Multi-step prediction for influenza outbreak by an adjusted long short-term memory

J. Zhang and K. Nawata

Department of Technology Management for Innovation, Graduate School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-Ku, Tokyo 113-8656, Japan

Abstract

Influenza results in approximately 3–5 million annual cases of severe illness and 250 000–500 000 deaths. We urgently need an accurate multi-step-ahead time-series forecasting model to help hospitals to perform dynamical assignments of beds to influenza patients for the annually varied influenza season, and aid pharmaceutical companies to formulate a flexible plan of manufacturing vaccine for the yearly different influenza vaccine. In this study, we utilised four different multi-step prediction algorithms in the long short-term memory (LSTM). The result showed that implementing multiple single-output prediction in a six-layer LSTM structure achieved the best accuracy. The mean absolute percentage errors from two- to 13-step-ahead prediction for the US influenza-like illness rates were all <15%, averagely 12.930%. To the best of our knowledge, it is the first time that LSTM has been applied and refined to perform multi-step-ahead prediction for influenza outbreaks. Hopefully, this modelling methodology can be applied in other countries and therefore help prevent and control influenza worldwide.

Introduction

Influenza, commonly known as flu, circulates worldwide and places a substantial burden on people’s health every year. The flu outbreak resulted in approximately 3–5 million annual cases of severe illness and 250 000–500 000 deaths [1]. In the USA, annual flu outbreak led to an average of 610 660 life-years lost, 3.1 million hospitalised days and 31.4 million outpatient visits. The total economic burden of annual flu outbreak using projected statistical life values amounted to $87.1 billion [2]. Flu is one of the costliest epidemics worldwide.

The flu vaccine is one of the best ways to reduce the risk of getting sick with flu and spreading it to others [3]. During the 2015–2016 flu season, flu vaccine prevented an estimated 5.1 million illnesses, 2.5 million medical visits, 71 000 hospitalisations and 3000 pneumonia and influenza deaths [3]. However, because flu virus undergoes high mutation rates and frequent genetic re-assortment [4–6], manufacturing flu vaccine suffers from a complicated process every year. In February, World Health Organization (WHO) assesses the strains of flu virus that are most likely to be circulating over the following winter. Then, vaccine manufacturers produce flu vaccines in a very limited time [7]. Usually, the first batch of vaccine is unavailable until Septembers [8, 9].

Moreover, hospital beds assignment to flu patients is also a challenging task due to the limited capacity of hospital beds, time dependencies of bed request arrivals and unique treatment requirements of flu patients [10]. Furthermore, flu seasons vary in timing, severity and duration from one season to another [7]. Therefore, flu hospitalisation also varies by sites and time in each season [11, 12], which makes beds assignment to flu patients more difficult for hospitals.

To help hospitals and pharmaceutical companies better prepare for an annual flu outbreak, we need an accurate model to perform multi-step-ahead time-series prediction for flu outbreaks. Multi-step-ahead time-series prediction, or simply ‘multi-step prediction’, is an analytical task of predicting a sequence of values in future by analysing observed values in the past [13]. Nonetheless, not many past papers studied multi-step prediction for flu outbreaks. The possible reason could be that multi-step prediction usually results in poor accuracy due to some insuperable problems, such as error accumulation. One compromising method is that one can aggregate raw data to a larger time unit and then use the single-step prediction to avoid performing multi-step prediction. For instance, if raw data are weekly based, we can aggregate weekly values to monthly values and then perform a single-step prediction for the coming month (that is roughly around 4 weeks). Although the single-step prediction avoids poor accuracy, it will hinder us from understanding the trend and variation during the coming month.

In this study, we leveraged the deep learning model of long short-term memory (LSTM). Our selection of LSTM was based on the theoretical and practical consideration. In theory, the LSTM is a special kind of RNN. Its elaborate structure (multilayers and gated cells) enables LSTM to learn simulate non-linear function, long-term dependencies [14] and refine...
time-series prediction [15]. In practice, we found that LSTM achieved the best accuracy in all the six models (autoregressive integrated moving average, support vector regression, random forest, gradient boosting, artificial neural network and LSTM) when we performed a single-step prediction for the US flu data (the same data source with those of this study) in one of our previous studies [16].

