2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

High-accuracy Time-series Prediction with Missing Data

Zisheng Li and Masahiro Ogino

***Abstract*— We propose a prediction model which combines time series learning and feature relationship learning on a cohort of aging trends study. We also propose an autoencoder-based method for missing data imputation and salient feature extraction. Evaluation experiments by using 1,645 samples of NHATS (National Health and Aging Trends Study) dataset show that the proposed method achieves high accuracy of physical capacity and cognition activities of elderly people. The concordance correlation coefficient (CCC) is higher than 0.89.**

***Clinical Relevance*— Experimental results show that the time series learning can give effective prediction results on both physical and cognitive scores. Moreover, the relationship learning between the measurement features and the prediction targets might help finding appropriate intervention suggestions on the purpose of improving elderly people's QoL.**

I. INTRODUCTION

We propose a method for predicting physical capacity and cognition activities of elderly people. The prediction model combines time series learning and feature relationship learning. Moreover, we integrate an autoencoder-based module for missing data imputation and salient feature extraction to the prediction model.

II. METHODS

We use the cohort dataset of the National Health and Aging Trends Study (NHATS) [1] to train the proposed prediction model. NHATS begun in 2011 and gathers information on elderly people of ages 65 and older annually. We collect the latest 9 rounds of NHATS data and extract 38 variables as input data. There are 709 samples who have valid data of the 38 variables during all the 9 rounds. And there are 936 samples who have less than 3 missing data points per variable. In total, we use the above 1645 samples to train the proposed model and perform cross validation test.

Fig. 1 shows the architecture of the proposed method. We generate "input data" by replacing the missing values with those of previous/future year. We propose an autoencoding compression module to reconstruct the input data by using a DNN-based autoencoder. We choose the output of an intermediate convolutional layer as "compressed features". We also propose a predicting compression module to predict the data of following time steps. We also choose the output of an intermediate convolutional layers as compressed features of this module. We use such compressed features as the main input for the following time-series learning module where we adopt GRU (gated recurrent unit). In order to catch and interpret "significant" and "meaningful" changes in the life status of elderly people, we add a convolutional self-attention module after the GRU module. On the purpose of learning connection between different features (factors of life status), we proposed to integrate a GAT (Graph Attention Network) module to the GRU module. We consider the NHATS data as graph data with an unknown structure and consider each variable as a node of the graph. We assume that all the nodes (features) are connected to each other and we set the adjacency matrix as all ones. We also apply dense layers to learn static features such as gender and race. Features of time-series learning, static feature learning and graph learning are concatenated to input to further dense layers for final prediction tasks.

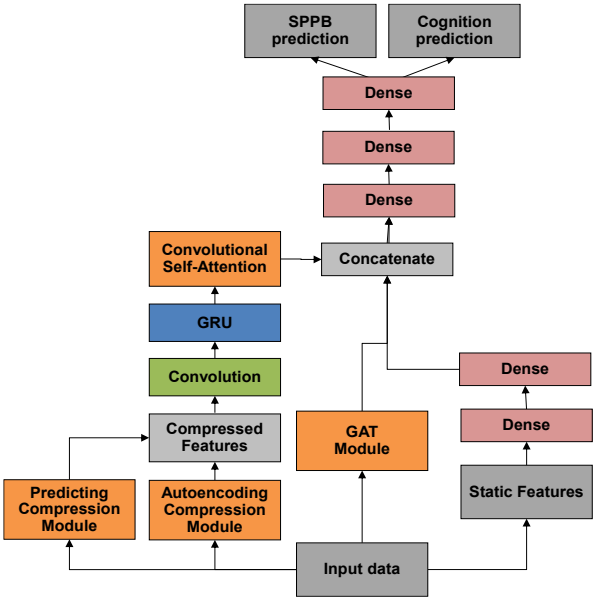

Figure 1. Architecture of the proposed method.

III. RESULTS & CONCLUSION

We run 128-fold of cross validation. We use data of the first 8 rounds for training and predict physical (SBBP, short physical performance battery) and cognitive capacity in the final round. CCC of SBBP and cognition scores is 0.895 and 0.896, respectively. Physical and cognitive capacity can be predicted in the same multi-task prediction model with missing data in high accuracy.

REFERENCES

[1] NHATS Public Use Data, sponsored by the National Institute on Aging through a cooperative agreement with the Johns Hopkins Bloomberg School of Public Health..

Zisheng Li is with Research & Development Group, Hitachi, Ltd., Tokyo, 1858601 Japan, e-mail: zisheng.li.fj@hitachi.com.