

事例研究 [論文]

時間付加グラフのベクトル表現を用いた タクシー軌跡データの解析

河上 佳太, 西村 直樹, 白鳥 友風, 工藤 晃太,
松岡 雄大, 最首 大輝, 渡邊 彰久, 高野 祐一

1. はじめに

タクシー事業では営業効率の低さが課題となっている。タクシー運転手の月間労働時間は全産業平均と比較して約 11%長いが、年間賃金水準は全産業平均と比較して約 36%低い [1]。また本研究の分析データ（詳細は 2 節を参照）でも、タクシー運転手の業務時間に対する実車時間の割合を表す実車率は平均 34.6%となっており、業務時間の半分以上が空車状態である。実車率の向上によって営業効率を改善することで、労働時間や賃金も改善していくと考えられる。

本研究では GPS タクシー軌跡データを用いて、タクシー運転手の行動パターンと営業効率の関係性を分析する。運転手の行動パターンはタクシーの移動軌跡、すなわち地図上の地点（頂点）と地点間の移動（枝）により構成されるグラフとして表現できる。また上述のように、実車率は運転手の営業効率を表す重要な指標である。したがって本研究の目的は、タクシー経路を表すグラフの特徴を抽出し、実車率向上に寄与する行動パターンを示すことと換言できる。

タクシー経路を分析する研究には、営業効率改善のための施策を提言する研究 [2-5] や、タクシー需要を予測して最適な経路を推薦する研究 [6, 7] などがある。タクシー経路の定量化手法は多岐にわたる。Ma et al.

[5] はタクシーが各区画を通った回数を要素にもつ疎行列によってタクシー経路を表現し、この行列の主成分スコアを用いてタクシー経路を分析した。Yao et al. [6] はタクシー需要のヒートマップを畳み込みニューラルネットワークにより定量化した。Huang et al. [7] は Ma et al. [5] と同様の疎行列を、スパース自己符号化器などの特徴埋め込み手法によりベクトル化した。

タクシー経路のようなグラフを分析する際に、グラフの特徴を表す数値ベクトルを生成し、機械学習手法を適用する方法が近年注目を集めている [8]。これまでに各頂点のベクトル化手法 [9] やグラフ全体のベクトル化手法 [10-12] が提案されている。その中でも graph2vec [12] はグラフのベクトル表現を生成する手法として、グラフのクラス分類やクラスター分析において従来手法よりも高い精度を示したことが報告されている。

本研究では、グラフの高性能ベクトル化手法 graph2vec [12] を利用して、GPS タクシー軌跡データを分析する。売上が高い運転手は、時間帯ごとに行動を変えていることが報告されている [2]。しかしながら著者らの知る限り、時間帯を考慮してタクシー経路を定量化した研究は存在しない。そこで本研究では、各頂点に時間帯の情報を付与した時間付加グラフのベクトル表現を提案する。実車率を目的変数とした回帰分析と、タクシー経路のクラスター分析を実施し、その結果から営業効率改善のための施策を提言する。

時間付加グラフを利用することで、実車率の回帰モデルの予測精度が向上した。また時間付加グラフを用いた回帰分析とクラスター分析の結果から、営業効率改善に寄与する行動パターンを獲得することができた。

本論文は以下のように構成される。2 節では分析データの概要を説明する。3 節ではタクシー経路のベクトル化手法を説明する。4 節では実車率の回帰分析の結果を示し、5 節ではタクシー経路のクラスター分析の結果を示す。最後に 6 節でまとめと今後の課題を述べる。

かわかみ けいた, にしむら なおき, しらとり ともかぜ,
くどう こうた, まつおか ゆうた
筑波大学大学院システム情報工学研究科
〒 305-8573 茨城県つくば市天王台 1-1-1
さいしゅ ひろき
筑波大学大学院システム情報工学研究群
〒 305-8573 茨城県つくば市天王台 1-1-1
わたなべ あきひさ
東京工業大学工学院経営工学系
〒 152-8552 東京都目黒区大岡山 2-12-1
たかの ゆういち
筑波大学システム情報系
〒 305-8573 茨城県つくば市天王台 1-1-1
受付 20.7.25 採択 21.1.12

2. 分析データ

本研究では、経営科学系研究部会連合協議会主催、令和元年度データ解析コンペティションで提供されたデータを利用した。2016年4月1日から2018年3月31日までの2年間のGPSタクシー軌跡データが提供され、会社IDや運転手IDと共に、平均して41.3秒の間隔で日時、車両の状態（実車、空車、回送など）、緯度・経度などの情報が記録されている。なお、提供データにはタクシー会社7社の記録が含まれているが、本研究では途中で記録が途絶えた会社や記録時間の間隔が長い会社を除いた3社（運転手89人）を分析対象とした。

本論文では以下の用語を使用する。

実車 乗客から運賃が支払われている状態とする。

分析データにおいては、「実車」「賃走」「割増」「貸切」「迎車」「支払」の6状態が対応する。

空車 ただちに乗客を獲得し、「実車」へ移行できる状態とする。分析データにおいては「空車」の状態のみが対応し、「回送」「休憩」「閉局」は含まない。

1乗車 各運転手について、ある乗客が乗車してから次の乗客が乗車するまでの区間とする。実車の状態だけでなく空車や回送の状態も含まれる。

1勤務 各運転手について、記録間隔が3時間未満で連続している区間とする。すなわち記録間隔が3時間以上空いた場合に別日の勤務とみなす。

実車率 タクシーの営業効率を表す指標とし、以下のように定義する。

$$\text{実車率} := \frac{1 \text{ 勤務における実車時間}}{1 \text{ 勤務における総勤務時間}}$$

データの前処理として、4時間以下または24時間以上の勤務は、タクシー業の一般的な勤務形態には該当しないためデータから削除した。また上述の平均記録間隔（41.3秒）を考慮すると、平均記録間隔が100秒以上となる勤務は情報が不正確であると考えてデータから削除した。

