CS 598 DLH final Project 2022

Song Chen
stchen4@illinois.edu

Group ID: 200, Paper ID: 288, Diffculty: Easy
Presentation link: https://youtu.be/rI3Sy2n4QM4
Code link: https://github.com/ycbq999/CS598DLH

1 Introduction

This paper is to introduce the rework of one research, Readmission prediction using deep learning on electronic health records [1], This study focus on predicting readmissions for the patients with congestive heart failure (CHF). In the original paper, it presents a deep learning model that group both human driven features and machine driven features as input and feed into a sequential modelling to achieve a better results. The original report also present a financial saving with the introduced prediction model.

2 Scope of reproducibility

In the original paper, the author presents four model characteristics:

HDF: Human derived features are fed as input to the model

MDF: Machine derived contextual embeddings are fed as input to the model

LSTM: The model captures the sequential visit patterns in the EHR

CA: The model adjusts for misclassification costs

The original paper claims that with the comparison of different combinations of four model characteristics, The best result come from the adding all these characteristics into the model prediction. Which reach a testing result of ROC-AUC 0.77 and F1 score 0.51. We successfully achieve the similar result.

Another strong claim from the paper is that the human derived features gives a significant 3% increase in the model AUC compared to a model that relied only on machine derived embeddings. With our reproduction work, we are able to verify the claim

The third claim in the original report 'Capturing the visits sequentially (as they appear in EHR) adds a significant 26% rise in the AUC compared to a memory-less neural network like multi-layer perceptron.' With our reproducion work, this claim is not able to be verified. However, we do find that the sequential visits make a great contribution on the F1 score and Recall values. The paper will discuss about the failure of the this part of rework.

2.1 Addressed claims from the original paper

In this paper, we focus on justifying the follwoing claims:

- All combinations of LSTM, Cost Adjustment, Human Derived Features and Machine Derived Features results in better performance than using either of them alone. Claim to be 0.77 ROC-AUC and 0.51 f1 score.
- HDF gives a 3% boost in AUC compared to a only MDF model
- LSTM gives a 26% boost in the AUC compared to a non LSTM neural network.

3 Methodology

We follow author's methodology to first collect data and form them into the correct format.The author's source code can be found from GitHub. The code is a useful source to grasp the concept of their work.

3.1 Model descriptions

The model first extract EHR data and combine the features into two groups which are then categorized as HDF and MDF.
HDF (Human Derived Features)includes the demographic features such as age, gender, compliance etc. MDF features are mainly ICD procedure and diagnosis codes but convert to embeddings. See the folowing table for HDF features.

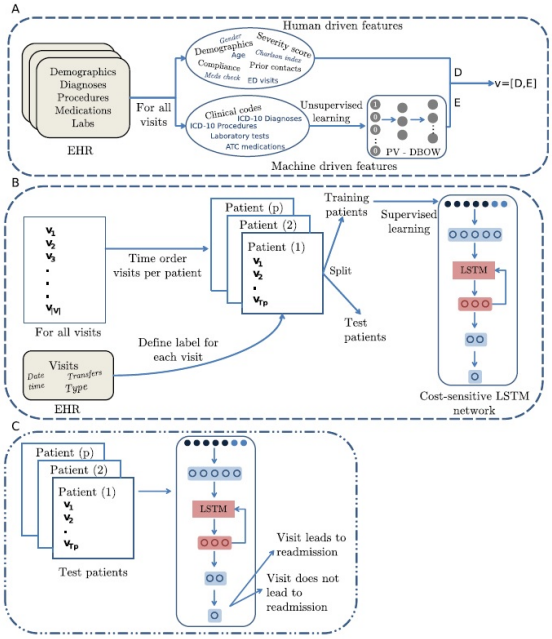

Figure 1: Figure 1: Diagram of Model Design

Age at the time of visit	Discrete
Gender	Binary
Total procedures performed	Discrete
Duration of stay	Discrete
Duration of all stays	Discrete
Type of visit	Discrete
Charlson comorbidity score	Discrete
Number of prior ICU visits	Discrete
Number of prior admissions	Discrete

