

センサネットワークのための データベース技術

Database Technologies for Sensor Networks



白石 陽

東京大学空間情報科学研究センター
siraisi@csis.u-tokyo.ac.jp

近年、データベースの分野においても、センサネットワーク技術に関する研究・開発が盛んに行われている。センサデバイスの小型化、無線インフラの整備に加えて、TinyDBをはじめとするセンサノード上の小型データベースシステムが出現したことにより、新しいデータベース技術が発展しつつある。また、広域に分散したセンサネットワークからの膨大なデータストリームを処理する枠組みも必要とされている。本稿では、センサネットワークのための新しいデータベース技術(「センサデータベース技術」)を取り上げ、解説を行う。

●● データベースとしての センサネットワーク

センサネットワークは、ユビキタスコンピューティングや環境モニタリングなどさまざまな応用が期待されており、近年、その要素技術の開発が盛んに行われている¹⁾。特に、多数のセンサノードの集合体であるセンサネットワークから膨大なデータを取得し、用途に応じて適切に加工するための情報処理技術は、センサネットワークを利用したアプリケーションを実現する上で不可欠である。

センサネットワークは、実世界をセンシングすることで、現在の状態や変化をモニタリングする、いわば、環境観測情報を蓄積するデータベースといえる。生活環境や自然環境の現在の状態や過去の時系列をデータベースに保持することで、いつでも、どこでも、誰もが、その情報を参照できることができれば、より便利なアプリケーションの実現が期待できる。しかし、一般に、センサネットワークといった場合、センシング機能と通信機能とわずかな処理能力を持った小型のセンサノードから構成される無線ネットワークを想定することが多く、そのままでは、センサネットワークへの問合せを、既存のデータベース管理システムと同様に行うことはできない。

データベースの分野では、近年、そういったセンサネットワークによって観測されたデータ(センサデータ)に対する処理を行うために、センサネットワークを1つの仮想的なデータベースと見なして、そのデータベースに

対する問合せ(クエリ)を発行し、その問合せを処理するための技術に関する研究が多数行われてきている^{1), 2)}。この場合、問合せを送るユーザ(アプリケーション)とセンサネットワークとの間に、その問合せを理解して、センサネットワークに働きかけ、センサネットワークからのデータを問合せの答えとしてユーザに返す仕組みが必要である(図-1)。本稿では、こういった技術を総称して、「センサデータベース技術」^{☆1}と呼ぶが、センサネットワークの持つ制約やセンサデータの性質から、センサネットワークのための新しいデータベース技術が必要とされる。また、図-1では、センサデータを発生する情報源として、無線センサネットワークだけでなく、インターネットに接続された計算機、カメラ、自動車など、より高機能なセンサノードも想定している。

データベースの分野では、このセンサデータベース技術の研究として、大きく分けて、次の2つの流れが存在する。

A) センサネットワーク内問合せ

ユーザから無線センサネットワークに発行された問合せを、どのように効率よくセンサネットワーク内で処

☆1 「センサデータベース」といった場合、TinyDB や COUGAR などの無線センサネットワークを対象としたデータベースを指すことが多いが、本稿では、新たな展開を期待して、データベース分野におけるセンサネットワークやセンサデータストリームに関する技術をまとめて、「センサデータベース技術」と呼ぶ。

