

修士論文

プローブカーデータを用いた
自動交通異常検出に関する研究

(A study on automatic traffic incident detection
based on probe car data)



赤塚 裕人

東京大学大学院 情報理工学系研究科 電子情報学専攻

指導教員 安達 淳

2013年2月6日提出

概要

筆者はリアルタイムな車の走行データであるプローブカーデータ (PCD: Probe Car Data) を用いた自動交通異常検出 (AID: Automatic Incident Detection) 手法を提案する。交通異常は深刻な渋滞の原因となり、その結果として経済的な損失や環境汚染、燃料浪費などを引き起こす。できる限り早く交通異常を検出する事は、交通マネジメントにおける非常に重要な課題の1つである。また、災害発生時には走行不能車や道路の寸断などの様々な交通異常が発生し、救急活動や災害復興の大きな妨げとなってしまうため、そのような観点からも交通異常の早期検出は非常に重要である。

従来の研究においては、あらかじめ道路に設置されたループディテクタと呼ばれる交通量を測定するセンサや、定点ビデオカメラから得られるデータを用いて交通異常検出を行っていた。しかし、それらのセンサはセンサを設置した特定の道路のデータしか得られず、また設置や維持に多大な費用と人員が必要となってしまうという問題があった。

これに対して PCD は、今ある車、そしてこれから販売される車に搭載されている GPS センサから得ることができるため、非常にデータの規模、範囲が大きく、また余計な費用をかけずに得られるという点で、従来の道路設置型センサに変わって大きな注目を集めている。しかし、PCD から得られるデータは道路設置型センサから得られるデータと異なっているため、従来の AID 手法を PCD に対して適用することはできない。そこで筆者は、PCD を用いた AID を実現するための新しい特徴量と、それを用いた交通異常検出手法を提案する。

PCD は GPS センサを搭載して道路網を動くセンサのような役割を果たすプローブカー (PC: Probe Car) から得られるデータである。よって、PCD を用いて AID を行うためには、1台1台の車から得られるデータを用いて対象道路の交通状態を把握する必要がある。そのような新しい特徴量として速度の“ゆらぎ”を提案する。この特徴量は対象道路を通過した1台のPCだけから得る事ができ、道路上の交通状態をうまく表す特徴量である。また、交通異常発生時の時間的、空間的な道路上の交通状態を分析することで、AID を行うための特徴量ベクトルを考案する。

次に，提案する特徴量ベクトル空間から実際に交通異常を検出するための2種類の異常検出手法を提案する．1つは特徴量ベクトル空間上で交通異常が発生された際に生成されるベクトルを含むようなハイパーレクタングルを定義する手法である．もう1つは，交通異常の際に生成されるベクトルは特徴量ベクトル空間上で外側に位置するという特徴を利用するスカイライン検索を用いた手法である．

実際のPCDと交通異常データを用いた評価実験の結果から，提案手法によって高い精度で交通異常を検出できることを確認した．これにより，PCDを使用したAIDシステム構築を実用的なものにできる可能性を示した．

目次

第1章 序論	1
1.1 研究の背景と目的	2
1.2 本論文の構成	5
第2章 関連研究	6
2.1 本章の構成	7
2.2 スカイライン検索	8
2.2.1 スカイライン	8
2.2.2 スカイラインの定義	9
2.2.3 スカイライン検索アルゴリズム	9
2.3 交通量－交通密度曲線	11
2.4 自動交通異常検出	14
2.4.1 AID 技術の評価法	14
2.5 道路設置型センサを用いた AID アルゴリズム	16
2.5.1 統計ベースアルゴリズム	16
2.5.2 交通パターンベースアルゴリズム	16
2.5.3 交通理論ベースアルゴリズム	17
2.5.4 予測ベースアルゴリズム	18
2.5.5 計算知能アルゴリズム	19
2.6 PCD を用いた自動交通異常検出	20
2.6.1 従来手法の問題点	22
第3章 交通異常検出のための特徴量ベクトル	23
3.1 研究の対象	24
3.2 Probe Car Data の特徴	25
3.3 特徴量に関する考察	26
3.3.1 平均速度・旅行時間の欠点	26
3.3.2 時間－空間的交通状態	26

3.3.3	速度の“ゆらぎ”	29
3.4	交通異常発生時の時間的, 空間的特徴	31
3.4.1	時間的特徴	31
3.4.2	空間的特徴	32
3.4.3	特徴量ベクトル	33
第4章	交通異常検出手法	34
4.1	交通異常検出手法の概要	35
4.2	ハイパーレクタングルを用いた異常検出手法	36
4.2.1	交通異常発生を検出	36
4.2.2	その他の異常状態を検出	36
4.3	スカイライン検索を用いた異常検出手法	38
4.3.1	手法概要	38
4.3.2	正常な範囲と異常な範囲の分割	38
4.3.3	異常の分類	40
4.3.4	先の道路区間の交通状態のチェック	41
4.3.5	手法の適用	41
第5章	評価実験	43
5.1	実験概要	44
5.1.1	実験環境	44
5.1.2	正解データの整理	44
5.2	しきい値の決定	47
5.2.1	ハイパーレクタングルの各しきい値	47
5.2.2	スカイライン検索手法のしきい値	48
5.3	実験結果	50
5.3.1	2手法の組み合わせ	51
5.4	考察	52
5.4.1	誤検出	52
5.4.2	検出できなかった交通異常	52
5.4.3	各手法の特徴	53
第6章	結論	55

謝辞	58
参考文献	59
発表文献	62

目次

2.1	東京都内の市区町村の本郷キャンパスからの距離（横軸）と家賃相場（縦軸）	8
2.2	最近某探索を用いたスカイライン検索アルゴリズムの動作例	10
2.3	交通量—交通密度曲線	12
2.4	交通異常検出のための <i>McMaster</i> テンプレート	18
2.5	旅行時間の変化を用いた交通異常検出手法 [LM05]	21
2.6	交通異常発生時の時間的，空間的な特徴を利用した交通異常検出手法 [ZWL09]	22
3.1	時間—空間的交通状態	28
3.2	事故発生区間における事故発生前（○）と発生後（×）の時間—空間的交通状態	31
3.3	事故発生区間に続く区間での事故発生前（○）と後（×）での時間—空間的交通状態	32
3.4	特徴量空間	33
4.1	2次元空間で n 層スカイライン検出を行った例	39
4.2	n 層スカイライン集合に対して k-means クラスタリングを行った例	41

表目次

5.1	対象道路区間	45
5.2	実験結果	50
5.3	2手法を組み合わせた性能評価	51

第 1 章

序論

1.1 研究の背景と目的

モータリゼーション（車社会化）の進行によって、自動車が広く社会に普及し、人や物の移動手段としてなくてはならないものとなってきた。しかしながら、それに伴って道路交通需要が大きく伸び、交通渋滞が深刻化してきている。国道交通省の推計によれば、日本全国で年間に発生する渋滞による経済的な損失は約 38.1 億人時間、1 人当たり年間 30 時間以上である。貨幣価値換算すると約 12 兆円となり、GDP の 2% 以上を占めている [kok06]。渋滞は経済損失だけでなく燃料の浪費や環境汚染など様々な問題を引き起こすため、渋滞緩和は様々な方面から求められている。また、渋滞は日本だけでなく世界中で問題となっており、渋滞緩和は最も大きな課題の 1 つとなっている。

交通渋滞には大きく分けて自然渋滞と交通異常による渋滞（異常渋滞）の 2 種類がある。道路の交通容量を超える車が流入すると自然渋滞が発生するが、国土交通省が中心となって道路整備を進めており、渋滞は少しずつではあるが毎年減少している [kok06]。

一方、異常渋滞は交通異常によって発生する。交通異常は交通事故や故障車、道路凍結、道路工事、その他、道路の通常の交通流を妨げる事象である。交通異常状態が 1 分間長く続くと、それによる渋滞時間はラッシュ時間帯以外で 4~5 分、ラッシュ時間帯であればさらに長くなり、それに伴って損失も大きくなる [TG05]。1998 年のアメリカでの調査では、交通異常検出の平均時間を 64 分から 41 分に減少させた結果、交通異常による経済損失が 44.6 百万ドル減少したという報告がある [A.G04]。したがって、できる限り早く交通異常を検出する事は、それによって引き起こされるその後の渋滞や、それに伴う損失を減らすために非常に重要である。

このような背景から、自動交通異常検出（AID: Automatic(Automated) Incident Detection）に関する研究が数多く行われてきた。従来の AID アルゴリズムでは、道路にあらかじめ設置されたループディテクタと呼ばれるセンサを用いて、道路上の車の台数や、ある地点を通過する車の台数を測定し、それを用いて異常検出を行ったり、定点ビデオカメラの映像に画像処理技術を適用して事故の検出を行っていた。しかし、これらの定点センサは道路にあらかじめ設置しておく必要があるため、センサを設置した特定の道路のデータしか得られず、また設置と維持に多大な費用がかかるという問題がある。

これに対して近年、Cyber Physical System(CPS) と呼ばれる実世界 (Physical Sys-

tem) とサイバー空間 (Cyber System) が緊密に結合されたシステムが登場した。安価なセンサの登場によって、個人が持っているスマートフォンに GPS センサや 3 軸加速度センサが搭載されるなど、実世界の事象が各種センサによって大量にデジタルデータとして観測できるようになった。これらのセンサは従来のように高い費用をかけて設置し維持する必要はなく、今ある物、これから発売される物にセンサが内蔵されて社会に浸透していくため、規模、範囲が非常に大きなデータを得る事ができる。CPS はそのような実世界に浸透したセンサーから得られる情報と、サイバー空間の IT 技術と組み合わせる事で、より効率のよい社会を実現するためのサービスおよびシステムのことである。

プローブカーデータ (PCD: Probe Car Data) はそのようなセンサから得られるデータの 1 つであり、GPS 受信機、GSM/GPRS 送信機を搭載したプローブカー (PC: Probe Car) のリアルタイムな車の走行データである。PCD は費用対効果の高さから現在ループディテクタやビデオカメラなどの従来の定点センサに変わって注目を集めている。PC は道路網を動くセンサのような役割を果たすため、従来のセンサのようにあらかじめ特定の道路に設置しておく必要がない。これにより、定点センサを設置した特定の道路だけではなく、広範囲の道路でのデータを得られるようになる。

以上の点を踏まえて、本研究の目的は交通異常を早期に緩和させるために、PCD を用いてリアルタイムに交通異常を検出する事である。リアルタイムに交通異常を検出するためには大量のストリームデータ処理が必要である。通常、異常検出においてはデータから何らかの特徴量を抽出し、機械学習等を用いて異常パターン、正常パターンを判断する事が多い。そのため、どのような特徴量をデータから抽出するかという点と、異常と判定する際にどのような指標を使うかが重要となってくる。PCD から得られる特徴量は、ループディテクタや定点カメラから得られる物とは異なるため、異常検出のための新しい特徴量が必要となる。また、リアルタイムに交通異常を検出するためには、過去のデータに対して 1 度学習を行えばよいわけではなく、時間と共に変化する状況に対応しなければならない。そのため、次々と到着する大量のストリームデータから適宜再学習を行う必要があり、そのような観点からも検討する。

そこで、筆者は PCD を用いて AID を行うための新しい特徴量と、それを用いた異常検出手法を提案する。本論文で提案する特徴量である速度の“ゆらぎ”は、異常検出の対象である道路区間を今通った 1 台の車のみから得る事ができ、道路上の交通状態を示す指標となる。また、交通異常発生時の状況を分析することに

よって、速度の”ゆらぎ”を用いた特徴量ベクトルを生成し、その特徴量ベクトル空間において異常検出を行う。実際のPCDに対して提案手法を適用した結果、正しく交通異常を検出できることが示された。

1.2 本論文の構成

本稿では、第2章にて異常検出において使用するスカイライン検索について簡単に紹介する。また、交通工学に簡単に触れた後、これまでに行われてきたAIDに関する研究について紹介する。第3章では、PCDから得られる特徴量について考察し、新しい特徴量である速度の“ゆらぎ”を提案する。また、交通異常発生時の時間的、空間的な状況を分析することで、AIDを行うための特徴量ベクトルを考案する。次に、第4章で提案する特徴量ベクトルを用いて交通異常を検出するための、異常検出手法について紹介する。第5章では、実際のPCDに対して提案手法を適用した結果を示す。最後に、第6章にて本稿についてまとめた後、今後の課題と方向性を示唆する。

第 2 章

関連研究

2.1 本章の構成

本章では関連研究について紹介する。まず、第2.2節で、異常検出手法のところで使用するスカイライン検索について、スカイラインの概要と定義、検出手法を簡単に説明する。

次に第2.3節で交通工学の分野から、道路上の交通状態を表す最も基本的な3つの特徴量について紹介し、3つの特徴量のお互いの関係について説明する。

最後に自動交通異常検出に関する関連研究を紹介する。まず、第??節でAID技術の評価法などを述べる。

AID アルゴリズムは大きく分けて2つの種類がある。1つ目は、従来の道路設置型センサを用いたアルゴリズムであり、ループディテク等のセンサから得られる交通量や交通密度といった特徴量を用いて交通異常検出を行う。2.5節で道路設置型センサを用いた従来のアルゴリズムを紹介する。2つ目は、PCDを用いたアルゴリズムである。PCDから得られるのは交通量や交通密度ではなく、個々のPCの位置や速度等の特徴量であり、これらの特徴量を用いて交通異常検出を行う。これについては2.6節で紹介する。

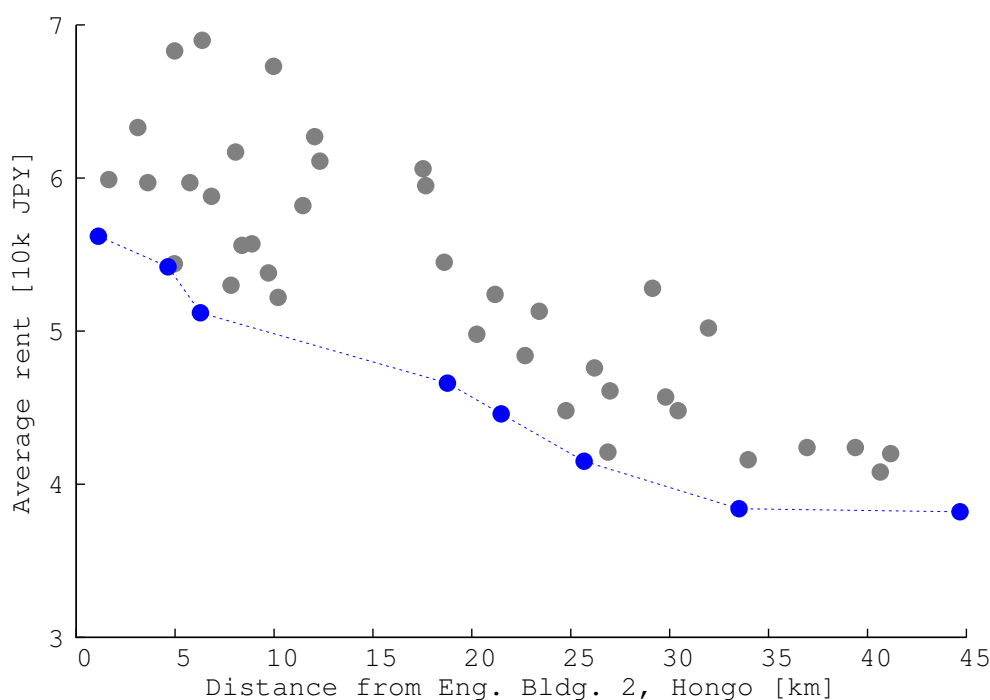


図 2.1: 東京都内の市区町村の本郷キャンパスからの距離 (横軸) と家賃相場 (縦軸)

2.2 スカイライン検索

2.2.1 スカイライン

スカイライン検索は単調な評価軸が複数存在する場合の意思決定の場面で用いられる [LYWL05; BKS01; PTFS05]. 例として, 東京大学本郷キャンパスに通う上でどこの市区町村が最も良いか考える場面を想定する. この場合の評価軸としては, 家賃相場, キャンパスからの距離, 治安の良さ, 人口密度, コンビニの数など様々なものが考えられるが, ここでは家賃相場とキャンパスからの距離を評価軸とする. 図 2.1 は東京都内のいくつかの市区町村について, 横軸を本郷キャンパスからの直線距離, 縦軸をその地域でのワンルームアパートの家賃相場としてプロットした物である.

