



UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA
Instituto Universitario de Sistemas Inteligentes
y Aplicaciones Numéricas en Ingeniería

Programa de Doctorado:
Sistemas Inteligentes y Aplicaciones Numéricas en Ingeniería

Tesis Doctoral
Aplicación de Redes Neuronales
para la Resolución de Problemas
Complejos en Confiabilidad y Riesgo

Mustapha Maarouf
2015

UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA (ULPGC)
Instituto Universitario de Sistemas y Aplicaciones Numéricas en Ingeniería
(SIANI)



Aplicación de Redes Neuronales para la Resolución de Problemas Complejos en Confiabilidad y Riesgo

Autor
Mustapha Maarouf

Director
Dr. Blas José Galván González

Programa de Doctorado:
Sistemas Inteligentes y Aplicaciones Numéricas en Ingeniería

Tesis Doctoral
Las Palmas de Gran Canaria
2015

Dedico este trabajo a mis padres.

Agradecimiento

Quiero agradecer a mi familia por su apoyo. A mi tutor, Dr. Blas Galván José González, quien ha aceptado dirigir mi tesis. Estoy muy agradecido por sus esfuerzos, por sus valiosos consejos académicos y no académicos, por su apoyo, por su insistencia, y que ha ayudado mucho para terminar mis estudios de investigación.

Agradezco también a los miembros del jurado, que no sólo accedieran a probar esta tesis, pero también que dieron su tiempo y esfuerzo para corregir mi tesis.

Agradezco sinceramente a todos los profesores e investigadores con quien he tenido contacto durante mis estudios y que hayan proporcionado documentos o datos, me orientaron en mis trabajos de investigación. No se me olvidan también los profesores que tuve en toda mi vida, el primer de ellos ha sido mi padre, mi gran maestro, del pasado, el presente y el futuro, siempre estará presente. Agradezco a mi familia sus esfuerzos y su confianza, a todos mis colegas dentro o fuera del trabajo, con los he compartido los altibajos a lo largo de los años.

Por último, un gran agradecimiento a los que no han sido nominados y que me han ayudado de alguna manera en mi camino personal y profesional.

Resumen

La ingeniería de Fiabilidad, Mantenibilidad, Disponibilidad y Seguridad/Riesgo (Ingeniería RAMS) se enfrenta a retos importantes debido a la creciente magnitud y complejidad de muchos de los modernos procesos industriales, de su logística, de las comunicaciones necesarias para su correcto funcionamiento, de la interacción con el ser humano, de las amenazas externas que comprometen el correcto funcionamiento de los sistemas técnicos que les dan soporte y de eventos no deseados que pueden afectar a los bienes, la naturaleza o al ser humano. Entre otros cabe citar la obtención dinámica de planes óptimos de mantenimiento basados en la condición de los equipos, la determinación en tiempo real del nivel de riesgo asociado a determinados eventos procesos críticos, la determinación del alcance de las consecuencias de sucesos no deseados (explosiones, nubes tóxicas, vertidos marinos, etc.) y la estimación de la propagación de eventos difícilmente controlables como los incendios (en industrias o en el medio humano: Ciudades, Bosques, etc.).

El modelado de tales procesos y eventos ha de realizarse en general usando formulaciones matemáticas muy complejas cuya solución implica frecuentemente el empleo de métodos numéricos y uso intensivo de recursos de computación. La solución de casos reales suele por tanto ser costosa en tiempos de CPU y tardar semanas o meses hasta alcanzar valores satisfactorios. Existen sin embargo situaciones en las que no es viable esperar semanas o meses para dar soluciones a un problema, tal es el caso de la predicción de la evolución de un vertido marino, una nube tóxica, de una avería en un equipo crítico o un incendio forestal, entre otros. En estas situaciones es necesario obtener respuestas satisfactorias (no necesariamente óptimas) en tiempos muy cortos (horas o minutos), lo que convierte al conjunto heterogéneo de estas situaciones en una clase de problemas con características específicas que requieren de un tratamiento diferenciado y en los que el uso de la Inteligencia Artificial (IA) adquiere especial relevancia. Entre las metodologías de IA y para esta clase problemas mencionados cabe destacar el uso de Redes Neuronales por su capacidad de

reflejar relaciones complejas entre variables y de dar respuestas rápidas aceptables. El trabajo presentado en esta tesis doctoral se enmarca por tanto de forma general en los problemas de Ingeniería RAMS y en particular en la clase de problemas complejos y/o críticos que requieren de una respuesta rápida no necesariamente óptima, siendo estos resueltos mediante Redes Neuronales. Como punto de partida se realiza un estudio del estado del arte de las Redes Neuronales con especial énfasis en sus aplicaciones para Ingenierías RAMS y la clase de problemas mencionados. Se estudiará también el estado del arte de los modelos matemáticos que dan soporte científico a esos problemas centrándose en dos modelos especialmente escogidos por su representatividad: Los de Vida Útil Remanente y los de Incendios Forestales. Las contribuciones de esta tesis radican no solo en novedades en el modelado de tales problemas sino también en ciertas mejoras en las Redes Neuronales que incrementan la calidad de sus soluciones cuando se aplican a esos problemas. Se ha diseñado un conjunto de experimentos que incluye situaciones reales complejas que son resueltos con diferentes tipos de Redes Neuronales y diversos métodos de entrenamiento. Diferentes métricas de interés se usan para comparar los resultados y alcanzar conclusiones contrastadas. El resultado final es la demostración con base científica de que es viable obtener un conjunto de Redes Neuronales entrenadas y validadas que permita dar respuestas rápidas aceptables para problemas relevantes de la clase de problemas abordados.

Abstract

Reliability engineering and risk analysis facing significant challenges due to the size, complexity and risks of many modern industrial processes, their logistics, of the communications necessary for its correct operation, from interacting with humans, from external threats that compromise the proper functioning of the technical systems, that support them and undesired events that may affect the property, nature or human beings. In addition, the realization of intelligent maintenance plans based on the actual condition of industrial equipment, determination of the risk level associated with hazardous industrial activities, calculation of the extent of the risk of fire, contaminating leaks, toxic clouds, the consequences of industrial accidents, etc.

The modeling of such processes has been generally performed using complex mathematical formulas, whose solution often involves the use of numerical methods and intensive computing resources. The solution of real problems is usually costly in terms of CPU time; processes can employ weeks or months to reach satisfactory solutions.

However, there are situations where it is not feasible to wait a long time to get solutions, as it is the case of forecasting the forest fire behaviour, a toxic cloud or a marine spill, etc. In these situations it is necessary to get satisfactory answers in a very short time (hours or minutes), therefore the use of techniques of Artificial Intelligence (AI) is especially relevant. Among the methodologies of AI and for the above problems, there is the use of neural networks for their ability to reflect complex relationships between variables and to provide acceptable answers in short time.

Therefore, we interest in performing a doctoral thesis on the application of Neural Networks to solve complex problems in Reliability and Risk. In this thesis, study and summarize of the state of the art in terms of neural networks, and industrial type applications with special emphasis on problems related to Reliability and Risk,

we will also be explored the mathematical models that give scientific support to the complex problems of Reliability and Risk, and the behavior of forest fires. Modeling and resolving these applications with different kind of neural networks, with different training methods, will be compared. Among the expected results of the proposed thesis, is to obtain a set of trained and validated neural networks, to provide rapid response to the applications considered, with an assessment of the reliability of their responses. The relevant results are validated by scientific publications in journals with impact index.

Résumé

L'ingénierie de la fiabilité, de la disponibilité, de la maintenabilité et de la sécurité (Ingénierie FDMS) fait face à des défis importants due à la croissance de la complexité de nombreux processus industriels modernes, de leurs logistique, des communications nécessaires pour leurs correcte fonctionnement, de leurs interaction avec les humains, des menaces extérieures qui mettent en péril le bon fonctionnement des systèmes techniques qui les soutiennent ou qui peuvent provoquer des événements indésirables.

La modélisation de ces processus doit être réalisée en utilisant des formules mathématiques complexes. Généralement, la solution implique souvent l'utilisation de méthodes numériques et des ressources qui nécessitent des calculs intensifs. La résolution des cas réels est généralement si coûteuse en termes de CPU, ainsi que les processus peuvent atteindre des semaines ou des mois pour pouvoir trouver des solutions satisfaisantes. Cependant, il y a des situations où ce n'est pas facile d'attendre des semaines pour résoudre tel problème, comme le cas de la prévision de l'évolution d'un feu de forêt, un nuage toxique ou les détritrus marins, etc. Dans ces situations, il est nécessaire d'obtenir des réponses satisfaisantes en peu de temps (question d'heures ou quelques minutes), alors que l'usage des techniques de l'intelligence artificielle (IA) est particulièrement pertinente. Parmi les méthodes de l'IA et pour les problèmes ci-dessus nommés, on trouve l'utilisation de réseaux de neurones, du a leur capacité de trouver des relations entre des variations complexes et qui peuvent fournir des réponses acceptables et rapides.

Par conséquent, notre objectif est de réaliser une thèse sur l'application des réseaux de neurones pour résoudre des problèmes complexes en matière de fiabilité et des risques. Dans cette thèse, le but est de réaliser et de résumer l'état de l'art en termes de réseaux de neurones, et des applications de type industriel ou écologique, avec une attention particulière sur les problèmes liés à la fiabilité et les risques.

Aussi bien, l'analyse du comportement des incendies de forêt sera également considérée. La modélisation mathématique et les résultats des modèles obtenues avec différents types de réseaux de neurones, avec différentes méthodes d'entraînement, seront analysés. Parmi les résultats attendus de cette thèse proposée, est d'obtenir un ensemble de réseaux de neurones entraînés et validés pour offrir des réponses rapides aux problèmes examinés, en tenant compte de l'évaluation de la fiabilité de leurs réponses. Les résultats scientifiques pertinents seront validés par des publications dans des revues avec indice d'impact.

Preámbulo

El uso de la Inteligencia Artificial (IA) se ha generalizado tanto en ciencia como en tecnología por lo que contribuye actualmente, de múltiples formas, al incremento del conocimiento así como a la mejora de habilidades necesarias para afrontar con éxito determinadas actividades relevantes para la humanidad. Son especialmente significativos los avances que el desarrollo de la IA ha permitido en la interacción evento-máquina como ayuda y en algunos casos sustitución de la intervención humana. La inherente complejidad del uso de la IA en esa interacción puede causar malentendidos, y sin duda puede ser interpretada de diferentes maneras. En efecto, si todo el mundo puede disfrutar cada vez más de las capacidades de las máquinas para realizar cálculos numéricos y operaciones simbólicas, la idea de que la máquina podría contener unas pocas onzas de inteligencia conduce naturalmente a muchas preguntas cuyas respuestas requieren en gran parte del desarrollo de mayores y más complejas tareas de investigación.

Una parte relevante de las investigaciones en IA se dedican a la mejora de la capacidad de las máquinas para adquirir información, razonar en situaciones estáticas o dinámicas, resolver problemas combinatorios, realizar diagnósticos, proponer decisiones o planes de acción y comunicar o explicar conclusiones. Como resultado de esas investigaciones hoy en día se acepta que la máquina debe estar equipada con métodos, procedimientos de ensamblaje de métodos y protocolos de actuación, cuyo nivel de generalidad permita la auto-adaptación a diversas clases de situaciones. Esto en sí mismo un enorme reto cuya factibilidad se va valorando conforme se van presentando nuevos avances exitosos.

La investigación en IA tampoco puede eludir evidentes relaciones con la ciencia cognitiva, ya que, en primer lugar, proporciona nuevos puntos de referencia para la concepción del significado de la inteligencia y su comprensión. En segundo lugar puede contribuir al aprendizaje de cómo funciona el cerebro y cómo es el proceso de razonamiento humano, aunque nada obliga a que la IA deba copiar la mente humana

en todas sus formas y procesos. Además, puesto que la máquina debe compartir sus resultados con los usuarios, es importante que se pueda expresar en términos cognitivamente significativos para ellos.

La investigación realizada en esta tesis se enmarca dentro de ese proceso, casi infinito, de mostrar evidencias con base científica de lo que la IA puede aportar a la humanidad, centrándose en determinadas aplicaciones de diagnóstico y predicción relevantes dentro de lo que se conoce como Ingeniería de Fiabilidad, Mantenibilidad, Disponibilidad, Seguridad y Riesgo.

Índice de Figuras	xvii
Índice de Tablas	xxi
1 Introducción	1
1.1 Introducción	2
1.2 Resumen de contribuciones de esta tesis	6
1.3 Publicaciones relacionadas con la línea de investigación	8
1.3.1 Proyectos	8
1.3.2 Publicaciones	8
1.3.3 Capítulos de libros	9
1.3.4 Congresos	9
1.3.5 Organización de Esta Memoria	10
2 Estado del arte	11
2.1 Redes Neuronales Artificiales “RNAs”	12
2.1.1 Historia de las Redes Neuronales	12
2.1.2 Ventajas de las Redes Neuronales Artificiales	14
2.1.3 Las Redes Neuronales Artificiales vs. los Equipos Convencionales	14
2.1.4 Las predicciones mediante las RNA	15
2.1.5 Presentación General, Definiciones y Teoremas	16
2.1.5.1 Definiciones y Teoremas	16
2.1.5.2 El Adaline	18

ÍNDICE

2.1.5.3	El Perceptrón	23
2.1.5.4	Conclusiones Fundamentales	27
2.1.5.5	Funciones de Activación	28
2.1.6	El Aprendizaje Numérico	32
2.1.6.1	Aprendizaje Supervisado	32
2.1.6.2	Aprendizaje no-supervisado	33
2.1.7	Conceptos básicos de una Red Neuronal	33
2.1.8	Proceso de Aprendizaje	34
2.1.8.1	Aprendizaje mediante Corrección del Error	34
2.1.8.2	Aprendizaje mediante la Regla de Hebb	35
2.1.9	Redes Neuronales Multicapa	37
2.1.9.1	Aproximación de una función	37
2.1.9.2	Clasificación	40
2.1.10	Retro-Propagación	41
2.1.10.1	Repaso Histórico del Algoritmo de Retro-Propagación	41
2.1.10.2	Introducción	41
2.1.10.3	Algoritmo de entrenamiento: Retro-Propagación y RPROP	42
2.1.10.4	Algoritmo RPROP	46
2.1.11	Fundamentos para la implementación de una RNA	47
2.1.11.1	Condición de Término	47
2.1.11.2	Fenómeno de Saturación	48
2.1.11.3	Selección de variables	49
2.1.11.4	Normalización de las variables	50
2.1.11.5	Inicialización de parámetros	50
2.1.11.6	Criterio de Parada del Entrenamiento	52
2.1.11.7	Selección del modelo	52
2.1.11.8	Validación Cruzada	54
2.1.11.9	Validación cruzada dejando uno fuera	55
2.1.12	Estimación de los Intervalos de Confianza en la Predicción de una RNA	56
2.1.12.1	Estimación con Mínimos Cuadrados	61
2.1.12.2	Intervalos de Confianza en la predicción de la red neuronal	65
2.1.13	Red Neuronal de Base Radial “RBF”	66
2.1.14	Máquinas de Soporte Vectorial “MSV”	67

2.1.15	Comparación entre las redes neuronales multicapas (MLP) y las redes de base radial (RBF)	68
2.2	Comportamiento de Incendios Forestales	70
2.2.1	Entorno del Incendio	71
2.2.2	Sistemas de evaluación del peligro de los incendios	72
2.2.3	Sistemas de un solo índice	73
2.2.4	Sistemas con múltiples índices	75
2.2.4.1	Sistema Canadiense de Evaluación de Peligro de Incendios Forestales	75
2.2.4.2	Sistema de evaluación de peligro de incendios de los Estados Unidos (NFDRS)	77
2.2.4.3	Conceptos Importantes	78
2.2.5	El uso de la Inteligencia Artificial en los Incendios Forestales	80
2.3	Diagnóstico / Pronóstico en Equipos Industriales	82
2.3.1	Gestión del ciclo de vida (GCV)	83
2.3.2	El Problema del Diagnóstico	84
2.3.3	El Problema del Pronóstico	86
2.3.4	Vida Útil Remanente “VR”	89
2.3.5	Clasificación de los modelos para la estimación de la VR	91
2.3.6	Observaciones importantes	94
2.3.7	Conclusión	94
2.4	Estimación de la Demanda de Energía Eléctrica	96
2.4.1	Introducción	96
2.4.2	Importancia y necesidad de una estimación eficiente de DEE	96
2.4.3	Revisión de los principales modelos aplicados	99
2.4.4	Problemática del estudio propuesto	100
3	Contribuciones	103
3.1	Introducción	104
3.2	Contribuciones en las Redes Neuronales Artificiales	104
3.2.1	Aportación para la definición de la estructura	104
3.2.2	Aportación sobre la inicialización de pesos	106
3.2.2.1	Aportación sobre las funciones de activación	112
3.2.2.2	Aportación sobre la fiabilidad	115

ÍNDICE

3.3	Contribuciones en la Aplicación de RNA en la Predicción de la Vida Útil Remanente en Equipos Industriales	116
3.3.1	Caracterización de la VR	117
3.3.2	Caracterización del caso de Información Incompleta	117
3.3.3	Características de las RNA propuestas para estimar la VR	119
3.4	Contribuciones en la Aplicación de las RNA en el Comportamiento de Incendios Forestales	122
3.5	Contribuciones en la Aplicación de las RNA en la estimación de la demanda de energía eléctrica	126
4	Experimentación y Aplicaciones	129
4.1	Experimentación en la Mejora del diseño y del uso de las Redes Neuronales Artificiales	130
4.1.1	Experimentación para la Inicialización de los pesos de una RNA	130
4.1.2	Experimentación en las funciones de activación	131
4.1.3	Experimentación para la definición de la estructura de una RNA	134
4.1.4	Conclusiones	136
4.2	Experimentación en el uso de las RNAs para la Predicción de la Vida Remanente Útil en Equipos Industriales	137
4.2.1	Problema propuesto	137
4.2.1.1	Análisis de Datos y selección de Entradas para el Modelo	138
4.2.1.2	Selección del modelo neuronal	140
4.2.1.3	Primer enfoque para la predicción del Índice de Salud	141
4.2.1.4	Segundo enfoque para la predicción del Índice de Salud	144
4.2.2	Comparación de los dos enfoques	149
4.2.3	Evaluación de la Fiabilidad - Rendimiento de los Modelos	150
4.2.4	Conclusiones	159
4.3	Experimentación en el uso de las RNAs para la Predicción del Comportamiento de Incendios Forestales	160
4.3.1	Estructura del modelo ANNFBP	160
4.3.2	Arquitectura del Modelo ANNFBP	160
4.3.3	Entrenamiento y Validación del ANNFBP	161
4.3.4	Test del modelo ANNFBP con otros casos de estudio	162
4.3.5	Conclusiones	169

4.4 Experimentación de las RNA para la estimación de la Demanda de Energía	
Eléctrica	170
4.4.1 Análisis de datos	170
4.4.2 Estimación de la DEE mediante regresión lineal y logarítmica	173
4.4.3 Estimación de la DEE mediante Inteligencia Artificial	174
4.4.3.1 Mediante MSV	174
4.4.3.2 Mediante AG	176
4.4.3.3 Mediante RNA	176
4.4.4 Eficiencia de los métodos	177
4.4.5 Medida de confianza para la RNA	178
4.4.6 Conclusiones	180
5 CONCLUSIONES Y LÍNEAS FUTURAS	181
Bibliografía	187

ÍNDICE DE FIGURAS

2.1	Matriz de memoria de correlación	16
2.2	ALC	18
2.3	Red con tres capas con funciones umbrales binarias	26
2.4	Ejemplo de una Red Neuronal Recurrente	27
2.5	Función lineal $f(x) = x$	29
2.6	Función Gaussiana $f(x) = \exp(-x)$	30
2.7	Función Sigmoidal $f(x) = \text{sigm}(x)$	31
2.8	Función Tangente hiperbólica $f(x) = \text{Tanh}(x)$	31
2.9	Red de una Neurona	34
2.10	Red Neuronal Multicapas (L capas) con n variables de entrada y m variables de salida	38
2.11	Red neuronal de tres capas	43
2.12	Modelo conceptual para la gestión de incendios forestales (Burrows 1994)	71
2.13	El Sistema Canadiense de Evaluación de Peligro de Incendios Forestales	76
2.14	El Índice Meteorológico de Peligro de Incendios (FWI)	78
2.15	Estructura del sistema de evaluación de peligro de incendios de los Estados Unidos	79
2.16	Visión general de la gestión del ciclo de vida para sistemas críticos	84
2.17	Estructura general del diagnóstico mediante RNA	87
2.18	Etapas del Pronóstico acorde con ISO 13381-1 (126)	89
2.19	Estrucutra general para el pronóstico con RNA	90
2.20	clasificación de metodologías para la Estimación de la Vida Remanente	92

ÍNDICE DE FIGURAS

3.1	Etapas para aplicar las RNA	105
3.2	Programa NNFINDER	107
3.3	Evolución del error (Datos 1)	109
3.4	Evolución del error (Datos 2)	109
3.5	Función α -Sigm	114
3.6	Función α -Tanh	115
3.7	Programa NNCI	116
3.8	Sensor 4	118
3.9	Estructura del modelo ANNFBP	124
3.10	Algoritmo para la reducción de entradas	125
3.11	El índice de FWI	126
3.12	Estructura del sistema FBP	127
3.13	El modelo neuronal para la estimación de la DEE	128
4.1	Evolución del error con Matriz de Hilbert y diferentes ejecuciones (Datos 1) . . .	130
4.2	Evolución del error con matriz de Hilbert y diferentes ejecuciones (Datos 2) . . .	131
4.3	Evolución del error con la función de activación sigmoideal y $\alpha = 2$ -Sigm y $\alpha = 3$ -Sigm (Datos 1)	132
4.4	Evolución del error con la función de activación sigmoideal y $\alpha = 2$ -Sigm y $\alpha = 3$ -Sigm (Datos 2)	133
4.5	Evolución del error con la función de activación Tanh y $\alpha = 2$ -Tanh y $\alpha = 3$ -Tanh (Datos 1)	133
4.6	Evolución del error con la función de activación Tanh y $\alpha = 2$ -Tanh y $\alpha = 3$ -Tanh (Datos 2)	134
4.7	Definición de la vida remanente útil	137
4.8	Sensor 1 vs sensor 2	139
4.9	Sensor 7 vs sensor 12	139
4.10	Dispersión (sensor1 - sensor2) y (Sensor 7- sensor 12)	140
4.11	Arquitectura de la red neuronal para la predicción de IS	141
4.12	IS modelo vs regresión lineal (Unidad 1)	142
4.13	IS modelo vs regresión lineal (Unidad 6)	142
4.14	IS modelo vs regresión lineal (Unidad 147)	143
4.15	IS Modelo vs Modelo Fit regresión (Unidad 1)	145

ÍNDICE DE FIGURAS

4.16 IS Modelo vs Modelo Fit regresión (Unidad 6)	145
4.17 IS Modelo vs Modelo Fit regresión (Unidad 147)	146
4.18 IS modelo vs regresión lineal (considerando $t - 2$, $t - 1$ y t) (Unidad 1)	146
4.19 IS modelo vs regresión lineal (considerando $t - 2$, $t - 1$ y t) (Unidad 6)	147
4.20 IS modelo vs regresión lineal (considerando $t - 2$, $t - 1$ y t) (Unidad 147)	147
4.21 IS modelo vs regresión lineal (Unidad 1)	148
4.22 IS Modelo vs Modelo regresión lineal ajustada (Fit) (Unidad 6)	148
4.23 IS Modelo vs Modelo regresión lineal ajustada (Fit) (Unidad 147)	149
4.24 IS: primer enfoque vs segundo enfoque (Unidad 1)	150
4.25 IS: primer enfoque vs segundo enfoque (Unidad 6)	151
4.26 IS: primer enfoque vs segundo enfoque (Unidad 147)	151
4.27 IS vs. IS_Modelo y el intervalo de confianza del modelo (Unidad 1)	153
4.28 Indicador 1 (Unidad 1)	154
4.29 Indicador2 (Unidad 1)	154
4.30 Indicador 3 (Unidad 1)	155
4.31 IS vs. IS_Modelo y el intervalo de confianza del modelo (Unidad 6)	155
4.32 Indicador 1 (Unidad 6)	156
4.33 Indicador 2 (Unidad 6)	156
4.34 Indicador 3 (Unidad 6)	157
4.35 IS vs. IS_Modelo y el intervalo de confianza del modelo (Unidad 147)	157
4.36 Indicador 1 (Unidad 147)	158
4.37 Indicador 2 (Unidad 147)	158
4.38 Indicador 3 (Unidad 147)	159
4.39 Comparación de la predicción de FWI: ANNFBP vs. Estación Mt Cook	164
4.40 Predicción de FWI y VP mediante ANNFBP-2 y Estación de Alberta	166
4.41 Dispersión matricial de los datos disponibles	168
4.42 DEE vs. POB, Islas Canarias	171
4.43 DEE vs. PIB, Islas Canarias	172
4.44 Estimación de la DEE mediante RLM	173
4.45 Estimación de la DEE mediante RLog	174
4.46 Estimación de la DEE mediante MSV con núcleo lineal	175
4.47 Estimación de la DEE mediante MSV con núcleo gaussiano	175
4.48 Estimación de la DEE mediante AG	176

ÍNDICE DE FIGURAS

4.49 Estructura de la RNA para predecir la DEE	177
4.50 Estimación de la DEE mediante RNA	178
4.51 Porcentaje de precisión para datos de ajuste	179
4.52 Porcentaje de precisión para datos de test	179

ÍNDICE DE TABLAS

4.1	Resultado de NNFINDER ejemplo 1	135
4.2	Evaluación Enfoque 1	152
4.3	Evaluación Enfoque 2	152
4.4	Arquitectura del ANNFBP	161
4.5	Rendimiento del Modelo ANNFBP	163
4.6	Predicción de VP mediante ANNFBP vs. Observado por La estación de Mt Cook basandose en modelos de Nueva Zelanda	165
4.7	Evaluación del modelo ANNFBP v. Estación de Alberta	166
4.8	Evaluación de la predicción del area quemada: ANNFBP-5 frente a las medi- ciones reales	167
4.9	Evaluación del error de estimación de los métodos	178

CAPÍTULO 1

INTRODUCCIÓN

1. INTRODUCCIÓN

1.1 Introducción

En los últimos años los avances tecnológicos han emergido exponencialmente en todos los aspectos de la sociedad del siglo XXI. Desde el diseño óptimo y la fabricación inteligente robotizada hasta el consumo masivo negociado basado en los nuevos paradigmas emergentes de interconexión y comunicación (“Internet of Things – IoT, Internet of Everything - IoE, o Internet of Anything – IoA”), la tecnología ofrece día a día posibilidades y servicios inimaginables pocos años atrás.

La creciente y cada vez más estrecha relación entre los consumidores y los proveedores que los nuevos paradigmas van descubriendo, propicia la aparición de nuevos equilibrios en esa relación que conllevan una constante revisión tanto de los requerimientos (1) como de las obligaciones (2) de ambas partes. Ofrecer productos o servicios sofisticados, de consumo masivo y alta disponibilidad, pero cercanos a los deseos de los usuarios finales, están entre los requerimientos esenciales para negocios rentables, mientras que ofrecer garantías de continuidad, ser respetuosos con el medio ambiente y mostrar claramente la preocupación por la seguridad de las personas, entre otras, van pasando a convertirse en obligaciones para negocios sostenibles. La reducción del consumo de energía puede considerarse tanto en (1) como en (2), marcando una tendencia imparable hacia el uso racional de los recursos energéticos.

La contribución de la Ingeniería de Fiabilidad, Mantenibilidad, Disponibilidad y Seguridad/Riesgo (Ingeniería RAMS - *I – RAMS* “**R**eliability, **A**vailability, **M**aintainability, **S**afety / **S**ecurity Engineering”) a los avances tecnológicos ha crecido mucho en las últimas décadas, ofreciendo novedades metodológicas que han contribuido de forma notoria a dar soluciones a (1) y (2) respecto a múltiples ejes (humanos, industriales y medioambientales entre otros). La Investigación y el Desarrollo también han crecido en *I – RAMS*, existiendo un gran volumen de publicaciones científicas que muestran aplicaciones de la *I – RAMS* a muy diversos tipos de productos, servicios y aplicaciones en diversas condiciones (por ejemplo: en (1) los autores discuten los retos de los métodos disponibles para la recopilación de datos y proponen una metodología para la recopilación de datos teniendo en cuenta el efecto de las condiciones ambientales, en (2) presenta nuevas técnicas probabilística de análisis que mejoran el análisis RAMS y como aplicación a sistemas de guías aéreas. En (3), los autores integran distintas técnicas de análisis y parámetros de incertidumbre para la evaluación del riesgo en los cambios que pueden afectar los requisitos de vigilancia. En (4), se ha aplicado una nueva estrategia de análisis de

fiabilidad basada en un modelo discreto en el lugar de modelos de diagramas de decisión binarios, dando así mejores resultados en dos casos reales de sistemas “phased-mission”. En (5), los autores presentan en su segunda edición de su libro, un nuevo material sobre la cultura de la seguridad, citando nuevos desarrollos de indicadores principales y modelos para la seguridad).

En el contexto de los paradigmas anteriormente citados la *I-RAMS* debe afrontar múltiples retos, entre ellos cabe destacar los derivados de la complejidad y la no linealidad de los modernos e interconectados procesos/sistemas/servicios industriales, así como de determinados fenómenos naturales relacionados con la seguridad de las personas, los bienes y el medio ambiente. Entre otros, tanto por su importancia como por su relación con los trabajos descritos en esta memoria, cabe destacar:

- El descubrimiento temprano de patrones insertados en señales masivas procedentes de equipos en funcionamiento. Miles/millones de equipos conectados a las redes y enviando múltiples señales procedentes de sensores embarcados. Con dos casos típicos: *Patrones de Comportamiento social* contenidos en las señales procedentes de terminales móviles de uso individual (ejemplo: teléfonos móviles) y *Patrones de Degradación* ocultos en las señales procedentes de sensores instalados en sistemas técnicos distribuidos por grandes áreas geográficas (ejemplo: miles de aerogeneradores continentalmente distribuidos, datos de degradación del medio ambiente y recursos naturales).
- Dificultad en la predicción de valores futuros en series temporales asociadas a procesos industriales (por ejemplo: predecir la vida útil remanente, o estimar la demanda de energía eléctrica futura) o ciertos fenómenos naturales (por ejemplo la estimación de la propagación de los incendios forestales), dado la alta no-linealidad y el tiempo limitado para dar respuestas. La predicción de la evolución en el tiempo de determinados procesos, implica la predicción de los valores futuros de la serie temporal que lo describe (6). Los cambios pueden ser tratados como una serie temporal. Sin embargo, la complejidad y la no linealidad del proceso plantean fuertes desafíos a los métodos estándar de análisis de series temporales.
- La creación de modelos *I-RAMS* realistas que contemplen la gran cantidad de variables, y la complejidad de las relaciones entre ellas, que intervienen en los modernos procesos productivos, que incluyen: *La maquinaria* (como el tipo, número de máquinas, la edad, la disposición de las máquinas respecto a la otra, la disposición de componentes en la máquina, defectos inherentes en los componentes), *las condiciones de funcionamiento*

1. INTRODUCCIÓN

(esfuerzo nominal o sobreesfuerzo, temperaturas variables, cargas inesperadas), *el factor humano* (nivel de habilidad y número de personal operativo, los hábitos de trabajo, las relaciones inter-personales, las ausencias, las medidas de seguridad, las condiciones ambientales, la peligrosidad de las tareas asignadas, los incidentes/accidentes durante las labores), *condiciones de mantenimiento* (la competencia y habilidad del personal de mantenimiento, la asistencia, hábitos del trabajo, medidas de seguridad, las relaciones inter-personales, defectos introducidos por las acciones de mantenimiento anteriores, la eficacia de la planificación y control de mantenimiento), e *infra-estructura* (repuestos, consumibles, herramientas comunes y especiales), entre otras.

El interés en $I - RAMS$ por el descubrimiento de patrones crece constantemente en la industria debido a los continuos avances tecnológicos en el desarrollo de sensores. El incremento de la fiabilidad de los sensores permite monitorear continuamente el estado de funcionamiento de los componentes operativos, pero a cambio se ha de gestionar una gran cantidad de datos, lo que favorece a su vez el interés por técnicas de análisis basadas directamente en ellos. Los enfoques basados en datos se pueden dividir en dos categorías: Técnicas Estadísticas (Métodos de Regresión, modelos ARIMA, etc.) y técnicas de Inteligencia Artificial (IA: Redes Neuronales, Lógica Difusa, etc.). Son de especial interés los modelos basados en datos para la estimación de la Vida útil Remanente (VR) que han de tener en cuenta los patrones de degradación presentes y normalmente ocultos en los datos.

Entre las muchas vertientes $I - RAMS$ son de mención especial aquellas que tienen que ver con la seguridad de las personas, la naturaleza y el medio ambiente. Todas ellas bajo los descriptores globales "Safety/Security". Uno de ellos es bien conocido por su alto impacto ambiental, económico y social: Los incendios forestales. El interés en la predicción del comportamiento de los incendios forestales y la evaluación del peligro, se ha incrementado a lo largo del siglo 20 y la ciencia de la modelización de los incendios ha hecho grandes avances. La predicción del comportamiento y la evaluación del peligro de los incendios forestales son elementos importantes en la gestión forestal, la extinción de incendios forestales, la planificación de emergencia, así como la predicción y evaluación de los efectos ecológicos. Existen muchos sistemas disponibles a nivel mundial para evaluar y controlar los incendios, basados en modelos matemáticos y físicos. La principal fuente de errores en las predicciones de estos modelos para el cálculo y la evaluación del comportamiento de los incendios forestales es la falta de la aplicabilidad, inexactitud interna y errores de entrada en los datos (7, 8, 9, 10, 11). En algunos casos, la mayoría de los parámetros de entrada para los modelos no suelen estar disponibles. Con frecuencia, el cálculo

de un promedio a partir de datos anteriores, la asimilación de la media con otras temporadas, o la comparación de la media de años anteriores u otras estaciones meteorológicas cercanas, son tareas que requieren tiempo, y búsqueda detallada en los datos históricos. Teniendo en cuenta que el tiempo es limitado para derivar decisiones en condiciones de alta actividad en tiempo real. Además, otra fuente de error dominante en general en las predicciones del modelo es la complejidad del combustible debido a la no uniformidad, la discontinuidad, la heterogeneidad y la presencia de múltiples capas en los bosques. Por no hablar de la alta no linealidad del viento y las condiciones metrológicas, que aumentan la no linealidad del incendio forestal.