3. タクシー経路のベクトル表現

本節ではタクシー経路のグラフを作成し、そのベクトル表現を生成する手法について述べる。さらに時間帯の情報を付与したグラフのベクトル表現を提案する。

3.1 タクシー経路グラフ

実車状態の行動は乗客の目的地に依存するため、運転手の乗客獲得能力とはほとんど関連がない。したがっ

て本研究では、空車状態における運転手の行動パターンに注目する。

タクシー経路グラフは各運転手の各勤務に対して定義する。1週間を単位として運転手の日常的な行動パターンを分析するために、勤務日の前の連続する100乗車分の空車時のタクシー経路をまとめてグラフ化する。1勤務あたり平均約29乗車であるため、100乗車はおおむね1週間（3～4勤務）に相当する。

タクシー経路グラフは、連続する記録時刻における地域メッシュの対を枝集合とした単純無向グラフとする。本研究では、worldmeshライブラリ¹を用いて地域メッシュを生成した。これは地図上の正方領域に対して一意に割り当てられた識別子であり、緯度・経度から求められる[13]。

単純無向グラフでは、枝の向き・多重枝・自己閉路が存在せず、タクシーの移動を正確に記述するという観点では情報の損失がある。しかしながら本研究では、「各時間帯にどの場所に移動すべきか」といった大局的な行動パターンの抽出を目的としており、単純無向グラフを用いてこの目的を達成する。

3.2 graph2vecの概要

本研究ではgraph2vec[12]を利用して、タクシー経路のベクトル表現を学習する。graph2vecでは、共通する部分グラフを多く含むようなグラフは類似していると考え、類似したグラフには類似した（距離の近い）ベクトルを与える。

グラフの集合 $\{g_i \mid i \in N\}$ に対して、部分グラフの集合を

$$S := \{s \mid s \in \text{sub}(g_i), i \in N\}$$

とする。ここで $\text{sub}(g_i)$ はグラフ g_i の各頂点を根とする根付き部分グラフの集合とする。グラフ g_i に対して i 番目の成分を1、それ以外の成分を0とするベクトル $\mathbf{x}(g_i) \in \{0, 1\}^{|N|}$ を定義する。部分グラフ $s \in S$ に対しても同様に、ベクトル $\mathbf{y}(s) \in \{0, 1\}^{|S|}$ を定義する。

図1に示すように、graph2vecでは2層のニューラルネットワークを考え、中間層のユニット数は d とする。このネットワークに対して、説明変数 $\mathbf{x}(g_i)$ と目的変数 $\mathbf{y}(s)$ のペアの集合

$$\{(\mathbf{x}(g_i), \mathbf{y}(s)) \mid s \in \text{sub}(g_i), i \in N\}$$

¹ 科学技術振興機構さがけ「グローバル・システムの持続可能性評価基盤に関する研究」の研究成果であり、世界メッシュ研究所により提供されている。

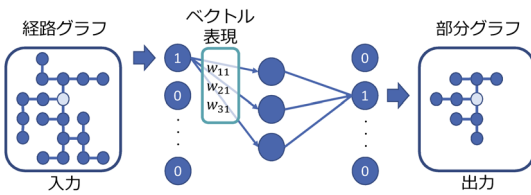


図 1 graph2vec の図解 ($d = 3$)

を与えて学習する。類似するグラフ（入力）には共通する部分グラフ（出力）が多く含まれるため、類似するグラフの入力層から中間層への枝重みは互いに近づく。したがって、入力層の i 番目のユニットから中間層への枝重みを成分とするベクトル

$$w(g_i) := (w_{1i}, w_{2i}, \dots, w_{di}) \in \mathbb{R}^d$$

が、グラフ g_i の d 次元のベクトル表現として得られる (図 1)。

本研究では、公開ソースコード²を改変して graph2vec を実装した。ベクトル表現の次元数は 16, 32, 64, 128 などの設定を試し、回帰分析の決定係数が最も向上した次元数 $d = 64$ を採用した。

3.3 頂点ラベル

根付き部分グラフの集合を出力する *sub* 関数の計算には、グラフの同型性を判定するアルゴリズム [10] を利用する。このアルゴリズムでは頂点を指定し、そこからグラフ上で隣接する頂点のラベルを結合してハッシュ化する操作を繰り返して、指定された頂点を根とする部分グラフに識別子を与える。各頂点のラベルとして、文献 [10, 12] では頂点の次数が使用されており、このラベルを単に「次数ラベル」と呼ぶ。

タクシー経路グラフの場合には、頂点の位置情報を分析に反映させるために、タクシーが通過した地域メッシュを頂点のラベルにする方法が考えられる。このラベルを「地域ラベル」とする (図 2)。地域ラベルを用いた場合の図 2 の経路グラフは枝集合

$$\{(C, A), (A, B), (B, D)\}$$

として定義する。

売上が高い運転手は、時間帯ごとに行動を変えていることが報告されている [2]。時間帯ごとの全体的な行動パターンの違いを調べるために、空車時のタクシーの位置分布 (各地域メッシュにおける記録回数) を変数として、24 時間を分割するクラスター分析を行った。その結果を参考にして、朝 : 6 時～8 時、午前 : 8 時～

