

# AIタクシー —交通運行の最適化をめざした タクシーの乗車需要予測技術—

サービスイノベーション部

かわさき さとし  
川崎 仁嗣いしぐろ しん  
石黒 慎ふかざわ ゆうすけ  
深澤 佑介

IoTビジネス部

ふじた まさのり  
藤田 将成すずき りょうへい  
鈴木 亮平まきしま あきひと  
槇島 章人

タクシードライバーが収益を増加させるためには、不慣れな地域や時間帯などにおいても効率的にお客様を獲得し、乗客を乗せている実車時間を最大化することが重要となる。そこで、ドコモは過去のタクシー運行データや、人々の統計的な位置情報から、エリアごとにタクシーの乗車需要を予測する技術を開発し商用提供を開始した。これにより、タクシー運転手はリアルタイムに変化する乗車需要を確認し、効率的な運行による生産性向上が可能となるほか、タクシーの乗客にとっても待ち時間の短縮が期待できる。

## 1. まえがき

タクシーの運行において、空車状態の時間を減らし、実車（乗客を乗せている状態）の時間を最大化することは、収益向上の観点で非常に重要である。不慣れな地域や時間帯、また、新たな道路の開通や商業施設のオープンなどの日々変化する環境、電車遅延など突発的な状況下においても効率的に乗客を見つかることができれば、実車時間の割合を増やすことができタクシードライバーの収益向上が期待できる。さらに、効率が改善されることで生産性が向

上し、より働きやすい環境となることも期待される。

このような状況に対応するためドコモは、過去のタクシーの乗車実績データ、および、人々の統計的な位置情報を用いて、場所ごとにタクシーの乗車需要を予測する手法を考案した。また、その手法を組み込んだ需要予測システムを開発し、その有効性をタクシー事業者と共同で実施した実証実験にて検証し、全期間連続で収益向上と実車率向上の効果が確認できたことからAIタクシー<sup>®\*1</sup> [1] として提供を開始した。需要予測システムの予測結果をタクシードライバーに提示することで、図1に示すように需

©2018 NTT DOCOMO, INC.  
本誌掲載記事の無断転載を禁じます。

\*AIタクシーで使用する人口統計は、エリアごとや属性ごと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、AIタクシーで使用する人口統計は、以下のモバイル空間統計ガイドラインを遵守しております。  
[https://www.nttdocomo.co.jp/corporate/disclosure/mobile\\_spatial\\_statistics/guideline/index.html](https://www.nttdocomo.co.jp/corporate/disclosure/mobile_spatial_statistics/guideline/index.html)

要の高いエリアへ配車することができる。結果として図2①に示すように乗客がタクシーをつかまえるまでの時間が短くなるだけでなく、図2②～④に示すようにタクシー事業の収益向上に繋がること期待できる。

## 2. タクシーの乗車需要予測技術

### 2.1 概要

本技術では、モバイル空間統計のリアルタイム版\*2、天気予報、過去のタクシー乗車実績などを入

力データとし、500mメッシュ\*3単位で30分先までのタクシー需要（当該メッシュにおける乗車の発生回数）を推定する。

一般的に、過去の実績から将来にわたる時系列のデータを予測するには、時系列予測モデルである自己回帰モデル（AR model：Auto Regressive model）\*4やそれを発展させた自己回帰和分移動平均モデル（ARIMA model：Auto Regressive and Moving Average model）\*5が用いられる。タクシーの乗車需要においても同様に時系列予測モデルを適用することが可能であるが、過去のタクシー乗車実績