Debido a que los modelos analíticos se vuelven muy complejos para el análisis de la interacción entre los diferentes factores que afectan a estos retos, ha surgido un gran interés en nuevos enfoques, entre los que cabe citar el uso de determinadas metodologías de la Inteligencia Artificial (Lógica Difusa, Redes Neuronales Artificiales, Algoritmos Genéticos, Máquinas de Soporte Vectorial) y de los Métodos Evolutivos de Optimización (Algoritmos Genéticos, Programación Genética, Recocido Simulado y Búsqueda Tabú, entre otros), representan metodologías clave muy importantes, que han ido resolviendo problemas *I – RAMS* complejos que incluyen el apoyo a la toma de decisiones en presencia de incertidumbre.

Recientemente, las Redes Neuronales Artificiales - RNA son objeto de atención creciente en el ámbito *I – RAMS* por parte de la industria y los negocios basados en los nuevos paradigmas, antes citados, de interconexión y comunicación, lo que por ende ha contribuido a estimular los avances científicos. Las RNA representan una técnica ampliamente basada en los datos, cuyas bondades frente a los métodos tradicionales de predicción de series temporales han sido ampliamente estudiadas y suficientemente contrastadas (12, 13, 14). La esencia del potencial de las RNA radica es su naturaleza flexible y su capacidad para capturar relaciones no lineales complejas entre los patrones de entrada y salida, usando para ello un aprendizaje apropiado aplicado a un conjunto de datos representativos que debe contener información suficiente del problema a tratar. Las RNA se basan en algoritmos sencillos capaces de realizar asignaciones funcionales entre las entradas y las salidas. Su característica de paralelismo, su carácter multi-paramétrico y sus requerimientos mínimos de almacenamiento, las convierten en una poderosa herramienta computacional, especialmente cuando los modelos físico-matemáticos subyacentes son complejos. Entre otros aspectos estas características justifican el hecho de que las RNA puedan ser utilizadas para modelar ciertos sistemas complejos cuyas descripciones matemáticas no están disponibles.

1. INTRODUCCIÓN

El diseño de las RNA se inspira en los sistemas neuronales biológicos, constando esencialmente de una colección de nodos de procesamiento (neuronas) interconectados entre sí. Las conexiones entre las neuronas definen la estructura de la red, conocida como arquitectura de la red neural. Es típica una arquitectura basada en capas de neuronas (Capa de Entrada, Capa(s) Intermedia(s), Capa de Salida) así como la existencia de mecanismos llamados algoritmos de entrenamiento, cuya función es ajustar los pesos de la red neuronal, estos pesos son coeficientes de las conexiones entre las neuronas. En los años 70's, Werbos originó el algoritmo de retro-propagación, que se usó en un modelo neuronal dinámico para predecir dos variables (15). Sin embargo, su gran trabajo ha sido casi desconocido para la comunidad científica hasta los 80's, cuando ha sido redescubierto por Packer y Rumelhart (16, 17). La investigación sobre las RNA ha permitido constatar (18, 19, 20, 21) algunas importantes características de las mismas que más tarde se citarán en la revisión del estado del arte, estando entre ellas las siguientes:

- Las RNA ofrecen soluciones viables debido a su capacidad de relacionar simultáneamente múltiples variables cuantitativas y cualitativas, así como su capacidad de formar modelos basados únicamente en datos mínimos representativos, en lugar de los supuestos de linealidad u otras relaciones analíticas estáticas.
- Las RNA son reconocidas como *aproximadores universales* de cualquier tipo de función dentro de un intervalo compacto, de acuerdo con el Teorema de Stone-Weierstrass (22). Por ejemplo una red neuronal con retro-propagación y una sola capa oculta es capaz de proporcionar aproximaciones precisas a muchas funciones complejas, siempre que haya suficientes neuronas ocultas (23), lo que posibilita una fácil implementación.

1.2 Resumen de contribuciones de esta tesis

Los trabajos y resultados descritos en este documento se inscriben en la línea de investigación en *I – RAMS* de la División de Computación Evolutiva y Aplicaciones (CEANI) del Instituto de Sistemas Inteligentes y Aplicaciones Numéricas en Ingeniería (SIANI), que pertenece a la Universidad de Las Palmas de Gran Canaria (ULPGC). Más explícitamente las contribuciones versan sobre aspectos novedosos que afrontan retos *I – RAMS* mediante Redes Neuronales Artificiales (RNA). Aunque serán descritas con detalle en capítulos posteriores, se presenta en esta sección un resumen de las principales contribuciones de esta tesis.

La primera contribución de esta tesis, investiga las etapas de preparar una red neuronal para su uso. Se aporta un método nuevo determinista de álgebra lineal para la inicialización

de los pesos de una RNA, este método se justifica con una base científica sólida, cumpliendo los requisitos del teorema de convergencia. También, se analizan las funciones de activación, estudiando el comportamiento de minimización del error. Por otra parte, se crea un modelo aleatorio para la selección de la estructura de la red, seguido por otro modelo que calcula la eficiencia de la red neuronal una vez entrenada, este modelo calcula dos métricas de confianza, la primera en base al conjunto de entrenamiento, y la otra en base a los pesos de la red neuronal, por lo tanto el resultado tiene en cuenta la información presente en el conjunto de entrenamiento y la información oculta en la red neuronal.

La segunda contribución de esta tesis, explora la utilidad de las RNA como técnica para resolver nuevos problemas complejos *I – RAMS* no satisfactoriamente resueltos hasta el momento debido a la inexistencia de información física o matemática, esos problemas son asociables a series temporales no lineales pero de naturaleza puramente industrial. En concreto, se aborda el problema de encontrar la curva de Vida Residual (VR) solamente con los datos disponibles, aplicando los resultados obtenidos del primer caso de estudio. Se aporta un criterio con base científica para la solución del problema mediante una combinación de métricas de eficiencia y de intervalos de confianza.

La tercera contribución analiza en profundidad la aplicación de las RNA como técnica RAMS en el control de fenómenos naturales, como es el caso de control de incendios forestales. En concreto propone crear un nuevo modelo neuronal llamado ANNFBP “Artificial Neural Network for Fire Behaviour Parameters”, analizando su capacidad para predecir algunos parámetros del comportamiento del incendio, utilizando todas las entradas optimizando los factores y por ende los cálculos, sin perder precisión en las predicciones. Además, se estudia con criterio científico las ventajas y desventajas de la utilización de este modelo neuronal como herramienta *I – RAMS* para la gestión y el control de los incendios forestales.

La cuarta contribución de esta tesis, trata sobre la estimación de la demanda de energía eléctrica. Los cambios sociales y las incertidumbres económicas, como las derivadas de la última crisis económica, han convertido la estimación futura de la energía eléctrica en un problema no lineal y ha despertado el interés de los investigadores en buscar alternativas a los modelos tradicionales que llevan décadas de uso, para afrontar estas nuevas dificultades. Por lo tanto se ha decidido de realizar una comparativa entre métodos de distinta naturaleza, y no obstante aplicar medidas de fiabilidad al modelo neuronal presentado en esta memoria, ofreciendo seguridad y precisión.

1. INTRODUCCIÓN

1.3 Publicaciones relacionadas con la línea de investigación

Algunas de las contribuciones de esta tesis han sido publicadas en un libro y en una revista científica. Además han sido presentadas en congresos y jornadas:

1.3.1 Proyectos

- Participación como miembro del equipo de investigación CEANI de la ULPGC, en el proyecto: “Logistic Efficiencies And Naval architecture for Wind Installations with Novel Developments – LEANWIND, ref 614020”, contribuyendo a la investigación de las Aplicaciones de las RNA en Diagnóstico y Pronóstico de averías en Aerogeneradores. (13 de marzo del 2014- Actualidad)
- Premio de Programa Innova Canarias 2020, para realizar el proyecto: “Uso de la Inteligencia Artificial para el Diagnostico, Fiabilidad de una Central Eléctrica en Canarias”. (01/01/2013-31/12/2013)
- Participación como becario de investigación en el proyecto CENIT Líderes en Energías Renovables Oceánicas OCEAN LIDER. Investigando y aplicando las redes neuronales para el Diagnóstico/Pronostico de los aerogeneradores. (01/02/2011 - 31/10/2012)
- Colaboración como personal investigador de apoyo a la división “CEANI” Instituto de Sistemas Inteligentes y Aplicaciones Numéricas en Ingeniera “SIANI” de la Universidad de Las Palmas de Gran Canaria “ULPGC”. (01/10/2010- Actualidad).

1.3.2 Publicaciones

- Winter-Althaus, G., Gonzalez-Landin, B., Pulido-Alonso, A., Galvan-Gonzalez, B., Maarouf, M. (2015) “Long Terms Predictions Of Electricity Demand: A Challenge For Computer Engineering”. DYNA DOI: <http://dx.doi.org/10.6036/7834> .
- Winter-Althaus, G., Gonzalez-Landin, B., Pulido-Alonso, A., Galvan-Gonzalez, B., Maarouf, M.. (2015). ”Forecasting Electricity Consumption Including Data Of The Present Economic And Financial Crisis”. Application To Canary Islands. Dyna Energía y Sostenibilidad, 4(1). 1-13. DOI: <http://dx.doi.org/10.6036/ES7782>
- M. Maarouf, B. Galván. ”Prediction of Wildfire Behaviour Parameters using an Artificial Neural Network Model”. International Journal of Wildland Fire (JCR: 2.429, Estado: revisiones menores con fecha 22-08-2015).

1.3 Publicaciones relacionadas con la línea de investigación

- M. Maarouf, B. Galván. (2015) “A weight initialization method for improving training speed in feedforward neural network”. Artificial Intelligence. Elsevier Editorial. Estado: Bajo revisión desde 15-01-2015)
- A. Pulido, G. Winter, B. Gonzalez-Landin B., Galvan-Gonzalez B., Maarouf M.. “Analysis of the impact of the meteorological, tourist-related and economic variables on monthly electricity demand: Gran Canaria Island (Spain)”. Journal of Energy in Southern Africa , Estado: Bajo revision con fecha 30-07-2015)

1.3.3 Capítulos de libros

- The Role of Artificial Neural Networks in Evolutionary Optimisation: A Review. M. Maarouf, A. Sosa, B. Galvn, D. Greiner, G. Winter, M. Mendez, R. Aguasca. Extended Abstracts Book: Evolutionary and Deterministic Methods for Design, Optimization and Control with Applications to Industrial and Societal Problems. Universidad de Las Palmas de Gran Canaria. EUROGEN 2013. October 7-9, 2013. pp 163-168. ISBN 978-84-616-6249-4
- Physical Activity Classification Using Resilient Backpropagation (RPROP) with Multiple Outputs. Mustapha Maarouf, Blas J. Galván González. Computer Aided Systems Theory - EUROCAST 2013. Lecture Notes in Computer Science Volume 8111, 2013, pp 77-83

1.3.4 Congresos

- The Role of Artificial Neural Networks in Evolutionary Optimisation: A Review. M. Maarouf, A. Sosa, B. Galvn, D. Greiner, G. Winter, M. Mendez, R. Aguasca. Extended Abstracts Book: Evolutionary and Deterministic Methods for Design, Optimization and Control with Applications to Industrial and Societal Problems. Universidad de Las Palmas de Gran Canaria. EUROGEN 2013. October 7-9, 2013
- Physical Activity Classification Using Resilient Backpropagation (RPROP) with Multiple Outputs. Mustapha Maarouf, Blas J. Galván González. Computer Aided Systems Theory - EUROCAST 2013. (Ponencia)
- Uso de Inteligencia Artificial en Diagnosis Remota de activos industriales: Caso de Aerogeneradores. Blas Galván, Mustapha Maarouf, Silvia Alonso, Juan Pedro Ramos. XV

1. INTRODUCCIÓN

Seminario últimos Avances en Informática UAI 2011. Universidad de La Laguna ULL. Tenerife. Septiembre 2011 (Ponencia).

- Identité numérique et Apprentissage en Réseau: Application PhD-ePortfolio. Mustapha Maarouf, Enrique Rubio Royo, Manuel Galán Moreno, Ginés Delgado Cejudo. Open Digital Space for the Mediterranean Conference. March 26-29 Agadir, Morocco. (Ponencia)

1.3.5 Organización de Esta Memoria

En el **capítulo 2**, lo más importante a citar aquí es que se realiza una completa revisión del estado del arte de las RNA en lo que pueda ser de interés para las aplicaciones $I - RAMS$ de esta tesis, además se definen los conceptos básicos y necesarios para tener una conceptualización general acerca de las RNA, tanto los principios de convergencia, los intervalos de confianza. En el **capítulo 3** se presentan los aspectos teóricos, metodológicos y los análisis realizados como contribuciones originales de esta tesis, ya descritos anteriormente en resumencontri. En el **capítulo 4** se presentan los resultados de la experimentación y aplicación de las aportaciones. Finalmente, en el **capítulo 5** se presentan las conclusiones generales de esta tesis doctoral y las futuras líneas de investigación.

CAPÍTULO 2

ESTADO DEL ARTE

2. ESTADO DEL ARTE

2.1 Redes Neuronales Artificiales “RNAs”

Una red neuronal es una herramienta potente de modelización de datos, que es capaz de captar y representar relaciones complejas entre entradas y salidas. La motivación para el desarrollo de la tecnología de las redes neuronales surgió del deseo de desarrollar un sistema artificial que pudiera realizar tareas “*inteligentes*” similares a las tareas realizadas por el cerebro humano. Las redes neuronales se asemejan al cerebro humano en dos aspectos:

- Una red neuronal adquiere conocimiento a través del aprendizaje.
- El conocimiento de una red neuronal se almacena dentro de las conexiones entre las neuronas conocidas como pesos sinápticos.

La potencia y la utilidad de las Redes Neuronales Artificiales se ha demostrado en múltiples aplicaciones, incluyendo la síntesis del habla, problemas de diagnóstico, medicina, negocios y finanzas, control de robótica, procesamiento de señales, visión artificial y muchos otros problemas que se incluyen en la categoría de reconocimiento de patrones. En algunas áreas de aplicación, se espera que los modelos neuronales logren en un futuro no muy lejano rendimientos similares a los del ser humano.

2.1.1 Historia de las Redes Neuronales

A continuación, se describe el desarrollo histórico de las RNA:

- McCulloch y Pitts (1943) desarrollaron modelos de redes neuronales basados en su conocimiento de la neurología y en supuestos acerca de cómo trabajan las neuronas. Sus redes se basaban en neuronas simples, que eran consideradas como dispositivos binarios con umbrales fijos. Los resultados fueron funciones lógicas simples, tales como “ $A \text{ o } B$ ” y “ $A \text{ y } B$ ”. Otro intento fue mediante el uso de simulaciones por ordenador.
- Rosenblatt (1958) despertó un interés y una actividad considerable en el campo cuando diseñó y desarrolló el perceptrón. El perceptrón tenía tres capas, con la capa intermedia conocida como la capa de asociación.

Otro sistema era el Adaline (Elemento Lineal Adaptativo “ADaptive LINEar Element”), que fue desarrollado en 1960 por Widrow y Hoff (en la Universidad de Stanford). El Adaline era un dispositivo electrónico analógico hecho de componentes simples. El método utilizado para el aprendizaje no solo era diferente al del perceptrón, sino que empleó la media de los Mínimos Cuadrados (LMS) como regla de aprendizaje.

- El período de frustración y desprestigio: en 1969, Minsky y Papert, escribieron un libro libro (24) en que se generalizan las limitaciones de los modelos del perceptrón con una sola capa y sistemas multicapas. Las conclusiones acentuaron el desencanto de los investigadores en el campo, causando un perjuicio considerable a las líneas de investigación y desarrollo.
- La era de Innovación: aunque el interés público y los fondos disponibles eran mínimos, varios investigadores continuaron trabajando en el desarrollo de la neurociencia basándose en métodos computacionales para problemas como el reconocimiento de patrones. Durante este período se generaron varios paradigmas que mejoraron el uso de las RNA. Anderson y Kohonen desarrollaron técnicas asociativas independientes. Henry Klopff en 1972 desarrolló una base para el aprendizaje en las neuronas artificiales fundamentada en la heterostasis, un principio biológico para el aprendizaje neuronal. En 1974, Paul Werbos desarrolló y utilizó el método de aprendizaje de retro-propagación, sin embargo, pasaron varios años antes de que este enfoque se popularizase (15). En 1988, Steve Grossberg y Carpenter Gail fundaron una escuela de pensamiento que explora algoritmos de resonancia. Ellos desarrollaron la Teoría de Resonancia Adaptativa como un arte de redes basadas en modelos biológicamente plausibles (25). Las redes de retro-propagación son probablemente las más conocidas y las más ampliamente aplicadas en las redes neuronales de hoy en día. En esencia, la red de propagación hacia atrás, es un perceptrón con múltiples capas, una función de umbral diferente en la neurona artificial, y una regla de aprendizaje más robusta. En 1976, Amari Shun-Ichi un investigador esencialmente dedicado a desarrollos teóricos, publicó un documento en el que estableció una teoría matemática como base de aprendizaje (método de corrección de errores), para la clasificación de patrones. Mientras, Fukushima en (26), dio un paso más entrenando una red neuronal multicapa para la interpretación de los caracteres escritos a mano. La red original se publicó en 1975 y fue llamada Cognitron (27, 28, 29).
- Los avances durante la década de los 70 y a principios de los 80 fueron importantes para la reaparición del interés en el campo de las RNA. Varios factores influyeron en este cambio de tendencia. Por un lado los libros publicados contribuyeron a difundir ampliamente los avances, mientras que por otro lado las conferencias globales proporcionaron foros técnicos de discusión para los especialistas. Siendo amplia y positiva la respuesta y acogida ambos tipos de actividades. Los medios de comunicación recogieron el aumento de la actividad y

2. ESTADO DEL ARTE

crearon tutoriales ayudando a difundir la tecnología. Aparecieron programas académicos y cursos que se introdujeron en la mayoría de las principales universidades (en EE.UU. y Europa).

- Hoy en día se ha hecho un progreso significativo en el campo de las RNA, suficiente como para atraer una gran atención de organismos y entidades, lo que permite financiar la investigación en este campo en muchos frentes. Como fruto de ello están surgiendo tanto conceptos neuronales como aplicaciones nuevas para resolver problemas complejos. Existen evidencias razonables, a favor de una percepción generalizada en muchos investigadores, de que en la actualidad se está en un periodo de transición hacia modelos neuronales mucho más avanzados.

2.1.2 Ventajas de las Redes Neuronales Artificiales

Muchas son las técnicas informáticas que pueden utilizar las redes neuronales artificiales, con su notable capacidad de entender el significado / naturaleza de los datos complejos o imprecisos, para extraer patrones y detectar tendencias de difícil detección. Las ventajas incluyen:

- Aprendizaje adaptativo: una habilidad para aprender a hacer las tareas basandose en los datos proporcionados para el entrenamiento o la experiencia inicial.
- Auto-organización: una RNA puede crear su propia organización, o representación, de la información que reciba durante el proceso del aprendizaje.
- Operaciones en Tiempo Real: los cálculos de las RNA pueden llevarse a cabo en paralelo, y en la actualidad se diseñan y fabrican dispositivos hardware especiales para aprovechar esta capacidad.
- Tolerancia de los fallos a través de la codificación de informaciones redundantes: la destrucción parcial de una red conduce a la correspondiente degradación del rendimiento. Sin embargo, algunas capacidades de la red pueden ser retenidas incluso con daños en la red principal.

2.1.3 Las Redes Neuronales Artificiales vs. los Equipos Convencionales

Las redes neuronales artificiales adoptan un enfoque diferente que el proporcionado por los ordenadores convencionales, para la solución de problemas, y se debe a que:

- Los ordenadores convencionales utilizan una aproximación algorítmica, es decir, el ordenador sigue un conjunto de instrucciones con el fin de resolver un problema, pero solo podrá hacerlo si son conocidos los datos específicos que el algoritmo necesita. Eso limita la capacidad de generalización. No obstante es evidente que los ordenadores serían mucho más útiles si pudiesen resolver problemas para los cuales se desconoce la aproximación algorítmica que permitiría resolverlos.
- Las redes neuronales por otra parte, procesan la información de una manera similar al cerebro humano. La red está compuesta de un gran número de elementos de procesamiento altamente interconectados (neuronas) que trabajan en paralelo para resolver un problema específico. Las redes neuronales aprenden con los ejemplos.
- La red neuronal puede descubrir cómo resolver un problema por sí misma, su funcionamiento puede llegar a ser impredecible. Mientras que los ordenadores convencionales utilizan un enfoque cognitivo totalmente predecible y si algo sale mal se debe a un fallo de hardware o software.

Las redes neuronales y los algoritmos convencionales no compiten, sino que se complementan entre sí. Hay tareas que son más apropiadas para un enfoque algorítmico, como las operaciones aritméticas, y tareas que son más adecuadas para un enfoque neuronal. Más aún, un gran número de tareas requieren de sistemas que utilizan una combinación de los dos enfoques (normalmente un ordenador convencional se utiliza para supervisar la red neural) para llevarlas a cabo con la máxima eficacia.

2.1.4 Las predicciones mediante las RNA

Predecir lo que puede suceder en el futuro siempre ha sido visto como una actividad misteriosa, que los científicos tratan de convertir en una actividad científica basada en teorías y modelos. En la sociedad moderna, la predicción se puede utilizar en muchos problemas del mundo real para poner a prueba nuestra comprensión científica del comportamiento de una variedad de sistemas o fenómenos complejos. También se puede utilizar la predicción como una guía potencial o base para la toma de decisiones, especialmente en la prevención de catástrofes y / o en la mitigación del alcance de sus consecuencias.

En la actualidad, la predicción científica es utilizada para abordar varios problemas difíciles como lo son la mayoría de problemas existentes en el mundo real, en una variedad de campos y aplicaciones tales como las previsiones financieras y las predicciones del medio ambiente.

2. ESTADO DEL ARTE

2.1.5 Presentación General, Definiciones y Teoremas

En esta sección se presentan los conceptos, fundamentos, teoremas y definiciones que son importantes a la hora de entender y aplicar las Redes Neuronales Artificiales.

2.1.5.1 Definiciones y Teoremas

Definición 1 La distancia de Hamming se define como:

$$\rho(a, b) = \frac{1}{2} \sum_{i=1}^n |a_i - b_i|$$

Teorema 1 Sea $\{x^{(1)}, x^{(2)}, \dots, x^{(p)}\}$ un conjunto aleatorio de patrones $x^{(i)} \in \{-1, 1\}^n$ bipolares. Si $x \in \{-1, 1\}^n$ se encuentra a una distancia de Hamming de $\frac{n}{2p}$ de un patrón $x^{(m)}$, es decir:

$$\rho(x, x^{(m)}) = \frac{n}{2p}$$

Entonces: $x^{(m)}$ es el patrón más cercano a x . $\forall i \in \{1, 2, \dots, p\} - \{m\} : \rho(x, x^{(i)}) \geq \frac{n}{2p}$

La condición $\frac{n}{2p}$ es suficiente para la convergencia, pero no es necesaria.

Definición 2 La matriz de correlación entre $\{x^{(1)}, x^{(2)}, \dots, x^{(n)}\}$ y $\{y^{(1)}, y^{(2)}, \dots, y^{(n)}\}$ $y_i = \sum_{j=1}^n M_{ij} x_i$, y una matriz $M = \sum_{m=1}^p y^{(m)} x^{(m)T}$

$$\Rightarrow M_{ij} = \sum_{m=1}^p (y^{(m)} x^{(m)T})_{ij} = \sum_{m=1}^p y_i^{(m)} x_j^{(m)T}$$

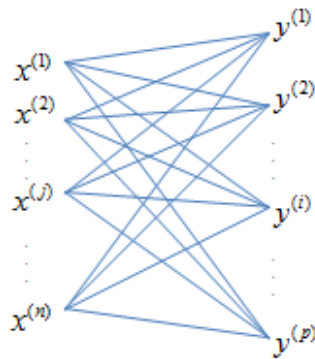


Fig. 2.1: Matriz de memoria de correlación

La cuestión es poder encontrar una matriz $M \in \mathbb{R}^{n \times m}$ de tal manera que: $Ma^{(i)} = b^{(i)}, \forall i \in \{1, 2, \dots, p\}$, es decir $Ma = b$.

Proposición 1 *La tres propiedades siguientes son equivalentes:*

$\forall A \in \mathbb{R}^{n \times p}, A^T A \in \mathbb{R}^{p \times p}, A$ es invertible

$$\Leftrightarrow \left[A = \begin{pmatrix} a^{(1)} & a^{(2)} & \dots & a^{(p)} \end{pmatrix}, v_1 a^{(1)} + v_2 a^{(2)} + \dots + v_p a^{(p)} = 0 \Leftrightarrow v_1 = v_2 = \dots = v_p \right]$$

$\Leftrightarrow A$ un vector linealmente independiente en \mathbb{R}^n

Teorema 2 *Sea $A \in \mathbb{R}^{n \times p}$, linealmente independiente. Entonces:*

$\forall B \in \mathbb{R}^{m \times p} : \exists M \in \mathbb{M}(n \times m)$ de tal manera que: $MA = B$

Ejemplo 1 *Se puede considerar por ejemplo la matriz como: $M = B(A^T A)^{-1} A^T$*

Definición 3 *Sea X es la matriz inversa generalizada de A*

$$X \text{ es la matriz inversa generalizada de } A \Leftrightarrow \begin{cases} AXA = A \\ XAX = X \\ (AX)^T = AX \\ (XA)^T = XA \end{cases}$$

Se suele anotar la matriz inversa generalizada como: $A^\#$

Proposición 2 $\forall A \in \mathbb{R}^{m \times n}, AA^\#$ es la proyección ortogonal de A si:

$$AA^\# = (AA^\#)^T = (AA^\#)^2$$

Nota 1 *Si $\text{rang}(A) = n \Rightarrow A^\# = (A^T A)^{-1} A^T$ Si $\text{rang}(A) = m \Rightarrow A^\# = A^T (AA^T)^{-1}$*

Definición 4 $\|\cdot\|_F$ *F-norm en $\mathbb{R}^{m \times n}$:*

$$\begin{aligned} \|A\|_F^2 &= \text{Tr}(A^T A) \\ &= \sum_{i=1}^n (A^T A)_{ii} \\ &= \sum_{j=1}^m \sum_{i=1}^n A_{ij}^T A_{ji} \\ &= \sum_{j=1}^m \sum_{i=1}^n A_{ij}^2 \end{aligned}$$

Por lo tanto se puede obtener:

$$\text{Tr}(AB) = \text{Tr}(BA) \text{ y } \text{Tr}(A^T) = \text{Tr}(A),$$

$$\text{Tr}(A^T B) = \text{Tr}(AB^T) = \text{Tr}(BA^T) = \text{Tr}(B^T A)$$

Teorema 3 *Sea $A \in \mathbb{R}^{n \times p}, B \in \mathbb{R}^{m \times p}$:*

$$X = BA^\# \in \mathbb{R}^{m \times n} = \min \|XA - B\|_F$$

Cualquier matriz X que satisface: $X = BA^\# B$, se considera como una solución mínima.

2. ESTADO DEL ARTE

2.1.5.2 El Adaline

Considerando un caso especial; si la salida deseada es un vector $y^{(1)}, y^{(2)}, \dots, y^{(p)} \in \mathbb{R}$, entonces para $x^{(1)}, x^{(2)}, \dots, x^{(p)} \in \mathbb{R}^l$, se busca una matriz de memoria $M \in \mathbb{R}^l$ que satisfice $Mx = y$ es decir:

$$y = \sum_{i=1}^l m_i x_i, \text{ siendo } M = (m_1, \dots, m_l) \quad (2.1)$$

En la figura 2.2, se representa el sistema conocido por ALC “Adaptive Linear Combiner”. En

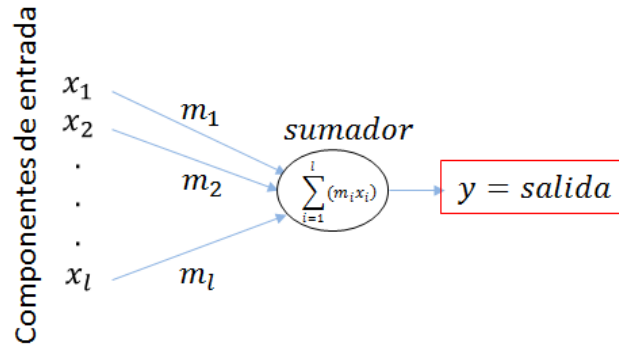


Fig. 2.2: ALC

los casos reales, no es tan fácil encontrar esa matriz de memoria M , que satisfice la ecuación 2.1.

La idea es encontrar u optimizar en cierto sentido M . Para ello un primer enfoque es mediante

el cálculo del error cuadrático: $\varepsilon = \frac{1}{2p} \sum_{i=1}^p |y^{(i)} - z^{(i)}|^2$

Minimizando el error (la diferencia) entre el valor deseado y el valor obtenido de la salida.

Recordando que si $A = (x_1, x_2, \dots, x_m)$ constituyen un conjunto ortonormal, entonces A tiene columnas independientes, por lo tanto AA^T tiene inversa, y se puede definir: $A^\# = (A^T A)^{-1} A^T = A^T$, i.e $BA^\# = BA^T = M$

El caso general sería: $M = BA^\# = \min \|B - MA\|_F^2$

Pero, se va a desarrollar un algoritmo para construir la matriz de memoria M

Como: $\varepsilon = \frac{1}{2p} \sum_{i=1}^p |y^{(i)} - z^{(i)}|^2$

$$\begin{aligned}
 \varepsilon &= \frac{1}{2p} \sum_{i=1}^p \left| y^{(i)} - \sum_{j=1}^l m_j x_j^{(i)} \right|^2 \\
 \varepsilon &= \frac{1}{2} \left[\frac{1}{p} \sum_{i=1}^p y^{(i)2} + \sum_{i,j,k}^{l,p} m_j x_j^{(i)} x_k^{(i)} m_k - 2 \sum_{i,j,k}^{l,p} y^{(i)} m_j x_j^{(i)} \right] \\
 \varepsilon &= \frac{1}{2} \left[\sum_{j,k}^p m_j A_{jk} m_k - \sum_{j=1}^p b_j m_j + c \right]
 \end{aligned}$$

Donde: A es simétrica,

$$\begin{aligned}
 A_{jk} &= \sum_{i=1}^p x_j^{(i)} x_k^{(i)} \\
 b_j &= \frac{1}{p} \sum_{i=1}^p y^{(i)} x_j^{(i)} \\
 c &= \frac{1}{p} \sum_{i=1}^p y^{(i)2}
 \end{aligned}$$

Por lo tanto:

$$\frac{d\varepsilon}{dm_i} = 0 \Leftrightarrow \sum_{k=1}^l A_{ik} m_k = b_i, \forall i \in \{1, 2, \dots, l\}$$

$$Am = b$$

$\sum_{k=1}^l A_{ik} m_k = b_i, \forall i \in \{1, 2, \dots, l\}$ se llama ecuación de Weiner-Hopf o la ecuación de Gauss-normal.

Si al escribir que: $\varepsilon = \frac{1}{2p} \|MX - Y\|_F^2$

Según el teorema 3, $M = YX^\#$ es un mínimo.

$$A = \frac{1}{p} XX^T, b^T = \frac{1}{p} YX^T, C = \frac{1}{p} Y^T Y$$

Si además A es invertible: $M = (XX^T)^{-1}XY^T \Rightarrow M^T = YX^\#$

Otro enfoque para encontrar $m^* = \min \varepsilon$, es mediante el gradiente descendente.

La idea es mover m a una distancia más pequeña $m - \alpha$ del gradiente de ε , repitiendo el proceso, se espera llegar al mínimo de ε .

La estrategia que se emplea trata de considerar una serie de vectores $m(n)$ que satisfacen:

$$m(n+1) = m(n) - \alpha \text{grad}(\varepsilon), n = 1, 2, \dots$$

$m(1)$ es arbitrario, α es la tasa del aprendizaje.

2. ESTADO DEL ARTE

Entonces: $m(n+1) = m(n) - \alpha(b - Am(n))$

Y como A es simétrica, por lo tanto puede ser diagonalizada: $UA^T U = D = \text{diag}(\lambda_1, \dots, \lambda_l)$

Asumiendo que: $\lambda_1 \geq \dots \geq \lambda_l$ $\varepsilon = \frac{1}{2}m^T Am - b^T m + \frac{1}{2}c$

Entonces: $\varepsilon = \frac{1}{2}m^T U^T U A U^T U m - b^T U^T U m + \frac{1}{2}c$

$\varepsilon = \frac{1}{2}Z^T D Z - v^T Z + \frac{1}{2}c$

$\varepsilon = \frac{1}{2} \sum_{i=1}^l \lambda_i z_i^2 - \sum_{i=1}^l v_i z_i + \frac{1}{2}c$

Donde: $Z = Um, v = Ub$

Así que: $m(n+1) = m(n) + \alpha(b - Am(n))$

Por lo tanto:

$$z(n+1) = z(n) + \alpha(v - Dz(n))$$

$$\Rightarrow z_j(n+1) = (1 - \alpha\lambda_j)z_j(n) + \alpha v_j$$

$$= \mu_j^n z_j(1) + (\mu_j^{n-1} + \mu_j^{n-2} + \dots + \mu_j + 1)\alpha v_j$$

Donde:

$$\mu_j = 1 - \alpha\lambda_j$$

Entonces si:

$$|\mu_j| < 1, (\text{es decir: } -1 < 1 - \alpha\lambda_j < 1 \Rightarrow 0 < \alpha\lambda_j < 2\forall j)$$

Por lo tanto si: $0 < \alpha < \frac{2}{\lambda_j}, \forall j \Rightarrow 0 < \alpha < \frac{2}{\lambda_{\max}}$, donde λ_{\max} es el valor propio máximo de A :

$$A_{jk} = \frac{1}{p} \sum_{i=1}^p x_j^{(i)} x_k^{(i)}$$

Existe un mínimo.

$\Leftarrow?$ (el otro sentido)

Si a partir de $m(1)$ arbitrario: $m^* = \lim_{n \rightarrow \infty} m(n)$, cuando $0 < \alpha < \frac{2}{\lambda_{\max}}$

Se obtiene: $m^* = m^* + \alpha(b - Am^*) \Rightarrow Am^* = b$

Lo que demuestra que cumple y satisface la minimización de ε .

