ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6.
Available from http://www.i6doc.com/en/.

Temporal modeling of ALS using longitudinal data and long-short term memory-based algorithm

Aviv Nahon and Boaz Lerner

Ben-Gurion University of the Negev
Department of Industrial Engineering and Management
Beer Sheva, Israel

Abstract. ALS is a neurodegenerative disease where factors such as disease progression rate and pattern vary greatly among patients. Since patient functionality deteriorates over time, we model ALS temporally to mimic the physician's reasoning by incorporating old with new information using a long-short term memory (LSTM) network. We demonstrate that the LSTM achieves a higher accuracy than a random forest in disease state prediction, and improves accuracy with data from additional clinic visits. Being an anytime predictor, our model can help physicians and caregivers to adjust patients' treatment and living environment along the disease period, improving patients' life quality.

1 Introduction

Amyotrophic lateral sclerosis (ALS) is a devastating neurodegenerative disease, with no significantly affective therapies. The disease is heterogeneous; its onset site and progression pattern can differ significantly among patients. Since there are no specific biomarkers which are good predictors of future disease state or deterioration rate, physicians cannot assess disease state well at a future time point. In addition, when running clinical trials in developing therapies, there is no proper tool that helps in choosing patients with similar deterioration, and thus a large sample of patients is needed, which raises the trial cost tremendously.

Two major drawbacks of previous models learned by machine learning for ALS disease state prediction were: (1) Linearity assumption of the disease deterioration rate, which cannot express complex deterioration patterns; and (2) Use of non-temporal models, which cannot fully exploit the richness of the longitudinal data.

We design and train a temporal model for disease state prediction of ALS patients using long-short term memory (LSTM) and longitudinal data without the need to aggregate or drop some of the data. The memory component of the LSTM exercises the practice of physicians to use past information about a patient to make more reliable future decisions. We examined the model's prediction capabilities when using three months of data regarding patients and predicting their disease state one year from the baseline. The results show that the LSTM model outperformed a random forest model. In addition, we examined an online prediction task: testing the model's ability to improve prediction when more data regarding a specific patient is available. The results show that, in many

ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6. Available from http://www.i6doc.com/en/.

cases, the predictions are accurate even in early clinic visits. In other cases, when the model had less accurate prediction in the early stage, it was able to improve its prediction as data from additional clinic visits became available.

2 Background

ALS is a fatal idiopathic neurogenerative disease of the human motor system. It has a highly uncertain pathogenesis whose inner workings and mechanisms remain unknown, more than a hundred years after its discovery [1, 2]. Known to be terminal within an average of three to five years from onset, the disease often takes a mental toll on its patients as well as a physical one [3]. Accurate prediction of future disease state can help assess disease deterioration, and hence help in: (1) managing clinical trials, allowing use of a relatively small sample of patients, since the individual deterioration rate would be well assessed; and (2) giving the patient and his caregivers a better understanding of his future condition, allowing necessary preparations of his environment in advance.

The Pooled Resource Open-Access ALS Clinical Trials Database (PRO-ACT) [4], which contains over 8,500 unique clinical patient records, was used in this research. In the database, a patient's medical condition is evaluated using ALS functional rating scale (ALSFRS). The ten ALSFRS items describe physical functionalities of the patient, e.g., breathing, speaking, and walking, each given a value between 0 for no functionality and 4 for full functionality [5]. A number of studies have used PRO-ACT and the ALSFRS measure and were focused on the task of prediction of the disease state or deterioration rate. For example, one study developed multiple models for different progression rates based on clustering the patients in the database into groups of faster and slower progressors [6]. After a new patient was grouped into a cluster, his ALSFRS value was predicted at the stage of 12 months later using a Weibull prediction model associated with this cluster. Another study performed crowd-sourced analysis of clinical trial data to predict ALS [7, 8]. Solvers were asked to use three months of clinical trial information to predict the future progression of the disease from the fourth month to the twelve month. The progression of the disease was assessed by the slope of change in ALSFRS values, assuming linearity. Four out of the top six teams employed variants of the random forest [9] approach, which is the benchmark algorithm for comparison of our model in this study.

