から論ずる。

(1) 前提

① 1つの中間結果を展開して直接の後続中間結果を1つ生成するのに要する時間を一定値 $u \text{ sec}$ とする。

② 1つの中間結果を別ノードに転送するために要する時間を $r \text{ sec}$ とする。したがって通信・処理時間比 c は、

$$c = r/u \quad \text{となる。}$$

③ 中間結果の展開処理と送受信処理は、図7に示すように別プロセッサによって完全に並列に実行されるものとする。

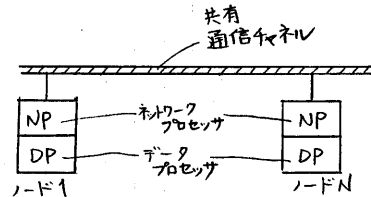


図7 通信と処理の並列実行

したがって、全体システムが最適実行状態となるためには、通信と処理の負荷がバランスしていなければならないということが予想される。

④ 確率変数 Z を1つの中間結果から発生するすべての直接の後続中間結果の総数とする。もし与えられた問題が解をもつのであれば $E(Z) = p < 1$ である。

⑤ 確率変数 X, Y をそれぞれ先行中間結果と同一達成域に属する直接の後続中間結果の総数、および異なる達成域に属する中間結果の総数とする。

⑥ $E(X) = \lambda, E(Y) = \mu$ とし、 $\sigma = \lambda / (\lambda + \mu) = \lambda / p$ とおく。 σ を探索のローカリティと呼ぶ。これは直接の後続中間結果が、先行中間結果と同一達成域に属する確率である。

⑦ 与えられた問題を O の世代、それから生じた中間結果の集合を $O+1$ 世代、... と呼び、各世代集合の要素数を、 $Z_0, Z_1, \dots, Z_n$... であらわす。但し、 $Z_0 = 1, Z_1 = Z$ である。

(2) 探索ロ-カリテ-の最適値

$\rho(z)$ を Z の分布の母関数, $\phi_X(z)$, $\phi_Y(z)$ をそれぞれ X, Y の分布の母関数とする。いま Z_m の分布の母関数を $\rho^{(m)}(z)$ で表すと, 変数 $Z_0, Z_1, \dots, Z_{m-1}$ は分枝過程 (branching process)⁴⁾ を構成するから,

$$\rho^{(m)}(z) = \rho(\rho^{(m-1)}(z))$$

が成り立つ。したがって, オ m 世代集合の要素数の平均 Z_m は;

$$Z_m = \left[\frac{d\rho^{(m)}(z)}{dz} \right]_{z=1} = \rho'(\rho^{(m-1)}(1)) \cdot \rho'(1)$$

$\rho(1) = \rho$ であるから $Z_m = \rho^m$ となる。したがって, オ 0 世代から解にいたるまでに生じた要素の総数は;

$$Z = \sum Z_m = 1/(1-\rho)$$

となる。したがって単一ノードで探索を行ったとき要する平均時間 T_{max} は,

$$T_{max} = u/(1-\rho) \quad (1)$$

である。

一方, $Z = X + Y$ であり, また展開された中間結果の属する達成域が先行中間結果のそれとは異なる場合1回通信を伴うから $m-1$ 世代から m 世代を生じたとき要する通信回数 Y_m は, 変数 Y の独立な Z_{m-1} 回の和となる。したがって, Y_m の分布の母関数は $\rho^{(m-1)}(\phi_Y(z))$ で与えられる⁴⁾。故に Y_m の平均値 y_m は,

$$y_m = \left[\frac{d\rho^{(m-1)}(\phi_Y(z))}{dz} \right]_{z=1} = \rho'^{(m-1)}(\phi_Y(1)) \cdot \phi_Y'(1) = \rho$$

探索が終結するまでに要した通信時間

の平均値 T_{com} は(3)式で与えられる。

$$T_{com} = cu(\sum y_m) = cu\mu/(1-\rho) \quad (3)$$

一方, ある達成域に属する1つの中間結果が同一達成域に後続中間結果を生成しなくなるまでの平均時間を T_{min} とすると, ρ を λ で置換えたときの T_{max} と同様にして $T_{min} = u/(1-\lambda)$ を得る。これは探索に要する最少時間の平均値を与えているとみなすことができる。

