Revista Cubana de Ciencias Informáticas
Vol. 13, No. 3, Mes Julio-Septiembre, 2019
ISSN: 2227-1899 | RNPS: 2301
Pág. 107-123
http://rcci.uci.cu

Tipo de artículo: Artículo de revisión
Temática: Reconocimiento de patrones
Recibido: dd/mm/aa | Aceptado: dd/mm/aa | Publicado: dd/mm/aa

Detección de anomalías basada en aprendizaje profundo: Revisión

Anomaly Detection based on Deep Learning: Review

Leyanis López-Avila[1*], **Niusvel Acosta-Mendoza**[1], **Andrés gago-Alonso**[1]

[1]FSA Cadastros Técnicos LTDA. Rua Alceu Amoroso Lima, 786, Sala 510, Caminho das Árvores, Salvador-BA, Brasil.

[*]Autor para correspondencia: leyalopez92@gmail.com

Resumen
La detección de anomalías es una técnica de Minería de Datos que permite el reconocimiento de nuevos patrones con comportamiento inusual, los cuales pueden ser traducidos como acciones no válidas o anómalas sobre los datos. La detección de anomalías ha permitido la identificación y prevención de actividades maliciosas como fraude e intrusiones, entre otros. El uso de técnicas tradicionales para la detección de anomalías ha reportado muy buenos resultados. Sin embargo, en los últimos años se han reportado resultados de mayor relevancia mediante el uso de técnicas de aprendizaje profundo. El objetivo de este reporte es la revisión de los principales y más recientes métodos del estado-del-arte para la detección de anomalías (fraude e intrusiones) basados en aprendizaje profundo (en inglés: Deep Learning), los cuales categorizamos según el tipo de red profunda que utilizan.

Palabras claves: Detección de anomalías basado en aprendizaje profundo, detección de fraude, detección de intrusiones, aprendizaje profundo

Abstract
Anomaly detection is a Data Mining technique that allows the recognition of new patterns with unusual behavior, which can be translated as non-valid actions or anomalies over the data. Anomaly detection has allowed the identification and prevention of malicious activities such as fraud and intrusions, among others. The use of traditional techniques of anomaly detection has reported very good results. However, in the last years, more relevant results have been reported through the use of deep learning techniques. The aim of this report is to give a revision of the principal and most recent state-of-the-art methods for anomaly (fraud and intrusions) detection based on the deep learning technique, which we categorized according to the kind of the used deep neural network.

Keywords: *Deep learning-based anomaly detection, fraud detection, intrusion detection, deep learning*

Introducción

La detección de anomalías es una técnica de Minería de Datos con un amplio espectro de aplicaciones enfocadas en la seguridad social, como, por ejemplo: el análisis de redes informáticas y sociales, análisis de transacciones bancarias, y análisis de datos sensoriales, entre otros (Chandola et al., 2009; Kumar, 2005; Aleskerov et al., 1997; Fujimaki et al.,

Revista Cubana de Ciencias Informáticas
Vol. 13, No. 3, Mes Julio-Septiembre, 2019
ISSN: 2227-1899 | RNPS: 2301
Pág. 107-123
http://rcci.uci.cu

2005; Spence et al., 2001). Esta técnica permite el reconocimiento de patrones que no se comportan de la manera esperada en los datos (Kesavaraj and Sukumaran, 2013; Chandola et al., 2009). En una red informática, patrones de comportamiento inusual podrían significar que una computadora pirateada está enviando datos confidenciales a un destino no autorizado (Kumar, 2005). Diferentes comportamientos en los datos de transacciones con tarjetas de crédito podrían indicar el robo de identidad o de la tarjeta de crédito (Aleskerov et al., 1997). Las lecturas de comportamientos inusuales de un sensor de nave espacial podrían significar un error en algún componente de la nave (Fujimaki et al., 2005). Incluso al trabajar con imágenes médicas, un cambio abrupto en la intensidad de los píxeles en lugares inesperados, puede indicar la presencia de tumores malignos (Spence et al., 2001). Todos estos patrones que no siguen el funcionamiento esperado son conocidos como anomalías y su detección permite la prevención de nuevos ataques, malos funcionamientos, así como la detección a tiempo de tumores. En la Fig. 1 se muestran varios de los campos donde se realiza la detección de anomalías.

Figura 1. Representación de la detección de anomalía mediante sus campos de aplicaciones.

Entre los campos nombrados en la Fig. 1, los más tratados son los de detección de fraude y detección de intrusiones, lo que los convierte en los de mayor interés para su análisis en este trabajo. La detección de fraude, según (Chandola et al., 2009), se refiere a la detección de actividades delictivas que ocurren en organizaciones comerciales como bancos, compañías de tarjetas de crédito, agencias de seguros, compañías de teléfonos celulares y mercado de valores, entre otros. La detección de intrusiones está enfocada en la detección de actividad delictiva (robos, hackeo, ataques informáticos, entre otros) en un sistema informático (Phoha, 2002). Ambos, detección de fraude y detección de intrusiones, pueden ser tratados de dos formas diferentes, detección de uso indebido y detección de anomalías (Javaid et al., 2016; Ahmed and Garcia, 2005; Aravindh et al., 2012). La detección de uso indebido solo puede identificar patrones de ataques ya conocidos mediante la correspondencia de reglas establecidas y patrones, mientras que la detección de anomalías aprende nuevos patrones de ataques basado en comportamientos normales (o anómalos) de los

datos (Kwon et al., 2017; Yu et al., 2017b). Debido a la naturaleza cambiante de los ataques, en este trabajo, nos enfocamos en la detección de anomalías porque permite ajustarse a los cambios en los datos.

Las anomalías han sido definidas de diferentes formas por la comunidad científica a lo largo de los años (Dixon, 1950; Grubbs, 1969; Elashoff, 1972; Barnett and Lewis, 1978). En (Dixon, 1950), los autores refieren una anomalía como un valor que es dudoso a los ojos del analista. En (Grubbs, 1969), una anomalía se definió como una observación que parece desviarse notablemente de otros miembros de la muestra en la que ocurre. Elashoff (Elashoff, 1972) definió una anomalía como una observación que es extrema en algún sentido o viola el patrón aparente de las otras observaciones. Barnett (Barnett and Lewis, 1978) presentó una anomalía como una observación (o subconjunto de observaciones) que parece ser inconsistente con el resto de los datos. En (Hawkins, 1980), una anomalía se definió como una observación que se desvía tanto de las otras observaciones como para despertar sospechas de que fue generada por un mecanismo diferente. Otra definición de anomalía se presentó como un punto de datos que es significativamente diferente de los datos restantes (Aggarwal, 2017). Todas estas definiciones de anomalías comparten el mismo núcleo de funcionamiento. Por lo tanto, en este trabajo, utilizamos la siguiente definición genérica: una anomalía es una instancia de datos que contiene valores muy diferentes al resto del conjunto de datos analizado.

