

平成22年5月28日現在

研究種目：若手研究（B）
 研究期間：2008～2009
 課題番号：20700078
 研究課題名（和文） 異常補正・状況認識・確率計算を支援する高性能なセンサ情報処理基盤に関する研究
 研究課題名（英文） Research on a high-performance sensor information processing infrastructure which supports noise correction, context recognition, and probability computation.
 研究代表者
 川島 英之（KAWASHIMA HIDEYUKI）
 筑波大学・大学院システム情報工学研究科・講師
 研究者番号：90407148

研究成果の概要（和文）：本研究では異常補正・状況認識・確率計算を支援する高性能なセンサ情報処理基盤に関する研究を行った。具体的には来歴管理，クリネ閉包処理，ベイジアンネットワークについて研究を行った。

研究成果の概要（英文）：This research tackled on a high-performance sensor information processing infrastructure which supports noise correction, context recognition, and probability computation. The work includes provenance management, Kleene closure processing, Bayesian networks.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2008年度	1,900,000	570,000	2,470,000
2009年度	1,400,000	420,000	1,820,000
年度			
年度			
年度			
総計	3,300,000	990,000	4,290,000

研究分野：データベース

科研費の分科・細目：情報学・メディア情報学・データベース

キーワード：ストリーム処理

1. 研究開始当初の背景

センサデバイスの発展に伴い，実世界で生じる現象をより深く広く理解すべく，MOTE/uPart 等の小型センサデバイスにより，屋内外のセンサデータを火山（ハーバード大），ボストン市内の水道管（MIT-CSAIL），スマートルーム（NTT-CS 基礎研，慶大 SFC-徳田研，筑波大-中内研），スポーツ選手運動時（慶大 SFC-仰木研）にて収集・解析するような取り組みが行われていた。

このような状況に対してセンサデータ基盤技術の研究は増加の一途である。海外の関連

研究には，信号処理に特化した WaveScope (MIT)，そして DBMS 内部でノイズ除去を行ったビューを提供する MauveDB (メリーランド大)，そして RFID データ処理専用である確率的 DBMS として PEEEX (ワシントン大) が挙げられた。国内の関連研究には，研究代表者自身による組込制御系向けセンサ DBMS である KRAFT，そして異種センサ情報統合基盤である StreamSpinner (筑波大-北川研) があった。

これらの研究の問題は，いずれもがセンサデータ処理の流れの一部に焦点を当てた研究であり，センサデータ処理全般 (図 1) を支

援しない点であった。そもそもセンサデータの使用目的は実世界のイベント発生検知にあり、センサデータが取得されてからイベント発生検知までの全処理過程は次のようにまとめられる。①観測データ取得、②異常補正(ノイズ修正・欠損値補間)、③状況認識(信号処理による単一イベント認識)、④確率計算(全イベントの発生確率計算)、⑤状況報告。既存研究の対象領域を述べると、WaveScope/KRAFT/StreamSpinnerは③のみ、MauveDBは②のみ、そしてPEEXは④のみとなっており、全般を支援する研究は存在しなかった。そのため、実世界の状況を連続的に求める、冒頭に述べたアプリケーション群を現在のデータベース技術は部分的にしか支援できていなかった。

2. 研究の目的

本研究では上記3機能(②~④)を全て有する、高性能なストリーム指向DBMSの設計原理確立を目指した。データモデルには、簡潔かつ強力な関係データモデルを基礎として、②~④を実現するために、異常補正演算子、状況認識演算子、そして確率計算演算子を新たに導入することを目的とした。これらの演算子は既存の演算子とは異なり、それぞれ特定のクラスの関数群を抽象化したものだった。

3. 研究の方法

まずは関係データモデルと確率的関係データモデルをもつDBMSカーネルを開発した。そして、同DBMSにおいて関係データに対して窓演算子及びセンサデータ処理用演算子を適用可能な問合せ処理系を開発した。

その後、センサデータ処理用演算子の効率的な実現方式について研究を行った。特に、確率的ストリームに対して、クリネ閉包を有する問合せ処理技法についての研究を行った。これはクリネ閉包がストリーム処理において極めて大きな役割を果たすことが知られていたからであった。また、確率推論技法であるベイジアンネットワークを関係データ処理の枠組みに導入する方法を研究した。

上記と並行して、非同期データ格納方式の研究に取り組んだ。KRAFT方式では基本的にダブルバッファリングに基づきディスクへの通常データ書き込みを高速化したのみだった。しかし本研究ではデータのみならず来歴データを対象にした効率的な書き込み技法を提案した。