各評価軸について, キャンパスからの距離は近ければ近いほど, 家賃は安ければ安いほど良いとすれば, ある市区町村よりも距離が遠く家賃が高い地域は明らかに最適解になり得ない. 最終的な最適解は距離と家賃のどちらに比重を置くかによって変わってくるが, 最適解は必ず距離, 家賃ともにより良い地域が存在しない地域の中, つまり図 2.1 の破線で結ばれた青い点の中に存在する. このようにして得られるデータの集合を**スカイライン**と呼ぶ.

2.2.2 スカイラインの定義

スカイラインはデータセット内の他のデータに *dominate* されていないデータの集合である。データセットにおいて、あるデータが他のデータを *dominate* しているとは、全ての評価軸において同じかより良く、また最低1つの評価軸でより良い場合である。

一般性を失わないので、検索対象データセット D 内の各データ p は d 次元のベクトルで各成分が非負の実数であるとする。データ点 p の第 i 成分 ($i = 1, 2, \dots, d$) を p_i で表す。また、ベクトルの各成分の値は小さければ小さいほど評価が高いとする。このとき、*dominate* の定義 (*dominance relation*) は次のようになる。

dominance relation 2点 $x, y \in D$ について x が y を *dominate* しているとは、次の2条件が同時に成立していることである。

1. $\forall i \in \{1, 2, \dots, d\}, x_i \leq y_i$
2. $\exists i \in \{1, 2, \dots, d\} \text{ s.t. } x_i < y_i$

データセット D のスカイラインは以下の式 (2.1) で定義される集合 $S \subset D$ である。

$$S \equiv \{p \in D \mid \neg(\exists p' \in D \text{ s.t. } p' \text{ dominates } p)\} \quad (2.1)$$

2.2.3 スカイライン検索アルゴリズム

スカイラインの検索アルゴリズムには分割統治法 [KLP75]、ブロック入れ子ループ [BKS01]、最近某探索法 [KRR02] など様々な種類がある。ここでは最近傍探索 (NNs: Nearest Neighbor search) に基づくアルゴリズム [KRR02] の2次元の場合の例を簡単に紹介する。データセット D 内で原点に最も近い点 (図 2.2 の NN) は、他の点に *dominate* されることがないため必ずスカイライン点である。最初のスカイライン点が見つければ、その点を基準として図のように空間を $A_1 \sim A_4$ に分割する事ができる。このとき、見つかったスカイライン点が *dominate* する領域である A_4 内の全ての点はスカイラインとなり得ないため探索する必要がない。また、当然であるが A_1 内にはデータ点は存在しない。残った A_2, A_3 領域内で原点に最も近い点を再び探索し同じ操作を繰り返す。NN アルゴリズムは、こうした分割と探索を繰り返して枝狩りを行いながらスカイラインを求めるアルゴリズムである。

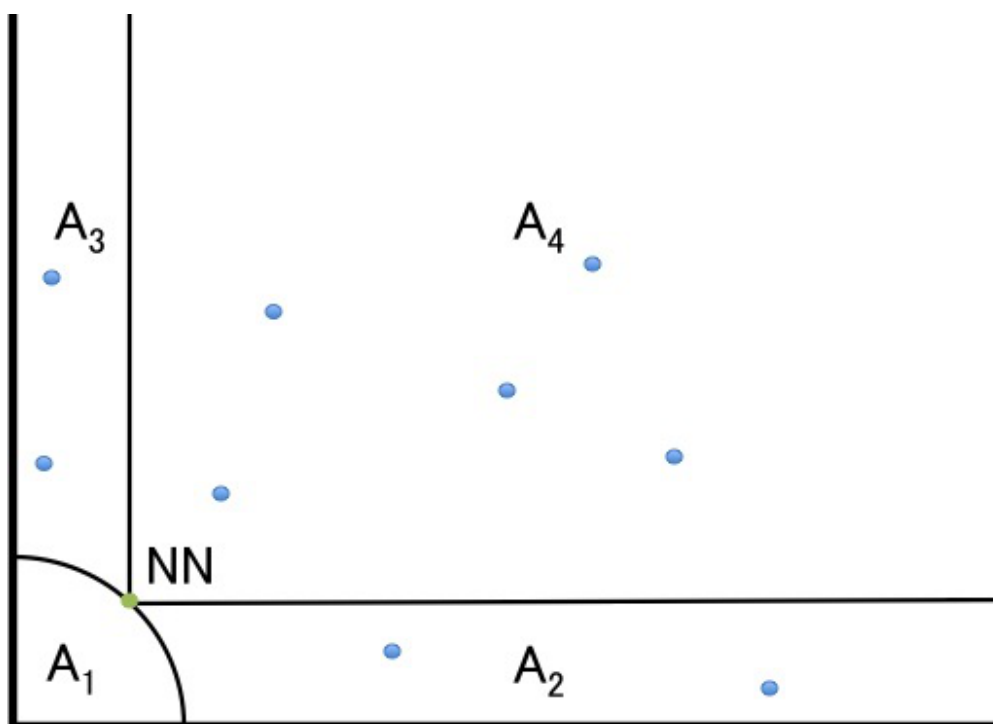


図 2.2: 最近隣探索を用いたスカイライン検索アルゴリズムの動作例

2.3 交通量—交通密度曲線

本節では交通工学に関する初期の研究 [Ric56; LW55] の内容から、基本的な事柄について簡単に説明する。

道路上の交通流の状態を表す代表的な特徴量として、交通量 q 、交通密度 k 、平均速度 v の3つがある。それぞれの詳細な定義を述べる。

交通量 交通量 q は道路上のある地点を単位時間当たりには通過する車の台数である。時刻 t 、時間間隔 τ をそれぞれ定め、時刻 $(t - \tau/2) \sim$ 時刻 $(t + \tau/2)$ の間に対象地点を通過した車の台数を n とするとき、流量 q は式 (2.2) で定義される。このとき、 τ は多数の車が通過できる程度に長い期間を取る必要がある。

$$q = n / \tau \quad (2.2)$$

交通密度 長さ dx の道路区間をとったとき、この道路区間をそれぞれの車が通過するのにかかった時間の合計を $\sum dt$ とする。このとき、交通密度 k は式 (2.3) で定義される。

$$k = \frac{\sum dt}{\tau dx} \quad (2.3)$$

つまり、交通密度 k は対象道路区間上の車の平均台数 ($\sum dt / \tau$) を区間の長さ dx で割ったものであり、道路上の単位長さ当たりの車の平均台数である。

平均速度 平均速度には2種類ある。まず、**空間—平均速度** v は以下の式 (2.4) で定義される。

$$v = \frac{q}{k} = \frac{dx}{\frac{1}{n} \sum dt} \quad (2.4)$$

空間—平均速度は、対象の道路区間の長さを、それぞれの車が区間を通過するのにかかった時間の平均で割ったものである。速度の遅い車、つまり対象区間により長い時間存在した車の速度により大きな重みがかかる。一方で、**時間—平均速度** v_t は以下の式 (2.5) で定義される。

$$v_t = \frac{1}{n} \sum (dx/dt) \quad (2.5)$$

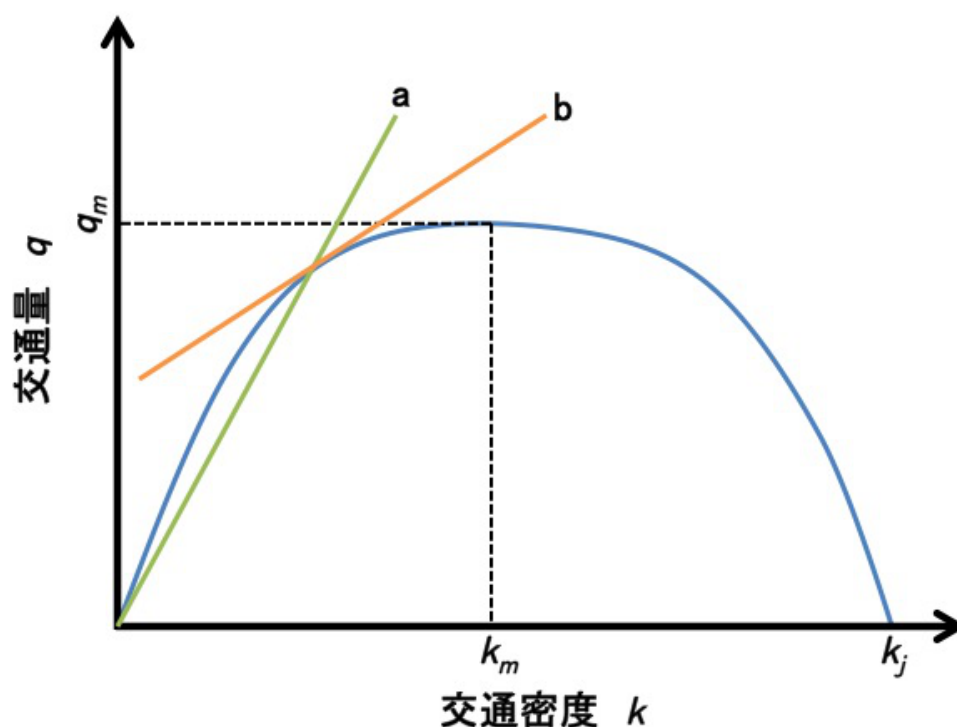


図 2.3: 交通量—交通密度曲線

これは対象の道路区間を通過した車の平均速度の平均であり，どの車の速度も重みは同じである．よって，時間—平均速度は空間—平均速度 v に比べて大きくなる．

例として，長さ $6[\text{km}]$ の区間をある車は $0.05[\text{h}]$ ，もう 1 台の車は $0.1[\text{h}]$ で通過した場合を考える．このとき，空間—平均速度は

$$\frac{6}{\frac{1}{2}(0.05 + 0.1)} = 80[\text{km/h}]$$

となる．一方で時間—平均速度は

$$\frac{1}{2}(6/0.05 + 6/0.1) = 90[\text{km/h}]$$

となり，時間—平均速度の方が大きくなることが分かる．

上記のように交通量 q ，交通密度 k ，平均速度 v を定義したとき，これらの特徴量の関係は図 2.3 のようになる．グラフ上の任意の点での平均速度 $v = q/k$ は原点とその点とを結ぶ接線の傾き (a) である．また交通量の交通密度に対する変化の割合 $c = dq/dk$ は接線の傾き (b) によって表される． $k = 0$ ，つまり道路上に 1 台も車がない場合，当然 q も 0 となる．また車両台数が多すぎて，車が動けなくなる

限界交通密度 $k = k_j$ のとき再び $q = 0$ となる。この両極端の値の間のある交通密度 $k = k_m$ で交通量は最大 $q = q_m$ となり、これを道路の交通容量と呼ぶ。交通密度 k が、車同士が互いの速度に影響を与え始める特定の値 $k = k_c$ を超えるまでは、 k が増えても v は影響を受けず、 $dv/dk = 0$ である。またそのような低交通密度の状態の時に限り $c = v$ である。 k が k_c を超えると、道路上の車両は互いの速度に影響を与え始める。そのような状態を道路が“混雑している”と表現する。

これは、道路が空いており、自分以外にほとんど車が走っていなければ自分の好きな速度で走行するが、ある程度他にも車が走行している場合は、前の車との車間距離によって速度を調整して走行するという我々の直観と一致していると言える。

2.4 自動交通異常検出

AIDの目的は、交通異常によって引き起こされるその後の損失を少しでも減少させるために、交通データを正しく分析し、可能な限り早く交通異常を検知する事である。本節では、AIDシステムを考える際の指標と、評価法について述べる。

2.4.1 AID技術の評価法

AIDシステムの評価 AIDシステムの構築を考える際の指標として以下の7つがある [A.G04].

