

Stream OLAP のための可視化ユーザインタフェース

坂元沙季[†] 中挾晃介^{††} 北川博之^{†††} 天笠俊之^{†††}

[†]筑波大学情報学群情報科学類

^{††}筑波大学システム情報工学研究科コンピュータサイエンス専攻

^{†††}筑波大学システム情報系情報工学域

1. はじめに

近年、株式市場データやセンサーデータなど、連続的に生成されるストリームデータが数多く存在する。このストリームデータに対して、高度な分析を行うニーズが存在する。静的なデータに対する高度な分析手法に OLAP 処理がある。筑波大学北川データ工学研究室では、ストリーム処理エンジンを用いたストリームデータに対する効率的な OLAP 処理についての研究[1]を行っており、Stream OLAP システムのプロトタイプを開発している。Stream OLAP システムはストリームデータを受取り、集約処理を行う。本研究では、Stream OLAP システムによる集約結果の表示を目的とした可視化ユーザインタフェースとなる Stream OLAP ビューワシステムの研究開発について報告する。本ビューワシステムを用いることで、ロールアップやドリルダウンなどをユーザが対話的に行うことが出来る。

2. Stream OLAP

従来の OLAP 処理[2]はデータウェアハウス内に蓄積された大量のデータに対して多次元分析を行う手法である。分析したい対象の数値データであるメジャーや、次元表に関連付けるキーが格納された事実表と、階層構造を持つ属性が格納された次元表からキューブを構築する。キューブとは多次元的なデータを参照できるモデルである。キューブに対して、ロールアップやドリルダウン、スライスなどの操作を行うことにより多次元分析を実現する。以上が静的なデータに対する OLAP 処理である。

一方で、Stream OLAP では時間の属性を持つストリームデータを対象にする。時間に対する集約処理を可能にするために、集約対象の時間幅を設ける。これを IoI (Interval of Interest) という。新しいデータが集約対象となると、古いデータは集約対象から外れ破棄される。ここで図1の事実表 Sales と次元表 Customer, Supplier, Time を持つデータについて考える。事実表 Sales にはメジャーである sale や各次元表と関連付けるキーである customerID, supplierID, timeID が格納されており、ストリームとして新しいデータが順次到着する。次元表 Customer, Supplier には name, nation, region, 次元表 Time には minute, hour の次元階層がそれぞれ格納されている。例えば IoI を 60 分

間とする場合は、60分間のデータを保持し集約処理を行う。図1では時刻13:59から14:58までのデータが集約対象である。時刻14:59のデータが新しく集約対象となると、時刻13:59のデータは集約対象から外れ、破棄される。またこのデータにおける、次元階層の組合せ Customer の name, Supplier の name, Time の minute に対応する集約結果は図1左上のキューブである。また Customer の name から region, Time の minute から hour へロールアップした結果が右下のキューブに対応する。

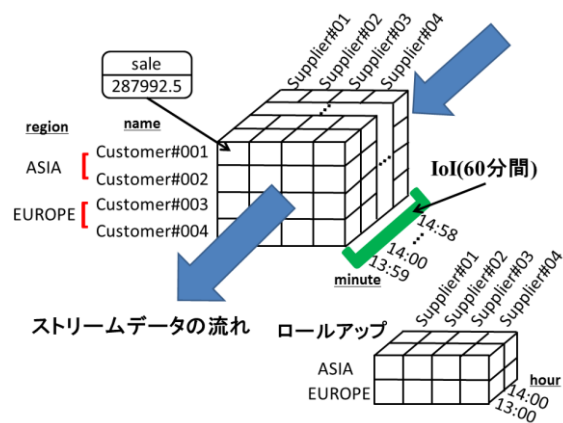


図1 キューブとロールアップ操作

3. Stream OLAP ビューワシステム

本研究で提案する Stream OLAP ビューワシステムは、Stream OLAP システムに対してユーザによる対話的な操作を実現する。本システムの設計すべき要件は、以下である。

- 1) ユーザによる集約対象の次元階層の組合せの選択
- 2) Stream OLAP システムへのリクエスト送信とレスポンス受信
- 3) 受信したレスポンスを結果として出力
- 4) ロールアップ、ドリルダウンなどの操作

ここで Stream OLAP システムに対するリクエストとは 1) でユーザが指定した次元階層の組合せのことであり、レスポンスは Stream OLAP システムにより生成されたグループキーとメジャーの値の組合せである。

図2に本システムの概要を示す。本システムを起動すると、Stream OLAP システムにおいて処理するデータのスキーマ情報を読み込み、表示する。図3では初期画面に対応する。ユーザが次元階層の組合せを選択後、選択された組合せを取得し、Stream OLAP システムへリクエス

A Visual User Interface for Stream OLAP

Saki Sakamoto[†], Kosuke Nakabasami^{††}, Hiroyuki Kitagawa^{†††} and Toshiyuki Amagasa^{†††}

[†]College of Information Science, University of Tsukuba

^{††}Graduate School of Systems and Information Engineering, University of Tsukuba

^{†††}Faculty of Engineering, Information and Systems, University of Tsukuba

トを送信する。レスポンスを受信後、集約結果をテーブル形式で出力する。図3のレスポンス結果出力画面にあたる。ロールアップとドリルダウン操作は1), 2), 3)を再度行うことで実現する。スライス、ダイス操作は出力されたテーブルからユーザが抽出したい属性とその値を指定し、条件に一致する組合せのみを表示することにより実現する。なお本システムではこの機能をフィルタリングとする。図3ではフィルタリング出力画面にあたる。

