CAPÍTULO

2. ESTADO DEL ARTE

En este capítulo se presenta un panorama de los trabajos de investigación más relacionados con el trabajo desarrollado en esta tesis y que, de alguna forma dieron pie a la presente investigación.

2.1 Redes Neuronales con Álgebra Geométrica

(Bayro-Corrochano y Vallejo 2001)

En este trabajo se presenta un nuevo conjunto de redes neuronales que utilizan operadores del álgebra geométrica. Las redes neuronales tradicionales, las complejas y las cuaterniónicas se pueden generalizar en el marco del álgebra geométrica. Para estos autores, los pesos, las funciones de activación y las salidas se representarán como multi-vectores[1]. El producto geométrico se utiliza para realizar las operaciones de la red, en lugar del producto escalar.

Se propone una función de activación que afecta cada elemento base de lo multi-vectores, la función de activación $f\ \cdot$ para un multi-vector A de dimensión n está dada por:

[1] Un multi-vector es el producto exterior de dos o más vectores, véase sección. **¡Error! No se encuentra el origen de la referencia.**.

$$
\begin{aligned}
f(A) = f(& A_0 + A_i\sigma_i + A_j\sigma_j + A_k\sigma_k + \cdots + A_{ij}\sigma_i \wedge \sigma_j + \\
& \cdots + A_{ijk}\sigma_i \wedge \sigma_j \wedge \sigma_k + \cdots + A_n\sigma_i \wedge \sigma_j \wedge \ldots \wedge \sigma_n) \\
= \;& A_0 \; + f \;\; A_i \;\; \sigma_i + f \;\; A_j \;\; \sigma_j + f \;\; A_k \;\; \sigma_k + \cdots + f \;\; A_{ij} \;\; \sigma_i \wedge \sigma_j + \\
& \cdots + f \;\; A_{ijk} \;\; \sigma_i \wedge \sigma_j \wedge \sigma_k + \cdots + f \;\; A_n \;\; \sigma_i \wedge \sigma_j \wedge \ldots \wedge \sigma_n
\end{aligned}
$$

donde $\wedge$ es el producto exterior[2], los valores de $f \;\cdot$ pueden ser de tipo gaussiano o sigmoidal.

Mientras que las neuronas de (McCulloch y Pitts 1943) usan el producto escalar del vector de entrada y su vector de peso, en ese modelo se propone una *neurona geométrica.* En dicha neurona se sustituye el producto escalar tradicional por el producto geométrico, esto es:

$$
W^t X + \theta \;\; \Rightarrow \;\; wx + \theta = w \cdot x + w \wedge x + \theta
$$

Una neurona geométrica puede ser vista como un tipo de operador de correlación geométrico que, a diferencia de las neuronas de McCulloch-Pitts, ofrecen interpolación no solo entre puntos sino entre multi-vectores de alto orden como planos, volúmenes, híper-volúmenes, etc.

Una *red neuronal geométrica tipo feed-forward* se construye al cambiar el producto interior por el producto geométrico y las funciones de activación como se mencionó anteriormente.

En ese trabajo, se presenta una *generalización multidimensional* de regla de aprendizaje del gradiente descendiente en álgebra geométrica. La cual se utiliza para el perceptrón geométrico multicapa y para ajustar los pesos de las funciones geométricas de base radial.

[2] Para más información sobre el producto exterior véase la sección **¡Error! No se encuentra el origen de la referencia.**.

2.2 Máquinas de Soporte Vectorial en Álgebra Geométrica

(Bayro-Corrochano y Vallejo 2001)

Las máquinas de soporte vectorial aplican métodos de optimización para el aprendizaje. Con estas máquinas se pueden generar redes de dos capas, redes de base radial y redes con otros *núcleos*. En ese trabajo se generan, con álgebra geométrica, redes neuronales usando máquinas de soporte vectorial. Las llamadas *máquinas de soporte multi-vectorial*.

Una máquina de soporte multi-vectorial mapea un espacio de entrada multi-vectorial en un espacio de características de alto orden. Estas máquinas construyen un híper-plano de separación óptimo en el espacio multi-vectorial de características.

