

LSTM ネットワークによる GPS データのモデル化と状態推定

城内 光平[†] 江口 浩二^{††}

[†] 神戸大学システム情報学研究科 〒 657-0013 兵庫県神戸市灘区六甲台町 1-1

^{††} 神戸大学システム情報学研究科 〒 657-0013 兵庫県神戸市灘区六甲台町 1-1

E-mail: †shirouchi@cs25.scitec.kobe-u.ac.jp, ††eguchi@port.kobe-u.ac.jp

あらまし 近年, GPS 移動軌跡データを分析することにより, 移動状態推定や人流予測を実現する試みが注目されつつある. この目的のもと, 従来は隠れマルコフモデル (HMM) などが用いられてきた. 本研究では, 再帰型ニューラルネットワーク (RNN) の一種で, 短期依存性だけでなく長期依存性をも捉えることが可能な長短期記憶ネットワーク (LSTM) を用いて GPS 移動軌跡データによる移動状態推定を実現し, 有効性の評価を行う.

キーワード 移動状態推定, GPS データ, LSTM, RNN

1. はじめに

近年, ディープラーニングの技術を用いた研究に対する注目が高まっている. 画像認識のコンテストでディープラーニングの技術を用いたグループが他のグループを大きく離して優勝したことをきっかけにしてディープラーニングの有効性が大きく話題になった. ディープラーニングの特徴は, 生のデータを与えるとモデルが自動的に特徴抽出から学習までを行ってくれるという点であり, 画像データや音声データ, GPS データなどをはじめとした様々なデータへの応用が考えられている.

本研究では特に時系列で変化するデータに対してディープラーニングにおいて検証する. 時系列データを解析する研究としては, GPS 移動軌跡から速度変化を利用して旅行者の現在の状態を推測する研究 [1] や隠れマルコフモデル (HMM) や Support Vector Machine (SVM) を用いて GPS 移動軌跡データからどのような手段で移動しているのかを分類する研究 [2], [3] や GPS の生のデータから Conditional Random Field (CRF) や SVM を用いて移動者の状態を予測する研究 [4] など, GPS 移動軌跡データを用いて状態を予測することには様々なニーズがあり, 様々な手法での研究がなされている. しかし, ディープラーニングの手法を GPS 移動軌跡データに適用し人流データを予測, 解析した研究事例は私の知る限りあまりない. 本研究の目的はディープラーニングの手法を用いて GPS 移動軌跡のデータを解析し, 状態を予測することである.

時系列データを扱うモデルとして再帰型ニューラルネットワーク (Recurrent neural network ;RNN) [5], [9] がある. しかし, RNN では勾配消失問題により長い系列のデータを扱うことが容易ではない. 長い系列のデータを扱うために RNN を拡張したモデルとして Long Short-Term Memory (LSTM) [6] がある. 本研究では LSTM ネットワークを用いて学習を行い, 移動者の状態の推定を行う.

2. 関連研究

本章では最も基本的な順伝播型ニューラルネットワーク (Feed-

forward neural network) [7] と, その時系列データのための拡張モデルに当たる再帰型ニューラルネットワーク (Recurrent neural network) 及び, Long Short-Term Memory (LSTM) [8] について述べる.

2.1 順伝播型ニューラルネットワーク

順伝播型ニューラルネットワーク (Feedforward neural network; FFNN) は, 層状に並べたユニットが隣接層間でのみ結合した構造を持ち, 情報が入力側から出力側に一方向にのみ伝播するニューラルネットワークである. ネットワークを構成する各ユニットは複数の入力を受け取り, 1 つの出力を計算する. ここで, ユニットの 4 つの入力 x_1, x_2, x_3, x_4 を受け取るとすると, その総入力 u は各入力にそれぞれ異なる重みを掛けたものを加算し, バイアス b を加算したものとなる.

$$u = w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + b \quad (1)$$

ユニットの出力 z はユニットへの総入力 u に対する活性化関数と呼ばれる関数 f の出力となる. 活性化関数にはシグモイド関数やソフトマックス関数などがよく用いられる.

