Vol.2017-MBL-84 No.21
Vol.2017-CDS-20 No.21
2017/8/30

リアルタイム人口を用いた Stacked denoising Autoencoder によるタクシー将来需要予測

石黒慎[†1]　菊地悠[†1]　深澤佑介[†1]

概要：タクシーの効率的な運行には，運転手に乗客に関する様々な情報を与えることが有効である．高効率なタクシー運行は，タクシー事業の収益向上に繋がるだけでなく，乗客が素早くタクシーに乗車することにも効果があり，社会全体の交通の効率化を期待できる．本稿では Stacked denoising Autoencoder を用いたタクシー将来需要の予測とそれに基づいた運行支援手法を提案する．提案法では，タクシー運行データ，リアルタイム人口統計データおよび雨量データを用いることで，MAPE による評価により，26.77%の誤差で予測が可能であることを示す．

キーワード：タクシー，需要予測，ニューラルネットワーク，位置情報

Real-Time Population based Taxi Demand Forecast using Stacked denoising Autoencoders

SHIN ISHIGURO[†1]　HARUKA KIKUCHI[†1]　YUSUKE FUKAZAWA[†1]

1. はじめに

タクシードライバーにとって利益を最大化するためには，空車の状態を減らし，乗車時間を最大化することが重要である．ベテランドライバーは，どこを運行すれば乗客の候補がいるかを暗黙的な知見として持っており効率的に収益を得ることができる．一方，新人のドライバーはそのような知見が少なく，結果として収益につながらず，離職などにつながる可能性がある．また，ベテランドライバーであっても，不慣れな場所に送迎した帰路は空車になる可能性が高い．従来，このような格差を埋めるためにマニュアルや講習などでノウハウを伝承していたが，効率的ではない．そこで，本研究では，過去のタクシーの乗車データおよび人々の統計的な位置情報から場所ごとにタクシーの乗車需要を予測するアルゴリズムを提案する．さらに，タクシー運転手にその情報を表示し，実際に乗車需要が予測できたかどうかを確認する．タクシーの効率的な運行は，タクシー事業の収益向上に繋がるだけでなく，乗客が素早くタクシーに乗車することにも繋がる．

近年では機械学習手法の一つであるディープラーニングが注目を浴びている．今日までにディープラーニングは画像認識，自然言語処理，音声認識など多様な分野で成果を挙げている[15] ．ディープラーニングでは，多層のニューラルネットワーク構造を用いて，低レベルデータの入力を元に，データ間の構造・関係性を高レベルな特徴量で表現することで，データ中の重要な要素の抽出を実現する．

本研究では，タクシーの需要を予測するため，リアルタイム人口，天気，過去のタクシーの乗車データなど様々な異なる情報を扱う．予測精度を上げるためには，これらを組み合わせた複雑で抽象的な特徴を設計する必要があるが，組み合わせの可能性が膨大になり人手では対応ができない．そこで，本研究では，これらの異種混合データから抽象的で複雑な特徴を自動で獲得するため，ディープラーニングを用いたタクシー将来需要予測手法を提案する．ここでは様々な研究で予測精度の効果が示されている Stacked denoising Autoencoder (SdA)モデル[1]を用いて個別に各層の学習を進める．モデル学習の中で，入力の時間情報と空間情報の関係性が非明示的に考慮される．

2. 関連研究

これまでタクシーの将来需要予測は，多くの研究者によって研究をされている．過去の関連研究には以下の様な方法がある．

Chang et al.[2]は，タクシー乗降履歴データのクラスタリングを行い，クラスタ領域内の道路とクラスタを対応付けることで，道路毎に将来需要を推定する手法を提案した．

Li et al.[5],，Powell et al.[6]は，位置情報データを時刻ごとグリッドごとに分け，グリッド毎の需要を推定する手法を提案した．

Lee et al.[3], Yue et al.[4]は，タクシーの乗降履歴データのクラスタリングを行い，タクシー需要のホットスポットを推定し，ホットスポットごとに需要を推定する手法を提案した．

このように，タクシーの将来需要予測には注目をされており，現在まで，様々な需要予測手法が研究されている．しかしながら，過去の研究では，タクシーデータのみを用

†1 株式会社 NTT ドコモ
NTT DOCOMO, INC.

