

# Informe final publicable de proyecto Detección de anomalías en sistemas de telecomunicaciones mediante métodos de aprendizaje continuo

Código de proyecto ANII: FMV\_1\_2019\_1\_155850

04/11/2022

**GÓMEZ SENA, Gabriel Pablo** (Responsable Técnico - Científico)

**FERNÁNDEZ PARDO, Alicia** (Co-Responsable Técnico-Científico)

**GARCÍA GONZÁLEZ, Gastón Darío** (Investigador)

**MARTINEZ TAGLIAFICO, Sergio Andres** (Investigador)

---

UNIVERSIDAD DE LA REPÚBLICA. FACULTAD DE INGENIERÍA (Institución Proponente) \\  
TELEFÓNICA MÓVILES DEL URUGUAY - MOVISTAR

## Resumen del proyecto

La detección de anomalías se aplica a múltiples contextos y es particularmente relevante en las redes de telecomunicaciones.

Los sistemas usuales de monitoreo permiten detectar fallas en los equipos o enlaces, pero es menos común disponer de capacidad para detectar la degradación de la performance de un servicio o posibles intentos de ataque o fraude.

El crecimiento y la complejidad de los servicios de telecomunicaciones actuales, requiere el análisis de múltiples fuentes de información. El tráfico de las interfaces de un enrutador, la cantidad de mensajes cursados por un equipo, la cantidad de llamadas, son ejemplos de variables relevantes a analizar.

Actualmente existen plataformas que permiten almacenar y analizar grandes volúmenes de datos (Big Data) que representan una oportunidad para incorporar analítica de detección de anomalías sobre grandes volúmenes de información.

Los datos de monitoreo de red generalmente consisten en cientos de contadores recopilados periódicamente en forma de series de tiempo, lo que resulta en un proceso multivariable (MTS) complejo de analizar. Los métodos tradicionales de detección de anomalías de series de tiempo tienen como objetivo el análisis de series de tiempo univariado, lo que hace que el análisis multivariante sea engorroso y prohibitivamente complejo cuando se trata de datos MTS. En el proyecto se propuso un enfoque novedoso para la detección de anomalías en datos MTS, que aprovecha las redes neuronales convolucionales (CNN) y los variational autoencoders (VAEs). Lo que permite detectar anomalías en los datos de MTS a través de un solo modelo, explotando la información temporal sin sacrificar recursos computacionales y de memoria. En particular, en lugar de utilizar redes neuronales recursivas, grandes filtros causales o muchas capas, el algoritmo propuesto se basa en convoluciones dilatadas (DC) para capturar fenómenos a largo y corto plazo en los datos, evitando arquitecturas profundas complejas y menos eficientes, simplificando así el aprendizaje. Se evaluó la detección de anomalías en el conjunto de datos TELCO, un conjunto de datos de monitoreo de red multidimensional a gran escala recopilado en Telefónica Uruguay, donde los eventos anómalos se detectaron manualmente. etiquetados por expertos durante un lapso de tiempo de siete meses, con una granularidad de cinco minutos. Usando este conjunto de datos, comparamos el desempeño con un amplio conjunto de detectores de anomalías de series temporales tradicionales provenientes de los dominios de procesamiento de señales y aprendizaje automático. Se evaluó también la capacidad de detección de cambios en el punto de operación y detección de "concept drift".

En aras de la reproducibilidad y como contribución adicional, se pondrá a disposición de la comunidad el conjunto de datos TELCO, con la autorización de Telefónica y se publicara en forma abierta el código de los algoritmos implementados.