Los métodos de detección de anomalías pueden ser clasificados acorde a la naturaleza de la entrada, el tipo de la anomalía, el etiquetado de los datos, o el tipo de salida que devuelve el método (Chandola et al., 2009; Baddar et al., 2014). La entrada de estos métodos es una colección de instancias (objetos, registros, puntos, patrones, entre otros) (Tan et al., 2005) la cual puede ser univariable o multivariable (Chandola et al., 2009). Los tipos de anomalías son tres, puntuales, contextuales y colectivas (Chandola et al., 2009; Baddar et al., 2014; Parmar and Patel, 2017). Dependiendo del grado de disponibilidad de las etiquetas, los métodos de detección de anomalías también se pueden clasificar en uno de los tres modos siguientes: detección de anomalías supervisadas, semi-supervisadas y no supervisadas (Chandola et al., 2009). Otra forma pudiera ser según la forma de retornar la anomalía detectada, usualmente, se hace en forma de puntuaciones o etiquetas (Chandola et al., 2009). Por último, una de las clasificaciones más comunes es de acuerdo con las técnicas utilizadas para la detección de anomalías, tales como análisis estadístico, aprendizaje automático, teoría de la información y teoría espectral, entre otras (Chandola et al., 2009; Baddar et al., 2014). En busca de una mejor asimilación de lo antes mencionado, en la Fig. 2 se muestra una distribución de las clasificaciones para los métodos de detección de anomalías.

Entre las técnicas de aprendizaje automático, el aprendizaje profundo ha tomado gran popularidad en la comunidad científica, debido a los muy buenos resultados alcanzados en disimiles temas como el procesamiento de imágenes, el procesamiento de rostros (Sun et al., 2014), dígitos (Lecun et al., 1998), texto (Jaderberg et al., 2014) y tipos de letras (Wang et al., 2015), así como en la detección de anomalías (Kwon et al., 2017; Kakanakova and Stoyanov, 2017; LeCun et al., 2015). Estas razones aumentan nuestro interés en el estudio y análisis del uso de esta técnica.

Revista Cubana de Ciencias Informáticas
Vol. 13, No. 3, Mes Julio-Septiembre, 2019
ISSN: 2227-1899 | RNPS: 2301
Pág. 107-123
http://rcci.uci.cu

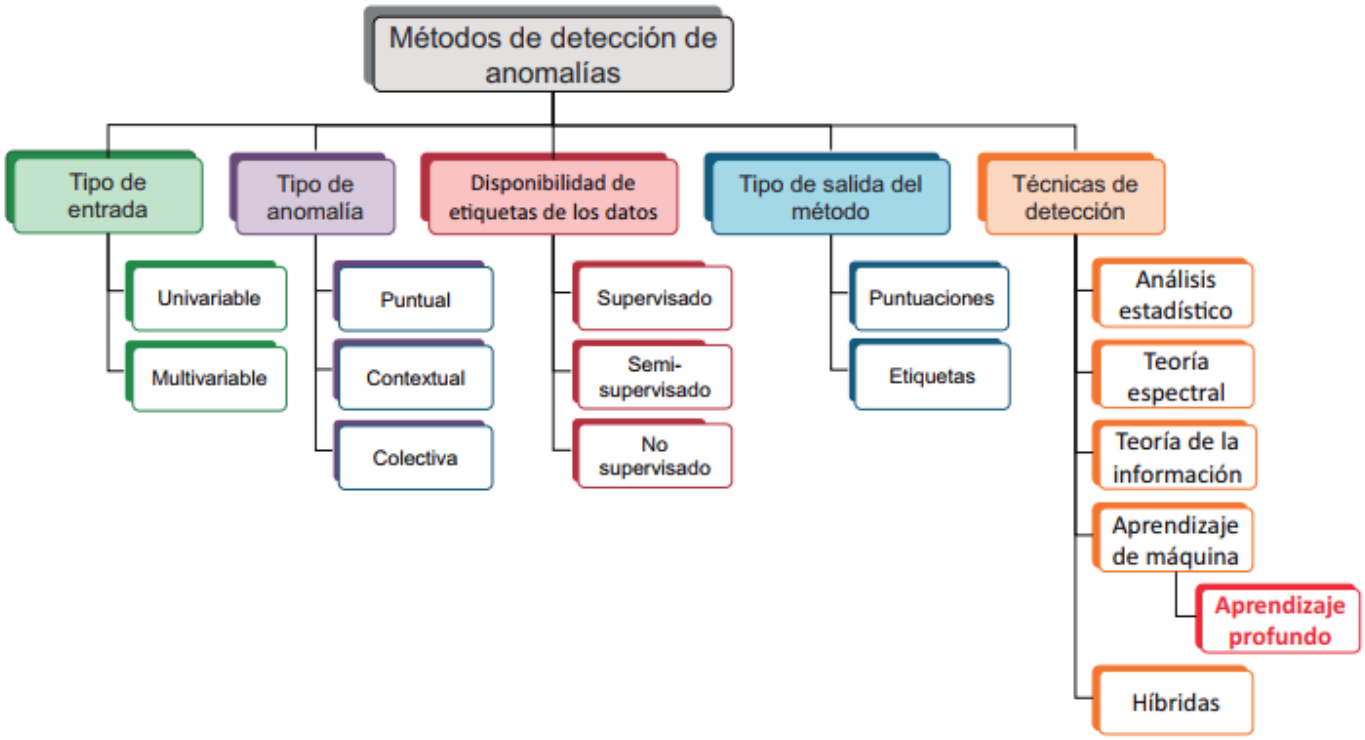

Figura 2. Clasificaciones para los métodos de detección de anomalías.

