

# 平成 22 年度 情報工学コース卒業研究報告要旨

石川 研究室	氏 名	加 藤 翔
卒業研究題目	科学データベースにおける関数近似問合せ処理の効率化に関する研究	

今日では、科学データベースの有効活用が重要となってきた。科学データベースでは連続的な値を取るデータを扱うが、観測データを関数として近似し、関数に対して直接的な問合せを可能とするシステムが開発されている。しかし、問合せ時に動的に関数値のサンプリングを行うため、問合せ時のオーバヘッドは小さくない。

本研究では、問合せ時のオーバヘッドを減らすために、一種の索引を提案する。この索引手法は、統計情報を集約表現する索引手法である STING の考え方に基づく。あらかじめ、近似関数データからグリッド状にサンプリングを行い、それらの統計情報からなる階層的な集約索引構造を生成する。

図 1 に提案手法の概念図を示す。実データに近似関数が当てはめられた結果を受けて集約索引の構築が開始される。まず、近似関数に対してグリッド状にサンプリングを行う。次に、サンプリング点が存在する空間をグリッド分割し、各セル内のサンプリング値の平均、最大値、最小値などの統計量を取り、各セルに対するエントリを構築する。これが集約索引のデータ構造で最も詳細なボトムレイヤとなる。上位の各レイヤについては、統計情報はそれぞれ一つ下のレイヤの統計情報を元に計算される。なお、すべてのセルはそのセルの中心の座標におけるサンプリング値も合わせて保持する。

問合せでは、集約索引に保存された統計情報を用いて、図 2 のように必要な領域を段階的に絞ることで、処理の効率化を図る。問合せの詳細レベルがボトムレイヤの詳細レベルを超えない限り、問合せ処理に必要な情報は集約索引にすべて納まっている。そのため、新たにサンプリング処理をする必要がなく、処理は高速に行われる。ボトムレイヤの詳細レベルを超える場合では、新たなサンプリング処理が必要となるが、サンプリングを行う領域が絞られているため、従来の問合せ処理に比べると効率的に処理できる。問合せ処理では、処理速度に重みを置いているため、問合せの選択条件から少しずれた値を返すことがある。また、関数近似を行っているため、必要な情報がノイズとして除去されていることがある。この点を補うために、実データに対しても近似関数のサンプリング点と同様、集約索引を構築する。ただし、セルはサンプリング点の代わりに、そのセルの統計情報の元となった実データへのポインタを含む。この集約索引は実データから構築されるため、実データに正確なデータを返す。ユーザはこの 2 つの集約索引を組み合わせることで、ユーザの目的に合った問合せ結果が得られる。

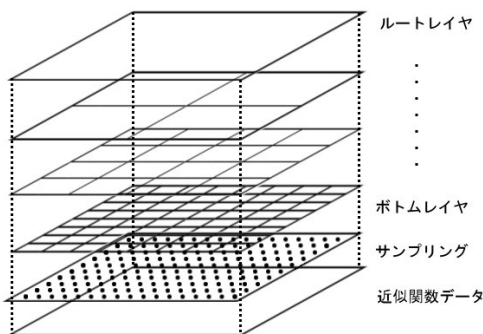


図 1 階層構造の構築

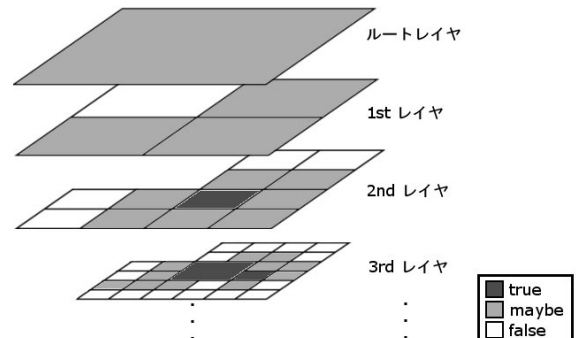


図 2 段階的な絞り込み