さらにストリーム処理の応用としてロボットの動作自動生成に関する研究に取り組んだ。

4. 研究成果

(1) 来歴管理について

本研究は、データストリームにおける来歴管理技法を示した。ストリーム処理エンジンの出力を受け取ったアプリケーションからの根拠を問い合わせる要求に応えるため、ストリーム処理エンジンに到着したダブルストリームの内、その出力の全来歴を永続化する。まず、データストリームの来歴となりうるデータは、実行木の出力タプルの来歴テーブルであることを示す。ストリーム処理環境では、実行木の出力テーブル以外のデータはユーザに提供されない。ユーザに提供されないテーブルは、ユーザに感知されない。感知されないテーブルは意味を持たない。次に、頻繁に到着するデータストリームを一括してディスクに転送する方式を提案する。結果集合を成すテーブルの全来歴テーブルを連続領域にマージングし、それをディスクへ一括転送することで、処理時間の長いディスクアクセスを減らすことができる。そして、プロトタイプSPEを作成して、ウィンドウ演算と選択演算から構成される実行木について、ウィンドウ幅を変えて来歴テーブル永続化に関する処理時間を計測した。実験の結果、来歴テーブル永続化処理には多大な時間を要すること、およびその原因はディスクアクセス回数であることを示した。

(2) クリネ閉包処理について

本研究では関係データ演算では処理できない時系列データ処理を支援するために、クリネ閉包を効率的に処理する技法を研究開発した。提案手法であるActive Instance Graphは従来手法であるActive Instance Stackを、確率的データストリームを処理できるように拡張した。提案手法の優位性を実験的に検証した。

(3) ベイジアンネットワークについて

本研究は確率推論技術であるベイジアンネットワークを拡張し、関係データストリーム処理に適合させる方法を提案した。ベイジアンネットワークをデータストリーム処理に適合させるために3つの貢献を行った。第一の貢献は、ベイジアンネットワークにウィンドウの概念を導入することであった。ベイジアンネットワークのイベントに生起時間幅を持たせ、各イベントの生起状態を管理する状態変化管理機構を構築した。これに合わせて、状態変化を用いたウィンドウ処理を提案した。第二の貢献は、ベイジアンネットワークで行われる確率推論処理を関係演算系に編み込むために、ベイジアンネットワークを抽象データ型として関係表で表現すること

と、演算木に処理を組み込むために、入出力がタプル形式になるようにベイジアンネットワークをモデル化することだった。これに際して、入力データストリームをベイジアンネットワークのイベントに関連付ける ASSOC、そしてベイジアンネットワーク内の各ノードの確率値等の情報をタプル形式で選択するメソッド群を提案した。第三の貢献は、確率推論処理を効率化するために、オンラインで利用者の問合せに基づきベイジアンネットワークの確率伝播を部分的に省略する方法を提案したことだった。実際に提案手法により確率推論処理時間の削減が行われているかを検証した。

(4) CPU キャッシュを考慮した高速化について

CPU キャッシュとメモリのアクセスコストの乖離によりキャッシュミスが RDBMS の性能低下を起こすことが近年明らかになりつつある。命令キャッシュミスが生じる一原因に、RDBMS 内のオペレータ群の合計フットプリントが L1 命令キャッシュに収まらないことがある。これを改善すべく Zhou らはオペレータ実行順序を変更するバッファリングオペレータを提案した。同技法は非常に優れたキャッシュミス改善技法であるが、最適化器にバッファリングオペレータ使用を決定させるアルゴリズムが示されていないため、Zhou らの技法のみでは RDBMS の性能を現実的に向上させられない。そこで我々は従来研究において Zhou らの技法をも含むクエリ処理計画を選択可能な最適化器である、CC-Optimizer を実現した。だが、CC-Optimizer は 1 アーキテクチャという限定的な状況しか使用することが出来なかった。我々は本論文において、更に、ユーザがパラメタを適切に設定できるよう変更を加え、複数アーキテクチャ上で CC-Optimizer が有用であることを示そうと試みた。

(5) ロボット動作の自動生成について

本研究の目的は、異種ロボット間でのジェスチャ情報を共有化することだった。本研究では、ジェスチャの本質的な特徴に注目しジェスチャを 2 種類に分類した。それらは直示ジェスチャと軌道重視ジェスチャである。直示ジェスチャはある一点を指差すジェスチャであり、動作の最終的な姿勢に意味がある。軌道重視ジェスチャは、ジェスチャの軌道自体に意味があり、軌道を平行移動させたり縮小した場合でも本質的な特徴を保存できる。それぞれの分類毎にジェスチャの変換アルゴリズムを作成した。提案手法の有効性を実験を通して示した。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 4 件)

