

WHAT IS CLAIMED IS:

1. A system for detecting an anomaly in a multivariate time series, the system comprising:

at least one processor programmed or configured to:

receive a dataset of a plurality of data instances, wherein each data instance comprises a plurality of features, wherein each feature of the plurality of features comprises a time series of data points;

determine a set of target data instances based on the dataset based on a target window segment of data points, wherein each target data instance of the set of target data instances is associated with a first time period, wherein the set of target data instances comprises a time series of data points for each feature of the plurality of features within the first time period, wherein a length of the first time period is equal to a length of the target window segment of data points;

determine a set of historical data instances based on the dataset, wherein each historical data instance of the set of historical data instances is associated with a second time period, wherein the second time period is prior to the first time period, wherein the set of historical data instances comprises a time series of data points for each feature of the plurality of features within the second time period;

generate, based on the set of target data instances, a true value matrix, a true frequency matrix, and a true correlation matrix, wherein, when generating, based on the set of target data instances, the true value matrix, the true frequency matrix, and the true correlation matrix, the at least one processor is programmed or configured to:

generate the true value matrix based on a number of time series of data points in the set of target data instances and a number of time steps in the target window segment of data points for the set of target data instances, wherein the number of times series of data points corresponds to the plurality of features, wherein a first dimension of the true value matrix is equal to a number of the plurality of features and a second dimension of the true value matrix is equal to the length of the target window segment of data points;

generate the true frequency matrix based on a discrete Fourier transform of the true value matrix; and

generate the true correlation matrix based on cosine similarity scores between a plurality of time series of data points of the set of target data instances;

generate a forecast value matrix based on a historical true value matrix, wherein a first dimension of the historical true value matrix is equal to a number of the plurality of features and a second dimension of the historical true value matrix is equal to a length of the second time period, wherein a first dimension of the forecast value matrix is equal to the number of the plurality of features;

generate a forecast frequency matrix based on the true value matrix and the historical true value matrix;

generate a forecast correlation matrix based on the true value matrix and the historical true value matrix;

determine an amount of forecasting error, wherein when determining the amount of forecasting error, the at least one processor is programmed or configured to:

determine the amount of forecasting error between:

the forecast value matrix, the forecast frequency matrix, and the forecast correlation matrix, and

the true value matrix, the true frequency matrix, and the true correlation matrix; and

determine whether the amount of forecasting error corresponds to an anomalous event associated with the dataset of the plurality of data instances.

2. The system of claim 1, wherein, when determining whether the amount of forecasting error corresponds to an anomalous event associated with the dataset of the plurality of data instances, the at least one processor is programmed or configured to:

determine whether the amount of forecasting error satisfies a threshold value of forecasting error; and

determine whether the dataset of the plurality of data instances includes an anomalous event based on determining whether the amount of forecasting error satisfies the threshold value of forecasting error.

3. The system of claim 1, wherein when determining the amount of forecasting error, the at least one processor is programmed or configured to:

concatenate the true value matrix, the true frequency matrix, and the true correlation matrix to generate a forecasting input matrix;

concatenate the forecast value matrix, the forecast frequency matrix, and the forecast correlation matrix to generate a forecasting output matrix; and

determine the amount of forecasting error based on the forecasting input matrix and the forecasting output matrix.

4. The system of claim 3, wherein when determining the amount of forecasting error, the at least one processor is programmed or configured to:

determine a measure of loss associated with a forecasting error mean for batches of input data instances of the dataset of the plurality of data instances based on the forecasting input matrix and the forecasting output matrix;

determine a measure of loss associated with a variance of forecasting error for each batch of input data instances of the dataset of the plurality of data instances based on the forecasting input matrix and the forecasting output matrix; and

determine the amount of forecasting error based on the measure of loss associated with a forecasting error mean for batches of data instances and the measure of loss associated with the variance of forecasting error for each batch of data instances.

5. The system of claim 1, wherein the at least one processor is further programmed or configured to:

generate a historical true value matrix based on a number of time series of data points in the set of target data instances and a number of time steps in a target window segment of data points.

6. The system of claim 5, wherein, when generating the forecast value matrix, the at least one processor is programmed or configured to:

provide the historical true value matrix as an input to a dilated convolutional neural network (CNN) to generate an output of the dilated CNN; and
generate the forecast value matrix based on the output of the dilated CNN.

7. The system of claim 5, wherein, when generating the forecast frequency matrix, the at least one processor is programmed or configured to:

generate a sequence of window segments based on the historical true value matrix;

generate a plurality of correlation matrices based on cosine similarity scores of the sequence of window segments;

provide the plurality of correlation matrices as an input to a convolutional long short-term memory (ConvLSTM) neural network to generate an output of the ConvLSTM neural network;

provide the output of the ConvLSTM neural network as an input to an attention mechanism to generate an output of the attention mechanism; and

generate the forecast frequency matrix based on the output of the attention mechanism.