Los algoritmos basados en la técnica de aprendizaje profundo están motivados por el campo de la inteligencia artificial, y tratan de emular la habilidad cognitiva del cerebro humano (Najafabadi et al., 2016). Comúnmente estos algoritmos hacen uso de la estructura de datos conocida como red neuronal (Rumelhart et al., 1985), a la cual se le han realizado modificaciones creando nuevos tipos de redes destinadas a trabajar con diferentes tipos de datos o funcionalidades específicas. Entre estas nuevas estructuras podemos mencionar: los *AutoEncoders* (AEs), las *Deep Neural Networks* (DNN) (Goodfellow et al., 2016; Chollet, 2017; Bengio, 2009), las *Restricted Boltzmann Machines* (RBM) (Smolensky, 1986; Wang and Raj, 2017), las *Deep Belief Networks* (DBN) (Wang and Raj, 2017; Hinton et al., 2006), las *Convolutional Neural Networks* (CNN) (Wang and Raj, 2017; Hubel and Wiesel, 1962), y las *Recurrent Neural Networks* (RNN) (Wang and Raj, 2017). Aunque estas estructuras sean diferentes, todas son redes neuronales porque mantienen la estructura básica de neuronas, capas y conexiones entre neuronas utilizando funciones de activación lineales y no lineales. Estas redes trabajan con más de dos capas de profundidad (Patterson and Gibson, 2017) donde la combinación de varias capas de activación, representación y muestreo (*pooling* en inglés) permiten la extracción automatizada de representaciones de datos complejos a altos niveles de abstracción (Najafabadi et al., 2016; Bengio, 2009; Bengio et al., 2013; Bengio, 2013).

Revista Cubana de Ciencias Informáticas
Vol. 13, No. 3, Mes Julio-Septiembre, 2019
ISSN: 2227-1899 | RNPS: 2301
Pág. 107-123
http://rcci.uci.cu

Las redes antes mencionadas, además de ser utilizadas por sí solas, pueden ser combinadas para un mejor comportamiento. Una de las estrategias más utilizadas es la GAN (de sus siglas en inglés *Generative Adversarial Network*) (Goodfellow et al., 2014), la cual consiste en vincular dos redes donde la salida de la primera es la entrada de la segunda. Usualmente la primera red es conocida como generador y la segunda como discriminador. La red generadora, provee datos en el espacio de los datos de entrenamiento y trata de confundir al discriminador el cual a su vez aprende a identificar las muestras falsas. Esta relación interactiva entre ambas redes logra una optimización simultánea a través de un juego de minimax para dos jugadores (en inglés *twoplayer minimax game*).

Por las razones antes mencionadas, este documento se enfocó en el estudio de los trabajos de detección de anomalías basados en la técnica de aprendizaje profundo, para los temas: detección de fraude y detección de intrusiones. Las contribuciones de este trabajo son una revisión del estado-del-arte de la detección de anomalía y un análisis crítico del mismo.

Materiales y métodos

En esta sección se realiza un análisis de los trabajos más recientes relacionados a la detección de anomalías, específicamente la detección de fraude e intrusiones, que se basan en técnicas de aprendizaje profundo. Para esto se inicia con una breve explicación del funcionamiento de los métodos para la de detección de anomalías.

Un flujo de funcionamiento para los métodos de detección de anomalías parte de un conjunto de datos, los cuales son representados de la forma más adecuada para su análisis, llámese extracción de características. Con estas nuevas representaciones se hace el análisis de los datos y entrenamiento de los algoritmos para que sean capaces de discriminar entre datos anómalos y normales. Estos modelos entrenados son capaces de discriminar nuevos datos en un futuro. En la Fig. 3 se muestra un ejemplo de este flujo de funcionamiento tomando como datos transacciones de tarjetas de créditos.

Detección de fraude

El fraude, de acuerdo a la Asociación de Examinadores de Fraude, se define como el uso de la propia ocupación para el enriquecimiento personal a través del uso indebido deliberado o la aplicación de los recursos o activos de la organización empleadora (Kou et al., 2004; Richhariya and Singh, 2012). También, el *Concise Oxford Dictionary* definió el fraude como un engaño criminal, el uso de representaciones falsas para obtener una ventaja injusta. Hay dos formas de combatir los fraudes: prevención de fraude y detección de fraude (Bolton and Hand, 2002; Sohony et al., 2018). La prevención de fraude consiste en un conjunto de medidas, reglas, procedimientos y protocolos para evitar que ocurran, cuando por otro lado, la detección de fraude consiste en identificar los fraudes lo más rápido posible cuando la prevención ha fallado y se han cometido fraudes (Bolton and Hand, 2002; Sohony et al., 2018). Existen varios tipos de detección de fraude que se han investigado, como detección de fraude con tarjeta de crédito, detección de fraude en teléfonos móviles,

Revista Cubana de Ciencias Informáticas
Vol. 13, No. 3, Mes Julio-Septiembre, 2019
ISSN: 2227-1899 | RNPS: 2301
Pág. 107-123
http://rcci.uci.cu

detección de fraude en reclamos de seguro, y detección de tráfico de información privilegiada (Chandola et al., 2009; Kou et al., 2004).

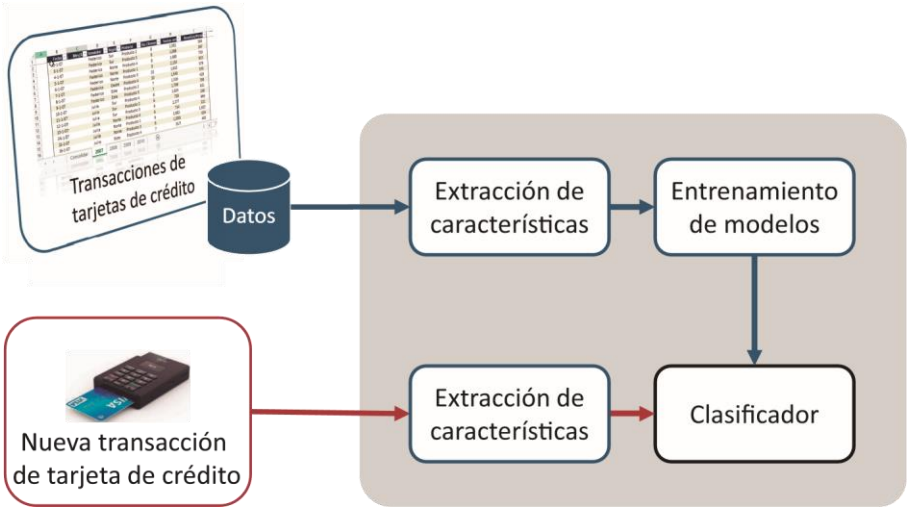

Figura 3. Flujo de un método de detección de anomalía tomando como datos transacciones de tarjetas de créditos.