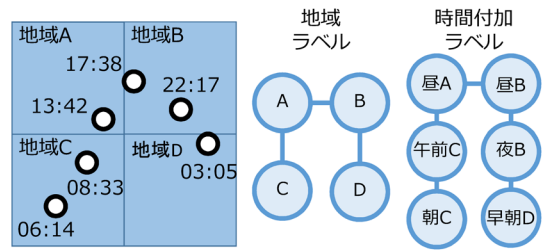


図 2 地域ラベルと時間付加ラベルの例

13 時、昼 : 13 時～21 時、夜 : 21 時～翌 3 時、早朝 : 3 時～6 時の五つの時間帯を定義した。

本研究では、各頂点に上記の時間帯の情報を付与した時間付加グラフのベクトル表現を提案する。タクシーが通過した時間帯と地域メッシュのペアを「時間付加ラベル」とする (図 2)。時間付加ラベルを用いた場合の図 2 の経路グラフは枝集合

$$\{(\text{朝 } C, \text{午前 } C), (\text{午前 } C, \text{昼 } A), (\text{昼 } A, \text{昼 } B), (\text{昼 } B, \text{夜 } B), (\text{夜 } B, \text{早朝 } D)\}$$

として定義する。

3.2 節で説明したように、経路グラフ g_i は説明変数 $x(g_i) \in \{0, 1\}^{|N|}$ の形式で graph2vec に入力し、64 次元のベクトルを生成する。

4. 実車率の回帰分析

本節では、勤務日とタクシー経路の情報を説明変数として、タクシーの実車率を予測する回帰分析の結果を説明する。

4.1 回帰分析の設定

L1 正則化線形回帰モデル (Lasso) [14] を利用し、1 週間を単位として営業効率を分析するために、各運転手の連続する 4 勤務日における平均実車率を予測する。平均実車率の値域は区間 $[0, 1]$ であり、線形回帰モデルの予測値の値域と整合させるために、平均実車率をロジット変換 [15] した値を目的変数とした。

説明変数として「勤務日の情報」と「タクシー経路ベクトル」を使用する。各説明変数は標準化した値を使用した。勤務日の情報は以下の 23 種類の変数とする。

総勤務時間 (1 種類) 勤務 4 回の総勤務時間

各月の勤務回数 (12 種類) 勤務 4 回のうち X 月に勤務した回数 ($X \in \{1, 2, \dots, 12\}$)

各曜日の勤務回数 (7 種類) 勤務 4 回のうち X 曜日に勤務した回数 ($X \in \{\text{日}, \text{月}, \dots, \text{土}\}$)

各勤務形態での勤務回数 (3 種類) 勤務 4 回のうち勤務形態 X で勤務した回数 ($X \in \{\text{日勤}, \text{夜}$

² <https://github.com/benedekrozemberczki/graph2vec>

勤, 隔日勤務 })

ただし, 日勤 (平均: 7 時~17 時), 夜勤 (平均: 17 時~翌 6 時), 隔日勤務 (平均: 8 時~翌 5 時, その後翌日まで休み) の 3 種類の勤務形態は, 勤務時間帯 (1 時間ごとのダミー変数) に基づくクラスター分析により勤務を分類して構成した.

タクシー経路ベクトルは, 予測対象となる 4 勤務日より前の勤務日における 100 乗車の空車時経路のベクトル表現とする. 64 次元のベクトルを生成し, 各成分を $w_1, w_2, \dots, w_{64}$ とする. 頂点ラベルとして, 「次数ラベル」「地域ラベル」「時間付加ラベル」を使用し, 性能を比較する.

4.2 予測精度の検証

2016 年度を訓練データ (総数: 6,927), 2017 年度をテストデータ (総数: 6,312) として使用し, 予測精度を検証する. 訓練データを用いて線形回帰モデルを学習し, 予測精度の評価指標として決定係数

$$R^2 := 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

を使用した. ここで, y_i と $\hat{y}_i$ は事例 i の目的変数の (ロジット変換後の) 実測値と予測値とし, $\bar{y}$ は実測値の平均とする. L1 正則化項の重みは訓練データの 10 分割交差確認により求めた.

地域メッシュの 1 辺の長さに対する線形回帰モデルの予測精度を表 1 に示す. ここでは頂点ラベルとして時間付加ラベルを使用している. 地域メッシュの 1 辺の長さが 125 m の場合にテストデータにおいて決定係数が最も高くなった. 次に地域メッシュの 1 辺の長さを 125 m とし, 各種の説明変数に対する線形回帰モデルの予測精度を表 2 に示す. 勤務日の情報に加えてタクシー経路ベクトルを使用することで, 予測精度が大きく向上した. また, 頂点ラベルとして時間付加ラベルを使用した場合に, テストデータにおいて決定係数が最も高くなった.

次数ラベルではタクシーの位置情報は無視されるが, 経路グラフの構造から営業形態 (後述する付け待ち, 流し, 周回など) を識別することができ, ある程度の予測精度を得ることができたと考えられる. 本研究で提案した時間付加ラベルは, タクシーの位置情報を考慮できることに加えて, 異なる時間帯における同じ地点を別の頂点とみなすことできる. それゆえ地域ラベルを用いた場合と比較して, 線形回帰モデルの予測精度が向上したと考えられる.

