Frequency-Aware Attention based LSTM Networks for Cardiovascular Disease

Hwin Dol Park
Bio-Medical IT Convergence Research Division
Electronics and Telecommunications Research Institute
Daejeon, Korea
park.hwin@etri.re.kr

Youngwoong Han
Bio-Medical IT Convergence Research Division
Electronics and Telecommunications Research Institute
Daejeon, Korea
hanhero@etri.re.kr

Jae Hun Choi
Bio-Medical IT Convergence Research Division
Electronics and Telecommunications Research Institute
Daejeon, Korea
jhchoi@etri.re.kr

***Abstract*— There are various medical features associated with cardiovascular disease in the EMR data, but the frequency of each medical feature is different. Less frequent feature may be considered as non-critical feature, although cardiovascular disease is closely associated in the cardiovascular disease risk prediction model. We propose a frequency-aware based Attention-based LSTM (FA-Attn-LSTM) that weighs on important medical features using an attention mechanism that considers the frequency of each medical feature. Our model predicts the risk for cardiovascular disease using the ejection fraction as a prediction target and shows RMSE = 3.65 and MAE = 2.49.**

***Keywords*—EMR, attention-based LSTM, time series analysis, deep learning, cardiovascular disease**

I. Introduction

Many hospitals have been able to systematically accumulate medical records for a large number of patients by introducing an electronic medical record (EMR) system. Deep learning has been successfully applied to medical field based on accumulated EMR data [1], [2]. In particular, many studies have been conducted to predict the risk of cardiovascular disease in order to prevent cardiovascular diseases with a high mortality rate globally [3]. Because EMR data is recorded according to the patient's visit to the hospital, it has a time series characteristic such as natural language. Therefore, studies predicting the risk of cardiovascular disease using recurrent neural network (RNN) showed high accuracy [4]. However, previous studies do not consider the frequencies of various medical features in EMR data. EMR data consists of a lot of medical features that appear in various medical test results such as computed tomography, magnetic resonance imaging, and general health check-up. Coronary angiography is an important test to determine the risk of cardiovascular disease, but it is difficult to receive regularly due to cost or time problems. Due to this problem, medical tests for cardiovascular diseases have different test cycles from general health check-up. The characteristic of EMR in which each medical feature is asynchronously recorded for each patient is illustrated in Fig. 1. Unless the frequency of each medical feature is considered in the prediction model, the less frequent feature can be considered as an insignificant feature even though the cardiovascular disease surrogate to be predicted is closely related.

In order to solve the above problem, this paper proposes a frequency-aware based Attention-based LSTM (FA-Attn-LSTM) model using attention mechanism. Attention

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (No.2017-0-00053, A technology development of artificial intelligence doctors for cardiovascular disease).

Patient 1	Feature A	O	O	NA	O	O
	Feature B	O	NA	O	NA	NA
	Feature C	O	O	O	O	NA
	Feature D	O	NA	NA	O	NA
Patient 2	Feature A	O	O	NA	NA	O
	Feature B	NA	NA	O	NA	O
	Feature C	NA	O	NA	O	O
	Feature D	O	NA	O	NA	O
Patient 3	Feature A	O	O	O	NA	O
	Feature B	NA	NA	O	O	NA
	Feature C	NA	NA	O	NA	O
	Feature D	NA	O	O	O	O
Visit		1	2	3	4	5

Fig. 1. Example of frequncy that each medical feature appears in EMR data.

mechanism is a way to obtain correlation between prediction target and input features [5], [6]. The FA-Attn-LSTM model predicts the risk of cardiovascular disease by weighting medical features that are associated with cardiovascular disease, considering the frequency of each medical feature through the attention mechanism.

