

Clasificación de modos de falla en rodamientos utilizando Variational Auto-Encoders

Camila Fernández [DIM] & Gabriel San Martín [DIMEC]

Resumen—En Ingeniería Mecánica, el problema de identificar el modo de falla de un componente necesita tanto de experiencia humana como de la detección de la máquina en cuestión para poder realizar la clasificación. El aprendizaje de máquinas aplicado a este problema resulta de gran ayuda debido a que quita la necesidad del factor humano, además de permitir técnicas mucho menos invasivas de diagnóstico, como la medición de señales vibratorias en los componentes. Los variational auto-encoders (VAE's) son un modelo generativo y de extracción de características relativamente nuevo dentro del campo, y con el cual es posible reducir la dimensionalidad de los datos para así mejorar la calidad y cantidad de la información contenida. En el presente paper se realizaron experimentos sobre el dataset sintético de MNIST, para ejemplificar las capacidades generativas de los VAE's, para luego utilizar datos de vibración en rodamientos para poder analizar las ventajas de usar este método como reductor de dimensionalidad y extractor de características en un problema de clasificación de modos de falla.

I. PRESENTACIÓN DEL PROBLEMA

Durante los últimos años, con el desarrollo de nuevas tecnologías que permiten tomar mediciones de forma más automatizada y guardar en la memoria enormes cantidades de información, la cantidad de datos a manipular ha crecido exponencialmente, lo que se puede observar en particular en los sistemas mecánicos.

En el área de gestión de activos físicos estos datos son lecturas de sensores que representan características de la operación de los equipos. La gestión de activos físicos está estrechamente relacionada con el mantenimiento y la fiabilidad, dos conceptos muy importantes dentro del mundo empresarial, pues es fundamental mantener en buen estado de funcionamiento los activos físicos de una empresa de modo de asegurar que se cumplan los servicios que esta ofrece.

Dentro de este contexto y con el fin de asegurar el funcionamiento de las máquinas es que es importante identificar los modos de falla de un componente, luego, se plantea que con un correcto análisis y tratamiento de los datos obtenidos por los sensores es posible obtener información relevante sobre el estado de un equipo que permita clasificar y predecir modos de falla.

Para realizar esto se plantea generar nuevos modelos basados en aprendizaje de máquinas que permitan un pronóstico certero del daño y vida remanente de elementos mecánicos, que en este caso se realizará a través de la medición de señales vibratorias de las componentes.

El método que se propone usar es el de Variational Autoencoders que tiene la característica que dado un conjunto de datos de muchas dimensiones estos pueden ser codificados en variables latentes que capturan algunas de estas dimensiones (las que representan características relevantes para el reconocimiento de los datos) de modo que al trabajar en este nuevo espacio se aliviana la carga de trabajo del computador y además es posible entrenar redes neuronales y otros métodos de aprendizaje de máquinas.

II. DISCUSIÓN DE MÉTODOS EXISTENTES

Para realizar clasificación general sobre algún dataset etiquetado, las redes neuronales se alzan como un método ya estudiado y que generalmente entrega muy buenos resultados para distintos tipos de datos. De esa manera, se propone como mecanismo de clasificación los siguientes dos modelos existentes de resolución, para posteriormente discutir sus desventajas.

- Red Neuronal sobre los datos originales.
- Red Neuronal sobre los datos reducidos en dimensionalidad mediante PCA.

El primer modelo es lo que cualquiera intentaría como primer intento en un problema de clasificación. Es un método sencillo, relativamente rápido de implementar y que generalmente no entrega malos resultados. Sin embargo, cuenta con dos principales problemas: primero, puede ser un método en extremo costoso en tiempo de cálculo y en recursos computacionales dependiendo del dataset. Si este es de un tamaño excesivo, la cantidad de parámetros de la red neuronal rápidamente crece y esta podría volverse demasiado compleja para poder ser resuelta dentro de un tiempo razonable, con hardware asequible. Segundo, los dataset podrían ser muy ruidosos, tanto en estructura como en los valores de sus datos. Por ejemplo, el dataset podría estar compuesto de mediciones de muchas magnitudes físicas, de las cuales una gran mayoría podría ser irrelevante para el problema en cuestión. Por otra parte, los sensores utilizados en la toma de muestras añaden ruido a las mediciones, y si este no es controlado, las desviaciones podrían alterar mucho los datos volviendo al problema en uno mucho más confuso para la red neuronal.