表 1 回帰分析の予測精度

(説明変数: 勤務日の情報+時間付加ラベル)

地域メッシュの 1 辺の長さ	決定係数	
	訓練	テスト
1,000m	0.552	0.477
500m	0.552	0.481
250m	0.559	0.512
125m	0.569	0.523
100m	0.584	0.516

表 2 回帰分析の予測精度

(地域メッシュの 1 辺の長さ: 125 m)

説明変数 種類	総数	決定係数	
		訓練	テスト
勤務日の情報のみ	23	0.035	0.004
勤務日の情報+次数ラベル	87	0.488	0.352
勤務日の情報+地域ラベル	87	0.584	0.518
勤務日の情報+時間付加ラベル	87	0.569	0.523

4.3 説明変数の考察

時間付加ラベルを用いた L1 正則化線形回帰モデルによって選択された, 説明変数の偏回帰係数と解釈を表 3 に示す. ここでは正則化項の重みを大きくして, 6 個の説明変数を選択した. 選択された説明変数はすべてタクシー経路ベクトルの成分であり, これらに対して (正則化項を含まない) 最小二乗法により偏回帰係数を推定した. ただし解釈を容易にするため, 目的変数と負の相関をもつ説明変数は符号を反転させた. 選択された 6 個の説明変数はいずれも 0.1% 水準で有意であった.

解釈は時間帯ごとの活動拠点や営業形態に基づく. ここで付け待ち営業とは, 駅や空港などのタクシー乗り場で乗客を待つ行動とする. 流し営業とは街中を走りながら乗客を探す行動とし, その中でも同じルートを何度も通るような行動は周回営業とする.

選択された 6 個の説明変数 (表 3) の値が大きい経路と小さい経路を図 3, 4 に示す. 各行が説明変数の値が上位・下位 0.5% (65 個) の経路に対応し, 各列が時間帯 (朝/午前/昼/夜/早朝) に対応する. 東京都で乗車人員の多い主要 6 駅 (新宿駅, 池袋駅, 東京駅, 品川駅, 渋谷駅, 新橋駅) と羽田空港の位置も記載した. 各点の大きさはその地点におけるデータの記録回数を表し, 大きな円の中心は付け待ち営業の拠点を示す.

説明変数 w_3 の値が大きい経路 (図 3: 1 行目) では, 午前・昼に東京駅の北側を中心に流し営業をしている. 昼は池袋駅, 夜は東京中心部 (主に山手線内側) の複数拠点で付け待ち営業をしており, 早朝には付け

表 3 選択された説明変数（タクシー経路ベクトルの成分）

説明 変数	偏帰帰 係数	解釈 (4.3 節)	
		変数の値が大きい（実車率が高い）経路	変数の値が小さい（実車率が低い）経路
w_3	0.059	午前・昼に東京駅北側を中心に流し営業	時間帯によって拠点を変えながら付け待ち営業
w_7	0.066	午前から夜に東京中心部で流し営業	午前・昼に世田谷区・目黒区付近で周回・付け待ち営業
w_{35}	0.066	午前・昼に渋谷駅付近で流し営業	全時間帯に東京駅周辺で付け待ち営業
w_{41}	0.060	午前から夜に東京の西から東へ流し営業	午前から夜に東京中心部から離れた地域で付け待ち営業
w_{45}	0.051	午前から夜に渋谷駅・品川駅周辺で流し営業	午前から早朝に東京駅の周辺や北東側で付け待ち営業
w_{62}	0.060	東京中心部で昼は流し営業、夜は付け待ち営業	午前・昼に東京駅周辺で付け待ち営業

待ち拠点が池袋駅付近と新橋駅付近に集中する。説明変数 w_3 の値が小さい経路（図 3：2 行目）では、時間帯によって拠点を変えながら付け待ち営業をしている。午前・昼には東京駅の東側、夜には新橋駅付近が拠点となっている。朝・早朝には複数拠点を付け待ち営業をしている。

説明変数 w_7 の値が大きい経路（図 3：3 行目）では、午前から夜に東京中心部で流し営業をしている。夜は東京中心部や品川駅で付け待ち営業もみられる。説明変数 w_7 の値が小さい経路（図 3：4 行目）では、午前・昼に世田谷区・目黒区付近で周回営業や付け待ち営業をしている。夜は複数拠点での付け待ち営業に加えて東京中心部でも活動しており、早朝は東京中心部で付け待ち営業が増える。

説明変数 w_{35} の値が大きい経路（図 3：5 行目）では、午前・昼に渋谷駅付近で流し営業をしている。夜は東京中心部で付け待ち営業をしていて、早朝には拠点が新橋駅付近に移動する。説明変数 w_{35} の値が小さい経路（図 3：6 行目）では、全時間帯で付け待ち営業が非常に多い。朝・午前は東京駅の北側に複数の付け待ち拠点が、昼・夜は東京駅付近の付け待ち営業に活動が集中する。

説明変数 w_{41} の値が大きい経路（図 4：1 行目）では、午前から夜に東京中心部を西から東へ横断するように流し営業をしている。実際に午前・昼は渋谷駅付近での記録回数が多く、夜は東京駅付近での記録回数が多。説明変数 w_{41} の値が小さい経路（図 4：2 行目）では、午前から夜に東京中心部から離れた複数拠点で付け待ち営業をしている。

説明変数 w_{45} の値が大きい経路（図 4：3 行目）では、午前から夜に渋谷駅・品川駅周辺で流し営業をしており、夜は東京中心部で付け待ち営業もみられる。説明変数 w_{45} の値が小さい経路（図 4：4 行目）では、午前から早朝に東京駅の周辺や北東側で付け待ち営業をしている。

説明変数 w_{62} の値が大きい経路（図 4：5 行目）で

は、午前・昼は東京中心部で流し営業や羽田空港で付け待ち営業をしており、夜・早朝は東京中心部で付け待ち営業が増える。説明変数 w_{62} の値が小さい経路（図 4：6 行目）では、午前・昼に東京駅周辺で付け待ち営業をしているが、一方で夜・早朝には付け待ち営業が少なくなる。

