

複数ドメインのデータストリームにおける意味的な複合イベント検出について Semantic Complex Event Processing over Data Streams from Multiple Domains

佐々木 勇和[†] 石川 佳治[‡] 杉浦 健人[‡]
Yuya Sasaki[†] Yoshiharu Ishikawa[‡] Kento Sugiura[‡]

1. まえがき

現代はビッグデータ時代と呼ばれるように、時々刻々と大量のデータが生成されている。時々刻々と生成される時系列的なデータは、データストリームと呼ばれ、センサ機器やスマートフォン、ウェブなど、様々なデータ源から多種多様なものが生成されている。データストリームは、逐次的にシステムに送信され、リアルタイムなデータ解析などに用いられる。例えば、スマートフォンからはユーザの位置情報や加速度データ、センサからは気温や湿度などが生成され、ユーザの行動分析や環境モニタリングなどが行われている。データストリームは、時系列的に生成されるイベントとみなすことができ、イベントを組み合わせることで、より複雑なイベントの検出を行う。他のイベントと結合していない独立なイベントは単純なイベントと呼び、単純なイベントの組合せで定義されるイベントは複合イベントと呼ぶ。この複合イベントを検出する複合イベント処理 (Complex Event Processing, CEP) が盛んに研究されている [2, 5]。データストリームを対象とする複合イベント処理では、リアルタイムなデータ処理が要求されており、効率的な計算が必要である。従来の複合イベント処理では、基本的に数値などの単純なデータ値のみを考慮していたが、近年では意味的な複合イベント処理 (Semantic CEP, SCEP) への関心が高まっている。意味的な複合イベント処理は、イベントがより多様で、セマンティック技術を用いるのが特徴である。オントロジを用いて、知識の階層構造の定義や、イベント間の関係性をより理解すること、宣言的なイベント処理などができる。意味的な複合イベント処理により、従来の複合イベント処理では検出できなかったイベントを検出することが可能である。

本研究では、データストリームから意味的な複合イベントを検出 (および、蓄積、解析) 可能なイベントベースの開発を目標としている [3, 4]。このシステムでは、様々なドメインのデータストリームが生成/送信されることを想定している。例えば、ドメインとして、ユーザの位置や行動を含むスマートフォンからのデータ、地域の温度や湿度を含むセンサからのデータ、天気情報などのウェブからのデータなどを想定し、オントロジにて登録されているものとする。また、場所の表現方法なども、統一性をもたせる必要はないが、地理オントロジなどの背景知識によって対応できるものを扱う。ユーザ (および、システム) が検出したいイベントは、オントロジに記述されており、イベントの結合や比較を行うことで、複合イベントを検出できる。複合イベントの検出パターンは、同時発生や、連続発生、地理的な近さなどの組み合わせを用いるが、オントロジに記述可能なもの、および問合せ言語で処理可能なものとする。多くの種類のドメインからのデータストリームがイベントベースに送信されると予想される。処理すべきイベントの数が多く、複合イベントの検出に大きな時間がかかると、リアルタイム性に影響がでてしまう。検出可能な複合イベントを減少させずに、効率的な処理が必要である。単純なイベントを減少させることにより、処理すべきイベント数が減少するため、複合イベント検出の計算量が減少する。オントロジの解析により、不必要なイベントのドメインを判断し、イベントのフィルタリングを行う。オントロジの解析は複合イベントに必要な単純なイベント

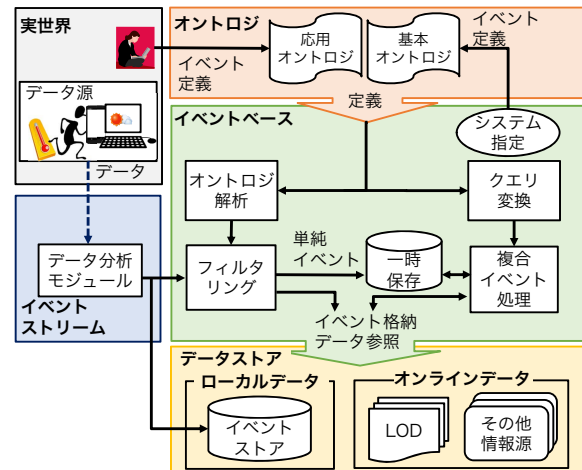


図 1: システム図

の導出、およびデータのフィルタリングはイベント内のドメインの値を確認するだけでよいため、単純な方法で処理が可能である。

本稿では、2. にてイベントベースの概要、3. にて具体的なオントロジの解析方法とフィルタリング方法について説明する。4. にて関連研究を紹介し、5. にて本稿をまとめる。

