

# 東京大学 量子ソフトウェア寄付講座

## 第1回 量子ソフトウェア産学協同ゼミ

2022年 3月 2日

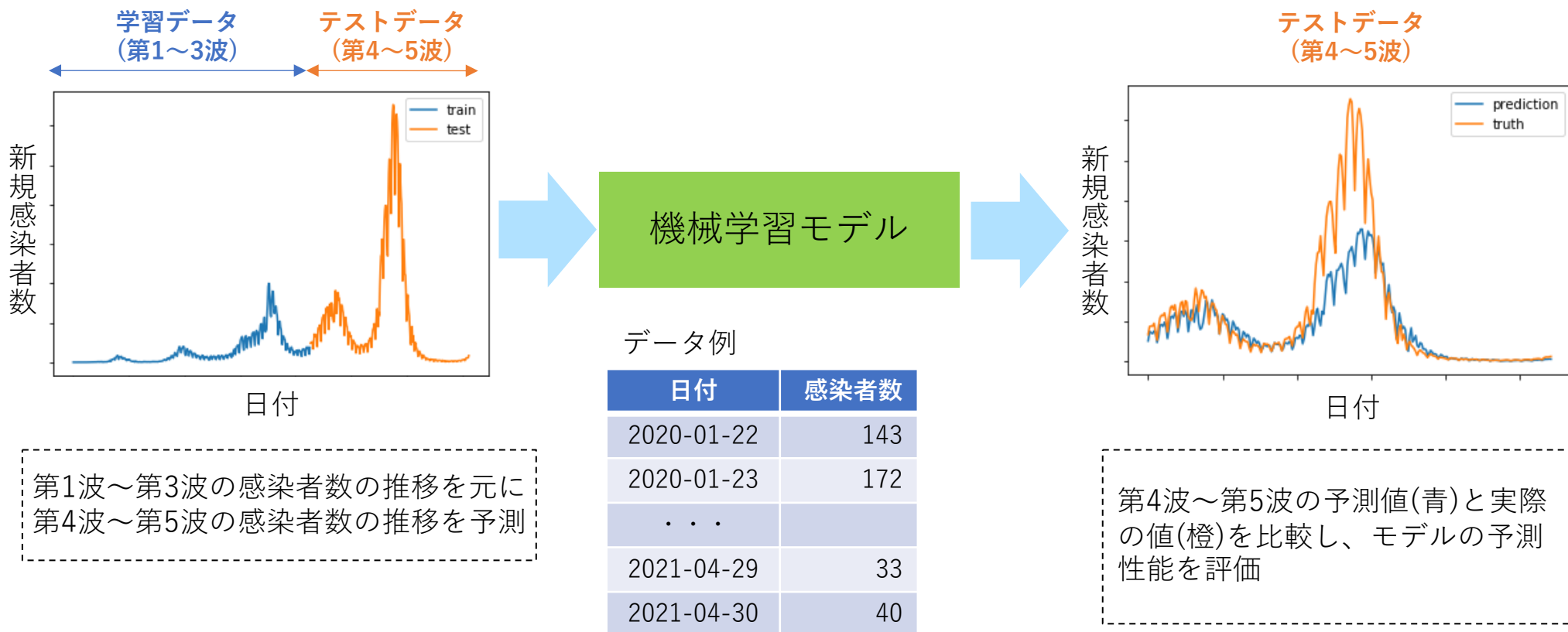
# ■ 本日のアジェンダ

---

1. はじめに
2. テンソルネットワークを使ったパラメータ圧縮
3. 演習課題と演習環境
4. RNN (Recurrent Neural Network) による時系列予測モデル
5. **【演習】** テンソルネットワークによるパラメータ圧縮  
**【討論】** パラメータ圧縮と予測結果の関係について
6. 結果共有、模範解答の確認・解説
7. **【演習】** 発展課題
8. 結果共有、模範解答の確認・解説
9. 総評、まとめ

