

位置情報処理へのストリームデータ処理基盤の応用

伊藤 大輔 西澤 格 今木 常之 檜山 俊彦 藤原 真二

株式会社日立製作所 中央研究所 〒185-8601 東京都国分寺市東恋ヶ窪一丁目 280 番地

E-mail: {daisuke.ito.mq, itaru.nishizawa.cw, tsuneyuki.imaki.nn, toshihiko.kashiyama.ez, shinji.fujiwara.yc}@hitachi.com

あらまし 近年、大量の時系列データを処理する手段としてストリームデータ処理技術が注目されている。一方で GPS 付き携帯電話や通信機能つきカーナビの登場により、プローブデータを用いた位置情報処理を行うインフラが整いつつある。本稿ではストリームデータ処理技術を位置情報処理に応用した一例としてプローブカーを用いた渋滞検出のデモを行い、インメモリの差分演算により集計処理が高速であり、かつ CQL によって処理内容を容易に修正できるというストリームデータ処理技術の長所が位置情報処理に適することを実証する。

キーワード ストリームデータ処理, 位置情報処理, プローブカー, 可視化

Stream Data Processing System and its Application to Location Information Processing System

Daisuke Ito, Itaru Nishizawa, Tsuneyuki Imaki, Toshihiko Kashiyama and Shinji Fujiwara

Central Research Lab., Hitachi Ltd., 1-280, Higashi-koigakubo, Kokubunji-shi, Tokyo, 185-8601, Japan

E-mail: {daisuke.ito.mq, itaru.nishizawa.cw, tsuneyuki.imaki.nn, toshihiko.kashiyama.ez, shinji.fujiwara.yc}@hitachi.com

Abstract A stream data processing system has been gathering attention as a means to process large scale time-series data. On the other hand, an infrastructure of a location information processing system using probe device data is being put into place along with the population of GPS cell phones and wireless car navigation devices. In this poster paper, we have made a prototype of a congestion detection system as an example of applications of location information processing systems using the stream data processing system and suggested that a high-speed calculation mechanism by using only differentials and an ease of development by using CQL are advantages of the stream data processing.

Keyword Stream Data, Location Information, Probe Car System, Visualize

1. はじめに

情報爆発時代の到来に伴い、大量の時系列データをリアルタイム処理する技術としてストリームデータ処理技術が注目されている。ストリームデータ処理技術の適用範囲は株の取引情報や株価情報に始まり、サーバやネットワークのログモニタリングや物流のモニタリングなど多岐におよんでいる。

また、情報爆発時代に特徴的なデータソースとして、GPS (Global Positioning System) 携帯電話や通信カーナビなど、位置情報を絶え間なく送信し続ける携帯端末があげられる。これら携帯端末をプローブとみなし位置情報を処理することで従来よりも詳細な交通情報を得ることができると期待されているが、一方で位置情報処理はデータ量が莫大で処理量が非常に多いという特性を持つ。

そこで、我々は位置情報処理にストリーム処理ミドルウェアを適用した例として、プローブカーを用いた

渋滞検出のデモアプリケーションを開発した。デモアプリケーションの開発を通して、ストリームデータ処理が高速であることと、一度実装した処理内容を容易に修正できアプリケーション開発期間を短縮できることを実証した。

2. 従来研究との比較

従来からプローブカーシステムに関する研究は行われてきたが、特に近年になってストリームデータ処理技術を応用した研究^[1]成果が発表されている。[1]ではストリームデータ処理ライブラリを用い数千台の実際のプローブカーからのデータをニアリアルタイムに処理しており、プローブカーシステムのリアルタイム処理の実用化に近い事を実証している。

[1]と比較し、本稿ではストリームデータ処理ミドルウェアである uCosminexus Stream Data Platform (以下, uCSDP と略す) を用いてリアルタイム処理を実現

した事が特長である。汎用ミドルウェアである uCSDP を用いたことで渋滞検出処理を CQL (Continuous Query Language) で記述できるようになり、その結果、渋滞検出処理の実装と修正が容易になった。

3. プローブカーを用いた交通情報処理

従来、交通情報は道路上に取り付けられた各種センサを用いて収集されてきた。日本国内では既に VICS (Vehicle Information and Communication System) として 1996 年から交通情報を車両に配信するサービスが提供されている。しかし道路上にセンサインフラを整備するコストは非常に高いため主要国道を中心とした限定された道路にしか設置されておらず、道路カバー率は十分に高いとはいえない。

一方で現在の自動車には速度センサや外気温センサ、さらには ABS (Anti-lock Brake System) の加速度センサやオートワイパーの降雨センサといった多数のセンサが導入されている。これらセンサ情報を多くの自動車から収集する事で渋滞情報や所要時間情報のみならず、気象情報や路面の凍結情報などさまざまな情報を取得できる。

自動車をあたかもセンサインフラの一部とみなして情報を収集することで、交通情報を収集するシステムを安価に構築できる。このようなシステムは、自動車を探針 (プローブ) になぞらえプローブカーシステムと呼ばれている。またセンサインフラの一部となり情報を収集する自動車はプローブカーと呼ばれている。

プローブカーシステムでは上記の通り既に自動車に搭載されたセンサを利用することを前提としているが、現状ではセンサ情報をリアルタイムかつ安価に収集する通信インフラとサーバシステムが無いため、一部の先進ユーザがバッチ処理的に利用するに留まっている。それでもなおプローブカーシステムは非都市部などへ安価に交通情報処理を展開する手段として有望であると注目を集めている。

