Prediction of Alzheimer's Disease Based on Bidirectional LSTM

Qiao Pan¹, Shiyu Wang² and Junhao Zhang ¹*

¹Computer science and technology, Donghua University, Shanghai, 201620, China
²Computer science and technology, Donghua University, Shanghai, 201620, China
¹*Computer science and technology, Donghua University, Shanghai, 201620, China
Shiyu Wang e-mail: 1248965945@qq.com

Abstract. Alzheimer's disease (AD) is a common disease in the elderly. It affects human life seriously and is difficult to cure, so if you can predict the occurrence of the disease and the development trend in advance, you can prevent or treat Alzheimer's disease as soon as possible. Mild cognitive impairment (MCI) is a syndrome that occurs in the preclinical phase of Alzheimer's disease (AD). It is a transitional state between normal aging and early AD and may be an early sign of AD. This article uses the basic information of the patient's neuropsychological test scale data, genetic data and tomographic data in first, six and twelve months as input data and bidirectional LSTM plus Attention mechanism as a model to obtain a three-dimensional model. The output of the vector is divided into normal (NL), mild cognitive impairment (MCI) and Alzheimer's disease (AD). The experimental results can predict the development of Alzheimer's disease (AD), and determined the model has a good performance.

1. Introduction
Alzheimer disease (AD), commonly known as senile dementia, is a slow progress and incurable course. It starts with memory impairment, gradually develops to loss of multiple cognitive abilities, changes in personality and behavior, and eventually loses basic bodily functions until death. Although the pathological changes in AD begin to appear very early, their typical clinical symptoms do not appear until later. But when the patient shows clinical symptoms, the condition has missed the best treatment opportunity. Therefore, if you can conduct an in-depth study of MCI, which is an early signal of the onset of AD, based on the patient's early diagnosis, it is hoped that the high-risk population of AD will be discovered and screened, and the best time window for AD treatment will be provided to prevent or delay the occurrence of AD.

This paper uses a bidirectional LSTM plus Attention mechanism as a model. In addition, the patient's basic information, genetic data and three time points of neuropsychology scale were used as input to predict the development trend of the patient's condition, which is normal (NL), mild cognitive impairment (MCI) or Alzheimer's disease (AD).

2. Related work
Medically, clinical/cognitive measures have been established to assess the cognitive status of patients and to refer to these scales as an important criterion for clinical diagnosis of Alzheimer's disease. For example, the mini mental state examination (MMSE) and the Alzheimer's disease assessment scale-cognitive subscale (ADAS-Cog) [1].
Duchesne et al. [2] studied the relationship between magnetic resonance image (MRI) and cognitive function changes, and used the component analysis (PCA) to perform the MRI image of the obtained nuclear magnetic resonance image. After dimension reduction, a linear regression model was used to predict the trend of MMSE changes in patients with Alzheimer's disease for one year. Stonnington et al. [3] used the correlation vector regression (RVR) method to study the relationship between MRI scans and related clinical scores in patients with Alzheimer's disease, on two independent medical data sets (Mayo). After analyzing the data in the Clinic and the Alzheimer's Disease Neuroimaging Initiative, a continuous model of predicting the patient's clinical score based on individual MRI scans of Alzheimer's patients was established.

All of the above methods are based on clinical data at a time point to predict clinical scores related to the condition of Alzheimer's disease. To improve predictive performance, Zhou et al. [4] proposed a simultaneous use of multiple time points.

During the collection process, there are often cases where some entries are missing, resulting in incomplete patient records. To solve these problems, Xiang et al. [5] proposed a unified two-layer efficient optimization model based on complete multi-source data. Another method is to use the multi-core learning (MKL) method [6] for data fusion. The disadvantage of this method is that only the source data layer is analyzed, without considering the feature layer and the data source layer, the sub-optimal solution is produced when the source is high-dimensional data.

In view of the above problems, this paper proposes a method to predict the development of AD by using bidirectional LSTM plus Attention mechanism as a model. Compared with the traditional method, it has the following advantages: firstly, this paper starts from the acquired neuropsychological data, analyzes the time availability of the data, and refers to the time series problem in deep learning to apply the key attribute of time to the model. This reduces the performance problems associated with making predictions at a single point in time. Secondly, this paper also makes use of the special methods in deep learning to attach weights to different attributes (chapter 3), so that the attribute values have their respective weights.