1. 交通異常を検出するまでにかかる時間
2. 交通異常検出の精度
3. システム構築，維持にかかるコスト
4. システムの維持のしやすさ
5. 人員が必要であるか
6. 交通異常検出に用いるデータが他の目的にも役立つか
7. 技術が開発されてから，実際に運用開始されるまでの時間

AIDアルゴリズムの評価 AIDアルゴリズム評価の際には、*DetectionRate(DR)*、*FalseAlarmRate(FAR)*、*MeanTimeToDetect(MTTD)*の3つの指標が用いられることが多い [JNL⁺10]。DRは観測期間中に発生した交通異常の中で、正しく検出されたものの割合であり、以下の式 (2.6) で定義される。

$$DR = \frac{\text{正しく検出された交通異常数}}{\text{観測期間中に発生した交通異常数}} \quad (2.6)$$

FARはアルゴリズムを使って異常かどうかを判定した全入力インスタンスに対する、誤って交通異常と判断されたものの割合であり、以下の式 (2.7) で定義される。

$$FAR = \frac{\text{交通異常と判定されたインスタンスの中で実際には異常ではなかった数}}{\text{アルゴリズムで判定したインスタンスの総数}} \quad (2.7)$$

$MTTD$ は交通異常が発生した瞬間から、AID アルゴリズムが異常を検出するまでの平均時間であり、総異常数を n として以下の式 (2.8) で定義される。

$$MTTD = \frac{1}{n} \sum_{i=1}^n [\text{異常 } i \text{ 発生時刻} - \text{異常 } i \text{ 検出時刻}] \quad (2.8)$$

DR と FAR は AID アルゴリズムの実効性、 $MTTD$ は効率性を示す指標である。3つの指標は互いに関係している。検出の際に異常と判定するしきい値を変化させて条件を厳しくする事で FAR を低くしようとする、それに伴って DR も低くなってしまふ傾向がある。逆に、異常と判定する条件を緩めて DR を高くしようとする、 FAR も高くなってしまふ。また、例えばより長く観測対象について見ることによって FAR を低くしようとする、当然 $MTTD$ が増加してしまふ。このように3つの指標は同時に考慮される必要がある。

2.5 道路設置型センサを用いた AID アルゴリズム

本節では、道路設置型センサから得られる交通量や交通密度といった特徴量を用いて交通異常検出を行う従来の AID アルゴリズムを、いくつかのカテゴリに分類し簡単に紹介する。

2.5.1 統計ベースアルゴリズム

統計アルゴリズムは異常検出に用いる特徴量に対して基本的な統計手法を適用し、急激な特徴量の値の変化やその他の異常な振る舞いを検出する。例えば、過去のデータに対して特徴量の値の平均値を求めておき、それに対して現在得られた特徴量の値があらかじめ定義したしきい値以上外れていたら異常とする。特徴量としては交通密度や平均速度が用いられる事が多い。

簡単な例として *StandardNormalDeviation(SND)* アルゴリズムを紹介する。*SND* アルゴリズムでは過去のデータと現在のデータから *SND* を計算し異常を検出する [A.G04]。特徴量の過去のデータでの平均値を $\hat{x}(t)$ 、標準偏差を S 、現在値を $x(t)$ としたとき、 $SND(t)$ を以下の式 (2.9) で定義する。

$$SND(t) = \frac{\hat{x}(t) - x(t)}{S} \quad (2.9)$$

こうして求まった *SND* があらかじめ定義されたしきい値よりも大きければ交通異常として検出する。このアルゴリズムの利点はシンプルなことであるが、交通異常発生時の交通状態の変化の流れを捉える事は難しい。

2.5.2 交通パターンベースアルゴリズム

交通パターンベースアルゴリズムでは、対象道路から得られるデータをもとに交通パターンを判別し、異常な交通パターンが発生したときそれを交通異常として検出する。例えば、交通事故の発生により通行可能車線が減少すると、上流（事故発生地点に向かう車の流れ）の交通密度は増大し、下流（事故発生地点から去る車の流れ）の交通密度は減少する。このような交通異常発生時の交通パターンを発見することで異常検出を行う。

例として *California* アルゴリズムを紹介する。*California* アルゴリズムは観測地点から得られるデータから複数の特徴量を抽出し、決定木によって交通異常を

検出する [A.G04]. 抽出する特徴量としては、異常検出の対象とする道路区間内にある観測地点での交通密度の値、対象道路区間の下流の道路区間内の観測地点での交通密度の値、対象道路区間と下流の道路区間の交通密度の差異、交通密度の時間的变化等がある。交通異常が発生した際には、対象道路区間での交通密度の値は1単位時間前の値に比べて上昇し、下流の交通密度は減少するといったような特定のパターンが見られる。*California* アルゴリズムを用いた AID システムでは、そのような特徴量を基に交通状態を判別する決定木をあらかじめ作成しておく。そして、抽出した特徴量を決定木にかけ交通状態を推定し、交通異常が発生しているかどうか判定する。

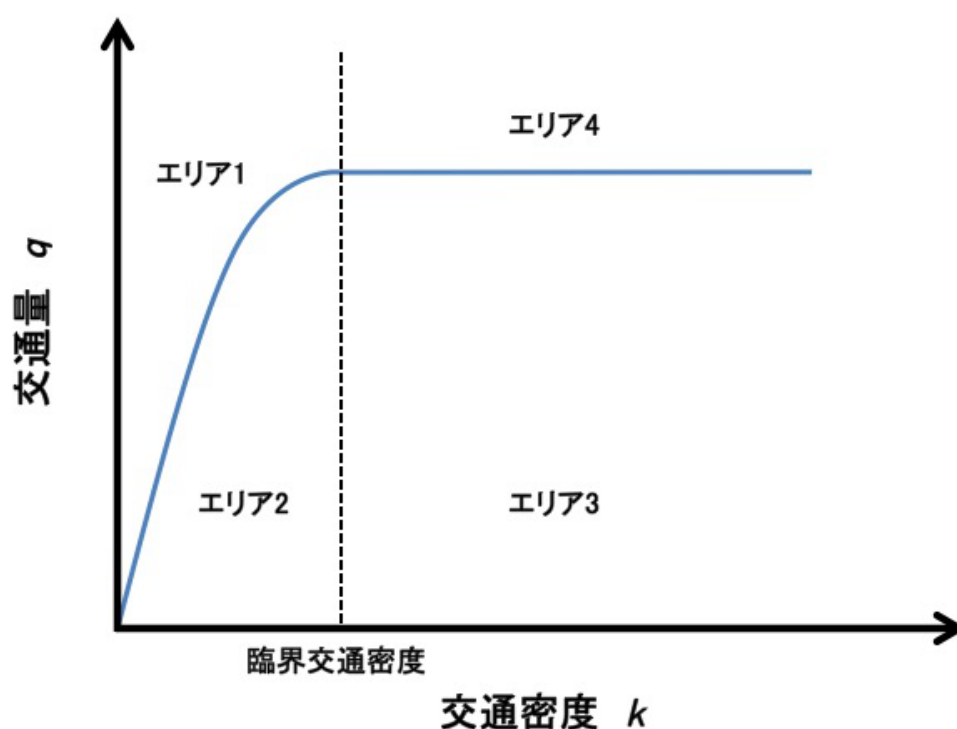
2.5.3 交通理論ベースアルゴリズム

2.3 節で交通量、交通密度、平均速度の関係について述べたように、道路上の交通状態を表す特徴量は影響を与え合うため互いに相関関係がある。交通理論ベースアルゴリズムでは、交通異常発生時には特徴量間の関係のルールが破られるという仮定を用いて交通異常検出を行う。

例として、*McMaster* アルゴリズムを紹介する。*McMaster* アルゴリズムは交通量—交通密度—平均速度の関係についてカタストロフィー理論を適用する [Coh94]。 “カタストロフィー” という単語は大変動や破局といった意味を持っており、秩序だった現象の中で不意に発生する無秩序な現象のことを指す。元の *McMaster* アルゴリズムは渋滞を検出するために考案されたものである。交通異常検出は渋滞をラッシュ時に発生するような自然渋滞と、交通異常によって発生する異常渋滞に区別することで行う。

McMaster アルゴリズムでは、各センサ設置点毎に図 2.4 のような *McMaster* テンプレートを生成する。ここで、エリア 1 は理想的な交通状態、エリア 2 は交通密度はそれほど高い状態ではないが平均速度が遅い状態、エリア 3 は交通密度が高く渋滞している状態、エリア 4 は交通密度は高いが交通量が多い状態である。交通異常検出対象地点のセンサと、その下流の地点のセンサでの値が入ったエリアを決定木にかけて交通異常が発生しているかどうか判定を行う。例えば、対象地点のセンサの値がエリア 2 もしくはエリア 3 に入った場合下流の地点のセンサの値をチェックする。そして、下流の地点のセンサの値がエリア 1 もしくはエリア 2 であれば交通異常が発生していると判定する。

しかし、先に挙げた交通パターンベースアルゴリズム、そして交通理論ベース

図 2.4: 交通異常検出のための *McMaster* テンプレート

アルゴリズム共に言える事であるが、交通流の特性は曜日、時間帯、天気、道路構造等様々な要因によって変化するため、モデル化することが非常に難しい。また、仮にモデル化できたとしても全ての交通異常の特徴をあらかじめ予想することは不可能である。

2.5.4 予測ベースアルゴリズム

予測アルゴリズムでは検出に用いる特徴量を時系列モデル化することによって値の予測を行い、現在値とモデルから得られる予測値との乖離度によって異常を検出する [CRS11; dFRV08]。

ここでは *ImprovedNomParametricRegression(INPR)-SND* アルゴリズム [TG05] について紹介する。 *INPR-SND* アルゴリズムでは k 近傍探索を用いて特徴量の値の予測を行い、予測値と現在値に対して統計アルゴリズムで紹介した *SND* の値を計算し異常検出を行う。検出に用いる特徴量は交通量か平均速度である。まず、以下の式 (2.10) のように 5 単位時間前の時刻から現在までの特徴量の値の列を現

在の状態ベクトル $X(t)$ とする.

$$X(t) = [q(t-5), q(t-4), q(t-3), q(t-2), q(t-1), q(t)] \quad (2.10)$$

ここで $q(t-i)$ は現在時刻 t から見て i 単位時間前の交通量の値である. 過去のデータについても状態ベクトルを求めておく. そして, $X(t)$ をクエリとして過去のデータに対して k 近傍探索を行い, 結果として得られた状態ベクトルの重み付き平均値を予測値とする. こうして得られる予測値と実測値の SND を計算し, あらかじめ定義されたしきい値よりも大きければ交通異常として検出する.

2.5.5 計算知能アルゴリズム

ニューラルネットワークやサポートベクタマシンなどの, 計算知能を用いた AID アルゴリズムが提案されている [CLLG10; SA01; Wan11].

例えば, サポートベクタマシンを用いた AID アルゴリズムでは, いくつかの特徴量を選んでサポートベクタマシンにかけ, 通常か異常か 2 つのクラスに分類することで交通異常検出を行う.

計算知能を用いたアルゴリズムでの問題点は, ほとんどの場合学習のための実正解データが十分得られないということである. そのため, 多くの研究では交通シミュレーションソフトを用いて交通異常をシミュレーションし, それを元に学習を行うという方法をとっている.

2.6 PCD を用いた自動交通異常検出

Probe Car Data には低コストで広範囲のデータが得られるという利点があるが、PCD から得られるのは交通量や交通密度といった統計値ではなく、個々の Probe Car から得られる走行データである。そのため、これまでの PCD を用いた研究においては、主に速度に関する特徴量が研究の対象となってきた。ある道路区間を通過する際にかかる旅行時間や、平均速度などのより正確な推定や予測の手法が研究されている [dFRV08; WTB⁺08]。また、対象道路の交通状態の推定や予測、ルート推薦などの研究も、特徴量としては旅行時間や平均速度など、速度に関するものを用いている [KDH⁺05; AGS10; YZXS11]。PCD を用いた AID アルゴリズムの研究も例外ではなく、異常を検出するための特徴量としては主に平均速度を用いている。

旅行時間を平均と比較する手法 Cheu らは現在の旅行時間と、対応する曜日、時間帯での旅行時間の平均値とを比べることで交通異常を検出する手法を提案している [RLCL02]。この手法は、交通異常発生時には通常時に比べて旅行時間が大きくなるという前提に基づいている。しかし、旅行時間は自然渋滞においても当然大きくなるため、この手法では交通異常によって発生する渋滞と自然渋滞を区別することができないと言える。

旅行時間の変化を捉える手法 Li らは旅行時間の変化を用いて急激な旅行時間の増加を捉え交通異常を検出する手法を提案している [LM05]。まず、単位時間間隔を 5 分に設定し、各時間間隔内で対象道路区間を通過した PC の旅行時間の平均値 T_i を求め、これを 1 つ目の特徴量とする。次に、 i 番目の旅行時間の平均 T_i と、 $i-1$ 番目の時間間隔での旅行時間 T_{i-1} との差分 $DT_i = T_i - T_{i-1}$ を求め、これを 2 つ目の特徴量とする。過去の学習データからこの 2 つの特徴量を求めて図 2.5 のように 2 次元空間にプロットし、全体の 99% の値が含まれる範囲、99.9% の値が含まれる範囲をそれぞれ求める。そして、現在値のプロット点が入らなかった場合に交通異常として検出する。

交通異常発生時の特徴を分析する手法 Zhu らは交通異常発生時の時間的、空間的な状況を分析して、それを元に特徴量ベクトルを生成し、得られた特徴量ベクトル空間に対して、フィルタリング、はずれ値検出を用いて交通異常を検出する手法を提案している [ZWL09]。まず空間的な特徴として、交通異常発生時には図

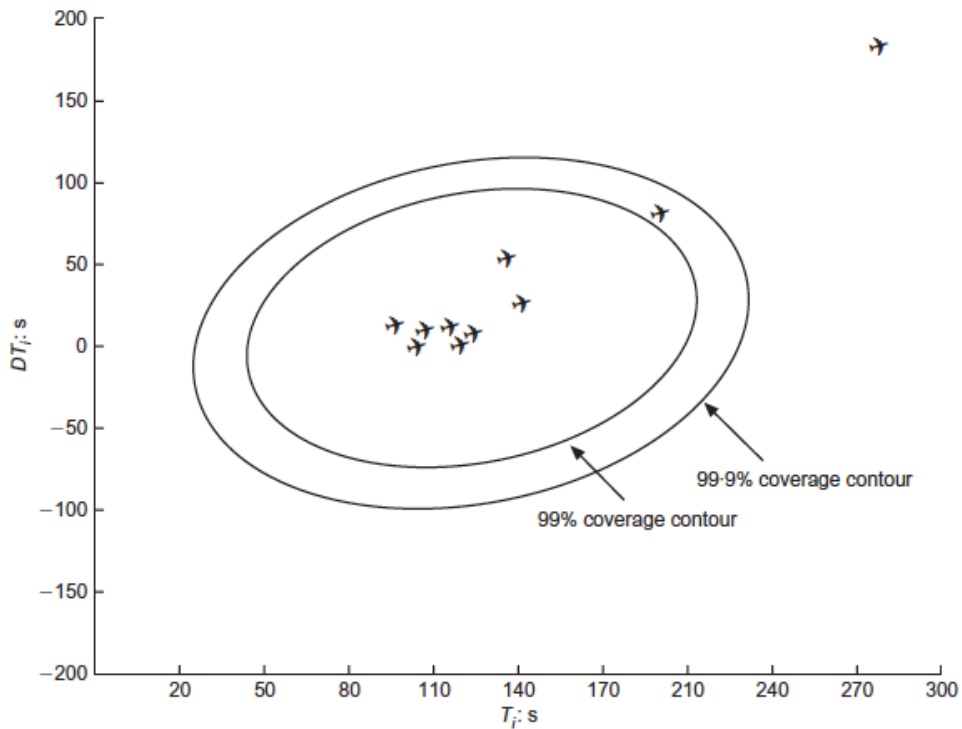


図 2.5: 旅行時間の変化を用いた交通異常検出手法 [LM05]

