An Attention-based Recurrent Neural Networks Framework for Health Data Analysis

Qiuling Suo[1], Fenglong Ma[1], Giovanni Canino, [2], Jing Gao, [1], Aidong Zhang[1], Agostino Gnasso, MD[3], Giuseppe Tradigo [2], Pierangelo Veltri [2]

[1]Department of Computer Science and Engineering, University at Buffalo, NY, USA
[2]Department of Surgical and Medical Sciences, Magna Graecia University, Catanzaro, Italy
[3]Metabolic Diseases Unit, Department of Clinical and Experimental Medicine, Mater Domini Hospital, Magna Graecia University, Catanzaro, Italy

(Extended Abstract of Recent Publication)

Abstract

In this paper we focus on prediction of health status of patients from the historical Electronic Health Records (EHR). We propose a multi-task framework that can monitor the multiple status of diagnoses. Patients' historical records are fed into a Recurrent Neural Network (RNN) which memorizes all the past visit information, and then a task-specific layer is trained to predict multiple diagnoses. Experimental results show that prediction accuracy is reliable if compared to widely used approaches [1]

1 Introduction

Disease monitoring is often limited by physician experience, test time, economic barriers and so on. The Electronic Health Record (EHR) is a valuable source for exploratory analysis to monitor diseases and assist clinical decision making. However, due to the complexity of EHR data, the efficient mining of EHRs is not trivial.

Recent work has made rapid progress in utilizing EHRs for predictive modeling tasks in healthcare, including predicting unplanned readmission [1], early prediction of chronic disease [3], adverse event detection [4] and monitoring disease progression [5]. The main idea here is to learn a good representation of a patient's historical health information, in order to improve the performance of the prediction for future risks.

In order to model the dependencies of diagnoses, deep leaning techniques, such as recurrent neural networks, can be employed. Recent work [10, 1, 8, 3, 9] shows that deep learning can significantly improve the prediction performance. To handle the temporality of multivariate sequences, dynamically modeling the

[1] An extended version of such a paper has been included in the proceedings of AMIA 2018, Washington DC[2]

sequential data is necessary. Recurrent neural networks (RNNs), in particular Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), have achieved state-of-the-art performance in handling long-term dependencies and nonlinear dynamics.

In this paper, our goal is to predict the status of multiple diagnoses (or observations), with each diagnosis having multiple severity levels. We form our problem as multi-task learning, which first learns a shared representation from all the features, and then performs task-specific predictions. We propose an attention-based RNN model to monitor patient's longitudinal health information. First, we use an RNN to memorize all the information from historical visits, and then attention mechanisms to measure visit importance. Based on the latent representation, we train multiple classifiers and each focuses on the prediction of a specific task. We perform our model on two applications: predicting chronic states for bone health, and monitoring BloodTest values for cardiovascular disease.

2 Method

The basic component of our framework is gated recurrent unit, which is a state-of-the-art deep learning architecture for modelling long range sequences. To further improve its performance, we apply attention mechanisms to measure the importance of historical sequences. To predict the status of multiple diagnoses, we add a multi-task classification layer on top of the learned representations.

We implement our RNN with Gated Recurrent Units (GRU) [16], which has been shown to have comparable performance as Long-Short Term Memory (LSTM), while employing a simpler architecture.

RNN can remember the past information for future prediction. However, it is limited to only a few latest steps, with more impact from later ones, and may not be able to discover major influences from earlier timestamps. Therefore, we apply attention mechanisms to memorize the effect from long-time dependencies, which have gained success in many tasks.

Our task is to predict the status of multiple measurement results at the time $(t+1)$ given the historical records from x_1 to x_t. Figure 1 shows a high-level overview of the proposed model. Given the information from time 1 to t, the i-th visit's health record x_i is fed into an RNN network, which outputs a hidden state h_i as the representation of the i-th visit. Along with the set of hidden states $\{h_i\}_{i=1}^{t-1}$, we compute their relative importance α_t, and then obtain a *context* state c_t. From the context state c_t and the current hidden state h_t, we can obtain an *attentional hidden* state $\tilde{h}_t$, which is used to predict diagnoses in the $(t+1)$-th visit. For the prediction, we use M softmax classifiers, which correspond to the M different diagnoses, to predict the severity level for each diagnosis. The representation h_t contains the visit information of all the input features, and the task-specific classifier focuses on the prediction of each diagnosis.