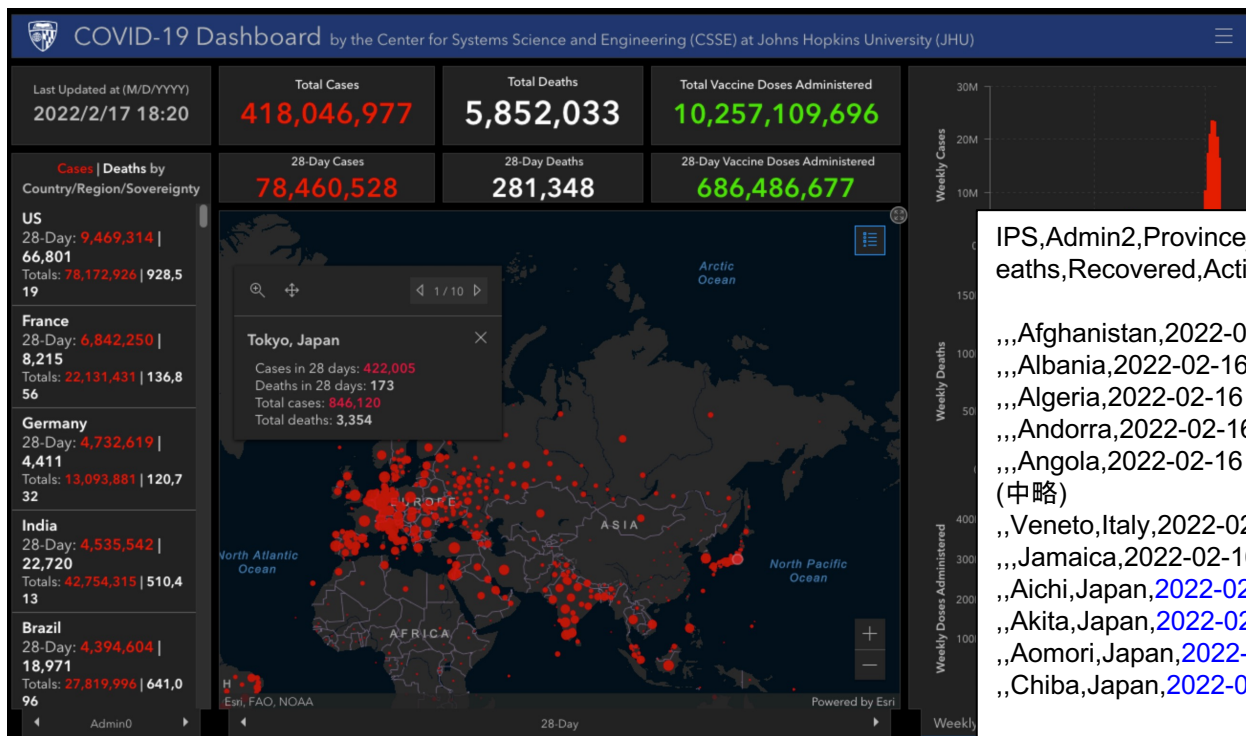
## ■ 演習課題

- COVID-19の過去の新規感染者数の推移から、未来(7日後)の新規感染者数を予測する。
  - 日本国内の日次の新規感染者数を対象とする。
  - 2020年1月22日～2021年4月29日 の期間で新規感染者数が既知である。(学習データ)
  - 2021年4月30日～2022年1月30日 の期間の新規感染者数の推移を予測する。(テストデータ)
- 予測には、時系列データの扱いに適した機械学習モデルを活用する。



## ■ 使用するデータについて

- データソースは Johns Hopkins 大学が github で公開しているcsv形式のもの
- 演習では、これを Shad Reynolds 氏が国別に集計したものを使用



COVID-19 Dashboard

IPS,Admin2,Province\_State,Country\_Region,Last\_Update,Lat,Long\_,Confirmed,Deaths,Recovered,Active

```
,,,Afghanistan,2022-02-16 04:21:10,33.93911,67.709953,171422,7504,,,
,,,Albania,2022-02-16 04:21:10,41.1533,20.1683,268491,3420,,,
,,,Algeria,2022-02-16 04:21:10,28.0339,1.6596,262570,6753,,,
,,,Andorra,2022-02-16 04:21:10,42.5063,1.5218,37361,150,,,
,,,Angola,2022-02-16 04:21:10,-11.2027,17.8739,98555,1899,,,
(中略)
,,,Veneto,Italy,2022-02-16 04:21:10,45.43490485,12.33845213,1276773,13607,,,
,,,Jamaica,2022-02-16 04:21:10,18.1096,-77.2975,127233,2750,,,
,,,Aichi,Japan,2022-02-16 04:21:10,35.035551,137.211621,254979,1367,,,
,,,Akita,Japan,2022-02-16 04:21:10,39.748678999999996,140.408228,6973,29,,,
,,,Aomori,Japan,2022-02-16 04:21:10,40.781541,140.828896,15773,50,,,
,,,Chiba,Japan,2022-02-16 04:21:10,35.510141,140.198917,216185,1141,,,
```

csvのイメージ

(参照URL)

