

Research session 4-4

On Historical Diagnosis of Sensor Streams

著者: Charu C. Aggarwal and Philip S. Yu.

担当: 高橋 翼 (NEC, CMU)

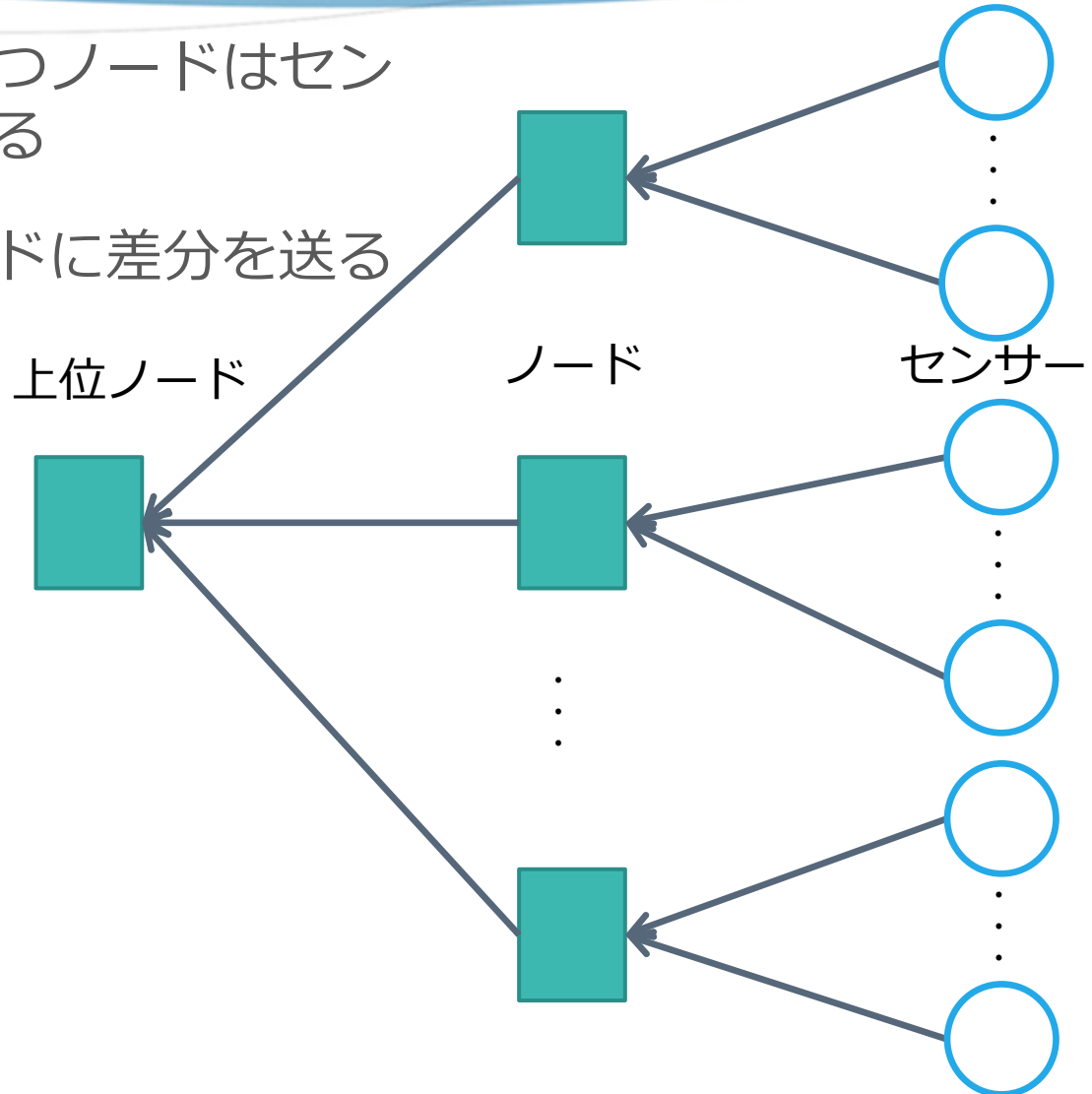
概要

- 一言で言えば
 - センサーストリームの時系列解析のためのスケッチ管理方法の提案
- 既存研究のフォーカス：aggregate frequency histograms
- 本研究のフォーカス：**履歴に対するtemporal behaviorの追跡**
 - すべての履歴を効率的にストアすることが必要
 - 解決策：**pyramidal sketches** (ピラミッド型スケッチ)
 - “This is the first paper which **provides error guarantees** for important queries which ... (略)”
- どのくらいすごいか
 - 100年分**(100x365x24x3600秒分)のストリームのスケッチを **31554** スナップショットに圧縮しつつ、**99.99%のaccuracy**で範囲検索 (horizon estimation) できる

