

5 AI タクシー：リアルタイム人口を用いたタクシー乗車の需要予測 — AI タクシーサービスへの深層学習の適用 —

石黒 慎 | (株) NTT ドコモ

リアルタイムな人口データを活用した タクシー乗車の需要予測

タクシードライバーの収益を向上させ、また、タクシー利用者の待ち時間を軽減するには、タクシードライバーの空車運転時間を減らし、運行を効率化することが必要である。タクシーの需要は周辺施設や状況に応じてさまざまに変化する。すべての場所と時間について、人手で網羅的に需要を予見することは難しい。そこで、タクシー乗車需要にかかわるさまざまなデータから機械的に乗車需要を予測し、タクシードライバーに配車エリアを推薦することで運転効率を向上させることを考える。我々は、モバイルネットワークの仕組みを用いることで、リアルタイムにエリアごとの人口を推定し、人口の推移からタクシーの乗車需要を予測する深層学習方式の開発を行った。

我々が利用したのはモバイル空間統計の計算を高速化し、より少ない遅延で取得できるリアルタイム人口統計データである。モバイル空間統計とはモバイルネットワークの基地局と携帯電話の位置関係を分析することで、携帯電話ユーザの位置情報を推定し、さまざまな統計処理を行うことでエリアごとの総人口を推定する技術である。モバイル空間統計の高速版では、直近 10 ～ 20 分の 500m ごとの人口を推定することが可能である。図-1 に三軒茶屋駅におけるタクシーの乗車数と人口の時系列遷移を比較したグラフを示す。図からタクシーの乗車数が人口の増加から 5 時間ほど遅れて増加していることが分かる。このエリアでは人口がピークから減少し始め

るに従って、タクシーの乗車が増え、人口が最も低くなったときに乗車がなくなることが分かる。おそらく、終電後にタクシー利用が増えていると考えられる。

人口の増減は人が移動をする際に生じるものであり、移動の中にはタクシー乗車による移動も含まれている。上述の例のようにタクシーと人口は互いに相関していると考えられるが、実際にはそれ以外の情報、たとえば、エリア特性や気象、イベントといったさまざまな状況が複雑に関係あって、需要が生じていると考えられる。このため、種々のデータを用いて、データ間の関係性を高度に抽象化し、タクシー需要にかかわる重要な情報を抽出する手法が必要となる。そこで、我々はディープラーニングによる需要予測モデルに着目した。

深層学習を用いた需要予測モデル

深層学習を用いたタクシー乗車の需要予測モデルについて説明する。ここでは、予測する対象を 500m のグリッドごとに、現在から 30 分先までのタクシーの乗車回数の合計を予測する設計とした。予測にあたって

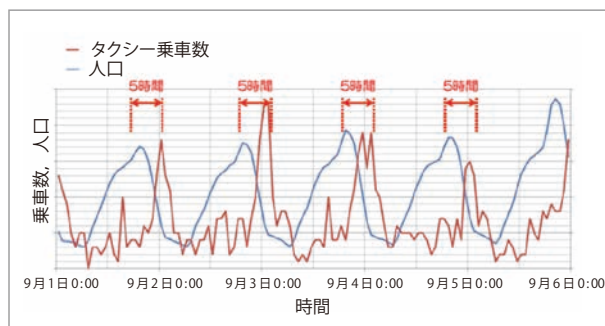
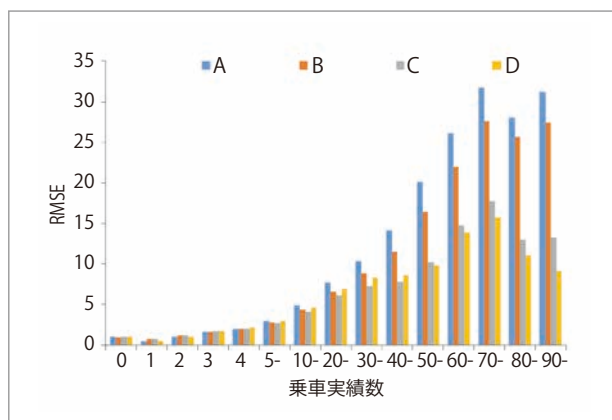