Todo este análisis, requiere un conocimiento detallado de A . Porque al saber sus valores propios, nos permite elegir α

El enfoque anterior, supone estudiar la matriz A , ahora se aborda el problema desde otro enfoque:

Sabiendo que A está definida por: $A_{jk} = \frac{1}{p} \sum_{i=1}^p x_j^{(i)} x_k^{(i)}$

Y también b como: $b_j = \frac{1}{p} \sum_{i=1}^p x_j^{(i)} y^{(i)}$

Se sabe que: $m(n+1) = m(n) + \alpha(b - Am(n))$

$$m_j(n+1) = m_j(n) + \alpha(b - Am_j(n))$$

$$m_j(n+1) = m_j(n) + \alpha(y^{(i)} x_j^{(i)} - \sum_{i=1}^n x_j^{(i)} x_k^{(i)} m_k(n)) m_j(n+1)$$

$$m_j(n+1) = m_j(n) + \alpha(y^{(i)} - \sum_{k=1}^n x_k^{(i)} m_k(n)) x_j^{(i)}$$

Se nota: $\delta^{(i)} = y^{(i)} - \sum_{k=1}^n x_k^{(i)} m_k(n)$

Por lo tanto: $m_j(n+1) = m_j(n) + \alpha \delta^{(i)} x_j^{(i)}$

$\delta^{(i)}$ se reconoce como “delta-rule”, “Windrow-hoff learning rule”, o “Least Mean Square Algorithm” (LMS).

Etapas del algoritmo “LMS”:

1. Escoger α arbitrario entre 0 y 1, en general se suele escoger $\alpha = 0,1$ o $\alpha = 0,05$
2. Empezar con $m_j(1) = 0, \forall 1 \leq j \leq l$
3. $m_j(n+1) = m_j(n) + \alpha \delta^{(i)} x_j^{(i)}, \forall 1 \leq j \leq l$, donde: $\delta^{(i)} = y^{(i)} - \sum_{k=1}^l x_k^{(i)} m_k(n)$
4. Ir constituyendo un ciclo de valores: $(x_k^{(1)}, y^{(1)}), \dots, (x_k^{(p)}, y^{(p)})$
La cuestión de la convergencia queda pendiente discutirla.

Observación

Si α es muy pequeño, \Rightarrow convergencia lenta.

Si se asume que hay convergencia, δ es pequeño, m no será pequeña, \Rightarrow convergencia lenta.

Si α es muy grande, m puede pasar la solución óptima.

En la práctica, es mejor empezar con α grandes e ir disminuyendo α gradualmente durante el avance del aprendizaje.

Como:

$$\varepsilon^{(i)} = \frac{1}{2} \left| y^{(i)} - \sum_{k=1}^l m_k x_k^{(i)} \right|^2$$

Y:

$$\frac{\delta \varepsilon^{(i)}}{\delta m_j} = - \left(y^{(i)} - \sum_{k=1}^l m_k x_k^{(i)} \right) x_j^{(i)} = -\delta^{(i)} x_j^{(i)}$$

Entonces: $m_j(n+1) = m_j(n) + \alpha \delta^{(i)} x_j^{(i)}, \forall 1 \leq j \leq l, m(1)$ arbitrario.

Que es exactamente el mismo resultado obtenido anteriormente.

Lo que demuestra que todavía no hay razón ninguna que garantiza la convergencia. De hecho, se puede decir que los m valores resultantes no establecen un valor límite.

Se considera ahora otro tipo de minimización.

2. ESTADO DEL ARTE

Sea x, y variables aleatorias ($x \in \mathbb{R}^l, y \in \mathbb{R}$), y se busca: $\min \left(E \left[(y - m^T x)^2 \right] \right)$?

Aplicando el mismo procedimiento anterior, solo se remplazará:

A_{jk} por $E[x_j x_k]$ y b_j por $E[y x_j]$

La idea consiste en observar x, y para construir A, b (la parte común de la teoría de la estadística)

$$m_j(n+1) = m_j(n) + \alpha \left(y^{(n)} x_j^{(n)} - \sum_{k=1}^l x_j^{(n)} x_k^{(n)} m_k(n) \right)$$

Por lo tanto: $E[m(n+1)] = E[m(n)] + \alpha (b - E[m(n)])$

Al saber que esto nos lleva a elegir $0 < \alpha < \frac{2}{\lambda_{\max}}$

Entonces: $E[m(n)] \rightarrow m^* = \min E \left[(y - m^T x)^2 \right]$

Convergencia

Considerando entonces el caso general:

$((a^{(1)}, b^{(1)}), \dots, (a^{(p)}, b^{(p)})), a^{(i)} \in \mathbb{R}^l, b^{(i)} \in \mathbb{R}^m, \forall 1 \leq i \leq p$ Si $m = 1$, es decir un ALC (sistema anterior).

La idea consiste en buscar un método para minimizar el error del sistema general:

El error global: $\varepsilon(M) = \sum_{i=1}^p \varepsilon^{(i)}(M) = \sum_{i=1}^p \frac{1}{2} \|b^{(i)} - M a^{(i)}\|^2$,

$\varepsilon^{(i)}(M)$: error correspondiente al patrón i

En apartados anteriores, se ha obtenido:

$$\varepsilon(M) = \frac{1}{2} \|B - MA\|_F^2 \Rightarrow M = BA^\#$$

Por lo tanto, como: $\varepsilon^{(i)} = \frac{1}{2} (y^{(i)} - \sum_{k=1}^l m_k x_k^{(i)})^2$

Así que:

$$\frac{\partial \varepsilon}{\partial M_{jk}} = \sum_{i=1}^p \frac{\partial \varepsilon^{(i)}}{\partial M_{jk}}$$

$$\frac{\partial \varepsilon^{(i)}}{\partial M_{jk}} = \frac{\partial}{\partial M_{jk}} \left(\frac{1}{2} \sum_{k=1}^m \left(b_k^{(i)} - \sum_{j=1}^l M_{kj} a_j^{(i)} \right)^2 \right)$$

$$\frac{\partial \varepsilon^{(i)}}{\partial M_{jk}} = - \left(b_j^{(i)} - \sum_{s=1}^l M_{js} a_s^{(i)} \right) a_k^{(i)}$$

$$\frac{\partial \varepsilon^{(i)}}{\partial M_{jk}} = \left((M a^{(i)} - b^{(i)}) a^{(i)} \right)_{jk}$$

Continuación de las etapas del algoritmo LMS:

5. $(a^{(i)}, b^{(i)})$ Patrones presentados unos tras otros: $(a^{(1)}, b^{(1)}), \dots, (a^{(p)}, b^{(p)})$, constituyendo un ciclo de patrones, que se repite.

6. Una vez obtenida la matriz $M_{(n)}^{(i)}$ en el ciclo n :

$M_{(n)}^{(i+1)} = M_{(n)}^{(i)} - \alpha_n \left(M_{(n)}^{(i)} a^{(i)} - b^{(i)} \right) a^{(i)T}$, donde α_n es la tasa de aprendizaje.

$M_{(n)}^{(i+1)} = M_{(n)}^{(1)}$, es decir el patrón $p + 1$ será el patrón 1 en el ciclo $n + 1$.

Y el gradiente del error total es siempre:

$$\frac{\partial \varepsilon}{\partial M_{jk}} = \sum_{i=1}^p \frac{\partial \varepsilon^{(i)}}{\partial M_{jk}}$$

En cada etapa (i) , $\varepsilon^{(i)}$ se reduce, pero puede que ε aumente. Porque este mecanismo no es el caso del algoritmo del gradiente descendente estándar, y los argumentos de convergencia no se pueden aplicar.

Nota 2 Si $\alpha_n = \alpha > 0, \forall n$, entonces $M_{(n)}^{(i)} \rightarrow M_\alpha^{(i)}, \forall i = 1, 2, \dots, p.$ / Y $\lim_{\alpha \rightarrow 0} M_\alpha^{(i)} = BA^\#$
En general: $\lim_{\alpha \rightarrow 0} M_\alpha^{(i)}$ es diferente para cada i .

Por lo tanto:

Teorema 4 (Teorema de Convergencia del algoritmo LMS) Considerando α_n la serie de tasa de aprendizaje. Si:

$$\sum_{i=1}^{\infty} \alpha_n = \infty \text{ y } \sum_{i=1}^{\infty} \alpha_n^2 < \infty$$

Entonces la secuencia de las matrices generalizadas, para el algoritmo con la condición inicial $m(1) = 0$, converge a $BA^\#$

Observación El ALC se puede usar para limpiar el ruido de las señales (utilizando el mecanismo de retardo).

Se ha señalado anteriormente que las redes neuronales con una sola capa no pueden resolver todos los problemas, excepto problemas de clasificación con parámetros linealmente separables. Pero las redes con multicapas pueden aumentar la eficiencia y resolver otros tipos de problemas.

2.1.5.3 El Perceptrón

El perceptrón es una red neuronal simple, inventada por Frank Rosenblatt en 1957 (30). El modelo de Rosenblatt fue refinado y perfeccionado en la década de 1960. Minsky y Papert en (24) han ilustrado claramente las limitaciones del algoritmo del perceptrón.

Con el desarrollo del perceptrón, surge el área de las Redes Neuronales Artificiales dentro de la Inteligencia Artificial. A principios de los años setenta, el tema fue revisado gracias al descubrimiento de métodos y técnicas para superar las deficiencias del perceptrón.

2. ESTADO DEL ARTE

Definición 5 Sea $S_1, S_2 \in \mathbb{R}^n$, linealmente separables

$$\Leftrightarrow \exists w \in \mathbb{R}^n, \exists \theta \in \mathbb{R} : \begin{cases} \sum_{i=1}^n w_i x_i > \theta, \forall x \in S_1 \\ \sum_{i=1}^n w_i x_i < \theta, \forall x \in S_2 \end{cases}$$

$\Leftrightarrow S_1$ y S_2 se encuentran en lados opuestos de un hiperplano: $\sum_{i=1}^n w_i x_i = \theta \in \mathbb{R}$

Teorema 5 (Teorema de Convergencia del Perceptrón Simple)

Sea $S_1, S_2 \in \mathbb{R}^n$, y $C = (C_1) \cup (-C_2)$ donde:

$S_1 \subset C_1 \in \mathbb{R}^{n+1}, S_2 \subset C_2 \in \mathbb{R}^{n+1}$ y $y = (-1, x) \in \mathbb{R}^{n+1}$

S_1, S_2 son linealmente separables.

Sea $y^{(k)}$ cualquier secuencia de los vectores $y^{(k)} \in C$

$$Y \text{ si } y \in C \text{ ocurre infinitas veces en } y^{(k)} \begin{cases} w^{(1)} \in \mathbb{R}^{n+1} \text{ arbitrario.} \\ w^{(k+1)} = \begin{cases} w^{(k)} & \text{si } w^{(k)} y^{(k)} > 0 \\ w^{(k)} + y^{(k)} & \text{si no.} \end{cases} \end{cases} \exists N \in \mathbb{N} : w^{(N)} \cdot y > 0, \forall y \in C.$$

Es decir, que después de N etapas finitas de iteraciones, todas las entradas y los pesos no cambian.

Demonstración (Teorema de Convergencia del Perceptrón Simple)

Al definir $\alpha^{(k)}$ de tal forma que: $w^{(k+1)} = w^{(k)} + \alpha^{(k)} y^{(k)}$

Donde:

$$\alpha^{(k)} = \begin{cases} 0 & \text{si } w^{(k)} y^{(k)} > 0 \\ 1 & \text{si no} \end{cases} \quad \forall k$$

Al tener en cuenta también que: $\alpha^{(k)} y^{(k)} w^{(k)} \leq 0 \quad \forall k$. Entonces:

$$w^{(k+1)} = w^{(k)} + \alpha^{(k)} y^{(k)}$$

$$w^{(k+1)} = w^{(k-1)} + \alpha^{(k-1)} y^{(k-1)} + \alpha^{(k)} y^{(k)}$$

$$w^{(k+1)} = w^{(1)} + \alpha^{(1)} y^{(1)} + \dots + \alpha^{(k)} y^{(k)}$$

Como S_1 y S_2 son L.S. $\Rightarrow \exists \hat{w} \in \mathbb{R}^{n+1} : \hat{w} y > 0 \quad \forall y \in C$ Sea $\zeta = \min \{ \hat{w} y, y \in C \}$

Entonces: $\zeta > 0$ Por lo tanto: $\hat{w} y^{(k)} \geq \zeta, \forall k$ Y así:

$$w^{(k+1)} \hat{w} = w^{(1)} \hat{w} + \alpha^{(1)} y^{(1)} \hat{w} + \dots + \alpha^{(k)} y^{(k)} \hat{w}$$

$$\Rightarrow w^{(k+1)} \hat{w} \geq w^{(1)} \hat{w} + (\alpha^{(1)} + \dots + \alpha^{(k)}) \zeta$$

Aplicando la inigualdad de Shwartz ($|w^{(k+1)} \hat{w}| \leq \|w^{(k+1)}\| \|\hat{w}\|$), se obtiene:

$$w^{(1)} \hat{w} + (\alpha^{(1)} + \dots + \alpha^{(k)}) \zeta \leq \|w^{(k+1)}\| \|\hat{w}\|$$

Por otra parte:

$$\begin{aligned}
 \|w^{(k+1)}\|^2 &= w^{(k+1)} w^{(k+1)} \\
 \|w^{(k+1)}\|^2 &= \left(w^{(k)} + \alpha^{(k)} y^{(k)}\right) \left(w^{(k)} + \alpha^{(k)} y^{(k)}\right) \\
 \|w^{(k+1)}\|^2 &= \|w^{(k)}\|^2 + \left(\alpha^{(k)}\right)^2 \|y^{(k)}\|^2 + 2 \underbrace{\left(w^{(k)} \alpha^{(k)} y^{(k)}\right)}_{\leq 0} \\
 \|w^{(k+1)}\|^2 &\leq \|w^{(k)}\|^2 + \left(\alpha^{(k)}\right)^2 \|y^{(k)}\|^2, \text{ (según la definición de } \alpha^{(k)}, (\alpha^{(k)})^2 = \alpha^{(k)}) \\
 \|w^{(k+1)}\|^2 &\leq \|w^{(1)}\|^2 + \alpha^{(1)} \|y^{(1)}\|^2 + \dots + \alpha^{(k)} \|y^{(k)}\|^2 \\
 \|w^{(k+1)}\|^2 &\leq \|w^{(1)}\|^2 + \left(\alpha^{(1)} + \dots + \alpha^{(k)}\right) K, \text{ siendo } k = \max \left(\|y\|^2, \forall y \in C\right) \\
 \|w^{(k+1)}\|^2 &\leq \|w^{(1)}\|^2 + N^{(k)} K, \text{ siendo } N^{(k)} = \alpha^{(1)} + \dots + \alpha^{(k)}
 \end{aligned}$$

Así que:

$$\begin{cases} \|w^{(k+1)}\|^2 \leq \|w^{(1)}\|^2 + N^{(k)} K < \left(w^{(1)} + \sqrt{N^{(k)}} \sqrt{K}\right)^2 \\ \|w^{(k+1)}\| \|\hat{w}\| \geq w^{(1)} \hat{w} + N^{(k)} \zeta \end{cases}$$

Entonces: $N^{(k)} \kappa \leq \left(w^{(1)} + \ell \sqrt{N^{(k)}}\right) \Rightarrow N^{(k)} \leq b + c \sqrt{N^{(k)}}$ Así que $N^{(k)}$ tiene un posible valor mayor (extremo), y como $N^{(k+1)} \geq N^{(k)}$ (una serie decreciente), entonces la serie $N^{(k)}$ converge.

$$\exists N \in \mathbb{N} : \alpha^{(N)} = \alpha^{(N+1)} = \dots = 0 \Rightarrow w^{(N)} y > 0, \forall y \in C$$

Después de N , los pesos no cambian.

Observación La condición de “linealmente separable” era suficiente para la demostración, y no hizo falta saber \hat{w} .

Definición 6 $S_1, S_2 (\in \mathbb{R}^l)$ son linealmente separables estrictamente (L.S.S).

$$\Leftrightarrow \exists! v \in \mathbb{R}^l, \theta \in \mathbb{R}, \delta > 0 : \begin{cases} v \cdot x > \theta + \delta \quad \forall x \in S_1 \\ v \cdot x < \theta - \delta \quad \forall x \in S_2 \end{cases}$$

Teorema 6 Sea $C \in \mathbb{R}^{l+1}$, $\exists \hat{w} \in \mathbb{R}^{l+1}, \delta > 0 : \hat{w} \cdot y > \delta \quad \forall y \in C \quad \forall y^{(k)} \in C$
 ($w^{(k)}$) secuencia: $w^{(k+1)} = w^{(k)} + \alpha^{(k)} y^{(k)}$;

$$\text{donde } \alpha^{(k)} = \begin{cases} 0 & \text{si } w^{(k)} y^{(k)} > 0 \\ 1 & \text{si no.} \end{cases}$$

$$\Rightarrow \exists M : \alpha^{(1)}, \alpha^{(2)}, \dots, \alpha^{(k)} = N_1^{(k)} \leq M \quad \forall k = 1, 2, \dots$$

Es decir $\alpha^{(k)} \rightarrow 0$, k suficientemente grande, los pesos no cambian.

Nota 3 Cuando S_1, S_2 son L.S.S, no necesariamente se puede separar todos los vectores de S_1 , y S_2 , porque puede que al elegir $w^{(1)}$ arbitrario, se obtiene: $w^{(1)} y^{(i)} > 0, \forall i$. Lo que demuestra la imposibilidad de obtener una clasificación del todo satisfactoria.

2. ESTADO DEL ARTE

Teorema 7 Sea m clases S_i linealmente separables en el sentido de que existe:

$$w_1^*, w_2^*, \dots, w_m^* \in \mathbb{R}^n : w_i^* x > w_j^* x \quad \forall x \in S_i.$$

Si $(x^{(k)})$ es una serie de $S = S_1 \cup \dots \cup S_m$, cada patrón de esta serie aparece a menudo en S .

Entonces, después de un número finito de iteraciones la serie converge.

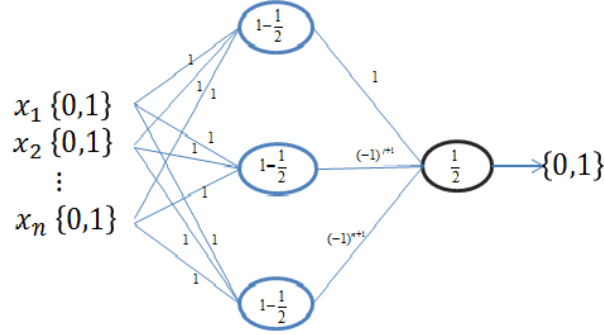


Fig. 2.3: Red con tres capas con funciones umbrales binarias

Definición 7 Sea f una función de paridad de orden n si:

$$f : [0, 1]^n \rightarrow \{0, 1\}$$

$$f(x) = \begin{cases} f(x_1, \dots, x_n) = 1 & \text{si } \sum_{i=1}^n x_i = 2k + 1 \\ f(x_1, \dots, x_n) = 0 & \text{si } \sum_{i=1}^n x_i = 2k \end{cases}$$

$$\text{i.e. } f(x_1, \dots, x_n) = \frac{1}{2}(1 - (-1)^{x_1 + \dots + x_n})$$

Teorema 8 No se puede implementar una función de paridad de orden n , como un perceptrón simple con n entradas.

Teorema 9 Se puede implementar una función de paridad de orden n , como una red de tres capas de funciones umbrales binarias con n entradas.

La figura 2.3, es un ejemplo gráfico de una red de tres capas de acuerdo con el teorema 9.

Método 1 (Algoritmo de Ho-Kashyap (31)) .

El objetivo es poder clasificar $S_1, S_2 \in \mathbb{R}^l$.

Sea $S_1 \subset C_1 \in \mathbb{R}^{l+1}, S_2 \subset C_2 \in \mathbb{R}^{l+1}$,

$C = C_1 \cup (-C_2)$.

Entonces hay que encontrar: $w, b \in \mathbb{R}^N, \forall y \in C : y = (y^{(1)}, \dots, y^{(N)})$ que satisfice $wy^{(i)} = b_i > 0 \forall i$.

Sea $b^{(1)}$ arbitrario, $b^{(1)}_j > 0 \forall 1 \leq j \leq N$.

$$w^{(k)} = y^\# b^{(k)} \quad y \quad b^{(k+1)} = b^{(k)} + \alpha \Delta b^{(k)}, \quad \alpha > 0$$

Como el error se define:

$$e^{(k)} = yw^{(k)} - b^{(k)}, \text{ entonces: } b^{(k+1)} = b^{(k)} + \alpha \begin{cases} e_j^{(k)} & \text{si } e_j^{(k)} > 0 \\ 0 & \text{si no} \end{cases}$$

Teorema 10 Si S_1, S_2 son L.S (linealmente separables) entonces:

si $\forall k, e_j^{(k)} \leq 0 \Rightarrow e^{(k)} = 0$. Se puede encontrar $e_j^{(k)} < 0$ entonces S_1, S_2 no son L.S.

Teorema 11 Si S_1, S_2 son L.S, $0 < \alpha < 2$. El algoritmo converge a una solución después de un número finito de iteraciones.

$$yw = b, \quad J = \|yw - b\|^2 = \frac{1}{2} \|yw - b\|_F^2 = \frac{1}{2} \sum_{j=1}^N (y^{(j)T} w - b_j)^2$$

$$\frac{dJ}{db_j} = -(y^{(j)T} w - b_j) \Rightarrow \frac{dJ}{db_j} = -(y^T w - b)$$

$$b^{(k+1)} = b^{(k)} + \alpha \frac{dJ}{db} = b^{(k)} - \alpha(y^T w - b), \quad y w^{(k)} = y^\# b^{(k)}$$

2.1.5.4 Conclusiones Fundamentales

- En (32) han demostrado que cualquier red neuronal “recurrente”, por muy compleja que pueda ser, se puede representar de una forma particular, llamada *Canónica*, que contiene una red neuronal “no recurrente” (sin bucles) donde algunas salidas vuelven a formar parte de las entradas (con un retardo de una unidad). Esta forma consiste, por lo tanto, en un gráfico cíclico (ver figura 2.4).

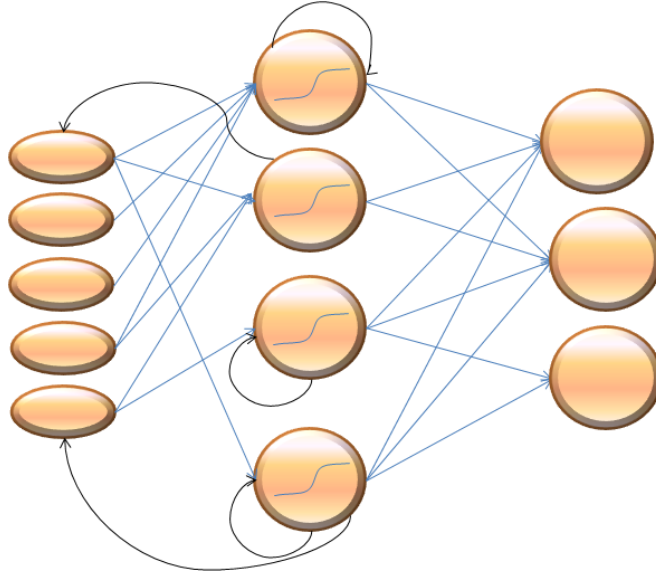


Fig. 2.4: Ejemplo de una Red Neuronal Recurrente

- En (22) Hornik et. al han demostrado que cualquier función suficientemente regular acotada, se puede aproximar uniformemente, con una precisión arbitraria, en un espacio

2. ESTADO DEL ARTE

finito de sus variables, mediante una red neuronal con una capa oculta de neuronas finitas, teniendo todas la misma función de activación y una neurona de salida lineal.

- En (33), se demuestra que el número de parámetros del modelo (en este caso la red neuronal) aumenta exponencialmente con el número de variables cuando se usan aproximadores lineales, y aumenta linealmente cuando se usan aproximadores no lineales.

La lentitud del entrenamiento es más notable cuando el número de entradas del modelo es más grande: para un modelo de una o dos entradas, se puede utilizar un modelo lineal con respecto a sus parámetros (por ejemplo: un polinomio) o un modelo no lineal respecto a sus parámetros (por ejemplo: una red neuronal).

2.1.5.5 Funciones de Activación

Entre las neuronas que forman la red neuronal artificial existe un conjunto de conexiones que las unen. Cada neurona transmite señales a aquellas que están conectadas con su salida. La función de activación de una neurona es la función que relaciona la información de entrada de la neurona con el siguiente estado de activación que tenga esa neurona.

Cuando se diseña una red neuronal deben establecerse los valores de activación para cada neurona, y decidir con qué función de activación la neurona procesará las entradas hacia las demás neuronas. Por lo tanto, la función de activación actúa sobre las señales de entrada, sobre los pesos sinápticos asociados y sobre el valor de activación que tenía la neurona en el momento de recibir las señales.

En esta sección se presentan las funciones de activación más comunes y usualmente utilizadas:

1. Función de escalón binaria

La función de escalón binaria o escalón de Heaviside “Binary threshold function”, también llamada función escalón unitario. Es una función discontinua cuyo valor es 0 para cualquier argumento negativo, y 1 para cualquier argumento positivo. Y se define como:

$$\varphi(v) = \begin{cases} 1 & \text{si } v \geq u \\ 0 & \text{si } v < u \end{cases}$$

2. Función de escalón bipolar

Llamada “bipolar threshold function”, se define como:

$$\varphi(v) = \begin{cases} 1 & \text{si } v \geq u \\ -1 & \text{si } v < u \end{cases}$$

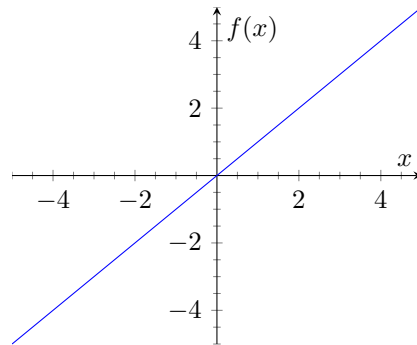


Fig. 2.5: Función lineal $f(x) = x$

3. Función lineal

$$\varphi(v) = tv$$

4. Función lineal a tramos

“Hard-limited linear function”

$$\varphi(v) = \begin{cases} 0 & \text{si } v < \frac{1}{2} \\ v + \frac{1}{2} & \text{si } -\frac{1}{2} \leq v < \frac{1}{2} \\ 1 & \text{si } v \geq \frac{1}{2} \end{cases}$$

5. Función de Activación competitiva

$$\varphi(v) = \begin{cases} 1 & \text{si n es máximo} \\ 0 & \text{si no} \end{cases}$$

6. Función Gaussiana

$$y = ce^{-\frac{x}{\sigma}}$$

7. Función Sigmoidal “Sigm” La función sigmoidal se define como:

$$y = 1/(1 + e^{-x}) = f(x) \tag{2.2}$$

Es una función continua, que tiende a 1 cuando $x \gg 0$, y tiende a 0 cuando $x \ll 0$,

2. ESTADO DEL ARTE

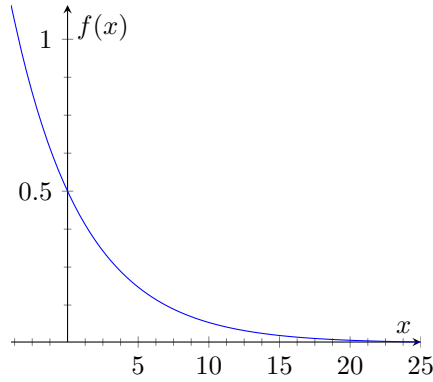


Fig. 2.6: Función Gaussiana $f(x) = \exp(-x)$

igual a $\frac{1}{2}$ cuando $x = 0$. Y calculando su derivada se obtiene:

$$\begin{aligned} \frac{dy}{dx} &= \frac{d}{dx} \left(\frac{1}{1 + e^{-x}} \right) \\ \frac{dy}{dx} &= \frac{d}{dx} (1 + e^{-x})^{-1} \\ \frac{dy}{dx} &= (-1) \times (1 + e^{-x})^{-2} \times (-1) \times e^{-x} \\ \frac{dy}{dx} &= \frac{1}{1 + e^{-x}} \times \frac{e^{-x}}{1 + e^{-x}} \\ \frac{dy}{dx} &= \frac{1}{1 + e^{-x}} \times \frac{1 + e^{-x} - 1}{1 + e^{-x}} \\ \frac{dy}{dx} &= \frac{1}{1 + e^{-x}} \times \left(\frac{1 + e^{-x}}{1 + e^{-x}} - \frac{1}{1 + e^{-x}} \right) \end{aligned}$$

Por lo tanto:

$$\frac{dy}{dx} = y(1 - y) \quad (2.3)$$

8. Función Tangente hiperbólica “Tanh”:

También cumple la condición de continuidad, y su primera derivada es fácil de calcular como en la ecuación 2.3, y se define como:

$$y = \tanh\left(\frac{x}{2}\right) = g(x)$$

$$y = \frac{1 - e^{-x}}{1 + e^{-x}} \quad (2.4)$$

Es decir: $y = \frac{2}{1 + e^{-x}} - 1$

$y = 2f(x) - 1 = g(x)$, siendo f la función sigmoideal anterior.

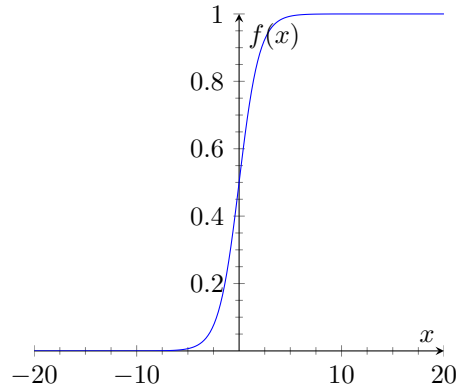


Fig. 2.7: Función Sigmoidal $f(x) = \text{sigmoid}(x)$

Calculando la derivada de la Tanh se obtiene:

$$\begin{aligned} \frac{dy}{dx} &= 2 \frac{df(x)}{dx} \\ \frac{dy}{dx} &= 2f(x)(1 - f(x)) \end{aligned} \quad (2.5)$$

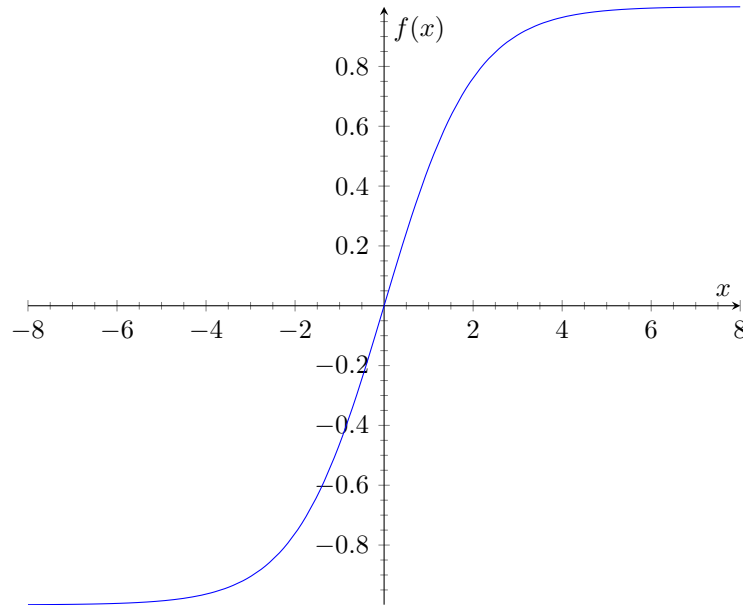


Fig. 2.8: Función Tangente hiperbólica $f(x) = \text{Tanh}(x)$

Nota 4 En (34), se presenta un estudio exhaustivo de los motivos por los que se usa la función *Tanh* para resolver los problemas no lineales.

2. ESTADO DEL ARTE

2.1.6 El Aprendizaje Numérico

Dentro las propiedades deseadas de una red neuronal, la más fundamental es su capacidad de aprender de su entorno y de mejorar su rendimiento a través de su aprendizaje. Pero ¿qué es el aprendizaje?

La noción del aprendizaje es clara e intuitiva para los seres humanos y los animales, es un proceso cognitivo que tiene que hacer el individuo para realizar una tarea especificada. Típicamente, estos procesos se suelen efectuar a partir de pruebas, hasta que se pueda aprender a realizar tal tarea de una manera efectiva. Por ejemplo, cuando se enseña a un niño leer o escribir, al final del aprendizaje el niño será capaz de leer y escribir, pero no solamente las formas de escritura de aquellos libros que se usaron para su enseñanza, sino otras debido a la capacidad de generalizar adquirida en el proceso de aprendizaje .

El aprendizaje numérico tiene el mismo objetivo, poder aprender y generalizar. Con la ayuda de procesos numéricos de programación y algoritmos, y realizar predicciones dependiendo de las entradas.

En el contexto de redes neuronales, se adopta esta definición:

Definición 8 *El aprendizaje es un proceso dinámico e iterativo que permite modificar los parámetros de la red como respuesta a los estímulos que recibe de su entorno. El tipo de aprendizaje se determina de cómo se producen los cambios en los parámetros de la neurona.*

En la mayoría de las referencias bibliográficas, el aprendizaje se traduce como el cambio o la modificación de la eficiencia sináptica, es decir los cambios de los parámetros (pesos) de la red.

La manera tradicional de cambiar los pesos es mediante su diferencia:

$$\Delta w_{ij}(t) = w_{ij}(t+1) - w_{ij}(t) \quad (2.6)$$

Un conjunto de reglas bien definidas permiten realizar tales procesos, lo que se llama “*Algoritmo de Aprendizaje de la Red*”.

En el contexto de la definición 8, se pueden distinguir dos tipos de aprendizaje: aprendizaje “supervisado” y aprendizaje “sin supervisión” (no-supervisado).

2.1.6.1 Aprendizaje Supervisado

Se ha visto en el apartado anterior, que una red realiza una relación algebraica entre sus entradas y salidas. Por lo tanto se puede asignar a la red una tarea, y realizar una función algebraica no lineal, que puede ser:

1. Analíticamente conocida: Por ejemplo, aproximar una función.

2. Analíticamente desconocida: Cuando se dispone de valores resultantes mediante mediciones, registros, etc. Por lo tanto, la red realiza lo que se llama un modelado estadístico, o un modelado de regresión.

