

大規模知識ベースマシン実験機の開発 (3) —ソフトウェア・シミュレーションによる評価 (第1報)—

酒井浩 柴山茂樹 岩田和秀 物井秀俊 森田幸伯 伊藤英則
(株) 東芝 総合研究所 (財) 新世代コンピュータ技術開発機構

1 はじめに

知識を、変数を含む項を要素とする表の形式で格納し、それに対する単一化結合などを基本演算とする知識ベースマシン [横田86] の検討を行なっている。本稿では、その主要な構成要素である単一化エンジン (UE) の評価 [森田87] をふまえ、システム構成や制御方式を検討し特性を明らかにする目的で行なったソフトウェアシミュレーションについて述べる。

2 シミュレーションの概要

2.1 評価の目的

今回の評価は、複数のUEとマルチポート・ページメモリ (MPPM) [田中84] を主要な構成要素とするシステムによる演繹データベース処理 (単一化結合等の反復で実現) について、最適なシステム構成や制御方式を検討するとともに、システムの特性を明らかにする目的で行なった。

2.2 システム構成

シミュレーション・モデルにおけるシステム構成を図1に示す。本システムでは、知識はMPPM上に表の形式で格納され、UEにより単一化結合などが施され、結果が再びMPPM格納されることになる。

UEとMPPMの諸元は次のとおりとした。

(1) UE

機能: MPPMの複数ページ間の単一化結合と制約
台数: 1 ~ 32
データ処理能力: 基本的に20 Mbyte/sec
入出力ポート数: 入力2ポート, 出力1ポート
処理方式: [森田87] の方式
最大入力量: ≤ 64 Kbyte

(2) MPPM

トラックサイズ: 512 byte
ページサイズ: 512 byteの整数倍
データ転送能力: 20 Mbyte/sec/port

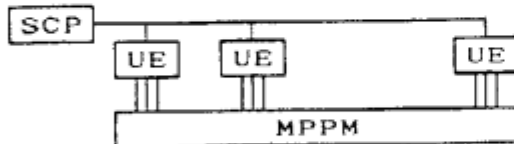


図1 システム構成

2.3 推論処理方式と制御方式

とりあげた推論処理方式は、or並列の入力演繹であり、定義節に対応する関係と第n回目の中間結果 (第0回目はゴール) に対応する関係の間で単一化結合を行ない、結果を第(n+1)回目の中間結果とすることにより推論を行なう。

本システムは、関係を固定長のページを単位としてMPPMに格納し、制御プロセッサ (SCP) がUEにページを指示して単一化結合などの処理をさせる。単一化結合では、分配則 $P \bowtie (Q + R) = P \bowtie Q + P \bowtie R$ が成立つので、SCPは関係間の単一化結合を何ページかごとに分割して複数のUEで並列処理させることが可能である。一般に、mページ分の関係とnページ分の関係の間で単一化結合する必要があるとして、問題を何個のどのような小問題に分割すれば良いかを決める必要がある。

本システムの制御上の主要な問題点は、この分割方法である。すなわち細かく分割しすぎるとシステム内のデータ転送の総量が増えたり、結果を少ししか含まないページが多数発生するため性能が低下する。逆に分割の仕方が足りないとUEの稼働率が低くなり、やはり性能が低下する。

また定義節および中間結果の量、UEのページ生成部の効果、単一化処理の負荷など問題自体が有する性質は、システムの処理特性を考える上で重要であり、異なる性質を有する問題について評価を行なう必要がある。

2.4 評価項目

システム構成や制御方式を変えた時に、システムの振舞いの変化を調べる尺度として、下記の項目をとりあげた。

所要時間: 全部の解が求まるまでの時間
台数効果: $1 / (\text{UE台数} \times \text{所要時間})$
MPPM使用量: 処理過程で必要なMPPMスペースの最大値
UE稼働率: 各UEが動作した時間の平均 / 所要時間
また、システムの振舞いを詳細に検討するため、UEの稼働状況とメモリ使用量について時間変化も調べた。

2.5 ソフトウェア・シミュレーションの方法

システム構成や制御方式を模擬するソフトを作成した。UEについては内部の構成と処理方式を正確に模擬し、またMPPMについてはそのデータ転送の特性を模擬することにより正確な処理時間が得られるようにした。ただし、SCPが制御に要する時間は0としている。

A Simulation Study of a KBM-Architecture

Hiroshi Sakai, Shigeki Shibayama, Kazuhide Iwata, Toshiba Corp.
Hidetoshi Nomoi, Yukihiko Morita, Ridenori Itoh, ICOT

