

データストリーム処理基盤を用いた交通渋滞情報提供システム

喜田 弘司・中村 暢達
今井 照之・藤山 健一郎

要旨

深刻化する交通渋滞や環境問題の解決に、ITを活用した高度道路交通システム（ITS）が期待されています。ITSの実現のためには、自動車（移動体）からセンターサーバにアップロードされるデータを大量に収集し、リアルタイムに分析する情報通信技術が重要です。本稿では、このような大規模データ収集・分析を実現するデータストリーム技術について紹介します。更に、本技術を渋滞情報提供システムに適用することにより、大量の車両からの位置・速度データを高速に処理し、鮮度の高い渋滞情報の提供が可能であることを検証しましたので、その結果を報告します。

キーワード

●データストリーム ●大規模データ収集・分析 ●ITS ●渋滞 ●運輸

1. はじめに

近年、交通渋滞は大きな社会問題となっています。燃費の悪化によるエネルギーのロスだけではなく、CO₂、NOx（窒素酸化物）、PM（粒子状物質）などの大気汚染物質の排出増加による環境の悪化が深刻化しています。この問題に対し、信号の制御といった交通コントロールが重要視されています。

交通コントロールの実現手段として、ITを活用して様々な情報を収集・分析し、ドライバーに提供する高度道路交通システム（ITS）の実現・普及が期待されています。例えば、道路を走っている1台1台の車両をセンサに見立て、取得した走行データをセンターサーバで収集・分析することにより、より正確でリアルタイムな渋滞情報を作り出すことができます。この情報をもとに、混雑している道路を回避するように走行経路を案内することで渋滞の悪化を防げます。

NECでは、このようなサービスを実現するために、多数のデータを効率的に収集・分析する基盤ソフトウェアが重要と考えています。現在、GPS、携帯電話は広く利用されており、位置、速度などの走行データをセンターサーバに集め、これらのデータを加工・活用するインフラは整いつつあるといえます。しかしながら、大量の自動車から、次から次へと走行データが発生すれば、データ量は膨大となります。そのため、データを収集・解析し、その結果を使ってリアルタイムにサービスを提供するには、一般に高性能な計算機が必要となります。

この問題を解決すべく、大量に連続発生するデータを収集しながら、流れ作業的な方法で高速に解析するデータストリーム処理基盤を開発しました。本稿では、データストリーム処理基盤を実現する要素技術を紹介し、本基盤を用いて試作した渋滞情報提供システムについて述べます。

2. データストリーム処理基盤

ログデータ、センサデータなど、一つひとつのデータは小さなサイズ（数百バイト）であっても、数百万カ所からデータを収集すると、1秒当たり数十万メガバイトのデータをセンターサーバで処理する必要が出てきます。通常のデータ処理の場合、これらのデータを一旦データベースに格納し、バッチ処理により解析を行います。大量のデータを処理するには、高性能計算機が必要となります。更に処理すべきデータが増大すれば、想定時間以内にバッチ処理が終了しない、過負荷のために性能が低下する、データあふれが発生する、データを紛失するといった問題が発生する可能性があります。

そこで、NECでは、図1に示すように、データベースを用いず、データをためることなく、データが発生した時点あるいは受信した時点で、順次データをパイプライン的に分散処理するデータストリーム処理技術の研究開発を進めています。データを収集する過程で、データをサンプリング、フィルタリング、クレンジング、平均化するといった処理を行う