Este último tipo de aplicaciones, función algebraica no lineal analíticamente desconocida, es el motivo principal por el cual se implementan mucho las redes neuronales mediante el aprendizaje supervisado. Los puntos para los que se dispone de valores que deben ser la salida de la red de acuerdo a las entradas correspondientes, son los que posibilitan el aprendizaje “supervisado”. Después del entrenamiento, la red neuronal se evalúa usando conjuntos de entrada y salida conocidos, proporcionándole sólo valores de entrada y midiendo la cercanía de las salidas de la red con respecto a las salidas correctas.

2.1.6.2 Aprendizaje no-supervisado

En el aprendizaje no supervisado, a la red neuronal no se le proporcionan las salidas objetivo durante el aprendizaje. Las redes neuronales no-supervisadas suelen realizar algún tipo de compresión de datos, tales como la reducción de dimensionalidad o agrupamiento. Las redes neuronales de aprendizaje no supervisado más estudiadas son las “mapas de Kohonen”.

2.1.7 Conceptos básicos de una Red Neuronal

Una red neuronal es una función no lineal, parametrizable con valores definidos. (Se mencionará en adelante la palabra red neuronal lineal, solo por distinguir que su función de activación es lineal).

Una representación sencilla de una neurona artificial se muestra en la figura 2.9 que se inspira del modelo de una neurona biológica (35).

$$y = f(w_0 + \sum_{i=1}^n w_i x_i)$$

Donde:

f : es la función de activación,

y : es la salida de la neurona,

$x_i, \forall 1 \leq i \leq n$ son las variables de entrada,

$w_i, \forall 1 \leq i \leq n$ son los parámetros (llamados también *pesos* de la red neuronal), y w_0 se llama *bias* o la polarización, que representa el umbral de activación de la neurona (en general $w_0 = -1$).

2. ESTADO DEL ARTE

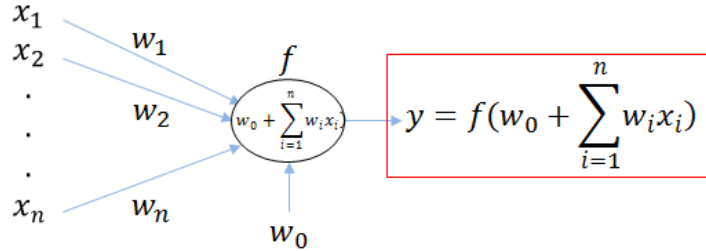


Fig. 2.9: Red de una Neurona

Nota 5 En las redes neuronales biológicas, las neuronas son asincrónicas, sin embargo las redes neuronales artificiales son sincrónicas (i.e tiempo es discreto, cada instante t se calcula $y = f(w_0 + \sum_{i=1}^n w_i x_i(t))$).

Como se ha señalado anteriormente, una neurona realiza simplemente una función no lineal, con un conjunto de variables de entrada. La ventaja de las neuronas consiste en las propiedades resultantes de su asociación en la red, es decir, la composición de las funciones no lineales llevadas a cabo por cada una de las neuronas.

2.1.8 Proceso de Aprendizaje

2.1.8.1 Aprendizaje mediante Corrección del Error

Se trata de corregir el error de la salida neuronal observada, en un instante t :

$$e(t) = y(t) - f\left(\sum_{i=1}^n w_i x_i(t) + w_0\right)$$

$$e_i(t) = y_i(t) - f_i\left(\sum_{i=1}^n w_i x_i(t) + w_0\right), \quad e(t) = (e_1(t), \dots, e_n(t))$$

El aprendizaje por corrección del error consiste en minimizar un índice de rendimiento F basado en los errores $e_i(t)$, con el objetivo de que las salidas de la red converjan hacia las salidas deseadas.

Por ejemplo, se puede calcular la suma de los errores cuadráticos:

$$F(e(t)) = \sum_{i=1}^n e_i(t)^2 = e^T(t)e(t)$$

Y minimizar F (obj. $F(t+1) < F(t)$):

$$F(t+1) = F(t) + \nabla F(t)^T \Delta w(t)$$

Donde: $\Delta w(t) = w(t+1) - w(t)$, ∇F es el gradiente de F .

Entonces para que $F(t+1) < F(t)$:

$$\nabla F(t)^T \Delta w(t) = \eta \nabla F(t)^T x(t) < 0$$

Donde $\eta > 0$, el vector $x(t)$ (dirección en la que se busca el mínimo) apunta en el sentido contrario del gradiente.

$$x(t) = -\nabla F(t)$$

$$\Rightarrow \Delta w(t) = \eta x(t) = -\nabla F(t)$$

$$\Rightarrow w(t+1) = w(t) - \eta \nabla F(t)$$

La expresión exacta de ∇F depende de la función de activación.

Nota 6 Esta regla de aprendizaje se puede usar solamente en un contexto de aprendizaje supervisado.

2.1.8.2 Aprendizaje mediante la Regla de Hebb

Esta regla ha sido inspirada en los trabajos neurofisiológicos de Donald Hebb (23):

“When an axon of cell A is near enough to excite a cell B and repeatedly or distantly takes part in firing it, some growth process or metabolic changes take place in one or both cells such that A’s efficiency as one of the cells firing B, is increased”. En este sentido neurológico, Hebb buscaba establecer una forma de aprendizaje asociativo a nivel celular. En las redes neuronales artificiales, se reformula la regla de Hebb en un aprendizaje de dos partes:

1. Si dos neuronas están conectadas mediante una sinapsis (conexión) están activadas simultáneamente (de una manera sincronizada), entonces la fuerza de esa conexión debe ser aumentada.
2. Si las mismas neuronas están activadas asincrónicamente, entonces la conexión debe ser debilitada o eliminada.

Tal mecanismo se llama: *aprendizaje Hebbiano*, y utiliza un espacio interactivo, en función del tiempo y del espacio, para aumentar la eficacia sináptica de una manera proporcional con la correlación de actividades pre-sinápticas y post-sinápticas.

De esta explicación, se destacan las siguientes propiedades:

2. ESTADO DEL ARTE

1. Dependencia temporal: las modificaciones de la sinapsis Hebbiana depende del momento exacto de las actividades pre y post-sinápticas.
2. Dependencia espacial: dada la naturaleza de la sinapsis que consitituye un lugar de transmisión de la información, el aprendizaje Hebbiano debe tener una continuidad espacial. Esa propiedad permite decir que el aprendizaje es *no-supervisado*.
3. Interacción: el aprendizaje de Hebb depende de la interacción de las actividades de ambos lados de la sinapsis.
4. Correlación: una interpretación de la regla de Hebb es que la condición que permite un cambio en la eficacia sináptica, que a su vez es una combinación de actividades pre y post-sinápticas. Otra interpretación más estadística se refiere a la correlación de estas actividades.

Se puede expresar la regla de Hebb en una forma sencilla mediante la fórmula siguiente:

$$\Delta w(t-1) = \eta p(t)y(t)$$

Es decir:

$$\Delta w_i(t-1) = \eta p_i(t)y(t)$$

Donde:

η constante positiva, que determina la velocidad del aprendizaje y $p_i(t)$ corresponde a la actividad pre-sináptica (la entrada i de la neurona) en el momento t .

$y(t)$ la actividad post-sináptica (salida de la neurona) en el momento t .

El problema es que $\Delta w_i(t)$ puede \nearrow o \searrow exponencialmente, lo que puede provocar una saturación en los pesos, por lo tanto se modifica de la manera siguiente:

$$\Delta w(t-1) = \eta p(t)y(t) - \alpha w(t-1), 0 \leq \alpha \leq 1 \quad (2.7)$$

Pero esta regla no resuelve todos los problemas, es esencial que los estímulos se repitan con regularidad, si no, las asociaciones aprendidas a través de la regla de Hebb en la ecuación 2.7 se pierden con el tiempo. Por lo tanto se añade una modificación a la ecuación 2.7:

$$\Delta w(t-1) = \eta p(t)y(t) - \alpha w(t-1)y(t)$$

si $\alpha = \eta$, se obtiene:

$$\Delta w(t-1) = \eta y(t) [p(t) - w(t-1)] \quad (2.8)$$

Una propiedad interesante del aprendizaje en la ecuación 2.8: si las entradas son normalizadas, tras el proceso de aprendizaje, los pesos convergerán a vectores normalizados.

2.1.9 Redes Neuronales Multicapa

Se ha visto que hay limitaciones en las capacidades computacionales del perceptrón. Por ejemplo, se sabe que cualquier función booleana se puede implementar mediante una red neural multicapa (de hecho, tres capas son suficientes). La pregunta que hay que hacerse para las redes multicapa es cómo se modifican los pesos.

Se discute, en esta sección, un método muy popular de aprendizaje de las redes neuronales feedforward con múltiples capas, [*A menudo, las redes neuronales disponen de capas de manera que las conexiones estan sólo entre capas consecutivas, y todas en la misma dirección, y no habiendo conexiones dentro de cualquier capa dada. Este tipo de redes neuronales son llamadas redes neuronales feedforward*] llamado algoritmo del error de propagación hacia atrás. Este último es un algoritmo de aprendizaje supervisado basado en un error establecido o con una función de coste, con valores determinados por las salidas reales y deseadas de la red, reducida al mínimo a través de un método dado como por ejemplo el método del gradiente descendente. Para esto, es necesario que las funciones de activación de las neuronas de la red sean diferenciables. La idea es muy simple, pero el análisis se complica debido a la abundancia de índices. Por ello, se va a tratar de definir y dar algunas características de los modelos de redes neuronales con multicapas. Para ilustrar el problema, se consideran las siguientes tres capas de feedforward (una simple red hacia delante con una capa oculta).

En este ejemplo hay: n entradas, l capas ocultas, m salidas en la tercera capa, como se ve en la figura 2.10.

2.1.9.1 Aproximación de una función

El objetivo de esta sección es llegar a la conclusión de que mediante una red neuronal multicapa, con una sola capa oculta. Se puede aproximar funciones con precisiones arbitrarias, a condición de disponer de suficientes neuronas en la capa oculta, para poder procesar la información.

Suponiendo que la red tiene n variables en la capa de entrada, y m variables en la capa de salida, entonces cualquier función dada: $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$, determina la relación entre los pares $(x, y = f(x))$.

Por lo tanto, el objetivo sería entrenar la red para que pueda aproximar tal función, y lo que es de interés es saber cómo se puede lograr tal objetivo.

2. ESTADO DEL ARTE

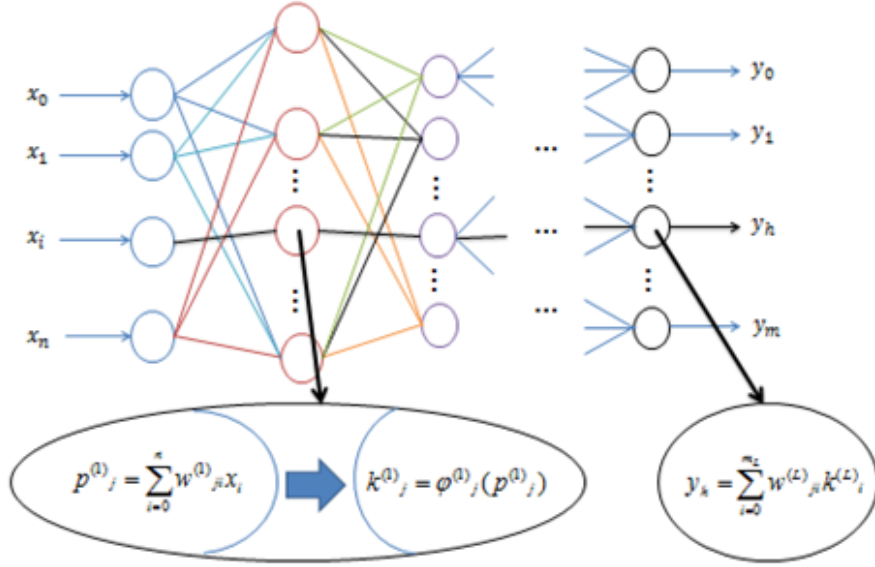


Fig. 2.10: Red Neuronal Multicapas (L capas) con n variables de entrada y m variables de salida

El teorema de Kolmogorov y de Sprecher, define la posibilidad de representar funciones continuas con multi-variables, en términos de una función continua de combinaciones lineales de funciones continuas de una sola variable.

Teorema 12 (Teorema de Kolmogorov)

Para cualquier función $f : [0, 1]^n \rightarrow \mathbb{R}$, existen funciones continuas: f_1, \dots, f_{2n+1} en \mathbb{R} y funciones continuas monótonas (crecientes) g_{ij} , $1 \leq i \leq n$ y $1 \leq j \leq 2n+1$, tal que:

$$f(x_1, \dots, x_n) = \sum_{j=1}^{2n+1} f_j \left(\sum_{i=1}^n g_{ij}(x_i) \right).$$

Las funciones g_{ij} no dependen de f .

Teorema 13 (Teorema de Sprecher)

Para cualquier función $f : [0, 1]^n \rightarrow \mathbb{R}$, existen funciones: h en \mathbb{R} , funciones continuas monótonas (crecientes) g_j , $1 \leq j \leq 2n+1$, y n constantes $\alpha_1, \dots, \alpha_n$ de modo que:

$$f(x_1, \dots, x_n) = \sum_{j=1}^{2n+1} h \left(\sum_{i=1}^n \alpha_i g_j(x_i) \right).$$

Las funciones g_j y las constantes α_i no dependen de f .

Este teorema implica que cualquier función continua f de un subconjunto compacto de $\mathbb{R}^n \rightarrow \mathbb{R}^m$, se puede realizar por unas tres capas de una red neuronal feedforward.

Por consiguiente, se demuestra que toda función continua en \mathbb{R}^n puede ser implementada en una red neuronal feedforward con 3 capas. Este resultado depende del teorema de Stone-Weierstrass.

Teorema 14 (Teorema de Stone-Weierstrass)

Suponiendo que Ω es un algebra de funciones reales continuas con valores en un compacto K , y Ω separa puntos de K y contiene constantes. Entonces Ω es uniformemente denso en $\zeta(K)$ espacio definido por todas las funciones reales continuas en K .

El teorema 14, implica que cualquier función continua en un compacto, se puede aproximar con suma de funciones de cosenos.

Nota 7

1. Ω es un algebra simple si $\forall f, g \in \Omega, \forall \alpha \in \mathbb{R}, fg \in \Omega$ y $\alpha f + g \in \Omega$
2. Ω separa puntos de K cuando:

$$\forall x_1, x_2 \in K, x_1 \neq x_2 \Rightarrow \exists f \in \Omega : f(x_1) \neq f(x_2)$$

3. Se puede decir que si Ω es un algebra simple y separa puntos de K , entonces Ω es suficientemente capaz de distinguir diferentes puntos de K .
4. El requisito de que Ω contiene constantes es para descartar la posibilidad de que todas las funciones puedan desaparecer en algún punto común.
5. La densidad uniforme de $\Omega \Rightarrow \forall f \in \zeta(K), \forall \varepsilon > 0, \exists g \in \Omega : |f(x) - g(x)| < \varepsilon, \forall x \in K$

Teorema 15

Suponiendo que: $f : [0, \pi]^2 \rightarrow \mathbb{R}$ es continua. $\forall \varepsilon > 0, \exists N \in \mathbb{N}, a_{mn}, (0 \leq m, n \leq N)$ entonces:

$$\left| f(x, y) - \sum_{m,n=1}^N a_{mn} \cos(mx) \cos(ny) \right| < \varepsilon, \forall x, y \in [0, \pi]^2$$

Teorema 16 Si: $f : [0, \pi]^2 \rightarrow \mathbb{R}$ es continua, $\forall \varepsilon > 0$ existe una red neuronal feedforward de 3 capas (la segunda es una capa oculta), siendo la capa de salida lineal, que implementa la función f en $[0, \pi]^2$ con una precisión ε .

Demonstración Aplicando el teorema 14:

$$\exists N \in \mathbb{N}, a_{mn}, (0 \leq m, n \leq N) : \left| f(x, y) - \sum_{m,n=1}^N a_{mn} \cos(mx) \cos(ny) \right| < \frac{\varepsilon}{2}, \forall x, y \in [0, \pi]^2$$

Entonces: $\forall (x, y) \in [0, \pi]^2$, sea $K = \max \{|a_{mn}| : 0 \leq m, n \leq N\}$. Por otra parte:

$$\cos(mx) \cos(ny) = \frac{1}{2} (\cos(mx + ny) + \cos(mx - ny))$$

y:

$$|mx \pm ny| \leq 2N\pi, \forall (x, y) \in [0, \pi]^2$$

Se puede aproximar:

$$\left| \frac{1}{2} \cos(t) - \tau(t) \right| < \frac{\varepsilon}{4(N+1)^2 K}$$

2. ESTADO DEL ARTE

Sea $\forall t \in [-2N\pi, 2N\pi]$:

$$\begin{aligned} & \left| \sum_{m,n=1} a_{mn} \cos(mx) \cos(ny) - \sum_{m,n=1} a_{mn} (\tau(mx + ny) + \tau(mx - ny)) \right| \\ & \leq \sum_{m,n=1} |a_{mn}| \left(\frac{\varepsilon}{4(N+1)^2 K} + \frac{\varepsilon}{4(N+1)^2 K} \right) \\ & \left| \sum_{m,n=1} a_{mn} \cos(mx) \cos(ny) - \sum_{m,n=1} a_{mn} (\tau(mx + ny) + \tau(mx - ny)) \right| < \sum_{m,n=1} \frac{\varepsilon}{2(N+1)^2} \\ & \left| \sum_{m,n=1} a_{mn} \cos(mx) \cos(ny) - \sum_{m,n=1} a_{mn} (\tau(mx + ny) + \tau(mx - ny)) \right| < \frac{\varepsilon}{2} \end{aligned}$$

Pero como τ se puede representar como: $\tau(t) = \sum_{j=1}^M w_j \text{step}(t - \theta_j)$.

Entonces:

$$\sum_{m,n=1} a_{mn} (\tau(mx + ny) + \tau(mx - ny)) = \sum_{m,n=1} a_{mn} w_j (\text{step}(mx + ny - \theta_j) - \text{step}(mx - ny - \theta_j))$$

Que es justo una representación de la formulación matemática de una red neuronal.

2.1.9.2 Clasificación

Un clasificador es un algoritmo que asigna automáticamente una clase (o categoría) a un determinado patrón. Antes de considerar el caso específico de clasificadores “neuronales”, es importante comprender las características básicas de los problemas de clasificación. Las redes con dos capas (una capa oculta, con función de activación sigmoideal), permiten generar fronteras de decisión convexas (abiertas o cerradas), mientras que las redes de tres capas ocultas, y funciones de activación no lineales, permiten generar fronteras de decisión cóncava o convexa (abierta o cerrada). Hay que tener en cuenta que una curva o superficie convexa no implica ningún cambio en el signo de la concavidad, mientras una cóncava implica un punto de inflexión. Los componentes de un problema de clasificación suelen ser:

1. un conjunto de N patrones;
2. n variables (o características) que describen los patrones y son relevantes para la tarea de clasificación;
3. un conjunto de C clases al que los patrones deben ser asignados (una de las clases puede ser una clase de rechazo en la que todos los patrones que no pueden ser asignados a las otras clases serán clasificados)

Por lo tanto, la solución de un problema de clasificación requiere encontrar una aplicación del conjunto de patrones que debe ser clasificado en un conjunto de clases. Es importante darse cuenta de que los clasificadores estadísticos, tales como las redes neuronales, no son apropiados para resolver todos los problemas de clasificación: hay muchos métodos de clasificación alternativos disponibles. En (36) se muestra una serie de ejemplos que ilustra el ámbito de la aplicación de las redes neuronales para la tarea de clasificación.

2.1.10 Retro-Propagación

2.1.10.1 Repaso Histórico del Algoritmo de Retro-Propagación

A finales de 1950, se introdujeron dos redes neuronales artificiales que han tenido un gran impacto en los actuales modelos de redes neuronales. La primera de ellas se conoce como el perceptrón y contiene unidades lineales con umbral, es decir, las salidas son cero o uno. La segunda contiene unidades lineales con umbral, a partir de unidades Adaline que tienen una salida lineal sin un umbral. Estas redes utilizan una regla de aprendizaje que se llama la regla delta (17).

El principal inconveniente de estos dos modelos de redes neuronales son sus limitaciones. En (24) Minsky y Pitts mostraron que tales redes son sólo capaces de asociar entradas linealmente separables. Minsky y Papert también observaron que estas limitaciones se pueden superar si se introduce una capa intermedia de más. En ese momento, sin embargo, no se conoce ninguna regla de aprendizaje eficiente para las redes con capas intermedias. En 1985, se reportaron varios esquemas de aprendizaje para la adaptación de estas conexiones como lo muestran los trabajos de Parker y LeCun (16, 37).

En esta apartado, se va a centrar exclusivamente en la regla delta generalizada que se introdujo en 1986. La aplicación de la regla delta generalizada requiere dos fases. En La primera fase, la entrada se propaga hacia adelante a las unidades de salida, donde se mide el error de la red. En la segunda fase, el error se propaga hacia atrás a través de la red y se utiliza para la adaptación de las conexiones. Debido a la segunda fase, este procedimiento se conoce hoy en día como el método de propagación hacia atrás del error (Retro-Propagación). Es importante señalar que este procedimiento es similar a un algoritmo descrito mucho antes por Werbos en 1974.

2.1.10.2 Introducción

El método de entrenamiento más popularmente usado es el algoritmo de retro-propagación, que usa esencialmente el gradiente descendente. Para el algoritmo del gradiente descendente,

2. ESTADO DEL ARTE

un tamaño de paso, que se denomina: la tasa de aprendizaje, debe ser especificado. La tasa de aprendizaje es crucial para la propagación hacia atrás de los cambios de pesos. Pero ante un cambio no adecuado en la tasa de aprendizaje, el entrenamiento puede sufrir problemas de convergencia lenta, ineficiencia y falta de robustez. Elegir tasas de aprendizaje pequeñas tiende a retardar el proceso de aprendizaje, mientras que las grandes tasas de aprendizaje pueden provocar oscilaciones de la red en el espacio de pesos. Una forma de mejorar el método de gradiente descendente original, es incluir un parámetro adicional llamado momentum, para permitir que tasas más grandes de aprendizaje, produzcan una convergencia más rápida y una reducción al mínimo de la tendencia a oscilar (17). La idea de introducir el término momentum consiste en dirigir el cambio de peso en el próximo paso, en la misma dirección que la etapa anterior, y por lo tanto reducir el efecto de oscilaciones mayores en las tasas de aprendizaje. In (38), Yu et al. describen un método dinámico de optimización adaptativo de la tasa de aprendizaje usando su derivada. También se ha mostrado que el paso se puede determinar efectivamente mediante el establecimiento de la relación entre la propagación hacia atrás y el método del gradiente conjugado.

2.1.10.3 Algoritmo de entrenamiento: Retro-Propagación y RPROP

El algoritmo de Retro-Propagación también conocido como BCP, o con el nombre propagación hacia atrás, en inglés “Back-Propagation”. La retro-propagación es un algoritmo supervisado de las redes multicapas. Requiere dos pasos de cálculo:

1. la propagación de la activación (paso hacia delante)
2. retro-propagación de errores (paso hacia atrás)

La retro-propagación funciona de la siguiente manera: se propaga la activación de la entrada a la capa oculta, y de la capa oculta a la capa de salida, se calcula el error en las unidades de salida, de nuevo se propaga el error en las unidades ocultas y luego a las unidades de entrada. BCP tiene un poder de aproximación universal, es decir, dada una función continua, existe una red de tres capas (una capa oculta) que puede ser entrenada mediante la retro-propagación con el fin de aproximar esta función. Aunque el algoritmo de retro-propagación es el algoritmo de entrenamiento conexionista más conocido y utilizado, es computacionalmente caro y a veces lento, y no resuelve satisfactoriamente los problemas de gran tamaño en general, incluso la solución encontrada puede ser atrapada en un mínimo local y no global. Estas dificultades se han intentado superar mediante sucesivas mejoras en el algoritmo.

2.1 Redes Neuronales Artificiales “RNAs”

Luego en resumen, dos parámetros importantes se utilizan para controlar el proceso de aprendizaje de una red neuronal mediante la retro-propagación. La *tasa de aprendizaje* se utiliza para especificar si en la red neuronal se van a hacer ajustes importantes después de cada ensayo de aprendizaje, o si sólo se van a hacer pequeños ajustes. El *momentum* es la tasa adicional de aprendizaje que se usa para controlar las posibles oscilaciones en los pesos, que podrían ser provocadas por las señales del error. Ahora se va tratar con el caso más general de una red

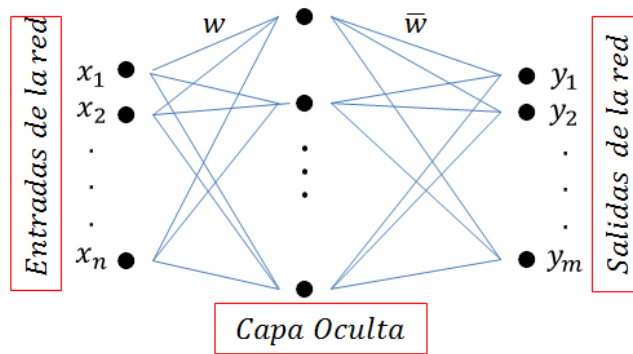


Fig. 2.11: Red neuronal de tres capas

neuronal de tres capas feedforward con n nodos de entrada, m neuronas en la capa oculta y l en la capa de la salida. Un análisis similar puede llevarse a cabo en general en una red neuronal multicapa (más capas ocultas). En primer lugar, se denota que:

- Las variables $\{x_1, x_2, \dots, x_n\}$ son las entradas que se encuentran en la capa de entrada.
- Las variables $\{d_1, d_2, \dots, d_l\}$ son las salidas deseadas.
- Las variables $\{y_1, y_2, \dots, y_m\}$ son las salidas de la red neuronal, que se encuentran en la capa de salida.
- El peso w_{ji} que conecta una neurona i de la capa de entrada con la neurona j de la capa oculta.
- El peso \bar{w}_{kj} que conecta una neurona j de la capa oculta con la neurona k de la capa de salida.
- La función $\varphi(\cdot)$ será la función de activación utilizada en la capa oculta.

2. ESTADO DEL ARTE

- La función $\bar{\varphi}(\cdot)$ será la función de activación utilizada en la capa de salida.

Por lo tanto cada neurona j de la capa oculta tiene como entrada:

$$v_j^{in} = \sum_{i=1}^n w_{ji}x_i + \theta_j = \sum_{i=0}^n w_{ji}x_i$$

Donde $\theta_j = w_{j0}$, $x_0 = -1$. (θ_j es el umbral de la neurona j)

Y como salida:

$$v_j^{out} = \varphi(v_j^{in}) = \varphi\left(\sum_{i=0}^n w_{ji}x_i\right)$$

Entonces, la entrada de la neurona k de la capa de salida, sería:

$$y_k^{in} = \sum_{j=1}^l \bar{w}_{kj}v_j^{out} + \bar{\theta}_k = \sum_{j=0}^l \bar{w}_{kj}v_j^{out}$$

Donde: $\bar{w}_{k0} = \bar{\theta}_k$ y $v_0^{out} = -1$. ($\bar{\theta}_k$ es el umbral de la neurona k)

Y su salida sería:

$$y_k = y_k^{out} = \bar{\varphi}(y_k^{in}) = \bar{\varphi}\left(\sum_{j=0}^l \bar{w}_{kj}v_j^{out}\right)$$

La función del error de la señal en la neurona de la salida está definido por:

$$\varepsilon = \frac{1}{2} \sum_{k=1}^m (d_k - y_k)^2$$

La estrategia consiste en intentar minimizar ε .

Se suelen usar diversos algoritmos, se empieza en este párrafo con el algoritmo del gradiente descendiente, aplicado a los pesos de la red neuronal:

$$w \mapsto w - \lambda \text{grad} \varepsilon \tag{2.9}$$

Donde λ es la tasa de aprendizaje.

Para concluir, se necesita calcular las derivadas siguientes.

$$\begin{aligned}
 \frac{d\varepsilon}{d\bar{w}_{kj}} &= \sum_{s=1}^l - (d_s - y_s) \frac{dy_s}{d\bar{w}_{kj}} \\
 \frac{d\varepsilon}{d\bar{w}_{kj}} &= - (d_k - y_k) \frac{dy_k}{d\bar{w}_{kj}} \\
 \frac{d\varepsilon}{d\bar{w}_{kj}} &= - (d_k - y_k) \frac{dy_k^{out}}{d\bar{w}_{kj}} \\
 \frac{d\varepsilon}{d\bar{w}_{kj}} &= - (d_k - y_k) \frac{d\bar{\varphi}(y_k^{in})}{d\bar{w}_{kj}} \\
 \frac{d\varepsilon}{d\bar{w}_{kj}} &= - (d_k - y_k) \frac{d\bar{\varphi}(\sum_{j=0}^l \bar{w}_{kj} v_j^{out})}{d\bar{w}_{kj}} \\
 \frac{d\varepsilon}{d\bar{w}_{kj}} &= - (d_k - y_k) \bar{\varphi}'(y_k^{in}) v_j^{out} \\
 \frac{d\varepsilon}{d\bar{w}_{kj}} &= - \bar{\psi} v_j^{out}
 \end{aligned}$$

Donde: $\bar{\psi} = (d_k - y_k) \bar{\varphi}'(y_k^{in})$. Y por otra parte:

$$\begin{aligned}
 \frac{d\varepsilon}{dw_{ji}} &= - \sum_{s=1}^l (d_s - y_s) \frac{dy_s}{dw_{ji}} \\
 \frac{d\varepsilon}{dw_{ji}} &= - \sum_{s=1}^l (d_s - y_s) \frac{dy_s^{out}}{dw_{ji}} \\
 \frac{d\varepsilon}{dw_{ji}} &= - \sum_{s=1}^l (d_s - y_s) \frac{d\bar{\varphi}(y_s^{in})}{dw_{ji}} \\
 \frac{d\varepsilon}{dw_{ji}} &= - \sum_{s=1}^l (d_s - y_s) \frac{d\bar{\varphi}\left(\sum_{j=0}^l \bar{w}_{sj} v_j^{out}\right)}{dw_{ji}} \\
 \frac{d\varepsilon}{dw_{ji}} &= - \sum_{s=1}^l (d_s - y_s) \frac{d\bar{\varphi}\left(\sum_{j=0}^l \bar{w}_{sj} \varphi\left(\sum_{i=0}^n w_{ji} x_i\right)\right)}{dw_{ji}} \\
 \frac{d\varepsilon}{dw_{ji}} &= - \sum_{s=1}^l (d_s - y_s) \bar{\varphi}'(y_s^{in}) \bar{w}_{sj} \frac{d\left(\varphi\left(\sum_{i=0}^n w_{ji} x_i\right)\right)}{dw_{ji}} \\
 \frac{d\varepsilon}{dw_{ji}} &= - \sum_{s=1}^l (d_s - y_s) \bar{\varphi}'(y_s^{in}) \bar{w}_{sj} \varphi'(v_j^{in}) x_i \\
 \frac{d\varepsilon}{dw_{ji}} &= - \psi x_i
 \end{aligned}$$

2. ESTADO DEL ARTE

Donde:

$$\psi = \sum_{s=1}^l (d_s - y_s) \bar{\varphi}'(y_s^{in}) \bar{w}_{sj} \varphi'(v_j^{in})$$

$$\bar{\psi} = \sum_{s=1}^l \bar{\psi} \bar{w}_{sj} \varphi'(v_j^{in})$$

Así que el algoritmo del error de retro-propagación se define siguiendo las etapas siguientes:

1. Inicializar los pesos (más detalles en 2.1.11.5)
2. Presentar todos los patrones de entrada y calcular todas las activaciones.
3. Calcular los discriminantes $\bar{\psi}$ y ψ y determinar las derivadas parciales del error respecto al peso.
4. Actualizar los pesos con el algoritmo del gradiente descendiente: $w \mapsto w - \lambda \text{grad} \varepsilon$
5. Repetir el proceso hasta llegar al termino de parada (ver 2.1.11.1).

La ecuación 2.9, en la etapa $t + 1$ del aprendizaje, seria:

$$w_{ji}(t+1) = w_{ji}(t) - \lambda \frac{d\varepsilon}{dw_{ji}}(t) \quad (2.10)$$

2.1.10.4 Algoritmo RPROP

Retro-propagación es el algoritmo más utilizado para el aprendizaje supervisado en las redes feedforward. La idea básica del algoritmo de aprendizaje de retro-propagación, (ver 2.1.10), es repetir una cadena de tasas de aprendizajes para el cálculo de la influencia de cada peso en la red respecto a la función del error de la red E .

El cálculo de la deriva $\frac{d\varepsilon}{dw_{ji}}(t)$ de la función del error respecto a los pesos, está dada en 2.1.10. Donde w_{ij} es el peso que relaciona la neurona j a la neurona i . Una vez que se conoce la derivada parcial para cada peso, el objetivo de minimizar la función del error, se consigue mediante la realización del gradiente descendente simple (ver la ecuación 2.10).

Obviamente, la elección de la tasa de aprendizaje λ , con la que se escala la derivada, tiene un efecto importante en el tiempo necesario para alcanzar la convergencia. Si se establece demasiado pequeña, serán necesarios demasiados pasos para llegar a una solución aceptable, por el contrario, una tasa de aprendizaje grande posiblemente dará lugar a una oscilación, evitando que el error caiga por debajo de un cierto valor, como se ha mencionado en 2.1.10, una propuesta para deshacerse de este problema es introducir el moméntum en la ecuación 2.10:

$$\Delta w_{ji}(t) = -\lambda \frac{d\varepsilon}{dw_{ji}}(t) + \mu \Delta w_{ji}(t-1) \quad (2.11)$$

Donde μ representa el momentum, que escala la influencia de las etapas anteriores en la actual. Se espera que la incorporación del momentum haga que el procedimiento de aprendizaje será más estable y acelere la convergencia en las regiones superficiales de la función del error. Sin embargo, en la práctica esto no es siempre cierto. Resulta, en efecto, que el valor óptimo del momentum es igualmente un problema dependiente como lo es la tasa de aprendizaje, y que no lo mejora generalmente (39). También nos encontramos con otro problema que es definir la tasa de aprendizaje adecuada, y el momentum adecuado.