以上をまとめると、実車率向上に寄与する行動として、

- 午前から夜に東京中心部で流し営業
- 午前から夜に東京中心部を西から東へ横断するように流し営業
- 夜・早朝に東京中心部で付け待ち営業

などが挙げられる。一方で、実車率を低下させる行動として、

- 東京中心部から離れた地域で周回営業や付け待ち営業
- 午前・昼に東京駅周辺で付け待ち営業

などが挙げられる。

5. タクシー経路のクラスター分析

本節では、タクシー経路のクラスター分析の結果を説明し、各クラスターの実車率の差や特徴的な経路について考察する。

5.1 クラスター分析の結果

タクシー経路（総数：13,239）は、時間付加ラベルを用いて生成した 64 次元のベクトルとして表現されている。平方ユークリッド距離に基づくワード法を利用して、これらのタクシー経路を分類する。

図 5 はクラスター分析による樹形図である。この樹形図に基づいて、タクシー経路を 6 種類のクラスターに分割した。

各クラスターに属する経路の時間帯ごとの中心座標の分布を図 6 に示す。各行がクラスター ID、各列が時間帯（朝／午前／昼／夜／早朝）に対応し、東京都で乗車人員の多い主要 6 駅と羽田空港の位置を記載した。クラスターの分割では 64 変数をすべて使用したが、ここでは実車率への影響が大きい 6 個の説明変数

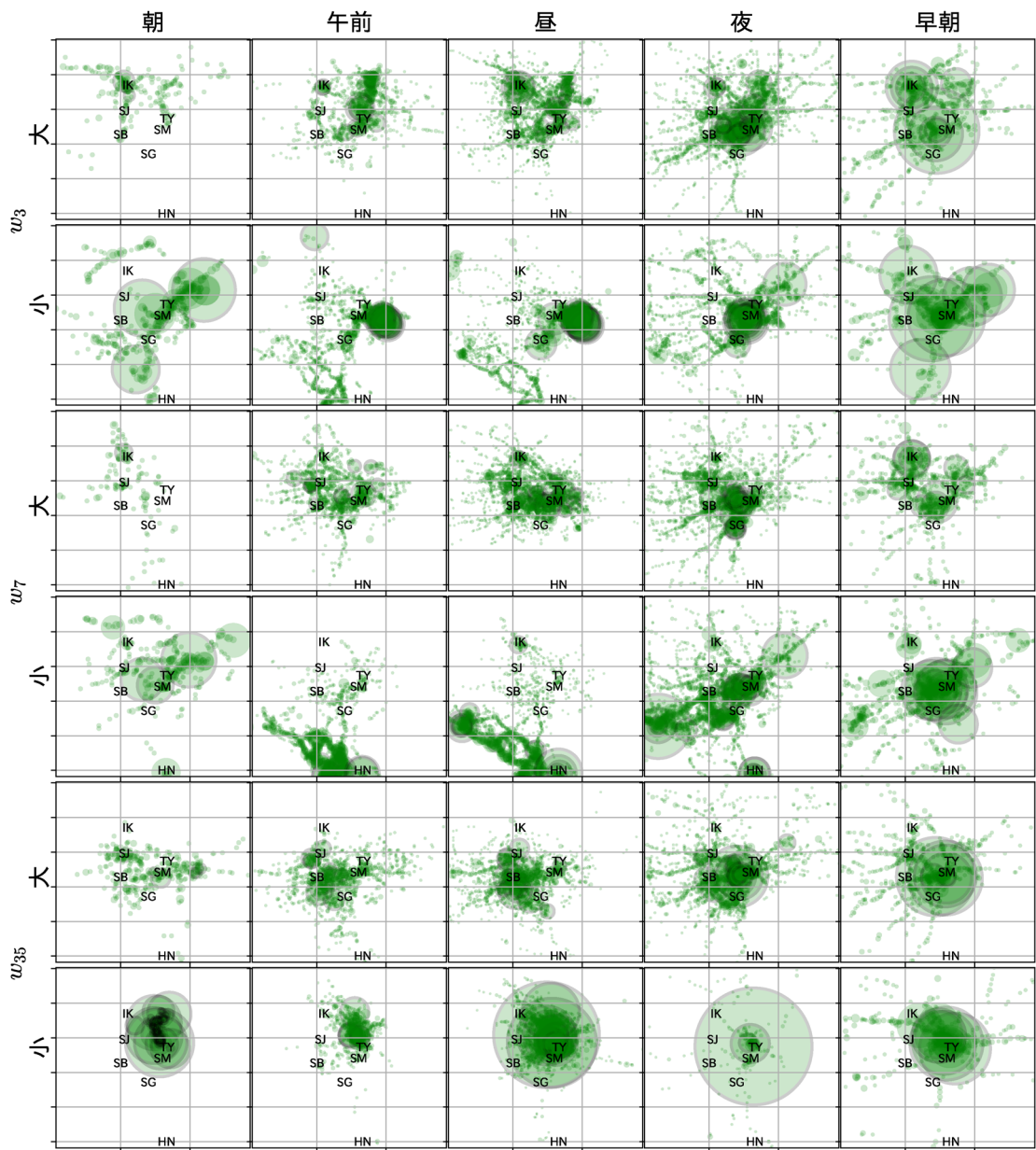


図3 選択された説明変数 (w_3, w_7, w_{35}) の値が大きい経路と小さい経路
(SJ : 新宿駅, IK : 池袋駅, TY : 東京駅, SG : 品川駅, SB : 渋谷駅, SM : 新橋駅, HN : 羽田空港)

(表3)に着目し、各クラスターの平均値を表4に示す。

タクシー経路のクラスターの概要を表5に示す。「実車率[%]」には、各クラスターに属する経路に対応する4勤務の実車率の平均と標準偏差を記載している。なお全経路の実車率の平均は34.6%、標準偏差は7.4%であった。次節では以上の結果に基づいて、各クラスターの解釈について説明する。