**Ingeniería y Tecnología / Ingeniería Eléctrica, Ingeniería Electrónica e Ingeniería de la Información / Telecomunicaciones / Procesamiento de señales, operación de redes**

**Palabras clave: detección de anomalías / aprendizaje continuo / big data telecomunicaciones /**

## Introducción

Hay múltiples revisiones sobre detección de anomalías en general [1]–[3], así como específicas para redes de telecomunicaciones [4], [5]. La diversidad de características de los datos y tipos de anomalías dan como resultado una carencia de modelos universales para la detección de anomalías. Como hay varios problemas que se pueden abordar como series temporales, hay un fuerte desarrollo de la comunidad científica en este campo [1], [15], y es común encontrar bibliotecas abiertas implementando los enfoques más tradicionales en la literatura. Un ejemplo es la biblioteca ADTK de python que usamos en este estudio es el python ADTK. Como se señaló en [1], la mayoría de los métodos para detección de anomalías en series temporales univariadas y multivariadas consiste en predecir un valor esperado tomando como base la información del pasado y encontrando un umbral de decisión para definir si la observación de la variable está cerca de la predicción. El cálculo automático y adaptativo de los umbrales de detección sigue siendo un problema de investigación abierto en la literatura.

En los últimos años han surgido enfoques para la detección de anomalías de series temporales basados en tecnologías de aprendizaje profundo [6], [16]. Los modelos generativos como VAEs [8] y GANs [17] han ganado relevancia en el campo de detección de anomalías [12], [13], [18]–[22]. Los modelos VAEs [8], [23], [24] representan una clase de modelos poderosa y ampliamente utilizada para aprender distribuciones de datos complejas. A diferencia de las GANs, una limitación potencial de VAE es la suposición previa de que las representaciones latentes de la muestras son independientes e idénticamente

distribuidas. Si bien esta suposición es común en la literatura, hay investigaciones en curso sobre los beneficios de considerar las covarianzas entre muestras en el tiempo y entre series temporales para mejorar el desempeño del modelo [25]–[28].

Por ejemplo, mientras que el trabajo original [8] asume que los parámetros y las variables latentes están isotrópicamente centrados de forma Gaussiana y las posteriores verdaderas son aproximadamente gaussianas con covarianza aproximadamente diagonal, [27] propone una aproximación capturando correlaciones temporales, considerando un proceso gaussiano en el espacio latente.

El modelado de secuencias de datos mediante una combinación de inferencia variacional y arquitecturas de aprendizaje profundo ha sido ampliamente investigado en otros dominios en los últimos años, principalmente mediante la extensión de VAE a redes neuronales recurrentes (RNN), con arquitecturas como STORN [29], VRNN [30] y Bi-LSTM [31] entre otros. Las capas convolucionales con dilatación también se han incorporado en algunos de estos enfoques [32], [33], permitiendo agilizar el proceso de entrenamiento en base a la posibilidad de paralelización que ofrecen estas arquitecturas.

Pocos artículos sobre detectores basados en aprendizaje profundo han abordado el problema de la detección en tiempo real. En [34], los autores consideran el retraso de alerta en la detección de las llamadas anomalías de rango —es decir, segmentos anómalos contiguos, y miden el desempeño del modelo tomando en consideración el F1-score y el retraso promedio en las alertas. La idea de detección de anomalías de rango es atractiva en la práctica para las aplicaciones del mundo real, ya que el operador generalmente no se preocupa por anomalías puntuales, y es aceptable para un algoritmo activar una alerta para cualquier muestra en una secuencia contigua de anomalías, en la medida en que el retraso de detección está limitado a cierto umbral máximo. El trabajo en [35] generaliza las medidas clásicas de Recall, Precision y F1-score para anomalías de rango y hemos considerado estas métricas de desempeño extendidas para evaluar la propuesta DC-VAE en el conjunto de datos TELCO.

La evaluación comparativa del desempeño del modelo utilizando un conjunto de datos reales etiquetados por experto, se destaca respecto a la mayoría de las propuestas de la literatura que utilizan conjuntos de datos públicos, como los conocidos Yahoo, Numanta [36], NASA [37], u otros, en los que las condiciones de operación son poco realistas, las anomalías pueden ser triviales y las etiquetas están mal asignadas en el proceso de etiquetado [7]. Obtener acceso a conjuntos de datos etiquetado por expertos en el dominio de un entorno operativo es insustituible para la evaluación realista de algoritmos.