2.2 再帰型ニューラルネットワーク

系列データは個々の要素が順序を持ち, その並びに意味が隠されているようなデータのことである. 系列の並びが時刻の並びとなっている系列データを特別に時系列データと呼ぶ. 系列データの例としては音声の波形, 動画, 文章 (単語列) などがある. それにより, 系列データの分類問題においてはある時刻, もしくは位置におけるデータの前後におけるデータとの関係性を上手く考慮することができれば, 結果が向上することが知られている. そして FFNN を系列データを扱うために拡張したモデルとして再帰型ニューラルネットワーク (Recurrent neural network ;RNN) がある. RNN における分類問題は入力の系列 $\mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^3, \dots$ から正解の系列 $\mathbf{y}^1, \mathbf{y}^2, \mathbf{y}^3, \dots$ を推定する問題として定式化することができる. このとき時刻 t における出力 \mathbf{y}^t はそれ以前の入力 $\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^t$ の影響を受けていると考えることができる.

RNN は内部に有向な閉路を持つニューラルネットの総称で

ある。RNNはこの構造により情報を一時的に記憶し、また振る舞いを動的に変化させることができる。これにより、系列データ中に存在する時間的な依存関係である「文脈」を捉えることができ、系列データに対して有効に処理することができる。RNNのグラフィカルモデルを図1に示す。

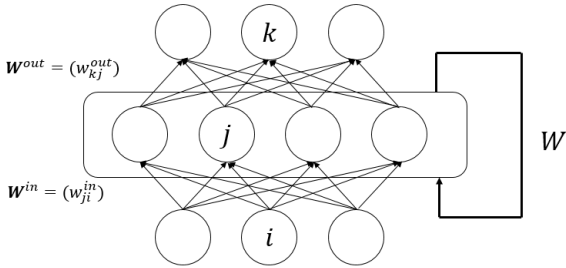


図1 Recurrent neural network.

図1にあるようにRNNは順伝播型ネットワークと同様の構造を持ち、ただし中間層のユニットの出力が自分自身に戻される「帰還路」を持っている。RNNの動作は各時刻 t につき1つの入力 \mathbf{x}^t を受け取り、また同時に1つの出力 \mathbf{y}^t を返すというものである。RNNは理論上、過去のすべての入力から1つの出力への写像を表現する。

系列データの分類問題においては、RNNの出力層は対象となるクラス数と同じ数のユニットを並べ、ソフトマックス関数を活性化関数に選ぶ。誤差関数は交差エントロピーを用いる。出力層 L の k 番目のユニットの出力はその下の層 $L-1$ 層からの入力を用いて以下のように定義される。

$$y_k \equiv z_k^{(L)} = \frac{\exp(u_k^{(L)})}{\sum_{j=1}^K \exp(u_j^{(L)})} \quad (2)$$

ここで、 $u_k^{(L)}$ は出力層の1つ下の層からの入力である。またこの出力 y_1, \dots, y_K は、総和が1になることに注意が必要である。次に、誤差関数は以下のように定義される。

$$E(\mathbf{w}) = \sum_n \sum_t \sum_k d_{nk}^t \log y_k^t(\mathbf{x}_n; \mathbf{w}) \quad (3)$$

ここで、 d_{nk}^t は n 番目のサンプルの時刻 t での目標出力、 y_k^t はそれと比較されるRNNの出力、 \mathbf{x}_n は n 番目のサンプルの入力系列である。RNNでは、中間層に十分な数のユニットがあれば、任意の系列から系列への写像を、任意の精度で近似することができることが証明されている。

2.2.1 順伝播計算

RNNにおいて入力系列から出力系列を得る計算を示す。図1に示すように、入力層、中間層、出力層の各ユニットのインデックスを i, j, k とする。またネットワークへの入力を \mathbf{x}^t 、中間層ユニットへの入力を \mathbf{u}^t 、中間層ユニットの出力を \mathbf{z}^t 、出力層ユニットへの入力を \mathbf{v}^t 、出力層ユニットの出力を \mathbf{y}^t とする。