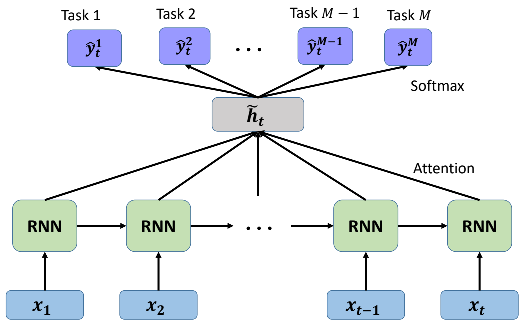

Fig. 1. Overview of the proposed model.

3 Experiments

We conduct experiments on two real-word datasets, and evaluate the performance of the proposed attention-based RNN models compared to other prediction methods.

Study of Osteoporotic Fractures Dataset. The study of osteoporotic fracture (SOF) [**?**] is the largest and most comprehensive study focused on bone diseases. It includes 20 years longitudinal data about osteoporosis of 9,704 Caucasian women aged 65 years and older. Potential risk factors and confounders belong to several groups such as demographics, family history, and lifestyle. We process people's bone health diagnoses of different areas using the bone mineral density (BMD) values by comparison with young healthy references [18], resulting in three BMD levels: normal, osteopenia and osteoporosis.

BloodTest Dataset. This dataset [20] contains multivariate blood tests of 3,000 patients affected by cardiovascular disease from the University Hospital of Catanzaro, Italy. For each patient, there are several blood tests during their in-hospital stay, such as hemoglobin, triglycerides, glucose, and calcium. As suggested by doctors, we pick 12 blood analytes variables which are important to cardiovascular. Each variable has a normal range provided by doctors. Knowing variable transitions in advance can alarm doctors to take actions before the abnormal occurs, in order to reduce the risk of diseases.

As a common issue of EHR, these datasets are irregularly sampled and sparse, so that data preprocessing is needed. For each person, we remove those visits without any monitored variables recorded, and remove patients with less than three visits. We use simple imputation to fill missing variables. For the SOF data, we fill the missing variables with the values in the previous visit. For the BloodTest data, we impute missing sequences (where a single variable is missing entirely) with a clinical normal value. The used datasets with statistics is shown in Table 1.

For each patient, we want to predict the diagnosis results of each visit based on his/her previous records. To validate the performance of the proposed models

Table 1. Statistics of datasets.

Dataset	SOF	BloodTests
Number of patients	5,318	2,055
Number of visits	22,313	18,758
Average number of visits per patient	4.19	9.13
Number of normal claims	25,145	221,642
Number of low abnormal claims	55,399	17,407
Number of high abnormal claims	31,021	79,837
Total number of features	42	17
Number of monitored diagnoses	5	17

in this diagnosis prediction task, we conduct experiments on two categories of methods: baselines and RNN-based models.

We set up two kinds of baselines. The first baseline is to use the median value of each monitored variable from V_1 to V_t to predict V_{t+1} for continuous variables. This is based on a heuristic assumption that the most frequent state is more likely to occur. For each patient, we use his/her most popular health status as the current status, regardless of time variations. The second baseline is a multi-task logistic regression (LR). To predict information at V_{t+1}, we feed the health records at V_t to a logistic regression model with multiple softmax classifiers. This can be viewed as a simplified model of Figure 1 without using RNNs and attention mechanism to learn latent states. This model only considers the effect from the previous one time step, rather than long time history.

Diagnosis Prediction Table 2 shows the accuracy of the proposed approaches in comparison with baselines on the two datasets. For each patient in the testing set, we predict the health conditions for the subsequent visits using his/her historical health records. For the SOF dataset, we predict the probability of BMD states of normal, osteopenia and osteoporosis for different measurements such as hip and femoral neck. For the BloodTest dataset, we predict the probability of each blood analyte falling into normal, low abnormal and high abnormal. The results are averaged over 5 random trials of 5-fold cross validation. *Avg.# Correct* represents the average number of correctly predicted claims of 5 random trials. *Accuracy* represents the ratio between correctly predicted claims and total number of claims to be predicted. For the two datasets, RNN_l, RNN_g and RNN_c can clearly outperform plain RNN. Since the prediction of RNN mostly depends on recent visits, it may not memorize all the past information. Through attention-mechanism, RNN_l, RNN_g and RNN_c can fully take all the previous visit information into consideration, assign different attention scores for past visits, and achieve better performance compared to RNN.