## Metodología/diseño del estudio

Los datos secuenciales, como las series de tiempo, generalmente se procesan a través de ventanas deslizantes de largo  $T$ . De esta manera definimos las muestras  $x$  como una matriz de  $M \times T$  entradas, donde  $M$  es el número de variables en el proceso MTS. También definimos  $x(t)$  como un vector  $M$ -dimensional, que representa el MTS en un cierto tiempo  $t$ . Para una entrada  $x$  dada, el modelo VAE entrenado produce dos predicciones diferentes, matrices  $\mu$  y  $\sigma$  del mismo tamaño que  $x$  correspondientes a la parametrización de la distribución de probabilidad que representa mejor la entrada dada. Si el modelo VAE se entrenó (principalmente) con datos que describen el comportamiento normal del sistema monitoreado, entonces la salida para una entrada no anómala no se desviaría de la media  $\mu$  más que un número entero específico  $\alpha$  multiplicado por la desviación estándar  $\sigma$ . Por el contrario, si la entrada presenta una anomalía, la salida no pertenecerá a la región determinada por la media y la desviación estándar previstas.

En el modelo VAE, se supone que las observaciones  $x$  dependen de una variable aleatoria  $z$  que proviene de un espacio latente de menor dimensión. El objetivo es maximizar  $P(x)$ , la probabilidad de las observaciones a través del modelo. Similar a  $x$ ,  $z$  también será una secuencia de longitud  $T$ , pero con un número menor de dimensiones  $J < M$ . En términos formales, dada una muestra de entrada  $x$  caracterizada por una distribución de probabilidad desconocida  $P(x)$ , el objetivo es modelar o aproximar la distribución real de los datos  $P$  utilizando una distribución parametrizada con parámetro  $\theta$ . Para explotar la dimensión temporal de las series, propusimos una arquitectura de codificador/decodificador basada en las CNN, utilizando convoluciones dilatadas (DC). DC es una técnica que expande la entrada insertando espacios entre sus muestras consecutivas. En términos más simples, es lo mismo que una convolución normal, pero implica omitir muestras para cubrir un área más grande de la entrada. Las convoluciones deben ser causales, para que la detección pueda implementarse en tiempo real. Debido a que tales arquitecturas no tienen conexiones recurrentes, a menudo son mucho más rápidas de entrenar que las RNN y no sufren problemas de explosión/desaparición de gradientes. El uso de DC en lugar de convoluciones estándar tiene varias ventajas para el análisis en tiempo real: (i) aumentan el llamado campo receptivo, lo que significa que se puede alimentar la detección con información más antigua; (ii) los DC son computacionalmente más eficientes, ya que brindan mayor cobertura al mismo costo de cómputo; (iii) al usar DC, se omiten los pasos de agrupación, lo que da como resultado un menor consumo de memoria; (iv) finalmente, para el mismo campo receptivo temporal, la arquitectura de red resultante es mucho más compacta.