3. Model
This section introduces a model of bi-directional LSTM attentional mechanism used to predict the patient's condition. Figure 1 shows the overall structure of the model. After the data is inputted, it is processed by dropout and transmitted to the hidden layer. In the hidden layer, the data processed by BI-LSTM is divided into two parts. One part is used for attention calculation of attention_vec and the other part is used directly for standby in attention_mul. The two parts are calculated into a unified matrix and output through the output layer.

Figure 1. Structure diagram.
3.1. Bidirectional long-term and short-term memory (BI-LSTM)
RNN has shown extraordinary ability to handle sequence problems. However, RNN will eventually cause the gradient disappearance. In order to solve this problem, scholars proposed RNN variant structure based on gate control mechanism, namely LSTM. [7] LSTM added three gates on the basis of RNN: input gate, forget gate, output gate, and a core memory module (cell).

Figure 2 shows the operation mode of the three gates. In the formula, w and h are weight matrix and a number of 0-1, indicating whether the t-th cell needs to be forgotten, 1 means complete retention, 0 means complete abandonment, and b is the value of paranoia.

The input gate is used to control how much of the current input can be transferred to the core memory module. The mathematical formula is as (1). The output gate controls how much of the current core memory module can be output to the hidden node (2). The Forgotten Gate is used to properly forget some things to improve the learning speed of the model when the context is low (3).

\[
i_t = \sigma(w_{xi}x_t + w_{hi}h_{t-1} + w_{ci}c_{t-1} + b_i) \quad (1)
\]
\[
o_t = \sigma(w_{xo}x_t + w_{ho}h_{t-1} + w_{co}c_t + b_o) \quad (2)
\]
\[
f_t = \sigma(w_{xf}x_t + w_{hf}h_{t-1} + w_{cf}c_t + b_f) \quad (3)
\]

The basic idea of Bi-LSTM is to propose two forward and backward LSTMs for each training sequence, and both are connected to the output layer. This structure provides complete past and future context information for each point in the input sequence of the output layer.

For the hidden layer of bi-lstm, forward calculation is the same as backward LSTM. But for the two hidden layers, the input sequence goes in the opposite direction. The output layer will wait until both hidden layers have passed before the output information is updated. Figure 3 shows the standard form of bidirectional LSTM. \(X_t\) is the input at different times and \(Y_t\) is the corresponding output, A is the forward LSTM cell, \(A'\) is the reverse LSTM cell, \(s_0\) to \(s_t\) is the change of the forward information transfer matrix, \(s_0'\) to \(s_t'\) is the change of the reverse information transfer matrix.

Figure 2. LSTM structure. Figure 3. Bidirectional RNN.

3.2. Attention
Currently, the Encode-Decode model is very popular with the public because it has achieved better results in many areas than other models. The usual practice is to encode an input sentence into a fixed-size vector and then as the initial state of the decode. Such a fixed encoding method causes some long sentences to lose features, and may also allow some shorter sentences to be mixed noise. In response to this problem, the attention mechanism is introduced. At this time, the decoder makes the input at each moment different according to the time. Attention mechanism also has its shortcomings, which leads to an increase in the computational complexity of the model and a longer training time for the model.
4. Experiment
The data in this paper is from the Alzheimer's Disease Neuroimaging Initiative (ADNI public data set), which mainly uses the data of ADNImerge.

In order to improve the practicability of the model, model uses the random method to separate test data and training data, and allocate them according to the proportion of the total ratio of 8:2. The dataset has 1405 patients and 4218 data. The number of patients with different symptoms is shown in figure 4. Figure 4 shows the number of patients with different symptoms, with 0 indicating normal, 1 indicating MCI and 2 indicating AD. There were 794 men and 612 women in the data. Age is also a closely related property of Alzheimer's disease.

