

社有車テレマティクスデータに対する 時系列分析技術の応用

稲田 陽光, 小西 伸之

昨今位置センサの低コスト化に伴いデータが大量かつ安価に取得できるようになってきた。インフラが発達すると同時に、取得した大量のデータを企業活動に資する情報として活用するための分析技術も発展してきている。本稿では移動体に搭載された位置センサデータを分析し、ルート営業を対象とした効率向上やコスト削減に向けての端緒を開くことができた事例を紹介する。

キーワード：IoT、テレマティクス、時系列分析

1. はじめに

昨今 IoT 技術の発展によりさまざまなデータが安価で取得できるようになってきた。IoT で取得できるデータの多くは時系列データであるが、従来に比べはるかに稠密かつ膨大な系列であり大規模なデータとなっている。そして、これら時系列データを取り扱うための分析技術も同時に発展してきている。たとえば、製造現場では各種設備の稼働データを取得し、異常検出や歩留改善に活用されている。また、物流現場においては運搬や荷役の効率性あるいは安全性の分析が行われている。とりわけ顕著なのが自動車に代表されるテレマティクスデータの分析である。車載機の低価格化に伴い、個人・企業を問わず多数の車両の秒単位でのテレマティクスデータを安価に取得可能であり、データの収集・分析と活用を始めている企業も少なくない。車両の現在位置を可視化して管理する運行管理や、加速度・アイドリング時間などの情報から危険運転の検知や燃費の改善を目指す分析などが行われている。

これらの分析は個々の設備や車両に注目して行われているが、企業にとって制度設計や業務標準化の面で有益な知見を得るためには、より俯瞰した観点でのデータの分析が重要である。つまり個々の設備・車両のデータ系列からだけでなく、複数のデータ系列を同時に、統合的に分析を行う必要がある。

本稿ではグループ会社で保有している社有車のルート営業の走行データに対して俯瞰した観点で行った分析を紹介する。ルート営業とは以下のような行動特徴をもつ、企業における営業活動の一形態を指す（図 1）。

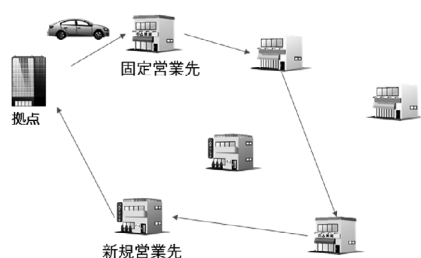


図 1 ルート営業概念図

- 拠点を出发して拠点に戻ってくるまでが営業活動の 1 単位となる
- 1 単位は定期的に訪問する固定の営業先と、新たに訪問する新規の営業先からなる
- 営業担当は固定の営業先への必要訪問頻度や、新規顧客の開拓必要性などを勘案して、日々ルートを決定する

分析の結果、個々の社有車のデータを分析するだけでは得られなかった、複数のルート営業を対象とした効率性向上やコスト削減に向けての端緒を開くことができた。本稿ではまず分析対象データの概要を述べ、分析方針の策定過程と、その方針に則った分析の中核である DTW（Dynamic Time Warping：動的時間伸縮法）[2] によるデータ類別、およびその適用に至るまでの前処理上での工夫、そして俯瞰的観点を具現化するための可視化について述べる。

2. テレマティクスデータの概観

分析対象となるテレマティクスデータの概要は以下のとおりである。

データ取得タイミング

エンジン起動／停止時

いなだ ひかる, こにし のぶゆき
キャノン IT ソリューションズ株式会社
〒550-0001 大阪府大阪市西区土佐堀 2-2-4



図2 走行データイメージ

代表データ項目

- 車両 ID
- 使用者コード
- 日時
- 緯度
- 経度
- 走行時間
- 停車時間

データの可視化イメージを図2に示す。ただし、本来のデータはエンジンの起動／停止時のみのデータから構成されるため、出発地／到着地を直線的に結んだイメージとなる。図2ではより現実的な経路に近づけるため、実際に通ったであろう経路を推定してデータに内挿している。