II. Materials and Methods

A. Materials

In this study, we used outpatient and inpatient visits data related to cardiovascular disease at Asan Medical Center located in Seoul from 2007 to 2016. We used ejection fraction obtained from echocardiography test as a predictive target to predict the risk of cardiovascular disease. The ejection fraction is a measure of heart's ability to contract and is one of the most important indicators of cardiac function. We also used lab test results as indirect components of cardiovascular disease. In the lab test results, we extracted 40 features, from 4551 patients aged 21 and older, clinically well-known features related to cardiovascular disease such as Cholesterol, Albumin, Glucose, and etc. This study approved by the Institutional Review Board (IRB) of the Asan Medical Center. Written Informed consents also from every patient were obtained.

As previously explained, the data consists of ejection fraction and features extracted from the lab test results. Before training the prediction model for risk of cardiovascular diseases, there are several pre-processing steps required [7]. Because outpatient visits are sparse, we aggregated the data by 6 month. If more than one value exists in one group, the last value is extracted. Due to the characteristics of the EMR data, there could not be no visit for six months.

Patient 3	Feature A	O	O	O	NA	O
	Feature B	NA	NA	O	O	NA
	Feature C	NA	NA	O	NA	O
	Feature D	NA	O	O	O	O
Visit		1	2	3	4	5
Medical record for patient 3						

Patient 3	Feature A	1	2	3	3	4
	Feature B	0	0	1	2	2
	Feature C	0	0	1	0	1
	Feature D	0	1	2	3	4
Visit		1	2	3	4	5
Visit-frequency for patient 3						

Fig. 2. Example of generating visit-frequency feature.

For each health record feature, if there are more than 50 missing values, we choose to exclude the feature. If the missing feature is less than 50%, data interpolation was performed using the k-NN interpolation method.

In this study, visit-frequency features for the FA-Attn-LSTM are generated while performing interpolation. The FA-Attn-LSTM model uses the frequency of each feature to calculate the correlation with the target to be predicted considering the frequency of features. As shown in Fig. 2, visit-frequency features represent the frequency of each health record feature. The visit-frequency feature is not increased if the patient is not examined when he visits. The frequency of each feature in the EMR data can be a meaningful feature because the medical examination items are determined according to the current patient status.

B. Methods

The goal of the FA-Attn-LSTM is to predict the value of ejection fraction for next visit. The proposed FA-Attn-LSTM networks structure is shown in Fig. 3. For the ejection fraction prediction, given input sequences $P = (P_1, P_2, ..., P_i)$ with $P_i \in \mathbb{R}^{j \times k}$, where j is the number of visits and k is the index of medical record features, we use P_i to represent medical records of a patient. Each patient P has visit variable such as:

$$P_i = (V_1, V_2, ..., V_j),\ \ V_j \in \mathbb{R}^k$$

$$V_j = (X_1, X_2, ..., X_k),\ \ X_k \in \mathbb{R}$$

In the FA-Attn-LSTM, the input vector V_j which is a visit of a patient consists of medical record features and X_j represents the features of lab test results such as blood sugar, blood pressure, potassium, and etc. We used LSTM networks which show powerful performance in time series analysis as base networks. Given the i-th input sequence P_i, the LSTM networks generates output $h = \{h_1,...,h_j\}$ for each visit. Each input sequence of visit-frequency P' has visit variable such as:

$$P'_i = (V'_1, V'_2, ..., V'_j),\ \ V'_j \in \mathbb{R}^k$$

$$V'_j = (F_1, F_2, ..., F_k),\ F_k \in \mathbb{R}$$

The input sequences of visit-frequency are $P' = (P'_1, P'_2, ..., P'_i)$ with $P'_i \in \mathbb{R}^{j \times k}$, where j is the number of visits and k is index of visit-frequency features when i-th patient sampled, and P'_i represents visit-frequency features of i-th patient. The input sequences of visit-frequency are input to frequency-aware attention model along with the LSTM output. We proposed the frequency-aware attention that adaptively selects the LSTM output relevant the target which is ejection fraction across all visits by referring to the visit-frequency features as follow:

$$e_j = S_e^{\top}\tanh(W_e h_j + U_e[V_j;V'_j]) \qquad (1)$$

$$\alpha_j = softmax(e_j) \qquad (2)$$

where $S_e \in \mathbb{R}^m$, $W_e \in \mathbb{R}^{m \times z}$, and $U_e \in \mathbb{R}^{m \times 2k}$ are a trainable vector and trainable projection matrices, respectively. The parameter m is the size of hidden state for attention model, and z is the dimensions of h_j for the LSTM hidden states. In the alignment model which is used in Equation (2), the weight α_j is score of importance of the i-th LSTM output for predicting the ejection fraction. The attentive context vector C calculated by:

$$C = \sum_{n=1}^{j} \alpha_n h_n \qquad (3)$$

We used the context vector $C \in \mathbb{R}^z$ as a weighted sum of all the LSTM output $h = \{h_1,...,h_j\}$ to predict the target as follow:

$$\hat{y} = S_y^{\top} W_y[h_j;c] + b_y \qquad (4)$$

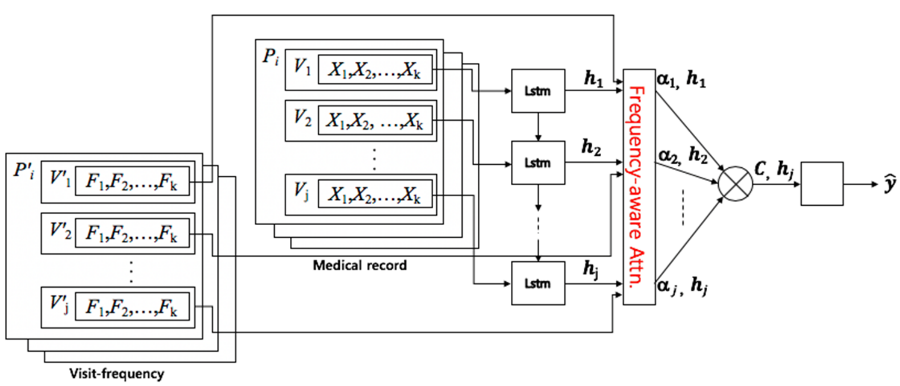

Fig. 3. Architecture of the FA-Attn-LSTM model.

where $S_y \in \mathbb{R}^q$, $W_y \in \mathbb{R}^{q \times 2z}$, and $b \in \mathbb{R}$ are a trainable weight vector, a trainable weight matrix, and bias, respectively. The parameter q is the hidden states for the final output. The final output $\hat{y}$ is the predicted ejection fraction.

III. Experiments

We compared performance of FA-Attn-LSTM with two baseline models. The first one is long short-term memory units (LSTM). The other one is attention-based encoder-decoder network (Attention LSTM) [8] which were developed for neural machine translation. The FA-Atten-LSTM networks have one-layer, each 32 cells, implemented. By assuming that the size of hidden states for LSTM networks and the input vector V_j are approximately one-to-one aligned, we set $m = k$ for simplicity. In order to train the FA-Attn-LSTM, we use ADAM optimizer with a minibatch [9]. The minibatch consists of 16 patients. We train the FA-Atten-LSTM for 10 epochs with an initial learning rate of 0.01 and dropout with probability 0.2 [10]. The FA-Atten-LSTM was trained by standard backpropagation with mean square error as a prediction error metric (we call it as the object function) [11]:

$$O(y, \hat{y}) = \frac{1}{N}\sum_{n=1}^{N} (y_n - \hat{y}_n)^2 \tag{5}$$

where N is the number of patients. All our models are implemented with TensorFlow framework [12] and trained on TITAN Xp GPU.

There are two evaluation metrics in order to compare our FA-Atten-LSTM to baseline models for the ejection fraction prediction. The first metric is root means square error (RMSE) that is denoted as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N} (y_n - \hat{y}_n)^2} \tag{6}$$

where y is the target and $\hat{y}$ is the predicted ejection fraction value. The other metric is mean absolute error (MAE) that is defined as:

$$MAE = \frac{1}{N}\sum_{n=1}^{N} |y_n - \hat{y}_n| \tag{7}$$

For performance evaluation, 40 medical record features extracted from the preprocessing step. The features with past ejection fraction values were used to predict the ejection fraction of the next visit through a total of 41 features.