Vol.2017-MBL-84 No.21
Vol.2017-CDS-20 No.21
2017/8/30

いた分析を行っており，人々の移動データおよび過去のタクシーデータの両方を用いた研究がなされていない. また，ディープラーニングを用いて精度向上を図った研究はなされていない.

3. 問題設定

本研究では，人口移動データ，天気データおよび過去のタクシー乗車数を入力とし，500m グリッド単位で 30 分先の需要（当該グリッドにおける乗車予測数）を出力するアルゴリズムを提案する．需要予測は連続値となるため，回帰問題としてモデル化する．また，グリッドごとにモデルを構築するのではなく，全グリッドで共通のモデルを作成する.

4. 手法

Stacked Autoencoder を用いたディープラーニングによるタクシー将来需要予測手法について説明する．Stacked Autoencoder[14]は Autoencoder[8]を積層することで, 深層化したニューラルネットワークにより，学習を行うモデルである.

4.1 Autoencoder による特徴抽出 (pre-training)

Autoencoder は入力データの復元を試みるニューラルネットワークの手法である．図 1 に Autoencoder の模式図を示す，この図では，入力層，隠れ層，出力層を 1 つずつ持つニューラルネットワークが表現される.

入力ベクトル$x = \{x_1, x_2, x_3, \dots, x_d\}$，$x_i \in R$が与えられるとき, Autoencoder は, まず式(1)のエンコーダー$y = f_\theta(x)$を用いて入力ベクトルxを変換した出力 y を求める. 次に式(2)のデコーダー$z = g_{\theta'}(y)$を用いて，隠れ層の出力yを入力として，$\tilde{x}$ を出力する．Autoencoder では，ニューラルネットワークの説明変数，目的変数にそれぞれ同じ値を入力した教師なし学習による事前学習を行うことで，エンコーダーとデコーダーに変換と復元の機能を実現する.

$$y = f_\theta(x) = s(Wx + b) \quad (1)$$

$$z = g_{\theta'}(y) = s(W'y + b') \quad (2)$$

式(1)のパラメータは$\theta = \{W, b\}$，式(2)のパラメータは$\theta' = \{W', b'\}$で表現される．ここで，Wは重み行列，bはバイアスベクトル，W'は逆変換の重み行列，b'は逆変換のバイアスベクトルとなる．$s(x)$には非線形の活性化関数を用いることができ，検証では，ReLU 関数$s(x) = max\,(0, x)$を用いた.

復元誤差$L(x, \tilde{x})$を最小化するため，下記の目的関数を用いて，モデルのパラメータθ, θ'を最適化する.

$$\theta, \theta' = \underset{\theta,\theta'}{\operatorname{argmin}} L(x, z)$$

$$= \underset{\theta,\theta'}{\operatorname{argmin}} \frac{1}{N} \sum_{i=1}^{N} \left(x_i - g(f(x_i))\right)^2 \quad (3)$$

このようにして獲得されるエンコーダーの出力yは，入力xを復元するための情報を保持したまま，隠れ層のノード数の増減を実現する．したがって，Autoencoder は元の入力から役立つ情報を抽出することを実現する機能を有する.

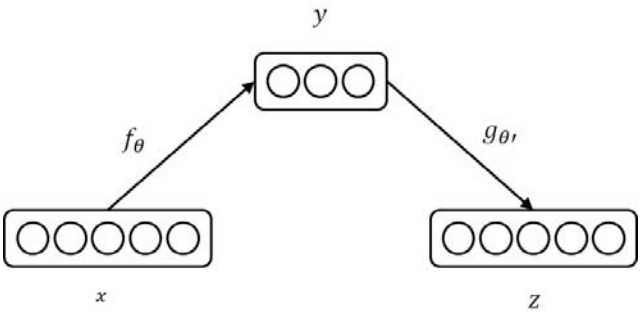

図 1. Autoencoder の模式図

4.2 Autoencoder の積層によるモデル作成 (fine tuning)

Stacked Autoencoder のモデルでは，下層の Autoencoder の出力結果を入力として用いることで, Autoencoder の層を積層し，深いネットワーク構造を表現する.

層数がl である Stacked Autoencoder を考える．初めの層は入力データを用いた Autoencoder を学習する．初めの層を獲得した後，k 番目の隠れ層の出力を，k+1 番目の隠れ層の入力として用いる．このようにして，複数の Autoencoder の積層を実現する.