Muchos algoritmos se han propuesto para hacer frente al problema de actualización adecuada de los pesos. Y la mayoría de las estrategias, que se emplearon, se pueden dividir en dos categorías: estrategias globales y locales. Las estrategias globales se basan en el conocimiento del estado de toda la red (por ejemplo, la dirección paso a paso de los pesos anteriores) para modificar los parámetros, mientras que las estrategias locales utilizan la información disponible sólo en el peso específico (por ejemplo, la derivada parcial), para adaptar el parámetro de este peso específico. Se ha demostrado que las estrategias de adaptación locales están más recomendadas que las estrategias globales (40). Algunos ejemplos de estas estrategias se pueden encontrar en (41),(42).

El algoritmo “RPROP” abreviación de “Resilient PROPagation”, propuesto por Martin Riedmiller y Heinrich Braun en (43). El RPROP actualiza el valor del peso w_{ji} directamente, es decir sin tener en cuenta el valor de la derivada parcial. EL algoritmo RPROP tiene diversas implementaciones: RPROP+, RPROP−, iRPROP+, iRPROP− (44). A continuación RPROP+ se define de la manera siguiente:

2.1.11 Fundamentos para la implementación de una RNA

2.1.11.1 Condición de Término

Se pueden implementar muchos criterios para ajustar el proceso de aprendizaje. Las estrategias más comunes consisten por ejemplo en fijar un número máximo de iteraciones, o fijar un error mínimo a alcanzar. Estas dos últimas técnicas, no son muy buenas si no se determinan adecuadamente. Por ejemplo, en la primera que consiste en fijar el número máximo de iteraciones, al implementarla con el algoritmo de retro-propagación, a veces no asegura alcanzar el mínimo global, pero puede converger a mínimos locales.

La segunda técnica, puede sufrir el fenómeno de sobre-aprendizaje, todo dependiendo de si los datos de entrada son buenos (i.e. que puedan representar el fenómeno que estamos aprendiendo o no), y de si los errores de medición, acompañados con los ruidos, representan o no la

2. ESTADO DEL ARTE

Algorithm 1 RPROP

```

1: Descripción del RPROP+.
2: for para cada  $w_{ij}$  do
3:   if  $\frac{\partial E(t-1)}{\partial w_{ij}} \cdot \frac{\partial E(t)}{\partial w_{ij}} > 0$  then
4:      $\Delta_{ij}^{(t)} := \min(\Delta_{ij}^{(t-1)} \cdot \eta^+, \Delta_{\max})$ ;
5:      $\Delta_{ij}^{(t)} := -\text{sign}\left(\frac{\partial E(t)}{\partial w_{ij}}\right) \cdot \Delta_{ij}^{(t)}$ 
6:      $w_{ij}^{(t+1)} := w_{ij}^{(t)} + \Delta w_{ij}^{(t)}$ 
7:   end if
8:   if  $\frac{\partial E(t-1)}{\partial w_{ij}} \cdot \frac{\partial E(t)}{\partial w_{ij}} < 0$  then
9:      $\Delta_{ij}^{(t)} := \max(\Delta_{ij}^{(t-1)} \cdot \eta^-, \Delta_{\min})$ ;
10:     $w_{ij}^{(t+1)} := w_{ij}^{(t)} - \Delta w_{ij}^{(t)}$ ;
11:     $\frac{\partial E(t)}{\partial w_{ij}} := 0$ ;
12:   end if
13:   if  $\frac{\partial E(t-1)}{\partial w_{ij}} \cdot \frac{\partial E(t)}{\partial w_{ij}} = 0$  then
14:      $\Delta_{ij}^{(t)} := -\text{sign}\left(\frac{\partial E(t)}{\partial w_{ij}}\right) \cdot \Delta_{ij}^{(t)}$ 
15:      $w_{ij}^{(t+1)} := w_{ij}^{(t)} + \Delta w_{ij}^{(t)}$ 
16:   end if
17: end for

```

naturaleza del problema a estudiar. Se puede solucionar este problema usando una técnica de validación cruzada “cross-validation”, (usando dos conjuntos independientes de datos para el entrenamiento: uno para el ajuste de los parámetros de la red, y el otro para su validación).

2.1.11.2 Fenómeno de Saturación

Una consideración más a tener en cuenta es la saturación de las neuronas. Las neuronas en la práctica pueden dejar de aprender. Por lo tanto, es importante adoptar medidas que intenten evitar esta situación. Entre ellas cabe citar la normalización de los datos de entrada y la inicialización de los pesos de la primera capa considerando un intervalo ajustado para las entradas. Por ejemplo: $\left[-\frac{1}{\max_q |p^q_i|}, \frac{1}{\max_q |p^q_i|}\right] \forall 1 \leq i \leq r$, donde q es un conjunto de entradas de aprendizaje, r es el número de entradas de la red.

Otra alternativa, es fijar todos los pesos a cero, esto puede arreglar el problema de saturación, pero no es viable porque el origen del espacio de los pesos corresponde a un lugar inestable en la función de error de la red, eso facilita la divergencia del algoritmo de retro-propagación.

También se puede utilizar el concepto de Agrupación, es decir, en vez de modificar (actualizar) los pesos en cada iteración, una enfoque alternativo es acumular las variaciones de los pesos en

un periodo de aprendizaje, y actualizarlos una vez con la media de estas variaciones (“Batching”). Sin embargo, el agrupamiento no es un remedio eficaz, porque ralentiza la convergencia. No obstante el agrupamiento puede contribuir a evitar que el proceso vaya en direcciones malas, y que quede atrapado en mínimos locales inadecuados.

2.1.11.3 Selección de variables

Se presenta a continuación un resumen del procedimiento para descartar variables irrelevantes:

1. Elegir el conjunto de los candidatos de entrada (variables primarias y secundarias).
2. Seleccionar la entrada más correlacionada con la salida, en el espacio de observación, proyectar todas las otras entradas, y la salida, sobre el sub-espacio nulo de la entrada seleccionada.
3. En el espacio nulo de las $m-1$ variables seleccionadas en anteriores iteraciones:
 - (a) Seleccionar el vector de entrada proyectado con más correlación con el vector de salida previsto.
 - (b) Calcular la probabilidad H_m de que la función de probabilidad sea más relevante que una de las m entradas seleccionadas previamente, y compararlo con el riesgo α elegido.
 - (c) Si H_m es menor que el riesgo, volver al paso 3.
 - (d) Si H_m es mayor que el riesgo, ir al paso 4.
4. Utilizar las variables seleccionadas como entradas de una red neuronal y entrenar como se indica en las siguientes secciones.

El primer paso en cualquier procedimiento del diseño de la red neuronal consiste en la reducción de la dimensión del espacio de entrada, haciendo dos preguntas.

- ¿La dimensión del vector de entrada es la más reducida posible, o es posible encontrar una representación de entradas más compacta, preservando al mismo tiempo la cantidad de información relevante?
- ¿Las entradas candidatas son todas relevantes para el modelado del valor objetivo?

La respuesta a la primera pregunta es proporcionada mediante el análisis de los componentes principales, o posiblemente mediante operaciones más complejas, como el análisis de componentes curvilineales o los mapas de auto-organización.

La respuesta a la segunda pregunta es proporcionada por métodos estadísticos, como el método

2. ESTADO DEL ARTE

de función de probabilidad.

Después de realizar la selección de las entradas, los parámetros del modelo se calculan como se explica en la siguiente sección.

2.1.11.4 Normalización de las variables

Antes de empezar el proceso del entrenamiento, las variables de las entradas deben normalizarse y estar centradas: si las entradas tienen diferentes órdenes de magnitud, los valores más pequeños no se tendrán en cuenta durante el entrenamiento. Por lo tanto, para cada vector de entrada V_i , la media μ_i y la desviación estándar σ_i de sus componentes deben ser calculadas, y se pueden presentar las nuevas variables como:

$V'_i = \frac{(V_i - \mu_i)}{\sigma_i}$ (o cualquier otro cambio similar) debe ser calculado; de ahí las nuevas variables están centradas, y su desviación estándar es del orden de 1.

También es recomendable, especialmente para los modelos de entrenamiento dinámicos (redes neuronales recurrentes), centrar y normalizar las salidas de una manera similar.

2.1.11.5 Inicialización de parámetros

Dado que los métodos de entrenamiento anteriores son iterativos, a los parámetros se les deben asignar valores iniciales antes del entrenamiento. Los argumentos siguientes son pautas para la inicialización:

- Los parámetros relacionados con los sesgos (entradas de valor constante e igual a 1) deben ser inicializados a cero, con el fin de asegurar que las neuronas ocultas se inicializan en torno a cero, y luego, si las entradas se han normalizado y se han centrado apropiadamente como se recomendó anteriormente, los valores de las salidas de las neuronas ocultas se normalizan y se centran también.
- Por otra parte, se debe comprobar que los valores de las salidas de las neuronas ocultas no están demasiado cerca de +1 o -1 (de estarlo, se dice que están saturados).
- Esto es importante debido a que el gradiente de la función de coste, que es la fuerza impulsora hacia la minimización durante el entrenamiento, depende de las derivadas de las funciones de activación de las neuronas ocultas con respecto al potencial. Si las salidas de las neuronas ocultas están inicialmente cerca de +1 o -1, las derivadas son muy pequeñas, por lo que el entrenamiento empieza muy lentamente.

Si n es el número de las entradas de la red, cada neurona oculta recibe $n - 1$ variables x_i . Los parámetros distintos de cero deben ser suficientemente pequeños para que el potencial de las neuronas ocultas tenga una variación del orden de 1, con el fin de evitar que los sigmoides entren en saturación. Suponiendo que las entradas x_i pueden ser elegidas aleatoriamente, idénticamente distribuidas, con variables centradas y normalizadas. Los valores iniciales de los parámetros deberían ser extraídos de una distribución centrada, cuya covarianza es desconocida. El parámetro relacionado con la desviación es igual a cero; el potencial de $v = \sum_{i=1}^n w_i x_i$ en cada neurona, la suma de $n - 1$ variables aleatorias que son los productos de variables aleatorias independientes, con media cero, que tiene la misma distribución.

Se puede demostrar, a partir de los elementos de las estadísticas proporcionadas, que se tiene $\text{var}(V) = (n - 1) \text{var}(w_i) \text{var}(x_i)$.

Con $\text{var}(x_i) = 1$, ya que las variables se han normalizado antes del entrenamiento.

Por lo tanto, si la variación deseada del potencial es de 1, los valores iniciales de los parámetros deben ser extraídos de una distribución centrada $\frac{1}{n-1}$.

Por ejemplo, puede ser conveniente elegir una distribución uniforme entre:

$$-w_{\max} \text{ y } w_{\max} : \text{var}(w_i) = \frac{w_{\max}^2}{3}, \text{ entonces: } w_{\max} = \sqrt{\frac{3}{n-1}}.$$

La discusión anterior es válida para perceptrones multicapa. Para las redes de base radial RBF (45) o wavelet (46), el problema de inicialización es más crítico, porque son funciones localizadas, si inicialmente están ubicadas lejos del dominio de interés, o si su extensión (desviación estándar) no es adecuada, en el entrenamiento generalmente se producirá un error. En (47) se presentan detalles más precisos al respecto.

Resumiendo, el procedimiento que se debe utilizar para el entrenamiento de una red neuronal feedforward con un número determinado de entradas y neuronas ocultas es:

- Inicializar los parámetros con el método descrito anteriormente.
- Calcular el gradiente de la función de coste por retro-propagación.
- Actualización de los parámetros iterativamente con un algoritmo de minimización apropiado (descenso del gradiente simple, el método de Broyden-Fletcher-Goldfarb-Shanno (BFGS), Levenberg-Marquardt, el gradiente conjugado, etc).
- Si un número máximo establecido de iteraciones se alcanza, o si la variación del módulo del vector de parámetros es menor que un umbral dado (los pesos ya no cambian significativamente), o si el módulo del gradiente es menor que un determinado umbral (un mínimo

2. ESTADO DEL ARTE

ha sido alcanzado), dar por terminado el procedimiento, en caso contrario, empezar una nueva iteración evaluando el gradiente.

2.1.11.6 Criterio de Parada del Entrenamiento

Como es habitual, el entrenamiento consiste en minimizar iterativamente una función de coste, por ejemplo con la función de mínimos cuadrados, cuyo valor se calcula en un conjunto de entrenamiento.

Si el entrenamiento terminó antes de alcanzar el error mínimo requerido en la función de coste, por lo tanto el proceso de entrenamiento no ha podido ajustarse bien con los datos de entrenamiento. Podría ser conllevado a regularizar el criterio de parada.

La dificultad que surge es: ¿cuándo se debe dejar de entrenar? El método más popular consiste en la supervisión de la variación del error estándar en la predicción del conjunto de validación, y terminar el entrenamiento una vez el error del entrenamiento empieza a aumentar.

2.1.11.7 Selección del modelo

Después de la selección de variables y el modo del entrenamiento, la selección del modelo es el tercer elemento importante en la metodología del diseño.

Suponiendo que varios modelos candidatos han sido entrenados, de los cuales se debe elegir uno. El modelo debe ser lo suficientemente complejo como para encontrar las relaciones deterministas existentes entre los valores que se modelan y los factores que tienen una influencia significativa en ellos, y no estar afectado por el fenómeno de sobreajuste.

En otras palabras, al modelo seleccionado se le debe incorporar el mejor equilibrio entre la capacidad de aprendizaje y la capacidad de generalización: si el modelo aprende demasiado bien, se ajusta con el ruido, por lo tanto, generaliza mal. Ese equilibrio se ha formalizado bajo el término del dilema de la varianza del sesgo (48). Desde un punto de vista teórico, el modelo que se busca es el modelo por el cual la función de coste es mínima. Esta cantidad se puede dividir en dos términos:

- El sesgo [Esto no se debe confundir con la entrada constante de un modelo, también llamada sesgo], que expresa el promedio, en todos los posibles conjuntos de entrenamiento (con todas las posibles realizaciones de las variables aleatorias que modelan el ruido) de la diferencia cuadrática entre las predicciones del modelo y los valores de la función de regresión;

- La varianza, que expresa la sensibilidad del modelo para un conjunto de entrenamiento dado.

Debido a que la función teórica de coste anterior no puede ser calculada, la función empírica de coste de mínimos cuadrados se minimiza durante el entrenamiento. Por lo tanto, un modelo muy complejo, con un gran número de parámetros ajustables, puede tener un sesgo muy bajo, es decir que puede tener la capacidad de ajuste de los datos sea cual sea el ruido presente, pero es candidato a tener una variación muy grande, es decir que depende fuertemente de la realización específica de la presencia de ruido en el conjunto de entrenamiento. A la inversa, un modelo muy simple, con un pequeño número de parámetros ajustables, puede ser insensible a la presencia de ruido en los datos de entrenamiento, pero los modelos simples resultan ser incapaces de aproximar la función de regresión.

Todas las técnicas que se deben emplear deberían tener como objetivo descartar obviamente el sobreajuste, y estimar el error de generalización a fin de encontrar el modelo que tenga el error de generalización más pequeño. Entre ellos, se pueden establecer en dos categorías:

- Métodos globales, que consisten en la estimación del error de generalización: como el método de la validación cruzada;
- Métodos locales mediante los cuales se estima la influencia de cada ejemplo en el modelo: el control local a través de sobreajuste del método de LOCI (49), que se basa en aprovechar la estimación y los intervalos de confianza en las predicciones del modelo.

Es importante señalar que no hay ningún problema en la combinación de métodos locales y globales al mismo tiempo.

Como se discutió anteriormente, el modelo de selección debe basarse en la comparación de los errores de generalización de los modelos candidatos, pero el error de generalización, así como la función de regresión, no se puede calcular: por lo tanto, deben ser estimados.

La idea consiste en la realización de la selección del modelo a base del error cuadrático medio en el conjunto de entrenamiento “The mean square error (MSE)”, $E_T = \sqrt{\frac{1}{N_T} \sum_{i=1}^{N_T} r_i^2}$ donde r_k es el error modelado en la iteración k : $r_k = y_p^k - g(x^k, w)$, y donde la suma se realiza sobre todos los ejemplos N_T del conjunto de entrenamiento. Como se indicó anteriormente, el error modelado en el conjunto de entrenamiento puede ser tan pequeño como se desee con sólo añadir neuronas ocultas, lo cual es perjudicial para la generalización. Así, el valor de E_T no es un criterio de selección adecuado.

2. ESTADO DEL ARTE

2.1.11.8 Validación Cruzada

La validación cruzada es una técnica para estimar el error de generalización de un modelo, a partir de datos que no se utilizan para la estimación de parámetros (durante el entrenamiento) (50). En primer lugar, el conjunto de datos disponibles se divide en D subconjuntos. Entonces, los pasos siguientes se llevan a cabo, para una familia de funciones que tienen la misma estructura (por ejemplo, redes neuronales, con un número determinado de neuronas ocultas):

- Iteración i , a realizarse D veces: construir un conjunto de entrenamiento con $D - 1$ subconjuntos de los datos disponibles; realizar varios entrenamientos, con diferentes valores iniciales de los parámetros, para cada modelo, se calcula el MSE en el conjunto de validación, de los ejemplos restantes, $E_V = \sqrt{\frac{1}{N_V} \sum_{i=1}^{N_V} r_i^2}$ almacenar en la memoria el más pequeño MSE calculado en cada E_{V_i} ;
- Calcular los valores de la validación cruzada de las cantidades E_{V_i} en cada iteración de D : $\sqrt{\frac{1}{D} \sum_{i=1}^D E_{V_i}^2}$

Por ejemplo, si $D = L$ se elige (el proceso se llama L veces validación cruzada), se construyen L diferentes particiones de la base de datos, para cada partición, 80% de los datos que están en el conjunto de entrenamiento y 20% en el conjunto de validación. Como se señaló anteriormente, el resultado de la validación cruzada es la raíz cuadrada de la media del MSE que se calcula en cada partición. Esta media, que se debe realizar en el 20% de la base de datos, puede no ser una muestra estadísticamente significativa de la distribución de todos los ejemplos posibles.

De una manera heurística, el procedimiento puede simplificarse mediante la realización de una única partición en la base de datos. Y la elección de un conjunto de validación que está tan cerca como sea posible a la distribución de los ejemplos disponibles. A tal efecto, se puede estimar la divergencia de Kullback-Leibler (51) entre dos distribuciones calculando la probabilidad p_1 y p_2 :

$$D(p_1 \parallel p_2) = \sum_x p_1(x) \log \left(\frac{p_1(x)}{p_2(x)} \right)$$

Que es una medida de cómo es diferente la probabilidad entre dos distribuciones de probabilidad (durante todo el espacio del mismo evento). Expresado en términos de la esperanza matemática, se obtiene:

$$D(p_1 \parallel p_2) = E_{p_1} \left[\log \left(\frac{p_1(x)}{p_2(x)} \right) \right]$$

Debido a que la expresión no es simétrica, una distancia más satisfactoria se define como:

$$\Delta = \frac{1}{2} (D(p_1, p_2) + D(p_2, p_1))$$

Se llevan a cabo varias particiones aleatorias en la base de datos, y se mantiene la partición para la que la distancia entre el conjunto de validación y el conjunto de entrenamiento sea más pequeña. Esa partición única satisfactoria puede ser utilizada para estimar el error de generalización.

La elección aleatoria de un gran número de particiones y el cálculo de la distancia de Kullback-Leibler es un proceso mucho más rápido que el entrenamiento de un modelo, el tiempo de cálculo se divide aproximadamente por un factor de L , en comparación con L -veces validación cruzada.

Aplicando la suposición de que las distribuciones son dos gaussianas: $p_1(\mu_1, \sigma_1)$ y $p_2(\mu_2, \sigma_2)$, la distancia de Kullback-Leibler se puede expresar como:

$$\Delta(p_1, p_2) = \frac{(\sigma_1^2 - \sigma_2^2) + (\mu_1^2 - \mu_2^2)}{(4\sigma_1^2\sigma_2^2)} (\sigma_1^2 + \sigma_2^2)$$

Ese procedimiento heurístico es muy útil para crear prototipos rápidos de un modelo inicial, que puede perfeccionar convencionalmente la validación cruzada o mediante el concepto de la validación cruzada dejando uno fuera como se explica a continuación.

El diseño del modelo se inicia a partir de modelos más simples (modelo lineal), y la complejidad aumenta gradualmente, para los modelos neuronales, al aumentar el número de neuronas ocultas).

También se puede aumentar el número de las capas ocultas; para problemas de modelado que pueden ser considerados en un segundo paso del diseño: si se ha encontrado un modelo satisfactorio con una capa oculta, se puede tratar de mejorar el rendimiento mediante el aumento del número de capas ocultas, a la vez que se disminuye el número de neuronas por capa. Este procedimiento conduce a veces a una cierta mejora, por lo general marginal.

Cuando se produce sobreajuste en el entrenamiento, el resultado de la validación cruzada se incrementa cuando la complejidad de los modelo aumenta. Por lo tanto, el procedimiento termina cuando el resultado comienza a aumentar, y se elige el modelo que tiene el MSE más pequeño.

2.1.11.9 Validación cruzada dejando uno fuera

La validación cruzada dejando uno fuera (“Leave-one-out cross-validation” LOOCV) es un caso particular de la validación cruzada, de los cuales $D = N$: en la iteración k , por ejemplo, k es

2. ESTADO DEL ARTE

retirada del conjunto de entrenamiento, y los entrenamientos se llevan a cabo (con diferentes valores iniciales en los parámetros) con $N - 1$ ejemplos del conjunto de entrenamiento, para cada modelo, se calcula el error de predicción en el ejemplo retirado k , y el error más pequeño de predicción en el ejemplo retirado, $r_k^{(-k)}$ se almacena.

El error sería:

$$E_t = \sqrt{\frac{1}{N} \sum_{k=1}^N (r_k^{(-k)})^2}$$

Que se calcula de la misma manera explicada en el apartado anterior, y los modelos serán diseñados, hasta que E_t comienza a aumentar y que aumente la complejidad.

El principal inconveniente de esta técnica es que es computacionalmente costosa, pero se puede demostrar que E_t es un estimador no-sesgado del error de generalización (52).

2.1.12 Estimación de los Intervalos de Confianza en la Predicción de una RNA

El uso de técnicas de estimación de la confianza en las salidas de las redes neuronales desempeña un papel importante y se aplican en muchas aplicaciones prácticas. Sin embargo algunas de estas técnicas no siempre tienen sensibilidad suficiente para considerar los niveles de ruido presentes en los datos, por lo que algunos autores han estudiado alternativas para superar esta deficiencia.

En el trabajo (53) se presenta una extensión del método de regresión no lineal para estimar los intervalos de predicción de las redes neuronales feedforward. La idea principal de su método es que la variación de los residuos debe ser estimada en función de los datos de entrada y no como una constante.

El error producido en una predicción mediante una RNA (o cualquier otro modelo empírico) como modelo de un sistema físico dependen en buena medida de la posibilidad de que se produzcan en dicho sistema valores (datos) incontrolables o no observables (54). Cuando las RNAs se utilizan en los problemas de regresión, su fiabilidad y facilidad de uso mejorarán si la precisión de los resultados de la red neuronal es apoyada por algún tipo de medida de confianza (55).

El error cuadrático medio "MSE" es sólo una indicación global de la exactitud del modelo y proporciona una indicación aproximada de la predicción de la red en toda la región donde los datos de entrenamiento están disponibles (56). El buen ajuste en el contexto de los modelos

neuronales puede empeorarse debido al ruido en los datos de entrenamiento o como consecuencia de la densidad de la información en algunas regiones (57), y/o su ausencia en los datos de entrenamiento. Muchos métodos y teorías han sido desarrollados para responder a este problema, teniendo en cuenta que algunas hipótesis deben estar presentes en el modelo, como: suponer que los residuos del modelo son independientes y tienen una distribución normal con media nula y no hay ningún error de observación (54), el número de los entrenamientos tiende a cero y la red neuronal converge.

En la mayoría de los casos prácticos donde si se suele disponer de la verdadera media o varianza de los residuos, pero sólo una muestra finita para poder estimarlos. La media puede ser distinta de cero debido a la falta de correspondencia entre los modelos, y los residuos de estas funciones son a menudo los valores de la entrada del modelo y deben ser modelados.

El histograma se ha presentado como el estimador de densidad no paramétrico sólo hasta la década de 1950, cuando se hacía la estimación de la densidad o la estimación de la densidad espectral. En un trabajo poco conocido de Fix y Hodges en 1951 (58), se introdujo el algoritmo básico de estimación de la densidad no paramétrica. Se abordó el problema de la discriminación estadística, cuando la forma paramétrica de la densidad de muestreo no era conocida. Durante la década siguiente, varios algoritmos generales y formas alternativas de análisis teóricas fueron introducidos por Rosenblatt (1956) (59), Parzen (1962) (60) y Cencov (1962) (61). Después, se ha seguido con muchas publicaciones teóricas importantes principalmente la de Watson y Leadbetter (62), Schwartz (63), Epanechnikov (64), Tarter y Kronmal (65), Wahba (66), etc. La generalización con multi-variables fue introducida por Cacoullos (67). En la década de los 70, tuvieron lugar las primeras aplicaciones prácticas de estos métodos se encuentran en (68) y (69), entre otros.

la utilización del histograma como técnica de representación de datos o como estimador de la densidad, la diferencia básica es que en este último caso debe estar normalizado para integrar 1. La estimación de densidad por núcleos puede ser interesante no sólo como un caso límite del histograma, sino también por otras técnicas. De hecho, prácticamente todos los algoritmos no paramétricos son asintóticamente métodos del núcleo, un hecho demostrado empíricamente por Walter y Blum (70), y Scott (71).

1. Problemas con el uso del histograma

La representación gráfica de un conjunto de datos es una ayuda indispensable para la interpretación de los mismos y su implementación en el entrenamiento en los modelos no lineales, las representaciones gráficas visuales pueden facilitar juicios acerca de la tendencia central, intervalos de confianza, etc. Permiten también adoptar decisiones de si la

2. ESTADO DEL ARTE

distribución de los datos se ajusta a los supuestos que subyacen a los resultados de un modelo lineal / no lineal. Las herramientas más utilizadas por los científicos de análisis para visualizar la distribución de datos de una variable son gráficas con puntos, y para grandes conjuntos de datos, el histograma.

El histograma es fácil de construir y proporciona una idea sobre la distribución de la densidad de los datos, si se utiliza una elección adecuada de las clases. El histograma es una estimación de la distribución de densidad de población. Sin embargo, la impresión visual obtenida a partir de un histograma puede depender en un grado no deseado de los intervalos seleccionados para las clases (es decir, el número y el punto medio de los datos). Una reconstrucción de la densidad de población más consistente que el histograma, por lo tanto, sería bienvenida. La estimación de la densidad del núcleo puede cumplir con este requisito (68, 72).

Observación *En esta sección se revisa brevemente el método de estimación de densidad por núcleo para una variable y el caso de multi-variables.*

Dada una secuencia de variables aleatorias independientes $\{x_1, \dots, x_n\}$ idénticamente distribuidas con una función de densidad $f(x)$, el problema de estimación de la densidad es la construcción de una secuencia de estimadores $\hat{f}(x, h)$ (h es un parámetro llamado “Band-width” ancho de banda de $f(x)$ basado en un muestreo de datos $\{x_1, \dots, x_n\}$).

2. La estimación de la densidad por núcleos para una variable

La estimación de la densidad es una técnica común y útil para realizar un análisis de datos. Dada una secuencia de variables aleatorias independientes de una distribución específica, el problema de estimación de la densidad es la construcción de una función de densidad de la distribución sobre la base de los datos extraídos de la misma.

La estimación de la densidad es un problema muy importante en el análisis numérico, la minería de datos y otros campos de investigación (69). Si se conoce la distribución de la densidad de una muestra del conjunto de datos, se puede tener una idea sobre la distribución que se debe implementar en el conjunto de datos. Así que en muchas aplicaciones de minería de datos, como el muestreo de la densidad y la agrupación basada en la densidad, la estimación de la densidad es un paso inevitable. La estimación de la densidad por núcleos es un método no paramétrico ampliamente estudiado para la estimación de la densidad (73).

La idea detrás del método es la generalización del método de Parzen de estimación mediante el histograma. En el segundo método, la densidad en un punto x se calcula en una

proporción de observaciones $\{x_1, \dots, x_n\}$ que están cerca de h . Para ello, se puede seleccionar una área alrededor de X cuya anchura es controlada por un parámetro de suavizado h , entonces se calcula el número de observaciones que caen en esta área. Esta estimación, que depende del parámetro de suavizado h , tiene buenas propiedades estadísticas, pero no es continua.

El objetivo del método de estimación por núcleos es conseguir la continuidad: para ello, se sustituye el área centrada en X y de ancho h , por una curva de campana centrada en X . Cuantas más observaciones se encuentran cerca del punto X a lo largo de la curva de campana se le asigna un valor numérico grande, por el contrario, se le asigna un valor numérico insignificante. El estimador está formado por la suma (o más bien la media) de las curvas de campana.

La idea simple de la estimación por núcleos, se basa en que para cada punto de los datos $x_i, i \in \{1, 2, \dots, n\}$ se sustituye por una distribución especificada (por lo general normal), centrada en el punto y con una desviación estándar designada h . Las distribuciones normales se suman y la distribución resultante se escala para tener una unidad de superficie (es una curva suave), y la estimación de la densidad del núcleo (Parzen), se define en forma general como:

$$\hat{f}(x, h) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right)$$

Donde K es el núcleo: una función simétrica, pero no negativa, que integra a 1: $\int K(x)dx = 1$, y $h > 1$ es un parámetro de suavizado llamado ancho de banda (band-width).

Un núcleo con el subíndice h se llama el núcleo de escala y se define como:

$$K_h(x) = \frac{k\left(\frac{x}{h}\right)}{h}$$

. Intuitivamente, uno quiere elegir h tan pequeño como los datos lo permitan, sin embargo siempre hay un equilibrio entre el sesgo del estimador y su varianza.

Una gama de funciones del núcleo son de uso general: uniforme, triangular, Epanechnikov, normal, El núcleo de la normal que se usa a menudo:

$$K(x) = \varphi(x)$$

donde φ es la función estándar de densidad normal.

La estimación por núcleo, cuando se calcula con un valor apropiado de h , da una buena estimación de la función de densidad en los datos, sin hacer ninguna hipótesis, por ejemplo

2. ESTADO DEL ARTE

cuando es una distribución normal. Esto es útil en los ejemplos de análisis científicos, donde la desviación de la normalidad es común. Los cálculos se pueden programar fácilmente. La única complicación es la determinación de un valor adecuado para h . Esta elección es un contexto específico y se le requiere experiencia.

Obviamente, la desventaja de este método es que es necesario obtener los n datos y mantenerlos en la memoria antes de hacer la estimación. Cuando los datos se presentan en forma de flujo continuo de datos, el gran volumen y la infinitud de datos hacen que sea imposible obtener todos los datos con antelación y guardarlos en la memoria. Es decir, la estimación de la densidad del núcleo sólo se puede calcular con conjuntos de datos estáticos, y no puede manejar flujos de datos.

3. Función de densidad por núcleos para multi-variables

El análisis teórico de los estimadores por núcleo para multi-variables es el mismo que el de una variable a excepción de algunos detalles.

Dado un conjunto de n datos p -dimensionales $x_i \in \mathbb{R}^P, i \in \{1, 2, \dots, n\}$.

La estimación de densidad mediante un núcleo fijo y de ancho de banda fijo para multi-variables (“Multivariate fixed with kernel density estimator”), con una función de núcleo fija φ (función de núcleo global) y de parámetro h , es definida por:

$$\hat{f}(x) = \frac{1}{nh^P} \sum_{i=1}^n \varphi\left(\frac{1}{h}(x - x_i)\right)$$

Donde:

$$\varphi(x) \geq 0, \int_{-\mathbb{R}^P}^{\mathbb{R}^P} \varphi(x) dx = 1$$

La función más utilizada es el núcleo Gaussiano:

$$\varphi(x) = \frac{1}{\sqrt{P} \sqrt{2\pi}} \exp\left(-\frac{1}{2} x^T x\right)$$

Normalmente, los datos observados no están igualmente extendidos en todas las direcciones. Por lo tanto, es muy deseable pre-escalar los datos para evitar estas diferencias. Un enfoque atractivo es la primera esfera (74) (mediante una transformación lineal con media cero y matriz de covarianza como unidad), es decir:

$$x \rightarrow z = S^{\frac{1}{2}} (x - E[x])$$

$E[x]$: La esperanza, que se calcula a través de la media de la muestra. Y $S \in \mathbb{R}^{P \times P}$ matriz de covarianza, es decir:

$$S = E[x - E[x]]$$

Entonces:

$$\begin{aligned} \hat{f}(z) &= \frac{1}{nh^P} \sum_{i=1}^n \varphi\left(\frac{1}{h}(z - z_i)\right) \\ \Leftrightarrow \hat{f}(x) &= \frac{1}{nh^P \det(S)^{\frac{1}{2}}} \sum_{i=1}^n \varphi\left(\frac{1}{h}S^{-\frac{1}{2}}(x - x_i)\right) \end{aligned}$$

Un ancho de banda óptimo del núcleo h^* se puede determinar a través de la minimización del error medio cuadrado (68). Por ejemplo usando la función Gaussiana, se propuso en (68), para la estimación de los datos distribuidos normalmente con la covarianza por unidad h^* es decir:

$$h^* = \left(\frac{4}{(2p+1)n}\right)^{\frac{1}{p+4}}$$

O también podemos usar el siguiente valor:

$$h^* = \left(\frac{4}{3}\right)^{\frac{1}{5}} \sigma n^{-\frac{1}{5}} \sim 1.06 \sigma n^{-\frac{1}{5}}$$

Esta aproximación se conoce como la aproximación con la distribución normal, la aproximación de Gauss, o la regla empírica de Silverman (68).

2.1.12.1 Estimación con Mínimos Cuadrados

1. Mínimos Cuadrados No Lineales:

Los mínimos cuadrados no lineales representan una forma de mínimos cuadrados especializados en la estimación de modelos no lineales, con m parámetros a partir de n observaciones ($n > m$).

Suponiendo que hay n observaciones $(x_i, y_i), i \in \{1, 2, \dots, n\}$, a partir de un modelo no lineal con una relación funcional conocida f , Así que:

$$y_i = f(x_i, \theta^*) + \varepsilon_i, i \in \{1, 2, \dots, n\}$$

Donde: ε es el error asociado con la función de modelar el sistema, que asumimos que es independiente e idénticamente distribuido con una varianza σ^2 , donde su distribución tiene la forma $N(0, \sigma^2)$, entonces $E[\varepsilon_i] = 0$, y $\text{var}(\varepsilon_i) = \sigma^2$. x_i es un vector $k \times 1$. θ^* es

2. ESTADO DEL ARTE

el valor correcto de θ .