分析対象のテレマティクスデータはルート営業を行う社有車のある期間での連続する走行データである。連続する走行データのままで是有用な分析を行うことは難しいため、これを適切に分割することを考える。先述のとおり、ルート営業では拠点を出て拠点に戻ってくるまでを1単位とするという特徴をもつ。そのため、連続する走行データもこの1単位で分割することが妥当だと考えられる。走行データによっては、日に幾度か拠点に戻ってくるデータや、遠方に複数日にわたって走行しているデータも見られるが、これらのデータも拠点から拠点までを1単位として見ることで分割が可能となる。以降ではこの分割された走行データを「走行ルート」と呼ぶ。

3. 分析の方向づけ

3.1 方針の設定

データ分析を企業活動に活かすために最も大切なことは、適切な方針を設定することである。故障検知や安全性分析など、分析のゴールが明確である場合には

そのゴールを達成するための指標を定めればよい。しかし保有データを使って“何らかの知見を出したい”といったあいまいな要求が多いのも事実である。このような状態のまま分析に着手すると、往々にして迷走し多量の労力を投入した挙句、何も得られなかったという結果に陥りがちである。であるからといって、分析に取りかかる前に知見の在処がわかっているわけではないため、明確なステップを構成することも難しい。したがって、分析としてステップを明確に規定することなく、しかも迷走を防ぐため方針の策定が極めて重要となってくる。

本稿の分析方針として、個々の営業の動きである走行ルートから、その集合体である組織としての活動状況を俯瞰的に把握することを第一義に据えた。抽象的ではあるが、この状況把握を足掛かりにして、将来的には営業活動の効率化や制度設計の指針の獲得につなげたい。そのためには、個々の時系列データのダイナミクスを保全しつつも、個々を集合体として把握するわかりやすさも併せもつ必要がある。

3.2 相対性による集合体の把握

集合体の動きを把握する観点の一つとして、個々の動きの時系列の相対性に着目する。個体の動きが相対的に類似しているか、乖離しているかを測り、その分布をとることで、集合体としてのまとめり具合や、孤立している個体の動きなどを同時に把握したい。乖離度を測る方法としてたとえば1日の稼働時間や立ち寄り先の件数などの特徴量を集計し、そのベクトルの近接度を測る手段が考えられる。しかしながら、このように時系列方向の情報を集計すると、行動の特徴を最も表現しているであろう時間推移のダイナミクスが失われてしまう。そこで移動した個体座標の時間推移を一对で直接的に比較し、それを測度として使用する。一対比較を全個体に対して適用すると計算コストは高くなるが、比較対象を適切に選定するなどして計算コストを適切レベルに抑制することは可能である。時系列データを一対比較する手法はさまざまであるが[1]、本分析ではDTWを活用した[2]。次節ではその具体的な適用方法と前処理について述べる。

4. DTW法による分析と知見発見

4.1 DTW

DTWとは二つの時系列データ間の乖離度を評価する方法の一つである。二つのデータ列をそれぞれ $x = (x_1, x_2, \dots, x_N)$, $y = (y_1, y_2, \dots, y_M)$ とし、 $d(x_i, y_j)$ をデータ x_i と y_j の距離としたとき、DTW

距離は以下によって求められる。

```

DTW( $x, y$ ) :
 $dtw(0, 0) = 0$ 
for  $i = 1 \dots N$ 
  for  $j = 1 \dots M$ 
     $dtw(i, j) = d(x_i, y_j) + \min(dtw(i-1, j),$ 
       $dtw(i, j-1), dtw(i-1, j-1))$ 
  ただし,  $dtw(0, *) = dtw(*, 0) = \infty$ 
return  $dtw(N, M)$ 

```

分析対象の走行ルートは、エンジンの起動／停止時のデータおよび間を補間したデータからなっており、走行時間や距離によってデータ長が異なるが、DTWを用いることで長さの異なるデータ列の乖離度を算出することができる。またデータ間の距離を任意に定義することができるため、分析の用途に応じて適した距離を定義することができる。

4.2 分析前処理

本節ではこれまで述べてきた分析の方針・手法を適用するためのデータの前処理について述べる。前節では走行ルートに対してDTWを適用して分析を行う方法を述べてきたが、実際の走行データからこの走行ルートを正確に抽出することは困難である。というのも、実データでは