Los AEs han sido de gran utilidad para la detección de fraude no supervisado, por lo que han sido utilizados en varios trabajos (Seng and Wong, 2017; Pumsirirat and Yan, 2018; Zheng et al., 2018b,a). En (Seng and Wong, 2017) se propuso un método basado en un enfoque de aprendizaje sensible a costos donde se utiliza un tipo de AE conocido como *Stacked Denoising Autoencoders* (SDAE) (Vincent et al., 2008) para identificar transacciones fraudulentas en un problema de detección de fraudes financieros. En este trabajo se realiza una básica selección de instancias en el paso de extracción de características teniendo en cuenta la cantidad de atributos no nulos de las transacciones. Además, introducen una modificación a la función de costo del SDAE con el objetivo de minimizar el costo de la clasificación errónea. De esta forma se identifican transacciones fraudulentas de manera eficaz y eficientemente. Los autores de (Pumsirirat and Yan, 2018) propusieron un método para la detección de fraude en tarjetas de crédito. Este método consiste en clasificar una petición de transferencia bancaria en tiempo real usando un AE, el cual se entrena teniendo en cuenta la información de transacciones realizadas con anterioridad. Los autores de (Zheng et al., 2018b) proponen 3 método para la detección de fraude en transacciones bancarias utilizando AEs. Entre las tres combinaciones la primera es un AE para la extracción de características y un clasificador tradicional y las otras dos AE-AE, AE-SDAE bajo la estrategia GAN, donde la primera red funge como extractora de características y la otra como el clasificador. En (Zheng et al., 2018a) se propone un método que usa un AE en el paso de extracción de características y siguen una estrategia GAN para la detección de fraude. En este trabajo se utiliza un AE para lograr una representación de los usuarios no maliciosos teniendo en cuenta la actividad de estos en línea (en inglés: *online*). Luego generan otra representación ficticia de usuarios no maliciosos mediante una DNN que se usa como la red generadora de la GAN. Finalmente,

mediante otra DNN (conocida como la discriminadora de la GAN) aprende a identificar los usuarios no maliciosos reales. De esta manera al procesar los datos reales, el método es capaz de separar los usuarios no maliciosos del resto.

En (Fu et al., 2016) se propuso un *Framework* para la detección de fraudes de tarjetas de crédito. En este trabajo, los autores introdujeron un nuevo tipo de característica basada en la ganancia de entropía durante un período de tiempo. Usando esta característica de entropía y otras siete características tradicionales, los autores construyeron una matriz de características como resultado del paso de extracción de características. Debido al desbalance existente en los datos, en este trabajo se propone un método de muestreo, que resuelve este problema, basado en el repoblado de las transacciones. Este aumento de datos ponderados tiene como objetivo evitar el sobre entrenamiento de la red hacia solo un tipo de datos. Estas matrices de características se utilizan como entrada de una CNN, la cual tiene como objetivo clasificar las transacciones como anómalas o normales.

Otras de las redes utilizadas para la detección de fraude son las conocidas RBMs (Pumsirirat and Yan, 2018; Luo et al., 2017). El método propuesto en (Pumsirirat and Yan, 2018), utiliza una RBM para la clasificación de fraude en tarjetas de crédito. En este método se valida una petición de transferencia bancaria en tiempo real entrenando la RBM teniendo en cuenta la información de transacciones realizadas con anterioridad. Los autores del trabajo (Luo et al., 2017) realizan un estudio comparativo entre algunos métodos de clasificación tradicionales (regresión logística multinomial, perceptrón multicapa y máquina de soporte vectorial) y un método basado en DBN con un RBM. En este trabajo se mostró la superioridad en eficacia del método que utiliza RBM para la clasificación de fraudes crediticios.

En (Wang and Xu, 2018) se propuso un Framework para la detección de fraudes en seguros de autos mediante una combinación de una técnica de minería de textos basada en LDA (del inglés: *Latent Dirichlet Allocation*) (Blei et al., 2003), información de datos categóricos y datos numéricos, así como una DNN. En este framework se usa una técnica de segmentación de palabras para el procesamiento de los textos y un modelo LDA para la extracción de los tópicos de los textos segmentados. Con estos tópicos, información categórica y numérica se confeccionan las características que se le pasan a la DNN para que aprenda de estas. De esta manera se identifica si una reclamación de accidente de auto es fraudulenta.

Detección de intrusiones

La detección de intrusiones es uno de los retos más tratados en la seguridad de las redes en sistemas informáticos y su objetivo es la identificación de actividad inusual o ataques a la seguridad de redes internas (Kwon et al., 2017). Con este fin se han desarrollado sistemas de detección de intrusiones que proveen tempranas alertas ante intrusiones que permiten prevenir o minimizar el daño (Sultana et al., 2018; Hodo et al., 2017; Lee et al., 2018; Yan et al., 2018). Los daños pueden ser causados por cuatro tipos principales de ataques: DoS (del inglés: *Denial of Service*), Probe, R2L (del inglés: *root-to-local*) y U2R (del inglés: *user-to-root*) (Lee et al., 2018). Un ataque DoS consiste en sobresaturar todos

los servicios y por lo tanto se deniegan todos los pedidos de acceso a la computadora. Un ataque de tipo Probe es cuando el atacante hace un escaneo de la computadora en busca de debilidades o vulnerabilidades que pudiera usar luego para comprometer el sistema. Un ataque R2L es cuando se envían paquetes desde una computadora remota a un usuario local con diferentes privilegios de uso para determinar la vulnerabilidad de la computadora local y colapsar los privilegios de acceso del usuario. Un ataque U2R es cuando el atacante comienza con una cuenta de usuario normal e intenta abusar de la vulnerabilidad del sistema para obtener privilegios de súper usuario.

Existen varios trabajos donde se proponen métodos para la detección de intrusiones basado en AEs (Javaid et al., 2016; Yu et al., 2017b; Farahnakian and Heikkonen, 2018). En (Javaid et al., 2016) se utilizan los AEs para la representación de los datos mediante las capas codificadoras de los mismos y así poder identificar ataques de tipo DoS. En este proceso de representación se propuso una modificación a los AEs profundos tradicionales, que consiste en la eliminación de la capa decodificadora. A estos AEs modificados los llamaron AEs profundos asimétricos (NDAEs, de sus siglas en inglés: *Nonsymmetric Deep AutoEncoders*). La estructura final del proceso de representación está formada por dos NDAEs en cadena utilizando como entrada del segundo NDAE la salida del primero. Luego de obtenida la representación de los datos utilizando la cadena de NDAEs anterior, se utiliza un clasificador tradicional (específicamente un *Random Forest*) para la identificación de las intrusiones. Los AEs también han sido usados para la detección de ataques de tipo U2R (Yu et al., 2017b). En este trabajo se utiliza los ya mencionados SDAEs para lograr una representación de los datos con baja dimensionalidad. Este SDAE se compone por tres AEs estructurados en cadena y entrenados previamente de manera no supervisada y posteriormente mediante un ajuste (en inglés: *fine-tunning*). Luego se utilizó un clasificador Softmax para identificar los ataques. De manera muy similar, en (Mighan and Kahani, 2018) se utiliza un SDAE para la representación reducida de los datos y luego mediante una máquina de soporte vectorial se realiza la clasificación de los ataques en redes de tráfico como *PU-IDS Dataset* (Singh et al., 2015). Los autores de (Farahnakian and Heikkonen, 2018) utilizan un AE profundo entrenado de manera glotona (en inglés: *greedy*) por capas para evitar el sobreajuste y los óptimos locales. De esta manera logran alta eficacia en la clasificación de los diferentes tipos de ataques.