Se define $\hat{\theta}$ como el mínimo de la suma del error al cuadrado:

$$\hat{\theta} = \min_{\theta} S(\theta)$$

$$S(\theta) = \sum_{i=0}^n [y_i - f(x_i, \theta)]^2$$

(Es importante señalar que esta suma tendrá mínimos locales a parte del mínimo absoluto $\hat{\theta}$)

Se indicara más adelante, que bajo ciertas suposiciones regulares, que: $\hat{\theta}$ y:

$$S^2 = \frac{S(\hat{\theta})}{(n-p)}$$

son estimaciones coherentes respectivamente de θ^* y σ^2 . Con otra condición de regularidad $\hat{\theta}$ es asintóticamente una distribución normal cuando $n \rightarrow \infty$. Además si ε_i tiene una distribución normal, entonces $\hat{\theta}$ es un estimador máximo de vecindad (75).

Como $\hat{\theta} = \min_{\theta} S(\theta)$, entonces:

$$\left. \frac{\partial S(\theta)}{\partial \theta_r} \right|_{\hat{\theta}} = 0, \forall r \in \{1, 2, \dots, p\}$$

Notaciones:

$$f_i(\theta) = f(x_i, \theta), f(\theta) = (f_1(\theta), \dots, f_n(\theta))^T$$

$$F(\theta) = \frac{\partial f(\theta)}{\partial \theta} = \left[\frac{\partial f_i(\theta)}{\partial \theta_j} \right]_{i,j}, \forall i \in \{1, 2, \dots, n\}, j \in \{1, 2, \dots, p\}$$

Sabiendo que:

$$S(\theta) = \sum_{i=0}^n [y_i - f(x_i, \theta)]^2$$

$$S(\theta) = \sum_{i=0}^n [y_i - f_i(\theta)]^2$$

$$S(\theta) = [y - f(\theta)]^T [y - f(\theta)]$$

$$S(\theta) = \|y - f(\theta)\|^2 \quad (2.12)$$

Y por lo tanto:

$$\left. \frac{\partial S(\theta)}{\partial \theta_r} \right|_{\hat{\theta}} = 0 \Rightarrow -2 \sum_{i=0}^n [y_i - f(x_i, \theta)] \left. \frac{\partial f_i(\theta)}{\partial \theta_j} \right|_{\hat{\theta}} = 0, \forall j \in \{1, 2, \dots, p\}$$

$$\begin{aligned}\Rightarrow F(\hat{\theta})^T (y - f(\hat{\theta})) &= 0 \\ \Rightarrow F(\hat{\theta})^T \hat{\varepsilon} &= 0\end{aligned}$$

Recordando que $y = f(\hat{\theta}) + \hat{\varepsilon}$.

La matriz idempotente: $\hat{P}_F = \hat{F}(\hat{F}^T \hat{F})^{-1} \hat{F}^T$, se puede reescribir la ecuación anterior de la forma siguiente:

$$\hat{F}^T \hat{\varepsilon} = 0$$

. Es decir:

$$\hat{P}_F \hat{\varepsilon} = 0$$

La ecuación 2.12 se llama la ecuación normal del modelo no lineal, normalmente es difícil de resolver analíticamente, y es necesario usar métodos iterativos (75).

2. Aproximación lineal:

Según la extensión de Taylor de primer orden, se puede escribir:

$$f_i(\theta) \approx f_i(\theta^*) + \sum_{r=0}^p \frac{\partial f_i(\theta)}{\partial \theta_r} \big|_{\theta^*} (\theta_r - \theta_r^*)$$

Es decir:

$$f(\theta) \approx f(\theta^*) + F(\theta - \theta^*) \quad (2.13)$$

Según la ecuación 2.13: $S(\theta) \approx \|y - f(\theta^*) - F(\theta - \theta^*)\|^2$

Se anota: $z = y - f(\theta^*) = \varepsilon$ y $\beta = \theta - \theta^*$, entonces:

$$S(\theta) \approx \|z - F\beta\|^2$$

Por lo tanto si:

$$S(\hat{\theta}) = 0 \Rightarrow \|z - F\hat{\beta}\|^2 = 0 \Rightarrow (z - F\hat{\beta}) = 0 \Rightarrow F^T z - F^T F \hat{\beta} = 0$$

Entonces:

$$\hat{\beta} = (F^T F)^{-1} F^T z$$

Cuando $\theta = \hat{\theta}$, se obtiene:

$$f(\hat{\theta}) - f(\theta^*) \approx F(\hat{\theta} - \theta^*) = F(F^T F)^{-1} F^T \varepsilon = P_F \varepsilon$$

Por otro lado:

$$y - f(\hat{\theta}) \approx y - f(\theta^*) - F(\hat{\theta} - \theta^*)$$

2. ESTADO DEL ARTE

$$\begin{aligned}
&= \varepsilon - F(\hat{\theta} - \theta^*) \\
&= \varepsilon - P_F \varepsilon \\
&\Rightarrow (n-p)s^2 = S(\hat{\theta}) \\
&= \|y - f(\hat{\theta})\|^2 \\
&= \|(I_n - P_F)\varepsilon\|^2 \\
&(n-p)s^2 = \varepsilon^T (I_n - P_F) \varepsilon
\end{aligned}$$

Del mismo modo:

$$\begin{aligned}
\|f(\hat{\theta}) - f(\theta^*)\|^2 &= \|P_F \varepsilon\|^2 = \varepsilon^T P_F \varepsilon \\
\hat{\theta} - \theta^* &\approx (F^T F)^{-1} F^T \varepsilon \Rightarrow \varepsilon^T F = (\hat{\theta} - \theta^*)^T (F^T F)
\end{aligned}$$

Así:

$$\begin{aligned}
\varepsilon^T P_F \varepsilon &= \varepsilon^T F (F^T F)^{-1} F^T \varepsilon \\
\varepsilon^T P_F \varepsilon &= (\hat{\theta} - \theta^*)^T (F^T F) F (F^T F)^{-1} F^T F (\hat{\theta} - \theta^*) \\
\varepsilon^T P_F \varepsilon &= (\hat{\theta} - \theta^*)^T F^T F (\hat{\theta} - \theta^*)
\end{aligned}$$

Entonces:

$$\begin{aligned}
S(\theta^*) - S(\hat{\theta}) &= \varepsilon^T \varepsilon - \varepsilon^T (I_n - P_F) \varepsilon \\
&\Rightarrow S(\theta^*) - S(\hat{\theta}) = \varepsilon^T P_F \varepsilon \\
&\Rightarrow S(\theta^*) - S(\hat{\theta}) = (\hat{\theta} - \theta^*)^T F^T F (\hat{\theta} - \theta^*)
\end{aligned}$$

Ahora según el teorema siguiente:

Teorema 17 Dado $\varepsilon \sim N(0, \sigma^2)$, y las condiciones de regularidad, para n grande, tenemos:

- (a) $\hat{\theta} - \theta^* \sim N_P(0, \sigma^2 C^{-1})$, $C = F^T F = F^T(\theta^*) F(\theta^*)$
- (b) $(n-p) \frac{s^2}{\sigma^2} \sim \varepsilon^T (I_n - P_F) \frac{\varepsilon}{\sigma^2} \sim \chi_{n-p}^2$
- (c) $\hat{\theta}$ es estáticamente independiente de S^2 y
- (d) $\frac{\frac{S(\theta^*) - S(\hat{\theta})}{p}}{\frac{S(\hat{\theta})}{n-p}} \approx \frac{\varepsilon^T P_F \varepsilon}{\varepsilon^T (I_n - P_F) \varepsilon} \frac{n-p}{p}$

2.1.12.2 Intervalos de Confianza en la predicción de la red neuronal

La salida de la RNA satisface un modelo de regresión no lineal: $y = f(x, \theta^*) + \varepsilon$, x representa el vector de las entradas, y el vector de salida, θ^* es el vector de los parámetros correctos, ε es el vector de los errores asociados a la modelización de la red neuronal después del entrenamiento.

Sea $\hat{\theta}$ el vector de los parámetros obtenidos después del entrenamiento (resultado de la minimización por mínimos cuadrados de la función del error):

$$\hat{\theta} = \min_{\theta} S(\theta) = \min_{\theta} \sum_{i=1}^n (y_i - f(x_i, \theta))^2$$

Entonces la predicción de la RNA para una entrada x_0 es:

$$\hat{y}_0 = f(x_0, \hat{\theta})$$

Se asume que ε es independiente y normalmente distribuido con media cero.

El porcentaje $100(1 - \alpha)$ de confianza del intervalo del valor predicho \hat{y}_0 es: $\hat{y}_0 \pm c$, donde c es calculado de la forma siguiente (54):

$$c = t_{n-p}^{\frac{\alpha}{2}} s \left(1 + f_0^T (F^T F)^{-1} f_0 \right)^{\frac{1}{2}}$$

Donde $t_{n-p}^{\frac{\alpha}{2}}$ es la inversa de la distribución t-student con $n - p$ grados de libertad, evaluado a $\frac{\alpha}{2}$, p es el número de los parámetros de la red, $s^2 = \frac{s(\hat{\theta})}{n-p}$

El vector f_0 es dado por:

$$f_0 = \left[\frac{\partial f(x_0, \theta^*)}{\partial \theta_1^*}, \dots, \frac{\partial f(x_0, \theta^*)}{\partial \theta_p^*} \right]^T$$

Siendo F la matriz Jacobiana:

$$F = \begin{bmatrix} \frac{\partial f(x_1, \hat{\theta})}{\partial \theta_1} & \cdot & \cdot & \cdot & \frac{\partial f(x_1, \hat{\theta})}{\partial \theta_p} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \frac{\partial f(x_n, \hat{\theta})}{\partial \theta_1} & \cdot & \cdot & \cdot & \frac{\partial f(x_n, \hat{\theta})}{\partial \theta_p} \end{bmatrix}$$

En (56), se incorporó la influencia de la distribución de los datos de entrenamiento en la estimación del intervalo de confianza, y utilizó el método wavelet para estimar la densidad de los datos de entrenamiento, y modificó el intervalo de confianza como:

$$C_s = \frac{2c}{1 + \frac{\rho}{\rho_{\max}}} \quad (2.14)$$

2. ESTADO DEL ARTE

Donde ρ es la densidad del dato, y ρ_{\max} es la densidad máxima de los datos.

En secciones anteriores se ha usado la función de error como suma de cuadrados, motivada principalmente por la simplicidad analítica. Hay muchas otras opciones posibles para la función de error, que también pueden ser consideradas, dependiendo de la particularidad de la aplicación.

2.1.13 Red Neuronal de Base Radial “RBF”

El diseño de una red neuronal puede ser visto como un problema de ajuste de curvas. De acuerdo con este punto de vista, el aprendizaje es equivalente a encontrar una superficie en un espacio multidimensional que proporciona un mejor ajuste de los datos de entrenamiento, mientras que la generalización es equivalente a la utilización de la superficie multidimensional capacitada para interpolar los datos de test.

Teorema 18 (Teorema de Cover) *Un problema de clasificación de patrones complejos emitidos en un espacio no lineal de dimensión grande es más probable que sea linealmente separable en un espacio con una dimensión pequeña.*

El teorema de Cover nos dice que se puede mapear el espacio de entrada a un espacio altamente dimensional, en el que se puede encontrar una función lineal.

La Teoría de Regularización de Tikhonov afirma que: en el contexto de un problema de la reconstrucción en un espacio de gran dimensión, la idea básica de regularización es estabilizar la solución por medio de alguna función auxiliar no negativa que incorpora información previa acerca de la solución. La Teoría de Regularización de Tikhonov nos dice que no podemos simplemente confiar en los datos de entrenamiento, sino que se deben introducir algunas limitaciones, por ejemplo, la función de suavizado.

Las redes RBF son redes feedforward entrenadas usando un algoritmo de entrenamiento supervisado. Se configuran normalmente con una sola capa oculta cuya función de activación se selecciona en una clase de funciones llamadas funciones básicas. En realidad, las RBF transforman el espacio de entrada en un espacio de gran dimensión de una manera no lineal, permitiendo encontrar una aproximación con un ajuste de curvas en un espacio de gran dimensión. La función de activación depende de la distancia euclídea entre los vectores de entrada y de salida. Una red RBF puede ser descrita por:

$$F(x) = \sum_{i=1}^n w_i f(\|x - x_i\|)$$

Donde f es la función de activación, la función gaussiana es la más usada:

$$f(x) = e^{\left(\frac{-\|x-c\|^2}{2\sigma^2}\right)}$$

c es el centro de la función de base, y σ es el parámetro de amplitud que controla cómo la curva se expande hacia fuera.

La capa de salida lleva a cabo una aplicación lineal desde el espacio oculto al espacio de salida. La salida de la red es la suma ponderada de las salidas de las neuronas ocultas. Un conjunto de datos de entrenamiento con sus correspondientes salidas se dan para realizar un aprendizaje supervisado:

$$\begin{bmatrix} f_{11} & \cdots & f_{1n} \\ f_{21} & \cdots & f_{2n} \\ \vdots & \ddots & \vdots \\ f_{n1} & \cdots & f_{nn} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix}$$

Es decir:

$$w = \phi^{-1}d$$

En la práctica, no se calcula la ϕ^{-1} , y se suelen aplicar dos etapas para el entrenamiento de las RBF, seleccionando los centros aleatoriamente o sobre la base de algunos criterios, se actualizan los pesos, típicamente usando un algoritmo de estimación lineal.

2.1.14 Máquinas de Soporte Vectorial “MSV”

Dado n muestras de entrenamiento $\{(x_1, y_1), \dots, (x_n, y_n)\}$ de un problema de dos clases, donde $x \in \mathbb{R}^n$, y $y \in \{-1, 1\}$, puede existir una estructura de subdivisiones de funciones de decisión: $\{f_\alpha(x) : \alpha \in V\}$, donde $f_\alpha : \mathbb{R}^m \rightarrow \{-1, 1\}$.

Sin embargo, la función óptima debe ser aquella con la que se puede lograr el margen máximo $\frac{2}{\|w\|}$.

El objetivo de MSV es reducir al mínimo w para obtener el margen máximo.

La solución óptima f_α se supone que proporciona el menor valor posible para el riesgo esperado $R(f_\alpha)$:

$$R(f_\alpha) = \int |f_\alpha - y| P(x, y) dx dy$$

Como no se conoce la distribución de probabilidad $P(x, y)$, el riesgo empírico $R_{empirico}(f_\alpha)$ se calcula en todas las redes neuronales convencionales, de la forma siguiente:

$$R_{empirico}(f_\alpha) = \frac{1}{n} \sum_{i=1}^n \frac{1}{2} |f_\alpha(x_i) - y_i|$$

Sin embargo, in (52) Vapnik-Chervonenkis indica la teoría de dimensión VC, que indica que para un número máximo de puntos de entrenamiento h que se pueden separar mediante un

2. ESTADO DEL ARTE

conjunto de funciones lineales f_α , existe un límite del riesgo esperado:

$$R(f_\alpha) \leq R_{empirico}(f_\alpha) + \sqrt{\frac{h \left(\ln \frac{2n}{h} + 1 \right) - \ln \frac{\eta}{4}}{n}}$$

Con la probabilidad $1 - \eta$, el segundo término es el ancho del intervalo de confianza.

Para un número fijo de muestras n , el error de entrenamiento (riesgo empírico) disminuye monótonamente a medida que la capacidad o la dimensión h aumenta, mientras que el intervalo de confianza aumenta monótonamente. En consecuencia, tanto el riesgo y la generalización del error, convergen a un mínimo. Por lo tanto, nuestro objetivo es encontrar una red neuronal de tal manera que la disminución de la dimensión VC se produce a costa del aumento más pequeño posible en el error de entrenamiento.

De manera similar en una red RBF, el espacio de entrada también se asigna a un espacio suficientemente grande en función de MSV, por lo que los patrones se convierten a patrones linealmente separables, por lo que un perceptrón simple en el espacio de características puede hacer la clasificación. Los MSVs son básicamente tipos de clasificadores diferentes que han atraído mucha atención últimamente, debido a la novedad de los conceptos que se llevan a cabo en el reconocimiento de patrones, mediante su base matemática y sus excelentes resultados en la práctica.

2.1.15 Comparación entre las redes neuronales multicapas (MLP) y las redes de base radial (RBF)

En (23) hay una sección detallada de una serie de comparaciones entre MLP y RBF. A continuación se presentan algunos puntos esenciales:

1. Las redes del perceptrón multicapas (MLP) y las RBF desempeñan papeles muy similares, ambas proporcionan técnicas para aproximar las asignaciones funcionales no lineales arbitrarias entre un vector x de entrada y un conjunto de salidas y .
2. Las conexiones de unidades ocultas del MLP dependen de la suma lineal ponderada de las entradas, transformadas mediante una función de activación sigmoide. Por el contrario, las unidades ocultas en una red RBF utilizan la distancia de un vector prototipo seguido por su transformación con una función de base.
3. El MLP forma una representación distribuida en el espacio de los valores de activación en las unidades ocultas, ya que para un patrón de entrada dado, muchas unidades ocultas

contribuyen en la determinación del valor de salida. Las representaciones de las unidades ocultas ajustadas durante el entrenamiento combinan los pesos de la capa final, y generan las salidas correctas para un rango de posibles valores de entrada. Sin embargo, una red RBF forma una representación en el espacio de las unidades ocultas, que es local respecto al espacio de entrada, debido a que para un patrón de entrada dado, sólo unas pocas unidades ocultas tendrán activaciones significativamente distintas a cero.

4. Todos los parámetros en un MLP suelen ser determinados al mismo tiempo como parte de una estrategia global de entrenamiento con un aprendizaje supervisado. La correcta aplicación de esta estrategia requiere un entrenamiento para la modificación de los pesos, un conjunto de validación para decidir cuándo hay que dejar de entrenarse y un conjunto de test para evaluar el desempeño de su generalización. Las redes RBF sólo requieren conjuntos de entrenamiento y test. Los datos de entrenamiento se usan sin etiquetas en la primera etapa de aprendizaje y luego los mismos datos (o un subconjunto de ellos, si no se etiquetan todos los datos de entrada) se utiliza para ajustar los pesos en la segunda capa. Dado que es una tarea de optimización lineal, existe un mínimo único para ese conjunto de datos, debido a las posiciones y los anchos del centro elegido. Por lo tanto una red RBF sólo requiere un conjunto para el entrenamiento y otro de prueba para caracterizar plenamente su desempeño.

2.2 Comportamiento de Incendios Forestales

Un incendio forestal es “un fuego que, cualquiera que sea su origen, se propaga sin control en terrenos rurales a través de vegetación leñosa, arbustiva o herbácea, ya sea viva o muerta”. Esto quiere decir que se queman ecosistemas constituidos por bosques, matorrales y/o pastos. Hoy en día, los incendios forestales no sólo son un problema del bosque, sino también se han transformado en un problema de gran impacto social, con capacidad para comprometer el orden y la seguridad interna de los ciudadanos. En cada época de verano, como resultado adicional de los incendios forestales se destruyen casas, construcciones y se pierden vidas tanto humanas como de animales. Por lo general, la población más afectada es la de escasos recursos que habita en áreas rurales y en la interfaz ciudad – bosque/matorral.

Para que se genere un incendio se deben combinar tres elementos básicos: Calor, Oxígeno y Material Combustible. Si se suprime algún elemento de este triángulo, no hay fuego. El calor es aportado por una causa inicial, esta puede ser natural (rayos o erupciones volcánicas) o generada por el hombre (fogatas, cigarros, fósforos y/o quemas de residuos agrícolas o forestales), o producto del propio incendio; el oxígeno del aire, el viento y el combustible constituido por la vegetación bajo condiciones de escasez de lluvias, o bajo contenido de humedad y alta temperatura ambiental se vuelve altamente inflamable.

El oxígeno es necesario para cualquier tipo de combustión, y el aire que nos rodea contiene oxígeno suficiente para mantener el fuego. Por tanto el movimiento del aire dentro y alrededor del fuego activo puede acelerar o retrasar el proceso de combustión.

El comportamiento del incendio se describe determinado cómo y donde empieza el incendio, la velocidad de su propagación, qué cantidad de calor libera y cuánto combustible consume. Entender el comportamiento de los incendios forestales es esencial para su control. Sin ese conocimiento es prácticamente imposible determinar el conjunto de acciones viables y adoptar las mejores decisiones para conseguir primero controlarlos y finalmente extinguirlos. En la figura 2.12, se presenta el modelo conceptual definido por Burrows (76) para la gestión de incendios forestales, donde cita las entradas, modelos con sus salidas para la gestión de los incendios forestales. A continuación se realiza una revisión de los modelos de comportamiento de incendios forestales más ampliamente usados.

2.2 Comportamiento de Incendios Forestales

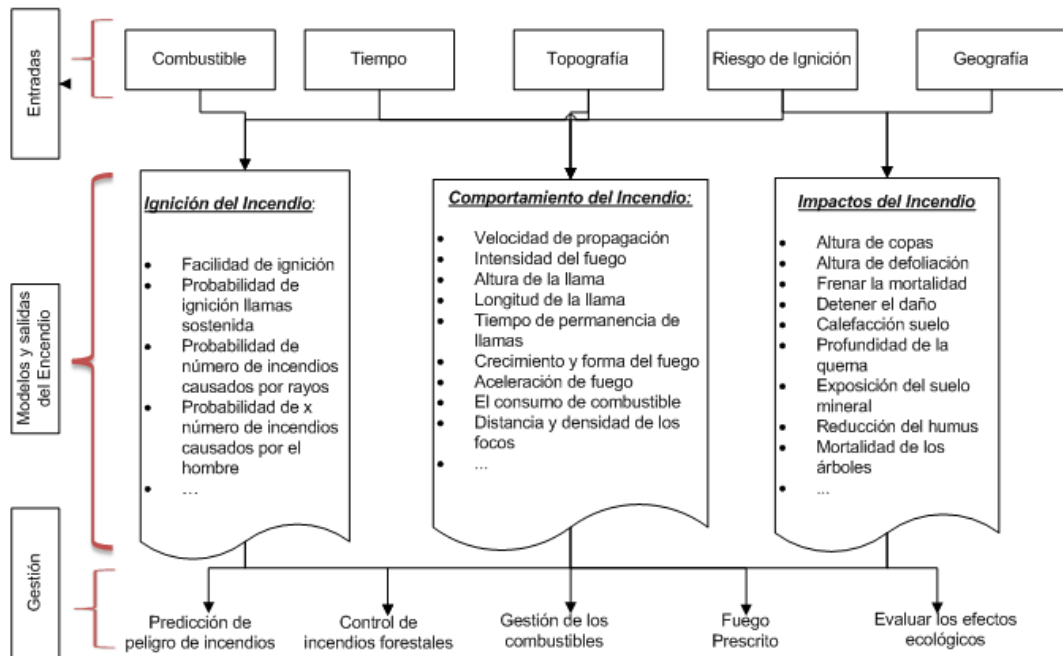


Fig. 2.12: Modelo conceptual para la gestión de incendios forestales (Burrows 1994)

2.2.1 Entorno del Incendio

Los incendios forestales ocurren dentro de un rango específico de condiciones descritas como el ambiente del incendio. En general, hay tres componentes o factores que componen el ambiente del incendio:

1. Combustible
2. Clima
3. Topografía

La carga de combustible es la cantidad de combustible disponible para quemar. Por lo general, se mide en términos de toneladas de combustible seco por unidad de superficie (acre, metro cuadrado, etc.). La carga de combustible incluye tanto la biomasa de la planta viva como de la muerta, pero sólo se debe estimar la que está disponible para contribuir al fuego, que no tiene porque coincidir necesariamente la biomasa total.

El viento y la humedad del combustible fino muerto son los principales determinantes de la tasa de propagación de un incendio. Si no hay viento el fuego va “hacia arriba” y el calor del fuego

2. ESTADO DEL ARTE

también va hacia arriba. Si el viento se sostiene con suficiente velocidad sobre una dirección constante, se pueden llegar a “doblar” las llamas en la dirección que sopla, lo que contribuye a dañar a los árboles.

La inclinación de la llama hace que la distinción entre la altura de la llama y su longitud se vuelva crítica. La altura de la llama es la distancia vertical desde la punta de la llama hacia abajo al centro de la superficie del combustible, mientras que la longitud de la llama es la distancia desde el punto medio de la cama de combustible a la punta de la llama. Sin viento, la longitud de la llama y su altura son iguales porque la llama es “hacia arriba”. Cuando la llama se “inclina”, la longitud de la llama será mayor que la altura. El punto medio del lecho de combustible se utiliza en lugar de la tierra porque la llama no siempre se extiende hasta el suelo.

Otro término que se aplica al viento respecto al fuego es la “velocidad del viento en el medio de la llama”. Durante un incendio es frecuente que tal velocidad sea estimada por un experto mediante “inspección visual”.

2.2.2 Sistemas de evaluación del peligro de los incendios

El peligro de incendio se determina en función de diversos factores, entre los más importantes se pueden citar: La facilidad de ignición de la vegetación, la dificultad para controlar el incendio una vez que se declare y la evaluación del daño puede hacer. Los denominados “*Sistemas de Clasificación de Peligro de Incendio*” (SCPI) producen indicadores cualitativos y / o cuantitativos del potencial del incendio, basándose para ello en los factores antes citados como componentes del ambiente del incendio: Combustibles, Clima y Topografía. Estos indicadores permiten evaluar con objetividad la eficiencia de las medidas de prevención existentes, así como identificar las zonas donde es necesario actuar tanto de forma preventiva (ejemplos: reducción de combustibles) como correctiva (ejemplo: decidir la asignación de recursos durante el desarrollo de los incendios).

Los SCPI son herramientas imprescindibles para la prevención y el control de los incendios. El desarrollo de tales sistemas requiere una gran inversión económica y de tiempo, acompañada casi siempre de rigurosos e intensivos proyectos de investigación rigurosa e intensiva. A continuación se presenta una breve descripción de los sistemas de prevención y control de incendios forestales más conocidos.

2.2.3 Sistemas de un solo índice

1. Índice de Monte Alegre o de Soarez

El índice FMA de Monte Alegre fue desarrollado para ecosistemas húmedos del sudeste de Brasil. Su ecuación básica es la siguiente:

$$FMA = I_{acum} * F + 100 * \sum \frac{1}{H}$$

Donde H la humedad relativa, F es el factor de corrección en base a la precipitación, I_{acum} es la sumatoria de los índices de los días anteriores.

2. Índice Angstrom

Es un índice de origen sueco, muy básico del peligro de incendios. Utiliza sólo la temperatura y la humedad relativa.

Los parámetros de entrada del índice de Angstrom I son: T temperatura del aire y R la humedad relativa.

$$I = \left(\frac{R}{20} \right) + \left(\frac{(29 - T)}{10} \right)$$

Que ha sido aplicado en los trabajos (77) y (78).

3. Índice de humedad del combustible (FMI)

El índice de humedad del combustible (FMI) es muy básico, pero demostró en un estudio presentado por su desarrollador Sharpes, su buen comportamiento en comparación con otros índices de humedad del combustible. Se utiliza la temperatura del aire y la humedad relativa para calcular un índice numérico de peligro de incendio. No existe una clasificación de riesgo de incendios para este índice. Cuanto menor sea el número, mayor es el riesgo de incendio, más información en (77).

Los parámetros de entrada:

T temperatura del aire y R la humedad relativa. Y se calcula de manera siguiente:

$$FMI = 10 - 0.25 * (T - R)$$

4. Índice de peligro de incendio forestal (FFDI)

El Índice de Peligro de Incendios Forestales (FFDI) fue desarrollado en la década de 1960 por el científico de AG McArthur para medir el grado de peligro de los incendios en los bosques australianos. El índice combina un registro de la sequedad, con base a las precipitaciones, la evaporación, y variables meteorológicas como la velocidad del viento, temperatura y humedad, más información en (79).

2. ESTADO DEL ARTE

5. Índice de Nesterov

El algoritmo de este índice fue desarrollado en Rusia. El índice de Nesterov es un índice acumulativo que suma el valor actual a los valores estimados en los últimos días hasta que haya una cierta cantidad de precipitación. Posteriormente se establece en cero. Este índice de peligro de incendios se utiliza actualmente también en Austria.

Los parámetros de entrada son:

V = vapor de saturación, D = punto de rocío, R = Humedad relativa del aire, T = temperatura, y se calcula de la manera siguiente:

$$P = \sum_{i=1}^W (T_i - D_i) \cdot T_i$$

P = índice de Nesterov y W = el número de días con precipitaciones $> 4mm$.

6. Índice de WBKZ

Este índice se utiliza en la parte noreste de Alemania. Fue desarrollado para mejorar la previsión del riesgo de incendios en las masas puras de pino. La base para el índice es muy similar al índice de Nesterov. Se tiene en cuenta la temperatura y la saturación de vapor y calcula un índice del incendio que se acumula cada día.

Los parámetros de entrada son:

V = vapor de saturación y T = temperatura.

$$WBKZ = \sum_{15.2}^{30.9} (T + 10) \cdot V$$

7. Índice de Thornthwaite “PET”

El cálculo del potencial de evapotranspiración “PET” se define como la cantidad de evaporación que se produciría si esta disponible suficiente agua. Se utiliza principalmente para la caracterización de paisajes y su régimen de agua. Como índice operacional de pronóstico de peligro de incendios diario no puede ser utilizado, ya que utiliza la temperatura media mensual y la cantidad anual de precipitación para su cálculo. Se calculan los valores mensuales de PET para describir el desarrollo de PET a lo largo de las temporadas de incendios, más información en (80).

Los parámetros de entrada:

Temperatura media mensual T_{mon} , suma anual de la precipitación, factor de corrección Latitud C_{lat} , y se calculan los valores mensuales de PET si la temperatura media mensual

> 0 .

$$I = \sum \left(\frac{T_{\text{mon}}}{5} \right)^{1.514},$$

$$a = (0.0675 * I^3 - 7.71 * I^2 + 1792 * I + 49239)$$

$$EP_{\text{mon}} = 1.6 * \left(10 * \frac{T_{\text{mon}}}{I} \right)^a * C_{\text{lat}}$$

Donde: EP_{mon} = evaporación potencial mensual, I = índice de calor del año,

a = un coeficiente empírico.

8. Índice de peligro desarrollado por Rodríguez y Moretti

Este índice fue desarrollado para la Región Andino Patagónica, en base al análisis de correlaciones entre las variables meteorológicas consideradas y la ocurrencia y magnitud de los incendios. Las variables de entrada: Temperatura, humedad relativa, viento y días consecutivos con o sin precipitación. Se asumió que cada una de las variables utilizadas explicaba un determinado porcentaje del peligro total de propagación del fuego.

2.2.4 Sistemas con múltiples índices

2.2.4.1 Sistema Canadiense de Evaluación de Peligro de Incendios Forestales

El sistema Canadiense de Evaluación de Peligro de Incendios Forestales (CFFDRS) proporciona un método uniforme, numérico de evaluación de peligro de incendios a lo largo de un área, que depende del clima y no tiene en cuenta diferencias en el riesgo, el combustible o la topografía. Este sistema consiste de cuatro módulos o subsistemas: Índice Meteorológico de Peligro de Incendios (FWI); Subsistema de Predicción de Comportamiento del Incendio (FBP); Subsistema de Predicción de Ocurrencia de Incendio (FOP) y Subsistema Accesorio de Humedad de Combustibles (AFM) 2.13. El Subsistema de “Accessory Fuel Moisture” (AFM), y el Sistema “Fire Occurrence Prediction System” (FOP), no se han desarrollado en su totalidad para el uso nacional, aunque varias versiones regionales de este último subsistema se encuentran entre otros en (81, 82). Los componentes del FOP en el CFFDRS están destinados a predecir el número de rayos y los incendios causados por el hombre, mientras que el objetivo es soportar aplicaciones especiales de los otros tres subsistemas del CFFDRS (83). El CFFDRS constituye uno de los pilares básicos para las otras guías y sistemas desarrollados ya sea para el personal de control de incendios u otros investigadores de incendios forestales (84).

El Índice Meteorológico de Peligro “Fire weather Index” (FWI) consiste en seis componentes. Los tres primeros son códigos de humedad del combustible que sufren cambios diarios en el contenido de humedad; los valores más altos representan un contenido de humedad menor y por lo

2. ESTADO DEL ARTE

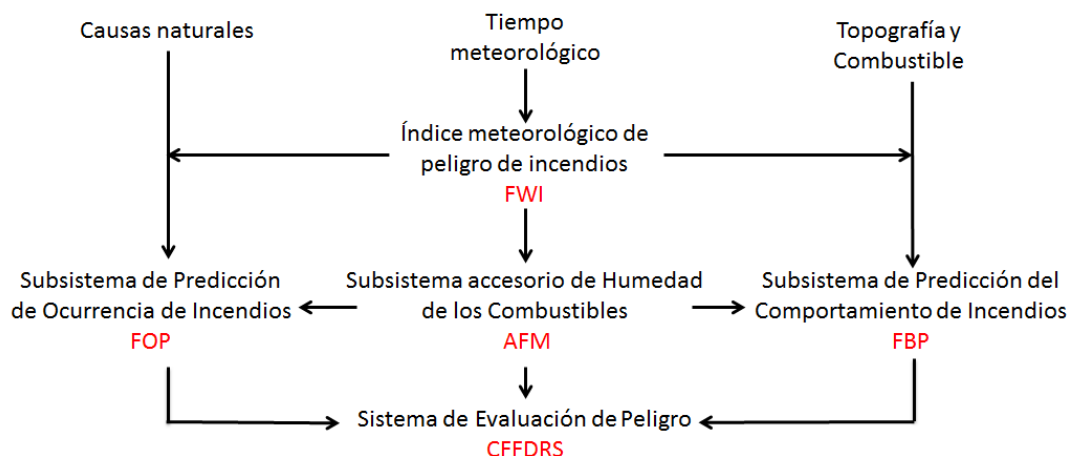


Fig. 2.13: El Sistema Canadiense de Evaluación de Peligro de Incendios Forestales

tanto mayor inflamabilidad. Los tres componentes finales son índices del comportamiento, que representan la velocidad de propagación, la cantidad de combustible disponible y la intensidad del fuego; sus valores aumentan a medida que empeoran las condiciones meteorológicas del fuego.