5.2 クラスターの解釈

クラスター1(図6:1行目)では、経路の中心座標

は朝は広範囲に分布しているが、昼にかけて東京中心部へまとまり、早朝にかけて再び広がる。午前から夜には中心座標の分布が羽田空港へと連なる。説明変数 w_3, w_{41} の値が最も大きいことから、主に流し営業をしていると考えられる(表3, 4)。

クラスター2(図6:2行目)は渋谷駅の南西側で主に活動している。説明変数 w_7 の値が最も小さいことから、主に世田谷区・目黒区付近で周回営業や付け待ち営業をしていると考えられる(表3, 4)。

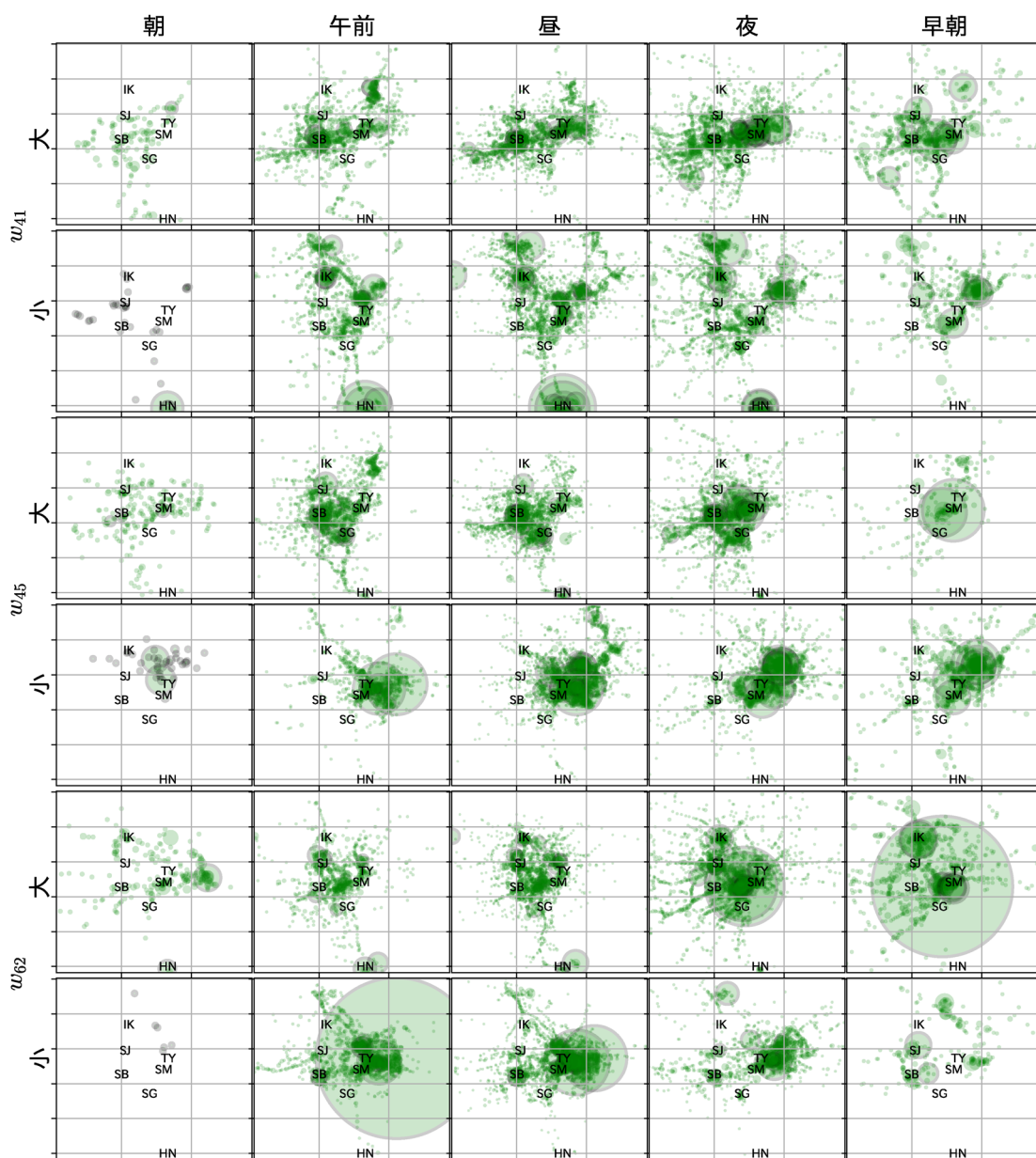


図4 選択された説明変数 (w_{41}, w_{45}, w_{62}) の値が大きい経路と小さい経路
(SJ: 新宿駅, IK: 池袋駅, TY: 東京駅, SG: 品川駅, SB: 渋谷駅, SM: 新橋駅, HN: 羽田空港)

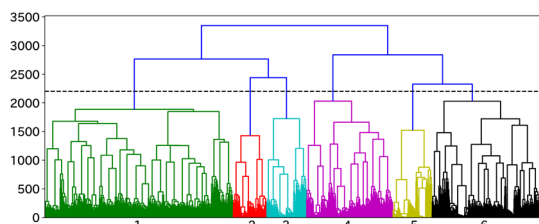


図5 タクシー経路の樹形図

クラスター3(図6:3行目)では、経路の中心座標は朝は東京中心部に広く分布しているが、以降の時間帯では主に品川駅付近を中心として分布が小さくまとまる。説明変数 w_3 の値が最も小さく、全時間帯での付け待ち営業が示唆される。一方、説明変数 w_{35}, w_{45}, w_{62} の値が最も大きいことから、午前から昼に流し営業をしていると考えられる(表3, 4)。

