

演繹データベースシステム

EX_DDBの概要

3K-1

吉田 敦* 中川 俊夫* 土井 晃一* 嶺野 和夫** 田中 英彦*

(* 東京大学工学部、 ** 富士通株式会社)

1. はじめに

関係データベースに推論機能を持たせたものを演繹データベースという。演繹データベースは、エキスパートシステムなど大量の知識を扱う知識処理システムを実現するための有力な手段の1つである。しかし、従来の演繹データベースに対する研究では推論実行方式の提案と実現可能性が示されたのみであり、推論実行方式の効率に対する考察はほとんどなされておらず、応用や言語に対する考察は全くといってよいほどなされていない。我々はエキスパートシステムや自然言語処理などの応用に演繹データベースを用いることを前提として、その上でのプログラミング環境や言語の実現方法、推論の実行方式、アーキテクチャを考える方針で研究を行なっている。この方針に沿って開発される演繹データベースがEX_DDBである。

2. 研究の概要

演繹データベースEX_DDBに対しては、応用、言語、推論実行方式、アーキテクチャの4つの側面を考え、各側面に対して以下に述べるような方針で研究を行なっている。

2.1 応用

応用面では、エキスパートシステムや自然言語処理など、大量の知識を扱う応用を記述する上で、プログラミングを支援する環境やプログラミングスタイルの確立を目的とする。

そのため知識エディタEX_DDB/KEDやデバッガEX_DDB/DBG等のツールを作成し、これを用いてプログラミングの方法を検討するとともにこれらのツール自体の機能についても考察を行なう^[1]。

2.2 言語

言語面では、演繹データベース上で動作する述語論理型言語として、ROTHORN(Relation Oriented HORN clause)^{[2][4]}と呼ばれる言語を提案し、開発を行なっている。

ROTHORNは、エキスパートシステムその他の応用を記述するために、四則・比較演算や入出力操作を組み込み述語として持ち、データ型として集合を導入しこれに対する操作を実行する述語を持つ。

2.3 実行方式

EX_DDBにおける処理には、質問を受ける前に実行される質問の前処理と、質問を受けてからの推論実行がある。前者はEX_DDBにおける推論実行の操作を、拡張された関係代数演算EXTRAを用いた処理の系列とみなし、その

効率良い実行方法を静的情報を基に前もって決定することである。具体的には定義の集合を基にrule/goal graphを生成し、解析することで拡張関係代数演算系列の生成や、reductionを実行する場合の各ruleに対するbody部リテラルの評価順序や単一化パタン検索の実行方法などを仮決定する。拡張関係代数演算系列におけるselectionの条件やreductionにおけるruleリテラル同士の評価順序の決定など、動的情報に基づくものはその時のリテラルのパタンおよび束縛情報に含まれる変数によって決定する。

推論の実行は、束縛を用いた即時評価方式^[5]と拡張関係代数演算系列の実行を併用する形で行なわれる。基本的には束縛を用いた即時評価を実行する形であるが、束縛情報によるリテラルに対する定義候補への絞り込み効果がなく、そのリテラルが再帰的定義を持たない場合には関係代数演算系列を実行し、得られた束縛情報とgoalに対する束縛情報の間でconsistency checkを実行することになる。

2.4 アーキテクチャ

前節で述べた推論実行方法の構成単位である拡張関係代数演算EXTRAは、結合演算を高速化するためhashとsortによるクラスタ分割を行ない、対応するバケット同士でのみ結合演算を実行するアルゴリズムを採用する。単一化を伴わない通常の関係代数演算に対してはこのアルゴリズムの高速性が証明されており、これを支援するマシンアーキテクチャも考察されている。EX_DDBにおいてはこのような関係データベースマシンのアーキテクチャを拡張関係代数演算を支援するように拡張することを考える。ここで問題となる点はタグ付き可変長レコードの格納と扱い、単一化・代入操作を実行するための処理モジュールの拡張などである。

3. 研究の現状

3.1 実行方式の検討と評価

関係データベース上で推論操作を効率良く実行するために関係代数演算を、単一化と代入を伴うように拡張し、これを用いた推論実行方法として、束縛情報を用いた即時評価方式を考案した。これに対して、幾つかの模式的例題における計算量の評価は既に済んでおり、比較的浅い探索木で、枝の広がり大きな例題に対しては遅延評価方式に比べて効率良い全解探索が実行出来ることが確認されている。但し、ruleによる展開が多く枝の広がりの少ない例題では遅延評価方式との間に顕著な差が出るわけではない。

ruleによる展開の多い例題でも、再帰的定義がない場合には関係代数演算系列を実行した場合の方が処理効率がよい場合がある。従って束縛を用いた即時評価と関係代数演算系列を併用することが処理効率化の有力な手段となる。