2.6 のように異常が発生した区間 l と、その前の区間 $l-1$ においては旅行時間が増加する。ある時刻 t 、区間 l における平均速度を $v(t, l)$ とするとき、これらはそれぞれ $v(t, l)$ 、 $v(t, l-1)$ という特徴量によって表される。逆に異常が発生した区間の先の区間では流入する車が減少することによって平均速度は上昇する。この上昇分を $\Delta v_{down}(t, l) = v(t, l+1) - v(t, l)$ という特徴量によって表す。また、交通異常発生時の時間的な特徴として、異常が発生した道路区間の平均速度が異常発生前と比べて減少する。このことは $dv(t, l) = v(t, l) - v(t-1, l)$ という特徴量によって表される。以上をまとめ、以下の式 (2.11) で表される特徴量ベクトル \vec{e}_d を生成する。

$$\vec{e}_d = \langle v(t, l), dv(t, l), v(t, l-1), \Delta v_{down}(t, l) \rangle \quad (2.11)$$

次に生成された各特徴量ベクトル \vec{e}_d をフィルタリングする。特徴量ベクトルの各値は交通異常発生時の特徴を表しているかどうかチェックされ、あきらかに交通異常でない場合はこのステップで除外される。1つ目のフィルタリング条件は、平均速度が1つ前の時間での値に比べて大きくなっているというもので、以下の条件式 (2.12) で表される。

$$dv(t, l) \leq 0 \quad (2.12)$$

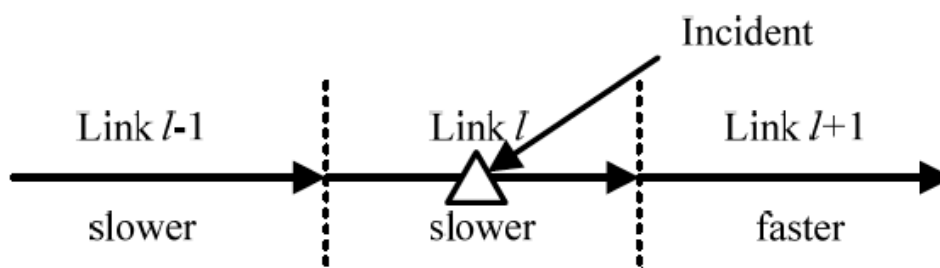


図 2.6: 交通異常発生時の時間的、空間的な特徴を利用した交通異常検出手法 [ZWL09]

2つ目のフィルタリング条件は、対象道路区間の先の区間での平均速度が対象区間の平均速度より大きいというもので、以下の条件式 (2.13) で表される。

$$\Delta v_{down}(t_k, l_k) > 0 \quad (2.13)$$

フィルタリングの結果残ったベクトルに対して、過去のデータを用いたはずれ値検出を行う。簡単に説明すると、過去のデータから得られる特徴量ベクトル集合のそれぞれのベクトルとの距離の和を計算し、あるしきい値を超えているものを交通異常として報告する。

2.6.1 従来手法の問題点

ここまで、PCD を使用した従来の自動交通異常検出を紹介してきたが、紹介した手法は特徴量として旅行時間を用いているため、旅行時間の精度によって大きく性能が変わってしまう。旅行時間の推定の精度は PC の台数に大きく依存するが [SJ96]、現状では PC だけから十分な精度の旅行時間を得る事は難しい。そのため、例えば Li らは対象道路に設置されたカメラの映像から得られるデータからランダムに選択した車を仮の PC として、旅行時間推定に用いる PC の台数を増やすということを行っている。しかし、それでは PCD の低コストで広範囲のデータを得られるという利点を生かす事ができていないと言える。

そこで、筆者は PCD の利点を最大限生かすため PCD だけから得られる特徴量について考察し、道路上の交通状態をうまく表す新しい特徴量を提案する。また、交通異常発生時の状況を分析することで AID のための特徴量ベクトルと、それを用いた異常検出手法を提案する。

第 3 章

交通異常検出のための特徴量ベクトル

3.1 研究の対象

本研究ではPCDを用いたAIDを行う。交通異常には交通事故や貨物落下等によって通行可能車線が一時的に減少する、道路工事によって長期間通行止めになる、悪天候や道路の凍結、積雪の影響で全体的に車の速度が減少するなど様々なものがある。

この中で、本研究では、交通事故や貨物落下等によって突発的に発生する交通異常をリアルタイムに検出することを目的とする。さらに、異常検出の対象を高速道路とする。この理由として、まず道路工事によって長期間通行止めや車線減少が発生する場合、事前に交通状態が変化することが分かるため、PCDを用いて検出する利点がない。悪天候等も同様である。高速道路を対象とするのは、一般道とは異なり車線が減少したからといって運転手は次の出入り口まではどうすることもできず、車線減少の影響を受けやすいという特徴があるためである。またPCDを用いてリアルタイムに異常を検出することで、より早い段階で通行止めとしたり、車線減少の情報を発信することができ、その後に発生する交通渋滞を緩和するのに大きな利点があるためである。

3.2 Probe Car Data の特徴

従来の道路設置型センサは設置や維持に多大な費用と人員が必要となってくる。また、センサが設置されていない道路を対象に交通異常検出を行おうとしても、センサが設置されるまでは不可能である。これに対してPCDは今ある車、そしてこれから販売される車から得る事ができる。PCは道路網を動くセンサのような役割を果たすため、そこから得られるデータは特定の道路に限定されることがなく、広範囲のデータを得る事ができる。

PCD技術はPCとPCDセンターとのデータのやり取りによって構成されている。各PCは一定時間毎に時刻、緯度経度による現在位置、瞬間速度、その他オプションデータをデータセンターに送信する。データを受け取ったPCDセンターは、道路網データを用いて対象車の追跡を行う。

PCDから抽出することができる特徴量は、対象道路区間の交通量や交通密度ではなく、個々のPCから得られる特徴量である。そのため、PCDを用いたAIDシステムの構築においては従来のAIDアルゴリズムとは違った視点が必要となってくる。従来の道路設置型センサから得られるデータを用いたAIDアルゴリズムにおいては、対象道路の交通量や交通密度といった“全体的な特徴量”を用いて交通異常検出を行っていた。しかしながら、PCDを用いたAIDにおいてはそれぞれのPCから得られる“個々の特徴量”を用いて交通異常検出を行わなければならない。だが、1台、もしくは非常に限られた台数の車から交通異常検出に必要な情報を得る事ができれば、PCDの利点を最大限に生かす事ができる。このような特徴量に必要な条件として、まず個人差の影響が小さいということが挙げられる。非常に限られた台数のPCDのみを用いるという状況においては、個人差が直接影響しないような特徴量が望ましいと予想される。また、他に必要な条件として、交通異常検出という目的に利用可能な特徴量であること、対象道路の交通状態をうまく表すような特徴量であることが挙げられる。次節で特徴量について考察する。

3.3 特徴量に関する考察

3.3.1 平均速度・旅行時間の欠点

平均速度は観測対象区間をある時間内に通った車の速度の平均、旅行時間は車が区間を通過するのに要した時間の平均である。渋滞や信号待ち、交通異常などによって車が区間内に滞在した時間が長いほど平均速度は小さくなる。また、当然であるが、区間長が長いほどそれに伴って旅行時間は増加する。どのルートを通っていけば一番早く目的に着けるかどうか知りたい場面等で、平均速度、旅行時間は重要な指標となってくる。

しかし、平均速度、旅行時間は車の速度の絶対値と直接関連しているため、個人差による影響が表れやすい。例えば高速道路であれば、走行車線を60km/hでゆっくり走る車もいれば、頻繁に追い越し車線を使って100km/hで走る車もいる。少ない台数の車から対象区間の交通状態を推測するという条件を考えると、平均速度や旅行時間は個人差の与える影響により問題があると言える。

また、交通異常検出のための特徴量という観点で見たときにもいくつかの欠点がある。まず、平均速度や旅行時間の絶対値がどのような意味を持つのが、対象とする道路区間によって変わってくるという点が挙げられる。例えば、旅行時間が10分であったということが分かったときに、それは通常より長いのか短いのかといったことを判断する必要があるが、そのためには、対象区間、対象時間帯毎の「通常状態」を定義する必要がある。次に、仮に旅行時間が普通より長いということが分かったときに、それはゆっくり走る車だったのか、道路が渋滞していたのか、天候条件が悪かったのか、事故によって突発的に車線が減少した結果なのか、といったことを区別する方法を編み出さなければならないという点が挙げられる。

3.3.2 時間一空間的交通状態

平均速度、旅行時間で挙げた問題点を解消するためには、個人差が直接影響してくる特徴量でなく、また、交通状態が悪いということがわかったときに、単なる自然渋滞と異常渋滞を区別できるような特徴量でなければならない。

そこで、車の速度の時間的な情報と空間的な情報の両方を含むような特徴量を考える。時間的交通状態上では、車がある地点で停止したり減速した場合に、それがどれだけの時間であったのかという時間特徴量を判断する。空間的交通状態上では、

道路のある地点だけで速度が遅くなったのか、それとも全体的にゆっくりであったのかという空間特徴量を判断する。このような、時間的な情報と空間的な情報の両方を含んだ特徴量として、**時間一空間的交通状態**が提案されている [YNL07]。本研究で提案する特徴量はこの時間一空間的交通状態を元に行っているため、まずこの概念について説明する。

時間一空間的交通状態は**時間的平均速度**と**空間的平均速度**の2つの速度の値で構成されている。

時間的平均速度 時間的平均速度 (TMS: Temporal Mean Speed) はいわゆる一般的な平均速度である。ある車が距離 L の対象道路区間を旅行時間 T で通過したとき、この車のこの道路区間における TMS は式 (3.1) で定義される。

$$TMS = \frac{L}{T} \quad (3.1)$$

空間的平均速度 空間的平均速度 (SMS: Spatial Mean Speed) は対象区間を一定の長さの小部分区間に分け、それぞれの小部分区間毎に車の平均速度を求め、対象区間全体でその平均をとったものである。距離 L の対象道路区間を長さ L/N の N 個の小部分区間に分割する。ある車がこの道路区間を通過した際に、 i 番目の小部分区間を通過するのにかかった時間を t_i とするとき、この車のこの道路区間における SMS は式 (3.2) で定義される。

$$SMS = \frac{\sum_{i=1}^N \frac{(L/N)}{t_i}}{N} \quad (3.2)$$

例として、総距離 $6[km]$ の道路区間をある車が $0.1[時間(h)]$ で通過した場合を考える。ただし、この車は区間中のある地点で $0.05[h]$ 停止し、それ以外は一定速度で走行したとする。簡単のため停止前後での加減速は考慮しない。また小部分区間の長さを $0.05[km]$ とする。このとき、この車は $6[km]$ の区間を $0.1[h]$ で通過したため、TMS は

$$\frac{6[km]}{0.1[h]} = 60[km/h]$$

となる。一方で、この車は停止していた時以外は一定速度 $6[km]/0.05[h] = 120[km/h]$ で走行したことになる。そのため、停止していた地点を含む小部分区間以外の残りの 119 の小部分区間での平均速度は $120[km/h]$ となる。停止していた地点を含む小部分区間での平均速度は、 $0.05[km]$ の小部分区間を $0.05[h] + 0.05[km]/120[km/h] =$

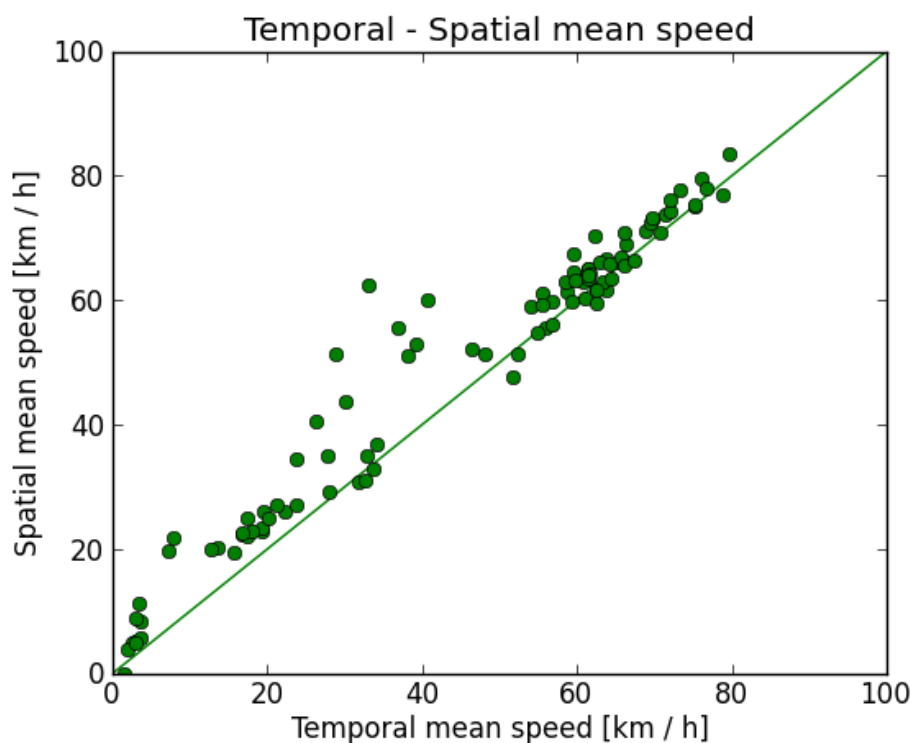


図 3.1: 時間—空間的交通状態

0.0504[h] で通過したため、 $0.05[\text{km}]/0.0504[\text{h}] \approx 1[\text{km}/\text{h}]$ となる。よって SMS は

$$\frac{120[\text{km}/\text{h}] * 119[\text{smallsegments}] + 1[\text{km}/\text{h}] * 1[\text{smallsegments}]}{120[\text{smallsegments}]} \approx 119[\text{km}/\text{h}]$$

となる。

このように、停止している時間が長ければ長いほど、また全区間に対して停止したり急激に減速する小分区間の割合が小さければ小さいほど、TMS に比べて SMS は大きくなる。2011 年 8 月 12 日に首都高速道路渋谷線のある道路区間を通過した各車について、TMS と SMS を求めその関係をプロットしたグラフを図 3.1 に示す。各プロット点が区間を通過したそれぞれの車にあたる。

Yoon らは信号の存在する一般道を対象に、グラフ上でプロット点が原点に近いほど交通状態が悪いとして、対象道路区間の交通状態の推定を行っている。信号の存在する一般道においては、信号で停止した車とそうでない車で TMS は大きく変化してしまう。SMS も考慮することで、単に信号で停止して TMS が遅くなった車が存在した場合に、それだけで対象道路が渋滞していると判定してしまうことを避けている。

しかし、Yoon らの研究の目的はあくまで対象道路の交通状態の推定であるため、

そのまま交通異常検出に用いることはできない。都市部の道路においてはラッシュの時間帯はもちろんのこと、それ以外の時間帯でも頻繁に渋滞が発生しているため、図 3.1 から分かるように多数の車が原点に近い位置にプロットされている。この日、この区間では交通事故が発生し、一時的に 2 車線ある道路のうち 1 車線が通行止めとなって異常渋滞が発生した。しかし、それぞれの速度の絶対値が小さくプロット点が原点に近いというだけでは、その他の自然渋滞と区別する事ができない。交通異常を検出するためには、多数の車の流入による自然渋滞が発生しているのか、交通異常によって車線が減少した結果、突発的な渋滞が発生したのかを区別する必要がある。