Visit Interpretation The attention mechanism can be used to understand the importance of historical visits to the current visit. As an example, here we analyze the concatenation-based attention mechanism on the SOF dataset. Figure 2 shows a case study for predicting the diagnoses in the sixth visit through the previous five visits.

Table 2. Prediction performance on two datasets.

2*Method	Osteoporosis fracture		Cardiovascular blood test	
(r)2-3 (r)4-5	Avg.# Correct	Accuracy	Avg.# Correct	Accuracy
Median	10,509	0.8209±0.0057	32,253	0.7616±0.0013
LR	10,125	0.7909±0.0069	34,836	0.8225±0.772
RNN	10,769	0.8412±0.0042	36,167	0.8540±0.0051
RNN_l	**10,822**	**0.8454±0.0031**	36,443	0.8605±0.0056
RNN_g	10,805	0.8440±0.0027	36,423	0.8600±0.0059
RNN_c	10,816	0.8449±0.0023	**36,560**	**0.8632±0.0051**

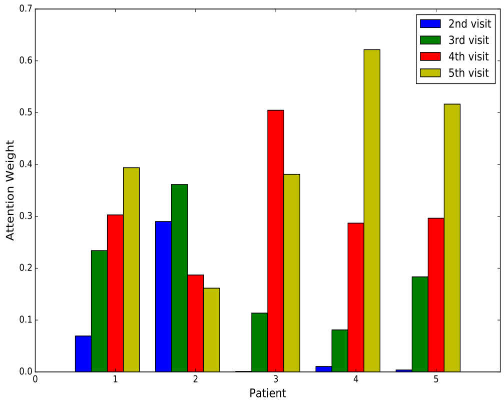

Fig. 2. Attention weights of five persons, each with four visits.

For chronic diseases, the last visit is often the most important since patients' health conditions change slowly. As in the figure, for the first, fourth and fifth patients, the importance of visit increases with time going on. However, this is not always the case due to the complexity of disease progression and impact from risk factors. Table 3 shows the variation of bone mineral density (BMD) diagnoses and attention scores of different visits of the second patient. In each visit, there are five different BMD diagnoses, and the values in the table indicate the severity of bone density loss. Although V_4 and V_5 are closer to V_6 in terms of time, V_2 and V_3 have the same condition as V_6. Thus health records of V_2 and V_3 are more important to V_6. We can see that the attention mechanism correctly assigns larger weights to V_2 and V_3. As for the BloodTest dataset, using attention mechanism to memorize all the past information is also important. An abnormal blood analyte can temporarily turn into normality via medicine, but it may fall back after some time. Therefore, interpreting visit importance through the attention mechanism can help to better monitor disease progression.

In diagnosis prediction, making decisions using very recent record is usually not enough, and it is important to lookup long term health information. To

Table 3. BMD diagnoses and attention scores of one patient with six visits on SOF dataset. 0 is normal, 1 is osteopenia, and 2 (osteoporosis) does not occur for this patient.

Diagnoses\Visits	V_2	V_3	V_4	V_5	V_6
Total hip	0	0	0	0	0
Femoral neck	1	1	0	0	1
Intertrochanteric	0	0	0	0	0
Trochanteric	0	0	0	0	0
Wards	1	1	1	1	1
Attention weights	0.290	0.361	0.187	0.162	–

understand the relationship between the length of patient medical history and the prediction performance, we select 1,000 patients from the BloodTest dataset with more than seven visits. Table 4 shows the accuracy of RNN_l in predicting the diagnoses from V_2 to V_7. We can see that with the number of visit increasing, the performance can often improve. We believe that it is due to the fact that RNN is able to learn better estimates of patient information as it memorizes longer health records.

Table 4. Prediction accuracy for V_2 to V_7 on BloodTest dataset.