MDF(Machine Derived Features) are a series of visits by patients. Each visit include a list of diagnosis code represented by ICD-10 and also ATC medication codes. In our reproduction work, we combine ICD9 codes, drug codes and lab codes into the one visit. Then we leverage the Paragraph Vector for Distributed Bag of Words(PV-DBOW) to do the embeddings with 200 dimensions as the final representation of each visit. The reason of chosing PV-DBOW is that it support dynamic window size wrt the size of diagnosis code.

Both HDF and MDF can be put as input separately or together and fed into a sequential model, Which will be implemented as a LSTM network with time distributed layers together they are configured as sequence to sequence prediction model. T

The model then add additional cost term to loss function to give extra penalty of for the misclassification.

The design of the model can be see in figure 1.

3.2 Data descriptions

The original study was conducted using EHR data from southern Sweden. Unfortunately, we could not get the exact data in time. Instead, we redo the work using MIMIC III data and generate the similar HDF and MDF. Because we use different dataset, there are some differences in terms of data extraction and parameter settings.

Data extraction contains two parts:

1. MDF feature representation:

VID	visit IDs (9235*42)
DiagID	diagnosis code (9235*42*[1-1818])

We feed Vids and the corresponding DiagIds into PV-DBOW model and generate embeddings with 200 dimentions named as d2vRNN0-d2vRNN199

2. HDF feature representation:

We also extract and summarize HDF feature shown in table 1.

3. weights and label

Since this study give a weight on the cost function for misclassification, a special attribute is a weight. The label is 30 days readmission represented by binary value.

The final dataset is 387878 rows and 213 columns

3.3 Hyperparameters

Based on the supplement paper, to convert input code to embeddings, author set the size of embedding K to be 200 and the window size was set to be the maximum number of code in one visit which is 1818.

In the combination of MDF and HDF modelling, the model take 209 dimensional vector. Layer 1, 3 and 4 are fully connected dense layer with 128, 32 and 1 node and sigmoid activation functions. The layer 2 of the LSTM layer has size of 64. All the layers contains their own weight and bias.

LSTM contain input, output and out gate that at each step it will update the parameters and propagate the output to the following layers. We feed a sequence of visit data into LSTM network. The sequence window is set to be maximum visits in one patient, which is 42.

The model takes binary cross entropy as error function and add additional cost term to address data imbalance issue. For misclassfication on the

readmission visit, we give max weight of 3. The loss minimization and weight are calculated and updated thought mini-batch gradient descent.

3.4 Implementation

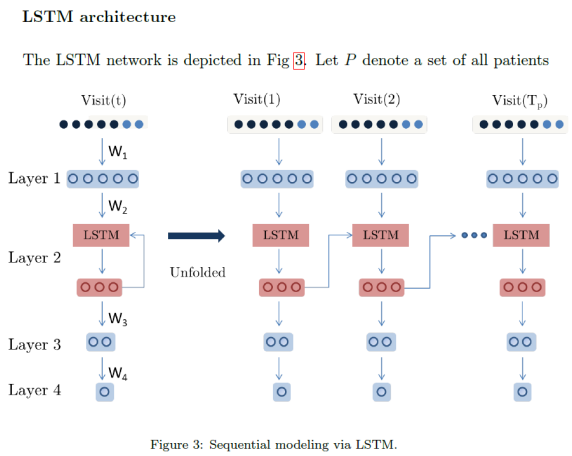

Figure 3: Sequential modeling via LSTM.

The implementation is stratightforward once the data is ready. The original work is done using keras. However We implement this architecture using pytorch. The neuron network contains four different layers. Layer 1, 3 and 4 are FC lienar layer with output feature of 128,32 and 1 respectively. Each linear layer is followed by sigmoid activation function. LSTM layer contains 64 hidden inputs.