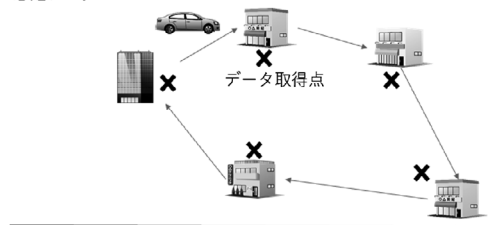
- GPSの誤差により停車位置が拠点からやや離れた位置で取得される
- 駐車場内での駐車位置変更など、ノイズとなるデータが存在する
- 拠点と離れた位置に駐車場が設置されている

などといったケースが頻発するからである（図3）。これらの課題を解決するために、以下に述べる前処理を実施している。

4.2.1 前処理：実拠点位置の自動判定

まずは実際の拠点として登録されている位置以外の拠点とみなせる位置を特定する。これはたとえば活動拠点として拠点からやや離れた位置に駐車場をもっているケースなどが該当する。これを特定するために、走行データから長時間停車しており、かつ登録されている拠点から離れている座標を抽出する。ただし、抽出した座標がすべて拠点とみなせるわけではなく、遠方への出張で宿泊を伴っているもの、自宅への車両の持ち帰りなどさまざまなケースが考えられる。可能であればこれらすべての座標について拠点とみなせるものかのチェックを行うべきであるが、候補の数が膨大なため、拠点とみなせる場所は登録されている拠点が

想定データ



実データ

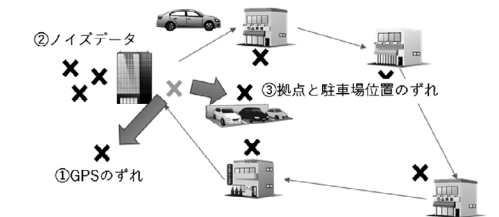


図3 分析課題

ら比較的近隣にあるという仮定を置き候補地点を絞った上でチェックを行い、拠点とみなせると判定したものを新たな拠点として登録した。

4.2.2 前処理：拠点到着判定

拠点の位置を登録できたところで、次は拠点の到着判定を行う。拠点の到着判定では登録された拠点位置から半径 R [m] の位置でエンジンを停止したものを拠点に到着したと判定する。この R の値はこの到着判定により分割された各走行ルートを確認して妥当性を判断する。拠点付近での長時間停車がある走行ルートが多ければ、本来分割されるべきルートが分割されていないとみなせるため R の値を大きくし、逆に拠点判定円内を移動しているような走行ルートが多ければ、本来分割されるべきではないルートとみなせるため、 R の値を小さくする。結果、各拠点で R の値は 100～300 [m] 程度の値となった。

4.2.3 前処理：ノイズデータの除去

最後に、作成した走行ルートの中からノイズとなる走行ルートの削除を行う。ノイズとみなした走行ルートは、拠点内を移動するような走行ルートである。この走行ルートはたとえば拠点の駐車場内の移動のような、客先で営業を行っていない走行ルートと考えられるため、分析の対象外とした。

