

## データストリームに対する連続的問合せの動的変更機能の提案

渡辺 陽介<sup>†</sup> 横田 治夫<sup>‡</sup>東京工業大学 学術国際情報センター<sup>†</sup>東京工業大学 大学院情報理工学研究科<sup>‡</sup>

## 1. はじめに

センサーデータやカメラなどのストリーム型の情報源が増加・多様化し、それらに対する連続的なデータ処理を実現するストリーム処理システムに注目が集まっている。ストリーム処理では、将来にわたって生み出され続けるデータへの処理要求を事前に登録するという特徴があり、1つの要求が実行される期間が非常に長い。そのため、実行中に実世界で起こる様々な環境変化にさらされる。データ分布やデータ発生頻度の変化、計算機リソース（CPU 負荷や遅延、計算機台数）の増減についてはこれまでの研究でも考慮されてきた。

しかし、処理要求レベルでの変化、例えば、処理対象となるストリーム情報源そのものを時間や位置・状況に応じて変更することや、実行中にストリーム情報源から届くデータの構造（スキーマ）が変更されることは想定されていない。また、実行中に利用者がフィルタリングの閾値やイベント検出のパターンを時間や位置・状況に合わせて差し替えることも考えられていない。既存のストリーム処理における処理要求定義方式 [1] では、そのような事態には実行を停止し、処理要求を再定義して新たに処理を開始する以外にはない。多様なストリームデータを統合利用するアプリケーションを長期安定稼働させていくためには、こうした変化さえも許容し取り込めるような柔軟さのある処理要求定義の仕組みが必要となる。

本研究では、情報源に関するメタデータを RDF グラフとして扱い、ストリームデータとメタデータの両方を用いたより柔軟性の高い要求定義方式を提案する。

## 2. 利用例

従来の処理要求定義方式では実現の難しい例として、「追跡対象者の移動に合わせて、距離 10 未満にあるカメラからの映像データを収集し続けてほしい」という要求を用いる（図 1）。

ここでの想定として、まず追跡対象者の位置情報は GPS からのストリームとして取得できるものとする。また、領域内には複数のカメラが設置され、データをネットワーク経由で取得可能とする。各カメラの設置座標、スキーマ（属性名、データ型）、管理サーバ名などのメタデータも与えられているものとする。

既存の処理要求定義方式の枠内で上記要求を実現するためには、追跡対象者の位置変化に応じて近傍のカメラが変わる度に処理要求定義を再登録することを繰り返さなければならない。または代替案として、領域内の全カメラからの映像を常時収集することで必要な情報を取り逃さないようにできるが、計算機資源やネットワーク帯域のコストの観点から現実的ではない。

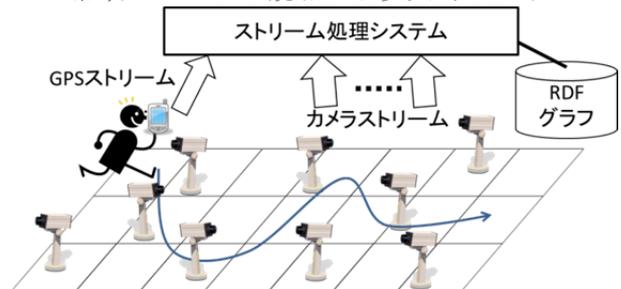


図 1 利用例

## 3. 求められる要件

上記の要求を扱うには以下の処理が必要である。

- 位置情報の取得：GPS からのストリームを受け取り、追跡対象者の現在位置を取り込む。
- 近傍のカメラの特定：カメラの設置座標を格納したメタデータから、追跡対象者の現在位置の近傍のカメラ  $I_{\text{nearest}}$  を検索する。
- カメラへの接続の確立：メタデータに基づいて、カメラ  $I_{\text{nearest}}$  のデータを提供する管理サーバ  $S_{\text{nearest}}$  への接続を行う。
- カメラからのデータ取得：メタデータに基づいてスキーマ情報  $I_{\text{nearest}}(A_1, \dots, A_n)$  から映像データを格納したカラム  $A_{\text{video}}$  を特定し、映像データを取り込む。
- カメラへの接続の終了：追跡対象者の最寄りのカメラ  $I_{\text{nearest}}$  が  $I'_{\text{nearest}}$  へ変化した場合、管理サーバ  $S_{\text{nearest}}$  への接続を終了する。

本研究では、メタデータの表現形式として RDF グラフデータを想定し、リレーション形式のストリームデータと RDF グラフのメタデータを統

A Proposal of Dynamic Modification Scheme for Continuous Queries over Data Streams

