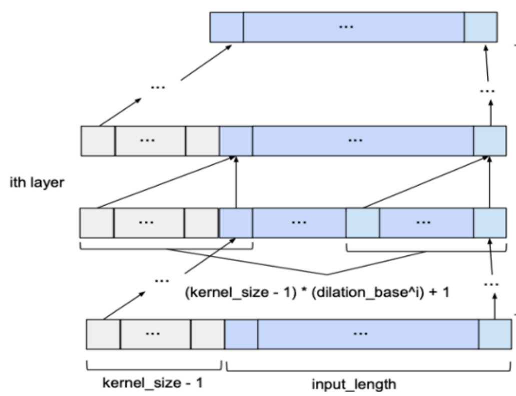
SCIENTIFIC REVIEW

- Rudraksh Mishra

1) What is the aim of the research?

- (Bai et al. 2018) This research aims to empirically evaluate the results of the common association between convolution and recurrent networks for sequence modeling. This is done by a generic temporal convolutional network (TCN) architecture to represent convolutional networks by taking it as the starting point or first layers of the sequential modeling task. By doing this, in a wide range of sequence modeling applications, TCN outperforms baseline recurrent architectures.



- (Niu et al. 2020) While using generative adversarial networks (GAN) for time series outlier detection, we face an issue in finding the appropriate mapping from real-time space to the latent stage for anomaly detection. It also takes a large amount of time for detection. LSTM-based variational autoencoder generation adversarial networks are the proposed alternative approach to solve the above problems.

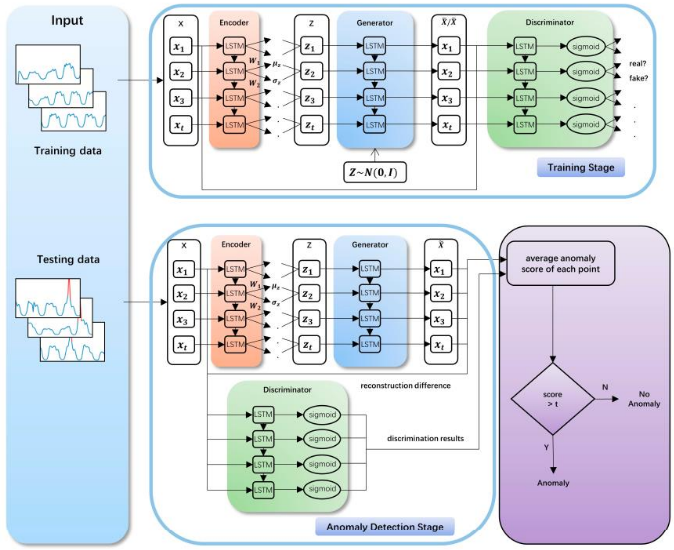


Figure2: LSTM-based VAE-GAN architecture.

- (Tatbul et al. 2018) As seen in the above paper (Niu et al. 2020) that the accuracy of time series anomaly detection was measured in terms of precision, recall, and f1score. The problem with that approach is that they are point-based evaluation metrics, and real-world time series are range-based. This paper focuses on expanding the precision and recall metrics to measure in ranges instead of points.