クラスター4(図6:4行目)は東京駅の周辺や北側で主に活動している。説明変数 w_{45}, w_{62} の値が最も

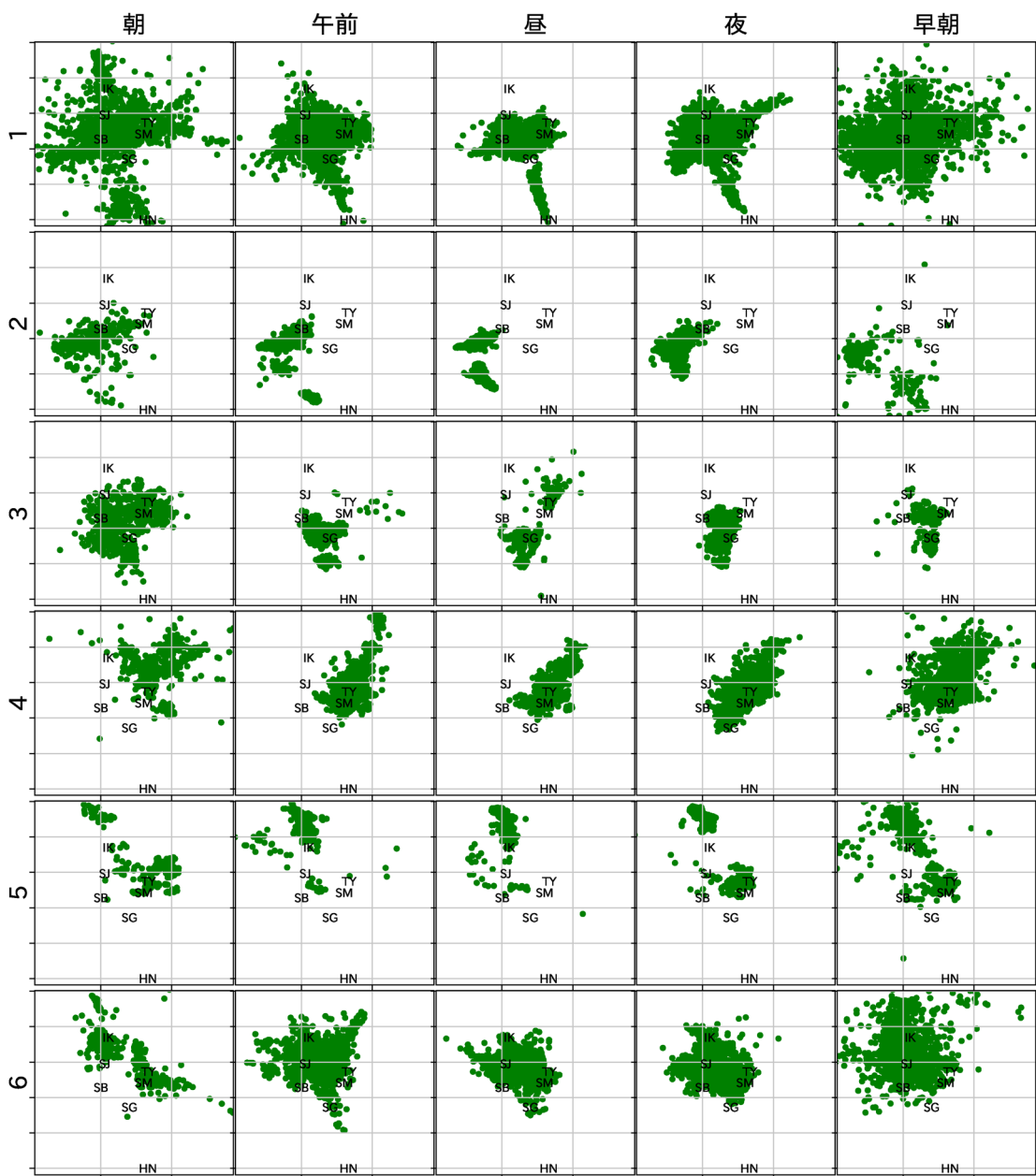


図6 各クラスター・各時間帯における経路の中心座標の分布
(SJ：新宿駅, IK：池袋駅, TY：東京駅, SG：品川駅, SB：渋谷駅, SM：新橋駅, HN：羽田空港)

小さいことから、主に付け待ち営業をしていると考えられる(表3, 4)。

クラスター5(図6:5行目)は池袋駅の北側で主に活動しており、夜から朝には東京駅の周辺でも活動している。説明変数 w_{35}, w_{41} の値が最も小さいことから、主に付け待ち営業をしていると考えられる(表3, 4)。

クラスター6(図6:6行目)では、朝は池袋駅と東京駅を中心に分布が2群に分かれているが、昼にかけて東京中心部に分布がまとまり、早朝にかけて池袋駅

周辺へ分布が広がる。説明変数 w_3, w_7 の値が最も大きいことから、主に流し営業をしていると考えられる(表3, 4)。

クラスター1とクラスター6の実車率が比較的高く(表5)、それらの分析結果を踏まえると、以下のような行動が実車率向上に寄与すると考えられる

早朝・朝 東京中心部と郊外との間を長距離輸送

午前から夜 東京中心部で流し営業や、東京中心部と羽田空港との間の輸送

表 4 各クラスターの説明変数 (表 3) の平均値

ID	w_3	w_7	w_{35}	w_{41}	w_{45}	w_{62}
1	0.24	0.20	0.40	0.34	0.39	0.18
2	-0.41	-1.14	-0.36	0.09	0.25	0.16
3	-0.84	-0.70	0.86	0.08	0.68	0.25
4	-0.01	-0.26	-0.41	-0.23	-0.82	-0.63
5	-0.61	-0.24	-0.62	-0.74	-0.50	-0.29
6	0.24	0.55	-0.34	-0.20	-0.16	0.16