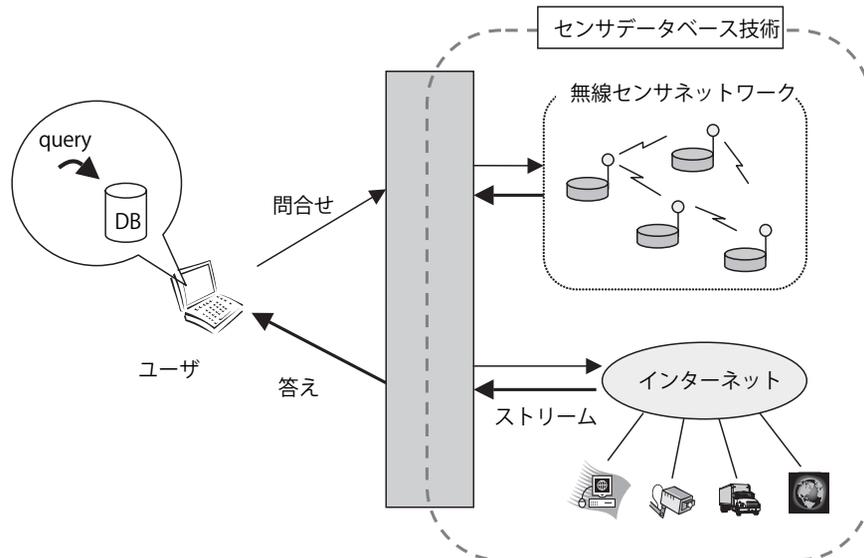


図-1 センサデータベース技術

理するかは重要な研究テーマである。特に、無線センサネットワークでは、電力消費や処理能力などのセンサネットワークの制約を考慮しながら、センサネットワーク内の問合せ処理を検討することが重要であり、このような問合せ処理を、“in-network query processing”（以下、ネットワーク内問合せ処理）と呼ぶ²⁾。ネットワーク内問合せ処理に関する研究は、センサネットワークのインフラ（ハードウェア、OS、ミドルウェア）の整備が進み、研究者が容易に、実験用のプラットフォームとして利用できるようになった頃から進められているが、特に、TinyDB³⁾ というセンサノード上の小型データベースシステムの登場により、データベースの分野でも多くの研究が行われるようになってきている。TinyDB は、TinyOS 上で動作するミドルウェアであり、SQL (Structured Query Language) に似た言語をプログラミングインタフェースとして提供している。SQLを利用したことのある研究者であれば、このインタフェースを利用することにより、センサネットワークのハードウェアやOSの詳細を知らなくてもアプリケーションを容易に記述することができる。

B) センサデータストリーム処理

データストリーム (data stream) は、時々刻々と流れてくるデータの流れであり、特に、センサネットワーク（あるいはセンサノード）を情報源とするデータを対象とするものを「センサデータストリーム (sensor data stream)」と呼ぶ。データベースの分野では、データストリームを効率良く処理するための技術に関する研究が盛んに行われている¹⁾。近年では、データストリームのための集約演算、結合演算など、より高度な演算を対象と

した問合せ処理手法やデータストリームを効率良く処理するアーキテクチャに関する研究が進められているが、特に、データストリームのためのマイニング手法に関する研究が注目されている⁴⁾。データストリームに関する研究の多くは、センサネットワークのみを対象としたものではないが、センサデータの性質を考えると、センサノードから送られる連続的なデータの流れを効率良く処理するための枠組みとして有用である。また、センサネットワークを1つのアプリケーションとして挙げている研究も多い。

●● 問合せの記述とデータモデル

センサネットワークを1つの「データベース」と考えた場合、既存のデータベース管理システム（たとえば、関係データベース管理システム）と同じように、センサネットワークからデータを取得し、処理を行うための問合せを記述できると便利である。また、膨大なセンサノードからのストリームであるセンサデータを、どのような形式で管理するかという点も重要である。

ここで、温度センサと湿度センサを持つセンサノードが多数ネットワークに接続されている環境を想定する。このような観測データを管理する最も単純な方法は、観測データが発生するごとに、ネットワーク上の単一のデータベース(ディスク)に時系列として蓄積していく方法である(図-2)。

図-2において、データベース (sensor_db) の各レコードは、センサノードID (node_id)、気温 (temp)、湿度 (humid)、観測時刻 (observ_time) といった属性から構成される。データベースの分野では、各レコードのこと

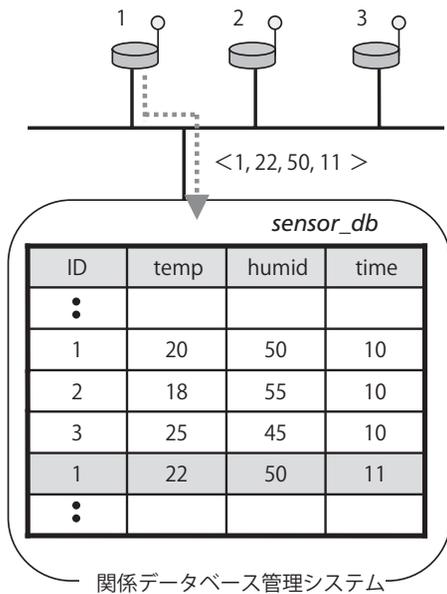


図-2 関係データベース管理システムによる時系列データの管理

をタプルとも呼ぶ。タプルは、属性値の組で表現することもできる。現在、気象情報や河川情報などさまざまな環境情報がWebサイトで閲覧できるようになってきているが、図-2のように、各ノードの観測データを時系列として管理することで、最新のデータのみならず、過去のある時点のデータを参照することも可能になる。