3.3.3 速度の“ゆらぎ”

本研究では、PCD を用いて交通異常を検出するために、前節で紹介した時間一空間的交通状態から新しい特徴量を抽出しそれを元に異常検出を行う。前節で述べたように、対象道路区間中で車が停止している時間が長ければ長いほど、また全区間に対して停止したり急激に減速する小分区間の割合が小さければ小さいほど、TMS に比べて SMS は大きくなる。

道路が空いている場合、ドライバーは自分の好きな速度で走行することができるため、速度の絶対値の大きさに違いはあるものの、ほぼ一定の速度で走行を続けるため TMS と SMS はほとんど同じ値となる。また、自然渋滞の場合においても、速度の増減を繰り返すことはあるものの、対象道路区間全体を通して比較的低い速度で走行を続けるため、TMS と SMS は共に小さくなるが、同じような値になると予想される。

一方で、交通事故が発生して走行可能車線が減少したときの状況を考えると、事故発生地点より前では走行可能車線の突発的な減少により、車の速度は大きく減少し、交通量が多い場合には長時間停止する事態にもなり得る。逆に、事故発生地点より先では、事故発生地点において交通流が妨げられる事によって流入する車の台数が減少して道路が空いた状態になり、ドライバーは好きな速度で走行できるようになる。そのため、交通異常発生時は TMS に比べて SMS が大きくなる事が予想される。

この TMS と SMS の違いの度合いは、時間一空間的交通状態のグラフにおいて TMS=SMS の比例直線 (図 3.1 の直線) と各プロット点との距離の大きさとして現れる。この距離は対象道路区間を通過した車の速度の“ゆらぎ”のようなものであ

り、対象道路区間の交通状態を示す指標となる。異常検出の対象とする道路区間を X とし、 X を i 番目に通過した車のこの“ゆらぎ” $dev_{i,X}$ を式 (3.3) で定義する。

$$dev_{i,X} = |TMS_{i,X} - SMS_{i,X}| \quad (3.3)$$

ここで $TMS_{i,X}$, $SMS_{i,X}$ は X を i 番目に通過した車の TMS, SMS である。 i については、対象区間 X を通過した車の順番が判定できればよい。つまり $i < j$ のとき i 番目の車は j 番目の車よりも早く対象区間 X を通過したことを示す。

この特徴量は同じ車の2つの種類の速度の値から求められるが、同じ車から得られる速度の違いを見ているため、速度の絶対値と違い個人差の影響を受けにくい。また前述したように、道路が空いている場合、自然渋滞の場合には TMS と SMS がほとんど等しくなるため“ゆらぎ”は小さくなるが、交通異常が発生した場合には TMS に比べて SMS が大きくなることにより“ゆらぎ”は大きくなる。よって、この特徴量によって自然渋滞と異常渋滞を区別する事が可能となる。

しかし、この特徴量もそれ単独では精度の高い交通異常検出には不十分である。図 3.1 で複数の点が比例直線から離れた位置にプロットされていることから分かるように、異常渋滞発生時以外でも“ゆらぎ”が大きくなることは十分起こりえることである。そこで、次節で交通異常発生時の時間的な特徴と空間的な特徴を分析する。

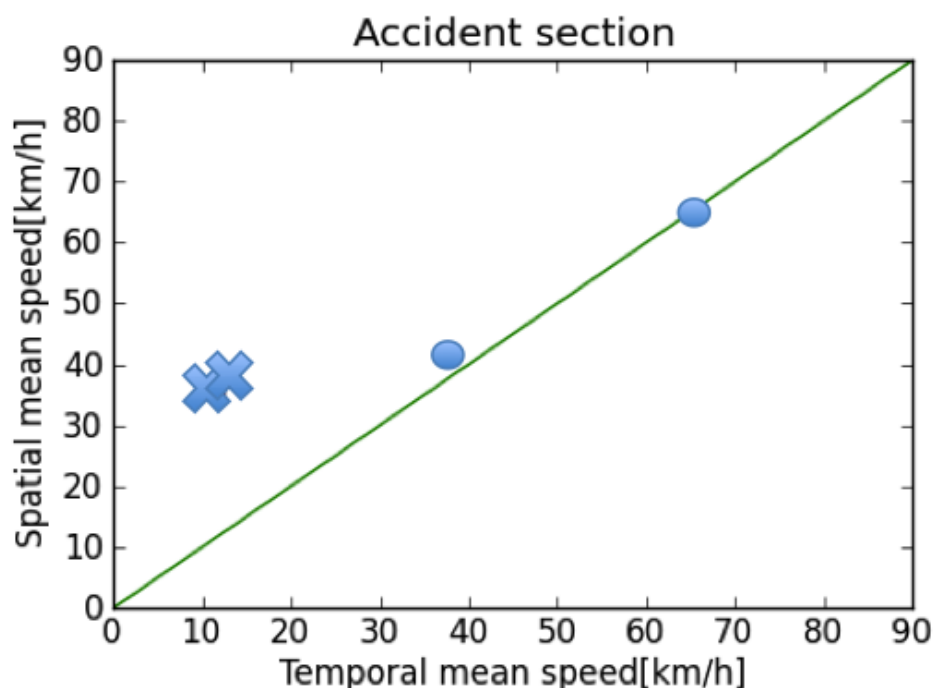


図 3.2: 事故発生区間における事故発生前 (○) と発生後 (×) の時間-空間的交
通状態

3.4 交通異常発生時の時間的, 空間的特徴

3.4.1 時間的特徴

ある道路区間 X で渋滞が発生したとする. この渋滞が自然渋滞である場合, 車の量が増えるに従って対象道路区間 X は徐々に混雑していき, それに伴って車の流れが徐々に悪くなっていく. しかし, 交通異常が発生し走行可能車線が減少した場合, それまで X において車の流れがスムーズだったとしても, 異常発生地点を先頭に急激に車の流れが悪くなり突発的に渋滞が発生する.

図 3.2 は図 3.1 から, 事故発生前に対象道路区間を通過した 2 台の車のプロット点を○で, 事故発生後に通過した 2 台の車を×で抜き出したグラフである. 図から分かるように, 事故発生前に対象区間を通過した車は速度の絶対値に違いはあるものの, 2 台とも TMS と SMS がほぼ等しく比例直線とプロット点との距離は非常に小さい. 一方で, 事故発生後に通過した 2 台の車は共に速度の絶対値が小さくなっているだけでなく, TMS に比べて SMS が大きくなったことによって, 比例直線とプロット点との距離が急激に大きくなっている. これにより, この区間で交通状態を急激に悪化させるような事象が発生したと推測できる.

このように, $dev_{i-1,X}$ と $dev_{i,X}$ によって, 対象道路区間 X における急激な交通状

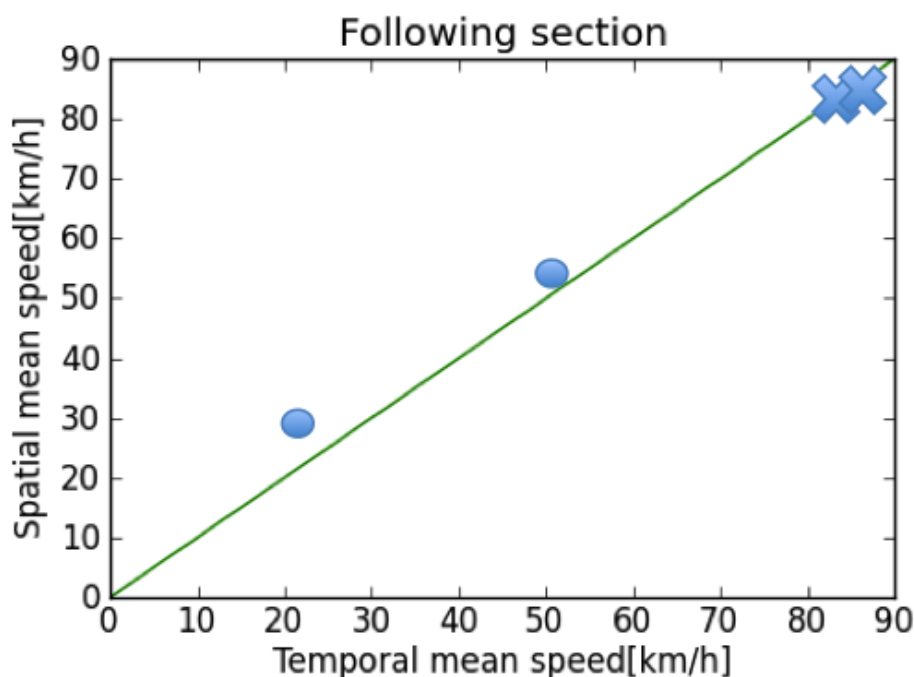


図 3.3: 事故発生区間に続く区間での事故発生前 (○) と後 (×) での時間-空間的交通状態

態の変化を捉えることができる。

3.4.2 空間的特徴

対象道路区間 X で発生した交通異常は X に続く道路区間 $downX$ の交通状態にも影響を与える。図 3.3 は図 3.2 と同じ 4 台の車が、 $downX$ を通過した際の時間-空間交通状態のグラフである。事故発生前に通過した 2 台の車は速度の絶対値が比較的 low、渋滞と判定するしきい値によっては渋滞と判定される可能性もある。一方で、事故発生後に通過した車は速度の絶対値が大きくなっている。そして、TMS と SMS がほぼ等しくなり比例直線上に乗っている。これは、この区間に流入する車の台数が減少し交通状態が理想的になることによって、ドライバーの好きな速度で走行できるようになったためであると推測できる。

このように、 $dev_{i,downX}$ と $TMS_{i,downX}$ により X に続く区間 $downX$ の交通状態が理想的な状態かどうかを判定できる。しかし、 $downX$ の交通状態が理想的であるからといっても、それが X での交通異常の発生を意味するわけではない。よって、この 2 つの指標は精度を向上させるためのフィルターとして使用する。

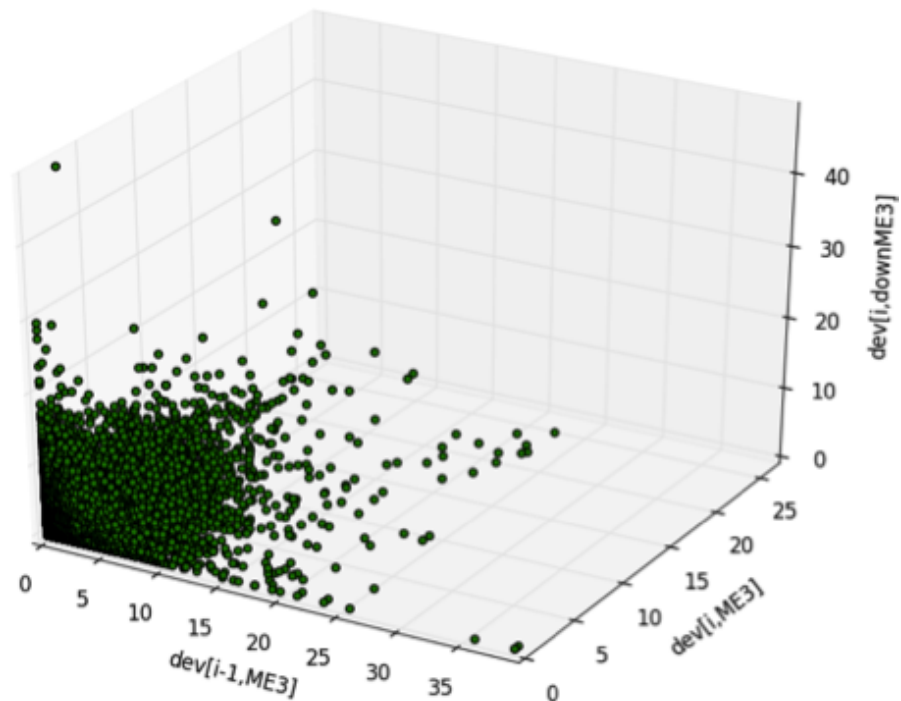


図 3.4: 特微量空間

3.4.3 特微量ベクトル

前節までの考察をまとめ, 対象道路区間 X で発生した交通異常を検出するための特微量ベクトル $\vec{f}_{i,X}$ を式 (3.4) で定義する.

$$\vec{f}_{i,X} = [dev_{i-1,X}, dev_{i,X}, dev_{i,downX}, TMS_{i,downX}] \quad (3.4)$$

本研究ではこの特微量ベクトルが最も重要であり, これを用いて交通異常検出を行う. 図 3.4 は 2011 年 2 月~2011 年 12 月までに首都高速渋谷線のある道路区間 (ME3) を通過した車について特微量ベクトルを求めプロットしたグラフである. 表示の関係で $TMS_{i,downX}$ は除外してある.