Methods

Source data and metrics

We used the US flu data from the 40th week of 2002 to the 30th week of 2017, collected from the ‘FluView’ Portal of Centre for Disease Control and Prevention (CDC) [17]. To remove any possible variations in populations, we used the influenza-like illness (ILI) rates as the response \( y \) of models. The formula for ILI rate is

\[
\text{ILI rate} = \frac{\text{The number of ILI}}{\text{Total number of Illness}}.
\]

Figure 1(a) illustrates the historical plot of the US flu data. We split the duration into two parts: the first 2/3 (from the 40th week of 2002 to the 44th week of 2012) for training and the last 1/3 (from the 45th week of 2012 to the 30th week of 2017) for testing. Figure 1(b) describes the histogram of the weekly ILI rates. The histogram is right-skewed. Generally, comparing accuracy of models by mean absolute percentage error (MAPE) mainly reflects the difference of the ‘median’; while comparing by root mean square error primarily reflects the difference of the ‘mean’. The right-skewed histogram and the Kolmogorov–Smirnov test \( (P < 0.05) \) showed that the US flu data followed a non-normal distribution. Therefore, we expected that the accuracy would reflect the difference of ‘median’ and thereby used MAPE as a key performance indicator (KPI) for comparing models.

\[
\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{F_t - A_t}{A_t} \right| \times 100\%.
\]

where \( A_t \) is the actual value and \( F_t \) is the forecast value.

Feature space and responses

For the feature space, we adopted the time lag of 52 weeks due to the result of one of our previous studies. In the previous study, we used the same US flu data and compared accuracy of models of the different time lags of 2, 4, 9, 13, 26 and 52 weeks and found that of 52 weeks grew out the best accuracy [16]. Moreover, we calculated the first-order differences as a part of the feature space, since some past study found that first-order differences helped improve accuracy of the prediction models for flu data [18]. In brief, for feature spaces, we used (I) the ILI rate of the current week, (II) the ILI rates of the past 52 weeks and (III) the 52 first-order differences. Totally, we have 105 features.

Regarding responses, we were forecasting the two-, three-, four-, five-, six-, seven-, eight-, nine-, 10-, 11-, 12- and 13-step-ahead ILI rates.

Model

For multi-step prediction, there are mainly two types of methodologies: (I) ‘recursive’ prediction and (II) ‘jumping’ prediction. Generally, the methodology of (I) predicts values step-by-step; the methodology of (II) predicts some-step-ahead values independently. The following (a), (b), (c) and (d) explain four multi-step prediction algorithms [19].

(a) Multi-stage prediction (MSP)

MSP is a ‘recursive’ prediction. MSP uses one single-output model, which is recursively applied in multiple-step prediction, feeding through the previous output as its new input [20]. In this study, as the first step, we predict \( X_{t+1} \) using the 53 historical values, i.e. \( X_0, X_{t-1}, X_{t-2} \ldots \) and \( X_{t-52} \) as the second step, we predict \( X_{t+2} \) based on \( X_{t+1} (X_{t+1} \text{ was predicted in the first step}) \), \( X_t, X_{t-1}, X_{t-2} \ldots \) and \( X_{t-51} \); as the third step, we predict \( X_{t+3} \) based on \( X_{t+2} (X_{t+2} \text{ was predicted in the second step}) \), \( X_{t+1}, X_t, X_{t-1}, \ldots \) and \( X_{t-50} \). Etc. Formula 1 describes the prediction process.

\[
\text{Formula 1. Algorithm of MSP}
\]

\[
\text{prediction}_{(t+1)} = \text{LSTM}_\text{MSP}_\text{MODEL}\#01(\text{observation}_{(t)}, \text{observation}_{(t-1)}, \text{observation}_{(t-2)}, \ldots, \text{observation}_{(t-52)})
\]

\[
\text{prediction}_{(t+2)} = \text{LSTM}_\text{MSP}_\text{MODEL}\#01(\text{prediction}_{(t+1)}, \text{observation}_{(t)}, \text{observation}_{(t-1)}, \ldots, \text{observation}_{(t-51)})
\]