La arquitectura de la red debe ser tal que los valores de salida dependan de todos los valores de entrada anteriores. La longitud  $T$  de la ventana deslizante juega un papel clave aquí, ya que debe garantizar que la salida en  $t$  dependa de la entrada en ese momento y en  $\{t(-1), t(-2), \dots, t(T+1)\}$ . La forma más sencilla de lograr esto es usar filtros de longitud  $F = 2$  y DCs con factor de dilatación  $d$  igual a  $F$  elevado a la  $h$ -ésima potencia, los cuales crecen potencialmente con la profundidad de capa  $h$  pertenece a  $[0, H-1]$ , donde  $H$  es el número de capas de la red. Posteriormente,  $H$  es el valor mínimo que verifica:  $T \leq 2 \times F$  elevado a  $H-1$ . Esta relación directa entre  $T$  y la arquitectura de la red tiene un fuerte impacto práctico, lo que facilita la construcción del codificador/decodificador en función de la profundidad temporal deseada del análisis. El entrenamiento del modelo se lleva a cabo sobre los datos de operación normal, para capturar la línea de base para la detección de anomalías. Una vez entrenado, el proceso de detección se ejecuta continuamente, rodando la ventana deslizante de longitud  $T$  en un paso de tiempo unitario. En cada tiempo  $t$ , el modelo DC-VAE toma como entrada la matriz  $x$ , construida a partir de las últimas  $T$  muestras observadas en el MTS, y produce como salida las matrices  $\mu$  y  $\sigma$ . A partir de estas dos matrices de salida, la detección de anomalías sólo considera sus valores en el tiempo  $t$ , correspondientes a dos vectores  $\mu(t)$  y  $\sigma(t)$ . Para cada una de las series de tiempo univariadas  $m$ , se detecta una anomalía en el tiempo  $t$  si su valor  $x(t)$  cae fuera de la región de operación normal, definida por su  $\mu(t)$  y  $\sigma(t)$  correspondientes. Más precisamente, se declara una anomalía en la serie de tiempo  $m$  en el tiempo  $t$  si:  $|x(t) - \mu(t)| > \alpha \times \sigma(t)$ , donde  $\alpha = (\alpha_1, \dots, \alpha_m, \dots, \alpha_M)$  es un vector de  $M$  umbrales de sensibilidad de detección, donde cada  $\alpha$  se puede establecer de forma independiente para cada serie temporal, lo que permite una calibración detallada por serie temporal del proceso de detección.

Un estudio reciente alerta sobre las limitaciones de evaluar algoritmos de detección de anomalías en conjuntos de datos de series temporales populares como Yahoo, Numenta o la NASA, entre otros. En particular, se observa que estos conjuntos de datos adolecen de fallas conocidas, como anomalías triviales, densidad de anomalías poco realistas, información básica mal etiquetada y sesgo de ejecución hasta falla. Por esta razón, decidimos evaluar DC-VAE en un conjunto de datos propio de MTS, correspondiente a mediciones reales recopiladas en un ISP móvil operativo que se publicará tan pronto sea aceptada la publicación enviada para evaluación. El conjunto de datos de TELCO corresponde a doce series temporales diferentes, con una granularidad temporal de cinco minutos por muestra, recopiladas y etiquetadas manualmente durante un período de siete meses, entre el 1 de enero y el 31 de julio de 2021. Dividimos el conjunto de datos completo en tres subconjuntos independientes ordenados por tiempo, utilizando mediciones de enero a marzo para el entrenamiento del modelo, de abril para la validación del modelo y de mayo a julio para la etapa de testeo.

Durante los experimentos pudimos observar que el DC-VAE puede rastrear correctamente diferentes tipos de comportamiento en la serie temporal, incluido el fuerte componente diario estacional, pero también la operación durante los días de semana y los fines de semana. Para aplicar DC-VAE para la detección de anomalías, debemos calibrar los umbrales de sensibilidad  $\alpha$ , lo que generalmente se realiza de manera supervisada, basándose en las anomalías etiquetadas disponibles en los conjuntos de datos de entrenamiento y validación. Este paso es el único que requiere cierto nivel de "supervisión" (en el sentido de disponibilidad de etiquetas), pero también podría realizarse de forma autosupervisada, etiquetando las anomalías mediante técnicas de detección de valores atípicos. En nuestro problema específico, cada umbral de sensibilidad  $\alpha$  se calibra por serie de tiempo, maximizando el valor F1 sobre los conjuntos de datos de entrenamiento y validación, haciendo una búsqueda de cuadrícula de valores enteros del 1 al 5. En pocas palabras, decidimos cuántas desviaciones estándar  $\sigma$  se considerarán como tolerancia para la variabilidad de operación normal de los datos. DC-VAE puede detectar diferentes tipos de anomalías presentes en los datos, de naturaleza más transitoria y puntiguda o de carácter más estructural. El etiquetado manual de datos por expertos es propenso a errores humanos, muchas veces debido a la falta de información concluyente disponible para que el operador tome una decisión adecuada. Estas observaciones son realmente críticas para considerar cuando se evalúan detectores de anomalías con datos reales de campo. Como nota relevante en esta dirección, tales complejidades en el proceso de etiquetar correctamente los datos y su interacción con el desempeño real del modelo basado en datos de IA/ML han originado una disciplina bastante novedosa conocida como IA centrada en datos (DCAI), que estudia el problema de la ingeniería sistemática de conjuntos de datos de alta calidad para entrenar modelos de aprendizaje automático. También realizamos un análisis de desempeño cuantitativo de DC-VAE en el conjunto de datos de prueba. Como métricas de desempeño, consideramos una versión elaborada de las métricas de evaluación por muestra utilizadas tradicionalmente, para considerar un enfoque más natural y práctico para aplicaciones de detección de anomalías reales, evaluando el desempeño de detección en forma de rangos temporales de anomalías. Las métricas tradicionales pueden tener sentido para anomalías puntuales donde un verdadero positivo corresponde a una detección correcta en el momento preciso. Sin embargo, muchas anomalías ocurren en forma de múltiples anomalías de puntos consecutivos, que definen un rango de anomalías. En tales escenarios, podría ser suficiente tener una superposición parcial entre el rango de anomalía real y el intervalo de anomalía previsto para considerar una detección correcta. Trabajos anteriores han considerado estas