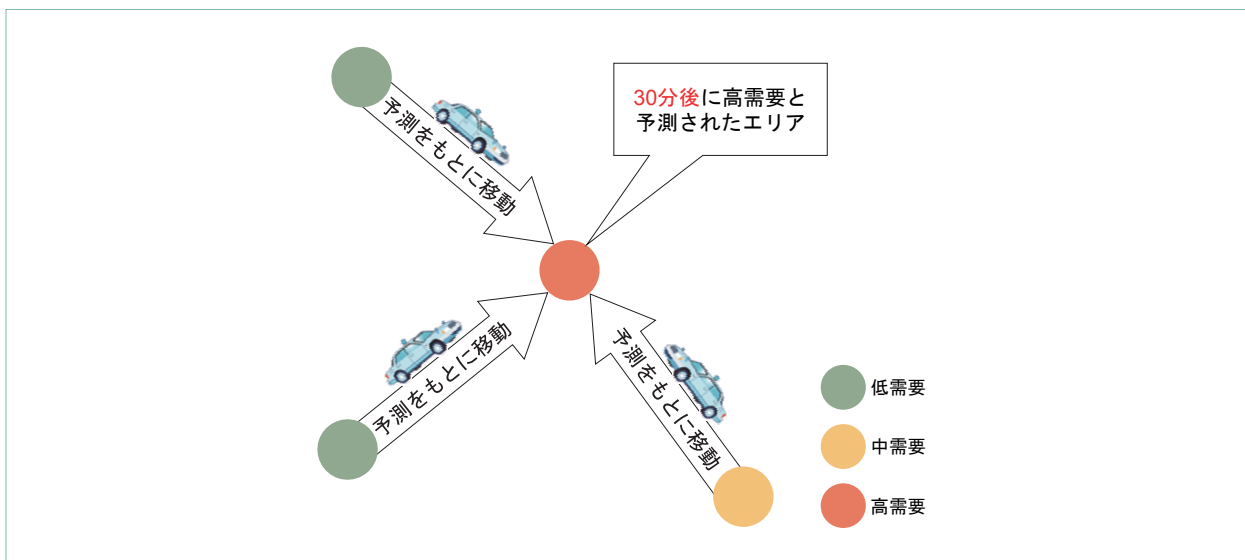


図1 需要予測とタクシー移動

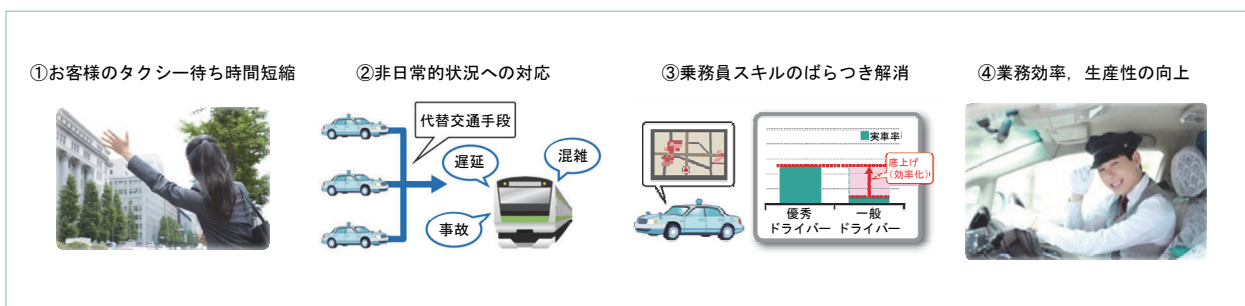


図2 AIタクシーによる期待効果

\*1 AIタクシー：(株)NTTドコモの登録商標。  
 \*2 モバイル空間統計のリアルタイム版：ドコモの携帯電話ネットワークの仕組みを使用して作成される人口の統計情報で、10分ごとに直近の日本全国の人口を500mグリッドで推定することが可能である。  
 \*3 メッシュ：緯度・経度に基づき、国土を網の目状に分けた区画。  
 \*4 自己回帰モデル（AR model）：回帰モデルでは、結果の値（目的変数）を、影響を与えた値（説明変数）で表す式としてモデ

ル化するが、その中でも説明変数の値として目的変数の過去の値を用いるモデル。  
 \*5 自己回帰和分移動平均モデル（ARIMA model）：自己回帰モデル（Auto Regressive model）と移動平均モデル（Moving Average model）、和分モデル（Integrated model）とを組み合わせた回帰モデル。移動平均モデルは過去の予測値と実績値との誤差の和、和分モデルは階差を利用してモデル化する。

