

自動車の時系列走行データを対象としたイベント抽出システムに関する研究

伊藤 貴之[†] 朝倉 宏一[‡]

大同大学大学院情報学研究科情報学専攻[†] 大同大学情報学部情報システム学科[‡]

1. 研究背景

時間の経過とともに記録されたデータのことを時系列データと呼ぶ。時系列データは株価や気象情報等、様々な分野でデータの時間的推移を解析する手法として用いられている。しかし、有益な情報が得られる反面、大量のデータを扱うので解析には労力を要する。そこで本研究では時系列データの値から解析できる特徴的な動きを「イベント」として定義し、抽出できるシステムの開発を目的とする。システムの応用例として、自動車のGPS時系列走行データを対象に、自動車の特徴的挙動をイベントとして抽出する実験を行う。

2. アプローチ

本稿では時系列データから得られる値の特徴的变化のことを「イベント」と定義する。時系列データの中から値の特徴的变化が起きた時区間（時間の区間）をイベントとして抽出するため、時区間論理に基づいた計算式を利用する[1]。時区間論理とは二つの時区間の関係を13の式で定義したものであり、本研究ではデータが記録された時刻を時区間と捉えて演算を行う。例えば、時間的に連続した動きを抽出する場合、時区間論理式の一つであるmeets式を利用する。meets式は以下の通りである。

$$\text{meets}(e1, e2) \text{ iff } e1.\text{etp} = e2.\text{stp}$$

meets式は「2つの時区間が連続している」ことを示す式である。ここでe1、e2は時区間を表しており、stp[start time point]は開始時刻、etp[end time point]は終了時刻を表している。この式を用いて2つのデータから1つのイベントを抽出する例を図1に示す。図1の四角形は時区間、楕円は時区間の関係を表しており、data1とdata2は9:00:00から9:00:05までに記録された2つのデータである。data1を時区間として扱う場合、記録時刻9:00:00がstpとなり、記録終了時刻9:00:03をetpと捉えることができる。同様にdata2にstpとetpを設定すると、meets

式のe1にdata1が、e2にdata2が当てはまることわかる。2つのデータは時間的に連続しているため、stpが9:00:00、etpが9:00:05となるイベントとして扱うことができる。

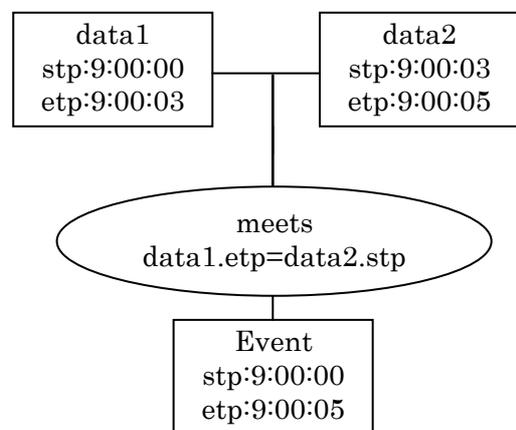


図1 meets式を用いたイベント抽出例

3. システム

イベント抽出システムはRubyで実装されている。ユーザがイベント定義としてRubyプログラムを記述し、抽出システム上で実行させる。イベント抽出システムではイベント定義に基づき、SQLデータベースへアクセスしてデータを参照し、データ中からイベントを抽出する。ここからは、プログラム上のイベントのデータ構造について説明する。

データベース上で取得した情報はデータの記録時刻からstpとetpを設定し、イベントとして配列に保存する。時区間論理の演算を行い新たなイベントを構成するが、このときイベント同士の関連付けは二分木構造で表現される。このデータ構造を図2に示す。e1~e5はイベントを表しており、四角形がデータベース上のデータから構成したイベント、角丸四角形がイベントを組み合わせて生成したイベントを示している。このように木を構成すると、イベント抽出処

理の段階で最初にデータとして生成される e1～e3 のような「データベース上から生成したイベント」が、必ず葉の部分になる。それらのイベントはデータベースから取得した情報を保持しているため、詳細なデータを損なわずにイベントを組み合わせた処理を行うことができる。また、イベントの親子関係は stp と etp を基準に構成しており、時区間論理式に基づいて 2 つのイベントから新しいイベントを生成する時、時系列的に早いイベントが左の子(stp)、遅いイベントが右の子(etp)として関連付けられる。