RNNでは、中間層の出力を自らの入力に戻すため、時刻 t

における中間層の各ユニットへの入力は、同時刻 t にて入力層から届くものと、時刻 $t-1$ の中間層の出力をフィードバックしたものとの和となり、以下の式で表すことができる。

$$u_j^t = \sum_i w_{ji}^{(in)} x_i^t + \sum_{j'} w_{jj'} z_{j'}^{t-1} \quad (4)$$

ここから中間層の出力は、この入力に対して活性化関数を計算した値となるため、活性化関数を f とすると以下のように表すことができる。

$$z_j^t = f(u_j^t) \quad (5)$$

式(5)を繰り返し計算することで任意の時刻 t における中間層の状態 \mathbf{z}^t を求めることができる。

出力層への入力 \mathbf{v}^t は、中間層の出力 \mathbf{z}^t を用いて以下の式で求めることができる。

$$v_k^t = \sum_j w_{kj}^{(out)} z_j^t \quad (6)$$

また出力層の出力 \mathbf{y}^t は出力層の活性化関数を $\mathbf{f}^{(out)}$ とすると以下のように表すことができる。

$$\mathbf{y}^t = \mathbf{f}^{(out)}(\mathbf{v}^t) \quad (7)$$

2.3 Long Short-Term Memory(LSTM)

RNNは系列データの文脈を捉えて推定を行うことができる。このとき、捉えることのできる文脈の長さ、すなわち現時刻からどれだけ遠い過去の入力を出力に反映させることができるかは重要である。しかし、RNNで実際に出力に反映することができるのは高々過去の10時刻分程度であると言われている。この限界は勾配消失問題により生じている。層の数の多い深いネットワークにおいて、誤差逆伝播法によって勾配を計算するとき層をさかのぼるにつれて勾配の値が爆発的に大きくなるか、あるいは0に消滅してしまう。長期にわたる記憶を実現するために提案されたモデルがLong Short-Term Memory(LSTM)である。LSTMでは上で述べたようなRNNに対し、その中間層の各ユニットをメモリユニットと呼ぶ要素で置き換えた構造を持っている。その他の構造は通常のRNNと同様である。メモリユニットの構造を図2に示す。

中央にメモリセル(図中記号a)があり、その周囲に5つのユニットが配置されている。メモリセルaは状態 s_j^t を保持し、これを1時刻隔ててメモリセル自身に帰還させることで記憶を実現している。この帰還路には途中で忘却ゲートが挿入されており、ユニットfの出力がゲートの値 $g_j^{F,t} \in [0, 1]$ となる。 s_j^t に $g_j^{F,t}$ をかけたものが伝えられ、 $g_j^{F,t}$ の値が1に近ければ現状態がそのまま記憶され、0に近ければリセット(忘却)される。

メモリユニットへの外部からの入力はユニットbが受け取り、

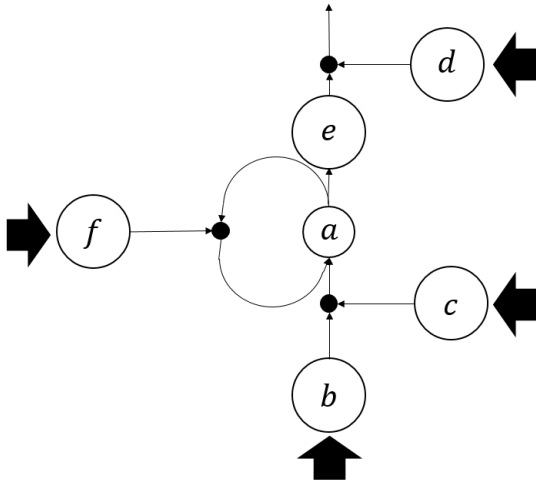


図2 Memory unit.