4.3 分析結果

このようにして得られた走行ルートデータに対してDTWを適用する。距離を算出する二つのデータ列内のデータ、

$$x_i = \{lat_i^x, lon_i^x, t_i^x\}, y_j = \{lat_j^y, lon_j^y, t_j^y\}$$

(lat は緯度, lon は経度, t は時刻)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
A	0	85.5	86.6	384.9	267.5	62.5	103.4	231.7	640.2	165.1	72.7	28.7	409.1	366.8	402.5
B	85.5	0	70.9	240.9	94.9	28.4	63	178	598.8	115.9	47.8	114.6	366.9	294	264.9
C	86.6	70.9	0	247.5	171.3	82.6	109.2	257	572	121.5	96.3	77.7	343.7	230.5	237.8
D	384.9	240.9	247.5	0	219.7	315.8	340.5	273.7	735.8	362.6	354.9	416.7	500.4	408.7	96.6
E	267.5	94.9	171.3	219.7	0	174	204.8	216.7	685.8	265.3	223	298.5	417.1	341.2	292.7
F	62.5	28.4	82.6	315.8	174	0	40	231.1	597.1	106.8	27.7	86.9	324.9	289	323
G	103.4	63	109.2	340.5	204.8	40	0	235.3	570.8	123.2	33.8	122.6	308.3	278.4	341.4
H	231.7	178	257	273.7	216.7	231.1	235.3	0	790.1	319.3	238.4	256.5	534.4	466.7	344.9
I	640.2	598.8	572	735.8	685.8	597.1	570.8	790.1	0	370.5	537	638.7	446.5	500.5	681.1
J	165.1	115.9	121.5	362.6	265.3	106.8	123.2	319.3	370.5	0	90.3	178.6	118.3	181.7	341.7
K	72.7	47.8	96.3	354.9	223	27.7	33.8	238.4	537	90.3	0	90.9	278.2	271.9	359
L	28.7	114.6	77.7	416.7	298.5	86.9	122.6	256.5	638.7	178.6	90.9	0	410.1	375.5	435.5
M	409.1	366.9	343.7	500.4	417.1	324.9	308.3	534.4	446.5	118.3	278.2	410.1	0	167.4	455.3
N	366.8	294	230.5	408.7	341.2	289	278.4	466.7	500.5	181.7	271.9	375.5	167.4	0	350
O	402.5	264.9	237.8	96.6	292.7	323	341.4	344.9	681.1	341.7	359	435.5	455.3	350	0

図4 DTW 距離行列

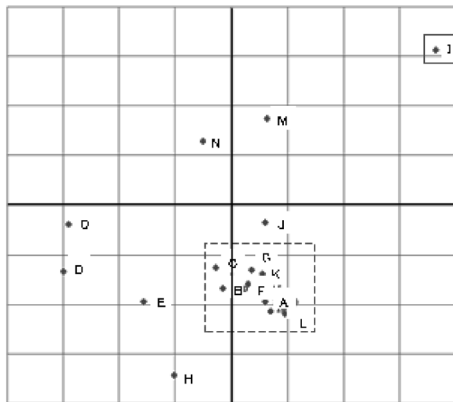


図5 走行ルート可視化

の距離は以下と定義した。

$$d(x_i, y_j) = \text{dist}(\text{lat}_i^x, \text{lon}_i^x, \text{lat}_j^y, \text{lon}_j^y) + f(t_i^x, t_j^y)$$

dist : 緯度経度から求まる実距離

f : 時間幅を距離換算する関数

f は分析の目的に応じて定義する。たとえば同一座標における時刻の差異を評価に組み込みたい場合、時刻の差異を標準的な速度で乗算し、距離の差異とするなどが考えられる。

本分析では、同一日以内での営業活動のダイナミクスを把握することを目的としたため、同一日以内であれば0、日が異なれば ∞ と変換する関数とした。DTWを適用して得られたDTW 距離行列を図4に示す。図中の行・列はそれぞれ一つの走行ルートを表している。各マスの濃淡は走行ルート間の乖離度を表し、色が濃いほど乖離度が高いことを表している。

4.4 分析結果の解釈

4.4.1 距離関係の可視化

図4により各時系列データの相対的な関連が得られた。さらにデータ全体を俯瞰するため、これらの距離関係を可視化した。可視化には多次元尺度構成法(MDS)を用いた。可視化結果を図5に示す。

このようにデータを可視化し俯瞰することで、デー

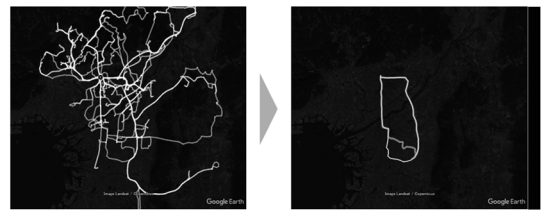


図6 DTW 法による類似走行ルートの抽出

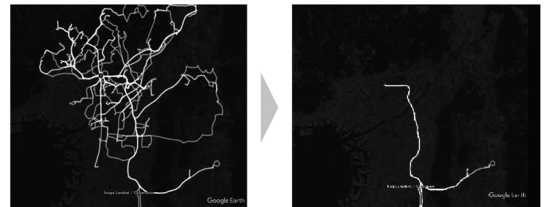


図7 孤立ルート例

タ同士のまとまりや孤立度合いといった、対象とする時系列データの集合としての特徴を捉えることができる。本稿では特徴を捉え現状を把握することに留まるが、たとえば複数のデータ集合の分布を比較することでデータ集合間の差異を見つけ出し、営業活動の効率化や制度設計の指針を得ることが可能となる。