このようなデータベースに対して、たとえば、SQLのSELECT-FROM-WHERE構文を用いて、次のような問合せを記述することができる。

問合せ例(1)

```
SELECT node_id, temp, observ_time
FROM sensor_db
WHERE observ_time > t1 AND observ_time < t2
```

この問合せによって、時刻 t1 から時刻 t2 の間に観測された気温データの値を、ノードID、観測時刻とともに取得できる（この場合、時系列データが取得される）。一般に、SQLでは、SELECTの後に取得したい属性を指定し、FROM句に参照するデータベース（テーブル）の名前、そして、WHERE句に条件を指定する。また、SELECTの後には、AVG（平均値）、SUM（総数）などの集約関数（aggregation function）を指定することもできる。

ネットワーク内問合せ処理およびデータストリーム処理の場合も、ほぼ同様のSQL構文によって、センサデータを取得する問合せが記述できる。ただし、いずれの場合も、問合せ対象であるデータベースの意味や内部構造は、大きく異なる。センサデータの更新間隔が、5分、

10分、1時間など比較的長い時間であれば、図-2の方法でも十分である。しかしながら、地震や災害など緊急性の高いデータであれば、より短い間隔でデータがストリームとして流れてくると考えられ、センサノードの数がより多くなれば、一度に到着するデータの数も増えるため、データベースを更新する際に遅延が生じる可能性があることは容易に想像できる。結果として、問合せの実行にも支障が発生し、特に、観測時刻とデータベース上のデータ更新時刻との間に大きな差が出てくることはリアルタイムアプリケーションにとって問題である。また、無線センサネットワークを考えた場合、各センサノードからリモートのデータベースに生データをストリームとして送信し続けるのは、メッセージ数が増大するため省電力の観点から好ましくない。

したがって、センサデータベース技術では、既存のデータベース管理システムを用いてセンサデータを管理する方法はとらず、別のデータ構造を用いて、各センサノードから連続的に流れてくるセンサデータを表現し、問合せの対象としている。

まず、ネットワーク内問合せ処理では、たとえば、次のように問合せが記述される。

問合せ例(2)

```
SELECT node_id, temp, observ_time
FROM sensors
WHERE temp >= 10
SAMPLE PERIOD 10 sec
```

ここで、“sensors”は、センサネットワーク全体が保持するセンサデータを表す仮想的なテーブルであり、図-2と同様のタプルから構成される（図-3）。実際には、各センサノード上の小型データベースシステムは、同じ属性からなるテーブルを持ち、一定期間ごとに、要求された問合せを実行することで、そのノードに接続されたセンサから観測データを取得し、タプルを生成する。結果として、タプルの生成が各ノード上で繰り返し行われることになるため、sensorsは、概念的には、無限のデータストリームを表現することになるが、各ノード上の小型データベースシステムは、生成されたタプルを短い時系列として保持することを指定されない限り、古いタプルを棄却する。ネットワーク内問合せ処理では、センサネットワークを表す仮想データベースの実体は、このような小型のデータベースシステムの集合体であり、各小型データベースシステムが互いに連携しながら問合せ処理が実行される。