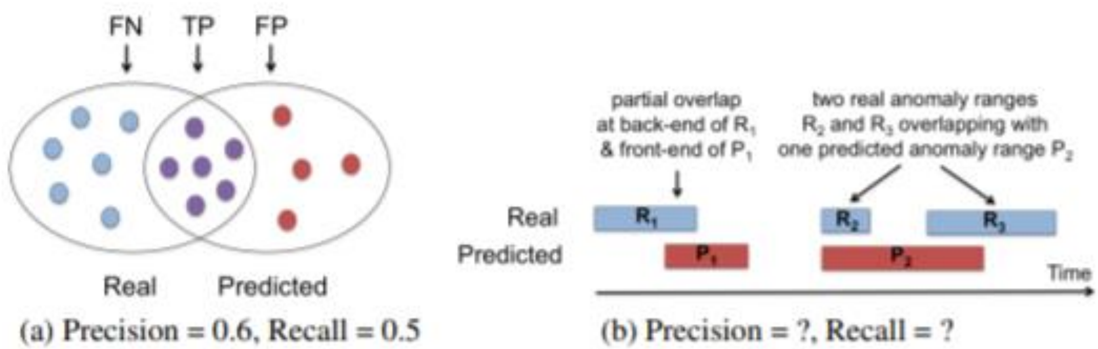


Figure3: Point-based vs range-based anomaly

- (Che et al. 2018) This research aims to construct a modified GRU system a state-of-the-art recurrent neural network (GRU-D) that exploits the missing patterns for effective imputation and better prediction performance by taking the missing patterns in two forms (masking and time intervals).

2) What are the scientific challenges?

- (Bai et al. 2018) There are two main challenges faced while using TCN as the starting point in networks:
    a. In the evaluation and testing phase, RNNs predict by taking the current input, for this they require to main only one single hidden state. Whereas TCNs must take in the raw sequence up to the duration of the effective history, which may necessitate extra memory during evaluation.
    b. Some applications require larger or smaller memory space for making the right prediction. Because TCN's receptive field isn't large enough, it may perform poorly while transferring the model from smaller memory requirements to larger memory requirements.
- (Niu et al. 2020) Some challenges while using LSTM-based VAE GANs are that they are accurate and fast at predicting the outliers/anomalies points. Hence for some domains where anomalies are to be detected in a successive/continuous manner they fail to perform well.
- (Tatbul et al. 2018) The only challenge faced during the development of the precision-recall time-series metrics is the consideration of the right values for bias functions, allowing to weigh several factors differently.
- (Che et al. 2018) One of the main challenges in handling missing values in multivariate time series data is the approach used to characterize missing not at random (MNAR) data. Another challenge faced by the GRU-D method is that the best accuracy is only achieved for larger datasets.

3) Describe the contribution of this research?

- (Bai et al. 2018) Through a different approach of using TCN in initial tasks, the neural network model exhibits longer memory than the traditional recurrent architectures. This helps in better prediction and lower losses in domain applications which requires longer usage of historical data for prediction.

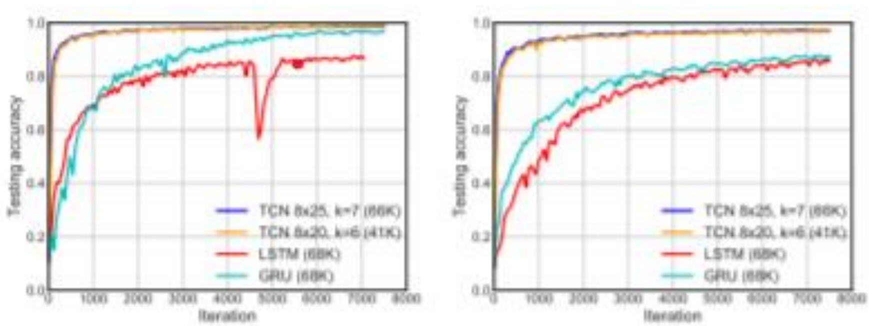


Figure 4: Results on Sequential MNIST and P-MNIST. TCNs outperform recurrent architectures.

- (Niu et al. 2020) By using LSTM based VAE GANs the encoders are built to produce a good mapping from the input time series to the latent space. This is done by using LSTM networks for encoders, generators, and discriminators. This leads to much faster detection of the anomalies with much higher accuracy compared to other similar methods like MAD-GAN, LSTM-AE, LSTM-VAE.