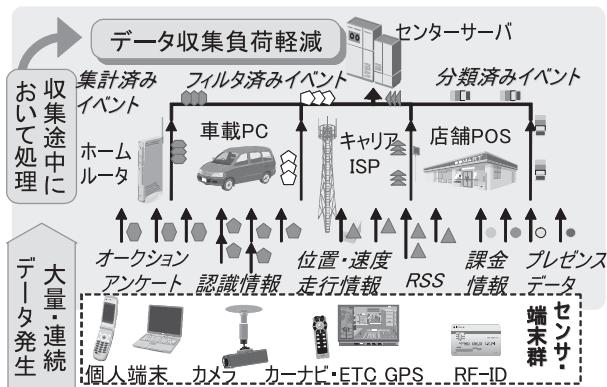


図1 データストリーム処理

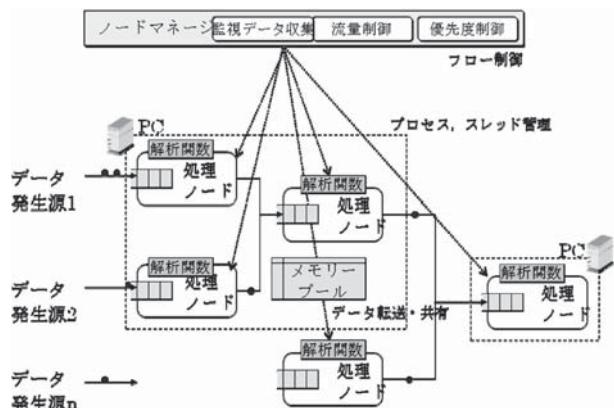


図2 データストリーム処理基盤アーキテクチャ

ことでデータ量を削減しセンターサーバーの負荷を軽減します。

図2にデータストリーム処理基盤のアーキテクチャを示します。各処理単位をノードと呼び、そのノードを組み合わせることで、1つのアプリケーションを構築します。各ノードはスレッドであり、独立して（非同期で）動作しますが、ノードマネージャと呼ばれる管理モジュールが、処理の開始・停止など全ノードの制御を行います。

ノードは、同一マシン内にあっても、異なるマシンにあっても構いません。異なるマシンにあるノード間で通信する場合には、送受信用のノードを使って、処理データの受け渡しを行います。同一マシン内にあるノードにおいては、共有メモリを用いた本基盤独自のメモリ管理機構により、データの受け渡しをします。このメモリ管理機構は、小さなサイズのデータを逐次処理することに最適化しており、高速なデータ

の受け渡し及びキューマネージメントを実現します。

現時点では、本基盤はC++のクラスライブラリとして提供しています。開発者は、このクラスライブラリを利用して、ノードごとにプログラムを実装します。各ノードでは、処理のトリガーとなる条件の指定と、データ処理のプログラムを記述します。そして、ノードの接続を指定しますが、複数の接続先を指定することも可能です。例えば、データの流れを分岐させるように指定することで、1つのデータソースに対して異なる処理を並列的に行う、またデータの流れを集約させるように指定することで、複数のデータを統合して高度な分析を行うといったことができます。

3. 渋滞情報提供システムの試作

第2章で述べたデータストリーム処理基盤を用いて渋滞情報提供システムを試作しました。約2,000本の道からなる道路ネットワーク上に、市販の交通シミュレーションソフトウェアNETSIMを使って仮想的に車両を走らせ、その位置・速度・方向をセンターサーバーに送信します。センターサーバーでは、車と道路との対応付け（マップマッチング）を行い、道路区間にごとに渋滞の度合いを算出します。図3は、渋滞状況を示した画面の例で、太線が渋滞、太点線が混雑、そして細線は車が順調に走行していることを表しています。

本システムの特徴は次の通りです。



図3 渋滞情報提供システム画面例

現状



幹線道路のみ、数分ごとに更新

将来(NECの新技術を適用)



細街区を含め、数十秒ごとに更新

図4 従来方式と本方式で提供する渋滞情報の比較

(1) きめ細かく渋滞情報を提供

渋滞情報提供の仕組みとしては、VICSがよく知られています。図4に本システムとVICSが提供する渋滞情報の違いを示します。VICSは、主に道路に設置されたセンサから情報を収集することで渋滞情報を作成しています。そのため、対象は主要幹線道路に限られ、範囲も大まかです。渋滞を解消するためには、主要幹線道路だけでは不十分で、細街区を使って、混雑している道路を回避するように経路案内をする必要があります。

本システムでは、細街区も含めた道路を対象に渋滞情報を生成することができます。

(2) 鮮度の高い渋滞情報を提供

従来の多くの渋滞情報は、20分から30分程度の間隔で更新さ

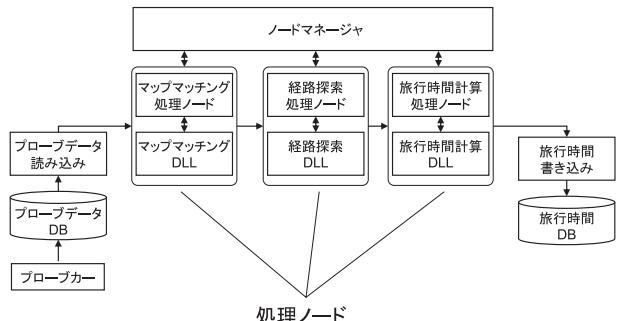


図5 データストリーム処理による渋滞情報生成

れています。つまり、ドライバーへ提示されている情報は20分から30分前の情報です。渋滞を解消するためには、渋滞の起き始めに対処することが重要であることが知られています。つまり、可能な限り渋滞情報の遅延を少なくすることが重要となります。

本システムでは、5分間隔で更新することが可能です。このように鮮度の高い渋滞情報を用いることで、信号の制御やカーナビによる経路案内において、より適切な交通コントロールを促すことができます。