前提

説明省略

- ◆ センサーを子に持つノードはセンサーの値を格納する
- ◆ ノードは上位ノードに差分を送る



Pyramidal Sketches

- ◆ センサーシグナルの量子化
- ◆ 通信の間引き (省略)
- ◆ Sketch based representation (省略)
- ◆ Pyramidal Storage Pattern
- ◆ スケッチを使ったQuery Estimation (省略、スライドもなし)

センサーシグナルの量子化

- まず各センサーで取得した値 $v(t)$ を中心化

- $z(t) = (v(t) - \mu) / \sigma$

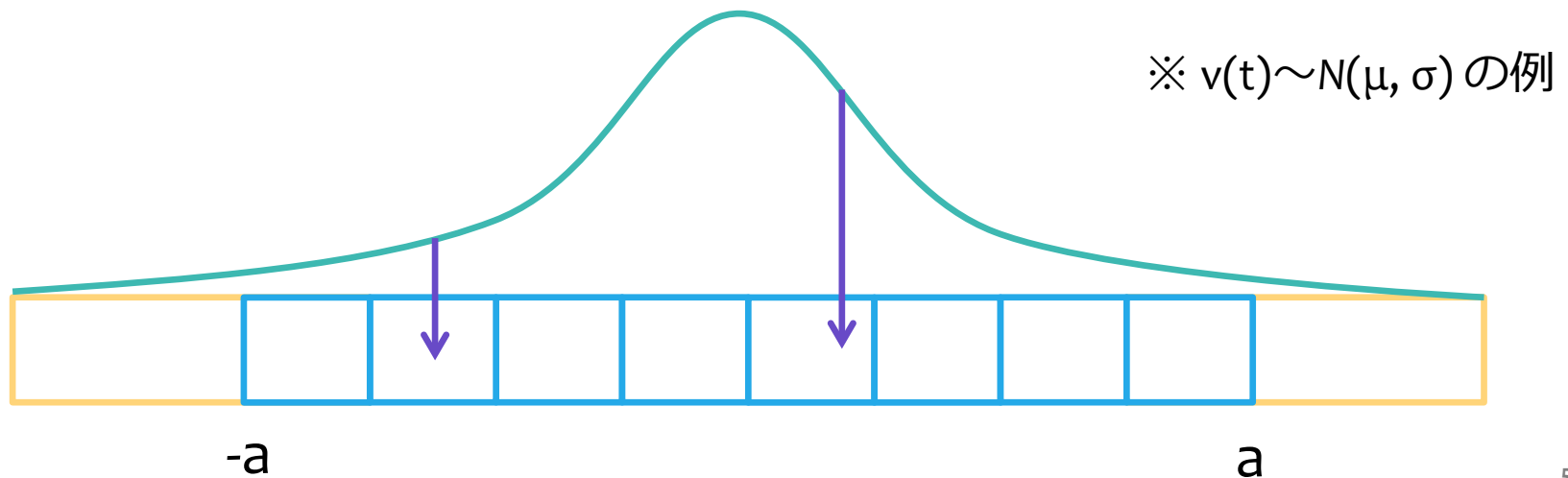
- Φ 個のバケットに区切る

- $[-a, a]$ に入るもの $\rightarrow \Phi-2$ 個に細く区切る

- a は3くらいがいいらしい ($\pm 3\sigma$ までをある程度正確に扱う?)

