

高木 潤一郎¹
Junichiro TAKAGI

川島 英之²
Hideyuki KAWASHIMA

猿渡 俊介¹
Shunsuke SARUWATARI

森川 博之¹
Hiroyuki MORIKAWA

東京大学 先端科学技術研究センター¹
RCAST, The University of Tokyo

筑波大学 計算科学研究センター²
Center for Computational Sciences, University of Tsukuba

1 はじめに

センサネットワーク技術の進展に伴い、多様な実空間情報を取得することが可能になり、データセンサモニタリングや構造/地震モニタリングなどが実現されている。取得したデータを蓄積し、過去のデータを参照することで、現在の状態の把握や傾向の予測などが可能となる。一方で、増え続けるセンサデータを蓄積する際にはデータサイズと問合せ処理性能の劣化が問題となる。筆者らは、センサデータの時系列性、定形性に特化した機構を導入することで、センサデータを効率よく蓄積、問合せ可能とするデータベースシステムについて研究を進めてきた [2]。本稿では、時系列データを蓄積するセンサデータベースに対して、高速な問合せ処理とデータサイズの削減を実現するためのセグメント化とデータ圧縮について述べる。

2 センサデータベースの問題点

センサデータを扱うデータベースには 2 つの問題が存在する。1 つ目の問題は、大量のセンサから連続的にデータが到達するためデータサイズが非常に大きくなることである。例えば地震モニタリングのために加速度センサを高層ビルに設置した場合、約 100 個のセンサから 100Hz で 8Byte のデータを集めることになり、そのまま蓄積すると 1 日で約 7GB になる。2 つ目の問題として、蓄積されるデータが膨大なることに起因して、データの問合せ速度が劣化するということが挙げられる。例えば、地震モニタリングの加速度センサデータ 1 日分に匹敵する 8 億レコードを既存のリレーショナルデータベースに登録し、地震 1 回分のデータの参照処理を行ったところ、3 秒程度の時間を要する。

3 問合せ処理高速化とデータサイズ縮小

センサデータベースの問い合わせ処理の高速化とデータサイズの縮小化に向け、センサデータベースに対する問い合わせが事前登録できる点に着目する。センサデータを利用するアプリケーションからの問合せは、多くが定期監視やイベントデータの取得を目的とした事前登録可能な問合せである。定期監視を目的とする問合せは、農業モニタリングのようなアプリケーションにおいて昨日 1 日分等のデータの参照を行い、過去のデータと比較することで状態の把握を行う。イベントデータに対する問合せは、地震モニタリング等のアプリケーションにおいて複数の地震での波形の比較を行い、それぞれの地震を分析することを目的とする。

既存手法 [1] では多くの問合せが予め決定されるアプリケーションを対象としたデータベースにおいて、問合せを事前登録し、問合せ結果を事前準備してデータの複製を持たせることで高速化を実現している。しかしながら、既存の手法をセンサデータベースに適用した場合、問合せ結果データの複製や重複により、データサイズが増大するという問題が発生する。

そこで本研究では、事前登録された問合せの情報を利用して、セグメント化による問い合わせ処理の事前準備と圧縮データブロックを実現する。問合せ結果データの事前準備において、既存手法で行われていたデータの複製を行わず、問合せ結果データをデータベース内のセグメントとして管理する。追記のみで更新や削除が行われないセンサデータを対象とすることで、データベース内セグメントのオンライン再編成を容易に実現することが可能となる。定期的な監視のための問合せに対する結果データは、新規センサデータの追記とともにデータセグメントの再編成を行うことで作成される。イベントデータ取得のための問合せに対しては、イベントの発生に伴いイベント情報に対する問合せを参照し、新たに結果データセグメントを作成する。

また、各セグメントを圧縮することでデータサイズの縮小を図る。センサデータはその定形性から、圧縮によりデータを高い圧縮率で縮小することが可能である。データを一定数毎にまとめて一様な圧縮を行う方式では、図 2 の 1 に示したように問合せの結果となるデータセグメントと圧縮データセグメントの不一致から生じる圧縮セグメント展開コストの増大が問題となる。それに対して、本研究では、図 2 の 2 に示したように、事前に登録された問合せに基づいてデータのセグメンテーションを行い、結果データセグメント、それ以外のセグメントをそれぞれ纏めて圧縮を行うことで必要なデータの展開コストを削減する。このような事前登録された問合せを利用したセグメント化と各セグメントの圧縮により、事前登録された問合せに対しては、事前に用意した問合せ結果データセグメントをそのまま転送することで高速な応答が可能となる。

4 センサデータベースシステムの実装

上記の 2 つの機能を実現するため、筆者らが実装を進めてきたセンサデータベースシステム [2] に対して機能の拡張を行った。センサデータベースシステムが持つデータ挿入機能・問合せ処理機能に対して問合せを事前登録するための拡張を行い、また図 1 に示すように新たにデータベースの

再編成を行う機能を追加した。ユーザは問合せ処理機能を通して事前に問合せを登録し、データベースに登録された問合せを元にデータの再編成を行う。以下にデータ挿入機構、問合せ処理機構、再編成機構のそれぞれの機能の説明を示す。