We set the visit sequence to be 42 as the maximum visits for one patient. This is a potient issue however, because in order to meet this requirement, we create many dummy visits for each patient. This in fact amply the implance data issue.

To address the imbalance issue, we follow original paper to add extra weight on the cost function

$$\min_{W} \sum_{p=1}^{|P|} \sum_{t=1}^{T_p} [-y'_{pt} log(y_{pt}) - (1 - y'_{pt}) log(1 - y_{pt})] \cdot C(y'_{pt}) \quad (1)$$

$C(y_{pt}{}')$in the equation 1 is the weight on the readmission visit.

3.5 Computational requirements

Hardware Specs: CPU: Intel(R) Xeon(R) CPU @ 2.20GHz Memory: 13GB RAM Space: 100GB Free Space Running platform: Google Colab. (GPU on demand)

4 Results

We randomly split the dataset into 70% as trainingset and 30% as testset. As we want to test the performance of different combination of models and data attributes, we list a table of their performance on F1 score and AUCs as follows.

Model Characteristic				Metric		
HDF	MDF	LSTM	CA	ROC-AUC	F1-SCORE	Recall
0	1	1	1	0.999	0.41	0.82
1	1	0	1	0.999	0.43	0.88
1	0	1	0	0.999	0.46	0.6
1	1	1	0	0.999	0.49	0.67
1	1	1	1	0.999	0.51	0.76

As we mentioned earlier, since we want to use maximum number of visit (42) as sequential data fed into the system, we create more than 90% of dummy visit. This caused so many dummy non readdimission visit. Therefore, the result especially roc-auc is always close to 1. Since the data was preprocessed it is hard to modify the sequence number, we will have to look further to rebalance the data. However, we can recreate the similar f1 scores as we can see from the table above.

4.1 Result 1

The statement that the combination of HDF, MDF, LSTM and cost adjustment has best performance, reaching 0.77 ROC-AUC an 0.51 f1 Score has been verified. As we can see from the table, the full model has the highest f1 score 0.51 and nearly 1 roc-auc score (the roc score issue has been explained earlier)

4.2 Result 2

HDF gives a 3% boost in AUC compared to a only MDF model. We fail to verify this statement as the imbalance data issue. ROC-AUC always close to 1 in our model. However, we can see that HDF features do help increase f1 scores significantly. All other condictions equal, the one without HDF has F1-score only 0.41 while the full model has F1 score of 0.51

4.3 Result 3

LSTM gives a 26% boost in the AUC compared to a non LSTM neural network. Again looking at ROC-AUC, all models perform same. But we can see from F1-score, the model without LSTM has only 0.43 F1-score, compared to full model, LSTM help increase F1-score to 0.51 which is significant.

Despite the fact that the imbalance sample issue. The final model produce a good recall value which is 0.76, which means the prediction on positive readdimmission is nearly 76%. We believe this model need to be improved for sure. However, The model itself give a good promise on predicting the readdmission patients.

5 Discussion

We follow the guidance from the original paper step by step. The mistake we took was to set a long sequence which is not necessary. This mistake cause us to create 90% dummy visit just to fit into the sequential model. The end result is causing more than 94% negative readdmission labels. So our model for sure has a high accuracy and ROCAUC plot.

However, seeing the final result on f1-score so close to the original paper, we think with a little extra modification we can 100% reproduce the similar performance and even better.

5.1 What was easy

The easy part of the this paper is the model design. It is using sequential model and combine with human select features and machine code embedding to work though the LSTM cells. It was not difficult to catch up the author's code and we manage to run the code for most of the part using the sample data.

However the approach the author using is a little out of date, especially on data processing. Author uss PV-DBOW embedding to process data before feed into the neuron network which cause a bit of trouble in term of flexibility. We believe, the similar work can also be done using RNN and normal embeddings.

5.2 What was difficult

I underestimate the difficult of the data extraction. First, Although I manage to contact to the author, I still haven't gain the access of the data. Instead, I use MIMIC III data and create a similar format. The extract the extact dataset is a challenging task is now taking 85% of my time.