SAMPLE PERIOD節は、ネットワーク内問合せ処理特有の表現であり、各ノード上のデータベースシステム

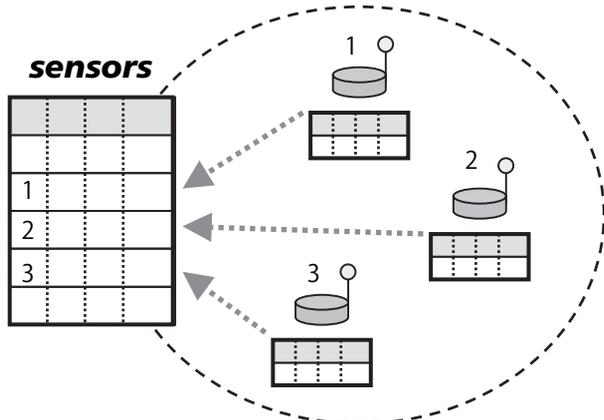


図-3 センサネットワークにおける仮想データテーブル

が問合せを実行する時間間隔が指定される。したがって、問合せ例(2)は、「気温データを10秒ごとに取得し、その値が10℃以上の場合に報告せよ」という意味になる。なお、TinyDB では、通常のSQLと同様の記述が可能であるが、CPUやメモリなど各ノードの計算資源の制約を受けるため、処理コストの大きな問合せ処理を実行するためには工夫が必要である。

次に、データストリーム処理に関する研究では、無線センサネットワークに接続された小型のセンサだけでなく、インターネットに接続された高機能の情報源を想定していることが多い。このようなネットワークに接続された多数の情報源から次々と到着するデータ（すなわちストリーム）を処理するシステムのことを、データストリーム管理システム（Data Stream Management System：DSMS）と呼ぶ。データストリーム処理の場合は、このDSMSに対して、センサネットワークに対する問合せを依頼することになる。DSMSは、無限のタプル列として表現されるデータストリームの中から、有限の部分集合を取り出して、メモリ上で処理を行う。たとえば、データストリームに対する問合せを、次のように記述することができる。

問合せ例(3)

```
SELECT COUNT (temp)
FROM S [RANGE 10 MINUTES]
WHERE temp > 20
```

この問合せは、「過去10分間で、気温20℃以上のタプルの数を数えよ」という意味になる。ここで、*S* [RANGE 10 MINUTES] は、ストリーム*S*を構成するタプルのうち、過去10分間に到着したタプルの列を表す。

このストリームの部分集合に対して、集約演算や結合演算などの問合せ処理が適用される。

いずれの場合の問合せ記述も、センサネットワークによって観測されたセンサデータを表示する方法の違いはあるものの、基本的には、関係データベース管理システムへの問合せ言語であるSQLを拡張したものである。このように、センサネットワークに対する問合せを宣言的(declarative)に記述するアプローチは、プログラマやアプリケーション開発者などが、既存のデータベースシステムと同様に、センサネットワークに対する問合せを容易に記述できるという利点がある。

なお、本稿では、センサネットワークやデータストリームに対する問合せ記述の方法として、代表的かつ理解しやすいと考えられるSQL拡張言語を利用するアプローチを取り上げて説明を行っているが、オブジェクト指向やXMLの考えを取り入れた、より記述性の高い問合せの方法も提案されている。

●● センサネットワークのための問合せ処理

■ センサネットワーク内の問合せ処理

センサネットワークに関する研究では、多くの場合、無線環境での設置を想定しており、センサノードはバッテリー駆動であり、処理性能や通信帯域も制限される。野生動物の習性モニタリングなど屋外に設置された無線センサネットワークを考えた場合、各センサノードのメンテナンス作業が大変であるため、バッテリーの寿命を伸ばすために、各センサノードの電力消費をいかに抑えるかが重要となる。センシングや計算処理よりも他のノードとの通信の際の電力消費が大きい場合、各ノードで発生するメッセージ数を抑えながら必要な問合せ処理を実現することが研究課題となる^{2), 3), 5)}。代表的な研究プロジェクトとしては、TinyDB^{☆2}やCOUGAR^{☆3}が挙げられる。ネットワーク内問合せ処理は、大きく分けて、次の2つのフェーズで実行される(図-4)。

- 1) 問合せ分配フェーズ(query dissemination)
- 2) データ収集・集約フェーズ(data collection/aggregation)

