

## Análisis de la utilización de Redes Neuronales para detección de curvas de difracción, tamaño de objetos y distancia para TAOS II.

Lo primero que es importante aclarar es que las Redes Neuronales (RN) no son “mágicas”, aunque claro está que la Inteligencia Artificial (IA) es una herramienta muy potente en cuanto a tareas que puede realizar, no es una solución que mágicamente va a resolver todos los problemas de forma fácil y como por arte de magia, requiere de un análisis para en primer instancia identificar si el problema amerita realmente que se utilice esta herramienta y una vez que se haya fundamentado la utilización de la misma, es importante analizar el problema para optar por una buena vía de solución.

En segunda instancia, es importante entender que lo que una red hace no es aprenderse un conjunto valores para así realizar una especie de comparación e identificar si los valores a los que se someten se asemejan. Por el contrario, una red sobre entrenada no representa una solución para un problema pues en caso de que se le someta a un conjunto de pruebas ligeramente distinto puede que la red no reaccione como se esperaría. La RN lo que hace es mapear un conjunto de valores de orden  $N$  a otro conjunto de valores de orden  $M$ . Viéndolo de este modo es como se logra entender el porqué una red puede resolver funciones analíticas y también se utiliza como solución para problemas en los que no existe un modelo o simplemente obtenerlo sería muy complicado.

Es de suma importancia el análisis previo a la elección de una RN, pues existen muchos tipos de redes y no todas responden a la misma realidad. Dicho análisis ayuda a entender el problema y la naturaleza a la que pertenece, sabiendo esto es de suma importancia respetar que la red que se vaya implementar también responda a la misma naturaleza, pues así será el comportamiento que va a tener al intentar mapear la solución. Si se está atendiendo un problema relacionado con el cálculo de alguna función booleana sencilla, no es necesario utilizar muchas neuronas en la capa oculta, y con menos razón varias capas ocultas y tampoco utilizar funciones de activación lineales o sigmoidales; si se utiliza una RN para calcular una función analítica ( $x^2$ ,  $\log x$ , etc), es conveniente utilizar capas ocultas y funciones de activación donde su comportamiento sea similar al del resultado esperado, esto depende si el resultado esperado se encuentra en todo  $\mathbb{R}$  (convendría utilizar lineales) o si se conoce el rango en que se encuentran los valores (con una sigmoide o rampa se puede hacer un escalamiento de la solución).

Si el problema a solucionar está relacionado a imágenes, conviene utilizar un tipo de redes que se llaman convolucionales, esta deriva de las que los humanos poseen en su cerebro y se encargan de la visión. Esta clase de redes lo que la diferencia es que en lugar de utilizar unidades de suma (suma de productos) o unidades de productos (productos de potencias), realizan operaciones a la imágenes como suavizados, erosión, dilatación, entre

otras que se le pueden realizar a las imágenes, y posteriormente se le somete a alguna función de activación. Esto las hace muy útiles para el tratamiento de imágenes que por métodos convencionales de funciones analíticas se dificultaría.

Por otro lado, si se tiene un problema que responde a una señal o valores que tiene en su naturaleza el tiempo o la variable temporal, se podría decir que obligatoriamente la red debe contemplar de alguna manera el tiempo. Justamente éste es el caso de las redes que se sugieren utilizar en el proyecto TAOS II. Existen muchas redes temporales que dependiendo del problema y lo que se espere obtener en el resultado se ajustan mejor. Existen las redes de recurrencia simple, en donde la salida de alguna de las capas se retorna a otra capa anterior.

Una de estas redes es la de Elman que se observa en la Figura 1, donde la salida de la capa oculta se convierte en una entrada más de la red en la capa de entrada, por lo que el cálculo anterior tiene relevancia en el resultado actual. De este modo se respeta el hecho de si en el fenómeno natural el tiempo es una variable que afecta al mismo y de algún modo influye en los posibles resultados que se pueda tener, en la red también se considera el tiempo y más que el tiempo como unidad de medida es que de algún modo los resultados se influyen por las soluciones que se dieron con anterioridad.