8. The system of claim 5, wherein, when generating the forecast correlation matrix, the at least one processor is programmed or configured to:

generate a sequence of window segments based on the historical true value matrix;

generate a plurality of frequency matrices based on a discrete Fourier transform of the sequence of window segments;

provide the plurality of frequency matrices as an input to a convolutional long short-term memory (ConvLSTM) neural network to generate an output of the ConvLSTM neural network;

provide the output of the ConvLSTM neural network as an input to an attention mechanism to generate an output of the attention mechanism; and

generate the forecast correlation matrix based on the output of the attention mechanism.

9. A method for detecting an anomaly in a multivariate time series, the method comprising:

receiving, with at least one processor, a dataset of a plurality of data instances, wherein each data instance comprises a plurality of features, wherein each feature of the plurality of features comprises a time series of data points;

determining, with the at least one processor, a set of target data instances based on the dataset based on a target window segment of data points, wherein each target data instance of the set of target data instances is associated with a first time period, wherein the set of target data instances comprises a time series of data points for each feature of the plurality of features within the first time period, wherein a length of the first time period is equal to a length of the target window segment of data points;

determining, with the at least one processor, a set of historical data instances based on the dataset, wherein each historical data instance of the set of historical data instances is associated with a second time period, wherein the second time period is prior to the first time period, wherein the set of historical data instances comprises a time series of data points for each feature of the plurality of features within the second time period;

generating, with the at least one processor and based on the set of target data instances, a true value matrix, a true frequency matrix, and a true correlation matrix, wherein generating, based on the set of target data instances, the true value matrix, the true frequency matrix, and the true correlation matrix comprises:

generating the true value matrix based on a number of time series of data points in the set of target data instances and a number of time steps in the target window segment of data points for the set of target data instances, wherein the number of times series of data points corresponds to the plurality of features, wherein a first dimension of the true value matrix is equal to a number of the plurality of features and a second dimension of the true value matrix is equal to the length of the target window segment of data points;

generating the true frequency matrix based on a discrete Fourier transform of the true value matrix; and

generating the true correlation matrix based on cosine similarity scores between a plurality of time series of data points of the set of target data instances;

generating, with the at least one processor, a forecast value matrix based on a historical true value matrix, wherein a first dimension of the historical true value matrix is equal to a number of the plurality of features and a second dimension of the historical true value matrix is equal to a length of the second time period, wherein a first dimension of the forecast value matrix is equal to the number of the plurality of features;

generating, with the at least one processor, a forecast frequency matrix based on the true value matrix and the historical true value matrix;

generating, with the at least one processor, a forecast correlation matrix based on true value matrix and the historical true value matrix;

determining, with the at least one processor, an amount of forecasting error, wherein determining the amount of forecasting error comprises determining the amount of forecasting error between:

the forecast value matrix, the forecast frequency matrix, and the forecast correlation matrix, and

the true value matrix, the true frequency matrix, and the true correlation matrix; and

determining, with the at least one processor, whether the amount of forecasting error corresponds to an anomalous event associated with the dataset of the plurality of data instances.

10. The method of claim 9, wherein determining whether the amount of forecasting error corresponds to an anomalous event associated with the dataset of the plurality of data instances comprises:

determining whether the amount of forecasting error satisfies a threshold value of forecasting error; and

determining whether the dataset of the plurality of data instances includes an anomalous event based on determining whether the amount of forecasting error satisfies the threshold value of forecasting error.

11. The method of claim 9, wherein determining the amount of forecasting error comprises:

concatenating the true value matrix, the true frequency matrix, and the true correlation matrix to generate a forecasting input matrix;

concatenating the forecast value matrix, the forecast frequency matrix, and the forecast correlation matrix to generate a forecasting output matrix; and

determining the amount of forecasting error based on the forecasting input matrix and the forecasting output matrix.

12. The method of claim 11, wherein determining the amount of forecasting error comprises:

determining a measure of loss associated with a forecasting error mean for batches of input data instances of the dataset of the plurality of data instances based on the forecasting input matrix and the forecasting output matrix;

determining a measure of loss associated with a variance of forecasting error for each batch of input data instances of the dataset of the plurality of data instances based on the forecasting input matrix and the forecasting output matrix; and

determining the amount of forecasting error based on the measure of loss associated with a forecasting error mean for batches of data instances and the measure of loss associated with the variance of forecasting error for each batch of data instances.