\[
\text{prediction}_{(t+3)} = \text{LSTM}_\text{MSP}_\text{MODEL}\#01(\text{prediction}_{(t+2)}, \text{observation}_{(t)}, \text{observation}_{(t-1)}, \ldots, \text{observation}_{(t-50)})
\]

\[
\text{prediction}_{(t+1), (t+2), (t+3)} = \text{LSTM}_\text{MSP}_\text{MODEL}\#01(\text{prediction}_{(t+1)}, \text{prediction}_{(t+2)}, \text{prediction}_{(t+3)}, \text{observation}_{(t)}, \text{observation}_{(t-1)}, \ldots, \text{observation}_{(t-49)})
\]

(b) Adjusted multi-stage prediction (AMSP)

AMSP is a refined version of MSP. Comparatively, when MSP predicting \( X_{t+p} \) \( (P \geq 2) \), it uses another model instead of using the same model repeatedly. Such a modification helps suppress error accumulation [21, 22]. Formula 2 illustrates the algorithm of AMSP.

\[
\text{Formula 2. Algorithm of AMSP}
\]

\[
\text{prediction}_{(t+1)} = \text{LSTM}_\text{AMSP}_\text{MODEL}\#01(\text{observation}_{(t)}, \text{observation}_{(t-1)}, \text{observation}_{(t-2)}, \ldots, \text{observation}_{(t-52)})
\]

\[
\text{prediction}_{(t+2)} = \text{LSTM}_\text{AMSP}_\text{MODEL}\#02(\text{prediction}_{(t+1)}, \text{observation}_{(t)}, \text{observation}_{(t-1)}, \ldots, \text{observation}_{(t-51)})
\]

\[
\text{prediction}_{(t+3)} = \text{LSTM}_\text{AMSP}_\text{MODEL}\#03(\text{prediction}_{(t+2)}, \text{observation}_{(t)}, \text{observation}_{(t-1)}, \ldots, \text{observation}_{(t-50)})
\]

\[
\text{prediction}_{(t+1), (t+2), (t+3)} = \text{LSTM}_\text{AMSP}_\text{MODEL}\#13(\text{prediction}_{(t+1)}, \text{prediction}_{(t+2)}, \text{prediction}_{(t+3)}, \text{observation}_{(t)}, \text{observation}_{(t-1)}, \ldots, \text{observation}_{(t-40)})
\]

(c) Multiple single-output prediction (MSOP)

MSOP is a ‘jumping’ prediction. MSOP directly predicts a \( p \)-step-ahead \( (P \geq 2) \) value only by historical values: \( X_0, X_{t-1}, X_{t-2}, \ldots, X_{t-p} \). Formula 3 explains the algorithm of MSOP.

\[
\text{Formula 3. Algorithm of MSOP}
\]

\[
\text{prediction}_{(t+1)} = \text{LSTM}_\text{MSOP}_\text{MODEL}\#01(\text{observation}_{(t)}, \text{observation}_{(t-1)}, \text{observation}_{(t-2)}, \ldots, \text{observation}_{(t-52)})
\]

\[
\text{prediction}_{(t+2)} = \text{LSTM}_\text{MSOP}_\text{MODEL}\#02(\text{observation}_{(t)}, \text{observation}_{(t-1)}, \text{observation}_{(t-2)}, \ldots, \text{observation}_{(t-52)})
\]

\[
\text{prediction}_{(t+3)} = \text{LSTM}_\text{MSOP}_\text{MODEL}\#03(\text{observation}_{(t)}, \text{observation}_{(t-1)}, \text{observation}_{(t-2)}, \ldots, \text{observation}_{(t-52)})
\]

\[
\text{prediction}_{(t+1), (t+2), (t+3)} = \text{LSTM}_\text{MSOP}_\text{MODEL}\#13(\text{prediction}_{(t+1)}, \text{prediction}_{(t+2)}, \text{prediction}_{(t+3)}, \text{observation}_{(t)}, \text{observation}_{(t-1)}, \ldots, \text{observation}_{(t-50)})
\]
prediction(t+13) = LSTM_MSOP_MODEL#13(observation(t), observation(t−1), observation(t−2), ..., observation(t−52))

(d) Multiple-output prediction (MOP)

MOP can be regarded as a merged version of MSOP. MOP uses one model to predict many some-step-ahead values all at once. In other fields, MOP was also implemented by multiple support vector regression [20, 23, 24]. Formula 4 outlines the algorithm of MOP.