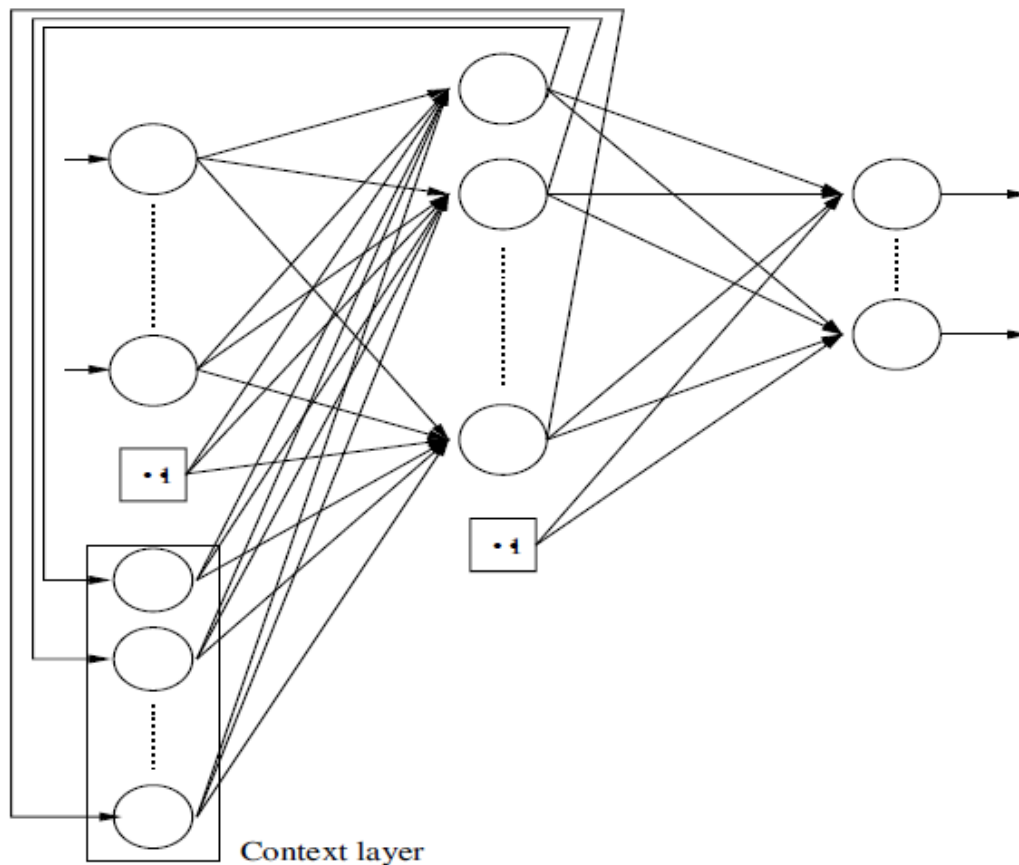


Figura 1: Red Neuronal de Simple Recurrencia de Elman

Hay otras redes que se le conocen como RN de Jordan, son similares a las de Elman pero con la diferencia que lo que se reingresa a la capa de entrada es el resultado de la capa de salida, por lo que en este caso también se le considera un tipo de red que en su naturaleza contempla el tiempo. La red de Elman o de Jordan se podría utilizar en el proyecto TAOS, ya que las señales de los patrones de difracción son dependientes del tiempo, sin embargo tal cual a como se observan en la Figura 1 o la Figura 2 se considera que no serían tan útiles, sino con un cambio.