$\sigma = \lambda/\rho$ から, (4)式, (5)式を得る。

$$T_{com} = cu\rho(1-\sigma)/(1-\rho) \quad (4)$$

$$T_{min} = u/(1-\rho\sigma) \quad (5)$$

図8に $c=10$, $\rho=0.99$ とした場合の T_{max} , T_{min} , T_{com} を示す。 σ の最適値は, 通信時間と処理時間がバランスしたところ, 図8の例においては, 0.895 から 0.975 までの間にあり, $\sigma > 0.975$ で処理時間ネック, $\sigma < 0.895$ で通信時間ネックとなることがわかる。

なお, これはノード総数, 達成域の重複の程度に無関係であり, ρ および c の値によってのみ定まる不変的な性質であることに注意されたい。

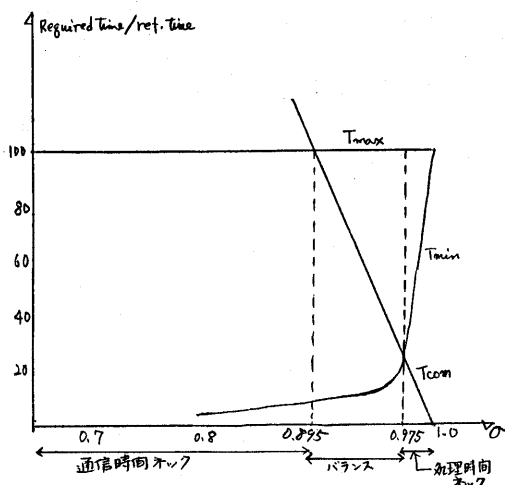


図8. $c=10$, $\rho=0.99$ の場合の T_{max} , T_{min} , T_{com}

一般に最適解 σ_{opt} は不等式 $\sigma_0 < \sigma_{opt} < \sigma_1$ を満す。ただし、 σ_0 は方程式 $T_{max} = T_{com}$ の σ についての解、 σ_1 は $T_{min} = T_{com}$ の σ についての解である。(4)式および(5)式と $\sigma \leq 1$ とから

$$\sigma_1 = (1 + p((1-p)^2 + 4(1-p)^{1/2}) / 2p \quad (6)$$

を得る。

同様に $\sigma_0 = 1 - 1/cp$ となる。 $0 \leq \sigma$ だから、 $cp < 1$ の場合 σ は定義されない (i.e. どのような σ の値に対しても通信ネックとはならない)。したがって σ_{opt} の範囲は一般に(7)式で表される。

$$\begin{aligned} \max(0, (1 - (cp)^{-1})) &\leq \sigma_{opt} \\ &\leq (1 + p - ((1-p)^2 + 4(1-p)/c)^{1/2}) / 2p \end{aligned} \quad (7)$$

4. 開発の現状と今後の技術課題

筆者等は、SD³システムの応用として略地図の発生システムの開発を行った。これは、ある特定地域における建物、交差点およびそれらの間の接続関係といった地図情報を関係データベースとして表現し、それらを適宜重複を許して各ノードに格納したEDBと、完全に重複させたIDBとから、与えられた2点間のパスを略地図として出力するという一種のプラン作成システムである。パスの導出は、評価値、全解探索のいづれによっても可能である。また現システムでは、初期状態で全編成となるようにEDB、IDBの分割を行っている。また各ノードの達成域は、2次元の領域として座標値によって記述されている。

システムの構成は、処理ノードとなる5台のパーソナルコンピュータと、

ネットワークノードとなる3台のネットワークアダプタ装置とから成っている。

ユーザは、どの処理ノードからでも問題の入力を行うことができ、問題解決は5台のプロセッサによって分散型で実行されることになる。但し、現システムでは、1つのノードにおいて問題が入力されるとその問題解決が終了するまで他のユーザからの入力はそのノードにおいても受け付けられない。以下にシステムの出力例を示す。

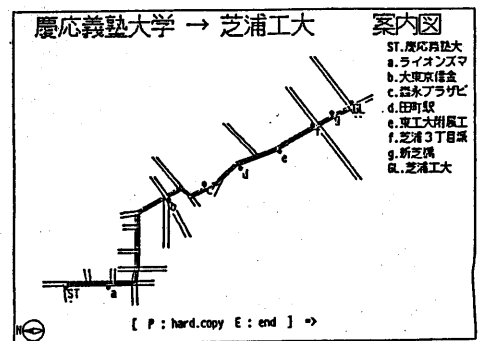


図9. システムの出力例(i)