The models which were used in previous studies are not temporal. In non-temporal models, when working with longitudinal data, such as the PRO-ACT, it is not possible or easy to exploit the data from all time points. Usually, some of the data are thrown away, aggregated, or partly used. In contrast, a temporal model that is fully capable of incorporating and utilizing longitudinal data, can take into account trends in our predictor variables (laboratory test results and vital signs) and give proper importance to each time point in our prediction. The relevance of past information to the prediction should dynamically adjust the algorithm during training according to the level of contribution of this information to the prediction. Using data from several clinic visits of a patient from the

ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6. Available from http://www.i6doc.com/en/.

PRO-ACT database and the LSTM [10] algorithm that combines past information about this patient with new information, we produced accurate prediction of the patient's future state.

3 Architecture and methodology

We examined ALS disease-state prediction using an LSTM-based network. The target variable is the total sum of the ten ALSFRS functionality values (i.e., an integer in the range 0–40).

Network architecture. The network includes two layers of LSTM, with 200 hidden units each. The output of the second LSTM layer is the input of a neural network layer, with 200 hidden units, which yields the system output.

Input data. The input data (Table 1) include static variables, i.e., variables which do not change through time such as onset site and gender, and temporal variables such as vital signs, forced vital capacity (FVC), and five laboratory test results, which were chosen based on their contribution for prediction, as presented in [11]. Only patients with no missing data have been used, 2,850 of them were used for training and 1,126 for testing.

Our methodology extends previous methodologies for ALS disease state prediction [7, 8] that use the first three months in a patient record for training, and the one-year ALSFRS value as the target. Instead of using patient representation based on only a single observation using specific past visits to predict disease state in a specific future visit, we created multiple observations, each referring to a prediction of a different future visit, using different past visits, and thereby extended the research question of previous studies to that of prediction for multiple time periods. Fig. 1 demonstrates the suggested enrichment: Fig. 1(a) presents all visits an example training patient has, whereas Fig. 1(b) demonstrates the previous methodology [7, 8] by which visit 3 (held on day 170 for this patient, after the three-month conventional training period has ended) does not contribute to the training. Our suggested enrichment methodology, shown in Fig. 1(c), makes six observations out of the patient record, allowing the model to use all data, and to learn a more flexible prediction model. Using this methodology, we are able to create $\frac{n^2-n}{2}$ observations for a training patient with n visits.

4 Experiment results

Three-months to one-year prediction. The results of using three-months data to one-year prediction was examined using the LSTM-based model, and the results are shown in Fig. 2, which presents real vs. predicted ALSFRS values. The predictions are close to the real values and are not clustered in a certain area in the graph, meaning that the model is not trying to reduce its error by making predictions based on common values in the data.

ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6. Available from http://www.i6doc.com/en/.

Table 1: An example for the input data variables for a specific patient. For simplicity, units of laboratory test results are not specified.

		Static variables			Temporal variables				
ID	**Visit's No.**	**Onset site**	**...**	**Gender**	**Days from diagnosis**	**Total ALSFRS**	**CK**	**...**	**FVC**
4390	1	Bulbar	...	Male	0	35	405	...	81
4390	2	Bulbar	...	Male	34	33	486	...	79
...	...	...	...	...	...	...	...	...	...
4390	12	Bulbar	...	Male	552	25	1096	...	51