TABLE I. EJECTION FRACTION PREDICTION RESULTS IN THE ASAN MEDICAL CENTER DATASET

Model	***RMSE***	***MAE***
LSTM	4.292	2.944
Attention LSTM	4.122	2.534
FA-Attn-LSTM	3.657	2.496

As shown in Table 1, the FA-Attn-LSTM model showed the best averaged results among the three models for unseen test set. The ejection fraction can have a value from 0 to 100, and the FA-Attn-LSTM achieved a RMSE = 3.65 and a MAE = 2.49. Unlike other base models, the FA-Attn-LSTM model gives the best results by weighting features with high relevance to the prediction target, considering the frequency of features.

IV. Conclusion

In this study, we introduced the cardiovascular disease risk prediction method using the features indirectly related to cardiovascular disease. To effectively perform the method, we proposed a novel LSTM networks called FA-Attn-LSTM. The FA-Attn-LSTM has the frequency-aware attention that gives weights to input data with high correlation between given input data and predicted object, considering the frequency of input data. To demonstrate the effectiveness of the FA-Attn-LSTM, we performed an experiment to predict the ejection fraction directly related to cardiovascular disease in the Asan medical center data. As a result, the FA-Attn-LSTM model showed the robust performance. Our model has the potential to be useful not only in medical data but also in other fields with asynchronous data.

References

[1] N. J. Yue-Hei, H. Matthew, V. Sudheendra, V. Oriol, M. Rajat, and T. George. Beyond short snippets: Deep networks for video classification. In CVPR, 2015.

[2] Z. Liang, G. Zhang, J. X. Huang, and Q. V. Hu, “Deep learning for healthcare decision making with emrs,” in Proc. Int. Conf. Bioinformat. Biomed., Nov 2014, pp. 556–559.

[3] J. Wang, H. Ding, F. Azamian, B. Zhou, C. Iribarren, S. Molloi, and P. Baldi, “Detecting cardiovascular disease from mammograms with deep learning,” IEEE Transactions on Medical Imaging, 2017.

[4] R. Miotto, L. Li, J. T. Dudley, "Deep learning to predict patient future diseases from the electronic health records" in Advances in Information Retrieval, Springer, pp. 768-774, 2016.

[5] K. Xu, J. Ba, R. Kiros, A. Courville, R. Salakhutdinov, R. Zemel, and Y. Bengio. Show, attend and tell: Neural image caption generation with visual attention. arXiv preprint arXiv:1502.03044, 2015.

[6] Z. Yang, X. He, J. Gao, L. Deng, and A. Smola. Stacked attention networks for image question answering. In CVPR, 2016.

[7] Y. Han, H. Park, J. Choi, H. Jung and M. Lim, "Prediction of left ventricular ejection fraction using time-series deep learning of EMR," Proceedings of KIIT Summer Conference, , pp. 199~200, Dec, 2017.

[8] D. Bahdanau, K. Cho, Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate", International Conference on Learning Representations, 2015.

[9] D. P. Kingma, J. Ba, "Adam: A method for stochastic optimization", Proc. of the International Conf. on Learning Representations (ICLR), 2015.

[10] W. Zaremba, I. Sutskever, O. Vinyals, Recurrent neural network regularization, Sep. 2014, [online] Available: http://arxiv.org/abs/1409.2329.

[11] Y. Qin, D. Song, H. Cheng, W. Cheng, G. Jiang, and G. W. Cottrell. “A dual-stage attention-based recurrent neural network for time series prediction”, In Proceedings of the 26th International Joint Conference on Artificial Intelligence, IJCAI’17, pp. 2627–2633. AAAI Press, 2017.

[12] M. Abadi, A. Agarwal, P. Barham, A. Davis, J. Dean, M. Devin, et al. "Tensorflow: Large-scale machine learning on heterogeneous distributed systems", 2016.