<sup>†</sup>Yousuke Watanabe, Tokyo Institute of Technology

<sup>‡</sup>Haruo Yokota, Tokyo Institute of Technology

合利用する扱う仕組みを実現する。処理要求定義方式およびシステムに求められる要件を一般化すると以下のようになる。

- メタデータの検索：例では近傍のカメラを探しているが、一般的な用途向けに RDF グラフに対する SPARQL [2] ライクな検索をサポートする必要がある。
- ストリームデータとメタデータの変換：ストリームデータ（例では位置情報）を用いて、メタデータの検索やメタデータの更新を行うことが求められる。
- メタデータを用いた情報源への動的接続：情報源に関するメタデータを用いて、情報源への接続を動的に確立・切断する。
- メタデータを用いた処理要求定義の記述：接続先の情報源が動的に変化するため、処理要求定義内で、固定の情報源名やカラム名を直接用いることができない。変数などを用いて処理要求定義を記述し、実行時に実際の情報源名やカラム名に置き換えられるようにしなければならない。

#### 4. 処理要求定義方式

前述の要件を満たすように、提案方式ではストリームデータに対しては SQL ベース、RDF グラフに対しては SPARQL ベースの問合せ言語を用いる。利用例における記述例を図 2 に示す。4 行目の SELECT-FROM 節は SQL ベースの構文で、カメラからの映像ストリームに対する処理を記述しているが、メタデータ変数として「?camera」と「?video」が用いられている。この変数の実際の値は、次の METADATA 節の中を繰り返し評価することによって変化する。METADATA 節は、SPARQL をベースとした問合せとなっており、検索条件を満たすカメラを探している。ここでは「ns:informationSource」は、ストリーム処理システムにおける情報源を表すクラスに対応し、「ns:locationX (Y)」は座標情報に対応する。VALUES 節は SPARQL の構文からの拡張で、ストリームに対する処理結果を変数に関連付けることができるようにしている。ここでは GPS からの座標が変数「?tx」, 「?ty」に関連付けられる。この記述例の評価は図 3 のように行われる。GPS からのストリームデータが届くたびに、変数?tx, ?ty の値が更新され、メタデータの検索が発生する。検索結果である?camera, ?video の値が前回の評価から変化した場合、接続する情報源が変化したことを意味する。このときシステムは対応するカメラへの接続を確立し、映像ストリームを受信する。

```

PREFIX rdf: <http://www.w3.org/... (略)...>
PREFIX ns: <http://... (略)...>

SELECT ?video
FROM ?camera [1 sec]
METADATA {
  SELECT ?camera , ?video
  WHERE {
    ?camera rdf:type ns:informationSource .
    ?camera ns:locationX ?cx .
    ?camera ns:locationY ?cy .
    FILTER DISTANCE(?cx, ?cy, ?tx, ?ty) < 10) .
    ?camera ns:hasAttribute ?video .
    ?video rdf:type ns:VideoData .
  }
}
VALUES (?tx, ?ty){
  SELECT Gps.x, Gps.y
  FROM Gps [1 sec]
}
    
```

図 2 記述例

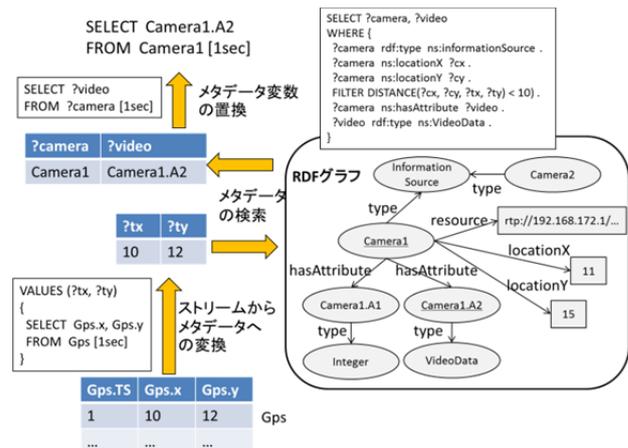


図 3 評価手順

#### 5. まとめ

本稿では、RDF グラフとストリームデータを連携させた要求定義方式の動的変更機能を提案した。今後は評価方式の効率化について検討していく。

#### 謝辞

本研究の一部は日本学術振興会科学研究費補助金基盤研究 A (#22240005, #25240014), 若手研究 B (#24700087) の助成により行われた。

#### 参考文献

[1] A. Arasu, S. Babu and J. Widom, "The CQL continuous query language: semantic foundations and query execution," *VLDB Journal*, vol. 15, no. 2, pp. 121-142, 2006.

[2] W3C, "SPARQL 1.1 Query Language," 21 3 2013. [Online]. Available: <http://www.w3.org/TR/sparql11-query/>.