

高速検索のための時空間データ管理方式

4 A A - 7

-動的データへの対応-

仲 篤起, 才脇直樹, 西田正吾
大阪大学大学院基礎工学研究科

1.はじめに

近年, コンピュータ内に構築した3次元仮想世界を, 様々な用途に役立てようとする研究が数多くされている. 特に, 3次元仮想世界を用いての現実疑似体験やアニメーション作成等を考えた際には, これら時空間データに対して高速なレスポンスが可能なデータ管理手法が必要になる. そこで, 我々は, 「ある時空間範囲に含まれるものを探す」あるいは「ある点に最も近いものを探す」などの時空間検索の高速化が可能なデータ管理手法の提案を研究の目的としている. さらには, このような機能を実現することにより, 3次元仮想世界やアニメーションでの直接的な検索条件の指示およびリアルタイムレスポンスの実現も目指している.

筆者らは, 従来より木構造に基づいた高速検索のためのデータ構造の研究を進めてきており[1], 地下配管の設備管理への適用検討等も行ってきた[2]. 本稿では, まず, 筆者らが提案しているデータ構造(以下, AT構造と呼ぶ)の概要について述べる. その後, 動的なデータを管理可能とするためのAT構造の拡張について, 具体的な検討を行う.

2.AT構造の概要[3][4]

多次元データの管理構造として, 静的データを管理対象としたk-d木[5], 動的データに拡張したMD木[7]等が提案されている. ここでの静的とは, データがあらかじめ全て与えられているという意味であり, 動的とは, 逐次データが投入・削除されていくという意味である. これらの手法を応用して時空間情報を持つデータを管理する場合, 代表的な手法としては,

- [1] バージョンごとに木構造を作成保持するMT構造
- [2] 全データを位置情報により分割した1つの木構造で管理し, 時間情報はデータの属性として扱うST構造
- [3] 時間情報も位置情報と同一視し, 位置(2次元) + 時間(1次元)の3次元空間内の時区間データと

して1つの木構造で管理する3D管理構造

などが考えられる. しかし, それぞれのデータ構造には一長一短があり, ST構造のように空間木を中心にデータ構造を作成し, 時間情報をデータ属性として取り扱おうと, 空間検索範囲が狭い場合には, 効率が良くなるのに対して, 空間検索範囲が広がると効率が悪化する. 一方, MT構造のように時間中心でデータ構造を作成すると, 時間検索範囲が狭い場合には効率が良いのに対して, 時間検索範囲が広がると効率が悪くなる. この「検索範囲が狭い場合には検索効率が非常に良いが, 検索効率が広がるにつれて検索効率が悪化する」という性質は, 木構造を用いる限り, 基本的には避けられないものである.

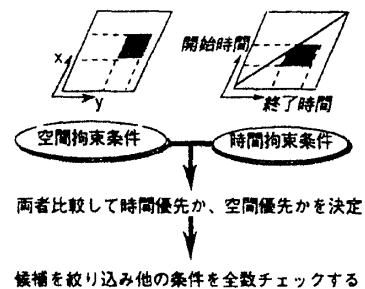


図1 AT構造の概要

そこで, 提案するデータ構造では, 空間木と時間木を用意しておいて, 空間と時間の検索範囲に応じて, どちらの木構造を優先して検索するかを適応的に切り換える方式を採用する. つまり, AT構造では, 図1に示すように, 任意の時空間検索範囲に対し, 「時間情報と位置情報の検索範囲の全体に対する割合を求め, 小さい方を採用する」というロジックで時間木を用いるか空間木を用いるかを決め, 他の条件については木構造より選択されたものの中から全数チェックを行う.

計算機実験では, AT構造, 従来のST構造, 3D管理構造について比較の結果, その有効性が確認されている[3]. また, 複数の状態をもつ対象への適応についても検討を行っている[4].