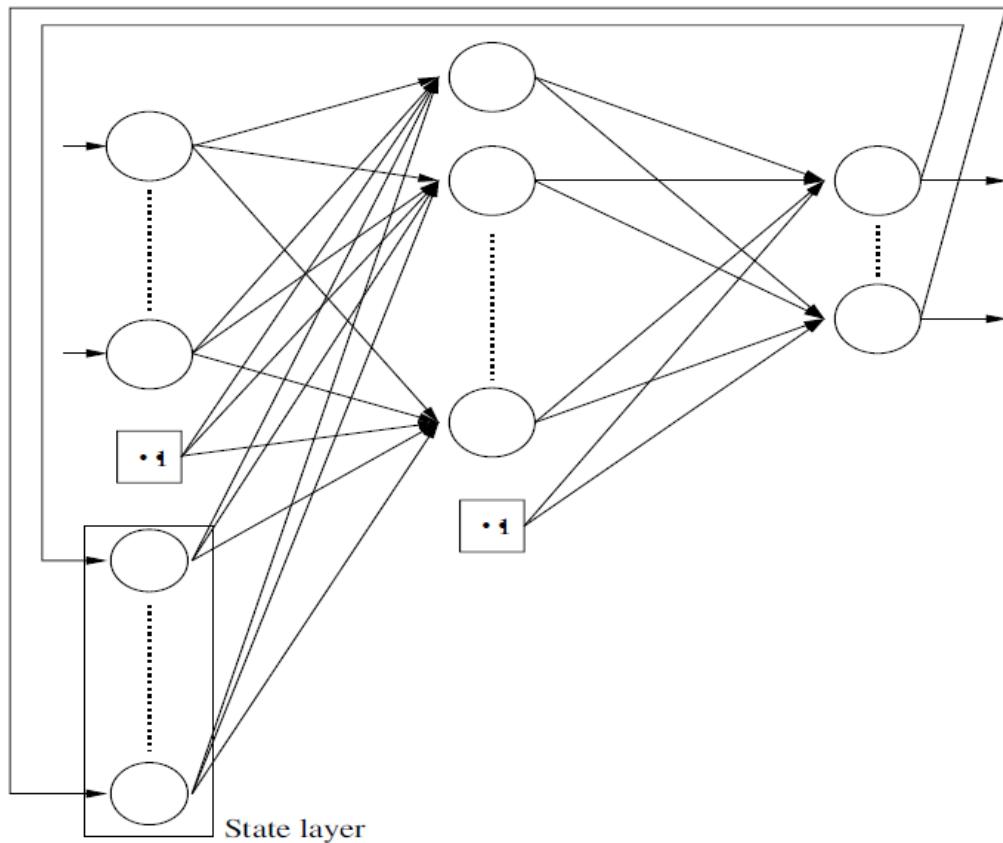


Figura 2: Red Neuronal de Simple Recurrencia de Jordan

El cambio que se menciona con anterioridad tiene que ver otro tipo de red que alberga la naturaleza temporal, conocida como red de tiempo de espera (Time-Delay Neural Network). En esta RN lo que ocurre es que la capa de entrada se comporta como un "buffer" de tipo FIFO, en la que una cada vez que ingrese un dato nuevo a la red este se coloca en la primer neurona de la capa de entrada, si en dicha neurona ya existía un dato entonces se traslada a la segunda neurona, así sucesivamente hasta que el dato que se encontraba en la última neurona se elimina para el cálculo que se realizará. En la Figura 3 se observa una neurona del tipo que se comenta; pero es solo una neurona y no una RN, sin embargo lo que variaría es que en lugar de todas las entradas ir a una sola neurona se repartirían hacia toda la red.