表 5 タクシー経路のクラスターの概要

ID	実車率 [%]	経路数	解釈 (5.2 節)
1	38.6 $\pm$ 6.3	5026	東京中心部・羽田空港流し
2	29.6 $\pm$ 7.2	896	渋谷駅南西側周回・付け待ち
3	34.4 $\pm$ 5.4	1061	品川駅周辺流し・付け待ち
4	30.5 $\pm$ 6.7	2306	東京駅周辺・北側付け待ち
5	28.2 $\pm$ 6.3	1040	池袋駅北側付け待ち
6	35.5 $\pm$ 6.9	2910	東京中心部流し

6. おわりに

本研究では、グラフの高性能ベクトル化手法 graph2vec [12] を利用して、GPS タクシー軌跡データを分析した。運転手の行動パターンが時間帯によって異なることから、時間帯の情報を頂点に付与した時間付加グラフのベクトル表現を提案した。タクシー経路のベクトル表現を利用した回帰分析では、時間帯の情報を付与した場合に実車率の予測精度が向上した。

本研究では回帰分析とクラスター分析の結果に基づき、実車率の向上に寄与する行動パターンを提示した (4.3 節, 5.2 節)。これらは時間帯ごとにどのような行動をすれば効率的な営業ができるかを示しており、営業効率の良い運転手の行動の特徴を整理することができた。当然ながら他の地域や他のタクシー会社では有効な行動パターンが異なると考えられるが、本研究の分析過程は同種の他のデータにも適用可能であり、このような汎用的な分析過程を提示したことも本研究の成果といえる。

今後の課題として、道路ネットワークに基づく経路グラフを分析することや、道路の混雑状況を頂点ラベルに付与することが挙げられる。また、他のベクトル化手法との性能の比較や、タクシー経路グラフの特徴を活用したベクトル化手法の開発についても今後検討したい。

謝辞 貴重なデータを提供していただいたデータ解析コンペティション関係者の皆様に御礼申し上げます。また大変有用なコメントをいただいた 2 名の査読者に

心より感謝申し上げます。

参考文献

- [1] 全国ハイヤー・タクシー連合会, Taxi today 2020 in Japan, http://www.taxi-japan.or.jp/pdf/Taxi_Today_2020.pdf (2020 年 7 月 10 日閲覧)
- [2] L. Liu, C. Andris and C. Ratti, “Uncovering cab-drivers’ behavior patterns from their digital traces,” *Computers, Environment and Urban Systems*, **34**, pp. 541–548, 2010.
- [3] B. Li, D. Zhang, L. Sun, C. Chen, S. Li, G. Qi and Q. Yang, “Hunting or waiting? Discovering passenger-finding strategies from a large-scale real-world taxi dataset,” *IEEE International Conference on Pervasive Computing and Communications Workshops*, pp. 63–68, 2011.
- [4] D. Zhang, L. Sun, B. Li, C. Chen, G. Pan, S. Li and Z. Wu, “Understanding taxi service strategies from taxi GPS traces,” *IEEE Transactions on Intelligent Transportation Systems*, **16**, pp. 123–135, 2015.
- [5] Q. Ma, H. Yang, H. Zhang, K. Xie and Z. Wang, “Modeling and analysis of daily driving patterns of taxis in reshuffled ride-hailing service market,” *Journal of Transportation Engineering Part A: Systems*, **145**, 2019.
- [6] H. Yao, F. Wu, J. Ke, X. Tang, Y. Jia, S. Lu, P. Gong, J. Ye and Z. Li, “Deep multi-view spatial-temporal network for taxi demand prediction,” In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*, pp. 2588–2595, 2018.
- [7] Z. Huang, G. Shan, J. Cheng and J. Sun, “Trec: An efficient recommendation system for hunting passengers with deep neural networks,” *Neural Computing and Applications*, **31**, pp. 209–222, 2019.
- [8] H. Cai, V. W. Zheng and K. C. C. Chang, “A comprehensive survey of graph embedding: Problems, techniques, and applications,” *IEEE Transactions on Knowledge and Data Engineering*, **30**, pp. 1616–1637, 2018.
- [9] A. Grover and J. Leskovec, “Node2vec: Scalable feature learning for networks,” In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 855–864, 2016.
- [10] N. Shervashidze, P. Schweitzer, E. J. V. Leeuwen, K. Mehlhorn and K. M. Borgwardt, “Weisfeiler-Lehman graph kernels,” *Journal of Machine Learning Research*, **12**, pp. 2539–2561, 2011.
- [11] P. Yanardag and S. V. N. Vishwanathan, “Deep graph kernels,” In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1365–1374, 2015.
- [12] A. Narayanan, M. Chandramohan, R. Venkatesan, L. Chen, Y. Liu and S. Jaiswal, “Graph2vec: Learning distributed representations of graph,” *arXiv preprint*, arXiv:1707.05005, 2017.
- [13] 伊藤彰彦, “地域メッシュ統計の紹介,” *行動計量学*, **4**, pp. 59–63, 1976.
- [14] R. Tibshirani, “Regression shrinkage and selection via the Lasso,” *Journal of the Royal Statistical Society: Series B (Methodological)*, **58**, pp. 267–288, 1996.
- [15] 小西貞則, 『多変量解析入門—線形から非線形へ—』, 岩波書店, 2010.