

予測分析を可能とするストリームOLAP分析フレームワーク

中川 遼太郎[§] 塩川 浩昭[†] 北川 博之[†]

[§]筑波大学情報学群情報科学類 [†]筑波大学計算科学研究センター

1 序論

近年のIoT (Internet of Things) 技術によるセンサやソーシャルネットワークサービスの発展により、それらから無制限でかつ急速に発生するストリームデータの分析が重要視されている。様々なストリームデータ分析の方法はあるが、中でもリアルタイム分析は必要性が高い。

データの分析方法としてOLAP(Online Analytical Processing) [1]がある。OLAPは蓄積された多次元データに対して、ユーザと対話的な手段を通して素早く分析結果を返答する方法である。我々はこれまで、蓄積されたデータに対する方法であったOLAPをストリームデータに拡張したStreamOLAP[2]を開発してきた。StreamOLAPは通常のOLAPと同様にストリームデータに対してSUMやAVERAGEといった集約・集計分析を行うことができる。しかし、StreamOLAPが分析の対象とすることができるデータは過去から現在までに既に配信されたデータに対してのみに限定されている。一般的にストリームデータは時系列的な情報を含むことから、その時系列的な変化に基づく将来の変動傾向予測分析もひとつの重要な要素技術となる。ところが、既存のStreamOLAPはストリームデータの予測分析に対応していない。

そこで本研究では、文献[2]で提案されたStreamOLAPの枠組みを拡張し、予測分析を可能とするストリームOLAPフレームワークを提案する。特に本研究では、自己回帰モデル[3]に基づく時系列データの予測手法を題材とし、StreamOLAPシステムへの統合方式を提案する。本方式を用いることにより、ストリームデータに対して将来予測や傾向の変動予測といった時系列による傾向などの特徴分析を可能とする。本稿では、提案するフレームワークの概要と気象庁が提供する気温データストリームを利用した評価実験結果について報告する。

2 提案フレームワーク

2.1 フレームワークの概要

StreamOLAP[2]に予測分析機能を組合せたアーキテクチャを提案する。図1に本研究で提案するフレームワークの概要を示す。提案するフレームワークはStreamOLAPと同様に、大きく分けてストリーム処理エンジン(SPE)とOLAPエンジンから構成されている。SPEでは時々刻々と配信されるストリームデータを受け取り、タプル毎にユーザが用意した次元表との結合を行う。OLAPエンジンではSPEから受け取ったタプルを基に分析処理を行う。本研究ではStreamOLAPとは異なり、予測分析機能をOLAPエンジンの内部に導入する。本稿で導入する予測分析機能の詳細については2.2節で述べる。

2.2 予測分析モデルの導入

本研究では自己回帰モデルによる予測分析機能をStreamOLAPに導入する。自己回帰モデルとは、時系列の沿ったデータ x の時間 t における値 x_t を目的変数とし、過去 p 個の連続した値 x_{t-1}, \dots, x_{t-p} を説明変数とした回帰分析である。

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + C + \varepsilon \quad (1)$$

本稿で提案するフレームワークでは自己回帰モデルによる予測分析を導入するために、OLAPエンジン内部にデータ保持領域を構築する。具体的には、新たに(1)予測データ保持領域、(2)モデル保持領域、(3)誤差保持領域の3種類のデータ保持領域をOLAPエンジンに実装する。各保持領域では保存するデータの詳細は次のとおりである。

予測データ保持領域 自己回帰モデルによる予測分析結果 \hat{x}_t を保持する領域。

モデル保持領域 予測分析に必要な自己回帰モデルのパラメータ $\phi_{t-1}, \dots, \phi_{t-p}, C$ を保持する領域。

誤差保持領域 予測データ保持領域のデータとストリームデータの誤差 ε をモデルごとに保持する領域。

OLAPエンジンでは上述したデータ保持領域を用いることで式(1)に基づく自己回帰分析を実行する。本フレームワークに基づく予測分析処理の詳細については2.3節で述べる。

Stream OLAP Framework for Predictive Analytics
Ryotaro NAKAGAWA[§], Hiroaki SHIOKAWA[†] and Hiroyuki KITAGAWA[†]
[§]College of Information Science, University of Tsukuba
[†]Center for Computational Sciences, University of Tsukuba

2.3 予測分析の実行

本フレームワークを用いた予測分析処理の流れについて述べる。本フレームワークは前提として、事前にオフライン学習して求めた自己回帰モデルのパラメータをモデル保持領域に格納しておく必要がある。このモデルに基づいて、提案フレームワークはストリームデータが到着する毎に次の手続きに従った予測分析を実行する。