Formula 4. Algorithm of MOP

prediction(t+1) = prediction(t+2) = ... = prediction(t+13) = LSTM_MOP_MODEL#01(observation(t), observation(t−1), observation(t−2), ..., observation(t−52))

Coding

We used Python and Keras package (Version 2.0.4) [25] based on Tensorflow (Version 1.1.0) [26]. We adopted an ‘early-stopping’ algorithm with a ‘patience’ of 100 epochs (for a total of 1000 epochs) and compared the predicting accuracy of LSTM models of the number of layers: 3–6 and 10.

Results

Results of MSP

Table 1(a) illustrates the MAPEs of LSTM with MSP algorithm. The three-layer LSTM with MSP algorithm achieved the predicting MAPEs of 19.757, 32.969, 49.096, 67.866, 89.892, 114.977, 143.999, 177.719, 217.721, 267.335, 332.577 and 426.828%, when forecasting the ILI rates of the coming second to 13th weeks. The MAPE increased by nearly 22 times as the number of predicting steps increased. Comparatively, the MAPEs of the four-layer LSTM with MSP algorithm increased limitedly, from 9.57% to 13.77% with some slight setbacks in 10-, 11- and 12-step prediction. The similar phenomena occurred in the MAPEs of LSTM of five, six and 10 layers with MSP (Fig. 2).

Results of AMSP, MSOP and MOP

Tables 1(b–d) display the MAPEs of LSTM with AMSP, MSOP and MOP algorithm. In AMSP, the average MAPE increased from 9.171% to 14.715% as the number of predicting steps increased from two to 13, and varied from 13.626% to 13.911% as the number of layers of LSTM increased from three to 10 layers.

The average MAPEs of LSTM of both MSOP and MOP had a slight upward trend as the number of predicting steps increased (from 8.911% to 14.064% in MSOP; from 10.747% to 26.389% in MOP); and varied limitedly as the number of layers of LSTM increased (from 12.935% to 13.386% in MSOP; from 17.910% to 20.684% in MOP).

In sharp contrast to MSP, the accuracy of AMSP, MSOP and MOP had little improvement when we used more layers of LSTM.
### Table 1. The MAPEs of LSTM with the multi-step predicting algorithms of MSP, AMSP, MSOP and MOP

#### (a) The MAPEs of LSTM with the MSP algorithm

<table>
<thead>
<tr>
<th>The numbers of multisteps</th>
<th>MAPEs of MSP of three-layer LSTM (%)</th>
<th>MAPEs of MSP of four-layer LSTM (%)</th>
<th>MAPEs of MSP of five-layer LSTM (%)</th>
<th>MAPEs of MSP of six-layer LSTM (%)</th>
<th>MAPEs of MSP of ten-layer LSTM (%)</th>
<th>Average MAPE of MSP of three-, four-, five-, six-, 10-layer LSTM (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>32.969</td>
<td>11.960</td>
<td>11.913</td>
<td>11.482</td>
<td>11.778</td>
<td>16.020</td>
</tr>
<tr>
<td>8</td>
<td>143.999</td>
<td>14.088</td>
<td>15.098</td>
<td>17.439</td>
<td>18.304</td>
<td>41.786</td>
</tr>
<tr>
<td>9</td>
<td>177.719</td>
<td>14.100</td>
<td>15.174</td>
<td>16.979</td>
<td>18.894</td>
<td>48.573</td>
</tr>
<tr>
<td>12</td>
<td>332.577</td>
<td>13.763</td>
<td>15.272</td>
<td>17.135</td>
<td>19.880</td>
<td>79.725</td>
</tr>
<tr>
<td>13</td>
<td>426.828</td>
<td>13.770</td>
<td>15.320</td>
<td>17.749</td>
<td>19.931</td>
<td>98.720</td>
</tr>
<tr>
<td><strong>Average MAPE of MSP of all multi-steps (%)</strong></td>
<td><strong>161.728</strong></td>
<td><strong>13.315</strong></td>
<td><strong>14.096</strong></td>
<td><strong>15.368</strong></td>
<td><strong>16.612</strong></td>
<td></td>
</tr>
</tbody>
</table>

MAPEs, mean absolute percentage errors; MSP, multi-stage prediction; LSTM, long short-term memory.