だけでなく、天気予報や周囲の人口変動も乗車需要に影響を与えていることから、より精度を向上させるためこれらのデータも組み合わせた自己回帰である、多変量自己回帰モデル\*6を利用している。しかし、多変量自己回帰モデルを含む自己回帰モデルでは、メッシュごとにモデル式内の重み値（パラメータ）として適切な値を人間が決めることが一般的である。サービスをより広範囲のエリアで展開していくにあたり、メッシュの数が増えることから機械的に重み値を決定できる手法が求められる。そこで、これらの異種混合データから精度向上に有用となるデータの組合せやパラメータを機械的に獲得するため、ディープラーニングと呼ばれる手法を用いることとした。商用サービスにおいては、図3に示すように多変量自己回帰モデルの予測結果とディープラーニングの予測結果とで精度の良い方を利用する構成となっている。

ディープラーニングは深層学習とも呼ばれる機械学習\*7手法の1つであり、学習処理などに高い計算能力が必要であるが、ここ最近のハードウェア性能の向上に伴い、これまでよりも低コストで実現可能と

なった事から注目を集めている。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他の機械学習手法を圧倒する高い性能を示し成果を挙げている。ディープラーニングでは、多層（一般的には4層以上）のニューラルネットワーク\*8を用いて、高度な特徴量抽出や選定などの事前のモデル設計を行わずに、単純な特徴量のデータを入力として、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。

## 2.2 モバイル空間統計のリアルタイム版の活用

需要予測においてモバイル空間統計のリアルタイム版を用いていることは本技術の大きな特徴である。これまで、モバイル空間統計\*9として過去の人口統計データを活用している事例はあったが[2]、タクシーの需要予測においては直前の人口統計を利用することが精度向上の観点でも重要であり、今回新たに直近の人口統計データを作成することが可能な、モバイル空間統計のリアルタイム版を用いている。

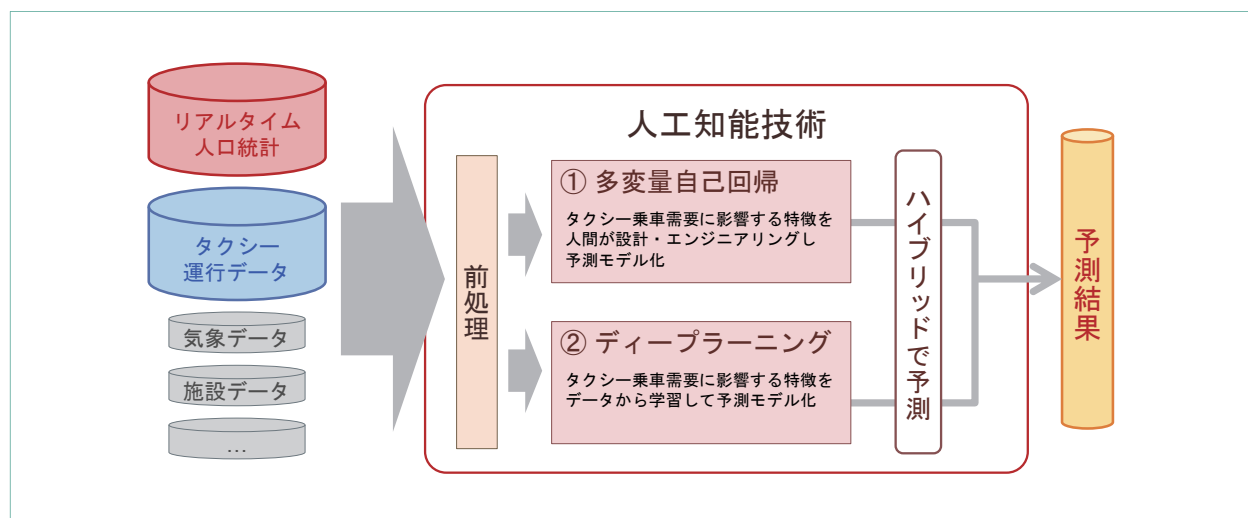


図3 タクシー乗車需要予測技術の構成

\*6 多変量自己回帰モデル：自己回帰モデルを多変量に拡張したモデル。ベクトル自己回帰モデルともいう。  
 \*7 機械学習：サンプルデータから統計処理により、有用な判断基準をコンピュータに学習させる枠組み。  
 \*8 ニューラルネットワーク：人間の脳機能における特性を計算機で模倣したものであり、脳では神経同士の結合が学習を繰り返すことで結合強度を変化させ正解を導けるように自己学習していくのと同様に、学習データからノード間の結合度合いを変化