以下ではデータを俯瞰して把握した傾向から、さらに掘り下げて個別の動態の挙動の確認に至るまでの流れについて二つ例を説明する。

4.4.2 類似ルートの発見

図5を参照すると、破線枠はルート群がクラスタを形成していることがわかる。このクラスタから二つの走行ルートを取り出して可視化したものを図6に示す。図6からは二つの走行ルートが南の一部を除いて同一の経路を走行している様子が見て取れるが、さらに走行時間についてもほぼ同一時刻で並走している。

日々このような類似ルートの抽出が可能になれば、それを継続的に観測する。そして、同一地域に類似ルートが繰り返し発生しているのであれば、そこは同行により効率を上げることが可能となる。さらにもう一歩踏み込んで、このようなルート群を取り出し、各ルートがもつ属性値(訪問内容等)を参照することによって、営業活動の括りである部門分けや担当品目の再編などを検討することも可能となる。

4.4.3 孤立ルートの発見

前節では類似した走行ルートについて述べたが、本節では図5の実線枠内のように、ほかのすべての走行ルートから距離のある、孤立した走行ルートに着目する。このルートを取り出し可視化したものが図7であ

る。これを見ると、ほかの走行ルートに比べ、顕著に遠距離でかつ孤立した地域の活動を行っていることがわかる。このように特定ルートを抽出したうえで個別に訪問先を確認すると、遠方かつ孤立地域ではあるが重要顧客であるため今後も継続した訪問が必要である場合が少なくない。顧客密着型が営業の強みというわけである。しかしながら昨今の働き方改革という視点から、長時間にわたる車両拘束を抑制する措置も必要とされる。このバランス達成のためには代替交通機関の検討や、別拠点への移管など個別事情に本目細かに配慮した効率化施策が必要である。さらに発展的に捉えるならば新たに営業拠点を設置する場合、立地の検討に勘案することも可能である。

5. おわりに

時系列データ分析の技術は目覚ましい勢いで発展しており、利益に直結する販売データについてはレコメンドや売上予測など多岐に活用されている。一方でヒトやモノといった経営資源が物理的に移動した実績については、データこそ集まっているものの、その活用についてはまだ限定的であり、さまざまな企業で試行錯誤が繰り返されている。本稿で紹介した内容もこの

ような組織としての試行錯誤の一例である。目的や指標が明確でない状況から問題点や知見を発掘するには、全体から部分への掘り下げを適切に行っていくことが鍵となる。そのためには俯瞰像を利用部門と共有しつつ、掘り下げる方向を都度修正し、再度共有するというサイクルが必要である。このサイクルを確立できれば、それを利用部門単体で継続運用する仕組みの構築もおのずと可能になる。本事例は社有車の移動分析について端緒を開いたに過ぎないが、その知見についての広い発展性に触れることができた。今後はこの視点をさらに広げ、製造・物流の現場である構内・構外の移動体、すなわち工場や倉庫内の作業員の人流およびトラック貨物などのテレマティクスデータ分析に拡大して、多様な分析手法の開発に挑戦しつつ、技術発展に寄与していきたい。

参考文献

- [1] P. Montero and J. A. Vilar, “TSclust: An R package for time series clustering,” *Journal of Statistical Software*, **62**, Issue 1, 2014. <https://www.jstatsoft.org/article/view/v062i01> (2019 年 11 月 2 日閲覧)
- [2] D. J. Bemdt and J. Clifford, “Using dynamic time warping to find patterns in time series,” AAAI Technical Report WS-94-03, 1994.