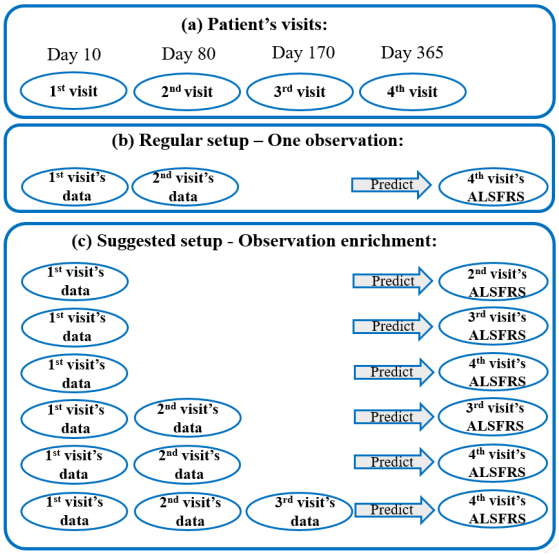

Fig. 1: (a) All documented visits for an example patient, (b) the training observation which is derived from the three-months to one-year prediction task, and (c) the suggested enrichment methodology

Table 2 presents result comparison of the LSTM-based model with a state-of-the-art prediction model, random forest (RF). RF is not a temporal model, but it was widely used in previous ALS machine learning competition [7, 8] and research. Three measures were used for evaluating the models' performance: (1) Root mean square error (RMSE), (2) Pearson's correlation coefficient[1] (PCC), and (3) Concordance index[2] (CI). The values in the table are the mean results of a 10-fold CV experiment. The LSTM-based model outperformed the RF model in all measures. All differences are significant (with a p-value lower than 0.001).

[1]Measures the correlation between the predictions vector and the real values vector. Should be maximized.

[2]The fraction of pairs in the test data, where the observation with the higher documented ALSFRS has the higher predicted ALSFRS. Should be maximized.

ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6. Available from http://www.i6doc.com/en/.

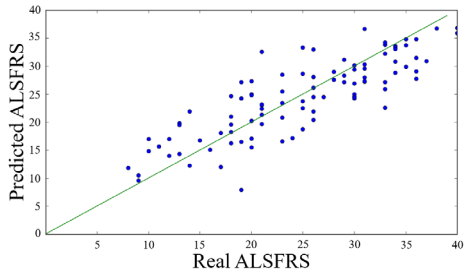

Fig. 2: ALSFRS predicted vs. real values for an LSTM-based model in a three-month to one-year prediction scenario.

Model	RMSE	PCC	CI
LSTM	**4.202**	**0.742**	**0.724**
RF	4.488	0.720	0.715

Table 2: Model comparison. Bold values represent outperformance.

Individual online prediction. The LSTM-based model can be used for individualized short-term (days or tens of days) and long-term (hundreds of days) predictions, due to our enrichment methodology. Fig. 3 shows the model's predictions for three arbitrary patients using only data of the first visit of each patient. In these cases, the model is able to make accurate short-term and long-term predictions. Not in all cases were predictions accurate starting from the first visit, but due to the temporal nature of the model, predictions can improve as more data regarding a patient are available. Fig. 4 demonstrates an online prediction for an arbitrary patient. In the first visit, the predictions were not accurate. As more data were available, starting from visit 2, the model was able to improve the predictions. This example illustrates a great advantage of temporal modeling since the model is able to be adjusted in real time.

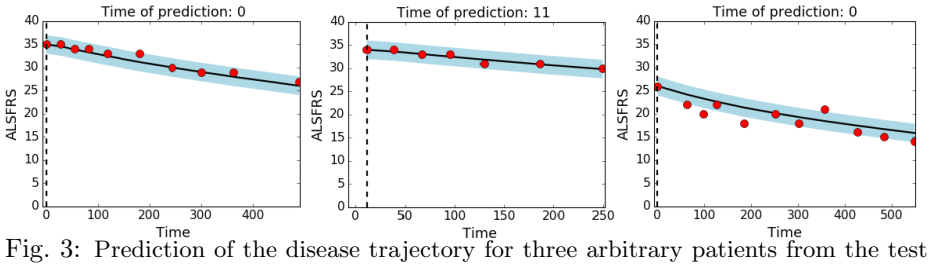

Fig. 3: Prediction of the disease trajectory for three arbitrary patients from the test set. Red dots are real ALSFRS observations, the black line is the predicted trajectory, and the blue band is an interval representing a deviation of 5% of the maximum value (of 40) of the ALSFRS. The vertical dashed line corresponds to the time of prediction in days (detailed in the graph title), such that any observation made after the line was unavailable to the model at prediction time.