Los trabajos (Roy et al., 2017; Kim et al., 2017b) muestran que también se pueden utilizar las redes del tipo DNN para la detección de intrusiones. En ambos trabajos se realizó un estudio del desempeño de este tipo de redes en el tráfico de una red física entre computadoras para la clasificación de cada uno de los tipos de intrusiones.

Otros trabajos parten del uso de las DBN para la detección de intrusiones (Zhao et al., 2017; Qu et al., 2017). En (Zhao et al., 2017) se reduce la dimensión de los datos mediante una representación obtenida con una DBN, en la cual se utiliza una cantidad óptima de neuronas para las capas ocultas. Esta cantidad de neuronas fue definida mediante el uso de un algoritmo genético basado en enjambre. De esta manera se mejora el rendimiento del proceso de aprendizaje de la red. Luego, haciendo uso de una red neuronal probabilística como clasificador se identifican cada uno de los tipos de ataques a una red. En (Qu et al., 2017) también se hace uso de una red DBN con un clasificador probabilístico al final,

para la etapa del fine-tunning, que realiza la clasificación de los datos. La cantidad de neuronas en las capas ocultas y entrada-salida fueron determinadas en este trabajo de forma empírica.

Los autores de (Vinayakumar et al., 2017; Yu et al., 2017a) se basaron en las CNNs para la detección de intrusiones. En (Vinayakumar et al., 2017) se evaluó la efectividad de la combinación de las CNNs y métodos de modelado secuencial de datos en el análisis y clasificación de todos los tipos de ataques a una red. Específicamente se utilizaron las siguientes combinaciones: CNN-RNN, CNN-LSTM (de sus siglas en inglés: *long-short term memory* (Sak et al., 2014)) y CNN-GRU (de sus siglas en inglés: *gated recurrent unit* (Chung et al., 2014)). De estas combinaciones la que mayor rendimiento reportó una CNN-LSTM, con la CNN de tres capas ocultas. En (Yu et al., 2017a) proponen una variante de las CNN, que nombraron como AEs convolucional dilatado (*DCA* del inglés: *Dilated Convolutional Autoencoders*), donde aprovechan las ventajas de los Autoencoders apilados, del inglés *Stacked Autoencoders* y las CNNs. La idea de este trabajo es realizar la disminución y recomposición de los datos, conocido como una convolución y deconvolución. Esta red sustituye las capas de agrupamiento por una convolucional dilatada. Esta variante no necesita de datos etiquetados para el entrenamiento, por lo que posteriormente se le realizó un fine-tunning utilizando una capa con un Softmax para la clasificación en los diferentes tipos de ataques.

Las RNN también han sido usadas para la detección de intrusiones (Loukas et al., 2018; Yin et al., 2017; Tang et al., 2018). En (Loukas et al., 2018) se propone un método para la detección de intrusiones en vehículos robóticos basado en RNN. Los autores de este trabajo extienden la RNN tradicional mediante la sustitución de la función de activación por un LSTM para evitar la pérdida de información cuando el gradiente se acerca a cero. Mediante esta modificación RNN-LSTM se representan los datos y luego se utiliza un perceptrón multi-capa profundo como clasificador. De esta manera se identifican los ataques de tipo DoS, de inyección y malwares. Otro trabajo basado en RNN es el propuesto por Tang *et al.* (Tang et al., 2018), donde se combina una RNN con un método de modelado secuencial de datos (GRU). Haciendo uso del GRU-RNN son capaces de clasificar con alta eficacia los diferentes ataques de intrusiones realizados sobre una red de softwares. En (Yin et al., 2017) se utilizó una RNN como un clasificador sin modificación alguna a la misma para la detección de los diferentes tipos de intrusiones. Nuevamente en (Kim and Kim, 2015) combinan una red RNN con un modelo secuencial de datos el LSTM usando un optimizador de descendente de gradiente estocástico y luego esta combinación fue mejorada en (Kim et al., 2017a) utilizando la misma combinación, pero con un optimizador Nadam.

Finalmente, en la Tabla 1 se muestra un resumen comparativo de las características de los trabajos relacionados.

Revista Cubana de Ciencias Informáticas
Vol. 13, No. 3, Mes Julio-Septiembre, 2019
ISSN: 2227-1899 | RNPS: 2301
Pág. 107-123
http://rcci.uci.cu

Tabla 1. Características de los trabajos relacionados.

Trabajo	Tipo de Red	Momento en que se usa la red	Aporte al proceso de Minería	Area de Aplicación
(Kim and Kim, 2015)	RNN-LSTM	Clasificación	No	Intrusiones
(Javaid et al., 2016)	AE	Representación	**Si**	Intrusiones
(Fu et al., 2016)	CNN	Clasificación	**Si**	Fraude
(Kim et al., 2017a)	RNN-LSTM	Clasificación	No	Intrusiones
(Kim et al., 2017b)	DNN	Clasificación	No	Intrusiones
(Luo et al., 2017)	RBM y DBN	Clasificación	No	Fraude
(Qu et al., 2017)	DBN	clasificación	No	Intrusiones
(Roy et al., 2017)	DNN	Clasificación	No	Intrusiones
(Seng and Wong, 2017)	AE	Clasificación	No	Fraude
(Vinayakumar et al., 2017)	CNN-(RNN, LSTM, GRU)	Clasificación	No	Intrusiones
(Yin et al., 2017)	RNN	Clasificación	No	Intrusiones
(Yu et al., 2017a)	CNN	Representación	**Si**	Intrusiones
(Yu et al., 2017b)	AE	Representación	No	Intrusiones
(Zhao et al., 2017)	DBN	Representación	No	Intrusiones
(Farahnakian and Heikkonen, 2018)	AE	Representación	No	Intrusiones
(Loukas et al., 2018)	RNN-LSTM	Representación	No	Intrusiones
(Mighan and Kahani, 2018)	AE	Representación	No	Intrusiones
(Pumsirirat and Yan, 2018)	RBM	Clasificación	No	Fraude
(Tang et al., 2018)	RNN-GRU	Clasificación	No	Intrusiones
(Wang and Xu, 2018)	DNN	Clasificación	**Si**	Fraude
(Zheng et al., 2018a)	AE y DNN	Representación y Clasificación	No	Fraude
(Zheng et al., 2018b)	AE y AE	Representación y Clasificación	No	Fraude