3.動的データへの適用手法

時空間データを管理するデータ構造を考える際に重要な点の一つとして, 対象のデータ集合が「静的」

であるか「動的」であるかが挙げられる。データ構造が動的データを管理可能であると、最新の入力データを管理できる点やデータ構造更新の際のコストを削減することができる等のメリットが生じる。

そこで、ここでは、動的データに対応できるようにAT構造を拡張する。以下、具体的なデータ構造の構成について述べる。

3.1 検索木決定ロジック部

検索木決定ロジック部は、静的データを対象とした場合には、時間木と空間木について、検索範囲の全体に対する割合の小さい方を検索する方式を採用していた[3]。動的データを管理対象とする場合も、基本的には同様の手法を用いて検索木を決定する。

3.2 空間木のデータ構造

位置情報によるデータ構造(空間木)については、kd木を用いてデータ構造を作成していたが、ここでは従来より提案されているBD木[6]やMD木[7]を用いることにより、容易に動的データが管理可能となる。

3.3 時間木のデータ構造

時間木では、開始時間・終了時間を軸にとった2次元平面に時間情報をマッピングすることにより、時区間情報を点情報に変換している。そして、その変換後の3角形領域を3分木で管理する手法を提案している[3]。しかし、動的データを管理対象に考えた場合、この手法では時間情報の管理領域がデータ更新毎に増加してしまうという問題点が生じる。

そこで、この増加領域を効率よく管理する手法が必要となるが、従来のトップダウン的に領域管理を行う平衡木等の考え方は利用できない。このため、以下のように一時的に管理領域の増加部分を管理する手法を提案する(図2参照)。

- (1) 時間情報の管理領域の情報(各ノードが所持)を最新情報へと拡張する。すなわち、図2において、一番右端の分割軸を終了時間軸と平行に延長することにより増加領域を管理する。

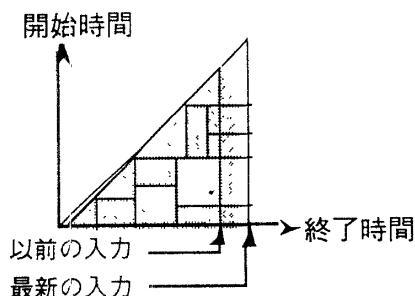


図2 時間データ構造の拡張方法

- (2) 入力データを格納すべきリーフを探索する。

- (3) リーフに格納されているデータ数がリーフのデータ容量より小さい場合はリーフにデータを追加して終了。データ容量と格納データ数が等しければ領域分割を行う。領域分割は、kd木の分割方法を利用して、 $(X \rightarrow Y \rightarrow X \rightarrow \dots \rightarrow X)$ のように巡回的に分割軸を選択し、リーフの管理領域を2等分することにより行う。

提案した時間木の動的データ管理手法の問題点としては、データが投入される毎にデータ構造の非平衡性が強くなるので、一定数のデータが投入された時点でデータ構造を再構築して平衡性を改良する必要があることが挙げられる。

4. まとめ

従来のAT構造は、静的データを対象としていたため、最新のデータを管理することが出来なかった。そこで、ここでは、動的データを管理可能となるようなAT構造の拡張方法について検討した。今後、シミュレーション実験を行い、本方法についての検討を重ねていきたいと考えている。

なお、本研究は一部文部省科研費重点領域研究(高度データベース)により行われたものである。

[参考文献]

- [1]寺岡, 丸山, 中村, 西田「空間検索を効率化した時空間データ管理構造の提案-多次元Persistent Tree-」信学論D-II, Vol.J78-D-II, No.9, pp.1346-1355, 1995.
- [2]玉田, 寺岡, 丸山, 西田「3次元仮想都市空間を用いた設備管理システム」電気学会論文誌C, Vol. 116-C, No.5, pp.517-523, 1996.
- [3]仲, 才脇, 辻本, 西田「時空間データの高速検索のためのデータ管理方式」, Proceedings of Advanced Database Symposium '96, pp.71-78, Dec.1996.
- [4]仲, 才脇, 辻本, 西田「高速検索のための時空間データ管理方式-状態が複数の場合の検討-」, 信学技報DE96-86, pp.73-78, Jan.1997.
- [5] J.L.Bentley, "Multidimensional Binary Search Trees Used for Associative Searching," Commun.ACM, vol.18, pp.509-517, 1975.
- [6]大沢, 坂内, 「良好な動特性を持つ多次元点データ管理構造の一提案」信学論D, vol.J66-D, No.10, pp.1193-1200, Oct.1983.
- [7]中村, 阿部, 大沢, 坂内, 「多次元データの平衡木による管理-MD木の提案」信学論D, vol.J71-D, No.9, pp.1745-1752, Sept.1988.