たとえば、問合せ例(2)のようなセンサネットワークに対する問合せが発生すると、センサネットワークに接続されたクライアントコンピュータ上で、まず、問合せ

☆2 <http://telegraph.cs.berkeley.edu/tinydb/>

☆3 <http://www.cs.cornell.edu/database/cougar/>

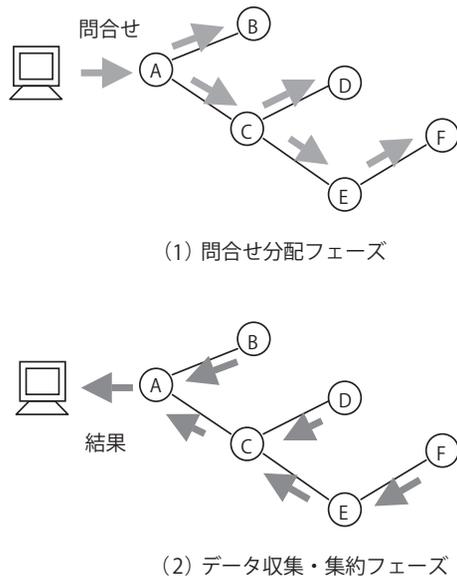


図-4 ネットワーク内問合せ処理

の最適化が行われて、問合せプランが生成される。この際、センサネットワークに関するメタデータが最適化のパラメータとして利用される。そして、ルーティングツリーが作成され、そのツリー構造に基づいて、根ノードから葉ノードに向けて、各ノードへ問合せが分配される。各ノード、すなわち、各小型データベースシステムは、SAMPLE PERIOD 節に指定された時間間隔で、与えられた問合せを実行し、センサデータを取得し、親ノードに通知するだけでなく、子ノードから受信したセンサデータを親ノードに転送する。結果として、各ノードによって観測されたセンサデータは、マルチホップで、根ノードに転送されることになる。この時、各小型データベースシステムは、WHERE句で指定された条件を満たすセンサデータのみを親ノードに転送するため、各ノードが観測データをそのまま送信する場合に比べて、センサネットワーク内を流れるメッセージ数を軽減することができる。

さらに、ネットワーク内問合せ処理では、データ収集と同時に、集約処理を行うことによって、メッセージ数を軽減することができる。たとえば、次のような問合せを考える。

問合せ例(4)

```
SELECT COUNT (id)
FROM sensors
SAMPLE PERIOD 1 min
GROUP BY temp/10
```

この問合せは、「1分ごとに、各データ範囲(0～10℃、10～20℃など)に含まれる値を出力するセンサノードの数を数えよ」という意味になる。各ノードが、子ノードからの観測データだけでなく、そのデータと自分自身の観測データを用いて計算した中間結果(この例の場合、各範囲ごとの集計表)を保持し、さらに親ノードに転送していくことで、データの収集と同時に、データの集約を行うことができる。

さらに、MITのS. Madden らのグループは、TinyDBの枠組みを拡張して、Acquisitional query processing (ACQP) を提案している³⁾。ネットワーク内問合せ処理では、各ノードがセンサデータを取得し、処理を行っている。そこで、そのデータの取得と処理を、いつ、どこで、どのような頻度で行うか、という点に着目し、センサノードの電力消費を考慮しながらさまざまな問合せ処理を行う手法を提案している。文献³⁾では、応用例の1つとしてウミツバメの習性モニタリングを挙げている。以下が、その問合せ例である。

問合せ例(5)

```
SELECT node_id, light
FROM sensors
LIFETIME 30 days
```

この問合せは、lifetime-based queryと呼ばれ、LIFETIME節を指定することにより、指定した期間(30日間)にわたって、各ノードの照度センサの観測値(light)を通知することを要求している。ACQPは、指定した期間の間、センサノードを稼働させ続けるために、サンプリングや通信にかかる消費電力を計算し、その結果を考慮して問合せプランを生成し、指定期間中の各ノード上の問合せ処理のスケジューリングを行う。

その他、TinyDBの枠組みでは、各ノード上のデータベースシステムが最近の時系列を保持する有限長の小さなテーブルを作成することができるため、過去数分間の観測データの平均値の算出や、過去の測定値より大きなデータの計数などを行うことができる。また、指定したイベントが発生した時に問合せを実行する仕組みも提供している。

■データストリーム処理

データストリーム管理システム(DSMS)では、連続して到着するデータストリームに対して、リアルタイムに応答するために、あらかじめデータストリームに対する問合せをDSMSに登録しておき、その問合せ処理をDSMS上で連続的に行う方法がとられる。一般に、このような問合せの形式を、連続的問合せ(Continuous