3 結果と考察

3.1 先祖検索の結果

定義節は大量で中間結果は比較的小量できる場合の典型として付録Aに示す問題をとりあげた。この問題は、ペア生成部が効果的に作用し、また単一化処理の負荷は軽い。

・問題の分割に関する評価

この問題は定義節の方が十分大きいので、 n 分割するすれば定義節をなるべく均等に分割すればよい。この n の候補として、現時点の空きUEの台数とUEの総数の2種類をとりあげて比較した。UEが4台の場合について、UEの稼働状況の時間変化を図2に示す。図からわかるように空きUE台数で分割する方式では、時々、少数のUEが大きな問題を抱え、他のUEは空いている状態が発生するため、UE稼働率が上がらず性能が低いことがわかる。

・問題の性質に関する評価

UEの台数を変え、台数に見合う性能を発揮するが調べた(図3)。この問題の処理上の特徴は、定義節入力に多く時間を要することである。例えばUEが1台では、約45%の時間を占める。また、複数のUEによる並列処理では、中間結果が小刻みに生成されるため、それと定義節の間で単一化結合する回数が増えるため、UEの台数に見合った効果が得られにくい。一層の性能向上をはかるには、定義節の人力時間を減らす努力が必要である。

3.2 8クイーンの結果

定義節は少量で、中間結果は大量にできる場合の典型として付録Bに示す問題をとりあげた。この問題では、ペア生成部の効果が無く、また単一化処理の負荷は変数の回数が多く非常に重い。

・問題の性質に関する評価

UE台数を変え、得られる性能を調べた(図3)。問題に関する上記の特徴により、単一化処理に多くの時間を要し、しかもそれは中間結果の量にはほぼ比例する。そのため、UE稼働率が大きくなるように中間結果を分割すれば、ほぼUE台数に見合った性能向上が得られることになる。

今後、この種の問題についてさらに性能改善を図るには、ペア生成部の機能を改良して、よりひろい範囲のデータに対して有効に作用させる必要がある。

4 おわりに

UEとMPPMを主要な構成要素とする知識ベースマシンに演算データベース処理をさせた場合の評価を行なった。今後、ハードウェア実験機を使用して、さらに検討を進める予定である。

参考文献

- [田中84] Tanaka, Y., "A Multiport, page-memory Architecture and a Multiport, Disk-Cache System", New Generation Computing, 2, Ghemsha, 1984.

[横田88] Yokota, H., Itoh, H., "A Model and Architecture for a Relational Knowledge Base", Proc. 13th Int'l Symp. Computer Architecture, 1988.

[森田87] 森田他, "知識ベースマシンにおける単一化専用装置の処理方式とその評価", 情報処理学会データベース・システム研究会, 57-4, 1987.

付録A: 先祖検索

定義節 $\text{ancestor}(A,B):-\text{father}(A,B).$
 $\text{ancestor}(A,B):-\text{mother}(A,B).$
 $\text{ancestor}(A,B):-\text{father}(A,C),\text{ancestor}(C,B).$
 $\text{ancestor}(A,B):-\text{mother}(A,C),\text{ancestor}(C,B).$
 関係のないルール 8 個
 父(母)子関係のファクト 1 8 0 0 個
 ゴール ? $\text{ancestor}(A,f992)$
 結果 128 件 (重複あり)

付録B: 8 queen

定義節 $q([A,B],b(\dots),b(\dots),b(\dots)):-$ } $\times 8$ 個
 $q(B,b(\dots),b(\dots),b(\dots)).$
 $q(\text{nil},A,B,C):-\text{nil}.$
 ゴール ? $q([1,2,3,4,5,6,7,8],A,B,C)$
 結果 92 件

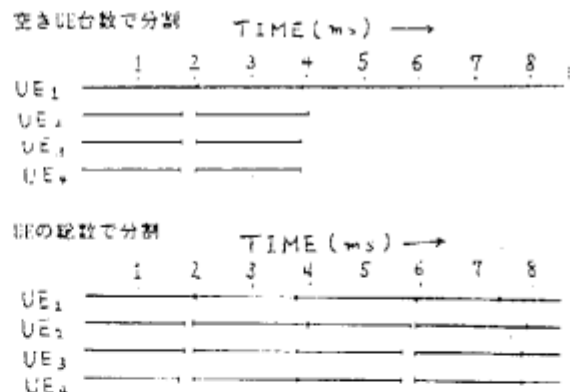


図2 UEの稼働状態の時間変化の比較

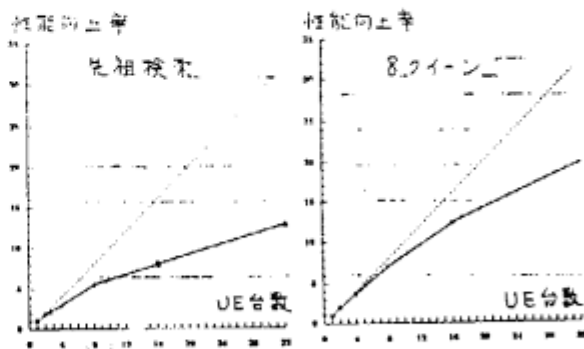


図3 先祖検索と8クイーンの台数効果