本稿のタクシー将来需要予測では，需要を回帰問題として解くアプローチを行う．このため，作成された Stacked Autoencoder モデルの最終層に回帰の予測器を加える必要がある．本論文ではロジスティック回帰の予測器を最終層に加え，目的変数としてタクシーの需要データを入力する教師あり学習を行う．ネットワークのファインチューニングを行うことで，深層ネットワーク全体でタクシー需要予測を行うアーキテクチャを実現する．図 2 に Stacked Autoencoder の予測器の模式図を示す.

4.3 Sparse Autoencoder

Autoencoder で表現されるエンコーダーの出力結果 y は，入力 x の情報を保存するが，それだけでは有用な情報が抽出されたことを保証することが出来ない．なぜなら，隠れ層のノード数が入力層と等しい Autoencoder の最適化では，恒等関数が学習される可能性があるからである[1]．したがって, 入力データのノイズから有用な情報を分離するには，更なる制約が必要である.

Vol.2017-MBL-84 No.21
Vol.2017-CDS-20 No.21
2017/8/30

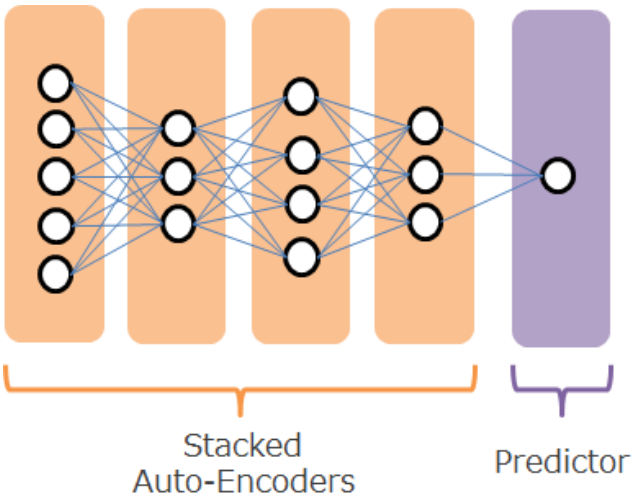

図 2. Stacked Autoencoder による予測器

入力層の次元数を d, 隠れ層の次元数を d' とした際, $d'<d$ とする変換を行う場合を考える．このようなエンコーダーの出力 y は入力 x よりも少ない次元数で x の復元を実現するため，次元圧縮と呼ばれる．しかしながら，低次元の y からの復元 x は，復元誤差が大きくなり，元の x よりも情報量が失われる可能性がある．

一方で，$d'>d$ とする変換を行う場合を考える．このエンコーダーの出力 y は，スパース表現を学習する．スパース表現では，次元圧縮より正確に元の情報を保持出来る場合があり，代替として注目を浴びてきた[12]．明示的に次元数を落とす場合と異なり，スパース表現は多くの 0 値を含むことで，隠れ層ノードで内的に次元削減を表現する．このようにデータがスパースに表現される Autoencoder を Sparse Autoencoder という[12].

スパース表現の獲得を促進するため，目的関数に制約を加えることを考える．本稿では重みWの平均を正則化項として加算する．正則化係数ρには 0 に近い小さな値を用いる．

$$L'(x,z) = L(x,z) + \frac{\rho}{D}\sum_{j=1}^{D}|W_j| \quad (4)$$

4.4 Denoising Autoencoder

次元圧縮，Sparse Autoencoder とは異なるアプローチとして，入力データにノイズを加えた上で，変換・復元を行う Denoising Autoencoder の手法がある[9]．Denoising Autoencoder では, 欠損が加わったデータから元データを復元する処理を行うことで，入力データが元々持つノイズや欠損に対してロバストになること．元データの復元において重要な情報を優先的に抽出することを期待できる．

Denoising Autoencoder のアルゴリズムでは，入力データに欠損を加えた上で，欠損が加わる前の元データの復元を行う．$q_D(\tilde{x}|x)$により，xに欠損を加えた$\tilde{x}$を取得する．$\tilde{x}$を用いて，通常の Autoencoder をかける．

$$y = f_\theta(\tilde{x}) = s(W\tilde{x} + b) \quad (5)$$
$$z = g_{\theta'}(y) = s(W'y + b') \quad (6)$$