- $[-a, a]$ に入らないもの $\rightarrow < -a$ と $a <$ にざっくり分ける

- チビチェフの不等式より $P(|z(t)| > k) < 1/k^2$



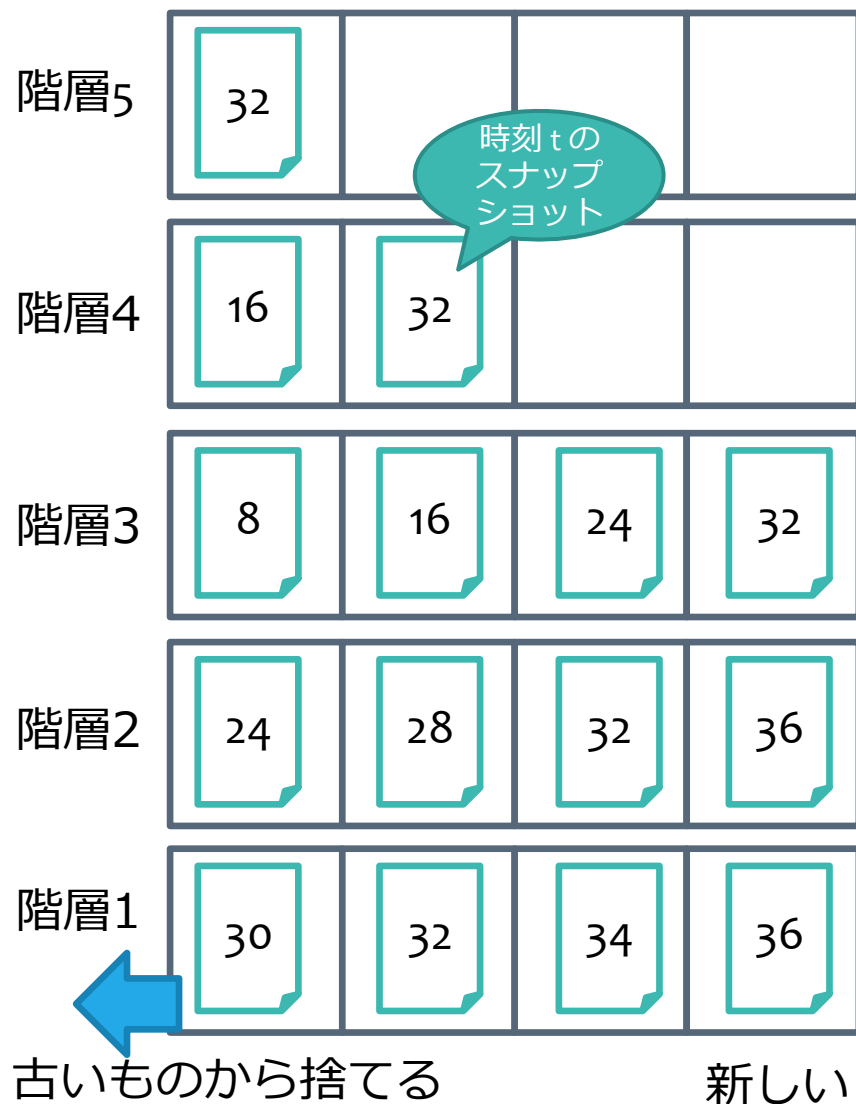
- ◆ センサーからノード
 - ◆ 送信するもの： <tickid, bucketid>
 - ◆ 送信タイミング：最後のtickidからbucketidが変化するときだけ
 - ◆ 筆者ら曰く “*In stable applications, sensor measurements change only gradually over time.*”
- ◆ ノードから上位ノード
 - ◆ 送信するもの： <tickid, sensorid, bucketid>
 - ◆ 送信タイミング：各センサーのbucketidが更新されたとき

Sketch based representation

説明省略

- ◆ センサーから取得したデータのスケッチをノードに格納
- ◆ **AMS Sketch** [STOC'96] を利用
 - ◆ 0次のスケッチ：値の種類
 - ◆ 1次のスケッチ：シーケンスの長さ
 - ◆ 2次のスケッチ：self-joinのサイズ(の予測に使える値)
 - ◆ . . .
 - ◆ ∞ 次のスケッチ：最頻の値
- ◆ 子孫のセンサー数 x x バケット数 ϕ x スケッチの次数 n のスケッチを各ノードが格納する