させ学習を行うことが可能な数式的なモデルである。これまでの機械学習では、入力データの特徴を人間が見つけて特徴量を設計していたが、ニューラルネットワークでは人間の脳と同様に学習データを与えることで機械的に特徴量を獲得していくことが大きな違いである。

\*9 モバイル空間統計：(株)NTTドコモの登録商標。

図4に示すように、通常よりも人口が多く集まっている時間・場所では乗客獲得率が高いことが分かっており、人口データを利用することで潜在的な需要を発見できる可能性がある。

人口の変動とタクシーの乗車需要の変動には、図5に示すように相関関係が見られる。メッシュAは人口の増加に伴い、乗車需要も高まっていることが分かる。このメッシュは駅などの他の交通機関を含ん

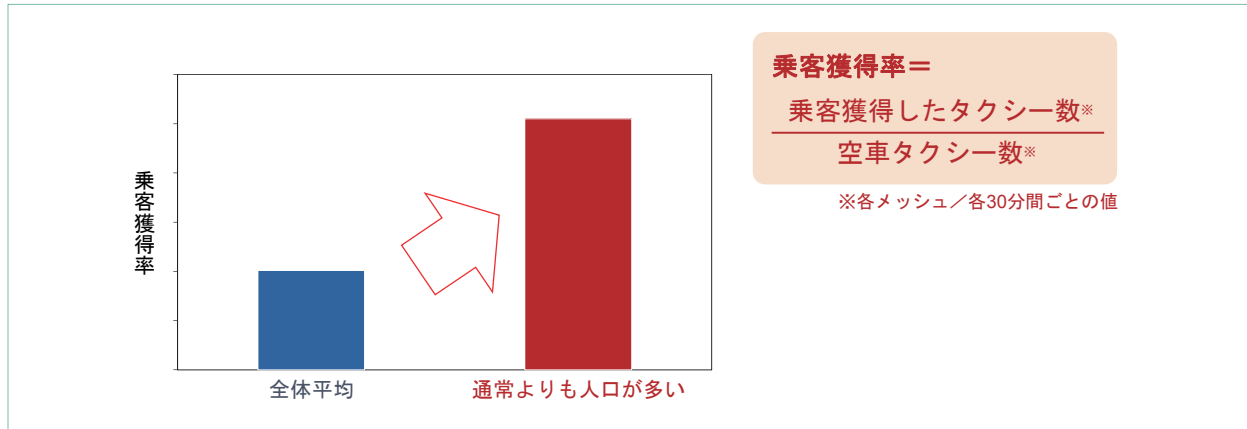


図4 タクシーの乗客獲得率

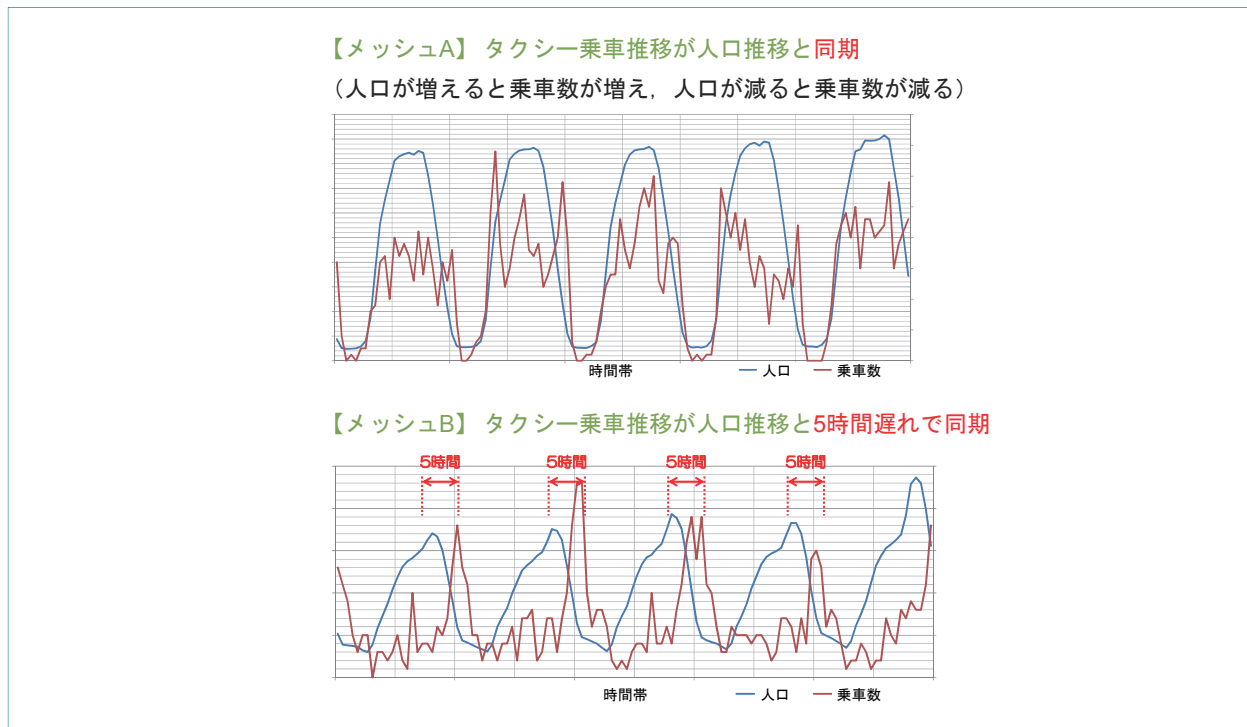


図5 タクシー乗車需要推移と人口推移の関係性

でいる可能性が高く、他のエリアから移動してきた人々の一部がタクシーに乗換えを行ったことで需要が高まっている（グラフ上で高い値になっている）と考えられる。一方で、メッシュBは人口の増加が見られてから5時間ほど経った後に乗車需要が高くなっている。