observaciones, definiendo nuevas métricas que priorizan la detección temprana o tardía, o centrándose principalmente en anomalías de rango. Por lo tanto, tomamos unas definiciones extendidas de recall y precision para generalizarlas a rangos de anomalías, considerando una detección correcta si al menos una de las muestras entre el inicio y el final de la anomalía real está marcada por el modelo. Nos referimos a estas métricas extendidas basadas en rangos como Rr, Pr y F1r, para recall, precision y valor F1, respectivamente. Finalmente, las evaluaciones se reportan de forma independiente para cada uno de los doce TSm de series temporales en el conjunto de datos TELCO. Para mostrar las ventajas de DC-VAE en comparación con el uso de una VAE estándar para la detección de anomalías en series temporales, definimos el VAE estándar (S-VAE) como un modelo de detección de anomalías basado en entradas instantáneas, donde la arquitectura del codificador/decodificador se basa en una red neuronal de alimentación directa totalmente conectada y de 3 capas estándar, y la entrada corresponde al MTS en el momento específico de la detección, es decir,  $T = 1$  en S-VAE. La primera observación es que los resultados logrados son, en general magras, alcanzando puntajes F1r de alrededor del 60% para ocho de las doce series temporales, y por debajo del resto. Esto contrasta en gran medida con las altas puntuaciones de F1 que normalmente se informan en la literatura, cuando se trata de conjuntos de datos simulados o defectuosos [20]. De hecho, como explicamos antes, lidiar con mediciones en campo y conjuntos de datos altamente desbalanceados etiquetados por humanos es más complejo de lo que los resultados en la literatura generalmente reportan: la detección real de anomalías MTS en la práctica es muy compleja. El desempeño es significativamente diferente para algunas de las series temporales, lo que corresponde a la naturaleza diferente y el comportamiento subyacente. Sin embargo, el desempeño superior de DC-VAE en comparación con S-VAE es destacable, mejorando en gran medida tanto la detección de anomalías (Rr) como el desempeño general (F1r), en casi un factor de dos en promedio.

Una evaluación preliminar sobre el bajo desempeño obtenido para algunas de las series temporales revela problemas relacionados con un etiquetado deficiente en algunos casos, así como falta de sensibilidad en otros (es decir, es posible que se necesiten valores de alfa de grano más fino). Aún así, los resultados de DC-VAE en términos de sus capacidades de modelado y seguimiento para datos de series de tiempo multivariados son prometedores, y su aplicación a datos reales permite además evidenciar las dificultades detrás de un problema ampliamente estudiado pero sin resolver. Una evaluación más profunda de DC-VAE en el conjunto de datos TELCO, incluida la evaluación comparativa de otros enfoques de detección de anomalías en este conjunto de datos, es parte del trabajo en curso. Asimismo se continúa con la evaluación de modelos de aprendizaje continuo que adapten el punto de operación en forma dinámica, detectando los "concept drift".