Visit	Accuracy
V_2	0.8579
V_3	0.8624
V_4	0.8706
V_5	0.8792
V_6	0.8780
V_7	0.8735

Acknowledgement This work was supported in part by NSF IIS-1218393 and IIS-1514204, and by SISTABENE POR project as PIHGIS POR project.

References

1. Nguyen P, Tran T, Wickramasinghe N, Venkatesh S. Deepr: A convolutional net for medical records. IEEE Journal of Biomedical and Health Informatics. 2016 Dec 1.
2. Qiuling Suo, Fenglong Ma, Giovanni Canino, Jing Gao, Aidong Zhang, Pierangelo Veltri, Agostino Gnasso A Multi-task Framework for Monitoring Health Conditions via Attention-based Recurrent Neural Networks. AMIA 2017, American Medical Informatics Association Annual Symposium, Washington, DC, November 4-8, 2017.
3. Cheng Y, Wang F, Zhang P, Hu J. Risk prediction with electronic health records: A deep learning approach. In Proceedings of the 2016 SIAM International Conference on Data Mining 2016 Jun 30 (pp. 432-440). Society for Industrial and Applied Mathematics.

4. Ma F, Meng C, Xiao H, et al. Unsupervised Discovery of Drug Side-Effects from Heterogeneous Data Sources. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
5. Wang X, Sontag D, Wang F. Unsupervised learning of disease progression models. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining 2014 Aug 24 (pp. 85-94). ACM.
6. Zhou J, Liu J, Narayan VA, Ye J. and Alzheimer's Disease Neuroimaging Initiative. Modeling disease progression via multi-task learning. NeuroImage. 2013 Sep 30;78:233-48.
7. Henriques R, Antunes C, Madeira SC. Generative modeling of repositories of health records for predictive tasks. Data Mining and Knowledge Discovery. 2015 Jul 1;29(4):999-1032.
8. Li H, Li X, Ramanathan M, Zhang A. Prediction and informative risk factor selection of bone diseases. IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB). 2015 Jan 1;12(1):79-91.
9. Suo Q, Xue H, Gao J, Zhang A. Risk Factor Analysis Based on Deep Learning Models. In Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics 2016 Oct 2 (pp. 394-403). ACM.
10. Che Z, Kale D, Li W, Bahadori MT, Liu Y. Deep computational phenotyping. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2015 Aug 10 (pp. 507-516). ACM.
11. Lipton ZC, Kale DC, Elkan C, Wetzell R. Learning to diagnose with LSTM recurrent neural networks. arXiv preprint arXiv:1511.03677. 2015 Nov 11.
12. Choi E, Bahadori MT, Sun J. Doctor ai: Predicting clinical events via recurrent neural networks. arXiv preprint arXiv:1511.05942. 2015 Nov 18.
13. Choi E, Bahadori MT, Sun J, Kulas J, Schuetz A, Stewart W. RETAIN: An Interpretable Predictive Model for Healthcare using Reverse Time Attention Mechanism. In Advances in Neural Information Processing Systems 2016 (pp. 3504-3512).
14. Choi E, Bahadori MT, Song L, Stewart WF, Sun J. GRAM: Graph-based Attention Model for Healthcare Representation Learning. arXiv preprint arXiv:1611.07012. 2016 Nov 21.
15. Ma F, Chitta R, Zhou J, et al. Dipole: Diagnosis Prediction in Healthcare via Attention-based Bidirectional Recurrent Neural Networks. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
16. Chung J, Gulcehre C, Cho K, Bengio Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555. 2014 Dec 11.
17. Zeiler MD. ADADELTA: an adaptive learning rate method. arXiv preprint arXiv:1212.5701. 2012 Dec 22.
18. Bonnick SL. Bone densitometry in clinical practice. Totowa, NJ: Humana Press; 1998 Jun 24.
19. Luong MT, Pham H, Manning CD. Effective Approaches to Attention-based Neural Machine Translation. In Empirical Methods in Natural Language Processing. 2015 Aug.
20. Canino G, Guzzi PH, Tradigo G, Zhang A, Veltri P. On the analysis of diseases and their related geographical data. IEEE journal of biomedical and health informatics. 2015 Oct 30.
21. Bergstra J, Breuleux O, Bastien F, et al. Theano: A CPU and GPU math compiler in Python. In Proc. 9th Python in Science Conf 2010 Jun (pp. 1-7).