Context-Aware Imputation for Clinical Time Series

Kejing Yin
Hong Kong Baptist University
Hong Kong SAR, China
Email: cskjyin@comp.hkbu.edu.hk

William K. Cheung
Hong Kong Baptist University
Hong Kong SAR, China
Email: william@comp.hkbu.edu.hk

I. Introduction

Missing data has been widely recognized as a key challenge of clinical time series analysis, which hinders the practical application of data-driven approaches to clinical data analytics [1], [2]. Various methods have been proposed to perform the time series imputation to alleviate this issue, yet most of them impose strong assumptions on the missing data, for instance, locality in Gaussian Process based models [3], low-rankness and temporal regularity in matrix/tensor factorization models [4], *etc*. More recently, researchers proposed to apply the Recurrent Neural Networks (RNNs) to tackle the missing data imputation problem for time series, where the RNNs try to capture and summarize the temporal dynamics using hidden state vectors [5]–[7].

Despite the recent success, RNNs have also been found likely to capture more the local properties rather than the global dependencies [8] which could be potentially critical for clinical time series imputation. For example, a patient with kidney disease may exhibit different temporal patterns of blood urea nitrogen (BUN) tests or blood pressure measurements than normal patients. We conjecture that capturing the patient's overall condition from the observed time series and performing imputation with reference to the captured condition would significantly improve the imputation accuracy.

To achieve this, we propose the **C**ontext-**A**ware **T**ime **S**eries **I**mputation (CATSI) framework, which is depicted in Fig. 1. CATSI consists of two major ingredients: the context-aware recurrent imputation and the cross-feature imputation. The former is designed based on the bi-directional RNN to model the longitudinal dynamics over time, and the latter utilizes the cross-feature relationships of the observed variables. Finally, we use a fusion layer to produce the final imputations based on the recurrent and cross-feature imputations.

II. Methodology

A. Notations

We denote the multi-variable time series of a single patient with missing data as $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T\}$, where $\mathbf{x}_t \in \mathbb{R}^D$ is the observation at time step t with the corresponding time stamp denoted as s_t, and T is the length of the time series. Following the existing works [5], [6], we use a masking matrix

This research is partially supported by General Research Fund 12202117 from the Research Grants Council of Hong Kong. The source codes are available via https://github.com/cscihkbu/CATSI.

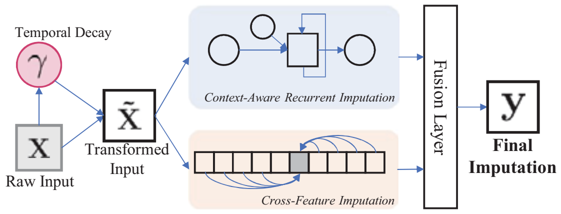

Fig. 1. The framework of Context-Aware Time Series Imputation (CATSI).

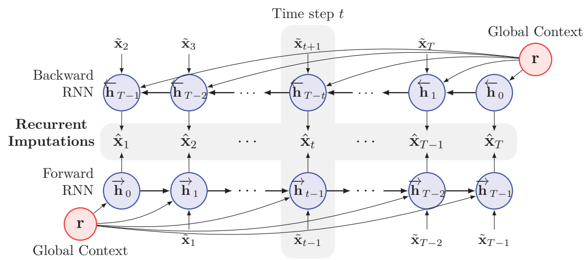

Fig. 2. The architecture of the context-aware recurrent component.

$\mathbf{M}$ to indicate the missingness, *i.e.*, m_t^d equals to one if the d^{th} variable is observed at time step t, zero otherwise. We introduce δ_t^d to represent the time gap between the current observation and the preceding one, *i.e.*,

$$\delta_t^d = \begin{cases} s_t - s_{t-1} + \delta_{t-1}^d & \text{if } t > 1, m_{t-1}^d = 0 \\ s_t - s_{t-1} & \text{if } t > 1, m_{t-1}^d = 1 \\ 0 & \text{if } t = 1 \end{cases} . \tag{1}$$

B. Temporal Decay

We first transform the raw input $\mathbf{X}$ with missing values into a "complete" version $\tilde{\mathbf{X}}$ by using a trainable temporal decay module proposed in [5]. In particular, we calculate a temporal decay factor $\boldsymbol{\gamma}_t$ based on the observation gaps $\boldsymbol{\delta}_t$ at time step t. If x_t^d is not observed, we complete it by decaying its last observation $x_{t'}^d$ towards the empirical mean $\bar{x}^d$. Formally,

$$\boldsymbol{\gamma}_t = \exp\{-\max(\mathbf{0}, \mathbf{W}_\gamma \boldsymbol{\delta}_t + \mathbf{b}_\gamma)\}, \tag{2}$$

$$\tilde{x}_t^d = m_t^d x_t^d + (1 - m_t^d)\left(\gamma_t^d x_{t'}^d + (1 - \gamma_t^d)\bar{x}^d\right). \tag{3}$$

C. CATSI: Context-Aware Recurrent Imputation

We build the recurrent imputation component based on a bi-directional RNN model. As shown in Fig. 2, the recurrent imputation at time step t is generated based on the hidden states of the forward and backward RNNs by:

$$\hat{\mathbf{x}}_t = \mathbf{W}_x[\overrightarrow{\mathbf{h}}_{t-1}; \overleftarrow{\mathbf{h}}_{T-t}] + \mathbf{b}_x, \tag{4}$$

where $[\cdot;\cdot]$ indicates the concatenation of the two hidden states.