第 4 章

交通異常検出手法

4.1 交通異常検出手法の概要

本章では、前章で述べた特徴量ベクトルを用いた2種類の交通異常検出手法について説明する。

1つ目は、ハイパーレクタングルを用いた検出手法で、特徴量ベクトル空間に交通異常発生時のベクトルを含むハイパーレクタングルを生成する。過去のデータから各対象道路区間の速度の“ゆらぎ”のリストを生成し、k-means法を用いてクラスタに分割する事で、ハイパーレクタングルのしきい値を決定する。

2つ目は、スカイライン検索を用いた検出手法である。本手法で用いる特徴量ベクトルでは、交通異常発生時に生成されるものは、特徴量空間上で外縁として現れることが多い。そこで、特徴量空間でスカイラインを抽出し、異常とする範囲を決定する。

4.2 ハイパーレクタングルを用いた異常検出手法

4.2.1 交通異常発生を検出

これまで述べてきたように、対象道路区間 X で発生した交通異常は、 X における交通状態を急激に悪化させ $dev_{i,X}$ を急激に増加させる。また、 X で発生した交通異常は先の区間 $downX$ の交通状態にも影響を与える。 $downX$ は流入する車の台数が減少することで理想的な交通状態になるため、 $dev_{i,downX}$ が非常に小さくなり、 $TMS_{i,downX}$ は大きな値となる。このことは、あるしきい値 d_1, d_2, d_3, v_{min} を用いて以下の条件式 (4.1) で表す事ができる。

$$(dev_{i-1,X} \leq d_1) \wedge (dev_{i,X} \geq d_2) \wedge (dev_{i,downX} \leq d_3) \wedge (TMS_{i,downX} \geq v_{min}) \quad (4.1)$$

条件式 (4.1) は特徴量ベクトル空間におけるハイパーレクタングルを生成する。交通異常発生時に生成される特徴量ベクトル $\vec{f}_{i,X}$ はこのハイパーレクタングル内に含まれると考える。ここで、 $dev_{i-1,X}$ に対して d_1 、 $dev_{i,X}$ に対して d_2 という2つの異なるしきい値を使用することに言及しておく。 $dev_{i-1,X}$ と $dev_{i,X}$ の差分に対して1つのしきい値を使用すれば X における速度の“ゆらぎ”の変化を捉えることはできる。しかし、それだけでは速度の“ゆらぎ”が大きな状態からさらに大きな状態に変化した場合と、小さな状態から大きな状態に変化した場合を区別する事ができない。よって $dev_{i-1,X}$ は小さく $dev_{i,X}$ は大きいということを保証するために異なる2つのしきい値を使用する。

生成された特徴量ベクトル $\vec{f}_{i,X}$ の各値が条件式 (4.1) を満たしたとき、 $(i-1)$ 番目の車が通過してから i 番目の車が通過するまでの間に、対象道路区間 X で交通異常が発生したと判断する。

4.2.2 その他の異常状態の検出

交通異常による渋滞が継続している、交通異常による渋滞が解消した、といった情報も、交通マネジメントにおいては重要であると言える。そこで、これらの異常状態についても、それぞれに合わせたハイパーレクタングルを生成し検出する。

まず、交通異常による渋滞が継続している場合、 $(i-1)$ 番目の車が通過した際も i 番目の車が通過した際も、対象道路区間 X の交通状態は悪い状態であるため、 $dev_{i-1,X}$ 、 $dev_{i,X}$ は共に非常に大きくなると考えられる。また、 $downX$ の交通状態は

交通異常発生後と同じく理想的であるため、 $dev_{i,downX}$ は非常に小さく、 $TMS_{i,downX}$ は大きな値をとると予想される。よって、交通異常による渋滞の継続状態は以下の条件式 (4.2) で表される。

$$(dev_{i-1,X} \geq d_2) \wedge (dev_{i,X} \geq d_2) \wedge (dev_{i,downX} \leq d_3) \wedge (TMS_{i,downX} \geq v_{min}) \quad (4.2)$$

次に、 $(i-1)$ 番目の車が通過してから i 番目の車が通過するまでの間に交通異常による渋滞が解消した場合、 $(i-1)$ 番目の車が通過した際は対象道路区間 X の交通状態は悪い状態であるため、 $dev_{i-1,X}$ は非常に大きいが、 i 番目の車が通過した際は交通状態は通常状態に戻っている。よって交通異常による渋滞が解消されたという事象は以下の条件式 (4.3) で表される。

$$(dev_{i-1,X} \geq d_2) \wedge (dev_{i,X} \leq d_1) \wedge (dev_{i-1,downX} \leq d_3) \wedge (TMS_{i-1,downX} \geq v_{min}) \quad (4.3)$$

4.3 スカイライン検索を用いた異常検出手法

4.3.1 手法概要

交通異常が発生すると対象道路区間 X の交通状態が悪化し、速度の“ゆらぎ”が大きくなる。そのため、交通異常が発生した後に X を通過した PC から生成される特徴量ベクトル $\vec{f}_{i,X}$ は、特徴量ベクトル空間において原点から遠い外側にプロットされると考えられる。そこで、外側のベクトル点を取ってくるスカイライン検索を用いて、特徴量ベクトル空間を正常な範囲と異常な範囲に分別することを考える。この異常な範囲は、先に述べたハイパーレクタングルによる手法と違い、交通異常の発生、交通異常による異常渋滞の継続、異常渋滞の解消といった全ての異常状態で生成される特徴量ベクトルを含む。そして、過去のデータで異常な範囲にプロットされたデータを k-means によってクラスタリングすることで、異常な範囲をさらに交通異常の発生、交通異常による異常渋滞の継続、異常渋滞の解消それぞれの範囲に分割する。

特徴量ベクトル $\vec{f}_{i,X}$ は、対象道路区間 X の交通状態を示す $dev_{i-1,X}$, $dev_{i,X}$ と、 X の先の道路区間 $downX$ の交通状態を示す $dev_{i,downX}$, $TMS_{i,downX}$ に分かれる。 X での交通異常の発生によって交通状態が悪化し速度の“ゆらぎ”が大きくなる道路区間は当然 X である。よって、スカイライン検索は $dev_{i-1,X}$, $dev_{i,X}$ の 2次元空間で行い、この 2次元空間を正常な範囲と異常な範囲に分割する。先にも述べたように、 $dev_{i,downX}$, $TMS_{i,downX}$ はフィルターとして機能するため、 $\vec{f}_{i,X}$ の $dev_{i-1,X}$, $dev_{i,X}$ が 2次元空間で異常な範囲にプロットされた場合に、 $downX$ の交通状態が理想的であるかを確認する目的で使用する。

4.3.2 正常な範囲と異常な範囲の分割

4次元の特徴量ベクトル空間のうち、 $dev_{i-1,X}$, $dev_{i,X}$ の 2次元空間をスカイライン検索を用いて正常な範囲と異常な範囲に分割する。基本的に、速度の“ゆらぎ”は大きければ大きいほど単純に異常度が高いと言えるため、特徴量ベクトル $\vec{f}_{i,X}$ が $\vec{f}_{j,X}$ ($i \neq j$) より $dev_{i-1,X}$, $dev_{i,X}$ の 2次元空間で優位である ($\vec{f}_{i,X}$ が $\vec{f}_{j,X}$ を *dominate* する) のは、次の 2条件が同時に成立している場合である。

1. $(dev_{i-1,X} \geq dev_{j-1,X}) \wedge (dev_{i,X} \geq dev_{j,X})$
2. $(dev_{i-1,X} > dev_{j-1,X}) \wedge (dev_{i,X} > dev_{j,X})$

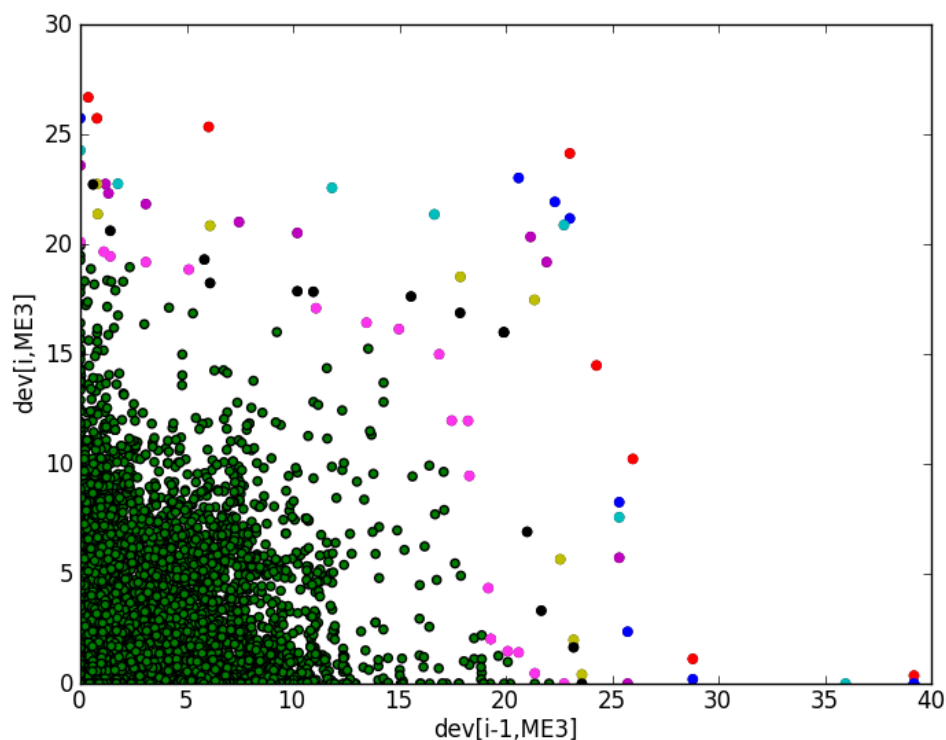


図 4.1: 2次元空間で n 層スカイライン検出を行った例

しかし、単純に一番外側のスカイラインを取ってきて、その外側を異常な範囲とするといった手法では問題があると言える。まず、ハイパーレクタングルを使った手法の説明で述べたように、交通異常が発生した直後に生成される特徴量ベクトル $\vec{f}_{i,X}$ の $dev_{i-1,X}$ と、交通異常による異常渋滞が解消した直後に生成される $\vec{f}_{i,X}$ の $dev_{i,X}$ は小さくなる。優位条件を $dev_{i-1,X}$, $dev_{i,X}$ が共に大きいとしているため、どちらかが小さいベクトルはもう一方がその近辺のベクトルで1番大きな場合しかスカイライン点となることができない。また、 $dev_{i-1,X}$, $dev_{i,X}$ 共に非常に大きく、他の全てのベクトルに対して優位になるような $\vec{f}_{i,X}$ が存在した場合、異常な範囲が原点から非常に遠い範囲になってしまい、異常と判断されるべき $\vec{f}_{i,X}$ も正常な範囲に含まれてしまう。

そこで、タマネギの皮を外側から何枚も剥いていくように、スカイラインを n 層抽出し、 n 層のスカイラインの中で一番内側（原点に近い側）のスカイライン集合が *dominate* する範囲を正常な範囲、そうでない範囲を異常な範囲とする。スカイラインを抽出する際は、まず対象とする特徴量ベクトル集合全体に対してスカイライン検索を行い、得られたスカイラインの集合を第1層のスカイライン集

合とする。次に、第1層のスカイライン集合を除いたベクトル集合に対して再びスカイライン検索を行い、得られたスカイラインの集合を第2層のスカイライン集合とする、という操作を繰り返して n 層のスカイラインを抽出する。図 4.1 は図 3.4 と同じ特徴量ベクトル集合を $dev_{i-1,X}$, $dev_{i,X}$ の2次元空間にプロットし、スカイラインとなるベクトルの n 層合計での数が全ベクトル数の1%を超えるまでスカイラインを抽出したものである。スカイライン以外のベクトルは緑色、スカイラインとなったベクトルは各層毎に色分けしてプロットしてあり、赤色が第1層、青色が第2層と続き、一番内側のピンク色が第7層である。

このように、あらかじめ過去のデータで n 層スカイラインを求め、最も内側の層のスカイライン集合が *dominate* する範囲を求めておく。そして、生成された特徴量ベクトルの $dev_{i-1,X}$, $dev_{i,X}$ がその範囲に入れば正常、そうでなければ異常と判断する。このようにして分割された異常な範囲には交通異常の発生、交通異常による異常渋滞の継続、異常渋滞の解消といった全ての異常状態で生成される特徴量ベクトルが含まれる。そのため、異常な範囲をさらにそれぞれの異常毎の範囲に分割する。

4.3.3 異常の分類

異常な範囲をさらに分割するため、 n 層スカイライン集合全体を k-means によって3つのクラスターにクラスタリングする。図 4.2 は図 4.1 の n 層スカイライン集合をクラスタリングした結果である。各クラスター毎に色分けしてあり、各クラスターのクラスター重心を黒色の四角点で表示している。

赤色の点で構成されるクラスターは $dev_{i-1,X}$ が比較的小さく、 $dev_{i,X}$ が大きな特徴量ベクトルを含むため、交通異常の発生を表す。黄色の点で構成されるクラスターは $dev_{i-1,X}$, $dev_{i,X}$ 共に大きな特徴量ベクトルを含むため、交通異常による異常渋滞の継続を表す。水色の点で構成されるクラスターは $dev_{i-1,X}$ 大きく、 $dev_{i,X}$ が小さな特徴量ベクトルを含むため、交通異常による異常渋滞の解消を表す。

異常な範囲にプロットされた特徴量ベクトルは、3つのクラスターのどのクラスター重心に近いかによって、何の異常を表すのか判断される。

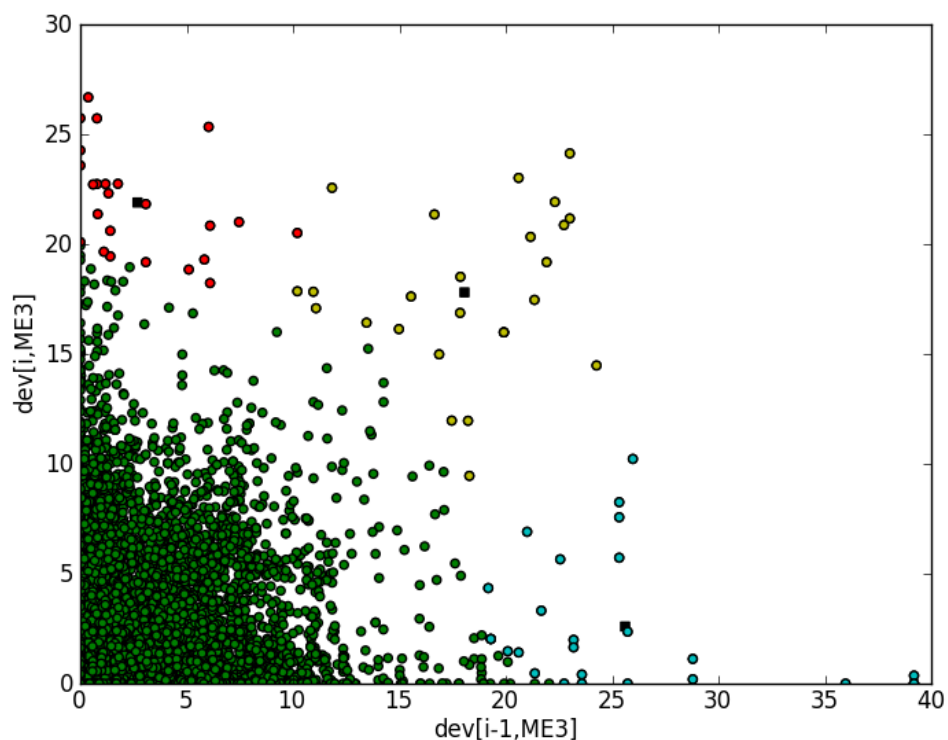


図 4.2: n 層スカイライン集合に対して k-means クラスタリングを行った例

4.3.4 先の道路区間の交通状態のチェック

交通異常検出の精度を向上させるため、先の道路区間の交通状態 $dev_{i,downX}$, $TMS_{i,downX}$ をチェックすることで、先の道路区間に流入する車の台数が減少しているかどうかを確認する。

ハイパーレクタングルを用いた手法で説明した条件式 (4.1) のうち、 $dev_{i,downX}$, $TMS_{i,downX}$ に関する部分をそのまま以下のフィルター条件式 (4.4) として用いる。

$$(dev_{i,downX} \leq d_3) \wedge (TMS_{i,downX} \geq v_{min}) \quad (4.4)$$

4.3.5 手法の適用

これまでに説明したスカイライン検索による異常検出手法をまとめると、適用手順は以下のようになる。

1. 過去のデータから特徴量ベクトルを生成し、 $dev_{i-1,X}$, $dev_{i,X}$ の 2次元空間で n 層スカイラインを抽出して正常な範囲と異常な範囲に分割する

