1. Introduction
- Multivariate time-series prediction often becomes challenging due to missing data caused by unreliable sensors and other issues.
- Inaccurate imputation of missing values can degrade the downstream prediction performance.
- We propose a novel approach that can automatically utilize the optimal combination of the observed and the estimated variables to generate not only complete, but also noise-reduced data by our own gating mechanism.
- By jointly training the downstream task module and gating mechanism with adversarial loss, our model produces realistic and helpful imputation to predict the downstream task.
- We also design a synthetic dataset with a known true distribution to verify our method.

2. Proposed method

2.1. Additional dropping of the input data
Colors indicate the status of each variable:
- white for observed variables,
- red for missing variables, and
- blue for additionally dropped variables.

2.2. Model overview
- An end-to-end multitask learning of missing value imputation and forecasting

2.3. GRU cell with decaying mechanism for missing values
- The missing pattern of a variable with respect to time should be considered.
- Thus, we propose to decay the hidden state vector of GRU cell if a variable has been missing for a long while:
  \[
  h_t = \gamma_t \odot h_{t-1} + (1 - \gamma_t) \odot \gamma_\text{decaying}
  \]
- The update functions of GRU with the decaying mechanism is as follows:
  \[
  h_t = \gamma_t \odot h_{t-1} + (1 - \gamma_t) \odot \gamma_\text{decaying}
  \]
  \[
  n_t = \sigma(W_n h_{t-1} + b_n) + \gamma_t \odot h_{t-1}
  \]
  \[
  \gamma_t = \max(0, W_{\gamma_1} y_t^2 + b_{\gamma_1})
  \]
  \[
  \gamma_\text{decaying} = \exp(-\sigma(W_{\gamma_2} h_t + b_{\gamma_2})^2)
  \]

2.4. Gating module for downstream classification task
- The generator output \( Y \) and the raw data \( X \) are mixed by the ratio of gating value, \( \lambda \).
- Since there is a shortfall in the scale of the missing values compared to the observed values, the mixed output \( S \) is compensated with the GRU weights.
- The combination of the observed and the estimated values is fed to the classifier.

3. Synthetic Dataset
- A toy multivariate time-series dataset designed for missing value imputation in time-series data.
- A model can be validated with completely known data distributions.
- Our proposed method outperforms the other imputation methods.
- The classification label is set to be one if the first and the second features without noise are both positive. Otherwise, the label is zero.

4. Experimental results

4.1. Imputation accuracy on PhysioNet Challenge 2012 dataset
- We randomly discard 10% of the validation set to measure the imputation performance.
- Every experiment is conducted five times. The average value and standard deviation (in parentheses) of performances are reported.

4.2. Mortality prediction performances on the PhysioNet dataset
- For the models only designed for imputation without a classifier, an additional training step on the downstream classification task is conducted after imputation.
- Our model performed better than other baselines, suggesting that the proposed model creates an imputation result helpful in predicting the downstream task.

4.3. The mortality prediction performances of ablated models
- Our model outperforms ablated models, indicating that every module of the proposed model has crucial roles in performing the downstream task.
- For example, the proposed decaying mechanism improves the model performance significantly since it helps the model consider the time gap between observations appropriately.

4.4. Effectiveness of the proposed dropping mechanism on the synthetic dataset
- Removing and reconstructing the data improves the model performance both in terms of the missing value imputation and the downstream classification.
- Regardless of the additional missing rate, the method significantly increases the imputation accuracy for unknown variables.
- MSE w/o noise is smaller than MSE w/ noise, indicating that our proposed model successfully captures the true data distribution removing the noise in the data.