TABLE I
INTERQUARTILE RANGE OF THE TRAINING DATA AND IMPUTATION ACCURACY BY NRMSD

	PCL	PK	PLCO2	PNA	HCT	HGB	MCV	PLT	WBC	RDW	PBUN	PCRE	PGLU	Mean
Interquartile range	100-108	3.7-4.4	22-28	135-142	26.8-32.7	8.9-11	86-94	129-330	7-14.1	14.5-17.4	16-44	0.7-1.9	100-148	--
CATSI	0.1738	0.2431	0.2026	0.1958	0.1436	0.1349	0.2534	0.1862	0.2270	0.2130	0.1574	0.2060	0.2602	**0.1998**
3D-MICE	0.2000	0.2632	0.2314	0.2145	0.1505	0.1488	0.2713	0.2294	0.2560	0.2458	0.1846	0.2338	0.2769	0.2235

To explicitly take the patient conditions into account, we propose to incorporate the "global context" of a given time series into the RNN model. In particular, we learn a "context vector" $\mathbf{r}$ from the given time series and at each time step, we update the hidden state based on both the input and the context vector. Formally, we have:

$$\overrightarrow{\mathbf{h}}_0 = \overrightarrow{\mathbf{W}}_h\mathbf{r} + \overrightarrow{\mathbf{b}}_h, \quad \overrightarrow{\mathbf{c}}_0 = \tanh(\overrightarrow{\mathbf{h}}_0), \tag{5}$$

$$\overleftarrow{\mathbf{h}}_0 = \overleftarrow{\mathbf{W}}_h\mathbf{r} + \overleftarrow{\mathbf{b}}_h, \quad \overleftarrow{\mathbf{c}}_0 = \tanh(\overleftarrow{\mathbf{h}}_0), \tag{6}$$

$$\overrightarrow{\mathbf{h}}_{t-1}, \overrightarrow{\mathbf{c}}_{t-1} = \mathrm{LSTM}(\tilde{\mathbf{x}}_{t-1}, \overrightarrow{\mathbf{h}}_{t-2}, \overrightarrow{\mathbf{c}}_{t-2}, \mathbf{r}), \tag{7}$$

$$\overleftarrow{\mathbf{h}}_{T-t}, \overleftarrow{\mathbf{c}}_{T-t} = \mathrm{LSTM}(\tilde{\mathbf{x}}_{t+1}, \overleftarrow{\mathbf{h}}_{T-t-1}, \overleftarrow{\mathbf{c}}_{T-t-1}, \mathbf{r}), \tag{8}$$

where $\mathrm{LSTM}(\cdot)$ is the standard LSTM model.

It now remains to learn the context vector $\mathbf{r}$ from the time series, and here we present two approaches of learning it.

a) Summarizing the Basic Statistics: We compute the basic statistics for each variable, including the empirical mean $\bar{x}^d$, the standard deviation σ^d, the missing rate p^d, and the length of the time series T. We then compute $\mathbf{r}$ using a function f that is approximated by a multi-layer perceptron (MLP) to summarize the overall baseline characteristics of the patient:

$$\mathbf{r} = f\left(\bar{\mathbf{x}}, \boldsymbol{\sigma}, \mathbf{p}, T\right). \tag{9}$$

b) RNN-based Encoder: We use another RNN model as an encoder to capture the more complex temporal dynamics of the time series. Briefly, we input the time series to the encoder RNN and collect the hidden states at the last time step as the context vector $\mathbf{r}$, in that it summarizes the entire time series. We use a standard GRU as the encoder RNN as it has a simpler structure than LSTM.

The two approaches can be applied standalone, and we use both of them together by concatenating the output of the two methods as the context vector.

D. Cross-Feature Imputation

We also incorporate a cross-feature component to allow the feature correlations being effectively utilized. Essentially, we can estimate the value of one variable based on other variables observed at the same time by:

$$\hat{z}_t^d = g(\mathbf{v}_t^d), \quad \mathbf{v}_t^d = \mathbf{W}_f^d\tilde{\mathbf{x}}_t + \mathbf{b}_f^d, \tag{10}$$

where $\hat{z}_t^d$ is the cross-feature imputation of the d^{th} feature at time step t. The d^{th} column of $\mathbf{W}_f^d$ is forced to be zeros to ensure that $\tilde{x}_t^d$ is not involved. $g(\cdot)$ is a (non-linear) function that is approximated by an MLP to fully explore the potentially complex feature correlations.