![Figure 4. Number of patients with different symptoms.](image)

4.1. Subsections
Java package of Weka and data of 25 dimensions was used as input. Weka used patient symptoms as labels, and set poolsize in CFS and lookuolocksize in BF to 1, searchtermination to 5, and then ran the code. The algorithm screened 7 common AD diagnostic attributes as shown in table 1.

| Serial number | ATTRIBUTES     | Serial number | ATTRIBUTES     |
|---------------|----------------|---------------|----------------|
| 1             | AGE            | 5             | RAVLT.forgetting |
| 2             | PTGENDER       | 6             | FAQ            |
| 3             | FDG            | 7             | CDRSB.bl       |
| 4             | MMSE           |               |                |

4.2. Baseline setting
In order to show the excellent performance of model, different experiments were used for comparison:
- A simple RNN model based on time series uses the attribute values mentioned above and combines the data of three time nodes to predict the last patient condition.
- The LSTM model, which is excellent in timing, also uses the same input data and functions as BI-LSTM model.
- The GRU model simplified by LSTM has the same input data and the same model function.

4.3. Experimental result
For different models, the parameters of the model are set to the same, the epoch (that is, the number of iterations of data per round when training the neural network) was set to 20, the dropout (training percentage) was set to 0.3, and the batch_size (that is, the amount of data input per round by the neural network) was set to 16.

Table 2 shows the training data set accuracy, test data set accuracy, and the model score provided by keras.

| Model         | Train-ACC | Test-ACC | Score |
|---------------|-----------|----------|-------|
| Simple RNN    | 0.45      | 0.50     | 1.03  |
|          | GRU  | LSTM | BL-LSTM attention |
|----------|------|------|-------------------|
| accuracy | 0.46 | 0.51 | 0.84              |
| precision| 0.51 | 0.56 | 0.87              |
| recall   | 1.06 | 1.03 | 0.41              |

The results show that the accuracy of the proposed model has been greatly improved. In the later stage, after further adjusting the parameters and changing the initialization mode inside the model, the accuracy was improved to over 90%.

While the accuracy of the model has been greatly improved, each round of epoch changed from 3 seconds to 15 seconds in the model training. Attention mechanism actually increased the model’s training time.

5. Conclusion
In this paper, the deep learning method was used to predict the development of the patient's condition from the neuroimaging scale, and at the same time the attention in BI-LSTM was added to let the attributes get their own weight. The patient's condition was successfully judged without using MRI.

Acknowledgments
This work was supported by the Science and Technology Development Foundation of Shanghai(16JC1400802, 16JC1400803, 18511102703), the Special Fund of Shanghai Municipal Commission of Economy and Informatization (RX-RJJC-08-16-0483,2017-RGZN-01004, XX-XXFZ-02-18-2666, XX-XXFZ-01-18-2604).

References
[1] Petrella, J. R., Coleman, R. E., & Doraiswamy, P. M. (2003). Neuroimaging and early diagnosis of alzheimer disease: a look to the future. Radiology, 226(2), 315-36.
[2] Duchesne, S., Caroli, A., Geroldi, C., Collins, D. L., & Frisoni, G. B. (2009). Relating one-year cognitive change in mild cognitive impairment to baseline mri features. Neuroimage, 47(4), 1363-1370.
[3] Stonnington, C. M., Chu, C., Klöppel, S., Jr, J. C., Ashburner, J., & Frackowiak, R. S. (2010). Predicting clinical scores from magnetic resonance scans in Alzheimer's disease. Neuroimage, 51(4), 1405-1413.
[4] Zhou, J., Yuan, L., Liu, J., & Ye, J. (2011). A multi-task learning formulation for predicting disease progression. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, Ca, USA, August (Vol.3, pp.814-822). DBLP.
[5] Xiang, S., Yuan, L., Fan, W., Wang, Y., Thompson, P. M., & Ye, J. (2013). Multi-source learning with block-wise missing data for Alzheimer's disease prediction. Paper presented at the Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, Chicago, Illinois, USA.
[6] Lanckriet GR, De Bie T, Cristianini N, Jordan MI, & Noble WS. (2004). A statistical framework for genomic data fusion. Bioinformatics, 20(16), 2626-35.
[7] Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. Computer Science.