図-1 三軒茶屋駅におけるタクシー乗車数と人口の遷移
(2016年9月1日～5日)

用いたデータは、時刻、タクシー、人口、雨量の4種のデータである。時刻データには時刻と休祝日の情報を用いた。タクシーデータは、各車両に取り付けられたGPS端末と運行記録端末により、顧客の乗車降車が行われた時刻と緯度経度が記録される。ここから、500mごと30分幅で、乗車回数と降車回数の累積回数を求めた。人口データは、500mごと、10分ごとの各エリアの総人口を推定したリアルタイム人口統計データを用いた。雨量データは、500mごと10分ごとの降水量を利用した。

これらを入力データとして、Stacked denoising Autoencoders (SdA)による需要予測モデルの実装を行った。SdAは、入力データの次元数を隠れ層で圧縮し、出力層で次元数を元に戻すことで、隠れ層中でより重要なデータの抽出を実現するAutoencoderの機能と、入力データにノイズを加えた上で、出力層で元のデータに復元させるノイズ除去を行うことで、ノイズではない重要な情報を優先して抽出するdenoising autoencoderの機能を持つ。SdAでは、このようなネットワークを積層することで、より重要な情報の抽出を行う。また、SdAにより得られたネットワークの重みを用いて、新たに最終層に回帰層を加えることで、需要予測問題を解く構造とした。

図-2に異なる入力データを用いた学習によるSdAの予測結果を示す。ここでは、タクシーデータのみ、



■図-2 異なる入力データによるSdAの乗車実績値ごとの予測精度
A: タクシーデータ, B: タクシーおよび人口データ, C: タクシーおよび天候データ, D: 全データ

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、すべてのデータの4種類の学習を行った。なお、時刻情報はすべてのモデルに用いた。結果から、タクシーデータ単体での学習に比べて、人口データおよび天候データを加えた場合に性能の向上を確認できる。これは、タクシーデータのみでは説明できない、人口や天候に含まれるタクシー需要に関する重要なデータがSdAにより抽出され、予測に利用されていると考えられる。また、グラフ中のAとBまたは、CとDを比較すると乗車実績数の大きい高需要の場所において、より大きな割合で誤差が軽減することが分かる。このことから、人口データは高需要な場所で性能向上に寄与することが確認できる。

以上、SdAによる需要予測問題の予測精度について示した。最後に、提案法を実装する上で苦労した点として、モデル性能向上のためのチューニングについて感想を述べる。ディープラーニングを新たな問題に適用する場合、選択したモデルによっては直ちには望ましい性能が出ない可能性がある。我々のアプリケーションにおいても純粋なSdAの実装のみでは十分な性能は発揮できず、ハイパーパラメータ探索や特徴量設計など、さまざまなチューニングの工夫を行っている。ディープラーニングでは、データと問題設計、利用モデルによって必要なチューニングも異なってくる。したがって実際の実装では、適用したモデルにおいて直ちに十分な性能が得られずともチューニングを試行錯誤してみることも重要だと考える。今後の予定として、タクシー需要予測のさらなる精度向上を図るため、CNN (Convolutional Neural Network) やRNN (Recurrent Neural Network) のアーキテクチャを組み合わせることで、時空間的にタクシーや人口などの需要の変化を予測するモデルを利用することが考えられる。そこで、より高度なネットワーク構造を用いたモデルの開発に取り組むことを考えている。

(2018年7月31日受付)

■石黒 慎 (正会員) shin.ishiguro.tb@nttdocomo.com

(株)NTTドコモAIスペシャリスト。人流データ分析に基づいた位置情報サービスに関する研究開発に従事。