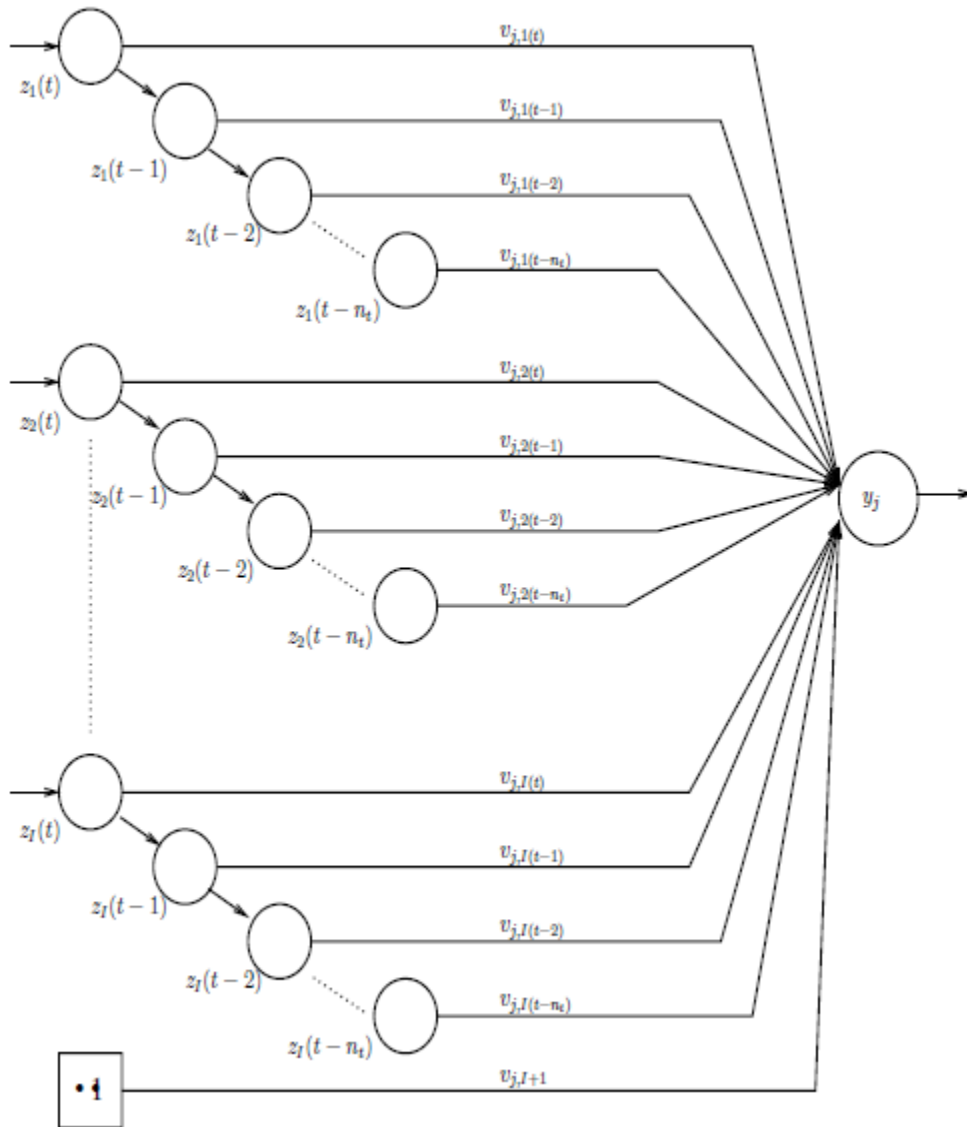


Figura 3: Neurona de tiempo de espera

Cabe destacar que este tipo de red contempla la naturaleza temporal porque en el cálculo actual utiliza valores de entrada pasados. Se puede utilizar este tipo de RN para analizar un rango de una señal, pues como a la red ingresan varios datos al mismo tiempo para ser procesados la información que se obtienen de ellos puede cubrir un espectro importante donde se representen características de interés.

Para el proyecto TAOS, dependiendo de lo que se desea obtener así debe ser la red a utilizar. Si se desea detectar si existe una ocultación o no, se debe tratar la información como una señal, pues la misma varía en el tiempo ya que por la naturaleza del fenómeno físico (patrón de difracción) los datos que se captan con el telescopio también varían en el tiempo, por ende la luminosidad de cada estrella lo hará. Es por esto que se recomienda utilizar (para detectar la ocultación) una RN como la de Elman (Figura 1) o como la de tiempo

de espera (Figura 3); o inclusive una combinación entre ambas. Pues una vez detectada la ocultación se vuelve más simple tratar la señal para determinar las características de la misma.