E. Imputation Fusion

After obtaining the recurrent and the cross-feature imputations, we use a fusion layer similar to [6] to produce the final imputation $\mathbf{y}_t$ by taking their convex combination:

$$\mathbf{y}_t = \boldsymbol{\beta}_t \odot \hat{\mathbf{z}}_t^d + (1 - \boldsymbol{\beta}_t) \odot \hat{\mathbf{x}}_t, \tag{11}$$

$$\boldsymbol{\beta}_t = \mathrm{sigmoid}(\mathbf{W}_\beta[\boldsymbol{\gamma}_t; \mathbf{m}_t] + \mathbf{b}_\beta). \tag{12}$$

F. End-to-end Training

We measure the estimation error by the mean squared deviation (MSD) on the observed entries, *i.e.*,

$$\mathcal{L}(\mathbf{Y}) = \|\mathbf{M} \odot (\mathbf{X} - \mathbf{Y})\|_F^2 / \|\mathbf{M}\|_F^2. \tag{13}$$

To accelerate the convergence, we also accumulate the estimation errors of the recurrent imputations $\hat{\mathbf{X}}$ and the cross-feature imputations $\mathbf{Z}$, leading to the final loss function:

$$\ell = \mathcal{L}(\mathbf{Y}) + \mathcal{L}(\hat{\mathbf{X}}) + \mathcal{L}(\mathbf{Z}). \tag{14}$$

The CATSI model is trained end-to-end by stochastic gradient-based optimization methods, and we normalize each variable of the input data using the min-max normalization by

$$\mathbf{x}^d = \frac{\mathbf{x}^d - \min(\mathbf{x}^d)}{\max(\mathbf{x}^d) - \min(\mathbf{x}^d)}. \tag{15}$$

III. EVALUATION AND RESULTS

We evaluate the performance of CATSI using the dataset provided by the ICHI 2019 DACMI organizers, which is derived from a publicly available real-world ICU dataset, MIMIC-III [9]. The dataset contains 13 common laboratory tests that are irregularly measured for 16,534 patients, half of which (8,267 patients) was provided as training data and the rest half was held out as test data. We summarize the interquartile range of the 13 analytes in Table I.

We use 80% of the training data to train the model and use the remaining 20% to determine the hyper-parameters. Then we freeze the model to generate the imputation for the test data. The performance is evaluated using nRMSD, which is essentially taking root of the MSD between the imputation and the ground truth that are both min-max normalized by Eq. (15) on the missing entries. We report the final evaluation results on the test set in Table I. The average nRMSD over all variables obtained by CATSI is 0.1998, whereas that obtained by 3D-MICE [2] is 0.2235. CATSI achieved 10.6% relative boost against the baseline.

IV. CONCLUSION

We presented a novel CATSI framework for clinical time series imputation, which utilizes the temporal dynamics with the patient conditions explicitly considered by incorporating a context vector in the RNN model. The feature correlations are fully explored by integrating a cross-feature imputation component. CATSI obtained 10.6% relative boost against the state-of-the-art model, which validates its effectiveness.

REFERENCES

[1] C. Xiao, E. Choi, and J. Sun, “Opportunities and challenges in developing deep learning models using electronic health records data: A systematic review,” *Journal of the American Medical Informatics Association*, vol. 25, no. 10, pp. 1419–1428, 2018.

[2] Y. Luo, P. Szolovits, A. S. Dighe, and J. M. Baron, “3D-MICE: Integration of cross-sectional and longitudinal imputation for multi-analyte longitudinal clinical data,” *Journal of the American Medical Informatics Association*, vol. 25, no. 6, pp. 645–653, 2017.

[3] T. Hori, D. Montcho, C. Agbangla, K. Ebana, K. Futakuchi, and H. Iwata, “Multi-task gaussian process for imputing missing data in multi-trait and multi-environment trials,” *Theoretical and Applied Genetics*, vol. 129, no. 11, pp. 2101–2115, 2016.

[4] R. Yu, D. Cheng, and Y. Liu, “Accelerated online low rank tensor learning for multivariate spatiotemporal streams,” in *Proceedings of International Conference on Machine Learning*, 2015, pp. 238–247.

[5] Z. Che, S. Purushotham, K. Cho, D. Sontag, and Y. Liu, “Recurrent neural networks for multivariate time series with missing values,” *Scientific Reports*, vol. 8, no. 1, p. 6085, 2018.

[6] W. Cao, D. Wang, J. Li, H. Zhou, L. Li, and Y. Li, “BRITS: Bidirectional recurrent imputation for time series,” in *Advances in Neural Information Processing Systems*, 2018, pp. 6775–6785.

[7] J. Yoon, W. R. Zame, and M. van der Schaar, “Estimating missing data in temporal data streams using multi-directional recurrent neural networks,” *IEEE Transactions on Biomedical Engineering*, 2018.

[8] A. B. Dieng, C. Wang, J. Gao, and J. W. Paisley, “TopicRNN: A recurrent neural network with long-range semantic dependency,” in *Proceedings of 5th International Conference on Learning Representations, ICLR 2017*, 2017. [Online]. Available: https://openreview.net/forum?id=rJbbOLcex

[9] A. E. Johnson, T. J. Pollard, L. Shen, H. L. Li-wei, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, and R. G. Mark, “MIMIC-III, a freely accessible critical care database,” *Scientific Data*, vol. 3, p. 160035, 2016.