予測分析の流れ (1) ストリーム処理エンジン内でユーザが用意した次元表と結合を行い、OLAP エンジンへ送る。(2) OLAP エンジン内の一番情報の粒度が細かい、データとして保持されるされる頂点 (以下「materialized node」と称する) の通常データ領域へ挿入される。この時、他の materialized node においては集約演算を行ってから挿入する。(3) 通常データ領域内から p 個のテーブルとモデル保持領域から次数 p の自己回帰モデルを使って IoP (Interval of Prediction) の分予測分析を行い、分析結果を予測データ保持領域に挿入する。(4) 予測データと同じ時間におけるテーブル (測定値) が到着した場合誤差計算を行う。その結果を誤差の記憶保持領域に挿入する。(5) ユーザから materialized node における分析結果の問合せがあった場合、その頂点における通常データ保持領域のデータと予測データ保持領域のデータをユーザに送る。(6) Materialized node 以外における分析結果の問合せがあった場合、一番近い materialized node の通常データ保持領域と予測データ保持領域の集約値をユーザに送る。

モデルの誤差計算 事前に学習した自己回帰モデルによる予測分析は時間が経つにつれて誤差が大きくなるのが考えられる。そのため、モデルを最新のデータで再学習する必要がある。本研究の提案手法にはモデルの再学習の機能はないがその前段階として、測定値と予測値の誤差を計算しその誤差結果をシステム内で保持する。これにより本フレームワークの利用者は、誤差結果を使ってモデルの再学習のタイミングを決めることが可能である。

3 評価実験

本実験では提案手法の実行した際の実行速度を評価する。本実験では評価のために、ストリームデータ1タプルの到着毎に予測計算および誤差計算にかかる計算時間を測定する。本実験で用いるデータセットは、気象庁が提供する日本全国の観測地点で計測された気温データストリームである。本実験で用いるストリームデータは1秒間に100タプル到着するものとする。

予測計算はIoPとモデルの次数を増やすと計算時間の増加見られたが、誤差計算はモデルの次数を大きくしても計算時間に大きな変化はなかった。誤差計算に

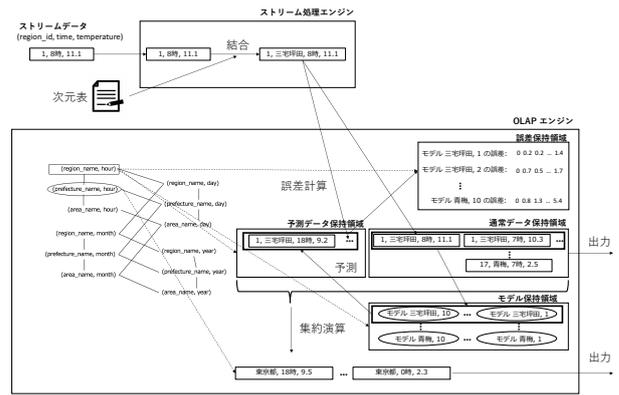


図 1: 提案フレームワークの概要

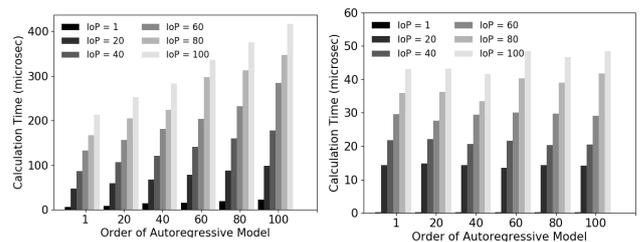


図 2: 予測計算の計算時間

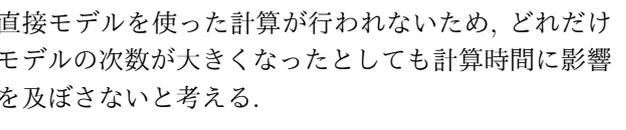


図 3: 誤差計算の計算時間

直接モデルを使った計算が行われなかったため、どれだけモデルの次数が大きくなったとしても計算時間に影響を及ぼさないと考える。

4 まとめ

本稿では、StreamOLAP の枠組みに自己回帰モデルを利用した予測分析が行える新たなシステムフレームワークを提案した。今後は予測計算にかかる時間の短縮とシステム内で計算された誤差の結果やモデルの再学習の頻度を考慮したモデルの再学習が行える機能を枠組み内に追加し高い精度の予測分析がいつでも行えるようにシステムを改善していく予定である。

謝辞

本研究の一部は、NICT 高度通信・放送研究開発委託研究「欧州との連携による公共ビッグデータの利活用基盤に関する研究開発」、文科省/理研「実社会ビッグ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] S. Chaudhuri et al., An Overview of Data Warehousing and OLAP Technology, ACM SIGMOD Record, 26 (1), pages 65-74, March 1997.
- [2] Kosuke et al., An Architecture for Stream OLAP Exploiting SPE and OLAP Engine. In Proc. IEEE Big Data 2015, pp. 319-326, 2015
- [3] Terence C. Mills, "Time Series Techniques for Economists", Cambridge University Press, 1990.