Se puede utilizar una red que solo posea 1 entrada (1 dato se analizaría por iteración) y 1 salida (indica si es ocultación o no, 0 ó 1). De ser así, RN Elman se adaptaría muy bien a esta situación y en este caso se tendría que analizar cuántas neuronas en la capa oculta serían convenientes utilizar, además de que se podrían utilizar varias capas de red oculta y podría mejorar el comportamiento de la RN; sin embargo, esto solo se puede determinar en el momento en que se entrene la red.

Para determinar la cantidad de neuronas en la capa oculta no existe un método analítico que se pueda utilizar, esto se hace a prueba y error y se verifica con el entrenamiento y las pruebas una vez entrenada la red; si no se obtiene el comportamiento deseado entonces conviene realizar los cambios. Y esto aplica también para conocer si se requiere utilizar más o menos capas ocultas de las que se planearon en un principio.

Es importante recalcar que no solo se puede utilizar una respuesta binaria de la red, también se puede entrenar a la RN para que en su resultado muestre un valor de probabilidad de una ocultación. Por ejemplo, se puede utilizar curvas de difracción para el entrenamiento con ciertos niveles de SNR, por lo que aprendería bien esos valores, y cuando se le someta a valores más altos de ruido daría probabilidades más bajas para cuando haya una ocultación pero también se disminuiría las falsas alarmas; por otro lado cuando el nivel de SNR es alto entonces el valor de probabilidad de ocultación sería alto, esto claro si se puede controlar por otros métodos el ruido.

Si no se utiliza la configuración de 1 entrada y 1 salida, se puede utilizar una RN donde se contemple la retroalimentación de capas posteriores a capas anteriores y que se analice un rango simultáneamente como lo hace la de tiempo de espera. Dicho esto, se podría utilizar una red de tiempo de espera donde se analice el espectro de la señal para cada estrella en el cual se prevea que tiene los suficientes valores de entrada con la información de la ocultación. Y la retroalimentación podría mejorar la respuesta de la red ante alteraciones en la señal, como lo es el ruido. Por otro lado, si se utilizara una red así se podría obtener información más compleja, ya que en lugar de utilizar la red para que detecte la ocultación se podría utilizar dicha red para que calcule el tamaño del objeto que realizó la ocultación, por lo que si el valor de respuesta de la red es muy bajo (unos cuantos metros) se sabría de antemano que no se ha producido ninguna ocultación sino que es solo ruido, pero una vez que éste valor aumente a cientos de metros se deduce que se detectó el objeto e inmediatamente se conoce su tamaño.

Para lo anterior conviene usar funciones de activación lineales, ya que estas no tienen un valor máximo o mínimo, por lo que ante situaciones como lo sería que un objeto ocultante sea de mucho mayor tamaño al que se preveía, la red podría dar un resultado

acercado a la realidad aunque en su conjunto de entrenamiento no se hubiera tomado en cuenta dicho tamaño. Si se utilizan porcentajes para valores ya establecidos del tamaño del objeto o de la distancia del objeto, funciones sigmoideas o rampas para la activación de las neuronas podrían ser convenientes usar, pero con el problema que si el objeto es más grande que los que se encontraban en el conjunto de entrenamiento dará como resultado un valor erróneo.

Una solución que se podría implementar es utilizar la foto tal cual se obtiene del telescopio y no las curvas para cada una de las estrellas, como se mencionó anteriormente. Esto sería conveniente realizar si el obtener los datos para todas las estrellas que se están analizando conlleva más gasto de recurso computacional y de tiempo. Para este caso la capa de entrada deberá ser del tamaño de píxeles que posee la imagen y así analizar toda la imagen, las redes convolucionales en este caso podrían utilizarse si la imagen presenta mucho ruido, ya que por medio de técnicas de visión se podría mejorar dicho problema. Posteriormente, es conveniente utilizar redes recurrentes pues aunque se analice la foto, en esencia se está entrenando a la red para que identifique por medio de un patrón de difracción ciertos parámetros, por lo que la naturaleza del tiempo (debido al patrón) debe mantenerse. Y para la salida de la red, esta podría indicar si existe una posible ocultación y además de las coordenadas en la imagen donde estaría la detección; podría existir el problema de que si la salida no se plantea bien se podrían perder ocultaciones si existen varias en una misma imagen.

