

令和5年度 コンピュータ科学科卒業論文要旨

石川 研究室	氏 名	杉 江 祐 介
卒 業 論 文 題 目	Universal Adaptive Radix Tree における空間分割の改善に関する研究	
<p>多次元索引は複数の次元で表されたキーによる検索を補助するデータ構造である．地理情報システムにおける空間オブジェクトの検索や，データベース管理システムにおける多次元データの選択演算効率化などに利用される．多次元データは複数の次元から成るが，メモリやストレージ上では1次元空間上に配置されるため，多次元空間上での局所性を適切に反映した索引構造が求められる．</p> <p>多次元索引は多次元空間を直接扱うものと1次元空間へ射影して扱うものの大きく2つに分けられ，本稿では後者を主に扱う．多次元空間を直接扱う索引の代表例はR木であり，多次元の包囲矩形などを用いて多次元空間を階層的に分割していく．1次元空間へ射影するものは主に空間充填曲線に基づいており，Universal B木（UB木）が代表である．後者の利点は，既存の効率的な1次元索引を流用でき，挿入・削除に伴う木の構造変更などが容易な点である．</p> <p>本稿ではUB木およびAdaptive Radix Tree（ART）を組み合わせた索引構造である Universal ART（UART）の改善について述べる．UARTは汎用的に使用可能な多次元索引だが，データに偏りがある場合に挿入と範囲検索の性能が低下するという課題を持つ．性能低下の原因は，不適切な空間分割による疎な部分空間の生成である．UARTではノードの空きスペースがなくなった際に対応する多次元空間を分割することでノードを分割するが，分割後の部分空間が十分な数のレコードを持つという保証がない．そのため挿入されるレコードに空間的な偏りがある際に，レコードを少数しか持たない疎なノードが多数生成され性能を低下させている．</p> <p>既存手法ではノードにレコードが入らなくなった時点で空間を分割しており，分割後に生成されたノードのレコード数には制限を与えていない．つまり，生成される部分空間の状態を無視して空間を分割するため，レコード数の少ないノードが大量に発生する場合がある．そこで提案手法では，部分空間にあたるノードの空間使用量がノード全体の三分の一を超えることを保証するよう空間を分割した．これにより，レコード数の極端に少ない疎なノードの生成を抑制し，性能向上を図る</p> <p>本実験では，実データを用いた有用性の評価と，偏りなどのパラメータに対する頑健性の評価を行った．比較対象は既存手法のUARTと，現在多次元索引としてよく使われるUB木とR木の計3つとした．有用性の評価には，日本，アジア，地球の規模の異なる3つの地理データセットを用いた．また，パラメータに対する頑健性の評価には，各次元の値がZipf関数に従う人工データを用いた．結果として，いずれのデータセットにおいてもUARTの改善による性能向上が確認できた．</p> <p>どのデータセットにおいても，UARTの改善前より改善後のほうが優れた性能を示した．今後の課題としては，ポリゴンデータにも対応させることが挙げられる．</p>		