De esa manera, si bien las redes neuronales son un método sencillo de realizar clasificación, al usarla sobre los datos originales podrían entregar malos resultados dependiendo de la cantidad y calidad del dataset.

El segundo modelo si bien podría ser una mejora con respecto al primero para los casos de un dataset de mucho

tamaño o uno con mucho ruido e información de sobra, el método que utiliza para lidiar con estos problemas es simple. El *principal component analysis*, o PCA, podría ignorar componentes importantes de los datos.

III. METODOLOGÍA PROPUESTA

El modelo que se propone es el de variational auto-encoders (VAE), un modelo generativo que usa inferencia variacional para producir datos que sean similares a los del conjunto de datos, este modelo busca representar la distribución de probabilidad asociada a los datos de modo de samplear nuevos datos desde esta distribución. Además, con este modelo podemos extraer características que sean clave y suficientes para la clasificación o representación de datos.

Los tradicionales autoencoders son modelos diseñados para entregar una reconstrucción del elemento de entrada. Específicamente, decodifica el input en variables ocultas o latentes y luego usa esta representación para reconstruir outputs que se asimilen al original. Estos procesos de decodificación y reconstrucción se llevan a cabo mediante redes neuronales.

Para efectos de notación, llamaremos X al conjunto de datos y z la variable latente. Queremos encontrar $P(X)$ la distribución de probabilidad de los datos usando $P(z)$ pues se espera que esta sea más fácil de estimar debido a la reducción de dimensionalidad. La relación que hay entre ambas está dada por la ley de probabilidades totales:

$$P(X) = \int P(X|z)P(z)dz$$

La idea de Variational Autoencoder es inferir $P(z)$ usando $P(z|X)$, esto pues solo nos interesa la parte de z que tiene que ver con los datos. Para inferir $P(z|X)$ se utiliza un método llamado inferencia variacional cuya idea principal es plantear la inferencia como un problema de optimización, modelando la distribución $P(z|X)$ a través de distribuciones más simples y fáciles de evaluar, como por ejemplo Gaussianas, y minimizar la diferencia entre estas dos. Esta diferencia estará determinada por la divergencia KL que se define a continuación:

$$D_{KL}(P(z|X)||Q(z|X)) = E(\log(Q(z|X)) - \log(P(z|X)))$$

Al usar el teorema de Bayes y manipular algebraicamente esta expresión llegamos a la función objetivo de la VAE que será:

$$\log(P(X)) - D_{KL}(Q(z|X)||P(z|X)) = E(\log(P(X|z))) - D_{KL}(Q(z|X)||P(z))$$

Donde $D_{KL}(Q(z|X)||P(z|X))$ representa el error que se quiere minimizar. Como esta expresión es más difícil de calcular haremos la optimización sobre el lado derecho de la ecuación, donde maximizar $E(\log(P(X|z)))$ es equivalente a

estimación de máxima verosimilitud.

Para $D_{KL}(Q(z|X)||P(z))$ consideraremos en primer lugar $P(z)$ como una normal de medio cero y varianza 1 y $Q(z|X)$ una normal de media μ y varianza Σ . Queremos hacer $Q(z|X)$ lo más cercano posible a $P(z)$ y con esta elección $D_{KL}(Q(z|X)||P(z))$ es muy fácil de computar.

Al implementar este proceso, $Q(z|X)$, que toma los datos X y devuelve μ y Σ los parámetros de la Gaussiana, representa la red de decodificación que será una red neuronal de una capa y $P(X|z)$ representa la red de codificación.

Finalmente nos enfrentamos al problema de como obtenemos la variable latente z desde los outputs de $Q(z|X)$. El problema se encuentra en que si queremos entrenar la VAE con gradiente descendiente el proceso de sampleo no es diferenciable, para arreglar esto utilizamos un truco de reparametrización que hace la red diferenciable, este truco básicamente consiste en desviar las operaciones no diferenciables fuera de la red para que esta pueda ser entrenada.

Con este modelo se pueden hacer principalmente dos cosas, generar datos desde las variables latentes, que observaremos a través de una implementación sobre la base de datos de MNIST, y codificar los inputs en variables latentes que representa una extracción de características sobre la que desea verificar si esta permite obtener mejor precisión para clasificación y regresión, lo que observaremos en la implementación de identificación de modos de falla de rodamientos.