2. イベントベースの概要

イベントベースを図 1 を用いて、説明する。まず、実世界のデータ源から時系列的にデータがイベントベースに送信される。このとき、データ分析モジュールを介すことにより、データを意味的な情報に変換することができる。例えば、加速度データを推測される行動に変換することや、緯度経度を滞在している場所などに変換する。実世界のユーザは検出したいイベントを応用オントロジに定義し、システムは多くの状況で必要となる基本的な定義を基本オントロジに記述する。イベントベースでは、オントロジからのイベントの定義を受けて、オントロジ解析とクエリ変換を行う。オントロジ解析では、複合イベントの定義を分解することにより、複合イベント処理に必要なドメインを求める。クエリ変換では、オントロジの複合イベントの定義をクエリに変換し、そのクエリを用いて複合イベント処理が複合イベントの検出を行う。フィルタリングでは、オントロジ解析の結果から必要なドメインを把握できるので、そのイベントのみを一時的な保存領域に挿入する。また、このとき、データストリーム上の全てのイベントは、イベントストアに蓄積され、イベントの解析などに用いられる。このイベントストアと一次保存領域はリレーショナルデータベースを想定している。一時的な保存領域は、複合イベントをリアルタイムに処理するのに必要なデータのみが保存されている。イベントが時間経過で不必要になった場合、そのイベントは削除される。この削除の間隔も、オントロジ解析によって求めることができる。リンクドオープンデータ (LOD) やその他の外部情報源を用いて、複合イベント処理は複合イベントを検出する。検出した複合イベントはイベントストアと一時的な保存領域に格納され、解析はもちろんだが、より高次の複合イベントの検出にも用いられる。

[†]名古屋大学未来社会創造機構

Institute of Innovation for Future Society, Nagoya University

[‡]名古屋大学大学院情報科学研究科

Grad. Sch. of Information Science, Nagoya University

3. オントロジ解析とフィルタリング

オントロジに記述してある複合イベントの定義を解析することにより、不必要なイベントを把握することができる。複合イベントは、基本的にデータの同時発生やデータ間の関連性によって定義され、イベントの下位上位関係なども定義可能である。例えば、外に滞在しており、近くの気温センサの値が高いユーザを検出する場合の複合イベントの定義を以下とする。

- StayHotArea ←
OutsideStay \sqcap Close(User.Loc, HighTempSensor)
- OutsideStay ←
 \forall -Move(User.Loc) \sqcap hasLocation.Outside
- HighTempSensor ← Sensor \sqcap hasTemperature.High

StayHotArea は、検出したい複合イベントであり、外に滞在 OutsideStay かつ、ユーザの位置 User.Loc と高い気温値をもつセンサ HighTempSensor が近い場合に検出される。さらに、OutsideStay はユーザの位置がしばらく動いていない \forall -Move(User.Loc) かつ、場所が外 hasLocation.Outside という複合イベントであり、HighTempSensor はセンサーが高い気温値をもっている hasTemperature.High という複合イベントである。また、StayHotArea は、OutsideStay の上位のイベントとして定義されている。この複合イベント StayHotArea の検出には、ユーザの情報（例えば、スマートフォンからのデータ）と気温センサの情報が必要となる。つまり、この2種類のドメイン以外のイベントは必要ないということになる。他にも、ウェブからの情報などがあっても、それらのドメインのイベントはフィルタリングすることができる。複合イベントは、単純なイベントの組合せなので、単純なイベントのドメインを検出することにより、必要なドメインを検出できる。これは、複合イベントの再帰的な分解で可能である。