## Resultados, análisis y discusión

El novedoso enfoque DC-VAE para la detección de anomalías en series de tiempo multivariadas, aprovecha las ventajas de convoluciones dilatadas y las VAEs. DC-VAE detecta anomalías en series de tiempo multivariadas, explotando la información temporal sin sacrificar los recursos computacionales y de memoria. En particular, en lugar de utilizar redes neuronales recursivas, grandes filtros causales o muchas capas, DC-VAE se basa en convoluciones dilatadas para capturar fenómenos a largo y corto plazo de los datos, evitando arquitecturas profundas complejas y menos eficientes y por tanto simplificando la fase de aprendizaje. La aplicación de DC-VAE a series reales recopiladas en un ISP móvil mostraron que su arquitectura subyacente es mejor que los VAEs estándar tradicionales para la detección de anomalías en series temporales, y muestran resultados prometedores.

La parametrización de la arquitectura DC-VAE se define básicamente por un solo parámetro (la longitud de la ventana deslizante utilizada para el análisis temporal) y la región de operación normal se puede adaptar fácilmente adaptable por serie mediante el ajuste de un solo valor entero. Estas características presentan importantes ventajas en la práctica.

El análisis de desempeño muestra además que DC-VAE tiene buenas propiedades para su implementación en producción: escalabilidad, fácil ajuste de la región de operación normal, robustez contra anomalías en otras series temporales, así como contra el "concept drift", que también se puede detectar. La comparación cuantitativa en los conjuntos de datos TELCO y SWaT muestran la complementariedad con otros métodos de detección, así como el desempeño a la par para detectores de anomalías MTS de última generación en la literatura. Las ventajas cuantitativas y cualitativas de DC-VAE con respecto a S-VAE evidenció el aporte de las capas de convoluciones para capturar un horizonte de tiempo más largo. La disponibilización del conjunto de datos de TELCO a la comunidad proporciona un entorno real y más representativo para la evaluación y comparación de detectores de anomalías, siendo esta una fuerte contribución para avanzar en el área.

## Conclusiones y recomendaciones

La conformación de un equipo de investigación y desarrollo en el que se amalgamaron perfiles complementarios que integran conocimiento del área de aprendizaje automático, telecomunicaciones con expertos del dominio fue una fortaleza

para lograr resultados en el estado del arte aplicables a un ambiente de producción local. La cooperación con la empresa Telefónica permitió adquirir una base de datos muy rica tanto por la cantidad de datos como por el etiquetado manual experto. La integración de un estudiante de doctorado potenció la propuesta de algoritmos novedosos y su evaluación en relación al estado del arte. La colaboración estrecha con el sector productivo aportó no sólo datos, sino requerimientos desafiantes (solución on-line, escalable para un problema con muchos datos y pocas etiquetas). La autorización para compartir en forma pública los datos etiquetados se considera una contribución importante para la comunidad académica. Asimismo el disponibilizar los algoritmos desarrollados y evaluados.