プロジェクト名	研究機関	URL
STREAM	Stanford Univ.	http://www-db.stanford.edu/stream/
Borealis / Aurora	MIT, Brown Univ., Brandeis Univ.	http://nms.lcs.mit.edu/projects/borealis/
TelegraphCQ / Fjord	UC Berkeley	http://telegraph.cs.berkeley.edu/
Nile	Purdue Univ.	http://www.cs.purdue.edu/Nile/
Gigascop	AT&T	http://public.research.att.com/viewProject.cfm?prjID=129

表-1 ストリーム研究プロジェクトの例

Query)と呼ぶ。

この問合せを、時々刻々と到着するデータストリームに対して、繰り返し、リアルタイムに近い速度で行うためには、図-2のように到着したデータを一度ストレージ(ディスク)に格納するのではなく、それらのストリームをメモリ上で処理することが必要となる。しかしながら、本質的にデータストリームは無制限のタプル列として表現されるので、DSMSにおいて、受信したデータストリームから、演算の対象となるタプル列を取り出す作業を行う必要がある。

この方法の1つが、問合せ例(3)に挙げたような、データストリームから連続したタプル列を切り出す問合せであり、一般に、スライディングウィンドウ(sliding window, 滑り窓)と呼ぶ。前述の例は、時間幅をパラメータとしており、time-based sliding window と呼ばれる。時間幅ではなくタプルの個数を指定した場合は、tuple-based sliding window と呼ばれ、 S [ROWS 1000] と指定した場合には、最新の1,000個のタプル列を切り出すことができる。問合せ例(3)では、1つのストリームから取り出したタプル列に対して集約演算を適用しているが、たとえば、FROM句の後に、複数のストリームを指定した場合には、それらのストリームを結合する処理が行われる。

データストリーム処理に関する研究は、DSMSのアーキテクチャに関する研究、データストリームに対する複数の問合せを最適化する方法から、ストリームマイニングといったより高度な問合せ処理に関する研究まで多岐にわたる。これまでも、多くの研究機関によってデータストリームに関する研究が行われてきている(表-1)。

イベントの検出や変化の予測などのモニタリングアプリケーションを考えた場合には、過去の時系列データと最新のデータストリームを比較したり、リアルタイムで到着するセンサデータと他の情報源のデータとを統合したりすることも重要である。最近では、データストリー

ムを時系列として蓄積していく方法に関する研究や、ネットワーク内問合せ処理の枠組みとデータストリーム処理の枠組みを統合する試みも行われており、今後の展開が注目される。

●●地理センサネットワーク

前章までは、主に、ネットワーク内問合せ処理やデータストリーム処理に関する説明を行ってきたが、その他のセンサネットワークのためのデータ処理技術に関する研究も行われている。本章では、その中で、筆者が特に注目している地理センサネットワークに関する研究について紹介する。

地理センサネットワーク(“Geo Sensor Networks”)は、地理的な空間に設置されたセンサネットワークのことを指すが、近年、その地理センサネットワークに関したさまざまな研究が行われつつある。その内容は、ルーティングなどのネットワークレベルの研究から、データ処理やアプリケーションに関する研究まで多岐にわたるが、地理的な位置情報を考慮することが重要となる。前述したネットワーク内問合せ処理に関する研究においても、地理情報システム(GIS: Geographic Information System)で提案されている手法を取り入れた研究が行われつつある⁵⁾。

環境モニタリングは、地理センサネットワークの応用の1つであり、問合せに空間的な範囲を指定することで、その範囲のセンサデータを収集し、センサノードの緯度経度座標に基づいて、地図上に点分布として表示することができる。しかし、モニタリングの対象が、気温分布や大気汚染状況などの連続的に分布するものである場合、センサノードの設置された地点のデータだけでなく、センサノードの設置されていない場所のデータを何らかの方法で推定する必要が生じる。