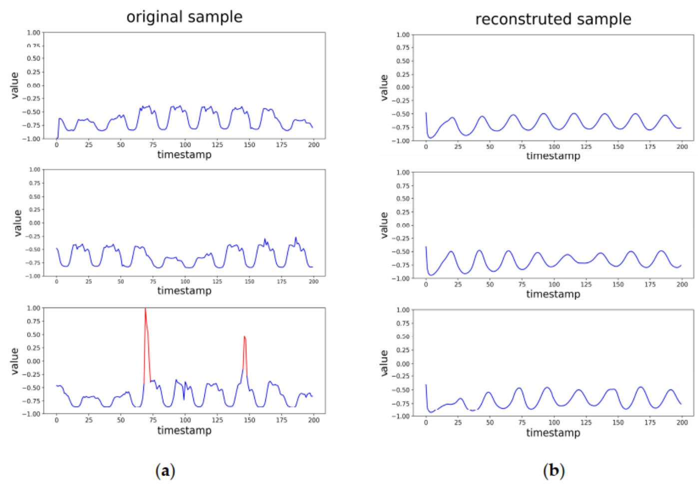


Figure5: a) The red areas represent anomalies, and the time subsequence is required to detect them. b) The reconstructed time subsequence that corresponds to part a's time subsequence.

- (Tatbul et al. 2018) Some of the great difficulties relating to ranges in time series evaluation were solved by taking partial overlaps between real and anticipated ranges into consideration, as well as their relative placements.
- (Che et al. 2018) By implementing masking and time intervals inside the GRU architecture, GRU-D can impute the informative missingness of time series data. This results in GRU-D outperforming GRU while making time series forecasting with a 2.5% higher AUC score on the MIMIC-III mortality case study.

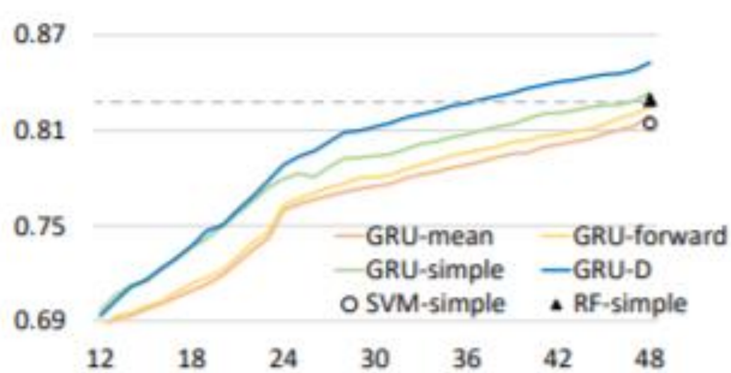


Figure 6: Performance for early predicting mortality on MIMIC-III dataset. x-axis, # of hours after admission; y-axis, AUC score

4) How this contribution differs from existing related work?

- (Bai et al. 2018) In this research, they distill the convolutional network design with a different starting point (TCN). An additive design of TCN is added in front of a convolution network which is made up of dilated, causal 1D convolutional layers that have the same input and output lengths. This method has unique characteristics like - The architecture's convolutions being causal, which makes sure that there is no information leakage from the future to the past. Just like an RNN, the architecture may take any length sequence and translate it to a length-matched output sequence.

- (Niu et al. 2020) All the other similar methods like MAD-GAN, LSTM-AE, LSTM-VAE do not make use of the similar RNN architecture as encoder, generator, and the discriminator of VAE-GAN. This proposed method trains and learns the distribution of the normal data by taking LSTM as encoder, generator, and discriminator for the VAE-GAN. After which the time series is divided into sub-sequences with the use of sliding windows and custom step size. Each of these subsequences is encoded to the vector in the latent space. Then the discriminator outputs another vector which tells whether the previous vector in real0time space follows the similar distribution of the normal training data or not. The method is divided into two stages: the model training stage and the anomaly detection stage. At the model training stage, our model is trained on normal time-series data to learn their distribution, and at the anomaly stage, it calculates the average anomaly score of each point in testing time-series data by determining whether the testing time-series data conform to the normal time-series data distribution.

- (Tatbul et al. 2018) As discussed above the model used in this research is to calculate the modified precision and recall on time series data. Previous approaches differ from this as they do not account for positional bias and do not give an adjustable model.