このメッシュは商業施設やイベント施設などを含んでいる可能性が高く、施設に来場した人々がある程度の時間滞在した後にタクシーへ乗車して他のエリアへ移動していると考えられる。このように人口の推移を入力データとして用いることで、将来の乗車需要の推移を予測できることを示している。しかし、エリアの特性に応じて人口推移に対する乗車需要推移の遅延間隔がそれぞれ異なっており、何時間前の人口データを重視して予測を行うべきかをエリアごとに判断しなければならない。そこでこのような複雑な相関関係をもつデータを扱うために、機械的に特徴を抽出することが可能なディープラーニングも用いることとした。

## 2.3 ディープラーニングによる需要予測手法

ディープラーニングとしてさまざまなネットワーク構造が研究されてきているが、本手法ではAutoencoder [3] を発展させたSdA (Stacked denoising Autoencoders) モデル [4] を用いている。Autoencoderとは、入力されたデータを、それよりも少ない数の中間データに圧縮し、そこから再び入力データと同じデータを復元させるネットワーク構造であり、入力データを表す入力層、中間データを表す隠れ層、復元データを表す出力層の3層から成る。隠れ層では、入力データの数よりも少ない中間データに圧縮されるため、データを復元する際に重要ではないデータの重みは小さくなり、入力データを表現する上で特に重要な要素が抽出されることが期待できる。