5 Discussion

In this paper, we proposed a temporal model based on an LSTM network, which is well-suited for analyzing the longitudinal data in ALS. This analysis demon-

ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6. Available from http://www.i6doc.com/en/.

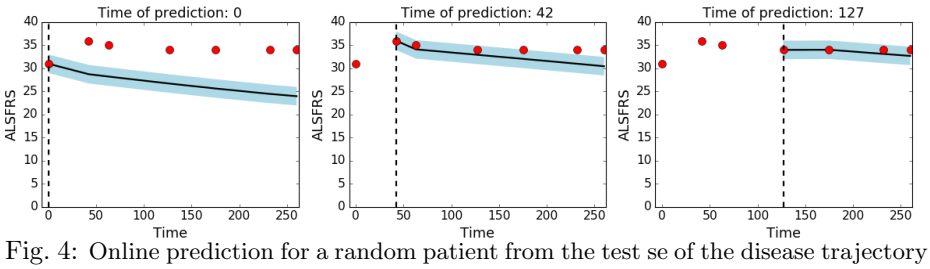

Fig. 4: Online prediction for a random patient from the test se of the disease trajectory through different time points.

strates the potential of implementing fully temporal techniques in modeling the disease. Our experiments demonstrate the benefits in this model as a state-of-the-art predictor. The temporal nature of the model allows online prediction to be made, by integrating new data in the model anytime it is available.

The model's capabilities can be used for several applications. It can be used by patients and caregivers for a better understanding of the disease progression pattern and making the necessary arrangements to meet disease deterioration. In addition, it can help in managing clinical trials by allowing patients with similar deterioration rates to be chosen and, by that, to decrease the sample size needed for trials and thereby the trial's cost. Also, accurate prediction of the disease state is important for fitting personalized medicine to patients.

References

[1] M. Kiernan, S. Vucic, B. Cheah, M. Turner, and A. Eisen. Amyotrophic lateral sclerosis. *Lancet*, 377:942–955, 2011.

[2] J. D. Rothstein. Current hypotheses for the underlying biology of amyotrophic lateral sclerosis. *Annals of Neurology*, 65:3–9, 2009.

[3] D. Mitchell and G. Borasio. Amyotrophic lateral sclerosis. *Lancet*, 33:51–59, 2007.

[4] N. Atassi, J. Berry, A. Shui, N. Zach, and A. Sherman. The PRO-ACT database. *Neurology*, 83:1719–1725, 2014.

[5] B. R. Brooks, M. Sanjack, S. Ringel, J. England, J. Brinkmann, A. Pestronk, and J. M. Cedarbaum. The amyotrophic lateral sclerosis functional rating scale-assessment of activities of daily living in patients with amyotrophic lateral sclerosis. *Archives of Neurology*, 53.2:141–147, 1996.

[6] R. Gomeni and M. Fava. Amyotrophic lateral sclerosis disease progression model. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, 15:119–129, 2014.

[7] R. Kuffner, N. Zach, R. Norel, and J. Hawe. Crowdsourced analysis of clinical trial data to predict amyotrophic lateral sclerosis progression. *Nature Biotechnology*, 33:51–59, 2015.

[8] T. Hothorn and H. Jung. RandomForest4life: A random forest for predicting als disease progression. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, 15:444–452, 2014.

[9] L. Breiman. Random forests. *Machine Learning*, 45.1:5–32, 2001.

[10] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.

[11] J. Gordon and B. Lerner. Exposing and modeling underlying mechanisms in ALS with machine learning. In *2016 23rd International Conference on Pattern Recognition*, pages 2168–2173. IEEE, 2016.