その出力がメモリセルに入力される。その間には入力ゲートがあり、ユニット c の出力がゲートの値 $g_j^{I,t} \in [0, 1]$ となる。ユニット b は通常の RNN の中間層のユニット 1 つに相当しており、これにゲートの値 $g_j^{I,t}$ をかけたものがメモリセルに伝えられる。

メモリユニットから外部への出力は、メモリセルからユニット e を経て行われる。出力との間には出力ゲートが挿入されており、ユニット d の出力がゲートの値 $g_j^{O,t} \in [0, 1]$ となる。この値が 1 に近ければメモリセルの出力が外部に伝達される。

3. LSTM ネットワークによる時系列データの分類

LSTM ネットワークに対して GPS 移動軌跡より得られる種々のデータ (マルチモーダルデータ) を入力として与え、対応する状態ラベルを推定する。ここで、状態ラベルは入力として与えたあるデータと対応する移動状態のラベルを指す。GPS 移動軌跡は多くの外れ値や欠損値を持っており、また広範囲に分布しているため生のままの GPS 移動軌跡を用いるとスパース性が問題となる。外れ値や欠損値の影響を少なくするため GPS 移動軌跡から速度を計算し、その系列を実験に用いた。また GPS 移動軌跡にはそのデータのユーザ ID や絶対的な時間など位置情報以外にも複数の情報が含まれている。マルチモーダルな入力データを扱うことで推定をより正確にできることが考えられる。実験で用いた LSTM ネットワークを図 3 に示す。

ここで、 x は入力データであり、インデックス n は入力データの次元数を表している。入力データとしては速度や加速度、ユーザ ID や位置情報などが考えられ、例えば速度とユーザ ID を入力とするならば $n = 2$ となる。 s^t は中間層の状態であり、中間層はメモリユニットで構成されている。 y^t は LSTM ネットワークの出力を表している。移動状態のラベル集合を $\mathcal{L} = \{l_1, l_2, \dots, l_\ell\}$ とすると目標出力は正解と対応するラベルの値のみが 1 をとり、その他の要素が 0 となる ℓ 次元のベクトルである。

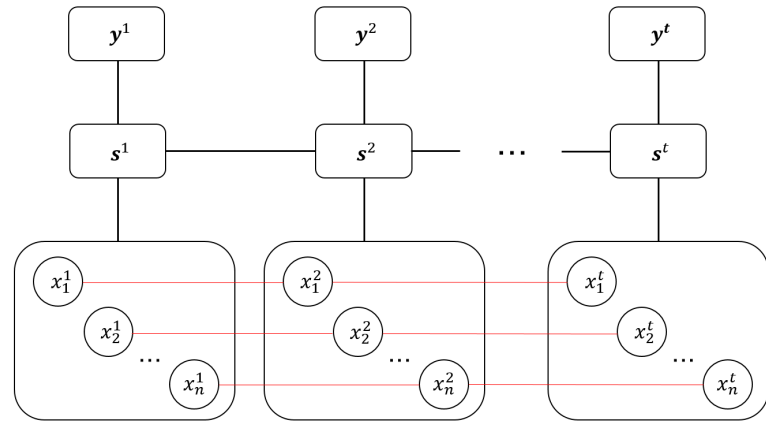


図3 LSTM network for multimodal data.

4. 実験

本章では GPS データから求めた速度データを用いた LSTM ネットワークによる状態推定の実験について述べる。速度データを用いて状態の推定を行い、その精度を測定し、性能を評価する。

4.1 データセット

本研究では、Microsoft Research Asia Geolife project により集められた、GPS センサを持った人間の移動の GPS 軌跡に関するデータ^(注1)を使用する。データはおおよそ 1-5 秒おき、もしくは 5-10 メートルおきに取得されている。データの内容としてはユーザ ID、経度、緯度、高度、取得時間の情報を持っている。また一部のデータにはラベルが付与されており、本実験ではラベルが付与されているデータを抽出し使用している。ラベルの種類としては、車、自転車、徒歩、駆け足、バス、地下鉄、電車、飛行機、ボート、バイク、タクシー、路面電車の 12 種類が存在している。また実験は速度のみを用いたものと、速度及びユーザ ID を用いて状態推定を行ったものの 2 種類について行った。