#### (b) The MAPEs of LSTM with the AMSP algorithm

<table>
<thead>
<tr>
<th>The number of multisteps</th>
<th>MAPEs of AMSP of three-layer LSTM (%)</th>
<th>MAPEs of AMSP of four-layer LSTM (%)</th>
<th>MAPEs of AMSP of five-layer LSTM (%)</th>
<th>MAPEs of AMSP of six-layer LSTM (%)</th>
<th>MAPEs of AMSP of ten-layer LSTM (%)</th>
<th>Average MAPE of AMSP of three-, four-, five-, six-, 10-layer LSTM (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Average MAPE of AMSP of all multi-steps (%)</strong></td>
<td><strong>13.626</strong></td>
<td><strong>13.544</strong></td>
<td><strong>13.873</strong></td>
<td><strong>13.409</strong></td>
<td><strong>13.911</strong></td>
<td><strong>13.615</strong></td>
</tr>
</tbody>
</table>

MAPEs, mean absolute percentage errors; AMSP, adjusted multi-stage prediction; LSTM, long short-term memory.
### (c) The MAPEs of LSTM with the MSOP algorithm

<table>
<thead>
<tr>
<th>The number of multi-steps</th>
<th>MAPEs of MSOP of three-layer LSTM (%)</th>
<th>MAPEs of MSOP of four-layer LSTM (%)</th>
<th>MAPEs of MSOP of five-layer LSTM (%)</th>
<th>MAPEs of MSOP of six-layer LSTM (%)</th>
<th>MAPEs of MSOP of ten-layer LSTM (%)</th>
<th>Average MAPE of MSOP of three-, four-, five-, six-, ten-layer LSTM (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Average MAPE of MSOP of all multi-steps (%)</strong></td>
<td><strong>13.221</strong></td>
<td><strong>12.935</strong></td>
<td><strong>13.133</strong></td>
<td><strong>12.930</strong></td>
<td><strong>13.386</strong></td>
<td></td>
</tr>
</tbody>
</table>

MAPEs, mean absolute percentage errors; MSOP, multiple single-output prediction; LSTM, long short-term memory.

### (d) The MAPEs of LSTM with the MOP algorithm

<table>
<thead>
<tr>
<th>The number of multi-steps</th>
<th>MAPEs of MOP of three-layer LSTM (%)</th>
<th>MAPEs of MOP of four-layer LSTM (%)</th>
<th>MAPEs of MOP of five-layer LSTM (%)</th>
<th>MAPEs of MOP of six-layer LSTM (%)</th>
<th>MAPEs of MOP of ten-layer LSTM (%)</th>
<th>Average MAPE of MOP of three-, four-, five-, six-, ten-layer LSTM (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>18.158</td>
<td>16.142</td>
<td>13.528</td>
<td>15.252</td>
<td>16.663</td>
<td>15.949</td>
</tr>
<tr>
<td>5</td>
<td>19.728</td>
<td>17.442</td>
<td>14.985</td>
<td>19.175</td>
<td>15.292</td>
<td>17.324</td>
</tr>
<tr>
<td>9</td>
<td>24.043</td>
<td>24.871</td>
<td>20.444</td>
<td>18.006</td>
<td>23.792</td>
<td>22.231</td>
</tr>
<tr>
<td><strong>Average MAPE of MOP of all multi-steps (%)</strong></td>
<td><strong>20.030</strong></td>
<td><strong>19.495</strong></td>
<td><strong>17.910</strong></td>
<td><strong>18.859</strong></td>
<td><strong>20.684</strong></td>
<td></td>
</tr>
</tbody>
</table>

MAPEs, mean absolute percentage errors; MOP, multiple-output prediction; LSTM, long short-term memory.
Comparison of the average MAPE of MSP, AMSP, MSOP and MOP

Figure 3 compares the average MAPEs of LSTM with multi-step predicting algorithms of MSP, AMSP, MSOP and MOP. The different numbers of the layers impacted the predicting accuracy tremendously in MSP (from 13.315% to 161.728%); slightly in MOP (from 17.910% to 20.684%), and barely in AMSP (from 13.626% to 13.911%) and MSOP (from 12.930% to 13.386%). Implementing MSOP in the six-layer LSTM structure achieved the best accuracy in this study. The MAPEs from two-step-ahead to 13-step-ahead prediction for the US ILI rates were all <15%, averagely 12.930%.