<https://www.arcgis.com/apps/dashboards/bda7594740fd40299423467b48e9ecf6> (Dashboard)

<https://github.com/CSSEGISandData/COVID-19> (csvデータ)

## ■ 演習環境

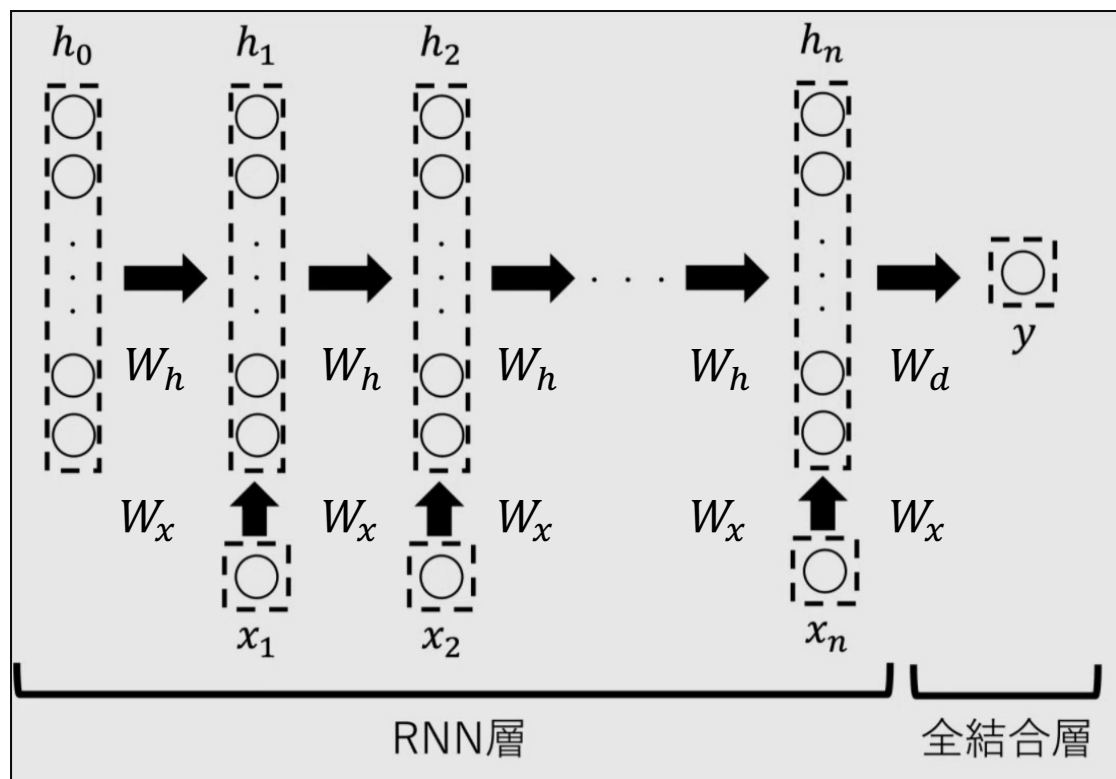
- 演習では Python を使って機械学習モデルを実装
- Google Collaboratory を使用することで環境構築作業が不要
- 事前に配布した Jupyter Notebook を Googleドライブに格納して開けばよい。

(参考) 独自に環境を構築する場合  
以下のパッケージをインストールしてください。 ※ バージョン指定は特になし

パッケージ名	用途
pandas	csv形式のファイルを読み込む
NumPy	行列・ベクトルの演算を行う
matplotlib	感染者数の推移等のグラフを描画する
TensorFlow	機械学習モデルを実装する
Keras	TensorFlowを使いやすくするライブラリ

## ■ 機械学習モデルについて

- 機械学習モデルは時系列予測に適したニューラルネットワークである RNN (Recurrent Neural Network) を使用
- RNN は Keras を用いて予め実装済 (本演習の対象外)