データ挿入処理機構はセンサからのストリームデータをデータベースに格納するための、データベースシステムの一般的な機構である。センサからのデータは、例えば 3 軸加速度データの場合 (*time, accelX, accelY, accelZ*) のように時刻とセンサ値の間からなるデータである。

問合せ処理機構は、ユーザからの問合せに応じて結果を返す機構である。例えば加速度センサのデータを蓄積したデータベースから、ある地域の地震の波形データの取得を行う問合せは、`GET(circle(latitude, longitude, radius), event(eqscale >=3), time(start_time, end_time))` のように書くことができる。問合せにはセンサの位置情報などのメタ情報、地震の震度などのイベント情報、そして時間範囲を指定する時刻情報が含まれる。一般的な問合せ処理機構に加えて、問合せの事前登録を行う機能と、データ参照の際に用意した問合せ結果データを参照する機能を実現した。図 1 に示されるように事前登録された問合せと、事前準備された結果データを新たに管理する。問合せが事前に登録されたものである場合、事前準備結果データに対する索引を利用して用意された問合せ結果データセグメントを参照し、対応するデータセグメントを直接返す。事前に登録されていない問合せに対しては、従来通りデータベースに登録されたセンサのメタ情報、イベント情報、時刻情報を元に索引を利用してデータの検索を行い、結果データセグメントを作成する。

再編成機能は、新規センサデータの追記やイベントの発生を契機として、データセグメンテーションによる問合せ結果データの作成とセグメントの圧縮を行う。まず事前登録された問合せを参照し、問合せ結果セグメントを実現するための新たなセグメンテーション案を決定する。次に事前登録された問合せの結果に該当するデータをファイルから読み込み、それぞれの問合せ結果データセットを 1 つのセグメントとして作成し、問合せ結果データセグメントを更新する。その際に、事前準備された問合せ結果データに対する索引を更新する。そして登録された問合せの結果に該当しないデータは時系列に並べられ、固定サイズ毎にセグメントとして保存される。各セグメントはそれぞれ圧縮されるが、登録された問合せに高速に応答するために、事前準備された結果データには圧縮を行わないことも可能である。また再編成中のデータに対する問合せは、一時的なデータのコピーを参照することで実現される。

5 評価

提案手法の有効性を示すため、圧縮を行わない手法、一様なブロックサイズで圧縮を行う手法、そして提案手法として、全てのセグメントの圧縮を行う提案手法 1、問合せ結果データのみ圧縮を行わない提案手法 2 の 4 種類の手法について、それぞれ問合せ処理性能とデータサイズがどのように異なるのか比較を行った。実験環境は地震モニタリングを想定し、加速度センサから収集したセンサデータ 1 億件と、実装したデータベースシステムを用いた。問合せ処理性能の比較では、事前登録した問合せと事前登録していない問合せについて、地震発生期間のデータとして 1 万件のデータ参照に要する時間を計測した。圧縮方式には、圧縮と伸長が高速であることが知られている LZ0 を用いた。図 3 に評価結果を示す。また、図 4 にそれぞれの手法における 1 億レコードのデータサイズの比較の結果を示す。図 3 より、提案手法 2 は登録された問合せに対して、圧縮を行わない場合と比べても高速な処理を実現することが分かる。また、登録されていない問合せに関しても、一様に圧縮を行う場合と同程度の処理性能を実現している。図 4 では提案手法 1、2 とともに圧縮を行わない場合の 1/4 程度のデータサイズを実現することが確認される。

6 おわりに

本稿では、センサデータベースにおけるデータサイズの削減と問合せ処理高速化を実現するためのデータセグメンテーション手法について提案し、その有効性について検証した。現在、問合せの事前登録を行うことなく多く利用される問合せを自動的に判定し、問合せ結果の事前準備を実現するセグメンテーション手法について検討を進めている。

参考文献

- [1] Stonebraker, M. et al. "The end of an architectural era: (it's time for a complete rewrite)," Proceedings of the 33rd International Conference on VLDB, pp.1150-1160, Sep, 2007.
- [2] 高木ほか, "時系列センサデータベースシステムの初期的検討," 信学ノ大, BS-5-4, Sep, 2009.

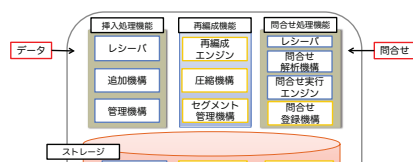


図 1 アーキテクチャ

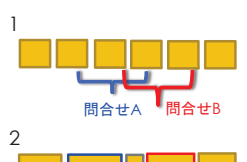


図 2 セグメンテーション

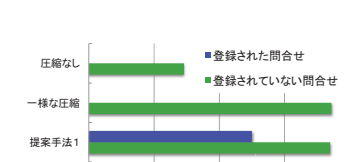


図 3 問合せ処理性能

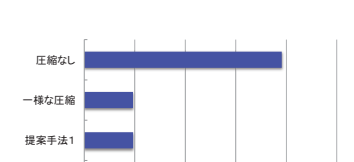


図 4 データサイズ