2. n 層スカイライン集合を k-means によってクラスタリングして各異常に分割し, 各異常のクラスタ重心を求める
3. 異常検出対象データから特徴量ベクトルを生成し条件式 (4.4) によってフィルタリングする
4. フィルタリングによって除外されなかった特徴量ベクトルを $dev_{i-1,X}$, $dev_{i,X}$ の2次元空間にプロットし, 正常, 異常どちらの範囲に入るかを判定する
5. 異常な範囲に入った場合, どの異常のクラスタ重心に近いかを判定し, 最も近かったクラスタ重心の異常の発生を報告する

第 5 章

評価実験

5.1 実験概要

本研究で提案する特徴量ベクトル，そしてそれを用いた2種類の異常検出手法の有効性を評価するため，実際のPCDと交通異常のデータを用いて高速道路上の異常渋滞を検出する評価実験を行う。

5.1.1 実験環境

実験に用いるPCDは2010年1月から2011年12月までの2年分である。2010年のデータを過去データとしてしきい値を決定し，性能評価には2011年のPCDを使用する。正解データとして用いるのは2011年の首都高速道路の交通データである。

特徴量ベクトルを生成する際， $(i-1)$ 番目の車と i 番目の車の通過時刻の差があまりに大きいと，対象道路区間において何度も交通状態が変化することが考えられる。また一方で，たまたま，ほとんど同時に対象道路区間をPCが通過した場合，対象道路区間における急激な交通状態の変化を捉える事ができない。よって，特徴量ベクトルを生成し異常検出手法を適用するのは， $(i-1)$ 番目の車と i 番目の車の対象道路への侵入時刻の差が3分以上40分以下の場合とした。

5.1.2 正解データの整理

正解データとして使用するのは2011年の首都高速道路の交通データであり，交通異常の記録そのものではない。そのため，各交通異常の発生時刻や発生地点，車線が通行止めになるなどの影響があったかどうか，その影響がどの程度続いたかといった詳細な情報は分からない。

交通データには首都高速道路の各道路区間の5分毎の交通状態が記録されている。交通状態とは各5分間隔での車両密度や速度の平均値，自由流や渋滞といった交通流に関する記述等である。そしてオプション情報として，その道路区間，もしくは付近の道路区間で交通異常が発生し，それがその区間の交通状態に何らかの影響を与えている場合は「事故」や「落下物」といった異常の種類が記録されている。

そのため，正解データを整理し本研究で提案する異常検出手法の検出対象となる交通異常を抽出する必要がある。そこで，まずは，各交通異常がどの区間で発生したのかを推測する。ある区間 X で交通異常が発生しそれが交通状態に影響を

表 5.1: 対象道路区間

No	道路区間名	検出対象交通異常数	$\vec{f}_{i,X}$ の生成数
1	3号渋谷線 上り 用賀一駒沢	43	26231
2	4号新宿線 上り 永福一幡ヶ谷	30	21179
3	5号池袋線 上り 護国寺一飯田橋	18	17930

与えている場合、その影響は X を先頭に徐々に X の前の (X に向かう) 道路区間に広がっていく。そのため、オプション情報として付加されている交通異常の先頭が X である場合に、その交通異常は X で発生したと見なす。

各道路区間で発生した交通異常は次の3つに分類される。

1. 異常発生前後の PCD が存在しない
2. 交通異常は発生したが異常渋滞は発生しなかった
3. 交通異常が発生し、異常渋滞が発生した

まず1番は交通異常が発生したと思われる時間の前後に PC が存在しなかったために特徴量ベクトルが生成されなかった場合である。この場合はそもそも検出されるべき特徴量ベクトルが存在しないため、交通異常を検出できない。

次に2番は交通異常が発生し、その発生時刻の前後に PC が存在するが、異常渋滞は発生しなかった場合である。交通異常が発生しても異常渋滞が発生しないのは、事故が発生してもすぐに路肩に退避できたために車線には特に影響がなかった場合や、車線が一時的に通行不能となっても交通量が少なかったために渋滞にまでは至らなかった場合である。このような交通異常を検出するためには、交通異常を起こした車そのもののデータが必要となり、異常渋滞の検出という目的から外れてしまうため、検出の対象からは除外する。

今回検出の対象とするのは3番の交通異常が発生し、その発生時刻の前後に PC が存在し、異常渋滞が発生した事象である。性能の指標には *Detection Rate(DR)* と *False Alarm Rate(FAR)* を用いるが、その際に検出対象事象とするのはこの3番の事象である。

今回評価実験を行ったのは表 5.1 に示す首都高速道路の3つの区間である。検出対象交通異常数が上記で述べた3番の事象数であり、 $\vec{f}_{i,X}$ の生成数は2011年のPCDから生成された $\vec{f}_{i,X}$ の合計数である。

先に述べたように、使用する正解データは首都高速道路で発生した各交通異常のレポートではなく、5分毎の交通データである。正確な交通異常の発生時刻やその影響がどれくらい続いたのかといったことは分からないため、今回評価を行ったのは交通異常渋滞の発生についてのみである。交通異常渋滞の継続、交通異常渋滞の解消については評価を行っていない。

DR 評価の際、検出対象とする各交通異常は、異常渋滞が発生している間に生成された特徴量ベクトルが1つでも異常として検出された場合に検出できたとし、1つも異常として検出されなかった場合は、その交通異常は検出できなかったとする。

FAR 評価の際、交通異常の発生として検出された各特徴量ベクトルは、その生成時刻に交通異常による渋滞が発生していた場合に正解とし、渋滞が発生していない場合は誤りとする。

5.2 しきい値の決定

異常検出手法を適用する際には、ハイパーレクタングルを用いる手法、スカイライン検索を用いる手法ともに過去のデータからしきい値を決定する必要がある。本章では評価実験の際に使用したしきい値について述べる。

5.2.1 ハイパーレクタングルの各しきい値

速度の“ゆらぎ”の分布は各道路区間毎に異なっているため、条件式(4.1)の各しきい値はそれぞれの道路区間毎に過去のデータを用いて自動的に決定されなければならない。そこで、k-means法を用いて過去のデータから自動的にしきい値を決定する。

まず、過去のデータから得られる $dev_{i,X}$, $dev_{i,downX}$ のリストを k-means 法を用いてそれぞれ4つにクラスタリングする。直観的に言うと、それぞれのクラスタはクラスタ重心が原点に近い順から、速度の“ゆらぎ”が、非常に小さい、小さい、大きい、非常に大きい、値のクラスタである。 $dev_{i,X}$ のリストの各クラスタ重心を原点に近い順から $c_{1,X}$, $c_{2,X}$, $c_{3,X}$, $c_{4,X}$, $dev_{i,downX}$ についても同様に $c_{1,downX}$, $c_{2,downX}$, $c_{3,downX}$, $c_{4,downX}$ とする。

対象道路区間において交通異常が発生していない場合、車は一定の速度で走行を続けるか、もしくは自然渋滞が発生している場合、比較的低速度で走行を続ける、または加減速を何度も繰り返して対象道路区間を通過するため、速度の“ゆらぎ”はそれほど大きくなることはない。よって、しきい値 d_1 を式(5.1)のように設定する。

$$d_1 = \frac{c_{2,X} + c_{3,X}}{2} \quad (5.1)$$

交通異常が発生すると、対象道路区間 X の交通状態は急激に悪化するため速度の“ゆらぎ”は非常に大きくなる。このとき、交通異常による影響の大きさや道路区間内のどの地点で異常が発生したかによって、“ゆらぎ”がどの程度大きくなるかが変わってくる。そこでしきい値 d_2 を2パターン用意し、それぞれのしきい値を用いた場合に結果がどのように変化するかを分析する。1つ目の条件では以下の式(5.2)のように大きく設定する。

$$d_{2,1} = c_{4,X} \quad (5.2)$$

2つ目の条件では以下の式 (5.3) のように1つめの条件よりは少し小さく設定する。

$$d_{2,2} = \frac{c_{3,X} + c_{4,X}}{2} \quad (5.3)$$

X で発生した交通異常によって、 X の次の区間 $downX$ に流入する車の台数は減少し、 $downX$ の交通状態は理想的になり、速度の“ゆらぎ”は非常に小さくなる。よって、しきい値 d_3 を式 (5.4) のように設定する。

$$d_3 = \frac{c_{1,downX} + c_{2,downX}}{2} \quad (5.4)$$

日本には速度制限が 50[km/h] の高速道路区間が存在する。よって v_{min} を式 (5.5) のように設定する。

$$v_{min} = 50[km/h] \quad (5.5)$$

以上をまとめると、ハイパーレクタングルを用いた異常検出手法では、まず以下の条件式 (5.6) を条件 1 として用いる。

$$\begin{aligned} (dev_{i-1,X} \leq \frac{c_{2,X} + c_{3,X}}{2}) \wedge (dev_{i,X} \geq c_{4,X}) \\ \wedge (dev_{i,downX} \leq \frac{c_{1,X} + c_{2,X}}{2}) \wedge (TMS_{i,downX} \geq 50[km]) \end{aligned} \quad (5.6)$$

次に少し緩い条件 2 として、以下の条件式 (5.7) を用いる。

$$\begin{aligned} (dev_{i-1,X} \leq \frac{c_{2,X} + c_{3,X}}{2}) \wedge (dev_{i,X} \geq \frac{c_{3,X} + c_{4,X}}{2}) \\ \wedge (dev_{i,downX} \leq \frac{c_{1,X} + c_{2,X}}{2}) \wedge (TMS_{i,downX} \geq 50[km]) \end{aligned} \quad (5.7)$$

5.2.2 スカイライン検索手法のしきい値

スカイライン検索を用いた手法では、過去のデータから n 層スカイラインを抽出する際に、何層まで抽出するかという点が性能に大きく影響すると考えられる。そこで、スカイライン検索を用いた手法においても2種類の条件を用意する。

まず、条件 1 では過去のデータから n 層スカイラインを抽出する際に、 n 層合計のスカイライン数が全データの 1% を超えるまで抽出を繰り返す。

条件 2 では 2% を超えるまで抽出を繰り返す。つまり、より多層のスカイライン

を抽出することになるため異常の範囲が大きくなり、結果的に異常と判定される可能性が高くなる。これは異常と判定される条件を緩めるということになるため、ハイパーレクタングルを用いる手法、スカイライン検索を用いる手法、共に条件1より条件2のほうが緩い条件である。

続く区間の交通状態が理想的な状態になっているか確認するフィルタリングのしきい値については、ハイパーレクタングルの $downX$ に関する部分をそのまま用いる。つまり、以下の条件式 (5.8) を使用する。

$$(dev_{i,downX} \leq \frac{c_{1,X} + c_{2,X}}{2}) \wedge (TMS_{i,downX} \geq 50[\text{km}]) \quad (5.8)$$

表 5.2: 実験結果

No	ハイパーレクタングル		スカイライン	
	<i>DR</i> (%)	<i>FAR</i> (%)	<i>DR</i> (%)	<i>FAR</i> (%)
	条件 1	条件 1	条件 1	条件 1
	条件 2	条件 2	条件 2	条件 2
1	37.2	<u>0.015</u>	<u>44.2</u>	0.019
	53.5	0.072	<u>60.5</u>	<u>0.050</u>
2	57.7	<u>0.085</u>	<u>65.4</u>	0.118
	57.7	<u>0.208</u>	<u>76.9</u>	18.057
3	38.9	0.022	<u>50.0</u>	<u>0.017</u>
	<u>66.7</u>	0.078	61.1	<u>0.061</u>

5.3 実験結果

表 5.1 の 3 つの道路区間についての実験結果を表 5.2 に示す。まず No. 2 の道路区間のスカイライン検索を用いた手法の条件 2 における *FAR* が非常に高い値となっている点であるが、これは過去データから 2% を超えるまで n 層スカイラインを抽出する際に“ゆらぎ”が大きなデータが少なかったため、“ゆらぎ”が非常に小さな特徴量ベクトルまでが n 層スカイラインに含まれてしまったのが原因である。そのことにより、一番内側のスカイライン点集合による正常な範囲と異常な範囲の分割がほとんど意味をなさなくなり、*FAR* が非常に高くなってしまっている。しかし、それ以外の場合においては *FAR* はおおむね 0.1% を下回っており、検出された結果の精度は非常に高いと言う事ができる。

一方で *DR* については、条件 2 を用いた場合には 60% 前後となっている。条件 1 と比べると条件 2 を用いた場合にはおおむね 10% ほど *DR* が上昇している。それに伴って *FAR* も当然上昇しているが、年間で数件程度しか増えていない。実際の交通異常検出システムにおいては *DR* が高いほうが望ましいため、より緩い条件を用いる方が良いと考えられる。

表 5.3: 2 手法を組み合わせた性能評価

No	<i>DR</i> (%)	<i>FAR</i> (%)
	条件 1	条件 1
	条件 2	条件 2
1	51.2	0.019
	67.4	0.091
2	69.2	0.142
	76.9	18.089
3	55.6	0.033
	72.2	0.084

5.3.1 2 手法の組み合わせ

各手法で検出される特徴量ベクトルは大部分は重なっているが、一方だけで検出できた交通異常渋滞や、一方だけで誤検出された特徴量ベクトルが存在する。

ハイパーレクタングルを用いた手法とスカイライン検索を用いた手法の結果を組み合わせた場合の性能評価を表 5.3 に示す。

両手法の結果を組み合わせることによって、条件 2 を使用した場合には *DR* が 70%前後まで上昇している。

5.4 考察

評価実験の結果についての考察を行う。

5.4.1 誤検出

全体的に FAR は非常に低く、異常検出手法によって検出された結果の精度は非常に高いと言える。スカイライン点集合による異常な範囲と正常な範囲の分割が意味をなさなくなり FAR が非常に高くなってしまったケースについても、“ゆらぎ”が一定以上小さい点がスカイライン点に含まれるようになった時点で n 層スカイライン抽出をストップする、などのヒューリスティクスを追加することによって防ぐ事ができる。

5.4.2 検出できなかった交通異常

検出できなかった交通異常について分析を行ったところ、大きく分けて次の3つの要因があった。

1. 異常発生前から非常に混雑していた
2. 交通異常が対象道路区間のちょうど終わりで発生した
3. 実際は渋滞してない可能性

異常発生前から非常に混雑していた 交通異常が発生する前から対象道路区間を含む一帯の道路が非常に混雑している場合、交通異常によって一時的に車線が通行不能になりそこを通過する車の台数が減ったとしても、それだけでは先の道路区間の交通状態がすぐには良くなる。その結果、交通異常発生地点を抜けた先でも速度が上昇することがなく、速度の“ゆらぎ”は大きくなるため交通異常として検出できない。