Resultados y discusión

El uso del aprendizaje profundo ha ganado importancia debido a su eficacia evaluando la seguridad de las redes. Esto ha permitido la inclusión de este tipo de aprendizaje en la mayoría de las aplicaciones de reconocimiento de patrones y minería de datos, mejorando considerablemente los resultados obtenidos con métodos tradicionales en diferentes tareas como la detección de anomalías. Además, cuando se desarrolla un método para la detección de anomalías existen varios retos a tener en cuenta debido a que la naturaleza de las anomalías es dinámica. Por lo que se necesita adaptabilidad en el método de detección, característica que está presente en las redes neuronales profundas.

En la literatura, como se describieron brevemente en la sección anterior, se han reportado numerosos trabajos donde se han dado evidencias del uso de los diferentes tipos de redes neuronales profundas para la detección de anomalías (Javaid et al., 2016; Yu et al., 2017b; Seng and Wong, 2017; Pumsirirat and Yan, 2018; Zheng et al., 2018b,a; Fu et al., 2016; Luo et al., 2017; Yu et al., 2017a; Farahnakian and Heikkonen, 2018; Mighan and Kahani, 2018; Roy et al., 2017; Kim et al., 2017b; Zhao et al., 2017; Qu et al., 2017; Vinayakumar et al., 2017; Loukas et al., 2018; Wang and Xu, 2018;

Yin et al., 2017; Tang et al., 2018; Kim and Kim, 2015; Kim et al., 2017a). En estos trabajos se pudo observar el uso de las redes neuronales profundas para tareas como la reducción de dimensionalidad y/o la representación de los datos y además como clasificador para la detección de anomalías, ya sea como fraudes en redes bancarias o ataques de intrusiones en redes de tráficos de datos.

La mayoría de los trabajos, analizados en la revisión presentada en este reporte y resumidos en la Tabla 1, hacen uso de los AEs profundos tradicionales y sus variantes. En esta misma tabla se puede notar que los AEs en la detección de fraudes fueron mayormente usados para la clasificación de los datos ya que en la mayoría de los casos la clasificación es binaria; mientras que en la detección de intrusiones fueron utilizados para lograr una representación más compacta y robusta de los datos. Esto último se debe a que en la detección de intrusiones la clasificación es multi-clase y no está demostrado el buen uso de los AEs para este tipo de clasificación.

Por otro lado, se puede mencionar que la estrategia GAN ha sido poco utilizada a pesar de los buenos resultados que se logran con su uso. Esto se debe a la complejidad que conlleva su implementación y entrenamiento. Sin embargo, fue mostrado el uso de varios tipos de redes como en (Zheng et al., 2018b) donde se utiliza la combinación de un AE profundo y un SDAE; y en (Zheng et al., 2018a) se combina un AE con una DNN separando los usuarios no maliciosos de los que sí lo son.

Finalmente, con la excepción de los trabajos (Javaid et al., 2016; Fu et al., 2016; Yu et al., 2017a; Wang and Xu, 2018), no existe un aporte científico desde el punto de vista de la Minería de Datos en los trabajos analizados en este reporte (ver Tabla 1). El aporte de los autores en (Javaid et al., 2016; Yu et al., 2017a) está enmarcado en la propuesta de nuevas alternativas de redes para un tipo de ataque específico en la detección de intrusiones. En (Fu et al., 2016) el aporte está en la proposición de nuevas características basadas en entropía y en (Wang and Xu, 2018) el aporte está enfocado en la selección de tópicos para el procesamiento del lenguaje natural. El resto de los trabajos (Yu et al., 2017b; Seng and Wong, 2017; Pumsirirat and Yan, 2018; Zheng et al., 2018b,a; Luo et al., 2017; Farahnakian and Heikkonen, 2018; Mighan and Kahani, 2018; Roy et al., 2017; Kim et al., 2017b; Zhao et al., 2017; Qu et al., 2017; Vinayakumar et al., 2017; Loukas et al., 2018; Yin et al., 2017; Tang et al., 2018; Kim and Kim, 2015; Kim et al., 2017a) solo proponen nuevas configuraciones obtenidas empíricamente para la optimización de las redes sobre los tipos de datos analizados.

Conclusiones

En este reporte se presenta una revisión de los métodos reportados para la detección de anomalías, específicamente detección de fraudes e intrusiones, basados en aprendizaje profundo y se enfatiza en el aporte científico brindado por estos métodos en el proceso de la detección de anomalías. En esta revisión se categorizaron los métodos reportados según el tipo de DNN usada. Esta categorización permitió identificar que los AEs profundos han sido los más usados, tanto para la representación de los datos como para la clasificación de los diferentes tipos de anomalías. Además, se

encontraron evidencias de la aplicación exitosa de todas las redes neuronales profundas actuales en tareas de detección de anomalías. Sin embargo, según la discusión realizada en este reporte, se puede llegar a la conclusión de que aún queda mucho por hacer e investigar para lograr un aporte significativo en la minería de datos, específicamente en la detección de anomalías, mediante el uso del aprendizaje profundo. Esto se debe a que solo unos pocos métodos aportan mejoras en el proceso de minería, más allá del uso directo de las redes neuronales profundas y sus configuraciones.

Por otro lado, no se encontró en el estado del arte un trabajo como este reporte, que realizara una revisión sobre la detección de anomalías basada en aprendizaje profundo desde el punto de vista de minería de datos. Por este motivo, se considera que mediante este reporte se brinda información valiosa que puede tomarse como guía para nuevas investigaciones sobre el tema.

Referencias

Charu C Aggarwal. *Outlier Analysis*. Springer, 2nd edition, 2017.

Aijaz Ahmed and Mario Garcia. Signature-Based Network Intrusion Detection System Using JESS(SNIDJ). In *EuroIMSA*, pages 281 – 286, 2005.