本手法で用いているネットワーク構造では、図6の右側に示すように複数のAutoencoderを積層 (Stack) させることで、より潜在的な特徴量を効果

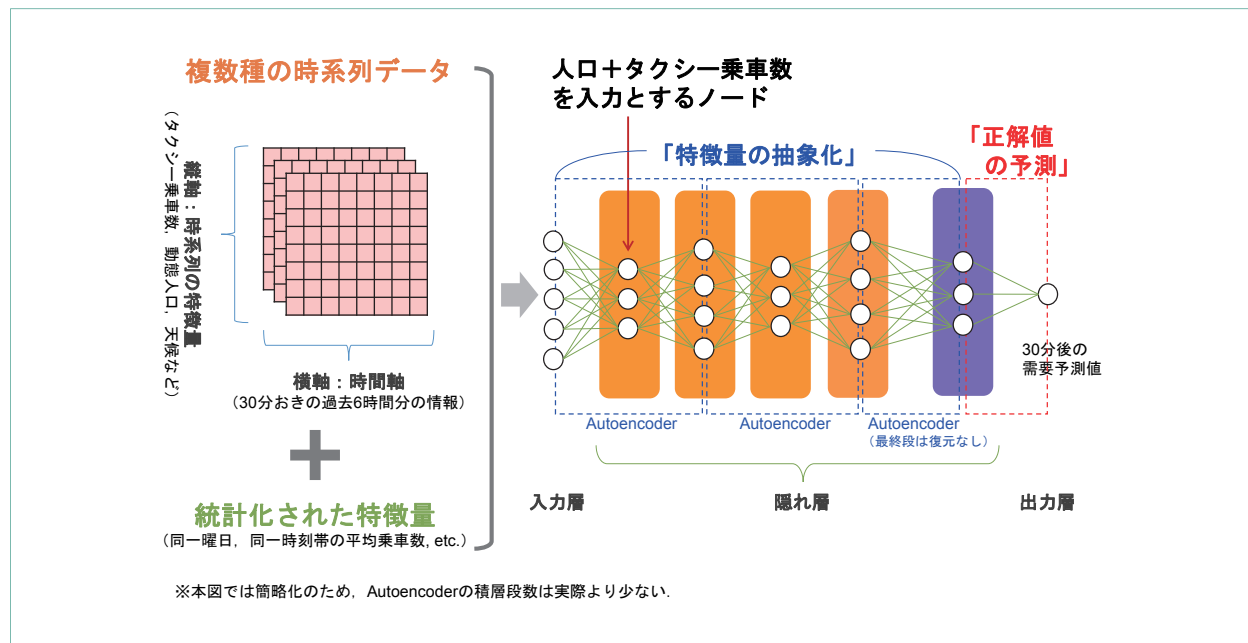


図6 タクシー乗車需要予測におけるネットワーク構造

的に抽出することができ、これにより精度向上を実現している。また、需要予測を回帰問題\*10として扱うために、最も重要な特徴量が抽出されている最終段のAutoencoderの隠れ層に回帰の予測器\*11を入れることで、最終的には需要予測結果の数値を出力する。図6の左側に示すように、過去のタクシー乗車実績データや人口などの時系列データに加え、同一曜日同一時刻帯の平均乗車数のような統計値も入力データとして入力層に与えており、複数の隠れ層を経由することで重要な特徴が抽出される。最後の隠れ層では各データに対し学習した重み値が掛け合わされ、その結果が足し合わされることから出力層には回帰での需要予測結果が出力される。なお、SdAモデルを利用した学習処理では、大量の学習用データから時間情報と空間情報の関係性が各ノード間のリンクにおける重み値として機械的に学習される。

また、Autoencoderの入力にノイズを加えて学習させることでノイズを除去する（denoising）特性も持たせている。Autoencoderでは入力データを復元しようとするが、欠損（またはノイズ）が加わったデータを入力データとして与え、欠損（またはノイズ）を加える前の元データを復元するよう学習を行うことで、入力データがもともとノイズや欠損の影響を受けづらくなることや、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。

### 3. 実証実験

本技術を用いて、東京でのタクシー需要を予測する実験を実施した。各500mメッシュに関して、ある対象の時刻から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できたかを計算する。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。その他のデータの詳細は表1のとおりである。運転席脇に設置したタブレット上に図7のような画面表示を行い、地図上の赤破線で区切られたエリア内での乗車需要回数の予測結果が数値で示されており、ドライバーは提示された需要の予測値が高いエリアを走行した。

実証実験期間における本技術の予測精度を検証したところ、エリアによって多少の上下はあるが93～95%となっており、高精度に需要を予測できていることが確認できた。実際の乗車実績と予測結果との比較を図8に示す。実証実験前の月と比較した実証実験対象月の売上上昇幅が、実証実験のすべての期間において、全ドライバーでの上昇幅平均値よりも需要予測システムを利用したドライバーの上昇幅平

表1 実証実験で用いたデータ

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	4,400台
	取得頻度	5～10秒ごと
人口データ	解像度	500m
	取得頻度	10分ごと
雨量データ	解像度	250m
	取得頻度	10分ごと