復元された z について, 欠損が加わる前のxとの平均誤差の最小化を実施する．これによって，欠損のない入力xを用いる場合よりも有用な情報を優先的に抽出することを実現する．

$$\theta, \theta' = \underset{\theta,\theta'}{\operatorname{argmin}} \mathrm{L}(x,z) \quad (7)$$

5. 実験

5.1 データの詳細

提案法のディープラーニングを東京のタクシーの将来需要予測に適用する．本稿では入力データとして，タクシーデータ，人口データ，雨量データの 3 種類のデータを用いて検証を行った．

タクシーデータとして，個々のタクシーに設置された GPS デバイスより 5～10 秒に 1 回の周期で，緯度，経度，客車状態(0, 1)の収集されたものを用いる．

人口データとして，携帯電話ネットワークの仕組みを利用して人口を推定した人口統計データを用いた．モバイル端末と基地局の位置関係から，東京都内では，500m の空間解像度，10 分毎の時間解像度で人口推定が可能であり，モバイルネットワークのデータから，日本の実際の全人口の推定を行ったデータとなる．*a

雨量データは，250m グリッドで降水量の推定を行なった高解像度降水ナウキャストのデータを用いた．各データの詳細は表 1 の通りである．

評価では， 10 分毎，各 500m グリッドに関して，ある対象の時刻から将来 30 分間に同エリアで何台のタクシーが乗客を乗せることが出来たかを計算する．2015 年 4 月 1 日～2016 年 8 月 31 日を学習用データとして用い，2016 年 9 月 1 日～2016 年 9 月 14 日を評価用データとして利用した．

図 3 にタクシーの乗車数毎の総出現回数を示す．このように，タクシーの乗車数は少量の乗車数が極めて多く出現する分布となっている．

a 本実験で使用する人口統計は、エリア毎や属性毎の集団の人数を示す情報であり、お客様個人を特定できる情報を一切含みません。したがって、この人口統計によりお客様の行動が他人に知られることはありません。なお、本実験で使用する人口統計は、モバイル空間統計ガイドラインを遵守しております。**モバイル空間統計ガイドライン**
https://www.nttdocomo.co.jp/corporate/disclosure/mobile_spatial_statistics/guideline/index.html

表 1. データの詳細について

データソース		東京都
期間		2015 年 4 月 1 日～ 2016 年 9 月 14 日
タクシーデータ	台数 取得頻度	4400 台 5～10 秒毎
人口データ	解像度 取得頻度	500m 10 分毎
雨量データ	解像度 取得頻度	250m 10 分毎

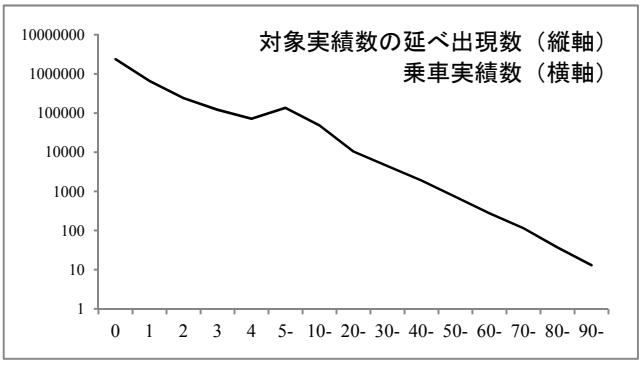

図 3. タクシー乗車数実績と総出現回数（対数スケール）
（2016 年 9 月 1 日～9 月 14 日）

5.2 データの前処理について

5.2.1 空間解像度/時間解像度の変換

各データで異なる空間解像度，時間解像度を合わせるため，各データの集約を行った．タクシーデータについては，500m/10 分毎に，将来 30 分間に渉る乗客の乗車数 ，乗客の降車数の総和を計算した．人口データは，500m/10 分毎の人口データへの集約を実施，雨量データでは，250m 毎の雨量の平均を求めることで, 500m への変換を行った．

5.2.2 時系列特徴量／統計特徴量

需要予測の精度を高めるには，タクシー需要に関係するデータを入力することが望ましい．タクシーの乗車数は季節，曜日（平日・休日・休前日），および時刻によって変化する．このような時系列による変化を予測するには，現在の時刻情報，短期間の時系列トレンド，長期間の時系列トレンドの情報をそれぞれ考慮することが有効である．