13. A computer program product for detecting an anomaly in a multivariate time series, the computer program product comprising at least one non-transitory computer-readable medium including one or more instructions that, when executed by at least one processor, cause the at least one processor to:

receive a dataset of a plurality of data instances, wherein each data instance comprises a plurality of features, wherein each feature of the plurality of features comprises a time series of data points;

determine a set of target data instances based on the dataset based on a target window segment of data points, wherein each target data instance of the set of target data instances is associated with a first time period, wherein the set of target data instances comprises a time series of data points for each feature of the plurality

of features within the first time period, wherein a length of the first time period is equal to a length of the target window segment of data points;

determine a set of historical data instances based on the dataset, wherein each historical data instance of the set of historical data instances is associated with a second time period, wherein the second time period is prior to the first time period, wherein the set of historical data instances comprises a time series of data points for each feature of the plurality of features within the second time period;

generate, based on the set of target data instances, a true value matrix, a true frequency matrix, and a true correlation matrix, wherein, the one or more instructions that cause the at least one processor to generate, based on the set of target data instances, the true value matrix, the true frequency matrix, and the true correlation matrix, cause the at least one processor to:

generate the true value matrix based on a number of time series of data points in the set of target data instances and a number of time steps in the target window segment of data points for the set of target data instances, wherein the number of times series of data points corresponds to the plurality of features, wherein a first dimension of the true value matrix is equal to a number of the plurality of features and a second dimension of the true value matrix is equal to the length of the target window segment of data points;

generate the true frequency matrix based on a discrete Fourier transform of the true value matrix; and

generate the true correlation matrix based on cosine similarity scores between a plurality of time series of data points of the set of target data instances;

generate a forecast value matrix based on a historical true value matrix, wherein a first dimension of the historical true value matrix is equal to a number of the plurality of features and a second dimension of the historical true value matrix is equal to a length of the second time period, wherein a first dimension of the forecast value matrix is equal to the number of the plurality of features;

generate a forecast frequency matrix based on the true value matrix and the historical true value matrix;

generate a forecast correlation matrix based on the true value matrix and the historical true value matrix;

determine an amount of forecasting error, wherein when determining the amount of forecasting error, the at least one processor is programmed or configured to determine the amount of forecasting error between:

the forecast value matrix, the forecast frequency matrix, and the forecast correlation matrix, and

the true value matrix, the true frequency matrix, and the true correlation matrix; and

determine whether the amount of forecasting error corresponds to an anomalous event associated with the dataset of the plurality of data instances.

14. The computer program product of claim 13, wherein the one or more instructions that cause the at least one processor to determine whether the amount of forecasting error corresponds to an anomalous event associated with the dataset of the plurality of data instances cause the at least one processor to:

determine whether the amount of forecasting error satisfies a threshold value of forecasting error; and

determine whether the dataset of the plurality of data instances includes an anomalous event based on determining whether the amount of forecasting error satisfies the threshold value of forecasting error.

15. The computer program product of claim 13, wherein the one or more instructions that cause the at least one processor to determine the amount of forecasting error cause the at least one processor to:

concatenate the true value matrix, the true frequency matrix, and the true correlation matrix to generate a forecasting input matrix;

concatenate the forecast value matrix, the forecast frequency matrix, and the forecast correlation matrix to generate a forecasting output matrix; and

determine the amount of forecasting error based on the forecasting input matrix and the forecasting output matrix;

wherein the one or more instructions that cause the at least one processor to determine the amount of forecasting error cause the at least one processor to:

determine a measure of loss associated with a forecasting error mean for batches of input data instances of the dataset of the plurality of data

instances based on the forecasting input matrix and the forecasting output matrix;

determine a measure of loss associated with a variance of forecasting error for each batch of input data instances of the dataset of the plurality of data instances based on the forecasting input matrix and the forecasting output matrix; and

determine the amount of forecasting error based on the measure of loss associated with a forecasting error mean for batches of data instances and the measure of loss associated with the variance of forecasting error for each batch of data instances.

16. The computer program product of claim 13, wherein the one or more instructions further cause the at least one processor to:

generate a historical true value matrix based on a number of time series of data points in the set of target data instances and a number of time steps in a target window segment of data points;

wherein the one or more instructions that cause the at least one processor to generate the forecast value matrix cause the at least one processor to:

provide the historical true value matrix as an input to a dilated convolutional neural network (CNN) to generate an output of the dilated CNN; and

generate the forecast value matrix based on the output of the dilated CNN.

17. The computer program product of claim 16, wherein the one or more instructions that cause the at least one processor to generate the forecast frequency matrix cause the at least one processor to:

generate a sequence of window segments based on the historical true value matrix;

generate a plurality of correlation matrices based on cosine similarity scores of the sequence of window segments;

provide the plurality of correlation matrices as an input to a convolutional long short-term memory (ConvLSTM) neural network to generate an output of the ConvLSTM neural network;

provide the output of the ConvLSTM neural network as an input to an attention mechanism to generate an output of the attention mechanism; and generate the forecast frequency matrix based on the output of the attention mechanism.