\*10 回帰問題：目的となる変数を導出するにあたり、影響を与える変数で表される式で回帰として扱うことが可能な問題。

\*11 予測器：与えられた入力をもとに、次に出現するものを推測する装置。

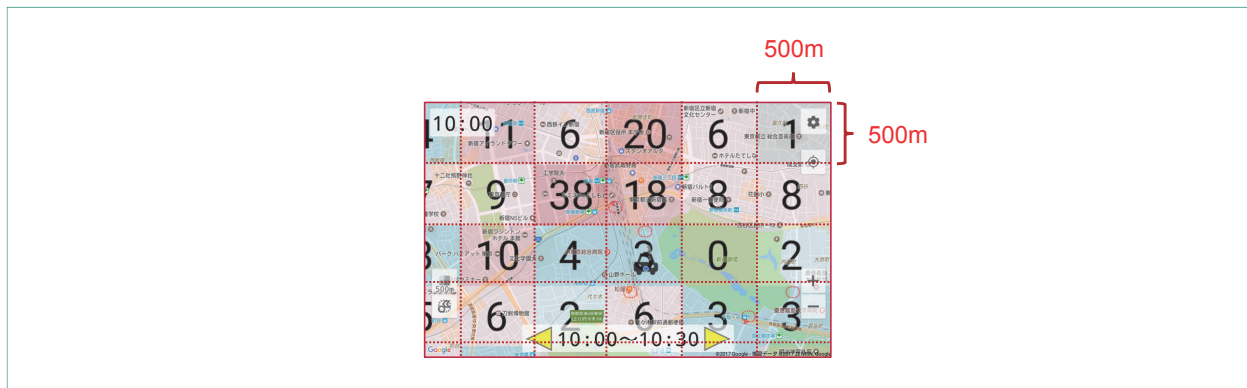


図7 需要予測結果の表示画面例

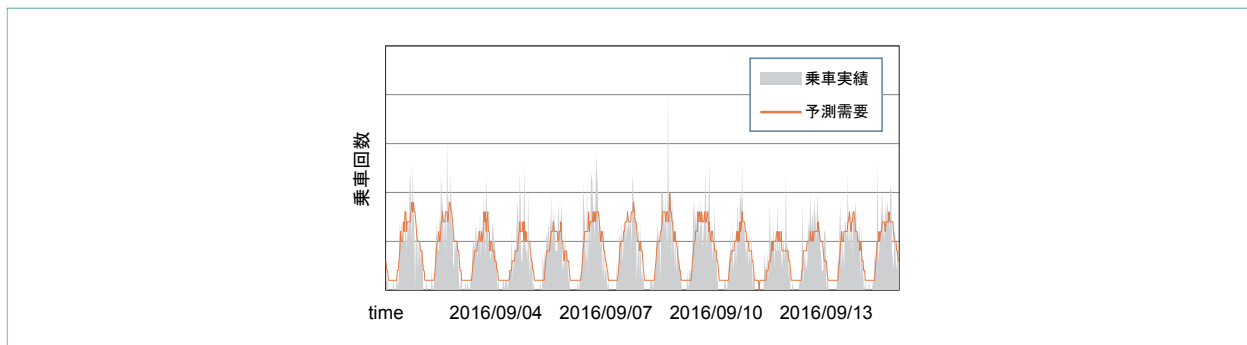


図8 需要予測結果と乗車実績の比較

均値が上回っており、1日あたり1人平均で約1,400円の売上増が確認できた。また、乗客獲得の効率性を示す指標である実車率の向上も見られた。

#### 4. あとがき

本稿では、AIタクシーサービスで用いている、タクシーの将来の乗車需要を予測する技術について、その詳細を解説した。運行の効率化を実現するためには、需要と供給のマッチングが重要となる。本技術により、あらかじめ需要が予測できていれば、それにあわせて空車タクシーの供給を最適化することが可能となる。今後は、全体最適を考慮した配車制御の仕組みについても検討を行っていきたい。

#### 文献

- [1] AIタクシーサービスWebサイト。  
<https://www.nttdocomo.co.jp/biz/service/aitaxi/>
- [2] 岡島, ほか: “特集: 社会・産業の発展を支える「モバイル空間統計」—モバイルネットワークの統計情報に基づく人口推計技術とその活用—,” 本誌, Vol.20, No.3, pp.6-44, Oct. 2012.
- [3] G. E. Hinton and R. R. Salakhutdinov: “Reducing the dimensionality of data with neural networks,” Science, Vol.313, Issue 5786, pp.504-507, 2006.
- [4] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio and P. A. Manzagol: “Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion,” Journal of Machine Learning Research, pp.3371-3408, Dec 2010.