Name:	Seng Chan You
Affiliation:	Department of Biomedical informatics, Ajou University School of Medicine, Suwon, Korea
Email:	applegna@gmail.com
Presentation type (s):	Poster

Development of Clinically Informing application based on Recurrent Neural Network (CIReNN)

Seng Chan You, MD, [1,*], Sanghyung Jin[1,*], Peter R. Rijnbeek, PhD[2], Rae Woong Park, MD, PhD[1,3]

[*]Two authors contributed equally to this work

[1]Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, Korea;

[2] Erasmus University Medical Center, Rotterdam, The Netherlands;

[3]Department of Biomedical Sciences, Ajou University Graduate School of Medicine, Suwon, Korea

Abstract

Temporal dependency and high-dimensionality are major challenges for prediction modeling in the medical field. The Recurrent Neural Network (RNN) model, which is a class of deep learning, can reflect temporal and non-linear relationship among features. We developed CIReNN (Clinically Informing application based on Recurrent Neural Network), which is a generic predictive model using RNN on top off the OMOP-CDM, and evaluated its feasibility and accuracy.

Introduction

Risk prediction for individual patient has been one of the most important themes in clinical research and patient care. Traditional risk prediction model such as the pooled cohort equation is used to identify patients for statins therapy[1]. These traditional approaches usually have used regression-based models, such as the logistic model and the Cox model. These models have been clinically useful and widely accepted because they use a small number of variables which can be easily obtained in clinical practice. However, these models cannot represent more complex relationship among individual predictors and do not model their temporal relationship. The Recurrent Neural Network (RNN) model can represent temporal and non-linear relationship among high-dimensional features.

Purpose

The objective is to build a predictive model based on a recurrent neural network by using features extracted from OMOP-CDM database: CIReNN (Clinically Informing application based on Recurrent Neural Network). CIReNN is expected to facilitate prediction of important clinical events by analyzing flexible and temporal relationships in health care data.

Methods

The following information is extracted from OMOP-CDM: age, sex, diagnosis history and drug history. For each patient, multi-hot label vectors are generated for representing the patient's medical history as shown in Figure 1[2]. These multi-hot label vectors are generated on the basis of individual health care visits, i.e., for each visit of a patient a row is added in the matrix.

Visit 1: Fever, Cough[condition], Tylenol[drug]	[1, 1, 0, 0, 0, …, 0, 0, 1, 0]
Visit 2: Pnuemonia [condition], Tylenol, Amoxicillin[drug]	[0, 0, 1, 0, 0, …, 0, 0, 1, 1]
…	…
Visit 10: Diabetes mellitus [condition], Insulin[drug]	[0, 0, 0, 1, 0, …, 0, 0, 1, 1]

Figure 1. Example of multi-hot vectors representing the occurrence of concepts during a visit

The overall number of features, which means the number of demographics, diagnosis and medication in the whole database, were 8,588 in the Korean national sample cohort (K-NSC) database. To overcome relatively small size and imbalance caused by rarity in medical outcome of interest, the features were selected from the whole feature set. Gradient tree boost is used to select top 800 features by their importance in the training set[3].

RNN with single-layer Gated Recurrent Units (GRU) is implemented[4]. The final model was evaluated in K-NSC database in the format of OMOP-CDM version 5.1.

Proof-of-concept

As a proof-of-concept, we predicted the occurrence of major adverse cardio/cerebrovascular events within 3 years in ischemia-naïve patients. Total target population was divided for training, validation and test set at a ratio of 7:2:1. Only 5,293 (1.7 %) among target patients developed the outcome. The result of CIReNN for this prediction problem is shown in Figure 2.

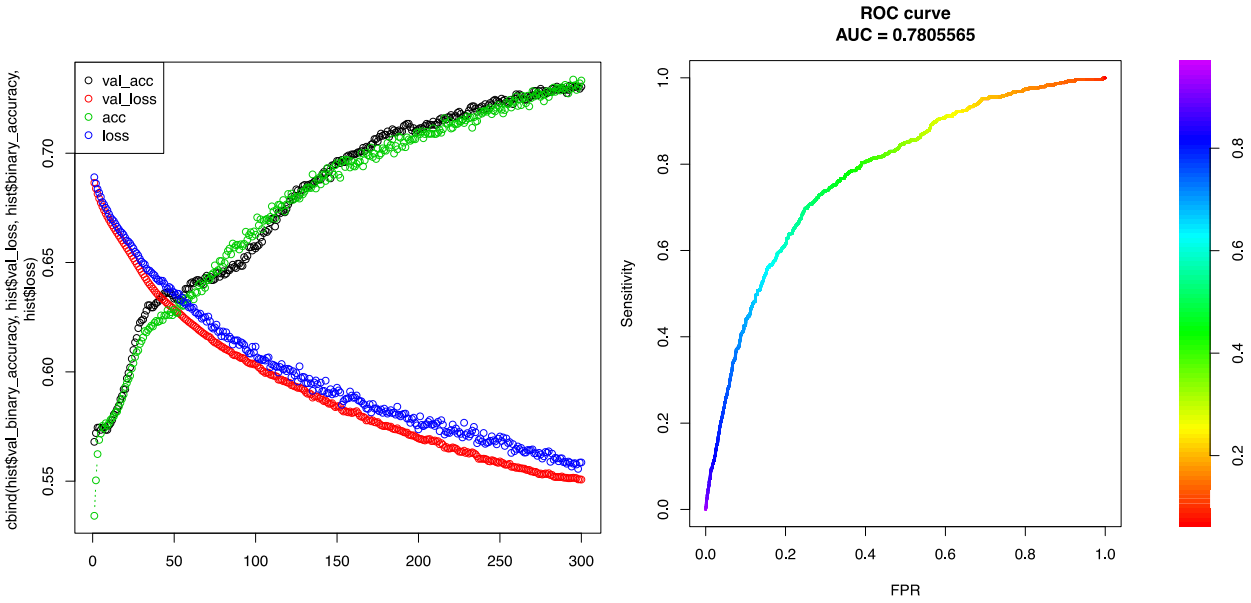

Figure 2. Loss and accuracy in training and validation data set for predicting atherosclerotic cardiovascular disease (left panel). The area under receiver operating characteristic (AUROC) in test set was 0.78 (right panel)

Conclusions

We developed a model for the prediction of atherosclerotic cardiovascular disease in the next 3 year using age, sex, diagnosis history and drug history obtained from health care visits using a recurrent neural network model. We demonstrated the feasibility of this approach on data stored in the OMOP-CDM and aim to further develop this method and compare its performance with other approaches.

References

1. Stone, N. J. *et al.* 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J. Am. Coll. Cardiol.* **63,** 2889–2934 (2014).
2. Choi, E., Bahadori, M. T., Schuetz, A., Stewart, W. F. & Sun, J. Doctor AI: Predicting Clinical Events via Recurrent Neural Networks. *ArXiv151105942 Cs* (2015).
3. Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. *ArXiv160302754 Cs* 785–794 (2016). doi:10.1145/2939672.2939785
4. Chung, J., Gulcehre, C., Cho, K. & Bengio, Y. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. *ArXiv14123555 Cs* (2014).