地理情報システム分野では、その空間推定の方法と

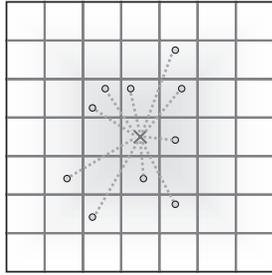


図-5 空間補間の例

$$z(x) = \frac{\sum_i w_i z_i}{\sum_i w_i}$$

$$w_i = \frac{1}{d_i^2}$$

して、空間補間 (spatial interpolation, 空間内挿) という手法が提案されており、実際に、GISのツールに、IDW (Inverse Distance Weighted) やクリギング (Kriging) といった手法が実装されている。クリギングは空間統計学に基づく手法であり、文献5)でも取り上げている手法であるが、ここでは、IDWを取り上げて、簡単に、空間補間について説明する。IDWは、周辺のデータに重み付けをすることによって、ある地点の推定値を計算する。

図-5のメッシュの中心点(x)の推定値(z)を求める場合、周辺の各ノードの値(z_i)を、中心点(x)までの距離(d_i)の逆数(w_i)で重み付けをして計算を行う。

この計算を、図-5のメッシュを構成する各セルの中心点ごとに適用することにより、メッシュ状の連続的な空間分布を生成することができる。これによって、任意の地点のセンサデータの値を推定することができる。

ネットワーク内問合せ処理に関する既存研究では、問合せに空間的な範囲が指定されても、収集されたデータに対して単純な集約演算を適用している場合が多い。たとえば、平均値を求める場合には、センサノードの分布が均等であっても、偏りがあっても、同じ結果が計算され、その結果が必ずしも意味のあるものとは限らない。そこで、文献5)では、連続的に分布する観測対象に対する問合せを実現するために、空間補間の手法をネットワーク内問合せ処理に適用している。ただし、一般に、空間補間の手法は、計算にかかるコストが大きく、通常は、1つの計算機で処理される。したがって、空間補間のための計算を、いかにセンサネットワーク上の各ノードに分散させ、それらのノード間のメッセージを少なくして処理を行うかがポイントとなる。文献5)で提案している手法では、遠くのセンサデータが近くのセンサデータよりも推定値に与える影響が少ないということに着目して、クリギングの計算に必要な周辺のセンサノードの部分集合を逐次的に計算することで空間推定を行っている。

●●まとめ

本稿では、センサネットワークのためのデータベース技術として、主に、ネットワーク内問合せ処理 (in-network query processing) とデータストリーム処理 (data stream processing) に関する解説を行った。また、センサデータベース技術に関する他の研究事例として、地理センサネットワークに関する研究について紹介した。

ユビキタスコンピューティング、高度交通システム、防災、農業、環境モニタリングなどのセンサネットワークの応用を考えた場合、その情報処理にかかわるセンサデータベース技術は、今後、ますます重要になっていくと考えられる。本稿では、基本的な考え方の説明を優先させたが、「センサデータベース技術」は、実に幅広い研究テーマを含む分野であり、これからの発展が期待される。より詳しい内容に興味のある読者、あるいは、最新の動向について知りたい読者は、文献1)などの参考文献およびSIGMOD, VLDB, ICDEなどのデータベース関係の国際会議の予稿集を参照されたい。

謝辞 本解説記事の執筆を進めるにあたり、筑波大学助教授石川佳治先生には、有用なご助言と貴重なご意見をいただきました。ここに感謝いたします。

参考文献

- 1) 安藤 繁, 田村陽介, 戸辺義人, 南 正輝 編著: センサネットワーク技術-ユビキタス情報環境の構築に向けて-, 東京電機大学出版局 (2005).
- 2) Gehrke, J. and Madden, S.: Query Processing in Sensor Networks, IEEE Pervasive Computing, Vol.3, No.1, pp.46-55 (2004).
- 3) Madden, S., Franklin, M. J., Hellerstein, J. M. and Hong, W.: TinyDB: an Acquisitional Query Processing System for Sensor Networks, ACM Transactions on Database Systems, Vol.30, No.1, pp.122-173 (2005).
- 4) 有村博紀, 喜田拓也, データストリームのためのマイニング技術, 情報処理, Vol.46, No.1, pp.4-11 (Jan. 2005).
- 5) シルビアニッテル, グァンジン, 白石 陽: センサネットワークにおけるネットワーク内空間推定, 電子情報通信学会論文誌A, Vol.J88-A, No.12, pp.1413-1421 (2005).

(平成18年3月2日受付)