Los seis códigos estándar e índices del Sistema del Índice Meteorológico de Peligro son:

1. Código de humedad de los combustibles finos (FFMC) es un código numérico del contenido de humedad de la hojarasca y otros combustibles finos curados. Este código indica la facilidad de ignición e inflamabilidad de los combustibles finos muertos (85).
2. Código de humedad del humus bruto (DMC) es un índice numérico del contenido promedio de humedad de las capas orgánicas poco compactas o poco profundas. Este código indica el consumo de combustible en las capas de humus bruto.
Dicho de otra manera, este código estima la cantidad de combustible medio disponible para arder, relacionada con el contenido de humedad de la capa orgánica poco profunda del suelo (85).
3. Código de sequía (DC) es un índice numérico del contenido de humedad media en las capas orgánicas profundas y compactas. Este código indica los efectos de la sequía estacional sobre los combustibles del bosque, y la cantidad de rescoldos en las capas profundas del suelo vegetal y en los troncos grandes. Es un indicador de la importancia que puede tener la combustión sin llama en dicha capa y en la madera de grandes dimensiones (85).

2.2 Comportamiento de Incendios Forestales

4. Índice de propagación inicial (ISI) es un código numérico que indica el ritmo pronosticable de propagación del incendio. Este combina los efectos del viento y del código de humedad de los combustibles finos sobre la velocidad de propagación pero excluye la influencia de las cantidades variables de combustible (85).
5. Índice de acumulación (BUI) es un indicador que realiza una clasificación numérica de la cantidad total del combustible disponible para la combustión, combina los códigos de humedad del suelo vegetal y de sequía. (También conocido como el código ajustado de humedad del suelo vegetal o ADMC entre 1969 y 1975). En resumen, representa el grado de disponibilidad del combustible para arder, combina los dos índices DMC y DC (85).
6. El Índice Meteorológico de Peligro de Incendios (FWI) se considera como un indicador de la intensidad de Byram (1959) , que se obtiene combinando el ISI y el BUI, y se emplea como índice general de peligro meteorológico de incendios en Canadá. Representa la intensidad de propagación del frente de llamas según la cantidad de energía desprendida por la unidad lineal del frente del incendio (85).

En la figura 2.14, se puede ver la relación de los seis componentes entre sí.

2.2.4.2 Sistema de evaluación de peligro de incendios de los Estados Unidos (NFDRS)

En 1914 los administradores en California determinaron que “la ocurrencia de incendios y el comportamiento del fuego estaban relacionados con la precipitación y la humedad relativa”. Harry Gisborne desarrolló una medida de riesgo de incendios en la década de 1930. En los siguientes 25 años se llevaron a cabo numerosos esfuerzos para desarrollar los sistemas de clasificación de riesgo de incendios. En 1958 existían al menos nueve sistemas diferentes utilizados por el Servicio Forestal, además de numerosos otros sistemas utilizados por las organizaciones de protección estatales y privadas.

En 1958, se tomó la decisión de desarrollar un solo sistema de evaluación de peligro para ser utilizado en todo el país. En 1968, el Servicio Forestal comenzó a trabajar en el desarrollo de un sistema de evaluación de peligro que se construye basándose en principios de la ciencia y la ingeniería en lugar de observaciones locales. La primera versión del NFDRS fue lanzada en 1972. Fue un sistema de funcionamiento manual compuesto por varias tablas de búsqueda y nomogramas. Esta primera versión incluía una disposición para la evaluación y actualización del sistema durante los primeros cinco años después de su publicación.

2. ESTADO DEL ARTE

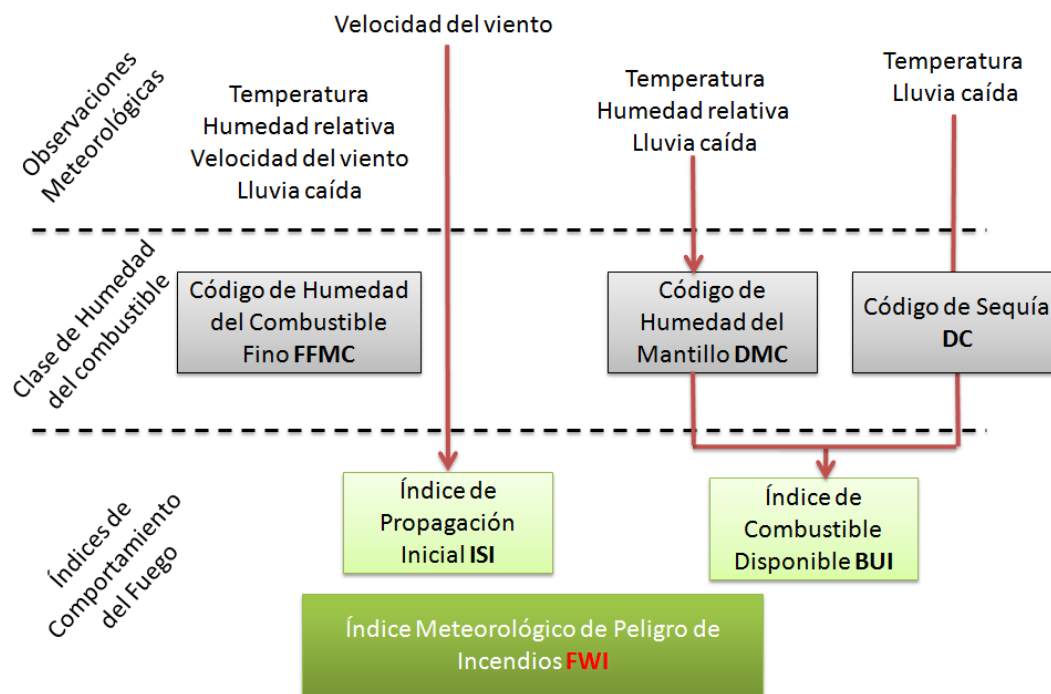


Fig. 2.14: El Índice Meteorológico de Peligro de Incendios (FWI)

En 1975, se instaló en un sistema informático accesible a nivel nacional llamado “Administrative and Forest Fire Information Retrieval and Management System” (AFIRMAS) una versión automatizada del NFDRS. En 1993, el AFIRMAS fue sustituido por el Sistema de Gestión de Información Meteorológica (“Weather Information Management System” WIMS) como procesador de información de peligro de incendios y en la actualidad está siendo utilizado por la mayoría de las agencias estatales y federales (86). En la figura 2.15, se pueden apreciar las entradas y las salidas del sistema de evaluación de peligro de incendios de los Estados Unidos:

2.2.4.3 Conceptos Importantes

1. ¿Cuál es el índice de riesgo del incendio?: el FWI fue desarrollado por el Servicio Forestal Canadiense y es utilizado por varios países en el mundo, especialmente en Europa. Mediante el uso de este índice, es posible estimar el riesgo de incendio a partir del estado de los diversos combustibles presentes, que se determina indirectamente a través de la observación de los elementos meteorológicos.

Para estimar el riesgo de incendio del sistema canadiense *FWI* se tienen en cuenta las

2.2 Comportamiento de Incendios Forestales

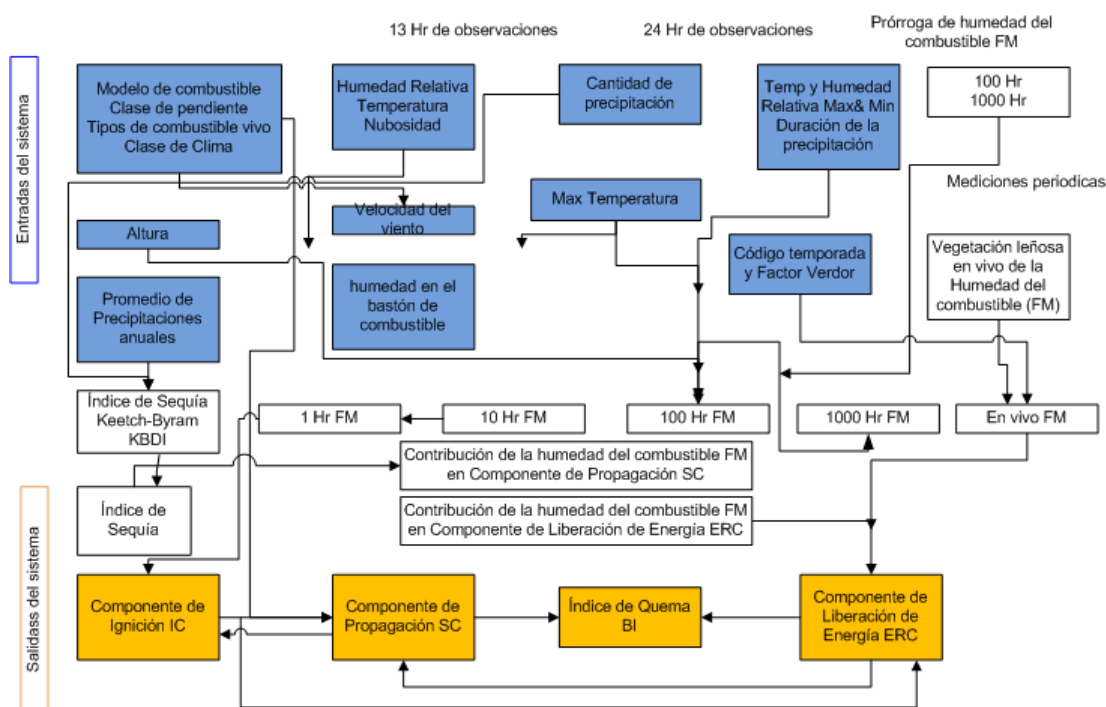


Fig. 2.15: Estructura del sistema de evaluación de peligro de incendios de los Estados Unidos

siguientes variables: Temperatura del aire, Humedad relativa, Velocidad del viento y Cantidad de la precipitación de las últimas 24h. El FWI es un índice acumulativo, lo que significa que el valor de índice para un día refleja tanto las condiciones observadas en ese día y su evolución a lo largo del tiempo transcurrido desde el comienzo de la estimación del índice.

2. La tasa de propagación (ROS): es la velocidad de avance lineal del frente de un incendio en la dirección perpendicular a la parte delantera del incendio.

La definición anterior permite el cálculo de la velocidad de propagación o intensidad de la línea del incendio para cualquier porción del perímetro del incendio. Cuando se estima la velocidad de propagación mediante la observación del intervalo de tiempo entre llamas y el paso frontal en dos puntos, es importante que los dos puntos estén orientados perpendicularmente a la parte delantera del fuego.

3. Intensidad del incendio "IF": probablemente la medida más importante del comportamiento del incendio es su intensidad. La "IF" representa el calor liberado por el frente de incendio en cada metro (kW / m). Se trata de una función de producción de calor (1)

2. ESTADO DEL ARTE

de combustible (kilojulios por kilogramo), (2) la cantidad de combustible por unidad de superficie Kg/m^2 y (3) la velocidad de propagación del frente de incendio (km/h).

Esta relación es descrita por la ecuación de Byram:

$$I = HWR$$

Donde:

I = intensidad (kW / m)

H = calor liberado de combustible (J / g)

w = combustible consumido (Kg/m^2)

r = propagación (m/sec)

4. Comportamiento extremo de un incendio: el comportamiento extremo del incendio se da cuando se alcanza tal nivel de actividad en el mismo que obstaculiza gravemente o impide su extinción. Por lo general implica uno o más de los siguientes factores:

- Alta velocidad de propagación y alta intensidad de fuego.
- Altura alta de coronación.
- Remolinos grandes de fuego.
- La columna de convección alta.
- etc.

2.2.5 El uso de la Inteligencia Artificial en los Incendios Forestales

La aplicación de la inteligencia artificial para resolver problemas complejos relativos a los incendios forestales se ha desarrollado en los últimos años. Los primeros sistemas inteligentes con aplicaciones en incendios forestales aparecieron en la literatura a finales de 1980. Según Kourtz (87) las técnicas de inteligencia artificial han sido reconocidas desde hace tiempo como herramientas adecuadas para el manejo forestal.

A continuación se citan algunos trabajos de la aplicación de la inteligencia artificial en el campo de incendios forestales: Alonso- Betanzos et al. (88) calcularon la tasa del riesgo diaria en los bosques y la clasificaron en cuatro categorías simbólicas de riesgo utilizando una red neural. En (89) Sakr et al. Presentaron un trabajo basado en dos métodos de inteligencia artificial (RNAs y MSV) en el cual trataron de predecir la ocurrencia de incendios forestales basándose en la reducción del número de entidades de supervisión y la eliminación de los mecanismos

2.2 Comportamiento de Incendios Forestales

meteorológicos de predicción. Çetin y Yusuf (90) mostraron un sistema de soporte de decisión con multi- agente para los incendios forestales, que se utiliza en el sistema de apoyo a la toma de decisiones relativas a incendios forestales (FOFDESS), utilizaron las RNAs, un clasificador bayesiano, la lógica difusa y el procesamiento de imágenes en su trabajo. En (91) Li Z. et al. desarrollaron técnicas de teledetección por satélite para identificar el humo de los incendios forestales mediante redes neuronales (que fueron exploradas para su aplicación con imágenes, con muy alta resolución, del radiómetro (AVHRR) a bordo de satélites NOAA). En (92) Armando et al. propusieron un nuevo método de clasificación en la detección de humos compuestos mediante redes neuronales. En (93) Wu J. et al. propusieron un nuevo algoritmo para la extracción de características del humo de los incendios forestales basado en imágenes, usando en una red neuronal de impulsos acoplada (PCNN). En (94) Cheng y Wang utilizan técnicas de minería de datos en espacio-temporales, su modelo se centra en la predicción de una zona quemada y utiliza una red neuronal recurrente para combinar los datos históricos de incendios y los datos meteorológicos. Un enfoque similar fue presentado por Cortez y Morais (95), donde exploraron la minería de datos (DM), los datos meteorológicos y usan una máquina de soporte vectorial (MSV), junto con los modelos de FBP y RNA para predecir la ocurrencia de incendios forestales y estimar la superficie quemada. En (89) Sakr et al. presentaron un algoritmo de predicción del riesgo de los incendios forestales, que se basa en MSV para predecir el nivel del riesgo. En (96, 97) se propuso un algoritmo de detección en tiempo real de incendios forestales mediante redes neuronales. En (98), Angayarkkani y Radhakrishnan presentaron un sistema inteligente para detectar la presencia de incendios forestales con los datos espaciales forestales utilizando RNAs. En (99) Deng et al. utilizaron un método basado en redes neuronales para el cálculo y la predicción de la tasa de liberación del calor en diferentes materiales para la seguridad durante los incendios. Otro enfoque para el modelado de la propagación del incendio usando un modelo difuso y neuronal es presentado por Vakalis et al. en (100, 101). La reducción de falsas alarmas también se presentó por Okayama (102).

Los trabajos citados anteriormente, entre otros, han demostrado la eficacia del uso de la inteligencia artificial para resolver diversos problemas en lo que respecta los incendios forestales.

2.3 Diagnóstico / Pronóstico en Equipos Industriales

Durante los últimos años ha crecido el interés por investigar y desarrollar algoritmos y herramientas para el diagnóstico y pronóstico de fallos par una gran variedad de sistemas de ingeniería (103). Determinadas tecnologías como el Mantenimiento Basado en la Condicion “Condition Based Maintenance, CBM” y la Gestión de Diagnóstico y Salud “Prognostics and Health Management, PHM”, siguen madurando y se desarrollan nuevos algoritmos para cumplir con los nuevos retos de la industria.

Algunos sectores industriales, debido principalmente a las exigencias de competitividad de los mercados globales, están inmersos en una etapa de transición que conducirá a la incorporación de las nuevas tecnologías CBM y PHM en sus activos críticos. Para llevar a cabo esta transición crucial son necesarias tecnologías adicionales de Validación y Verificación que permitan certificar el comportamiento de los módulos de Diagnóstico y Pronóstico, como paso indispensable para su incorporación a los procesos industriales.

El diagnóstico de fallos trata de identificar, localizar, determinar y clasificar la gravedad del fallo en un equipo/componente, mientras que el pronóstico es esencialmente el proceso de predecir la vida útil remanente “VR” (104). El diagnóstico y el pronóstico de fallos se han convertido en puntos clave en un mundo en el que el impacto económico de las cuestiones relacionadas con la fiabilidad y el funcionamiento rentable de los activos críticos es cada vez mayor. Diagnóstico de fallos implica la detección de condiciones anormales en el sistema, la identificación de sus localizaciones y sus causas, una tarea fundamental en cualquier sistema para la gestión de la salud (health management system HMS) de los activos físicos. El pronóstico (como una extensión natural del problema de detección e identificación de fallos) tiene como objetivo caracterizar la evolución en el tiempo del fallo detectado, lo que permite la estimación de la vida útil remanente de los subsistemas o componentes afectados.

La tecnología para el diagnóstico y pronóstico (DyP) ha sido desarrollada durante muchas décadas. Sin embargo, con el fin de comprender la importancia de la tecnología desarrollada, es esencial establecer las definiciones de estos términos. Estas definiciones varían dependiendo de las aplicaciones, por ejemplo, los términos de DyP difieren ligeramente según se aplicación a la tecnología electrónica, las redes de computación, las plantas de energía nuclear o los sistemas mecánicos (motores, máquinas rotativas, tales como helicópteros, máquinas de fabricación, maquinaria pesada, etc.). En el contexto de esta tesis, se usaran las definiciones para el “diagnóstico” y el “pronóstico” como:

- (a) Los diagnósticos son los procesos de detección, identificación y aislamiento de fallos o averías.
- (b) Los pronósticos son los procesos de predicción de un estado futuro en base a las condiciones actuales e históricos. La estimación de la vida útil remanente es un tipo de pronóstico.

Es creciente el interés por la investigación en el pronóstico debido al impacto positivo en el negocio de la predicción con suficiente anticipación de posibles fallos en los equipos. Sin embargo, muchos de los modelos de pronóstico desarrollados no se han implementado a nivel industrial debido a que solo funcionan satisfactoriamente en un entorno experimental (105). Una de las principales razones de esta deficiencia es que el modelo de pronóstico de fallo se basa en un modelado lineal (106, 107), mientras que los fallos de los equipos implican procesos no lineales. En ese caso, disponer de un enfoque no lineal es mucho más apropiado.

Progresivamente la industria va incorporando equipos modernos que disponen de múltiples sensores que monitorean continuamente las condiciones de funcionamiento. La información recabada por los sensores suele almacenarse en bases de datos (108). Sin embargo, debido a múltiples causas es frecuente que los datos no sean utilizados plenamente. A medida que se aprende a filtrar y correlacionar los datos con los procesos de degradación se va incrementando el rendimiento de los modelos de DyP (109). Ese aprendizaje en base a datos y correlaciones parece muy indicado para el uso de la Redes Neuronales y por lo tanto, existe un creciente interés en la exploración de la utilidad de las redes neuronales para predecir fallos y tiempos de vida útil remanente, especialmente para ayudar en la toma de decisiones asociadas al mantenimiento.

2.3.1 Gestión del ciclo de vida (GCV)

La gestión del ciclo de vida “GCV” (Life Cycle Management, LCM) para equipos y sistemas y críticos es una materia de importancia creciente a todos los niveles de la industria. Los principales beneficios de la GCV incluyen:

- Maximizar la disponibilidad
- Asegurar la disponibilidad, la eficiencia y el rendimiento esperado
- Optimizar el mantenimiento y reducir el coste del ciclo de vida
- Garantizar la seguridad y la fiabilidad de funcionamiento.

2. ESTADO DEL ARTE

Una visión general de la GCV que muestra su relación con el diagnóstico y el pronóstico se muestra en la figura 2.16

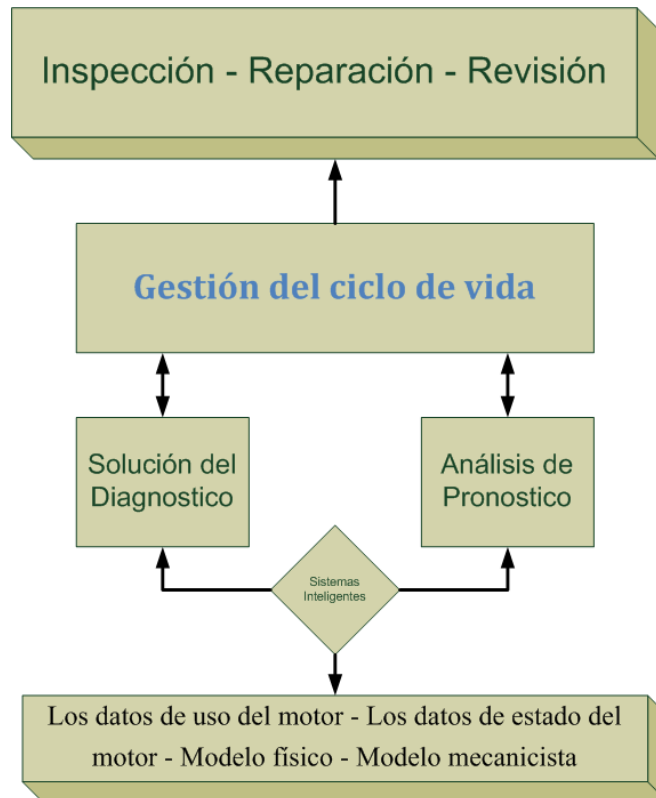


Fig. 2.16: Visión general de la gestión del ciclo de vida para sistemas críticos

2.3.2 El Problema del Diagnóstico

Un objetivo importante en la industria es conseguir que cada uno de los equipos /sistemas funcionen en un estado normal durante el mayor tiempo posible. Para mantener una condición normal de funcionamiento es muy importante disponer de sistemas eficientes de diagnóstico, especialmente en los equipos que funcionan de forma continua durante un largo tiempo. Cuando se producen anomalías estas deben ser detectadas lo antes posible, ya que si estos equipos funcionan continuamente en condiciones anormales, se pueden reproducir graves daños e incluso la pérdida de vidas humanas.

Hasta la fecha, a pesar de que se han desarrollado muchos tipos de métodos de diagnóstico, la mayoría de ellos se han basado en el método tradicional de establecer un modelo matemático (110) para el análisis de un conjunto de parámetros y, a continuación, concluir (diagnosticar)

2.3 Diagnóstico / Pronóstico en Equipos Industriales

sobre el estado de funcionamiento del equipo. Sin embargo, debido a la complejidad de los equipos, la incertidumbre en las condiciones de funcionamiento y muchos factores de no linealidad, en la mayoría de los casos, es muy difícil establecer un modelo matemático del equipo y caracterizar adecuadamente las condiciones de funcionamiento (normal y anormal). Además, en algunos casos, es incluso imposible detectar los fallos que se producen cuando el equipo está en funcionamiento. Así que, muchos investigadores se sienten atraídos recientemente por nuevos enfoques no tradicionales (111, 112). Originalmente, la mayor dificultad está en la comprensión clara de los procesos que crean daños, que pueden ser iniciados por procesos de degradación, contaminación o, simplemente, debido a causas externas.

En la literatura, el problema del diagnóstico de fallos en los sistemas se puede dividir según cuatro aspectos clave:

- Detección de fallos. En este caso la atención se centra en la identificación de los fallos en curso;
- Clasificación del fallo. El objetivo es definir qué tipo de fallo va a ocurrir;
- Localización del fallo. El interés es determinar el sector/componente defectuoso
- Identificar la causa de los fallos.

Los métodos de diagnóstico de fallos utilizan principalmente técnicas de clasificación y métodos de razonamiento (113). Las técnicas de clasificación incluyen principalmente métodos estadísticos, redes neuronales y la agrupación difusa. Los métodos de razonamiento incluyen generalmente la lógica de primer orden, la lógica difusa y redes bayesianas.

El diagnóstico de un mal funcionamiento de un sistema físico es la tarea de identificar aquellos componentes cuyos fallos son responsables de las discrepancias entre el comportamiento observado en el equipo/sistema y el comportamiento normal que debe seguir. La razón típica para realizar el diagnóstico es reparar el sistema mediante la sustitución de las piezas/componentes/equipos defectuosos por otros que no defectuosos para que el equipo/sistema pueda volver a su funcionamiento normal (114).

El sistema de diagnóstico tiene que cumplir con ciertos requisitos, de los cuales los más importantes según (115) son:

- Precisión - un diagnóstico preciso si el conjunto de los presuntas anomalías contiene la verdadera causa

2. ESTADO DEL ARTE

- Resolución - la capacidad del sistema de diagnóstico para proporcionar el número mínimo de fallos candidatos en base de la información disponible
- Sensibilidad - la capacidad de detectar pequeños fallos que se manifiestan sólo mediante leves cambios de señal
- Estabilidad de diagnóstico - el diagnóstico debe dar como resultado un conjunto estable de fallos a pesar de la existencia de condiciones cambiantes
- Fiabilidad - el sistema de diagnóstico debe proporcionar un diagnóstico exacto para todos los fallos, incluyendo fallos imprevistos para los cuales no hay experiencia previa.

En la figura 2.17, se muestra una estructura general para usar redes neuronales para tareas de diagnóstico

2.3.3 El Problema del Pronóstico

Con los recientes avances en la tecnología, las industrias y los investigadores evolucionan en el desarrollo de metodologías avanzadas de mantenimiento que tienen como objetivo la mejora de la fiabilidad y disponibilidad de los activos críticos así como la reducción de los costes globales. Más precisamente, se supone que hoy en día las acciones adecuadas (ya sean tareas de mantenimiento o cambios de perfil de carga,...) deben realizarse en el momento oportuno para que los fallos críticos, que conducen a grandes averías, pueden ser evitados. Dentro de este marco el Mantenimiento Basado en la Condición (CBM) parece ser un enfoque prometedor que utiliza los datos de monitorización para facilitar indicadores y mostrar tendencias que muestran la salud del sistema. El uso de la tecnología CBM permite llevar a cabo las acciones de mantenimiento en el momento adecuado (116). Sin embargo, surge la pregunta: “¿de qué forma se han de utilizar los datos de monitorización de la salud de un activo para determinar o predecir su tiempo de vida?” y esto se puede lograr a través del Pronóstico. De acuerdo con ISO 13381 – 1, el pronóstico se define como “la estimación de tiempo hasta el fallo y el riesgo de uno o más modos de fallo existentes y futuros” (117). También se conoce como la “predicción de la vida útil de un sistema”, ya que es un proceso cuyo objetivo es predecir la vida útil remanente ante un caso de fallo, dado el estado actual del equipo y su antiguo estado de funcionamiento. El pronóstico debe tanto permitir evitar los riesgos de fallo y el desgaste inoportuno (118). Existen diversos enfoques para el pronóstico (105, 119, 120). Sin embargo, los sistemas de pronóstico que se usan para enfrentar los desafíos industriales son todavía escasos. Esto puede

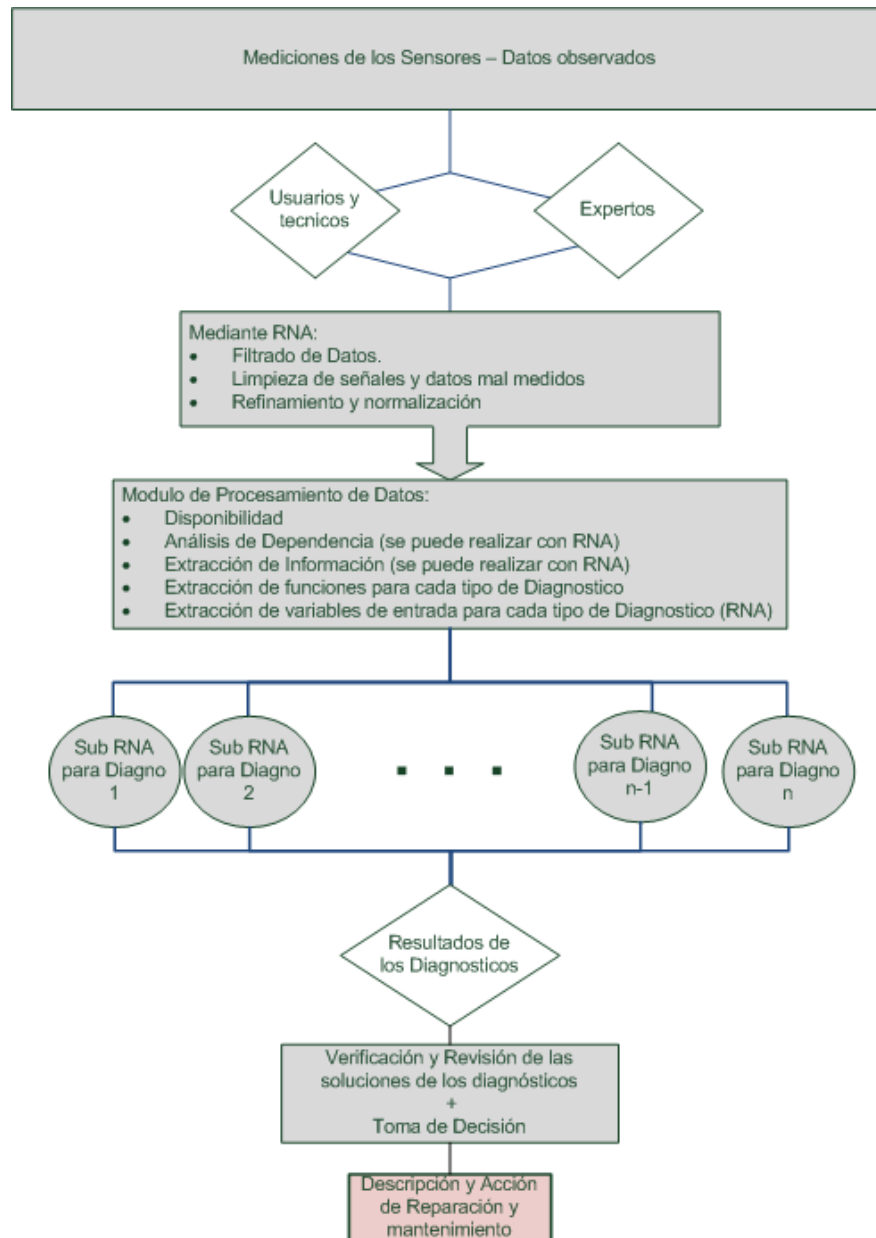


Fig. 2.17: Estructura general del diagnóstico mediante RNA

2. ESTADO DEL ARTE

ser debido al entorno operativo que puede ser muy complejo y la no linealidad de los sistemas industriales reales, lo que hace difícil establecer enfoques de pronóstico efectivos, lo suficientemente robustos para tolerar la incertidumbre, y lo suficientemente fiables para mostrar un rendimiento aceptable en diversas condiciones. Últimamente se presta una gran atención a enfoques de pronóstico basado en datos (principalmente las técnicas de inteligencia artificial). Tales métodos tienen como objetivo transformar los datos en información apropiada mediante la realización de una modelización no lineal de los sistemas reales. Sin embargo, la aplicabilidad de los enfoques basados esta limitada por los requisitos industriales, tales como la complejidad de la aplicación práctica, las decisiones humanas y la parametrización inicial (121, 122). Los problemas de robustez, fiabilidad y aplicabilidad de los pronósticos, siguen siendo áreas abiertas que deben ser abordados mediante tareas de investigación y desarrollo.

- Robustez; puede ser definida como la “capacidad de un enfoque de pronóstico de ser insensible a las variaciones de los datos de entrada”. En la práctica, un enfoque sólido debe ser lo suficientemente preciso para capturar el comportamiento dinámico de un sistema que se degrada.
- Fiabilidad; puede ser definida como la “capacidad de que un método de pronóstico sea coherente en situaciones en las que se presentan datos nuevos o desconocidos”.
- Aplicabilidad; puede ser definida como la “capacidad de que un método de pronóstico pueda ser aplicado prácticamente bajo las limitaciones industriales”. La aplicabilidad puede considerarse frente a diferentes aspectos: requisitos de aplicación, tiempo de cálculo, límites teóricos del enfoque o la verosimilitud de hipótesis.

Otras definiciones acerca del pronóstico industrial existen en (116, 123, 124), y los principales aspectos que se destacan son: el estado del sistema actual, la proyección (o extrapolación) del estado actual y la estimación del tiempo restante antes del fallo. Estas definiciones se han normalizado por la norma ISO 13381 – 1 (125) en el que el pronóstico se define como la estimación del tiempo de operación antes del fallo y el riesgo futuro de la existencia o aparición de uno o varios modos de fallo. Esta norma define los entornos del pronóstico, identifica los datos necesarios para realizarlo y establece los umbrales de alarma y los límites del sistema. Los pasos principales en la realización del pronóstico, tal como se define en la norma citada, se resumen en la figura 2.18. A pesar de que el diagnóstico y el pronóstico se refieren a la evaluación de la salud, los resultados del diagnóstico conducen a decisiones reactivas sobre acciones

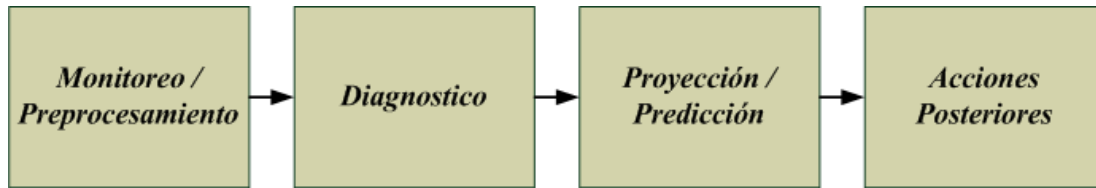


Fig. 2.18: Etapas del Pronóstico acorde con ISO 13381-1 (126)

correctivas, mientras que los resultados del pronóstico conducen a decisiones proactivas acerca de las acciones preventivas y/o evasivas, con el objetivo de aumentar la vida útil de los equipos. Los diferentes enfoques de pronóstico que se han aplicado con éxito para diferentes tipos de problemas se han basado en la experiencia, análisis de tendencias, uso de Inteligencia Artificial la AI y estimadores de estado (127).

En general, los pronósticos pueden ser clasificados según tres enfoques principales: modelos basados en la física, basados en la experiencia y basados en los datos (128). Los modelos basados en la física requieren modelos matemáticos exactos que se construyen a partir del principio físicos de los modos de fallo del sistema (129). Este enfoque se centra en la evaluación de residuos entre la medición detectada en el equipo y la salida de los modelos matemáticos. Este enfoque es el método más preferible cuando hay calidad en los datos medidos durante el funcionamiento y los modelos físicos de los fallos son buenos. Sin embargo, para desarrollar un modelo matemático preciso son necesarios conocimiento global y conocimiento sobre la mecánica teórica de los equipos monitoreados. El pronóstico basado en la experiencia depende únicamente de las opiniones de expertos y es de menor complejidad que el modelado basado en la física (129).