Usar álgebra geométrica con máquinas de soporte vectorial ofrece nuevas herramientas y un nuevo conocimiento sobre las mismas para aprendizaje multidimensional. Los autores realizan varios experimentos como estimación de movimiento rígido tridimensional y separación de dos conjuntos de puntos en tres dimensiones.

2.3 La neurona híper-esférica

(Banarer, Perwass y Sommer 2003)

En este trabajo se propone una neurona de *alto orden* que, para separar clases, usa una superficie no lineal. Se presenta una extensión simple de un perceptrón, cuya superficie de decisión no es un híper-plano sino una híper-esfera. El perceptrón híper-esférico multi-capa separa el espacio de entrada en regiones donde la clasificación es invariante a la orientación.

Esta híper-esfera se representa en el espacio conforme del álgebra de Clifford y para decidir si un vector de entrada se encuentra dentro o fuera de ella se utiliza un producto escalar estándar. Para clasificar un vector de entrada X en una híper-esfera normalizada S se utiliza el signo del producto interior $X \cdot S$.

Para construir un perceptrón cuya superficie de decisión es una híper-esfera se utilizan las componentes de S como los pesos correspondientes.

La función de propagación de una neurona híper-esférica se puede implementar como un producto escalar estándar. Durante la fase de entrenamiento, las componentes de S se consideran independientes. La neurona híper-esférica puede verse como un perceptrón estándar con un segundo *sesgo*. La complejidad se puede comparar con añadir un perceptrón adicional a cada capa en un perceptrón multi-capa convencional.

El resultado del producto escalar es la entrada de la función de activación, los pesos en una neurona híper-esférica se tratan como componentes independientes y representan una híper-esfera no normalizada.

Por ejemplo; sea X la representación conforme de un vector de entrada $\boldsymbol{x}$ y sea S la representación de una neurona híper-esférica con centro $\boldsymbol{c}$ y radio γ, además sea φ un escalar. Se considerar la función sigmoidal $\sigma\ \lambda, z\ =\ 1+\boldsymbol{e}^{-\lambda z}\ ^{-1}$ como la función de activación de la neurona. El entrenamiento de la neurona híper-esférica, para clasificar a $\boldsymbol{x}$ como perteneciente a la clase que representa esa neurona consiste en variar $\boldsymbol{c}$, γ y φ tal que $\sigma\ \lambda, X\cdot S\ > 1-\varepsilon$ donde ε es el umbral de decisión.

Para un ε, $\boldsymbol{c}$, y φ fijos; el radio de la híper-esfera depende del parámetro λ de la función sigmoidal. Mientras más pequeño sea λ, más grande será el radio.

2.4 Agrupando formas irregulares con neuronas de alto orden

(Lipson y Siegelmann 2000)

En este trabajo se presenta un método de *clusterización* de arreglos de datos con formas irregulares con neuronas de alto orden. Las formas analíticamente complejas se modelan al reemplazar los pesos sinápticos clásicos de la neurona con tensores de alto orden en coordenadas homogéneas. Las formas de alto orden se formulan siguiendo el principio de activación de máxima correlación y la regla de aprendizaje de Hebb.

En una neurona tradicional, para seleccionar la neurona que exhiba la máxima correlación con la entrada, se utiliza un criterio de distancia mínima euclidiana. Este criterio conlleva varias dificultades, al utilizar el criterio de distancia mínima implica que las características del dominio de entrada deben ser esféricas.

El trabajo presenta un intento de generalizar el esquema esférico/elíptico a métricas más generales. En general, la restricción de forma de las neuronas clásicas se puede evitar al reemplazar el vector de pesos o la matriz de covariancia de una neurona clásica con un tensor de alto orden, capaz de formar correlaciones multilineales entre las señales asociadas a las neuronas. Esto permite crear formas con hoyos y/o áreas separadas.

Se utilizan, además las coordenadas homogéneas para combinar las correlaciones de diferente orden en un tensor sencillo. Estas neuronas de alto orden presentan buena estabilidad y buen desempeño al entrenarlas con aprendizaje Hebbiano simple.