## Referencias bibliográficas

- [1] A. Blázquez-García, A. Conde, U. Mori, and J. A. Lozano, "A review on outlier/anomaly detection in time series data," *ACM Comput. Surv.*, vol. 54, no. 3, Apr. 2021. [Online]. Available: <https://doi.org/10.1145/3444690>
- [2] M. Gupta, J. Gao, C. Aggarwal, and J. Han, "Outlier detection for temporal data," *Synthesis Lectures on Data Mining and Knowledge Discovery*, vol. 5, no. 1, pp. 1–129, 2014.
- [3] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. 3, Jul. 2009. [Online]. Available: <https://doi.org/10.1145/1541880.1541882>
- [4] M. Ahmed, A. N. Mahmood, and J. Hu, "A survey of network anomaly detection techniques," *Journal of Network and Computer Applications*, vol. 60, pp. 19–31, 2016.
- [5] W. Zhang, Q. Yang, and Y. Geng, "A survey of anomaly detection methods in networks," in *2009 International Symposium on Computer Network and Multimedia Technology*. IEEE, 2009, pp. 1–3.
- [6] G. Pang, C. Shen, L. Cao, and A. V. D. Hengel, "Deep learning for anomaly detection: A review," *ACM Comput. Surv.*, vol. 54, no. 2, Mar. 2021. [Online]. Available: <https://doi.org/10.1145/3439950>
- [7] R. Wu and E. Keogh, "Current Time Series Anomaly Detection Bench-marks are Flawed and are Creating the Illusion of Progress," *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–1, 2021.
- [8] D. P. Kingma and M. Welling, "Auto-encoding Variational Bayes," *arXiv preprint arXiv:1312.6114*, 2013.
- [9] A. v. d. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, "Wavenet: A generative model for raw audio," *arXiv preprint arXiv:1609.03499*, 2016.
- [10] A. P. Mathur and N. O. Tippenhauer, "SWaT: AWater Treatment Testbed for Research and Training on ICS Security," in *IEEE International Work-shop on Cyber-physical Systems for Smart Water Networks (CySWater)*, 2016, pp. 31–36.
- [11] G. García González, S. Martínez Tagliafico, A. Fernández, G. Gómez, J. Acuña, and P. Casas, "DC-VAE, Fine-grained Anomaly Detection in Multivariate Time-Series with Dilated Convolutions and Variational Auto Encoders," in *Proceedings of the 7th Workshop on Traffic Measurements for Cybersecurity (WTMC)*, 2022 (to appear), pp. 1–6.
- [12] D. Li, D. Chen, B. Jin, L. Shi, J. Goh, and S.-K. Ng, "MAD-GAN: Multivariate anomaly detection for time series data with generative adversarial networks," in *International Conference on Artificial Neural Networks*. Springer, 2019, pp. 703–716.
- [13] G. García González, P. Casas, A. Fernández, and G. Gómez, "On the usage of generative models for network anomaly detection in multivariate time-series," *SIGMETRICS Perform. Eval. Rev.*, vol. 48, no. 4, p. 4952, may 2021. [Online]. Available: <https://doi.org/10.1145/3466826.3466843>
- [14] H. Zenati, C. S. Foo, B. Lecouat, G. Manek, and V. R. Chandrasekhar, "Efficient GAN-Based Anomaly Detection," *CoRR*, vol. abs/1802.06222, 2018.
- [15] M. Braei and S. Wagner, "Anomaly Detection in Univariate Time-series: A Survey on the State-of-the-Art," *CoRR*, vol. abs/2004.00433, 2020. [Online]. Available: <https://arxiv.org/abs/2004.00433>
- [16] K. Choi, J. Yi, C. Park, and S. Yoon, "Deep Learning for Anomaly Detection in Time-Series Data: Review, Analysis, and Guidelines," *IEEE Access*, vol. 9, 2021.
- [17] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," in *Advances in Neural Information Processing Systems*, vol. 27, 2014.
- [18] S. Zavrak and M. Iskefiyeli, "Anomaly-based intrusion detection from network flow features using variational autoencoder," *IEEE Access*, vol. 8, pp. 108 346–108 358, 2020.
- [19] H. Zenati, C. S. Foo, B. Lecouat, G. Manek, and V. R. Chandrasekhar, "Efficient GAN-based anomaly detection," *arXiv preprint arXiv:1802.06222*, 2018.
- [20] R.-Q. Chen, G.-H. Shi, W. Zhao, and C.-H. Liang, "A joint model for IT operation series prediction and anomaly detection," *Neurocomputing*, vol. 448, pp. 130–139, 2021.
- [21] J. Donahue, P. Kr̄ahen̄uhl, and T. Darrell, "Adversarial feature learning," *arXiv preprint arXiv:1605.09782*, 2016.
- [22] A. Geiger, D. Liu, S. Alnegheimish, A. Cuesta-Infante, and K. Veeramachaneni, "TadGAN: Time series anomaly detection using generative adversarial networks," in *2020 IEEE International Conference on Big Data (Big Data)*. IEEE, 2020, pp. 33–43.
- [23] C. Doersch, "Tutorial on variational autoencoders," *arXiv preprint arXiv:1606.05908*, 2016.