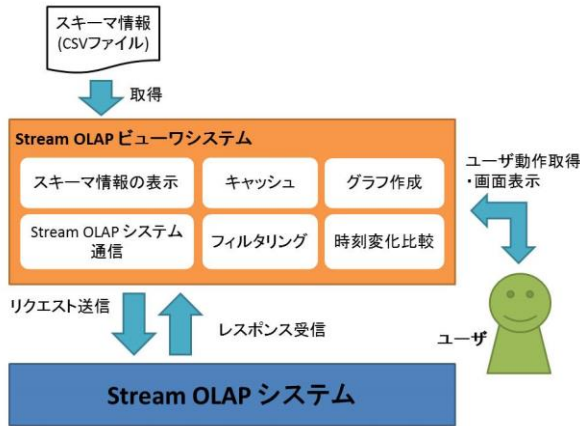


図2 Stream OLAP ビューワシステム

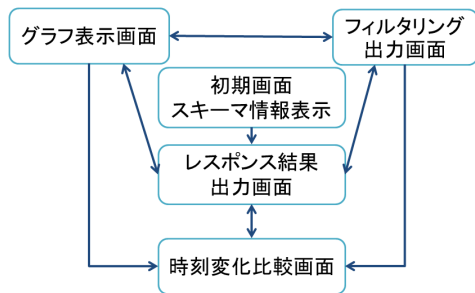


図3 状態遷移図

また Stream OLAP システムで扱うデータはストリームデータであるため、メジャーの値の時刻変化を比較する機能を実現する。これは出力されたテーブルから属性とその値を指定することにより、条件に一致する値のメジャーと時刻を軸にグラフを作成する。図3では時間変化比較画面にあたる。この画面では、指定された属性と一致するデータのみをテーブルとグラフで出力する。図4は時刻変化比較において出力されるグラフイメージ例である。左側のグラフは Customer の region, Supplier の region, Time の hour を集約対象の組合せとして選択した場合のレスポンスから作成したグラフである。ここから時間次元に対するドリルダウンを考える。Time の hour から minute へ次元階層を変更し、Customer の region, Supplier の region, Time の minute を組合せとして画面上部のプルダウンメニューより選択する。この組合せに対応するレスポンスから作成したグラフが右側にあたる。グラフは横軸を時間軸、縦軸をメジャーである sale としたグラフであり、左側グラフの各点は一時間の sale の平均を、右側グラフの各点は一分間の sale の平均にあたる。例えば左側グラフの 14:00 の点は 14:00 から 15:00 までの

sale の値の平均にあたる。

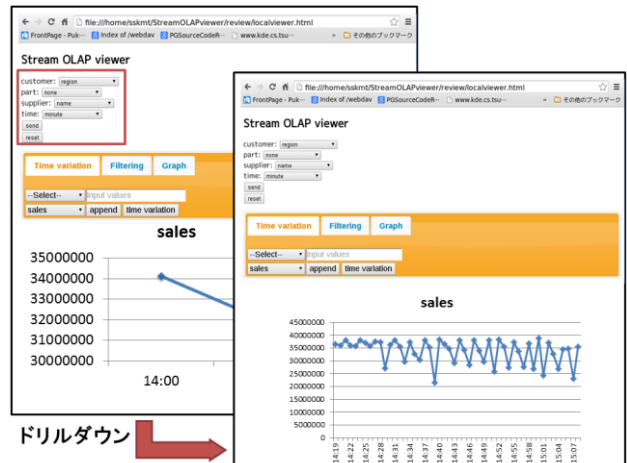


図4 時間次元に対するドリルダウンイメージ図

4. 実装

実装には HTML と JavaScript を用い、web ブラウザ上に結果を出力する。web ブラウザから HTML ファイルを開くと、スキーマ情報が格納された csv ファイルを読み込み、web ブラウザ上に次元ごとにプルダウンボタンを用いて表示する。

Stream OLAP システムとのやりとりには web サーバを介して行う。ユーザにより選択された次元の組合せを取得し、xmlHttpRequest により web サーバへリクエストを送信する。同様に、Stream OLAP システムにおいて処理されたレスポンスを web サーバから受信する。web サーバとのやりとりには REST を用いる。リクエストは URL 文字列、レスポンスは JSON である。

5. まとめ

本稿では、Stream OLAP システムに対する問合せ結果の表示を目的とした Stream OLAP ビューワシステムについて述べた。本システムにより、ユーザが Stream OLAP におけるロールアップやドリルダウンなどの対話的な操作が可能となる。

今後の課題としては、Stream OLAP システムより送信されるレスポンス情報を Stream OLAP ビューワシステム内に格納するキャッシュ機能の実装を進めるとともに、ユーザによるリクエスト送信からレスポンスが表示されるまでの応答時間の計測を行い、キャッシュを利用する場合の応答時間との比較実験を行う。

謝辞

本研究の一部は、文部科学省「実社会ビックデータ利活用のためのデータ統合・解析技術の研究開発」による。

参考文献

- [1] Kosuke Nakabasami., Hiroyuki Kitagawa., Salman Ahmed Shaikh., Toshiyuki Amagasa., Cost-based Optimization of Stream OLAP. (submitted)
- [2] Suragit Chaudhuri., Umeshwar Dayal., An Overview of Data Warehousing and OLAP Technology. ACM SIGMOD Record, Vol. 26, No. 1, pp. 65–74, March 1997.