

D-014

センサデータ集約を用いた実時間クエリ処理機構の設計と実装

A Real-time Query Processing System using Aggregated Sensed Data

森田 達也[†] 金井 圭介[‡] 石塚 宏紀[§] 戸辺 義人^{†*}
 Tatsuya Morita Keisuke Kanai Hiroki Ishizuka Yoshito Tobe

1. はじめに

近年、センサネットワークの普及とともに、センサデータベースの蓄積と処理を行うためのセンサデータベース技術が必要となってきた。センサデータは時々刻々と複数センサから取得されるため膨大なデータ量となる。それに伴ってユーザの予期するクエリ処理時間を大きく逸脱することがある。そこで我々は、ユーザが指定した時間内に応答を返す実時間応答を提案する。提案手法は、まず、データベース内の全センサデータから粒度の違う集約データセットを複数作成する。次に、ユーザが指定した処理時間内に応答可能なデータセットの推測を行い、最後に、適した集約データセットを導き出し、ユーザに応答する。我々は、提案手法のプロトタイプを実装し、簡易実験による実時間応答の検証を行った。

2. 研究課題

膨大なセンサデータを利用するサービスにおいて、クエリ処理に時間がかかり、予期した時間内に応答がないとサービスに支障が生じてしまう。そこで我々は、このようなサービスに対応するために、クエリを分割処理することで一部のデータだけでも実時間応答を可能にする実時間分割応答[1]を提案した。しかし、この手法では、応答するデータの優先度は考慮していないため、要求するデータが優先的に返ってくるとは限らない。そこで本研究では、応答するデータの優先度を考慮した実時間応答を提案する。

3. 提案手法

実時間応答するデータの優先度を考慮するために、我々は、センサデータベース本体から粒度の違うデータセットを抽出することで、様々な粒度のデータセットを作成する。このデータセットから、ユーザの嗜好にあったデータを応答することで実時間応答を実現する。

3.1 処理の流れ

図1にシステム概要図を示す。本システムでは、まず、センサデータベース本体からデータセット群およびデータセット管理テーブルを作成する。次に、ユーザはセンサデータベースに対して制約時間 T 、優先度 P を付加したクエリ $q(T,P)$ を発行する。 $q(T,P)$ を受信したセンサデータベースは、 $q(T,P)$ をプロファイリング機構に送る。プロフ

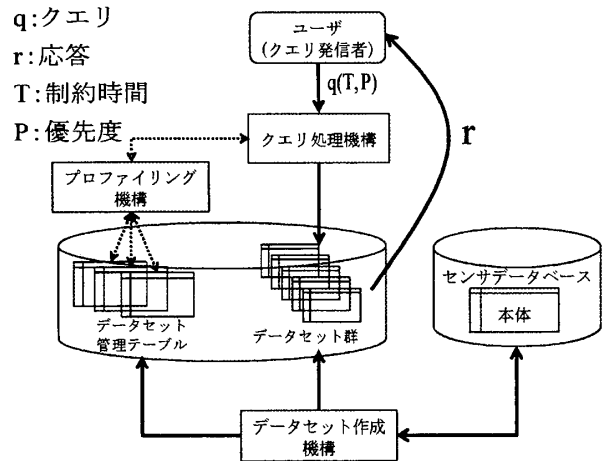


図1: システム概要図

ファイリング機構では、 q および P に適したデータセット管理テーブルと応答順序を決定し、それに基づいてデータセット管理テーブルにアクセスすることで、 T 内に処理可能なデータセットを導き出す。最後に、クエリ処理機構は、プロファイリング機構によって得られた T 内に応答可能なデータセットに対してクエリを発行し、応答する。

3.2 データセット作成機構

データセット作成機構では、センサデータベース本体のすべてのセンサデータをスキャンすることで、粒度の違うデータセット群とデータセット群の状態を管理するデータセット管理テーブルを作成する。そして、ユーザが要求するデータセットから順に応答できるように、データセット管理テーブルでは、データセットを木構造で管理する。

3.3 クエリ処理機構

図2にクエリ処理アルゴリズムを示す。クエリ処理機構では、まず、データセット管理テーブルにおいて、木構造で管理しているデータセット管理テーブルの各節を辿り、データセットが制約時間 T 内に処理可能であるかを比較していく。そして、処理可能ならば応答可能リストに加える。この際、 P により節を辿る順序を変えることで、ユーザが要求するデータから優先的に応答していくことを可能にする。次に、応答可能リスト内のデータセットに対して q を順に発行し、応答していく。

4. プロトタイプの実装と評価

我々は、提案手法の実装を行い、簡易実験を行った。本論文では、センサデータ取得時間に注目した実験を行う。テストセンサデータとして、UScan[2]で実際に街の中

[†] 東京電機大学 工学部 情報メディア学科
[‡] 東京電機大学大学院 工学研究科
[§] 東京電機大学大学院 先端科学技術研究科
^{*} 科学技術振興機構 CREST