では、具体的な処理について説明していく。まず、データストリームのイベント形式は、RDF 形式など様々な様式が考えられるが、以下のような形式を考える。

```
Event = {
  Domain = Smartphone,
  User = Alice,
  Time = 2015-08-01T13:30:00,
  LocationLng = 35.153784,
  LocationLat = 136.967597,
  Symbolic Location = Outside,
  Activity = Walking
}

Event = {
  Domain = Sensor,
  Time = 2015-08-01T13:30:00,
  LocationLng = 35.153784,
  LocationLat = 136.967597,
  Temperature = 37.2,
  Temperature Period = degree Celsius,
  Humidity = 40
}
```

このデータは、スマートフォンとセンサによって生成されたイベントを表す。スマートフォンによって生成されたデータは、ユーザ、時間、緯度、経度、意味的な位置、行動が含まれ、センサによって生成されたデータは、時間、緯度、経度、気温、気温の単位、湿度が含まれている。各要素は、ドメインによって異なるが、システムは把握できているものとする。このデータをフィルタリングする場合は、ドメインを確認するだけでよ

く、必要なものであれば一次保存領域に挿入するという処理で単純にフィルタリングできる。

OutsideStay の \forall -Move(User.Loc) という項は、ユーザがある一定時間 (T 秒) 動いていないことを表す。この T の値もオントロジによって定義される。一つ前のイベントの位置との比較を再帰的に行うことにより、 T 秒間動いていないという複合イベントを、ユーザの位置という単純なイベントから検出することができる。このとき、 T 秒以上前のイベントは OutsideStay を検出するには必要ない。そこで、一定時間経過後に一次保存領域から削除することにより、不必要なデータを定期的に削除することができる。

一次的な保存領域へのイベント数を、フィルタリングと定期的な削除によって、削減することにより、効率的な複合イベント処理を可能とする。

4. 関連研究

文献 [2] にて、複合イベント処理に関する大規模なサーベイがされている。文献 [1] では、データストリームにおける意味的な複合イベント処理システム ETALIS を提案している。ETALIS では、背景知識を用いて、意味的な複合イベントを検出しているが、オントロジなどを用いた柔軟な複合イベントの検出は提供していない。筆者らの知る限り、データストリームにおいて、包括的かつ柔軟に複合イベント検出できる意味的な複合イベント処理システムはまだ提案されていない。

5. むすび

本稿では、データストリームのドメインが複数ある場合において、オントロジの解析により、不必要なドメインのデータストリームをフィルタリングすることを可能とする。また、時間経過で複合イベント検出に不必要となったデータを削除することも可能となる。これにより、複合イベント検出の際に処理されるイベント数が減少するため、効率的に複合イベントを検出できる。このフィルタリング技術を現在開発中のイベントベースに組み込み、リアルタイムなデータストリーム処理を実現する。今後の課題は、一連の流れをシステム実装することである。

[謝辞]

本研究は独立行政法人科学技術振興機構 (JST) の研究成果展開事業「センター・オブ・イノベーション (COI) プログラム」、および日本学術振興会科学研究費 挑戦的萌芽研究 (26540043) の支援によって行われた。ここに記して謝意を表す。

参考文献

- [1] Darko Anicic, Sebastian Rudolph, Paul Fodor, and Nenad Stojanovic, "Stream reasoning and complex event processing in ETALIS", *Semantic Web*, Vol. 3, No. 4, pp. 397-407, 2012.
- [2] Cugola, Gianpaolo, and Margara, Alessandro, "Processing flows of information: from data stream to complex event processing", *ACM Computing Surveys*, Vol. 44, No. 3, pp. 15:1-15:62, 2012.
- [3] 石川 佳治, 佐々木 勇和, 築井 美咲, 高橋 正和, 杉浦 健人, "意味的な複合イベント処理を可能とするイベントベースについて", 情報処理学会研究報告, Vol. 2014-DBS-160, No. 22, pp. 1-8, 2014.
- [4] 高橋 正和, 築井 美咲, 佐々木 勇和, 石川 佳治, "オントロジとデータベース技術を活用した複合イベント処理システム", 第7回データ工学と情報マネジメントに関するフォーラム (DEIM 2015), 2015.
- [5] Eugene Wu, Yanlei Diao, and Shariq Rizvi, "High-performance complex event processing over streams", in *SIGMOD*, pp. 407-418, 2006.