- [24] D. P. Kingma and M. Welling, "An introduction to variational autoencoders," arXiv preprint arXiv:1906.02691, 2019.
- [25] F. P. Casale, A. V. Dalca, L. Saglietti, J. Listgarten, and N. Fusi, "Gaussian Process Prior Variational Autoencoders," in *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montr al, Canada, 2018*, pp. 10 390–10 401.
- [26] L. Girin, F. Roche, T. Hueber, and S. Leglaive, "Notes on the use of variational autoencoders for speech and audio spectrogram modeling," in *DAFx 2019-22nd International Conference on Digital Audio Effects*, 2019, pp. 1–8.
- [27] V. Fortuin, D. Baranchuk, G. Ratsch, and S. Mandt, "GP-VAE: Deep Probabilistic Time Series Imputation," in *International conference on artificial intelligence and statistics*. PMLR, 2020, pp. 1651–1661.
- [28] S. Ramchandran, G. Tikhonov, K. Kujanp a, M. Koskinen, and H. L ahdesm aki, "Longitudinal variational autoencoder," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2021, pp. 3898–3906.
- [29] J. Bayer and C. Osendorfer, "Learning stochastic recurrent networks," arXiv preprint arXiv:1411.7610, 2014.
- [30] J. Chung, K. Kastner, L. Dinh, K. Goel, A. C. Courville, and Y. Bengio, "A recurrent latent variable model for sequential data," *Advances in neural information processing systems*, vol. 28, 2015.
- [31] S. Shabanian, D. Arpit, A. Trischler, and Y. Bengio, "Variational biLSTMs," arXiv preprint arXiv:1711.05717, 2017.
- [32] Z. Yang, Z. Hu, R. Salakhutdinov, and T. Berg-Kirkpatrick, "Improved variational autoencoders for text modeling using dilated convolutions," in *International conference on machine learning*. PMLR, 2017, pp. 3881–3890.
- [33] G. Lai, B. Li, G. Zheng, and Y. Yang, "Stochastic wavenet: A generative latent variable model for sequential data," arXiv preprint arXiv:1806.06116, 2018.
- [34] H. Xu, W. Chen, N. Zhao, Z. Li, J. Bu, Z. Li, Y. Liu, Y. Zhao, D. Pei, Y. Feng et al., "Unsupervised anomaly detection via variational auto-encoder for seasonal KPIs in web applications," in *Proceedings of the 2018 World Wide Web Conference*, 2018, pp. 187–196.
- [35] N. Tatbul, T. J. Lee, S. Zdonik, M. Alam, and J. Gottschlich, "Precision and Recall for Time Series," *32nd Conference on Neural Information Processing Systems (NeurIPS)*, 2018.
- [36] S. Ahmad, A. Lavin, S. Purdy, and Z. Agha, "Unsupervised Real-time Anomaly Detection for Streaming Data," *Neurocomputing*, vol. 262, pp.134–147, 2017.
- [37] K. Hundman, V. Constantinou, C. Laporte, I. Colwell, and T. Soderstrom, "Detecting spacecraft anomalies using lstms and non-parametric dynamic thresholding," in *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, 2018, pp. 387–395.
- [38] A. Lavin and S. Ahmad, "Evaluating real-time anomaly detection algorithms—the numenta anomaly benchmark," in *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2015, pp. 38–44.
- [39] A. Shahid, G. White, J. Diuwe, A. Agapitos, and O. OBrien, "SLMAD: Statistical Learning-Based Metric Anomaly Detection," in *International Conference on Service-Oriented Computing*. Springer, 2020, pp. 252–263.

## Licenciamiento

Reconocimiento 4.0 Internacional. (CC BY)