$x_t$  : 時刻 $t$ の入力

$y$  : 出力

$h_t$  : 隠れ層

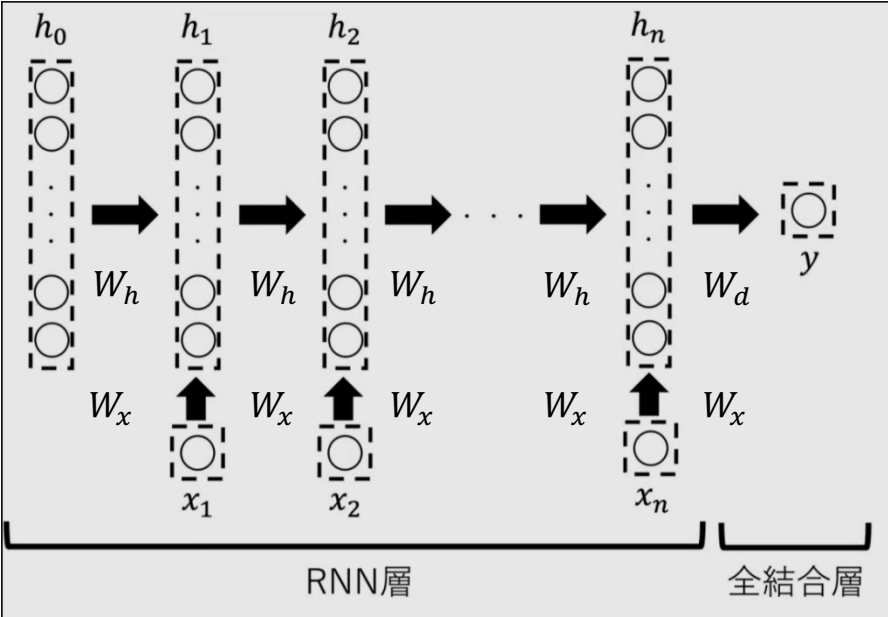
$W_h, W_d, W_x$  : 重み行列( $t$ に非依存)

# RNNの推論（順伝播）

- 事前知識として新規感染者数は曜日に依存する点を踏まえ、時系列長を  $n=7$  とする。
- 推論の計算は単純な行列・ベクトルの積和計算であり、実装が容易
- 本演習ではこの計算にテンソルネットワークを適用することにチャレンジする。

## 変数の詳細

変数	次元数	設定要領
$x_t$	1	日付tの新規感染者数
$y$	1	予測したい日付の新規感染者数
$h_t$	32	ネットワークの規模を決めるパラメータで、問題の複雑さに応じて調整。 $h_0 = 0$ で初期化。
$W_h$	32x32	重み行列。一様分布で初期化。
$W_d, W_x$	32	
$b_r$	32	RNNのバイアス項
$b_d$	1	全結合層のバイアス項



## 推論の計算

- $h_t = W_h h_{t-1} + W_x x_t + b_r \quad (t = 1, \dots, 7)$
- $y = W_d^T h_n + b_d$

## RNNの学習（誤差逆伝播）

- 勾配降下法にて重みを更新していく。
- 実装が複雑となるため、本演習では学習済の重みを使用する前提とする。

重みの更新式（予測誤差： $L$ 、学習率： $\eta$ 、バイアスは省略）

- $w_{ij}^{new} = w_{ij} - \eta \frac{\partial L}{\partial w_{ij}}$
- $\frac{\partial L}{\partial w_{ij}} = \frac{\partial L}{\partial h_t^{(i)}} \frac{\partial h_t^{(i)}}{\partial w_{ij}} = \delta_t^{(i)} \frac{\partial}{\partial w_{ij}} \sum_j w_{ij} h_{t-1}^{(j)} = \delta_t^{(i)} h_{t-1}^{(j)}$
- $\delta_t^{(i)} = \frac{\partial L}{\partial h_t^{(i)}} = \sum_j \frac{\partial L}{\partial h_{t+1}^{(j)}} \frac{\partial h_{t+1}^{(j)}}{\partial h_t^{(i)}} = \sum_j \delta_{t+1}^{(j)} \frac{\partial}{\partial h_t^{(i)}} \sum_i w_{ji} h_t^{(i)} = \sum_j \delta_{t+1}^{(j)} w_{ji}$



## ■ RNNで学習～予測を実行

- データを読み込み、予め実装済のRNNを使って時系列予測モデルを学習
- 学習によって得られた重みを使って、推論の計算を実行
- 詳細は Jupyter Notebook を参照

Keras による実装の詳細については TensorFlow のチュートリアルを参照  
<https://www.tensorflow.org/guide/keras/rnn>