4. ストリーム処理技術

従来のデータ処理システムでは一般的に関連データベースシステム (以下、RDB と略す) が用いられて来た。RDB を用いて大量のデータを短いレスポンスタイムで処理する際の問題点は主に以下の 2 点である。

1. 一度データをディスクに保存する
2. 保存したデータに対して後からバッチ的にクエリを適用する

近年においては主記憶コストが下がりインメモリデータベースが普及した事から問題 1 の影響は少なくなっているものの、問題 2 は RDB を用いる限り常につきまとう問題である。

そこで、従来の RDB とは発想を逆転した以下の特徴を持つデータ処理システム [2] が提案されている。

1. あらかじめクエリをシステムに登録する
2. データがシステムに到来すると同時に差分的にクエリを適用する

このシステムは流れてくるデータを逐次処理する事から、ストリームデータ処理システムと呼ぶ。またその基幹技術をストリームデータ処理技術と呼ぶ。

ストリームデータ処理技術ではスライディングウインドウを用いてストリームデータとリレーショナルデータに効率よく変換する。またリレーショナルデータに対して SQL と同等の処理を容易に記述できるよう、SQL にストリームデータとリレーショナルデータの変換演算を追加した CQL というクエリ記述言語を用いる。さらにリレーショナルデータに対する処理のうち特に集計処理が高速に行えるよう、集計演算をインメモリで差分のみ演算する。

ストリームデータ処理技術によって絶え間なく流入し続けるデータを短いレスポンスタイムで処理できるようになると、我々は「今」を分析できるようになる。さらに分析結果から得た知見を元に、あたかも PDCA サイクルを回すように分析クエリを逐次改善し続ける事で、より正確な分析結果を得られる。日立製作所ではストリームデータ処理ミドルウェアである uCSDP を製品化した。uCSDP は従来のデータ処理システムと比べ 2 桁以上の高速化を達成している。

5. 位置情報処理へのストリームデータ処理の応用

プローブカーシステムはデータ量が膨大かつストリーミング的に絶え間なく流入する事から、我々はストリームデータ処理技術が適すると考えている。そこで uCSDP を用いたプローブカーシステムのプロトタイプを作成した。本システムはノート PC 上でも実行可能なほど少ないリソースで実行可能でありながら数千台の車からのプローブデータをリアルタイムに処理して車両の速度計算と渋滞の検出を行える。また処理結果を KML (Keyhole Markup Language) に変換することで、各種 GIS (Geographic Information System) アプリケーションとも連携可能である。図 1 にプローブカーシステムプロトタイプの構成図を、図 2 に GIS ツールを用いて渋滞検出結果を可視化した例を示す。

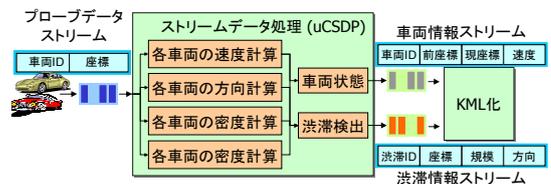


図 1: プローブカーシステムプロトタイプの構成図

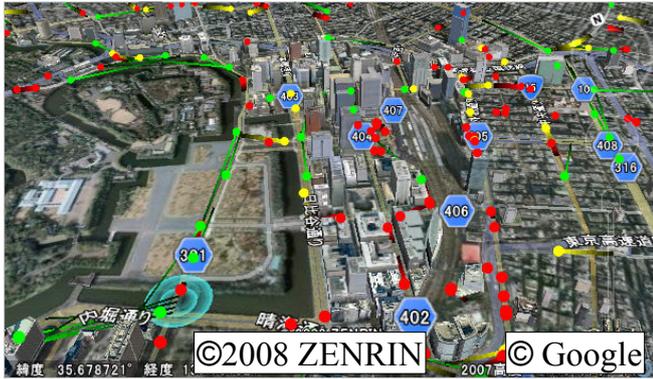


図 2 : GIS ツールを用いて渋滞検出結果を可視化した例

6. まとめと今後の課題

本稿ではストリームデータ処理ミドルウェアを位置情報処理のアプリケーションであるプローブカーシステムに応用し、その有用性を示した。ストリームデータ処理ミドルウェアを用いるメリットは主に以下の 2 点であり、継続的に進化するシステムを構築する上で都合が良いことが分かる。

1. インメモリの差分演算による高速な集計処理
2. CQL によって処理内容を容易に修正可能

今後の課題として、ストリームデータ処理技術に位置情報処理に特化した機能を導入する事が考えられる。例えば RDB に位置情報向けの R-Tree インデックスが導入されたように、ストリームデータ処理技術にも位置情報処理向けの機能を導入する事で、位置情報処理の更なる高速化が期待できる。

参 考 文 献

- [1] 喜田弘司, 藤山健一郎, 今井照之, 中村鴨達, “データストリーム処理による大規模プローブカーシステムの開発と評価”, 信学会 ITS Vol.2008 No.83, pp.1 - 8, Sep. 2008.
- [2] A. Arasu, et al., “STREAM : The Stanford Stream Data Manager”, IEEE Data Engineering Bulletin, Vol.26, Mar. 2003.