Con respecto a la solución planteada con anterioridad, se podría utilizar la red también para que por medio de ella se obtenga el patrón de difracción de cada estrella y luego procesarse en otra RN.

Si por la complejidad que representa utilizar la imagen para entrenar la RN y se puede obtener por algún otro método la curva de difracción de cada estrella entonces se recomienda usar una red de tiempo de espera y es de suma importancia tener en cuenta la cantidad de valores en la de entrada que sean representativos a la señal que se desea analizar. Ya que si se utilizan pocos o muchos datos, cuando se le someta a una señal donde su patrón de difracción no se ajuste a lo que se utilizaron en el entrenamiento lo que podría ocurrir es que ya sea que la red no detecte la ocultación o que por el contrario, detecte siempre ocultaciones que no existen, y esto se agrava cuando se le agrega ruido a la señal que se obtuvo luego de ser procesada de la imagen original. Esto anterior se menciona recalcando lo que ya se había comentado con anterioridad.

Cabe destacar que las RN no solo se utilizan para el procesamiento de los datos de un problema (analizándolo como un todo) para obtener la información final que justamente se busca; también se utilizan en el procesamiento. Se puede utilizar para la disminución de ruido, de hecho gracias a que el cerebro humano se compone de RN es que son buenos en el análisis de información con ruido. Por lo que se podría implementar una solución donde la entrada sea la imagen o la señal ya obtenida para cada estrella, y en su salida el resultado

que se obtiene es la misma información pero sin ruido. Por lo que si se utiliza la imagen completa se podría luego implementar otros métodos para su análisis, ya sea someterla a funciones utilizadas en los sistemas de visión (se mencionaron algunas al inicio del documento) o inclusive a otra red para que se encargue de la obtención de la información final. Si lo que se utiliza es la señal proveniente de cada estrella, se podría luego de filtrarse el ruido con la RN utilizar métodos analíticos para el análisis de la señal, con ello se podría reducir tiempo de cómputo pues el filtrado de la señal no requeriría una RN muy compleja en cálculos y para el resto del análisis, una programación convencional con funciones analíticas no requieren mucho tiempo de procesado pues hay software especializados para realizar estas operaciones (Matlab, Octave, entre otros).

Uno de los factores que afectan en gran medida al comportamiento de una RN es el conjunto de entrenamiento al que se vea sometido y a dicho entrenamiento. El conjunto de entrenamiento debe ser lo más semejante a la realidad posible pero además debe ser representativo, y con esto se hace referencia a que no solo se le debe de proporcionar lo que se desea que aprenda, sino también se le debe de introducir los valores que no se desea que aprenda. Por ejemplo, si se hiciera una red para identificar una ocultación y solo se le facilitan datos de entrenamiento donde hay ocultaciones y siempre se tiene una respuesta positiva, cuando se le introduzca una señal sin ocultación podría detectarla aunque no exista, pues nunca se le permitió ajustar sus valores para que también no identifique una señal sin ocultación. Por otro lado, hay que cuidar el hecho de no sobreentrenar la red, pues si se da un sobre entrenamiento se “aprendería los valores de memoria” y esto puede incurrir en que cuando se le someta a una señal con una ocultación ligeramente diferente a las que se entrenó la red no responda como se esperaba, sino que su comportamiento sea muy definitivo, decisivo o arbitrario.

Existen varias formas de entrenamiento para una RN. Se recomienda investigar en las comunes pues dependiendo del conjunto de entrenamiento y del tiempo de cómputo que se tenga, se podría implementar diferentes formas para así verificar con cual se obtiene mejor resultado y realizar con esa el entrenamiento de la configuración de red definitiva.

Como se observó existen multitud de tareas que se pueden realizar con IA, pero requiere de un análisis previo para determinar la mejor solución a seguir y de poseer un conjunto de entrenamiento lo más representativo posible. Es indispensable que la red se ajuste a la naturaleza del problema, ya que de no ser así la respuesta podría ser errónea.