IV. EXPERIMENTO

Se cuenta con 11753 datos vibratorios tomados con acelerómetros en rodamientos durante su operación. Los rodamientos pueden presentar cuatro estados: falla en el anillo exterior, falla en el anillo interior, falla en las bolas y no presentar falla alguna. De esa manera, el problema de clasificación tendrá cuatro clases. Los datos han sido previamente procesados para transformar la señal vibratoria en el dominio del tiempo a una imagen de 96 por 96 píxeles que es una representación gráfica del espectrograma de la señal. De esa manera, el dataset son 11753 imágenes con etiqueta indicando el modo de falla que representa cada una. Como ejemplo, en la figura 1 se muestra un *datapoint* al azar.

Se realizará clasificación sobre el dataset utilizando tres modelos diferentes:

- Red neuronal sobre los datos originales.
- Red neuronal sobre los datos con dimensión reducida mediante PCA.
- Red neuronal sobre los datos con dimensión reducida mediante Variational Auto-Encoder.

para dos modalidades diferentes:

- Sobre el dataset completo, dónde tanto el Variational Auto-Encoder como la red neuronal son entrenados con todos los datos.

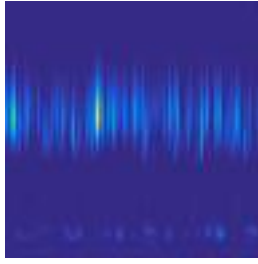


Figura 1. Espectrograma de una señal vibratoria perteneciente al dataset.

- Sobre el dataset reducido a 2000 datos tomados al azar de los 11753. En este caso, se trata de emular la situación en la que existen muchos datos, pero solo una pequeña parte de ellos tiene etiquetas. Esta situación es en realidad la que con mayor frecuencia se presenta dentro de la industria: tomar datos es muy sencillo, pero etiquetarlos requiere que las máquinas detengan su operación y los componentes por separado sean revisados por expertos. De esa manera, el Variational Auto-Encoder se entrena con los 11753 datos ya que para su entrenamiento no son necesarias etiquetas, sin embargo, la red neuronal de clasificación posteriormente es entrenada con sólo 1000 datos y validada con otros 1000 datos. En total, 2000 datos etiquetados.

En la tabla I se muestran las arquitecturas de las diferentes redes neuronales utilizadas en los modelos. Primero, la red neuronal de clasificación es la misma para cada modelo y cuenta con tres capas ocultas, de 500, 400 y finalmente 300 neuronas. Tanto el encoder como el decoder del Variational Auto-Encoder utilizado son redes neuronales simples de dos capas ocultas. Finalmente, la dimensionalidad latente del VAE es 4.

Cuadro I. DATOS SOBRE ARQUITECTURAS DE LOS MODELOS.

Red Neuronal de Clasificación	
Capas ocultas	3
Neuronas por capa	500-400-300
Tasa de Aprendizaje	0.001
Encoder del VAE	
Capas ocultas	2
Neuronas por capa	500-250
Decoder del VAE	
Capas ocultas	2
Neuronas por capa	250-500
Variational Auto-Encoder	
Tasa de Aprendizaje	0.001
Dimensión Latente	4

V. RESULTADOS

V-A. Generación de muestras con VAE entrenada con MNIST

Como primer resultado, se mostrarán las capacidades de generación de imágenes mediante el Variational Auto-Encoder. Para esto, se utiliza el dataset de MNIST, y se busca producir números nuevos, que no existan dentro del dataset y que

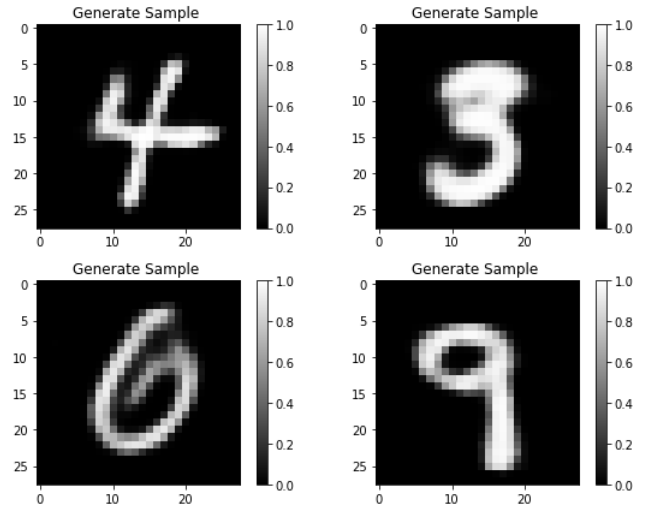


Figura 2. Imágenes generadas mediante el VAE.

se vean lo más reales posibles. Las imágenes generadas se muestran en la figura 2.