交通異常が対象道路区間のちょうど終わりで発生した 対象道路区間 X のちょうど終わり付近で交通異常が発生した場合、異常発生地点を抜けて速度が上昇する頃には車はすでに次の区間 $downX$ 内を走行しているため、 X における時間的平均速度と空間的平均速度には速度の“ゆらぎ”は表れず、結果として交通異常として検出できない。

この問題には区間同士が重なるように区間を設定し、ある区間の終わりが次の区間の中間に位置するように区間の取り方を工夫する事で対応できると考えられる。ある区間 X の終わりで発生した交通異常は X においては速度の“ゆらぎ”が上昇せず検出できなくても、次の区間では異常発生地点が中間に位置するために速度の“ゆらぎ”を発生させ、結果として検出可能となる。

実際は渋滞していない可能性 正解データでは交通異常が発生しており、交通状態が渋滞となっても、実際のPCDを見ると特に速度を落とさずに車が通過していく場合がある。このような事が起きる原因としてまず考えられるのは、交通状態が渋滞となっても実際には交通密度が高いだけで、車の流れはそこまで悪くはないという場合である。これについては、実際どうであったかを知る事はできない。もう1つ考えられるのは、交通事故等が発生したものの、すぐに路肩に退避したために車線が通行不能になるほどの影響はなかったが、路肩に破損した車が存在することによってドライバーが速度を落とすという場合や、左車線だけ少し通行しにくくなるために速度を落とす必要があるが、右車線は全く問題なく走行できる状態であり、Probe Carはその右車線を通行した場合などである。このような場合には速度の“ゆらぎ”は交通異常として検出できるほどには大きくならないため検出できない。

5.4.3 各手法の特徴

各種法の特徴について考察を行う。

ハイパーレクタングルを用いた手法は v_{min} を除いても3つのしきい値があるため、細かい調整が可能である。しかし、その裏返しとしてチューニングコストの問題や、オーバーフィッティングを起こしてしまう可能性がある。

一方でスカイライン検索を用いた手法は何層までスカイラインを抽出するかによって異常となる範囲の大きさをコントロール可能であるが、あまりに小さいものは排除する等の処理を入れておかないと範囲分割が意味をなさない場合がある。

表5.3で示したように両手法を組み合わせることでDRが上昇した。ハイパーレクタングルを用いた手法では検出できなかったが、スカイライン検索を用いた手法では検出できた交通異常としては、 $dev_{i-1,X}$ が少し大きかったためにハイパーレクタングル手法では d_1 によって排除された特徴量ベクトルが、スカイライン検索手法では交通異常発生のクラスタ重心に一番近かったために検出されたというも

のがある。

一方で逆のパターンもある。スカイライン検索を用いた手法では異常な範囲と正常な範囲を分割する際にスカイライン検索を用いている。このとき、 $dev_{i-1,X}$ 、 $dev_{i,X}$ 共に大きな点ほど有利になるが、異常発生時に生成される特徴量ベクトルの $dev_{i-1,X}$ は小さい。そのため、道路区間によっては $dev_{i-1,X}$ が小さい場合、 $dev_{i,X}$ が非常に大きくないと異常な範囲に入らないということが起こりえる。このような場合には、スカイライン検索を用いた手法では異常な範囲に入らないために検出されないが、ハイパーレクタングルを用いた手法であれば検出されるということになる。

このように両手法は互いの欠点を補うことができるため、両手法を組み合わせることによって FAR を低く抑えたまま DR を上昇させることが可能である。

第 6 章

結論

Probe Car Data は非常に大規模な車の走行データを低コストで得る事ができる情報源である。これまでのPCDを用いたAutomatic Incident Detection アルゴリズムでは平均速度や旅行時間といった特徴量を用いていたが、これらの特徴量は個人差の影響を受けやすく、また自然渋滞と交通異常による異常渋滞を区別するのが難しいという問題があった。そこで筆者は、PCDを用いたAIDのための新しい特徴量と、それを用いた異常検出手法の研究に取り組んだ。

特徴量の提案では、限られた台数のPCから得られるデータだけから、異常検出の対象とする道路区間の交通状態を正しく把握するための特徴量として車の速度の“ゆらぎ”を提案した。また、交通異常発生時の時間的、空間的な交通状態の変化を分析することで、AIDを行うための特徴量ベクトルを考案した。

異常検出手法については、特徴量ベクトル空間から実際に交通異常を検出するための2つの手法を提案した。1つは、特徴量ベクトル空間上で異常となる点を含むハイパーレクタングルを生成する手法である。もう1つは、異常となる点は特徴量ベクトル空間で外側に存在するという特徴を利用したスカイライン検索による手法である。実際のPCDと首都高速道路で発生した事故データを用いて評価実験を行い、提案手法によって非常に高い精度で交通異常検出を行えることを確認した。これにより、実際にリアルタイムにPCDが流れてくる状況において、対象道路区間を車が通過した際に特徴量ベクトルを生成し異常検出手法を適用することで、リアルタイムに交通異常検出を行える事を示した。

今後取り組むべき課題は以下の通りである。まず、提案手法の特徴、限界等についてさらに詳細に評価する必要がある。評価実験では現状、首都高速道路だけを対象としているが、今後はさらに他の道路における交通異常のデータを入手し、さらなる評価をしていく必要がある。車線数や付近の出入り口数といった道路の構造や、事故の影響の大きさ、季節、時間帯、曜日などの要因が、提案手法の性能にどのような影響を与えるのかを詳細に分析する必要がある。

次に、異常検出の手法についてさらに詳細な検討が必要となる。現状はk-meansを用いてクラスタリングを行っているが、他のクラスタリング手法を用いた場合どうなるのかといった検討が必要である。実際のシステムにおいては、データはストリームデータとして次々と流入してくることが想定されるため、処理速度や必要なメモリ量について考察し実用に耐え得る手法にする必要がある。どのようにデータを保持しておくか、どれだけ過去のデータを残しておくか、など検討の余地は多い。また、しきい値をどのように更新していくかは非常に重要である。最初に決めたしきい値を使い続けるのか、データの流入に従って学習を繰り返して

徐々に変えていくのかを検討し、変えていく場合は大量のストリームデータからオンライン学習を効率良く行うためのアルゴリズムを考える必要がある。その場合、チューニングコストなども検討の対象となってくる。

また、多くの正解データが入手できたため、計算知能等の機械学習を用いた手法も今後検討したい。

最後に、現状は高速道路だけを対象としているが、実際にシステムを構築する場合には一般道においても異常検出を行える方がより役立つシステムとなる。今後はこれらの課題に取り組み実現していきたい。

謝辞

本研究を行うにあたり、多くの方々に大変お世話になりました。

指導教員である安達淳教授には、恵まれた研究環境を与えていただき、今日に至るまで充実した研究生活を送る事ができました。また、様々な知識に触れる機会や絶え間ないご指導を頂き、お陰様で修士研究をまとめあげる事ができました。深く感謝いたします。

高須淳宏教授には、修士研究における具体的な内容について常にご指導頂き、様々な分野の幅広い知識をご教授いただきました。ありがとうございます。

相原健郎准教授には、様々な企業の方々との共同研究の現場を見せていただき、面白い分野の研究に触れる機会を与えてくださいました。ありがとうございます。

鈴木貴敦先輩には、論文の書き方を丁寧にご指導いただき、内容についてもとても詳しいアドバイスをいただきました。また、研究以外にも私生活で役に立つ知識を教えてくださいました。ありがとうございます。

事務の久芳藍様には、NIIでの研究生活を支えていただき、事務手続きで大変お世話になりました。また、竹中明子様にはリサーチアシスタントや出張手続きで大変お世話になりました。

最後に、安達研究室の皆様には大変お世話になりました。ありがとうございます。特に、木村光樹先輩には、研究室配属当初から修士研究の進め方についてご指導を頂きました。また研究に対して様々な意見をおっしゃっていただいたり、参考となる本や論文等を教えてくださいました。ありがとうございます。英語を書くときにアドバイスいただいた Chu Yimin 先輩、一緒に修士研究を進めた同期の山本敬介君、ネットワークに詳しく研究室の研究環境をいつも整えてくれた木下僚君、大変お世話になりました。

赤塚 裕人

参考文献

- [A.G04] A.Guin. An incident detection algorithm based on a discrete state propagation model of traffic flow. *M.S. thesis, Georgia Inst. Technol.*, 2004.
- [AGS10] Chang Ande, Jiang Guiyan, and Niu Shifeng. Traffic congestion identification method based on gps equipped floating car. In *Intelligent Computation Technology and Automation (ICICTA), 2010 International Conference on*, pages 1069 –1071, may 2010.
- [BKS01] Stephan Börzsönyi, Donald Kossmann, and Konrad Stocker. The skyline operator. In *Proceedings of the 17th International Conference on Data Engineering*, pages 421–430, Washington, DC, USA, 2001. IEEE Computer Society.
- [CLLG10] Xueqing Cheng, Wenfang Lin, Enxiang Liu, and Dan Gu. Highway traffic incident detection based on bpnn. *Procedia Engineering*, 7(0):482 – 489, 2010.
- [Coh94] S. Cohen. Comparative assessment of conventional and new incident detection algorithms. *IEE Conference Publications*, 1994(CP391):156–159, 1994.
- [CRS11] Hao Chen, H.A. Rakha, and S. Sadek. Real-time freeway traffic state prediction: A particle filter approach. In *Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on*, pages 626 – 631, oct. 2011.
- [dFRV08] C. de Fabritiis, R. Ragona, and G. Valenti. Traffic estimation and prediction based on real time floating car data. In *Intelligent Transportation Systems, 2008. ITSC 2008. 11th International IEEE Conference on*, pages 197 –203, oct. 2008.
- [JNL⁺10] Guiyan Jiang, Shifeng Niu, Qi Li, Ande Chang, and Hui Jiang. Automated incident detection algorithms for urban expressway. In *Advanced Computer Control (ICACC), 2010 2nd International Conference on*, volume 3, pages 70 –74, march 2010.

- [KDH⁺05] B.S. Kerner, C. Demir, R.G. Herrtwich, S.L. Klenov, H. Rehborn, M. Alek-sic, and A. Haug. Traffic state detection with floating car data in road networks. In *Intelligent Transportation Systems, 2005. Proceedings. 2005 IEEE*, pages 44 – 49, sept. 2005.
- [KLP75] H. T. Kung, F. Luccio, and F. P. Preparata. On finding the maxima of a set of vectors. *J. ACM*, 22(4):469–476, October 1975.
- [kok06] 国土交通省, 効果的な渋滞対策の推進. <http://www.mlit.go.jp/road/ir/ir-perform/h18/07.pdf>, 2006.
- [KRR02] Donald Kossmann, Frank Ramsak, and Steffen Rost. Shooting stars in the sky: an online algorithm for skyline queries. In *Proceedings of the 28th international conference on Very Large Data Bases, VLDB '02*, pages 275–286. VLDB Endowment, 2002.
- [LM05] Y. Li and M. Mcdonald. Motorway incident detection using probe vehicles. *Proceedings of the ICE - Transport*, 158(1):11 – 15, February 2005.
- [LW55] M. J. Lighthill and G. B. Whitham. On kinematic waves. ii. a theory of traf-fic flow on long crowded roads. *Proceedings of the Royal Society of Lon-don. Series A, Mathematical and Physical Sciences*, 229(1178):pp. 317–345, 1955.
- [LYWL05] Xuemin Lin, Yidong Yuan, Wei Wang, and Hongjun Lu. Stabbing the sky: Efficient skyline computation over sliding windows. In *Proceedings of the 21st International Conference on Data Engineering, ICDE '05*, pages 502–513, Washington, DC, USA, 2005. IEEE Computer Society.
- [PTFS05] Dimitris Papadias, Yufei Tao, Greg Fu, and Bernhard Seeger. Progres-sive skyline computation in database systems. *ACM Trans. Database Syst.*, 30(1):41–82, March 2005.
- [Ric56] Paul I. Richards. Shock waves on the highway. *Operations Research*, 4(1):pp. 42–51, 1956.
- [RLCL02] Hongtu Qi Ruey Long Cheu and Der-Horng Lee. Mobile sensor and sample-based algorithm for freeway incident detection. *Transportation Re-search Record: Journal of the Transportation Research Board*, 2002.
- [SA01] A. Samant and H. Adeli. Enhancing neural network traffic incident-detection algorithms using wavelets. *Computer Aided Civil and Infras-tructure Engineering*, 16(4):239–245, 2001.

- [SJ96] Karthink K. Srinivasan and Paul P. Jovanis. Determination of number of probe vehicles required for reliable travel time measurement in urban network. *Transportation Research Record 1537*, pages 15–22, 1996.
- [TG05] Shuming Tang and Haijun Gao. Traffic-incident detection-algorithm based on nonparametric regression. *Intelligent Transportation Systems, IEEE Transactions on*, 6(1):38 – 42, march 2005.
- [Wan11] Qi Wang. Traffic incident detection based on artificial neural network. In *Communication Software and Networks (ICCSN), 2011 IEEE 3rd International Conference on*, pages 657 –659, may 2011.
- [WTB⁺08] D.B. Work, O.-P. Tossavainen, S. Blandin, A.M. Bayen, T. Iwuchukwu, and K. Tracton. An ensemble kalman filtering approach to highway traffic estimation using gps enabled mobile devices. In *Decision and Control, 2008. CDC 2008. 47th IEEE Conference on*, pages 5062 –5068, dec. 2008.
- [YNL07] Jungkeun Yoon, Brian Noble, and Mingyan Liu. Surface street traffic estimation. In *Proceedings of the 5th international conference on Mobile systems, applications and services, MobiSys '07*, pages 220–232, New York, NY, USA, 2007. ACM.
- [YZXS11] Jing Yuan, Yu Zheng, Xing Xie, and Guangzhong Sun. Driving with knowledge from the physical world. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '11*, pages 316–324, New York, NY, USA, 2011. ACM.
- [ZWL09] Tongyu Zhu, Jifang Wang, and Weifeng Lv. Outlier mining based automatic incident detection on urban arterial road. In *Proceedings of the 6th International Conference on Mobile Technology, Application & Systems, Mobility '09*, pages 29:1–29:6, New York, NY, USA, 2009. ACM.

発表文献

- [1] 赤塚裕人, 高須淳宏, 安達淳, “プローブカーデータを用いた自動交通異常検出”, 電子情報通信学会技術研究報告, Vol. 122 No. 202, ITS2012-15, 2012.
- [2] 赤塚裕人, 高須淳宏, 相原健郎, 安達淳, “プローブカーデータを用いた高速道路における自動交通異常検出”, 情報処理学会第75回全国大会, 5V-1, 2013. (発表予定)
- [3] Hiroto Akatsuka, Atsuhiko Takasu, Kenro Aihara, Jun Adachi, ”HIGHWAY INCIDENT DETECTION BASED ON PROBE CAR DATA”, IADIS Information Systems 2013 (IS 2013) Conference, Lisbon, Portugal, 2013. (発表予定)