現在の時刻情報として，曜日情報および時刻情報を用いる．曜日情報は当日の平日・休日判定，翌日の平日・休日判定に基づいた 2bit の 1-of-k のデータで表現した．時刻情報は 0～24 時の形で用いると 0 と 24 が同時刻であるにも関わらず大きな差を生じてしまう．従って本稿では，下記式に基づいて時刻を(α, β)の 2 次元で表現した．

$$\alpha = \sin\frac{\pi}{12}t, \qquad \beta = \cos\frac{\pi}{12}t, \qquad \{0 \le t < 24\} \quad (8)$$

短期間のトレンドとしては，タクシーの乗降履歴，人口，雨量を対象の時刻から 30 分前，60 分前，…6 時間前までのラグ値を利用する．長期間のトレンドとして，過去 1 年間のタクシー乗降履歴，人口のデータについて各グリッド，曜日，時刻毎に平均を計算したデータを利用する．

5.2.3 データの正規化

ニューラルネットワークの活性化関数は，入力データが 0 付近にあるときに機能する．上記手順で作成された各データについて，式 X の計算を行うことで，$-1 \le x_i \le 1$を満たすようにデータの正規化を行った．

$$\hat{x_i} = \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)} \quad (9)$$

5.2.4 データのミニバッチ化

ネットワークの汎化性能を上げる手法としてミニバッチという手法が知られている[13]．本手法でもミニバッチを採用する．これまでの手順で作成された各データを一つのデータセットとしてまとめ，500m/10 分毎にデータセットからランダムに取得したミニバッチを取得する．

またバッチ毎のデータの分布の違いを考慮するため，活性化関数の前段で Batch Normalization[10]を用いることで，データの正規化を実施した．実験の SdA では，このようにデータを学習データとして用いる．

5.3 Stacked denoising Autoencoder のパラメータ探索

Denoising Autoencoder の深層ネットワークの予測性能はネットワーク構造を規定する各ハイパーパラメータの設定に依存する．

ハイパーパラメータは各層毎に多数あり，ハイパーパラメータ同士も互いに関係している為，選択パターンが多岐にわたる．探索方法として良く知られているグリッドサーチによる全探索は，離散的に設定されたパラメータ群から探索を行う物であるが，探索範囲が限られるため，十分に探索が実施された場合，ランダムサーチの精度に劣ることが知られている[11]．そこで本稿では Python の hyperopt モジュールを利用し，ランダムサーチによるハイパーパラメータ探索を 実施した[16].

探索対象としたハイパーパラメータは, 各層のノード数, denoising Autoencoder のノイズ係数，Sparse Autoencoder の正則化係数，Dropout の割合，バッチサイズである．

それぞれのパラメータについて表 2 の範囲で探索を実施した．

Vol.2017-MBL-84 No.21
Vol.2017-CDS-20 No.21
2017/8/30

表 2. ハイパーパラメータの探索範囲

ハイパーパラメータ	探索範囲
Denoising Autoencoder の ノイズ係数	$0 \leq q_D \leq 0.2$
Sparse Autoencoder の 正則化係数	$0 \leq \rho \leq 0.02$
Dropout の割合	$0.3 \leq \varepsilon \leq 0.7$
バッチサイズ	$50 \leq BS \leq 200$
ノード数	$10 \leq N \leq 1000$

5.4 評価指標について

提案法の有用性を評価するため，絶対平均誤差（MAE），2 乗平均平方根誤差（RMSE），平均絶対誤差率（MAPE）の 3 種類の指標を用いて評価を行う．これらの定義を以下に示す．ここで，t_iは，タクシーの実際の乗車量，$\hat{t}_i$は予測したタクシーの乗車量である．

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|t_i - \hat{t}_i| \qquad (10)$$

$$\mathrm{RMSE} = \left(\frac{1}{n}\sum_{i=1}^{n}(t_i - \hat{t}_i)^2\right)^{\frac{1}{2}} \qquad (11)$$

$$\mathrm{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\frac{|t_i - \hat{t}_i|}{t_i} \qquad (12)$$