Emin Aleskerov, Bernd Freisleben, and Bharat Rao. Cardwatch: A Neural Network Based Database Mining System for Credit Card Fraud Detection. In *Computational Intelligence for Financial Engineering (CIFEr), 1997., Proceedings of the IEEE/IAFE 1997*, pages 220 – 226. IEEE, 1997.

S Aravindh, S Venkatesan, and A Kumaravel. Online Credit Card Fraudulent Detection Using Data Mining. *Asian Journal of Computer Science and Technology (AJCST) (UGC Approved Journal)*, 1(2):9 – 15, 2012.

Sherenaz W Al-Haj Baddar, Alessio Merlo, and Mauro Migliardi. Anomaly Detection in Computer Networks: A State-of-the-Art Review. *JoWUA*, 5(4):29 – 64, 2014.

Vic Barnett and Toby Lewis. *Outliers in Statistical Data*. John Wiley & Sons, 1978.

Yoshua Bengio. Learning Deep Architectures for AI. *Foundations and trends* R *in Machine Learning*, 2(1):1 – 127, 2009.

Yoshua Bengio. Deep Learning of Representations: Looking Forward. In *International Conference on Statistical Language and Speech Processing*, pages 1 – 37. Springer, 2013.

Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation Learning: A Review and New Perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798 – 1828, 2013.

D.M. Blei, A.Y. Ng, and M.I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3, 2003.

Richard J Bolton and David J Hand. Statistical Fraud Detection: A Review. *Statistical Science*, pages 235 – 249, 2002.

Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly Detection: A Survey. *ACM Computing Surveys (CSUR)*, 41(3):15, 2009.

Francois Chollet. *Deep Learning with Python*. Manning Publications Co., 2017.

Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*, 2014.

Wilfred J Dixon. Analysis of Extreme Values. *The Annals of Mathematical Statistics*, 21(4):488 – 506, 1950.

Janet Dixon Elashoff. A Model for Quadratic Outliers in Linear Regression. *Journal of the American Statistical Association*, 67(338):478 – 485, 1972.

F. Farahnakian and J. Heikkonen. A deep auto-encoder based approach for intrusion detection system. In *2018 20th International Conference on Advanced Communication Technology (ICACT)*, pages 178–183, Feb 2018.

Kang Fu, Dawei Cheng, Yi Tu, and Liqing Zhang. Credit Card Fraud Detection Using Convolutional Neural Networks. In *International Conference on Neural Information Processing*, pages 483 – 490. Springer, 2016.

Ryohei Fujimaki, Takehisa Yairi, and Kazuo Machida. An Approach to Spacecraft Anomaly Detection Problem Using Kernel Feature Space. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, pages 401 – 410. ACM, 2005.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*, volume 1. MIT press Cambridge, 2016.

Frank E Grubbs. Procedures for Detecting Outlying Observations in Samples. *Technometrics*, 11(1):1 – 21, 1969.

Douglas M Hawkins. *Identification of Outliers*, volume 11. Springer, 1980.

Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. A Fast Learning Algorithm for Deep Belief Nets. *Neural Computation*, 18(7):1527 – 1554, 2006.

E. Hodo, X. J. A. Bellekens, A. Hamilton, C. Tachtatzis, and R. C. Atkinson. Shallow and deep networks intrusion detection system: A taxonomy and survey. *ACM Survey*, 2017.

David H Hubel and Torsten N Wiesel. Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex. *The Journal of Physiology*, 160(1):106 – 154, 1962.

Max Jaderberg, Andrea Vedaldi, and Andrew Zisserman. Deep features for text spotting. In *European conference on computer vision*, pages 512–528. Springer, 2014.

Ahmad Javaid, Quamar Niyaz, Weiqing Sun, and Mansoor Alam. A Deep Learning Approach for Network Intrusion Detection System. In *Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS)*, pages 21 –16. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2016.

Irina Kakanakova and Stefan Stoyanov. Outlier Detection via Deep Learning Architecture. In *Proceedings of the 18th International Conference on Computer Systems and Technologies*, pages 73 – 79. ACM, 2017.

Gopalan Kesavaraj and Sreekumar Sukumaran. A Study on Classification Techniques in Data Mining. In *Computing, Communications and Networking Technologies (ICCCNT), 2013 Fourth International Conference on*, pages 1 – 7. IEEE, 2013.

Jihyun Kim and Howon Kim. Applying recurrent neural network to intrusion detection with hessian free optimization. In *International Workshop on Information Security Applications*, pages 357–369. Springer, 2015.

Jihyun Kim, Howon Kim, et al. An effective intrusion detection classifier using long short-term memory with gradient descent optimization. In *Platform Technology and Service (PlatCon), 2017 International Conference on*, pages 1–6. IEEE, 2017a.

Jin Kim, Nara Shin, Seung Yeon Jo, and Sang Hyun Kim. Method of intrusion detection using deep neural network. In *Big Data and Smart Computing (BigComp), 2017 IEEE International Conference on*, pages 313–316. IEEE, 2017b.

Yufeng Kou, Chang-Tien Lu, Sirirat Sirwongwattana, and Yo-Ping Huang. Survey of Fraud Detection Techniques. In *Networking, Sensing and Control, 2004 IEEE International Conference on*, volume 2, pages 749 – 754. IEEE, 2004.

Vipin Kumar. Parallel and Distributed Computing for Cybersecurity. *IEEE Distributed Systems Online*, 6 (10), 2005.

Donghwoon Kwon, Hyunjoo Kim, Jinoh Kim, Sang C Suh, Ikkyun Kim, and Kuinam J Kim. A Survey of Deep Learning-Based Network Anomaly Detection. *Cluster Computing*, pages 1 – 13, 2017.

Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition.

Proceedings of the IEEE, 86(11):2278–2324, Nov 1998. ISSN 0018-9219. doi: 10.1109/5.726791.

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep Learning. *Nature*, 521(7553):436, 2015.

Brian Lee, Sandhya Amaresh, Clifford Green, and Daniel Engels. Comparative study of deep learning models for network intrusion detection. *SMU Data Science Review*, 1(1):8, 2018.

George Loukas, Tuan Vuong, Ryan Heartfield, Georgia Sakellari, Yongpil Yoon, and Diane Gan. Cloud-based cyber-physical intrusion detection for vehicles using deep learning. *IEEE Access*, 6:3491–3508, 2018.

Cuicui Luo, Desheng Wu, and Dexiang Wu. A deep learning approach for credit scoring using credit default swaps. *Engineering Applications of Artificial Intelligence*, 65:465–470, 2017.