Pyramidal Storage Pattern



- “temporal behavior”が対象
 - スケッチ間の差分計算が必要
 - スケッチのスナップショットの膨大な蓄積が必要
- Pyramidal Storage Pattern** [著者ら, VLDB'03]を利用
 - 格納間隔: α^i i : 階層レベル
 - 各階層の最大格納数: α^i
 - 時刻 T で高々 $\alpha \log_{\alpha}(T)$ 個
 - 現在時刻 t_c から時間 h 遡ったスナップショットの探索
 - $|h-h'|/h < 1/(2 \cdot \alpha^{l-1})$ を満たす h' の探索で実現

評価

データセット3種

Intel-L, Intel-H, Chlorine

データサイズ x スケッチのサイズ

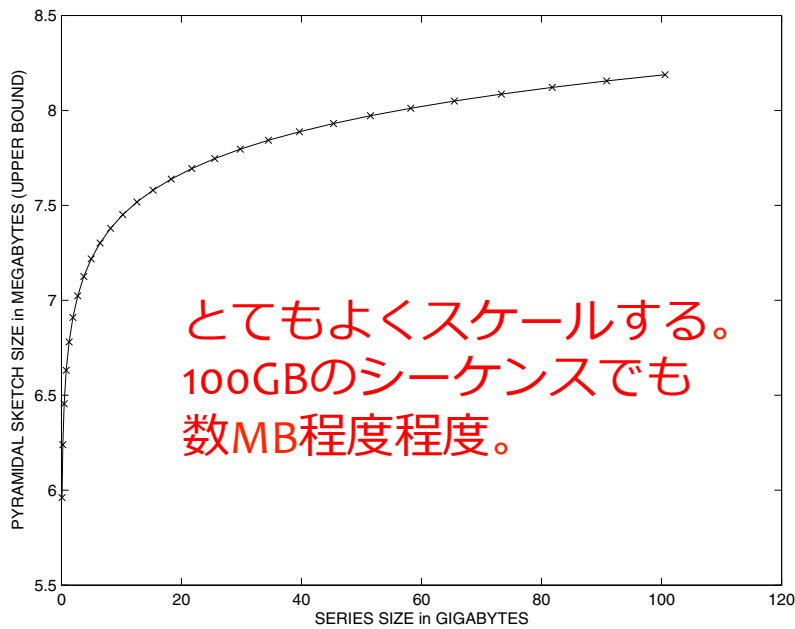


Fig. 1. Storage Requirement of Pyramidal Sketch with increasing series size (Upper Bound)

スケッチのサイズ x 誤差(horizon estimation)

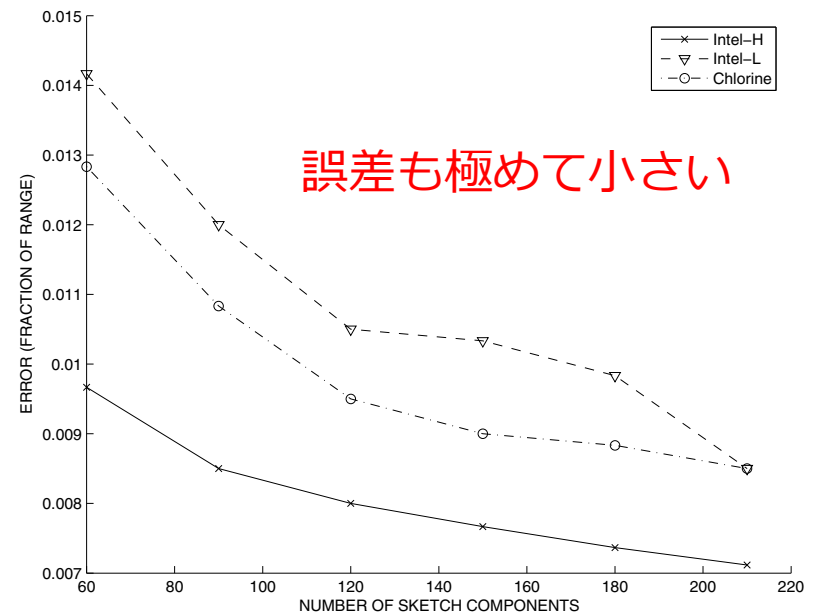


Fig. 2. Average estimation over a historical horizon (Varying number of Sketch Components)

図は論文からの引用です。