以上のように、きめ細かく鮮度の高い渋滞情報は、図5に示したデータストリーム処理により実現しています。

解析の種類ごと、すなわちマップマッチング、経路探索、旅行時間の計算それぞれが処理ノードです。これら処理ノードは、本基盤のノードマネージャの管理下で並列実行します。各処理ノードは互いに独立に解析できるわけではなく順序関係があるため、他の処理ノードの処理状況次第では、「待ち」状態に陥ることがありパフォーマンスが低下します。こういったことが発生しないように、ノードマネージャは処理ノードの処理待ち状態を極力少なくするように処理全体を制御します。

以上のように、解析の種類単位に処理ノードとし、これらの実行をノードマネージャで調整することで高速な並列分散解析を可能としています。

4. 性能評価

開発したストリームデータ収集・分析技術は、シミュレーション及び名古屋市内で実施された実証実験において有効性を確認しています。

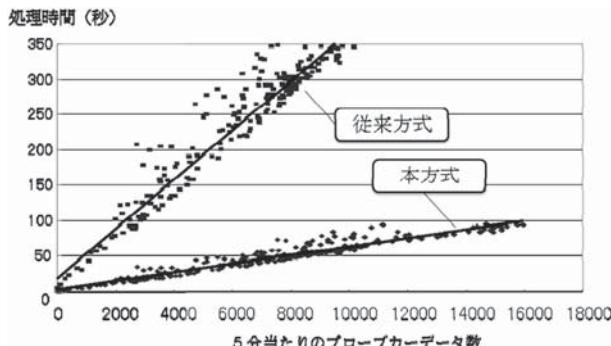


図6 渋滞情報処理に要する時間

シミュレーションでは、交通シミュレータにより5万台の自動車の走行データを生成します。このデータに対して本技術を用いて、1台のPC (CPU : Pentium 4 3.0GHz、RAM : 512Mバイト、OS : Windows Server 2003) で走行データを収集・分析し、6万ある全道路区間での平均速度を計算し、地図上に表示します。1万台規模の自動車を対象とした場合、従来方式では数十分の遅延が発生していましたが、本方式では1分以内に縮まり、鮮度の高い解析結果を提供できることが分かりました。

また、名古屋市内のタクシーを使って実施されていたITS実証実験において、1台のPC (CPU : Pentium Xeon 2.8GHz、RAM : 1.5Gバイト、OS : Windows Server 2003) 上で本技術を部分的に適用する実験を行いました。図6に遅延時間の計測結果を示します。本技術を適用することで、約6倍の高速化に成功しました。

5. おわりに

以上に述べましたデータストリーム処理技術に関する研究としては、ブラウン大学のAuroraプロジェクト、スタンフォード大学のSTREAMプロジェクトなどがあります。これらの研究は、主にクエリ言語、処理アーキテクチャ、分析アルゴリズムに関するものです。一方、製品としては、Apamaが金融分野への適用でよく知られています。また、内部統制に関連して、多くのログ収集製品が最近目立っています。このように、国内外で大量のデータを効率良く処理する技術の研究開発が活発化してきています。

NECで研究開発を進めていますデータストリーム処理基盤

は、大量かつ連続発生するデータを収集・分析する技術に関し、計算機資源の効率性（少ない計算機でシステムを実現）、高スループット性（リアルタイムで解析を実行）を特徴としています。

本稿では、渋滞情報提供システムへの応用例を述べましたが、本処理基盤は、大規模サーバの省電力運用管理といった環境改善を始め、膨大なデータを高速に処理する必要がある様々な問題への適用が可能です。今後は、基盤そのものの機能・性能を向上させるとともに、環境対策を始めとする様々な問題の解決に貢献したいと考えています。

*VICSは、財団法人道路交通情報通信システムセンターの登録商標です。

*Apamaは、プログレス ソフトウェア・コーポレーションの米国およびその他の国における商標または登録商標です。

*Netsimは、株式会社フェニックスリサーチの製品です。

*地図データはインクリメントP株式会社の製品MapDKを用いて作成されたものです。

*Pentium、Xeonは、米国およびその他の国における、Intel Corporationまたはその子会社の商標または登録商標です。

*Windows、Windows Serverは、米国Microsoft Corporationの米国及びその他の国における登録商標または商標です。

執筆者プロフィール

喜田 弘司
サービスプラットフォーム研究所
主任

中村 嘉達
サービスプラットフォーム研究所
主任研究員

今井 照之
サービスプラットフォーム研究所

藤山 健一郎
(独) 情報通信研究機構