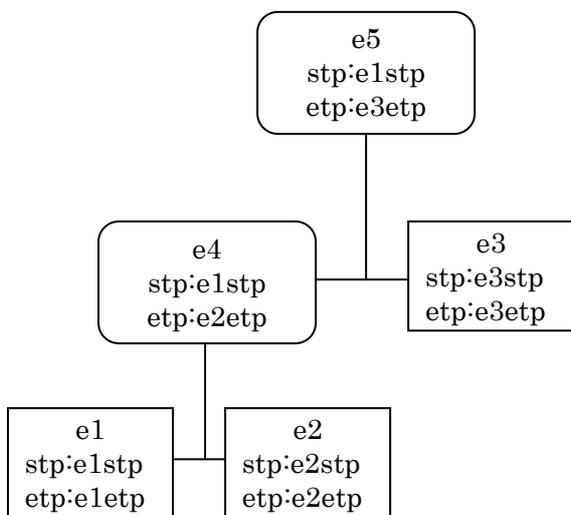


図2 イベントのデータ構造

4. 評価実験

本実験ではデータベース上に保存された自動車の時系列走行データ (GPS データ) の中から自動車の特徴的挙動をイベントとして抽出し、抽出処理にかかった時間、取得したイベントを地図に照合した結果からシステムの評価を行う [2]。対象となるデータは、経路 10 本分の自動車の走行データで、データ数は 21042 点である。走行データとして経度、緯度、進行速度、進行方向が保存されている。実験には CPU が Intel Core i3 3.20GHz、メモリが 2GB である計算機を使用した。時系列データは別サーバ上の SQL データベースに格納されており、ネットワークを介して参照する。実験で抽出したイベントは以下の通りである。

- ・ 停車：進行速度が 3.6km/h 以下の状態が 20 秒以上継続する (3.6 という数値設定は GPS データの計測誤差を考慮した結果)
- ・ 左折：進行方向が 1 秒間で 30 以上減少する

- ・ 停車後左折：停車イベントの後に左折イベントが発生する

停車後左折イベントは停車イベントと左折イベントを抽出した後、2 つのイベント集合の中から overlaps 関係にある (時区間の一部が重なっている) イベントを複合して抽出する。こうすることで自動車が停車した後に左折するという連続した動きが抽出できる。このとき、「停車イベントが時系列的に先である」、「重なっている時間は 1 秒以上」という 2 つの条件を追加する。以上の条件でイベントを抽出した結果、停車イベントは 77 件、左折イベントは 1085 件、停車後左折イベントは 76 件抽出し、抽出処理完了までに要した時間は 92 秒 (5 回計測の平均) であった。抽出したイベントを地図に照合した結果、左カーブのポイントにマーキングされたので GPS データから自動車の特徴的挙動をイベントとして抽出することは可能であるといえる。

5. まとめ

本研究では、時間とともに記録されたデータ (時系列データ) から、値の特徴的变化をイベントとして抽出するシステムを開発した。データベース上のデータからイベントを抽出するために、時区間論理に基づいたイベント抽出アルゴリズムを設計した。システムの評価実験として自動車から発信した GPS データの中から自動車の特徴的挙動をイベントとして抽出することを試みた。実験の結果、自動車の複雑な挙動をイベントとして抽出可能であることを確認できた。しかし、計 1238 件のイベントを抽出するために 92 秒という時間を要しており、大量のデータを扱う場合に実用的ではないという結果となった。そのため、時区間論理式の組み込み方法や SQL 命令文によるデータの絞込み方法に関して改良を行い、処理時間の短縮を行うことが今後の課題として挙げられる。

文献

[1] J.F.Allen:“Time Interval Logic”,
Communications of ACM, Vol.26, No.11,
pp.832-858 (1983).

[2] アイティティ:” GoogleMapsAPIv2 活用リファレンス”, 技術評論社, P.195(2007).