- (Che et al. 2018) Various ways to deal with missing values in time series have been developed throughout the years. Missingness in RNNs has been addressed in recent work (Lipton et al., 2016; Choi et al., 2015) by concatenating missing entries or timestamps with the input or doing simple imputations. There have, however, been no studies that systematically model missing patterns into RNN for time series classification tasks. In this approach, masking and time interval are two representations of informative missingness patterns that can be used. Time interval captures the input observation patterns, whereas masking informs the model which inputs are noticed (or missing).

Figure 7: Graphical representation of the difference between GRU and GRU-D model

5) Assess the weakness and the strength points?

- (Bai et al. 2018) Since TCNs architecture is newly implemented, there are not many well-established schemes for regularizing and optimizing them. Though TCNs have higher requirements for memory usage during evaluation and testing, they manage well with a long input sequence as they treat, they process it as a whole, rather than processing it sequentially as in RNN. As we know TCN's prediction accuracy drop with a switch in the high memory required domain to lower memory required domain, but they perform well when the memory size of a domain is fixed. This better performance is due to its capability of altering the size of its receptive field in a variety of ways. For example, increasing the filter size, stacking more dilated convolutional layers, or utilizing greater dilation factors are all possible options. TCN also takes care of the exploding and vanishing gradients, which is a huge problem faced in RNNs.
- (Niu et al. 2020) The weakness of using VAE-GAN based on LSTM for outlier detection is that they are not very accurate in detecting a continuous flow/subsequence of outliers as the calculated anomaly score determines only animality points. Though they fail to detect subsequences flow, they are more accurate in other cases as compared to other similar methods (MAD-GAN, LSTM-AE, LSTM-VAE) and detect anomalies with greater speed.
- (Tatbul et al. 2018) Through this approach, we can evaluate the performance of our VAE-GAN based on LSTM by using the modified precision and recall metrics for time series data.
- (Che et al. 2018) By the two representations of informative missingness patterns, GRU-D's use modeling masking and time intervals to characterize missing patterns are not missing completely at random time series data.

6) Explain how this paper is relevant to your project? What are the learned lessons from your reading?

- (Bai et al. 2018) Since the w5 dataset is a huge dataset with requirements of long memory, TCNs can be implemented to see if we get a better improvement in the time-series forecasting than compared to LSTM and GRU.
- (Bai et al. 2018) In our Walmart sales dataset in addition to yearly seasonality, we also have spikes that occur at monthly intervals. By integrating TCN architecture we can add extra time series components to the current time series that encode the current day of the month because the TCN model enables multiple input channels. This may aid in the faster convergence of our model.
- (Niu et al. 2020) Since we will be training our deep neural network on the Walmart product sales dataset, we would need it to detect anomalies in the product sales. As some days might have abnormal recordings of sales based on various reasons. As these abnormalities won't be in general subsequences, we can try to detect the anomaly points using LSTM-Based VAE-GAN.
- (Tatbul et al. 2018) With the help of modified precision and recall metrics, we will be able to determine if LSTM based VAE-GAN (Niu et al. 2020) works well with the identification of outliers in our case study or not.
- (Che et al. 2018) The Walmart product sales time series that we are working on has a lot of intermittent values (short periods of consecutive days with very low or missing sales). I believe

with the help of the GRU-D model we will be able to better forecast the product sales as it will help us fill those random short periods of missing values.

7) References:

[1] Bai, Shaojie, J. Zico Kolter, and Vladlen Koltun. "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling." arXiv preprint arXiv:1803.01271 (2018).

[2] Niu, Zijian, Ke Yu, and Xiaofei Wu. "LSTM-based VAE-GAN for time-series anomaly detection." Sensors 20.13 (2020): 3738.

[3] Tatbul, Nesime, et al. "Precision and recall for time series." Advances in neural information processing systems 31 (2018).

[4] Che, Zhengping, et al. "Recurrent neural networks for multivariate time series with missing values." Scientific Reports 8.1 (2018): 1-12.

[5] Francesco Lässig." Temporal Convolutional Networks and Forecasting (2020).https://medium.com/unit8-machine-learning-publication/temporal-convolutional-networks-and-forecasting-5ce1b6e97ce4