S. N. Mighan and M. Kahani. Deep learning based latent feature extraction for intrusion detection. In *Electrical Engineering (ICEE), Iranian Conference on*, pages 1511–1516, May 2018.

Maryam M Najafabadi, Flavio Villanustre, Taghi M Khoshgoftaar, Naeem Seliya, Randall Wald, and Edin Muharemagc. Deep Learning Techniques in Big Data Analytics. In *Big Data Technologies and Applications*, pages 133 – 156. Springer, 2016.

Jagruti D. Parmar and Jalpa T. Patel. Anomaly Detection in Data Mining: A Review. *International Journal*, 7(4):32 – 40, 2017.

Josh Patterson and Adam Gibson. *Deep Learning: A Practitioner's Approach*. O'Reilly Media, Inc., 2017.

Vir V Phoha. *Internet Security Dictionary*, volume 1. Springer, 2002.

Apapan Pumsirirat and Liu Yan. Credit Card Fraud Detection Using Deep Learning based on Auto-Encoder and Restricted Boltzmann Machine. *International Journal of Advanced Computer Science and Applications*, 9(1):18 – 25, 2018.

Feng Qu, Jitao Zhang, Zetian Shao, and Shuzhuang Qi. An intrusion detection model based on deep belief network. In *Proceedings of the 2017 VI International Conference on Network, Communication and Computing*, pages 97–101. ACM, 2017.

Pankaj Richhariya and Prashant K Singh. A Survey on Financial Fraud Detection Methodologies. *International Journal of Computer Applications*, 45(22):15 – 22, 2012.

Sanjiban Sekhar Roy, Abhinav Mallik, Rishab Gulati, Mohammad S Obaidat, and PV Krishna. A deep learning based artificial neural network approach for intrusion detection. In *International Conference on Mathematics and Computing*, pages 44–53. Springer, 2017.

David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning Internal Representations by Error Propagation. Technical report, California Univ San Diego La Jolla Inst for Cognitive Science, 1985.

Ha¸sim Sak, Andrew Senior, and Fran¸coise Beaufays. Long short-term memory recurrent neural network architectures for large scale acoustic modeling. In *Fifteenth annual conference of the international speech communication association*, 2014.

Kruy Seng and Man-Leung Wong. Cost-Sensitive Deep Neural Networks for Financial Fraud Detection. In *8th Annual International Conference on ICT: Big Data, Cloud and Security*, 2017.

Raman Singh, Harish Kumar, and RK Singla. A reference dataset for network traffic activity based intrusion detection system. *International Journal of Computers Communications & Control*, 10(3):390–402, 2015.

Paul Smolensky. Information Processing in Dynamical Systems: Foundations of Harmony Theory. Technical report, Colorado Univ at Boulder Dept of Computer Science, 1986.

Ishan Sohony, Rameshwar Pratap, and Ullas Nambiar. Ensemble Learning for Credit Card Fraud Detection. In *Proceedings of the ACM India Joint International Conference on Data Science and Management of Data*, pages 289 – 294. ACM, 2018.

Clay Spence, Lucas Parra, and Paul Sajda. Detection, Synthesis and Compression in Mammographic Image Analysis with a Hierarchical Image Probability Model. In *Mathematical Methods in Biomedical Image Analysis, 2001. MMBIA 2001. IEEE Workshop on*, pages 3 – 10. IEEE, 2001.

Nasrin Sultana, Naveen Chilamkurti, Wei Peng, and Rabei Alhadad. Survey on sdn based network intrusion detection system using machine learning approaches. *Peer-to-Peer Networking and Applications*, pages 1–9, 2018.

Yi Sun, Xiaogang Wang, and Xiaoou Tang. Deep learning face representation from predicting 10,000 classes. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.

Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. *Introduction to Data Mining*. Pearson Addison Wesley, 2005.

T Tang, SAR Zaidi, D McLernon, L Mhamdi, and M Ghogho. Deep recurrent neural network for intrusion detection in sdn-based networks, March 2018.

R Vinayakumar, KP Soman, and Prabaharan Poornachandran. Applying convolutional neural network for network intrusion detection. In *Advances in Computing, Communications and Informatics (ICACCI), 2017 International Conference on*, pages 1222–1228. IEEE, 2017.

Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th international conference on Machine learning*, pages 1096–1103. ACM, 2008.

Haohan Wang and Bhiksha Raj. On the Origin of Deep Learning. *arXiv preprint arXiv:1702.07800*, 2017.

Yibo Wang and Wei Xu. Leveraging Deep Learning with LDA-Based Text Analytics to Detect Automobile Insurance Fraud. *Decision Support Systems*, 105:87 – 95, 2018.

Zhangyang Wang, Jianchao Yang, Hailin Jin, Eli Shechtman, Aseem Agarwala, Jonathan Brandt, and Thomas S Huang. Real-world font recognition using deep network and domain adaptation. *arXiv preprint arXiv:1504.00028*, 2015.

J. Yan, D. Jin, C. W. Lee, and P. Liu. A comparative study of off-line deep learning based network intrusion detection. In *2018 Tenth International Conference on Ubiquitous and Future Networks (ICUFN)*, pages 299–304, July 2018.

Chuanlong Yin, Yuefei Zhu, Jinlong Fei, and Xinzheng He. A deep learning approach for intrusion detection using recurrent neural networks. *IEEE Access*, 5:21954–21961, 2017.

Yang Yu, Jun Long, and Zhiping Cai. Network intrusion detection through stacking dilated convolutional autoencoders. *Security and Communication Networks*, 2017, 2017a.

Yang Yu, Jun Long, and Zhiping Cai. Session-based network intrusion detection using a deep learning architecture. In *Modeling Decisions for Artificial Intelligence*, pages 144–155. Springer, 2017b.

Guangzhen Zhao, Cuixiao Zhang, and Lijuan Zheng. Intrusion detection using deep belief network and probabilistic neural network. In *Computational Science and Engineering (CSE) and Embedded and Ubiquitous Computing (EUC), 2017 IEEE International Conference on*, volume 1, pages 639–642. IEEE, 2017.

Panpan Zheng, Shuhan Yuan, Xintao Wu, Jun Li, and Aidong Lu. One-Class Adversarial Nets for Fraud Detection. *arXiv preprint arXiv:1803.01798*, 2018a.

Yu-Jun Zheng, Xiao-Han Zhou, Wei-Guo Sheng, Yu Xue, and Sheng-Yong Chen. Generative adversarial network based telecom fraud detection at the receiving bank. *Neural Networks*, 102:78–86, 2018b.