5.5 実験結果

実験の深層学習では，最適化関数として Adam を用いた学習を実施した．3 層，4 層，5 層の SdA について 1000 モデルずつ，パラメータ探索および学習を行い，精度評価を行った．表 3 および図 4 に評価結果を示す．結果から，乗車実績値の小さい領域では 3 層の SdA の精度が高く，一方，乗車実績値の大きい領域では，4 層の SdA の精度が高いことがわかる．これは，乗車実績数の小さい領域においては，総数の少ない 3 層のネットワークでも将来需要を説明する重要なパラメータを抽出することができていると考えられ，一方で，乗車実績値の大きい領域では，より表現力の高い 4 層の SdA によるデータ構造の抽象化が上手く機能し，精度を向上することができたと考えられる．また，5 層の SdA ではモデルの有する表現力が高く，探索範囲が広がり過ぎた結果，同じ学習時間では十分な性能を得ることが出来なかったと考えられる．

また，4 層のネットワークによる予測精度について，データを変更して学習を行った．表 4 および図 5 にタクシーデータのみ，タクシーデータと人口データ，タクシーデータと気象データ，3 種全てのデータでの学習結果をそれぞれ示す．結果から，タクシーデータ単体での学習に比べて，人口データおよび雨量データを加えた学習を行った場合に性能が向上することが確認できる．これは，タクシーデータのみでは説明できないタクシーの将来需要に関する重要なデータを SdA により抽出し，予測に利用出来ていると考えられる．また，特に雨量データが全域の乗車実績値の精度向上に大きく寄与していることがわかる．全データによる学習では，タクシーおよび雨量データによる学習に比べて乗車実績値の小さい領域では精度が劣り，一方で乗車実績値の大きい領域では精度が勝ることがわかる．これは，人口データが乗車実績値の大きな場所により影響する傾向があり，性能向上に寄与したことが考えられる．結果として，全データを用いた学習結果では，幅広いグリッドおよび時間帯で高い 需要予測精度を期待できる．

表 3. 層数毎のモデル評価結果

手法	MAE	RMSE	MAPE
3 層	0.727	1.295	38.78
4 層	0.990	1.378	26.77
5 層	0.962	1.399	31.74

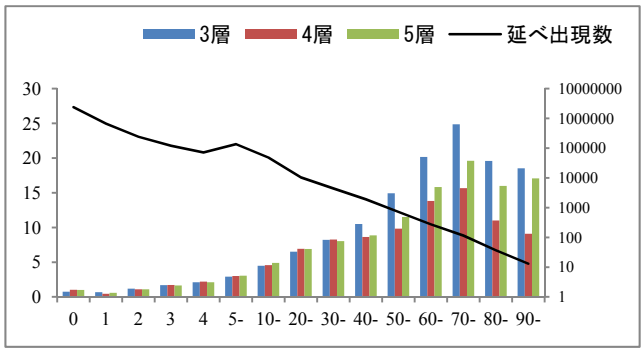

図 4. 各層数の SdA による乗車実績値毎の予測精度（RMSE）
左縦軸：RMSE，右縦軸：対象実績値の延べ出現数（対数スケール），横軸：乗車実績値

表 4. 異なる入力データによるモデル評価結果

データ	MAE	RMSE	MAPE
タクシー	1.042	1.513	28.74
タクシー ＋ 人口	0.877	1.370	36.75
タクシー ＋ 雨量	0.994	1.351	32.03
全データ	0.990	1.378	26.77

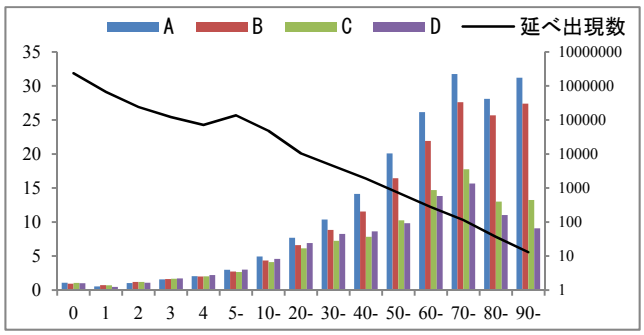

図 5. 異なる入力データによる SdA の乗車実績値毎の予測精度（RMSE）
左縦軸：RMSE，右縦軸：対象実績値の延べ出現数（対数スケール），横軸：乗車実績値，A: タクシーデータのみ利用，B: タクシーおよび人口データ利用，C: タクシーおよび雨量データ利用, D: 全データを利用