Fig. 2. The MAPEs of LSTM with MSP. The y-axis represents the MAPE of the predictions and the x-axis represents multi-steps of the predictions. The (a–e) illustrate the MAPEs with the MSP algorithm of three-, four-, five-, six-, 10-layer LSTM, respectively.

Fig. 3. The average MAPEs of LSTM with MSP, AMSP, MSOP and MOP. The y-axis represents the MAPE of the predictions and the x-axis represents the models of three-, four-, five-, six- and 10-layer LSTM with the multi-step predicting algorithms of MSP, AMSP, MSOP and MOP. Implementing MSOP in the six-layer LSTM achieved the lowest average MAPE of 12.930% in this study.
Discussion

Past studies

We did not find past studies that performed auto-regression in the multi-step prediction for flu outbreaks. Regarding multi-step prediction for studies in other fields, MSP is one of the most popular methodologies probably because many types of models can be used for this purpose, such as linear regression, support vector regression [27], random forest, gradient boosting, artificial neural network [28], etc. However, any such model inevitably introduces errors and tends to suffer from error accumulation problem when the predicted period is long. This is because the bias and variance from previous predictions impact future predictions [20]. These compounding errors change the input distribution for future prediction steps, breaking the train-test independent and identically distributed assumption common in supervised learning [20].

Comparison of accuracy of MSP, AMSP, MSOP and MOP

When comparing four different multi-step predicting algorithms, we found that the MAPEs of AMSP were less than those of MSP, which demonstrated AMSP suppressed the accumulated errors effectively based on its refined algorithm. Besides, the MAPEs of MSOP are less than those of MOP. As we mentioned in section ‘Methods’, to predict the ILI rates of the coming second to 13th weeks, MOP trained only one model while MSOP trained 13 models. As a result, MSOP can predict with no necessity of sharing neurons in LSTM structure, while MOP has to share neurons in LSTM structure. Consequently, the accuracy of MSOP performed better. Moreover, the average MAPEs of MSOP are slightly less than those of AMSP. The explanation is that MSOP does not accumulate errors at all, while AMSP just adjusted its accumulated errors by training new models. Therefore, MSOP performed best.

Other features

In our opinion, including other features in multi-step predicting models impacts models’ accuracy positively and negatively. For one thing, when predicting future values, other features could help predict more accurately, especially at turning points, such as an abrupt decrease in temperature. For another thing, before forecasting future ILI rates, we need to forecast other features (e.g. we need weather forecast for temperature and humidity). The error in former prediction could enlarge the error in later prediction. The mechanism is similar to MSP, which accumulates error step by step. In conclusion, whether the accuracy improves or deteriorates might depend on different data in different season from different countries.

In this study, we only performed auto-regression based on two pieces of consideration. First, we regard historical values as a response of all related features, such as temperature, humidity, etc. Therefore, to some extent, taking historical values as feature space includes all related features/factors in models. Besides, how to include temperature or humidity of the whole country in the models is a challenge work. Simply averaging temperature or humidity of all the places (cities and towns) of the USA might bring other problems, such as overlooking in population size, population density, life styles, etc. in different places.

Conclusion

In this study, we adjusted the LSTM model by the four multi-step prediction algorithms. The result showed that implementing MSOP in a six-layer LSTM structure achieved the best accuracy. The MAPEs from two-step-ahead to 13-step-ahead prediction for the US ILI rates were all <15%, averagely 12.930%. Hopefully, this accurate modelling approach will positively help hospitals, pharmaceutical companies, individuals and governments better prepare for the flu seasons and therefore prevent and control flu outbreaks worldwide.

Declaration of Interest. None.

Ethical Standards. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

References