2.5 Superficies de decisión esféricas usando el modelo conforme

(Perwass, Banarer y Sommer 2003)

Al incrustar el espacio euclidiano en un espacio conforme, las híper-esferas se pueden expresar como vectores. El producto escalar entre puntos y híper-esferas en el espacio conforme da una medida de qué tan lejos está un punto dentro o fuera de una híper-esfera. Los autores muestran que una neurona híper-esférica se puede implementar como un perceptrón con dos *sesgos*. También argumentan que un perceptrón multicapa basado en estas neuronas es similar a una red de base radial.

La idea básica de este trabajo es separar el espacio de entrada en dos clases utilizando híper-esferas. La separación lineal, en general, no es suficiente para fines prácticos, ya que lo datos están separados en varias clases y cada clase cubre una región en particular del espacio de entrada. Dependiendo de la estructura de los datos puede ser más útil usar neuronas que utilicen una superficie de separación no lineal en lugar de usar perceptrones. Dichas neuronas se conocen como *neuronas de alto orden*.

Los autores esperan que superficies de separación más complejas resuelvan una tarea con pocas neuronas, sin que el costo computacional se vea afectado. Para esto proponen una extensión simple del perceptrón, donde las superficies de separación son híper-esferas en lugar de híper-planos. La ventaja de esta representación es que, un producto escalar estándar es suficiente para decidir sí un vector de entrada está dentro o fuera de la híper-esfera.

En términos de complejidad concluyen que la complejidad de una red multi-capa híper-esférica es la misma que la de un perceptrón multi-capa pero agregando un perceptrón a cada capa. Además, se puede interpretar como una extensión de una red de base radial con igual número de neuronas.

2.6 Cálculo neuronal con algebra geométrica

(Buchholz 2005)

En esa tesis se presenta al álgebra geométrica como un marco de trabajo para el cálculo neuronal, o dicho de otra manera, el autor desarrolla un modelo de redes neuronales basado en el álgebra geométrica. Su principal objetivo es demostrar la utilidad del álgebra geométrica para el cálculo neuronal, debido a la naturaleza geométrica del mismo. En particular el autor desarrolla las llamadas *neuronas de Clifford*; parecidas a las neuronas clásicas, pero con las diferencias que los pesos y el *sesgo* se reemplazan por multi-vectores y se utiliza el producto de Clifford en lugar del producto tradicional.

Una neurona de Clifford básica se define de la siguiente manera:

$$Y = \sum_{i=1}^{n} W_i \otimes X_i + B,$$

donde *Y*, *X*, *B y* W_i son multi-vectores y $\otimes$ es el producto de Clifford. Estas neuronas son parecidas a las neuronas de alto orden, ya que desde un punto de vista formal, ambas procesan funciones polinomilaes a las entradas.

Una neurona de Clifford básica se puede ver como un asociador lineal; si se está trabajando con un álgebra geométrica de dimensión *1,* éstas se comportan como neuronas tradicionales. Sin embargo, se desarrollaron neuronas de Clifford de orden superior, como las neuronas de Clifford básicas complejas, las hiperbólicas o las duales; las tres de dimensión 2. Con estas neuronas se pueden calcular transformaciones de varias entidades geométricas.

Algo interesante es que por medio de estas neuronas básicas se puede desarrollar un *asociador de Clifford*, que es simplemente una red neuronal con neuronas de Clifford en lugar de las neuronas tradicionales y utiliza, como se mencionó anteriormente, el producto de Clifford, en lugar del producto tradicional.

Además de las neuronas básicas, se presentan las llamadas neuronas de Clifford *espinoriales*. A diferencia de su contraparte, la neurona básica, éstas se definen mediante la siguiente función:

$$Y = \sum_{i=1}^{n} W_i \otimes X_i \otimes \boldsymbol{f}\ \ W\ \ + B,$$

donde $\boldsymbol{f}\ W$ es la función reversión, conjugación o inversión. Las neuronas espinoriales son una arquitectura neuronal basada en dos productos geométricos. De ésta se derivan las *neuronas de Clifford espinoriales cuaterniónicas* y las *isomórficas*.

Ambas neuronas, las básicas y las espinoriales, pueden considerarse como un punto de partida para futuros trabajos sobre redes de perceptrones multicapa de segundo orden (Buchholz, Tachibana y Hitzer 2007).