

報告番号	※甲	第	号
------	----	---	---

主論文の要旨

論文題目 確率的データストリームにおける柔軟なイベント検出に関する研究
氏名 杉浦 健人

論文内容の要旨

センサ技術及び機械学習技術の発展により、センシングデータの収集及びそれらの解析による情報検出が容易となった。特に、スマートフォンやウェアラブルデバイスを用いた行動認識は活発に研究されており、より実用的な応用が期待されている。しかし、機械学習技術を用いてセンシングデータを解析したとき、センサのノイズや機械学習手法の制限により、解析結果には不確実性が含まれる。つまり、センシングデータの解析結果は、各時刻で各イベントの生起が確率的に表される確率的データストリームとなる。行動認識などの実用化を考えたとき、そのような確率的データストリームからいかに目的のイベントを検出するかが課題となる。

そこで、本研究では、確率的データストリーム上での正規表現によるパターン照合に取り組む。機械学習技術による解析の多くは主に単純なイベントのみを検出するため、現実的な環境で応用するには、そうした単純なイベントを組み合わせる複合イベント処理を適用しなければならない。しかし、確率的データストリームに対する既存研究の多くは選択や射影、集約などを対象としており、確率的な複合イベント処理は未だ発展途上である。したがって、本研究では非確率的データストリームで主に用いられているパターン照合に基づく複合イベント処理を適用し、その基盤的な技術を提案する。

確率的データストリームに対するパターン照合を考えたとき、以下の2つの課題がある。まず、確率的なパターン照合では大量のマッチが検出されてしまう。比較的単純な問合せパターンであっても検出されるマッチの数は指数関数的に増加するため、全てのマッチを出力するのは有効性・効率性両方の観点から望ましくない。つまり、何らかの手段によって出力するマッチを削減もしくは妥当な形に集約する必要がある。

次に、Kleene 閉包を含むパターン、つまり連続して生起するイベントを検出しようとしたとき、既存の評価指標では適切な照合結果が得られない点が問題となる。既存研究では適切なマッチを選ぶための評価指標として、主にマッチの生起確率を用いている。しかし、連続的なイベントの生起確率は小さくなりやすいため、単純な生起確率の絶対値では、本来検出したいマッチを見逃してしまう可能性がある。そこで、本研究ではグループ化によるマッチの集約手法と情報理論に基づくマッチの評価指標を提案し、確率的データストリームにおける柔軟なイベント検出を目指す。

本論文は全 6 章から構成される。各章の概要を以下に示す。

1 章では、前述した本研究の背景と課題についてより詳細にまとめ、本研究の目的と貢献を示す。

2 章では、本研究に関連する既存研究を俯瞰的にまとめるとともに、本研究の意義及び位置づけについて述べる。

3 章では、確率的データストリームや問合せパターンなど、本研究において基礎となる概念を定義する。また、確率的データストリームにおけるパターン照合の基本的なアルゴリズムについても紹介する。

4 章では、確率的なパターン照合によって生成される大量のマッチに対し、時間的なオーバーラップに基づくグループ化手法を提案する。確率的データストリームでは多数のマッチが検出されるが、実世界において同じパターンイベントの生起を示すような、本質的には区別する必要のないマッチもまた数多く存在する。そこで、そのような区別する必要のないマッチをグループとして集約し、パターン照合の結果として妥当なサマリの生成を目指す。具体的には、まずグループ化のセマンティクスを 2 つ定義し、各セマンティクスに応じたマッチのグループ化アルゴリズムを提案する。また、マッチのグループとしての生起確率を定義し、正規表現とオートマトンを活用した効率的な生起確率の計算手法を提案する。最後に、実データを用いてグループ化手法の検出性能と効率性を確認し、提案手法の有効性を示す。

5 章では、照合結果として有益なマッチを効率的に検出するために、情報理論に基づくマッチの評価指標を提案する。既存研究では生起確率の大きいマッチを優先的に検出しているが、Kleene 閉包による連続的な生起を考えた場合、生起確率では照合結果として本来出力したいマッチを見逃してしまう可能性がある。そこで、情報理論に基づくマッチの評価指標を定義し、ある時区間における相対的なマッチの尤度を定量化する。また、既存研究で用いられている照合セマンティクスに提案した指標を適用し、効率的な照合アルゴリズムを提案する。最後に、実データを用いて情報理論に基づく評価指標の検出性能を、シミュレーションによって効率性を評価し、提案手法の有効性を示す。

6 章では、本研究の提案とその成果について再度まとめ、今後の課題について述べる。