- ① 川島英之, 北川博之, 寺島裕貴「ストリーム処理エンジンにおける効率的な来歴管理」日本データベース学会論文誌, Vol. 8, No. 1, pp. 101-106. 2009 年 6 月. 査読有
- ② 佐藤亮, 川島英之, 北川博之, 「データストリーム処理へのベイジアンネットワークの導入」日本データベース学会論文誌, Vol. 8, No. 1, pp. 137-142. 2009 年 6 月. 査読有
- ③ Zhitao Shen, Hideyuki Kawashima, Hiroyuki Kitagawa, “Efficient Probabilistic Event Stream Processing with Lineage and Kleene-plus”, Special Issue of International Journal of Communication Networks and Distributed Systems (IJCNDS) On Data Management in Ad Hoc and Sensor Networks for the Information-explosion Era, Vol. 2, No.4 pp. 355 - 374, 2009. 査読有
- ④ 佐藤亮, 川島英之, 北川博之, “ベイジアンネットワークを用いた確率的データストリーム処理システムの提案”, 日本データベース学会論文誌, Vol. 7, No. 1, pp. 157-162, 2008 年 7 月. 査読有

[学会発表] (計 8 件)

- ① Yoshishige Tsuji, Hideyuki Kawashima and Ikuo Takeuchi, “Optimization of Query Processing with Cache Conscious Buffering Operator”, Proc. 6th International Workshop on Databases in Networked Information Systems (DNIS 2010), University of Aizu, Japan, March 29 - 31, 2010.
- ② Yuki Terajima, Hideyuki Kawashima and Hiroyuki Kitagawa, “Providing Persistence to Provenances on Stream Processing Environment”, Proc. 3rd International Workshop on SensorWebs, Databases and Mining in Networked Sensing Systems (SWDMNSS 2009), Carnegie Mellon University, Pittsburgh, USA, June 17 - 19, 2009.
- ③ Ryo Sato, Hideyuki Kawashima and Hiroyuki Kitagawa, “Associating Bayesian Networks with Stream Data Processing”, Proc. 3rd International Workshop on SensorWebs, Databases and Mining in Networked Sensing Systems (SWDMNSS 2009), Carnegie Mellon

University, Pittsburgh, USA, June 17 - 19, 2009.

- ④ Xin Li, Zhitao Shen, Hideyuki Kawashima and Hiroyuki Kitagawa, “Pattern-based Window: A Novel Window Operator to Support Event Detection for Data Stream Processing”, Proc. International Workshop on Sensor Network Technologies for Information Explosion Era (SeNTIE 2009), Taipei, Taiwan, May 18 - 21, 2009.
- ⑤ Kenshiro Hirose, Hideyuki Kawashima, Satoru Satake and Michita Imai, “Sharing Gesture Contents among Heterogeneous Robots”, Proc. International Workshop on Data Management for Information Explosion in Wireless Networks (DMIEW 2009), Fukuoka, Japan, March 16 - 19, 2009.
- ⑥ Hideyuki Kawashima, Ryo Sato, Hiroyuki Kitagawa “Models and Issues on Probabilistic Data Stream Systems with Bayesian Networks”, Proc. 2nd International Workshop on Sensor Webs, Databases and Mining in Networked Sensing Systems (SWDMNSS 2008), pp. 157-160. Turku, Finland, July 28 - August 1, 2008.
- ⑦ Shen Zhitao, Hideyuki Kawashima, Hiroyuki Kitagawa, “Lineage-based Probabilistic Event Stream Processing”, Proc. International Workshop on Sensor Network Technologies for Information Explosion Era (SeNTIE 2008), Beijing, China, April 27, 2008.
- ⑧ Ryo Sato, Hideyuki Kawashima, Hiroyuki Kitagawa, “The Integration of Data Streams with Probabilities and a Relational Database using Bayesian Networks”, Proc. International Workshop on Sensor Network Technologies for Information Explosion Era (SeNTIE 2008), Beijing, China, April 27, 2008.

[図書] (計1件)

- ① Hiroyuki Kitagawa, Yousuke Watanabe, Hideyuki Kawashima, and Toshiyuki Amagasa, “Stream-based Real World Information Integration Framework”, in “Wireless Sensor Network Technologies for Information Explosion Era” (Springer book series “Studies in Computational Intelligence”) 2010 (to appear).

[その他]

ホームページ等

<http://www.kde.cs.tsukuba.ac.jp/~kawashi ma>

6. 研究組織

(1) 研究代表者

川島 英之 (KAWASHIMA HIDEYUKI)

筑波大学・大学院システム情報工学研究
科・講師

研究者番号：90407148