6 Communication with original authors

We have successfully contacted to the author who gave me some insight on the data mining. But to gain the access of the data, I still need to ask different department. Instead of waiting, we use MIMICIII data. We manage to process the data and eventually form the same way the author has, although it took a bit of time.

References

[1] Ashfaq, Awais, et al. "Readmission Prediction Using Deep Learning on Electronic Health Records." Journal of Biomedical Informatics, Academic Press, 24 July 2019, https://www.sciencedirect.com/science/article/pii/S1532046419301753

[2] Data Resource Profile: Regional healthcare information platform in Halland, Sweden Ashfaq et al. https://academic.oup.com/ije/article/49/3/738/5701527

[3] C.K. McIlvennan, Z.J. Eapen, L.A. Allen, Hospital readmissions reduction program, Circulation 131 (20) (2015) 1796–1803.

[4] S. Kripalani, C.N. Theobald, B. Anctil, E.E. Vasilevskis, Reducing hospital readmission rates: current strategies and future directions, Annu. Rev. Med. 65 (2014) 471–485.

[5] H. Zhou, P.R. Della, P. Roberts, L. Goh, S.S. Dhaliwal, Utility of models to predict 28-day or 30-day unplanned hospital readmissions: an updated systematic review, BMJ Open 6 (6) (2016) e011060.

[6] S. Basu Roy, A. Teredesai, K. Zolfaghar, R. Liu, D. Hazel, S. Newman, A. Marinez, Dynamic hierarchical classification for patient risk-of-readmission, Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2015, pp. 1691–1700.

[7] B.A. Goldstein, A.M. Navar, M.J. Pencina, J. Ioannidis, Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review, J. Am. Med. Inform. Assoc. 24 (1) (2017) 198–208.

[8] C. Xiao, T. Ma, A.B. Dieng, D.M. Blei, F. Wang, Readmission prediction via deep contextual embedding of clinical concepts, PloS One 13 (4) (2018) e0195024.

[9] W. Farhan, Z. Wang, Y. Huang, S. Wang, F. Wang, X. Jiang, A predictive model for medical events based on contextual embedding of temporal sequences, JMIR Med. Inform. 4 (4).

[10] Y. Maali, O. Perez-Concha, E. Coiera, D. Roffe, R.O. Day, B. Gallego, Predicting 7- day, 30-day and 60-day all-cause unplanned readmission: a case study of a Sydney hospital, BMC Med. Inform. Decis. Making 18 (1) (2018) 1.

[11] H. Wang, Z. Cui, Y. Chen, M. Avidan, A.B. Abdallah, A. Kronzer, Predicting hospital readmission via cost-sensitive deep learning, IEEE/ACM Trans. Comput. Biol. Bioinform.

[12] R. Caruana, Y. Lou, J. Gehrke, P. Koch, M.

Sturm, N. Elhadad, Intelligible models for healthcare: predicting pneumonia risk and hospital 30-day readmission, Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2015, pp. 1721–1730.

[13] M. Jamei, A. Nisnevich, E. Wetchler, S. Sudat, E. Liu, Predicting all-cause risk of 30-day hospital readmission using artificial neural networks, PloS One 12 (7) (2017)e0181173.

[14] S.B. Golas, T. Shibahara, S. Agboola, H. Otaki, J. Sato, T. Nakae, T. Hisamitsu, G. Kojima, J. Felsted, S. Kakarmath, et al., A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: a retrospective analysis of electronic medical records data, BMC Med. Inform. Decis. Making 18 (1) (2018) 44.

[15] A. Rajkomar, E. Oren, K. Chen, A.M. Dai, N. Hajaj, M. Hardt, P.J. Liu, X. Liu, J. Marcus, M. Sun, et al., Scalable and accurate deep learning with electronic health records, npj Digital Med. 1 (1) (2018) 18.