Algorithm QueryProcessing Algorithm

```

1. procedure q(T,P)
2.   var
3.     Rquery //クエリで指定される条件の範囲
4.     list //応答可能リスト
5.     tablesize //データセット管理テーブルの項目数
6.     Ti:0 //クエリ処理にかかる時間の合計
7.   for i:=0 to tablesize do
8.     r:=getDatasetInfo(i,P)
           //管理テーブルを辿りデータセット情報を取得
9.     if compareRegion(r.region,Rquery)=0
           //データセットが管理する範囲と重なるか判定
10.      Ti:Ti + getProcessingTime(r)
11.      if(T>Ti) then
12.        addList(list,r)
13.      end if
14.    end if
15.  end for
16.  for j:=0 to list.length do
17.    qj := makeQuery(list[j])
18.    exec(qj);
19.  end for
20. end procedure
    
```

図2: クエリ処理アルゴリズム

にセンサを配置し、細粒度に取得したセンサデータを利用する。本実験では、この UScan のデータの中でも、2007年8月1日から2007年8月20日までの650万タプルのデータを利用した。

4.1 実験データセット

センサデータ取得時間に注目したとき、広い範囲のデータ要求に対して、偏ったデータの応答ではなく、全体のイメージがつかめる応答であることが望まれる。そこで、我々は、時系列データであるセンサ取得時間から、年、月、日のそれぞれの範囲で特徴点を抽出し、それぞれデータセットに格納する。具体的に、時間の変り目のデータを年のデータセットへ、分の変り目のデータを月のデータセットへ、そして、残りのデータを日のデータセットへ格納し、図3のように管理する。この際、上の階層に含まれるデータは下の階層のデータセットに含まれないようにする。このデータセットに対して、幅優先探索を用いて、クエリ処理を行った。

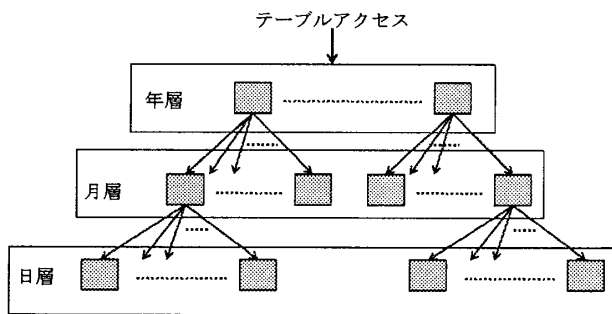


図3: データセット管理テーブルの構造

4.2 実験・評価

本実験では、センサデータ取得時間に着目したデータセットを作成したため、時間幅指定によるデータの要求をする。

```

> 時間幅指定によるデータの要求
SELECT * FROM data WHERE time BETWEEN
'2007-08-01 00:00:00' AND '2007-08-10 00:00:00';
    
```

このクエリを、制約時間 1s から 75s までずらしながら発行し、提案手法の評価を行った。図4に実験結果を示す。

実験結果は、横軸に制約時間、縦軸に各制約時間指定によるクエリ処理時間を示す。実験結果より、本手法を用いることで、制約時間内に応答可能なデータを返していることがわかる。また、制約時間がある値を超えると、対象となるデータセットすべてが応答可能となるため、クエリ処理時間は一定となっている。

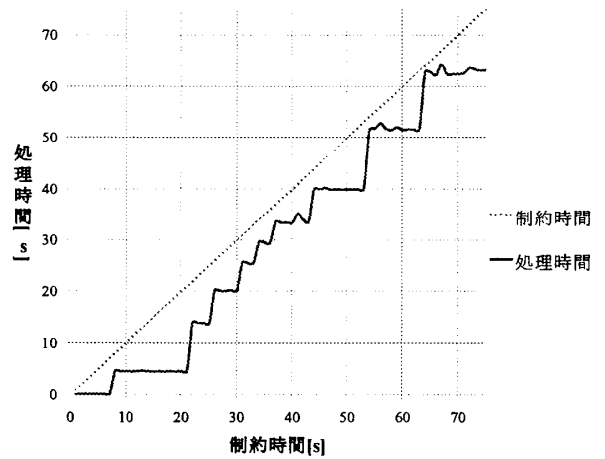


図4: 実時間応答の精度

5. まとめ

本研究では、センサデータベースに対して、粒度の違うデータセット群を作成することで、優先度を考慮した実時間応答を可能とする手法を提案した。また、プロトタイプを作成することで簡易実験を行い、実時間応答の精度を評価した。

今後、次々と取得されてくるセンサデータを想定し、動的なセンサデータベースに対して適応する手法を検討していく必要がある。また、センサデータベース本体からデータセットを作成する手法についても検討していきたい。

参考文献

[1] 金井 圭介, 石塚 宏紀, 戸辺 義人: センサデータベースにおける実時間分割応答の提案, 情報処理学会 第70回全国大会 (2008).
 [2] UScan <http://uscan.osoie.jp>