多くの例題ではruleに対する定義の数は比較的少なく、ruleに対する単一化パターン検索においてhashとsortを用いた方法による高速性が充分でない可能性がある。また、現在の実行方式ではgoalの表現が関係とは別のデータ型になっているため、そのデータ型と関係のtupleとの間の相互変換や、OR並列やAND並列を実行する場合の資源割り付けの問題が生じる。そこで、現在goalを関係で表現してOR並列やAND並列を実行する方法も検討中である[33]。

3.2 前処理方式の検討

前処理ではまず定義をもとにrule goal graphを生成し、loopの検出により再帰リテラルと非再帰リテラルの区別を行なう。非再帰リテラルに対しては関係代数演算系列を生成する。次に各リテラルに対してreductionを実行する場合のruleのbody部リテラルの評価順序を決定する。最後に関係代数演算系列に対して静的情報による最適化を施す。動的な部分に対しては推論処理の時にリテラルパターンと束縛情報を参照して最適化を行なう。

3.3 言語の実装

ROTHORNは、四則・比較演算や入出力操作、集合・関係操作述語の実装を終了している。現在、エディタやデバッガとの一体化と、組み込み述語体系の検討を行なっているところである。

3.4 その他

エディタやデバッガに対しては基本的機能を実行する部分が完成している。これらのツールの高機能化は今後の課題であるがそのためにプログラム作成の方法を検討中である。

アーキテクチャについては、処理モジュールの機能に対する要求とその実現方針が固った段階である。その他のモジュールに対しては、機能分担は決まっているがまだ具体的な機能の実現方法は検討段階である。

4. 今後の課題

EX_DDBに関する今後の研究課題として、以下にあげるようなものがある。

① プログラミング支援環境の整備

演繹データベースにおけるプログラミング支援機能の高度化の1つに、エディタに対して関係の正規性等を用いたfactの設計に関する知識や、関係で表現されたfactを扱うruleの設計に関する知識を持たせることが考えられる。このような知識をエディタに持たせることでプログラムの生産性をあげることは将来的に重要な課題である。

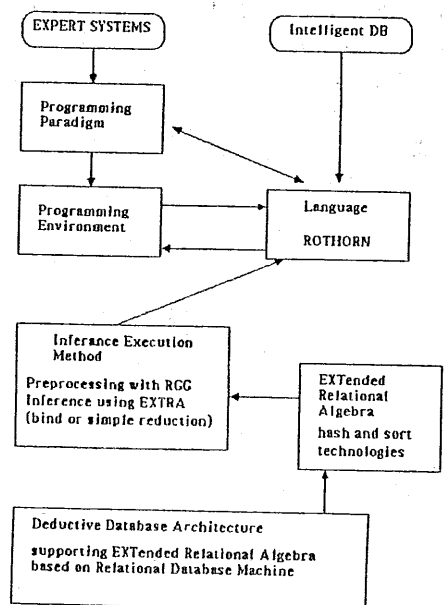
② 推論実行方式の多様化と統合

EX_DDBでは現在述語論理型言語における推論の実行を前提としている。この場合でも束縛を用いた即時評価の実行や関係代数演算系列の実行などの推論方法を統合する必要

がある。また、述語論理型言語以外の推論方式、例えば、CF値を用いた推論やPSなどとの統合も考えられる。

③ 協調型分散知識処理システムへの発展

画像理解や知的CADなどに演繹データベースを用いる場合、専用のシステムと演繹データベースとの協調型分散システムにする必要がある。また、複数の知識表現形式を併用する場合、演繹データベースではフレームのように構造の複雑な知識を効率良く扱えないため、専用マシンと演繹データベースをつないだ分散システムを構築する必要がある。このような協調型分散知識処理システムに対する要求は今後一層大きくなると思われる。



EX_DDBの概要

参考文献

- [1] 嶺野、中川、土井、吉田、田中：「演繹データベースEX_DDBにおける応用支援環境」第34回情報処理学会全国大会3k-3
- [2] 吉田、中川、土井、嶺野、田中：「演繹データベースシステムEX_DDB上の言語ROTHORN」第34回情報処理学会全国大会3k-2
- [3] 中川、土井、吉田、嶺野、田中：「演繹データベースシステムEX_DDBにおけるゴール表現の検討」第34回情報処理学会全国大会3k-4
- [4] 吉田、嶺野、田中：「大規模知識処理支援環境EX_DDB」電子通信学会人工知能と知識工学研究会A186-33
- [5] 吉田、土井、嶺野、田中：「演繹データベースにおける推論実行方式の評価」電子通信学会人工知能と知識工学研究会A186-24