図9は慶応大学を出発点、芝浦工大を目的点としたときの出力結果である。図中出発点はSTで、目的点はGLで示されている。経路上にある主要な建物等は、それぞれa, b, c等の記号で、右欄にその名称を表示している。

図10のaは品川駅を出発点、高輪消防署を目的点としたときの出力であり、同図bは、出発点と目的点を入れ換えたときの出力結果である。aからもわかるとおり経路は非常に複雑であり、このシステムは、必ずしも最短経路を保証するものではないため、aとbとは全く異なる経路導出を行っていることに注意されたい。

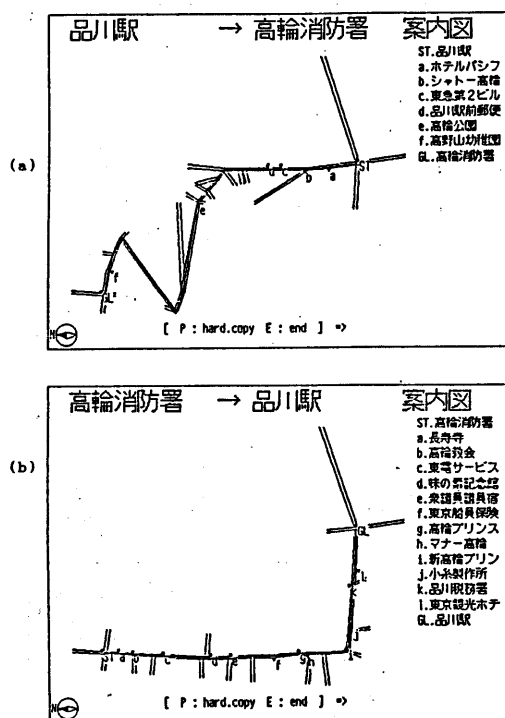


図10 システムの出力例(2)

実行時間は、生成中間結果数が100以上ある相当複雑な経路探索で3~5分、20以下の簡単なもので10~30秒を要しており、パーソナルコンピュータベースのプラン策定システムとしては、ほぼ満足いく性能を得ることができた。

現在、地図情報として試験的に格納されているファクトデータは、インデックス部分も含め総容量約160Kバイト、地契数約3000ヶ所であり、東京都港区田町駅周辺の主要な建造物、企業、官庁等はすべて収録されたものとなっている。

分散型問題解決システム⁶⁾⁷⁾という技術は極めて新しいものであり、未だ確立された方式も存在しないため、残された技術課題も多いが、筆者等は本稿で述べた基本方式のもとに、以下のよう

考えている。

(1)他の応用プログラムの開発

SD³システムの具体的適用例である略地図発生システムにおいては、問題解決の一形態であるプラン作成を扱った。SDの基本アーキテクチャおよび方式を検証する意味でも今後は診断システムや質問応答システムあるいは類似図形検索といった列種の応用システムの開発を行う予定である。

(2)分散データベース機能の拡張

SD³システムにおいては、IDB更新は現在ローカルな操作によってなされているが、これはIDBが分散データベースシステムの管理下におかれていないことに起因する。今後は分散データベースの機能拡張を行うことによりメタデータだけでなくルールを更新を含む広義の分散データベースによりこれらの統合を行うことを考えている。

参考文献

- 1) 岸田他：分散処理向きローカルネットワークについて、情報学会、分散処理システム研究会資料14-4(1982)
- 2) Yamazaki et al: A proposal for Broadcast Architecture Network (BANET), proc. 6th ICCS PP115-120 (1982)
- 3) 山崎：分散型演繹データベースシステムSD³とそのプロトコル、情報学会論文誌、Vol.26, No.2, PP288~295 (1985)
- 4) Feller, W: An Introduction to Probability theory and Its Application, Vol.1 Wiley, New York (1950)
- 5) 吉田他：演繹機構を持つデータベースによる略地図発生システム、情報学会 28回全国大会 PP735-736 (1984)
- 6) Smith, R.: The Contract Net Protocol: High level Communication and Control in a Distributed Problem solver, IEEE Trans Comput. Vol. C-29, No.12, PP1104~1123 (1980)
- 7) Lesser, V. and Eppan, L.: Distributed Interpretation: A Model and Experiment, IEEE Trans Comput., Vol. C-29, No.12, PP1144~1163 (1980)