また，図 6 に東京のある地点における予測値と実測値の時系列グラフを示す．図から，時系列に従って変化するタクシーの需要を上手く捉え，予測することが出来ていることが分かる．

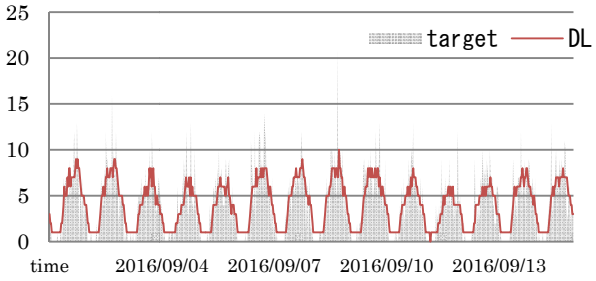

図 6. ある地点におけるタクシー需要の予測値・実測値の比較グラフ（横軸：時刻，縦軸：乗車数）

6. まとめ

本論文では，Stacked denoising Autoencoder を用いたタクシー需要予測手法について提案した．実験では，タクシー乗降履歴，人口，雨量を組み合わせた位置情報データを用いた学習を行い, MAPE では 26.77%での予測精度を示した．

提案では，タクシーの需要を予測する方法を示した．しかしながら全てのドライバーに同じ情報を見せるとタクシーの向かう場所が不均衡になり，客車効率が下がる可能性がある．したがって，個々のドライバーの配車制御の効率化が今後の課題の一つである．

謝辞

本研究を実施するにあたり，タクシーデータの提供や様々な相談をさせて頂いた東京無線協同組合の皆様に謹んで感謝の意を表する．

参考文献

[1] Vincent, Pascal, et al. "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion." Journal of Machine Learning Research 11.Dec (2010): 3371-3408.

[2] Chang, Han-wen, Yu-chin Tai, and Jane Yung-jen Hsu. "Context-aware taxi demand hotspots prediction." International Journal of Business Intelligence and Data Mining 5.1 (2009): 3-18.

[3] Lee, Junghoon, Inhye Shin, and Gyung-Leen Park. "Analysis of the passenger pick-up pattern for taxi location recommendation." Networked Computing and Advanced Information Management, 2008. NCM'08. Fourth International Conference on. Vol. 1. IEEE, 2008.

[4] Yue, Yang, et al. "Mining time-dependent attractive areas and movement patterns from taxi trajectory data." Geoinformatics, 2009 17th International Conference on. IEEE, 2009.

[5] Li, Bin, et al. "Hunting or waiting? Discovering passenger-finding strategies from a large-scale real-world taxi dataset." Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011 IEEE International Conference on. IEEE, 2011.

[6] Powell, Jason W., et al. "Towards Reducing Taxicab Cruising Time Using Spatio-Temporal Profitability Maps." SSTD. 2011.

[7] Li, Xiaolong, et al. "Prediction of urban human mobility using large-scale taxi traces and its applications." Frontiers of Computer Science 6.1 (2012): 111-121.

[8] Hinton, Geoffrey E., and Ruslan R. Salakhutdinov. "Reducing the dimensionality of data with neural networks." science 313.5786 (2006): 504-507.

[9] Vincent, Pascal, et al. "Extracting and composing robust features with denoising autoencoders." Proceedings of the 25th international conference on Machine learning. ACM, 2008.

[10] Ioffe, Sergey, and Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift." International Conference on Machine Learning. 2015.

[11] Bergstra, James, and Yoshua Bengio. "Random search for hyper-parameter optimization." Journal of Machine Learning Research 13.Feb (2012): 281-305.

[12] Ng, Andrew. "Sparse autoencoder." CS294A Lecture notes 72.2011 (2011): 1-19.

[13] Li, Mu, et al. "Efficient mini-batch training for stochastic optimization." Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2014.

[14] Lv, Yisheng, et al. "Traffic flow prediction with big data: a deep learning approach." IEEE Transactions on Intelligent Transportation Systems 16.2 (2015): 865-873.

[15] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." Nature 521.7553 (2015): 436-444.

[16] Bergstra, James, Dan Yamins, and David D. Cox. "Hyperopt: A python library for optimizing the hyperparameters of machine learning algorithms." Proceedings of the 12th Python in Science Conference. 2013.