Claramente, son números que representan dígitos y si uno observa los demás datos dentro de MNIST, serían prácticamente indistinguibles.

V-B. Clasificación de Modos de Falla utilizando datos de CWR

En la tabla II se muestra la máxima exactitud de clasificación alcanzada en el conjunto de validación para una red neuronal entrenada sobre los datos originales y la misma red neuronal entrenada sobre los datos con dimensión reducida mediante el Variational Auto-Encoder y PCA, tanto con el dataset completo como el reducido. Como se aprecia en la tabla, la exactitud alcanzada con el modelo del Variational Auto-Encoder es comparable (y mejor en algunos casos) a los otros dos modelos. La mayor diferencia se evidencia en el caso extremo de tener un 1% de los datos del dataset de entrenamiento etiquetados. En dicho caso, la exactitud del Variational Auto-Encoder es claramente mejor que la exactitud alcanzada tanto por el modelo que utiliza PCA como por el modelo que no reduce dimensionalidad.

Este resultado es importante porque indica que una reducción de dimensionalidad utilizando Variational Auto-Encoder permite obtener muy buenos resultados en el caso de tener pocos datos etiquetados para el entrenamiento del método de clasificación (en este caso una red neuronal), pero muchos datos no etiquetados para el entrenamiento del VAE.

La matriz de confusión para el caso extremo de un 1% de datos de entrenamiento etiquetados y el modelo que utiliza Variational Auto-Encoder se muestra en la figura V-B

Cuadro II. EXACTITUD DE CLASIFICACIÓN FINAL EN EL EXPERIMENTO PARA LOS TRES MODELOS.

Training Dataset	PCA-MLP 3L	VAE-MLP 3L	MLP 3L
100 %	95.60 %	99.04 %	97.04 %
25 %	94.18 %	99.01 %	99.5 %
5 %	92.72 %	97.41 %	96.8 %
1 %	87.2 %	90.68 %	87.7 %

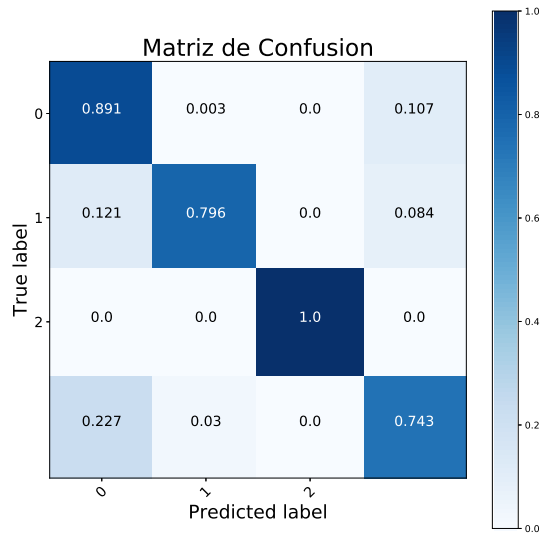


Figura 3. Matriz de Confusión para un 1 % de datos etiquetados y modelo VAE-MLP 3L.

VI. CONCLUSIONES

Este informe presenta un modelo de clasificación de modos de falla en rodamientos mecánicos realizando primero una extracción de características mediante un Variational Auto-Encoder, para luego aplicar una red neuronal de 3 capas ocultas. Los principales resultados del estudio son que en el caso en que se cuenta con muchos datos etiquetados, el modelo que realiza reducción de dimensionalidad mediante VAE se equipara a una red neuronal entrenada sobre los datos originales. sin embargo, para el caso en que se cuenta con sólo una pequeña fracción de los datos etiquetados, y la gran mayoría sin etiqueta alguna, el método desarrollado presenta grandes ventajas en el porcentaje final de exactitud con relación a una red neuronal pura y una red neuronal entrenada en base a datos con dimensión reducida con PCA.

REFERENCIAS

- [1] Tutorial on Variational Autoencoders, Carl Doersch, Carnegie Mellon, UC Berkeley, August 2016
- [2] Auto-Encoding Variational Bayes, Diederik P. Kingma, Max Willing, Universiteit Van Amsterdam, 2014