4.2 パラメータ設定

実験は単一の中間層を持つ LSTM ネットワークで行った。LSTM ネットワークに対して速度データ、あるいは速度データとユーザ ID を入力として与え、移動状態を推定しその正解率を評価する。入力層は速度だけを用いる場合には、ユニット数が各タイムステップにおいて 1 であり、速度とユーザ ID を用いた場合にはユニット数は 2 となる。中間層のユニット数は 200 とし、活性化関数は \tanh とした。出力層はラベルの種類が 12 種類のためユニット数を 12、活性化関数をソフトマックス関数とした。誤差関数は交差エントロピーを用い、学習は確率的勾配降下法により行った。

4.3 結果

実験の結果を表 1 に示す。

ここで、ランダムは 1-12 までの数字をランダムに生成し、そ

(注1) : <http://research.microsoft.com/en-us/downloads/b16d359d-d164-469e-9fd4-daa38f2b2e13/>

表 1 Accuracy results.

	accuracy
LSTM(速度)	0.3186 ± 0.00036
LSTM(速度+ID)	0.4337 ± 0.00031
ランダム	0.0823

の正解率について求めたものである。表 1 において、LSTM(速度) は LSTM ネットワークに対して速度の系列のみを与えて予測を行った結果を示しており、LSTM(速度+ユーザ ID) は速度の系列と合わせてユーザ ID の系列を入力として与えたものである。ランダムは 0 から 11 までの数字を一様分布よりランダムに求め、それにより正解ラベルと一致した個数となっている。結果としては、LSTM ネットワークを用いて予測した結果、完全にランダムに予測した結果よりも良い結果を残すことができた。また、速度の系列のみを与えた場合よりも速度の系列に加えてユーザ ID を与えた場合により良い結果を残すことが分かった。

5. おわりに

本論文では再帰型ニューラルネットワークの拡張である LSTM ネットワークを用いて、GPS 移動軌跡データによる時系列で与えられた速度データを用いた状態推定を行った。実験においては、LSTM ネットワークを用いて予測を行った結果がランダムにラベルを選択した場合よりも精度が高い結果となった。また、速度だけではなく速度にユーザ ID を付与した場合の方がより高い精度を得た。これにより、LSTM ネットワークが速度データのラベルによる違いを学習することができたことを示すことができた。また速度データとユーザ ID をともに用いたことで精度が上がったことは、今回のデータセットにおいては、人によるデータの数の違いやラベルの種類の偏りが大きかったため、ユーザ ID を付与したことであるユーザ ID とラベルとの間の関係を学習したのではないかと考えられる。

謝 辞

本研究の一部は科学研究費補助金基盤研究 (B) (15H02703) の援助による。

文 献

- [1] Hidekazu Kasahara, "Tourist Status Estimation from GPS Trajectory using Environmental Constraints", the Japan Society of Artificial Intelligence(2015).
- [2] Yong-Joong Kim, Sung-Bae Cho, "A HMM-Based Location Prediction Framework with Location Recognizer Combining k-Nearest Neighbor and Multiple Decision Trees" Volume 8073 of the series Lecture Notes in Computer Science pp 618-628(2013)
- [3] Bolbol, A., Cheng, T., Tsapakis, I., Haworth, J., "Inferring hybrid transportation modes from sparse GPS data using a moving window SVM classification", Computers, Environment and Urban Systems.,pp.526-537(2012)
- [4] Yu Zheng and Xing Xie, "Learning Transportation Mode from Raw GPS Data for Geographic Application on the Web", WWW(2008)
- [5] Felix A. Gers,"Learning to Forget: Continual Prediction with LSTM", Neural computation Vol. 12, No. 10, Pages: 2451-2471(2000)
- [6] Goller, C.; Kuchler, A. "Learning task-dependent distributed representations by backpropagation through structure". Neural Networks, IEEE, (1996).
- [7] F.Rosenblatt, "The perceptron:A probabilistic model for information storage and organization in the brain", Psychological Review,Vol.65,No.6,pp.386-408,(1958)
- [8] Takayuki Okatani,"Deep Learning"(2015)
- [9] T.Mikolov,M.Karafiát,J.Cernocký,andS.Khudanpur,"Recurrent neural network based language model.", Proc.Interspeech,(2010).