El enfoque del pronóstico basado en datos utiliza datos históricos para aprender automáticamente el comportamiento del sistema (130). El modelo se utiliza para predecir la VR 2.19. El reto principal con este enfoque es obtener suficientes datos normales históricos para el desarrollo de un modelo de pronóstico.

2.3.4 Vida Útil Remanente “VR”

La vida útil remanente “VR” (Remaining Useful Life “RUL”) llamada también la vida residual, se refiere al tiempo que queda antes de se produzca el fallo o una avería debido a la edad y el estado actual de la máquina. En algunas situaciones, especialmente cuando hay un fallo, sería deseable predecir la probabilidad de que una máquina funciona sin un fallo o un fracaso hasta algún tiempo futuro (por ejemplo, intervalo de inspección siguiente) dada la condición

2. ESTADO DEL ARTE

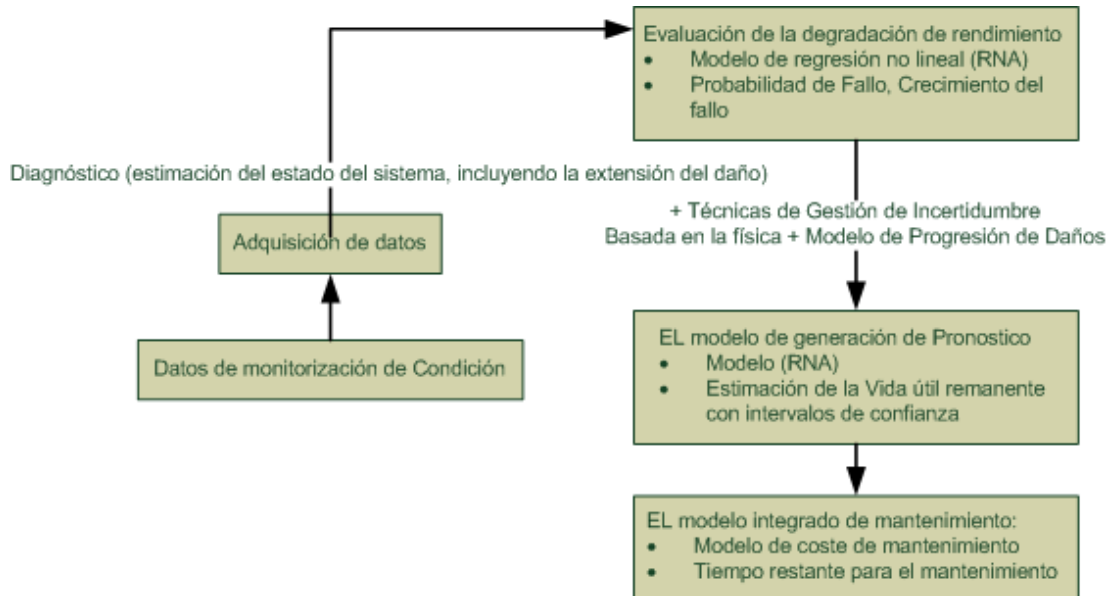


Fig. 2.19: Estructura general para el pronóstico con RNA

actual de la máquina y el perfil de funcionamiento anterior. La estimación de la VR es una tarea importante como parte de las labores de Pronóstico que se incluyen en el proceso de determinación del estado de salud de equipos industriales.

En realidad, en cualquier caso, la probabilidad de que un equipo funcione sin fallos hasta un intervalo siguiente de inspección podría ser una buena referencia para el personal de mantenimiento, para determinar si el intervalo de inspección es apropiado o no. La mayoría de los artículos en la literatura de pronóstico versa sobre la estimación de la VR (131, 132).

La estimación de la Vida útil remanente es la tarea más común en el campo de investigación sobre pronósticos y la gestión de la salud. El enfoque basado en datos para la estimación de la VR normalmente se basa en la disponibilidad de datos de funcionamiento hasta el fallo, base sobre la cual se puede estimar la VR, ya sea directamente a través de un modelo multi-variable, o indirectamente mediante la estimación de daños por la extrapolación de la progresión de fallos (133). Existen diferentes estrategias para la estimación de la vida útil remanente. En los métodos dirigidos por datos “data-driven”, la estrategia estima directamente la VR mediante la aplicación de un modelo multi-variable definido sobre los datos. Otra estrategia es estimar la VR indirectamente de los fallos, luego realizar una extrapolación adecuada para la progresión del fallo y calcular la VR mediante la intersección de los fallos extrapolados y el criterio de fallo. Este último enfoque requiere la definición de los fallos y del criterio de error, que muchas

veces es muy difícil de establecer (133).

L.C. Tang en (134) considera la vida residual como una variable aleatoria, y presenta su comportamiento asintótico cuando la función de fiabilidad está representada por distintas funciones de distribución. En (135), Lim J. y Park D. H. estudian el comportamiento ascendente de la vida residual, y ponen a prueba la hipótesis nula de que la vida residual no es monótona. Siddiqui y Caglar (136) tratan la vida residual como una variable aleatoria, y dan una representación de su función de distribución. Cuando la distribución es Gamma, o Weibull, los autores calculan la media y la varianza de la variable. Bradley y Gupta también estudian el comportamiento asintótico de la vida residual en (137).

Los pronósticos más avanzados se centran en el seguimiento de los resultados y la evaluación de la degradación, por lo que los fallos se pueden predecir y prevenir. Para cumplir con el objetivo del pronóstico, tres pasos fundamentales son necesarios:

- En primer lugar, tener la capacidad de detectar la anomalía en una etapa temprana.
- En segundo lugar, el funcionamiento del equipo/sistema debe evaluarse robustamente mediante un seguimiento continuo.
- Por último, la vida útil remanente y el modo de fallo posible deben ser predichos con un grado de fiabilidad alto.

2.3.5 Clasificación de los modelos para la estimación de la VR

Inicialmente, los modelos de estimación de la vida remanente (EVR) se clasificaron en la Figura 2.20. A continuación se presentan algunas conclusiones importantes como fruto de una revisión bibliográfica.

Diversas metodologías para la estimación de la VR y sus aplicaciones prácticas se han propuesto en la literatura. En la figura 2.20 se presenta una clasificación detallada de todas estas metodologías. Las redes neuronales representan una buena estrategia para resolver problemas complejos sin ningún conocimiento o suposiciones acerca de la composición interna de la estructura o algunas partes de un sistema técnico. La metodología de las RNAs se muestra mediante una descripción formal de las reglas de transformación que vinculan las entradas y las salidas o mediante la construcción de un prototipo del comportamiento que aproxima el funcionamiento del sistema (138).

En (139) y (140) las redes neuronales han sido utilizadas para desarrollar las metodologías de la VR, Shao y Nezu (141) propuso un nuevo concepto llamado progresión basada en la predicción

2. ESTADO DEL ARTE



Fig. 2.20: clasificación de metodologías para la Estimación de la Vida Remanente

2.3 Diagnóstico / Pronóstico en Equipos Industriales

de la vida útil remanente (PPRL) para estimar la VR de un rodamiento. Este nuevo concepto manipula las variables de entrada a partir de mediciones en línea a través de un modelo neuronal. Los pasos del concepto PPRL incluyen:

- La aplicación de diferentes normas logísticas para comprobar el estado del componente en la etapa inicial de funcionamiento;
- Dividir el trabajo de predicción en predicción de evolución general y predicción de VR, y la construcción del modelo de predicción en tiempo real utilizando el método de ventana móvil;
- La obtención de los límites de la VR.

También se puede citar los trabajos de Gebraeel et al. (139, 140) que desarrollaron una red neuronal para calcular la VR de los cojinetes de elementos rodantes mediante el control de sus vibraciones. Huang et al. (142) presentaron un ejemplo de un método para predecir la VR de un cojinete de bolas, basado en mapa auto-organizados (SOM) y de un nuevo método de propagación de redes neuronales. El procedimiento utilizado en su modelo, se compone de los siguientes pasos:

Paso 1: Dividir de los datos históricos disponibles de los fallos, que incluye los valores de edad y los valores reales de monitoreo en puntos de inspección de cada histórico del fallo.

Paso 2: Cada serie de mediciones para un historial de fallos se fija con la función generalizada Weibull-FR. Los valores de edad y los valores de medición insertados, en los puntos de control para todos los historiales de fallo, se utilizan para construir el conjunto de entrenamiento de la RNA.

Paso 3: El conjunto de validación RNA se construye utilizando los valores de edad y los valores de medición reales en los puntos de inspección de todo el histórico de fallos.

Paso 4: Capacitar el modelo de la RNA por un conjunto de entrenamiento y el conjunto de validación mediante el algoritmo de Levenberg-Marquardt (LM).

Después del paso 4, el proceso de entrenamiento de RNA se ha completado, y el modelo RNA entrenado está disponible para la predicción de la VR. Para un componente de equipo que está siendo monitoreado, los pasos siguientes sirven para la predicción de la VR para el equipo.

Paso 5: En un punto de inspección dado, se obtiene la serie de los valores de medición hasta el momento actual. Los valores de medición ajustados en los puntos de inspección actuales y anteriores, así como los valores de edad en estos puntos, se utilizan como las entradas al modelo

2. ESTADO DEL ARTE

neuronal entrenado

Paso 6: El porcentaje de la vida residual previsto en el punto actual se calcula utilizando el modelo neuronal entrenado.

Paso 7: la VR se calcula basándose en la edad actual de los componentes y el porcentaje de la VR prevista. Por ejemplo, si la edad actual es de 400 días y el porcentaje de vida útil prevista es del 80%, el tiempo de fallo previsto sería de $400/80\% = 500$ (días), y la VR sería de $500 - 400 = 100$ (días).

Paso 8: Cuando nuevos datos de inspección están disponibles, repetir el Paso 5 – 7 sobre la base de los datos disponibles, y actualizar la predicción de la VR.

Ventajas: Las redes neuronales son técnicas de modelado muy sofisticadas capaces de modelar funciones extremadamente complejas y de gran alcance para la predicción de la VR. Las otras ventajas son la facilidad de uso, la robustez y la generalización libre del modelo para la estimación de la VR (143).

Desventajas: Las redes neuronales son a veces lentas en la fase de entrenamiento cuando hay muchos datos.

2.3.6 Observaciones importantes

Durante la realización de la revisión general de las metodologías utilizadas para el cálculo de la VR, se han recopilado las siguientes observaciones:

- La estimación de la VR es muy dependiente del sistema de ingeniería / producto considerado. No todas las metodologías son aplicables a todos los sistemas / productos.
- En algunos casos, múltiples metodologías se han fusionado y como resultado la estimación de la VR suele ser más precisa que cuando se aplican una única por separado.
- Independientemente de los tipos de modelos de predicción de la VR, por lo general es necesario predecir la evolución futura del indicador de degradación para la predicción de la VR.

2.3.7 Conclusión

La precisión de la estimación de la VR depende altamente de la cantidad y la calidad de los datos. En contraste, las metodologías basadas en modelos físicos dependerán de la comprensión fundamental de la física del fallo del sistema. Los modelos físicos contienen un mapeo

2.3 Diagnóstico / Pronóstico en Equipos Industriales

funcional de los parámetros del sistema y su comportamiento, en consecuencia la precisión de las metodologías basadas en modelos físicos es sustancialmente mayor que las metodologías basadas en datos.

La combinación de los enfoques basados en características físicas y los enfoques basados en datos, proporciona una capacidad de pronóstico más completa que la que se obtiene de aplicaciones no combinadas. El resultado es una información más valiosa para la planificación de la inspección de los componentes/equipos durante períodos específicos.

2.4 Estimación de la Demanda de Energía Eléctrica

2.4.1 Introducción

Desde principios de los 70s, a raíz de la primera crisis del petróleo, la energía llamó la atención de los responsables políticos, lo que produjo un aumento y gran interés en la investigación de la demanda de energía “DE”, con el fin de superar la comprensión limitada de la misma y de dar respuesta satisfactoria a la DE incluso en circunstancias extraordinarias (144). En la actualidad, la sociedad afronta acontecimientos climáticos, sociales y económicos inestables que ocasionan alteraciones en la DE así como en las posibilidades de satisfacerla eficientemente. La DE mundial crece debido al aumento de la población mundial. Por ejemplo, el consumo de petróleo en Estados Unidos, como uno de los más grandes consumidores de energía del mundo, se ha incrementado en un 20,4% entre 1990 y 2005, su consumo de gas natural también ha aumentado un 16.32% según la administración de información de la energía “EIA” (ver <http://www.eia.gov>). Los gobiernos y los responsables políticos demandan pronósticos precisos de la DE, especialmente de las predicciones a largo plazo, para la toma de decisiones a gran escala, como la planificación de la inversión para la generación y distribución de energía. El debate continuo, entre los ingenieros y los economistas, llevó a importantes desarrollos metodológicos que enriquecieron el proceso de toma de decisiones sobre la energía en su conjunto, y se desarrollaron una amplia variedad de modelos para el análisis, previsión y predicción de la demanda de energía eléctrica “DEE”. Actualmente, nos encontramos en la era de la tecnología y la información. Las nuevas tecnologías de la información y computación ofrecen técnicas de predicción cuyo objetivo es obtener estimaciones de valores futuros de una serie temporal a partir de la información histórica contenida en la serie observada hasta el momento actual. Estas técnicas requieren la especificación de los factores que determinan el comportamiento de las variables. Tras determinar el efecto que aporta cada variable explicativa, se puede considerar que en un futuro su influencia permanecerá inalterable, por lo que si se cuenta con una predicción de la evolución futura global, podrá extrapolarse el comportamiento futuro de la variable de estudio.

2.4.2 Importancia y necesidad de una estimación eficiente de DEE

Son varios los tipos de empresas los que requieren las estimaciones a largo plazo de la DEE:

- Las empresas que de forma regulada o en monopolio estatal, como la distribución y el transporte de electricidad, buscan realizar la cobertura de la demanda de electricidad, logrando un mínimo coste con una fiabilidad y calidad razonable mediante la explotación

2.4 Estimación de la Demanda de Energía Eléctrica

centralizada del sistema. Para ello tienen que plantear nuevas inversiones que permitan absorber los incrementos futuros previstos. En el caso del transporte eléctrico lo costea el Gobierno Central, en el caso de la red de distribución el gasto depende de lo que lo ocasione, pudiendo tener que asumir la inversión la compañía de distribución o el Gobierno Autonómico o usuarios y promotores.

- Las empresas en libre competencia, como las de generación en régimen ordinario o especial, que están interesadas en conocer cuál va ser el volumen de negocio al que optan con sus inversiones actuales y futuras, realizando una valoración de las mismas. Considerar el mantenimiento operativo de instalaciones existentes, o su desmantelamiento, así como posibilidades de ampliación o realización de nuevas inversiones. Si las predicciones resultan muy bajas podrían aparecer deficiencias en el suministro de energía, con inconvenientes para distintos sectores económicos, siendo los costes muy superiores a la energía no abastecida. Si por el contrario, las estimaciones han resultado excesivas, los costes de oportunidad presentados serían elevados, presentándose una elevada inversión económica improductiva. En ambos casos, ya sea por defecto o por exceso, se encarecería el precio que paga el consumidor. Si se produjera una carencia, a fin de ajustar la curva de producción con la demanda subiría el precio de la energía consumida, como bien de este comercio. Si por el contrario se produjera un exceso de generación, los costes asociados a la misma, a pesar de no emplearse, se repercutirán al consumidor igualmente, dado que en la mayoría de los casos el cliente no cuenta con opción a prescindir de este bien, cambiándolo por otro.
- Las comercializadoras, han visto aumentar el número de clientes a los que suministrar electricidad, con el consiguiente crecimiento de la incertidumbre a la hora de gestionarles la energía. Para minimizar el riesgo del precio las comercializadoras pueden firmar contratos de compra/venta de energía a medio y largo plazo en distintos mercados. Antes de acudir a los mercados de medio y largo plazo (físicos o financieros) es necesario conocer, con la menor incertidumbre posible, el consumo de su cartera de clientes, y así poder ajustar las cantidades de energía para la que desea cubrir su riesgo de precio. En este punto es cuando surge la necesidad de disponer de un buen modelo de previsión de energía eléctrica a medio o largo plazo.
- A las administraciones públicas les corresponde garantizar el fluido eléctrico a la población, para mantener su nivel de vida, y todas las actividades económicas, sean del sector pri-

2. ESTADO DEL ARTE

mario, comercial o industrial (145). Además debe realizarse de manera compatible con el medio ambiente y se ha de realizar una ocupación del territorio compatible con otros usos. Todo ello lleva a una necesidad denominada Planificación Energética. En un emplazamiento, la inversión y las grandes infraestructuras desde que se plantea su desarrollo hasta que se encuentran operativas pasan entre 5 y 10 años. La elaboración del proyecto, la obtención de autorizaciones de la obra, autorizaciones medio ambientales, con exposición pública del estudio y respuestas a las alegaciones. Más el suministro del material necesario, y ejecución de las obras, establecen estos largos plazos de ejecución.

- Además de las empresas citadas se afectaría a otras, como las suministradoras de combustible, necesario para la generación en régimen ordinario.

Por lo que la estimación de la DE a largo plazo es realizada por varias empresas para proceder a invertir su capital, y prever su desarrollo futuro. Es fundamental, por tanto, actualizar los escenarios y las proyecciones correspondientes cada cierto tiempo, especialmente, en momentos de incertidumbre de los parámetros económicos y demográficos y de volatilidad de los precios. Las características propias de la energía eléctrica, hacen de ella un bien, o servicio reticente a su comercialización en régimen de competencia perfecta. El mercado eléctrico presenta algunas particularidades respecto al resto de suministros energéticos, al no ser almacenable de forma competitiva ni en cantidades apreciables, se debe producir en cada instante la misma cantidad que se consume más las pérdidas. Esto podría cambiar ligeramente en un futuro, pues la dificultad de integración de energías no gestionables en el sistema eléctrico, como la eólica y la solar, ha impulsado un plan en Canarias, España, y a nivel europeo de construcción de centrales hidroeléctricas de bombeo, como almacenamiento energético. La investigación en otros sistemas de almacenamiento a todos los niveles: baterías para uso doméstico, los vehículos eléctricos como almacenamiento distribuido y baterías multimegavatios, etc, podrían cambiar algo el panorama en las próximas décadas. Y por otro lado, el interés emergente de energías renovables (energía eólica, solar, hidráulica, biomasa, fuentes de energía fósil y nuclear, etc.), también catástrofes naturales, o accidentes industriales, crean confusión, indecisión y dificultad a la hora de producir o almacenar la energía. En resumen, disponer de modelos que permiten estimaciones eficientes es necesario y una necesidad urgente.

2.4.3 Revisión de los principales modelos aplicados

Entre los métodos de la estimación de la DEE se encuentran las series temporales (como los métodos de análisis de tendencias), econométricos (son métodos que intentan cuantificar la relación entre el parámetro de interés (variable de salida) y un número de factores que afectan a la variable de salida como los análisis estructurales), End-use (métodos de análisis de ingeniería), series temporales de predicción que se pueden dividirse en tres, 1) métodos de regresión lineal: adaptándose a la mejor línea recta con los datos históricos y se supone que el futuro va a seguir esa línea, como por ejemplo mediante método de regresión lineal con mínimos cuadrados. 2) métodos de regresión polinomial, usando cualquier orden polinomial; cuadrado, cubo, etc, pero para ordenes más grandes, suelen ser innecesariamente complejos, 3) métodos de regresión logarítmica, usando transformaciones logarítmica. Sin embargo, la mayoría de estos métodos se ven afectados cuando hay mucha variabilidad en los datos, o existen problemas de co-linealidad. Entre los métodos que han tenido en cuenta el problema de la variabilidad en los datos, podemos citar: modelos de ajuste funcional, técnicas de suavizado (media móvil ponderada o exponencial, filtros de Kalman, Box-Jenkins, etc.). La ventaja de estos modelos es principalmente su sencillez. Por otra parte, identificar estos modelos de aproximación y que respondan de una manera realista no suele ser una tarea fácil, debido a la no linealidad de las relaciones entre las variables que afectan a la DEE y su comportamiento.

En los últimos treinta años, se han desarrollado diferentes métodos para la predicción de la demanda de energía. El modelo autorregresivo integrado de media móvil (ARIMA) es uno de los métodos basados en series temporales más populares que se ha utilizado para la predicción del consumo de energía a largo plazo. Por ejemplo, Haris Y Liu propusieron un modelo ARIMA en (146). ARIMA y ARIMA estacional (SARIMA) fueron utilizados también por Ediger y Akar para estimar predicciones futuras del consumo de energía primaria en Turquía del periodo 2005-2020 (147). También se han propuesto modelos de regresión lineal para la predicción del consumo de energía (148).

Sin embargo, cuando aparece no linealidad en la DE, los enfoques lineales pueden no captar la dinámica no lineal del proceso. En la última década, modelos, basados en la inteligencia computacional, han estado en el centro de atención en las aplicaciones de predicción, como el consumo de energía y la predicción de la demanda. Los métodos basados en inteligencia computacional pueden capturar con eficacia y modelar el comportamiento no lineal de series temporales. La lógica difusa y redes neuronales artificiales, algoritmos genéticos, máquinas de soporte vectorial son las técnicas principales que se han usado en muchas aplicaciones para la

2. ESTADO DEL ARTE

modelización y predicción.

Una técnica híbrida del conocido sistema de inferencia difuso Takagi-Sugeno y la regresión difusa se ha propuesto para la predicción a corto plazo de las variaciones de la demanda eléctrica por Shakouri et al. (149). En su estudio, propusieron una máquina III TSK de inferencia difusa combinada con un conjunto de regresiones difusas lineales y no lineales, para modelar los efectos del cambio climático sobre la demanda de electricidad. Padmakumari et al. combinaron las RNAs y un modelado difuso para la predicción de carga de distribución a largo plazo (150). Emplearon redes de funciones de base radial (RBFN).

Un número considerable de estudios de revisión del estado del arte para la predicción de la DEE, también han sido publicados. Una encuesta exhaustiva de la literatura en la predicción de la DEE utilizando técnicas de inteligencia artificial se ha presentado en (151). Este estudio ha presentado una biografía amplia de numerosos trabajos, se concentró en las técnicas de predicción de carga basadas en los sistemas expertos, algoritmos difusos, genéticos y RNAs.

Otra revisión se ha realizado por Hahn et al. en (152). Ellos han proporcionado una visión general sobre los métodos de series temporales y de regresión clásica, inteligencia artificial y enfoques de inteligencia computacional. También varios modelos tradicionales y basados en inteligencia artificial se han presentado en (153).

2.4.4 Problemática del estudio propuesto

La predicción de la DEE es un componente esencial para la planificación energética, la formulación de estrategias y la recomendación de políticas energéticas. La tarea no es un reto sólo en los países en desarrollo donde los datos son necesarios, sino también en regiones aisladas. Los sistemas eléctricos se hallan fuertemente inter-conectados en muchos países. España por ejemplo, se halla interconectada con 4 países: Marruecos, Portugal, Francia y Andorra. Pudiendo intercambiar fluido eléctrico con cualquiera de ellos. No obstante en las islas canarias y dada la gran profundidad que las separa, en la actualidad está formada por 6 sistemas eléctricos independientes (cada uno se autoabastece), por ser imposible que un sistema apoye a otro. Esto hace que la predicción de energía una situación mucho más crítica que en el resto del Territorio español, a día de hoy las Islas Baleares se hallan eléctricamente unidas a la península. Las proyecciones de las demandas energéticas tienden a menudo a desviarse de las demandas reales debido a las limitaciones en la estructura del modelo o supuestos inapropiados. En Estados Unidos, evaluando la predicción de la demanda de energía, Craig et al. (154) mostraron que la mayoría de las predicciones sobre-estimaron la demanda futura. Los modelos empleados sufren

2.4 Estimación de la Demanda de Energía Eléctrica

de muchas limitaciones, a menudo son llamados cajas negras, difíciles de evaluar y con pobre posibilidades de producir buenos resultados. Muchos investigadores tenían la percepción de que un modelo complejo con gran cantidad de datos de entrada, era capaz siempre de producir resultados más precisos. Sin embargo a veces, simples modelos tienden a dar resultados más precisos que modelos complejos (155). Por otra parte Koomey (156) señaló que los modeladores de la demanda de energía, primero deben preguntarse si la herramienta de modelado conduce o apoya el proceso de desarrollo de un escenario coherente y ofrece credibilidad para hacer frente a las incertidumbres.

Los modelos de energía antiguos seguían ciertos escenarios mundiales que han ido desapareciendo, por ejemplo los modelos de energía en los países en desarrollo han seguido las mismas características de los países industrializados, pero con un retraso en el tiempo diferente (157). Sin embargo, esto no ha resultado cierto o mejor dicho correcto en todos los escenarios, ya que China por ejemplo, ha mantenido un nivel de crecimiento económico durante décadas, y se ha convertido en un factor global importante. Otro ejemplo, la incertidumbre económica tras la crisis del 2008, donde las predicciones económicas son más inciertas que nunca, y más vinculadas a situaciones internacionales.

Lo cierto es que en el pasado la demanda eléctrica ha sido creciente a lo largo de los años, debido principalmente a la proliferación de dispositivos eléctricos que han facilitado la vida, sin importar el consumo. Últimamente han aparecido una serie de nuevos factores que habría que considerar, pues podrían irrumpir con fuerza introduciendo grandes cambios en la evolución de la demanda futura:

- La incertidumbre económica tras la crisis del 2008, ha ocasionado que las predicciones económicas sean más inciertas que nunca, y más vinculadas a situaciones internacionales.
- La pobreza energética de la población, agravada por la subida brusca del precio de la electricidad, unido a la alta tasa de desempleo y precariedad laboral en nuestro país (158).
- El incremento en la eficiencia de los equipos eléctricos, surgiendo la clasificación energética de todos los electrodomésticos, de los edificios, del alumbrado, etc...
- La posible incorporación del vehículo eléctrico en los mercados podría producir un incremento en el consumo de esta energía. En caso de comercializarse, los vehículos a la red “Vehicle-to-grid (V2G)” podrían aportar un almacenamiento energético distribuido. Además influirá la gestión que se imponga en el sistema eléctrico.

2. ESTADO DEL ARTE

- La implantación de sistemas de gestión de la demanda, que se materializará con la incorporación en toda Europa de los contadores inteligentes en 2019, desplazará temporalmente el consumo eléctrico, pero podrá afectar también a la cuantía del mismo, al producir una toma de más conciencia por parte de los ciudadanos (159).
- La evolución de la aerotermia (Energía renovable basada en intercambio de energía térmica con el aire), podría provocar la incorporación de la electricidad como energía en procesos térmicos de baja temperatura, desplazando a algunos combustibles fósiles (160).

Por lo que a día de hoy resulta más difícil que nunca el predecir una evolución futura de la demanda eléctrica.

CAPÍTULO 3

CONTRIBUCIONES

3. CONTRIBUCIONES

3.1 Introducción

En el capítulo 2, se han introducido los conceptos y definiciones y deficiencias de las RNAs, un repaso bibliográfico sobre el comportamiento de los incendios forestales, la vida remanente y la demanda de energía eléctrica en los que se centrarán las aplicaciones prácticas de esta memoria. Este capítulo presenta las contribuciones originales que se han podido llevar a cabo durante este periodo de investigación. Para dar mayor claridad a la exposición se ha decidido describir cada contribución en un apartado diferenciado, realizando una exposición detallada de las mismas, atendiendo a los aspectos teóricos y metodológicos que se completarán con los resultados del diseño experimental que se presentará en el capítulo siguiente.

3.2 Contribuciones en las Redes Neuronales Artificiales

Las aportaciones que a continuación se citan han surgido de las necesidades específicas y complejas (retos *I – RAMS*) descritas en el capítulo de introducción. Para responder a los nuevos retos es necesario mejorar deficiencias y limitaciones observables en las aplicaciones de las actuales RNA, especialmente en lo que respecta a su arquitectura, estabilidad y a la fiabilidad de sus respuestas.

Como primera etapa se ha realizado un análisis exhaustivo de los conceptos, definiciones y teorías, así como de las consideraciones teórico-prácticas para la aplicación de las redes neuronales artificiales (RNA), ya expuestas en 2.1.

Como segunda etapa se han comparado los resultados de ese análisis con las necesidades de los casos de aplicación afrontados en la línea de investigación *I – RAMS* en la que se enmarca este trabajo.

En la etapa final se han definido las necesidades y posibilidades abordables cuya investigación ha permitido establecer las contribuciones que aquí se describen. Como referencia para la descripción de las contribuciones en el uso de las RNA, se seguirá el esquema mostrado en la figura 3.1 que muestra las etapas estándar para crear una RNA cuyo diseño permita la resolución de problemas complejos.

3.2.1 Aportación para la definición de la estructura

Se ha mostrado en el análisis del estado del arte que no existe un criterio único para definir el número de neuronas ocultas necesarias y suficientes para garantizar mejor convergencia. Existen varias propuestas para hacer frente a este problema, entre ellas se pueden mencionar:

3.2 Contribuciones en las Redes Neuronales Artificiales

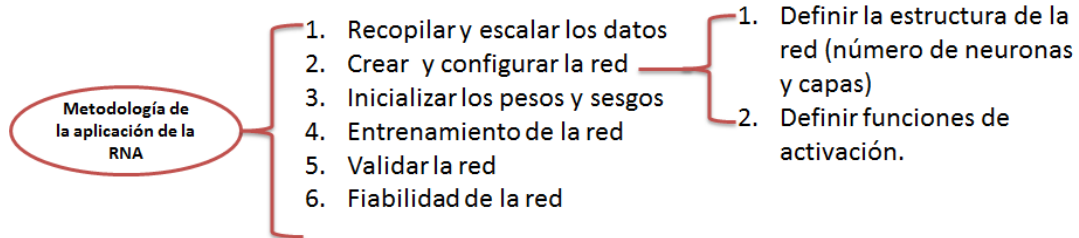


Fig. 3.1: Etapas para aplicar las RNA

Comenzar con una red de gran tamaño y luego disminuir su tamaño hasta lograr una menor complejidad computacional y un rendimiento mejor en la generalización 2.1.5.4, aplicar la eliminación de Gauss-Jordan con mínimos cuadrados lineales que se ha discutido en (161), usar la descomposición de Cholesky (162), aplicar el algoritmo clásico de Gram-Schmidt (163), usar mínimos cuadrados ortogonales (164) y emplear métodos evolutivos (165), (166), (167), etc.

La comparación de resultados entre las diferentes propuestas mencionadas no es concluyente, por lo que sigue siendo generalizado el uso de la experimentación para elegir el número de neuronas y capas ocultas. En general se procede entrenando varias redes neuronales con diferente número de capas ocultas y de neuronas en cada capa, midiendo a continuación el desempeño, por ejemplo mediante validación cruzada (detallada en 2.1.11.8). Por ello tanto los procesos de prueba-error como los de búsqueda (exhaustiva, restringida, optimizando algún objetivo, etc.) son a priori utilizables. Para el tipo de problemas abordados en este trabajo se han ensayado varios algoritmos hasta concluir que el que a continuación se describe contribuye a la mejora de resultados.

El algoritmo denominado “NNFINDER” (Neural Network Finder) es de tipo restringido y permite evaluar distintos rangos de neuronas ocultas, así como el número de las capas ocultas, el número de iteraciones máximas y el error máximo. La descripción del algoritmo NNFINDER se muestra en 2 y consiste esencialmente en la realización de iteraciones sucesivas hasta que se cumpla un criterio de parada. El criterio de parada no es fijo por lo que habrá que ajustarlo en función del tipo del problema abordado. El criterio general utilizado para ajustar el criterio de parada en la resolución de los problemas abordados en esta memoria se basa en dos magnitudes: El error y el número máximo de iteraciones. El entrenamiento no para hasta alcanzar el número máximo establecido de iteraciones o que el error sea constante o creciente. Como resultado se obtienen varias redes candidatas, y se elige la mejor empleando un criterio de decisión adaptado al problema abordado. El algoritmo NNFINDER 2 es de orden $(n \times m)$, un poco elevado sobre

3. CONTRIBUCIONES

Algorithm 2 Algoritmo NNFINDER

```
1: Descripción del Algoritmo NNFINDER.
2: for  $i = 1, 2, \dots, m$  do
3:   for  $j = 1, 2, \dots, n$  do
4:     Entrenar;
5:     if Criterio de Parada = correcto then
6:       Red Entrenada
7:     end if
8:     if Criterio de Parada = falso then
9:       Error
10:    end if
11:  end for
12: end for
```

todo cuando hay muchas entradas. Para superar este defecto, se propone modificar el criterio de parada eligiendo menos iteraciones de entrenamiento. Una vez se elige la mejor estructura (por ejemplo: mejor error de entrenamiento y menos neuronas ocultas), se puede ir aumentando el número de iteraciones para disminuir el error global de entrenamiento y el de validación.

En la figura 3.2 se muestra una captura de pantalla del ejecutable NNFINDER que se utiliza en todas las aplicaciones de esta investigación con el propósito de encontrar la mejor estructura y evaluar su eficacia.

3.2.2 Aportación sobre la inicialización de pesos

La generación de los valores iniciales de los pesos en una red neuronal, generalmente, se hace mediante un generador de números pseudoaleatorio llamado GPAN. GPAN es un algoritmo que produce una sucesión de números aleatorios, y el estado del GPAN no es exactamente aleatorio en el sentido de que la sucesión queda completamente determinada por un conjunto de valores iniciales relativamente pequeño. Otros enfoques son las implementaciones de “Nguyen Widrow Weight Initialization” y la de “Fan-In Weight Randomization”.

La técnica “Fan-In Weight Randomization” fue introducida por Simon Haykin en (168). Aquí los pesos se inicializan usando la siguiente fórmula: $w = \frac{\min}{n} + \alpha \left(\frac{\max - \min}{n} \right)$ Siendo α un número aleatorio entre -1 y 1 . Y n es el número de neuronas ocultas. Por otra parte, la técnica de “Nguyen Widrow Weight Initialization”, fue introducida por Derrick Nguyen y Bernard Widrow en (169), y para implementar esta técnica, primero se inicializan los pesos de la red neuronal con valores al azar en un rango específico. Luego se calcula un parámetro β de la

3.2 Contribuciones en las Redes Neuronales Artificiales

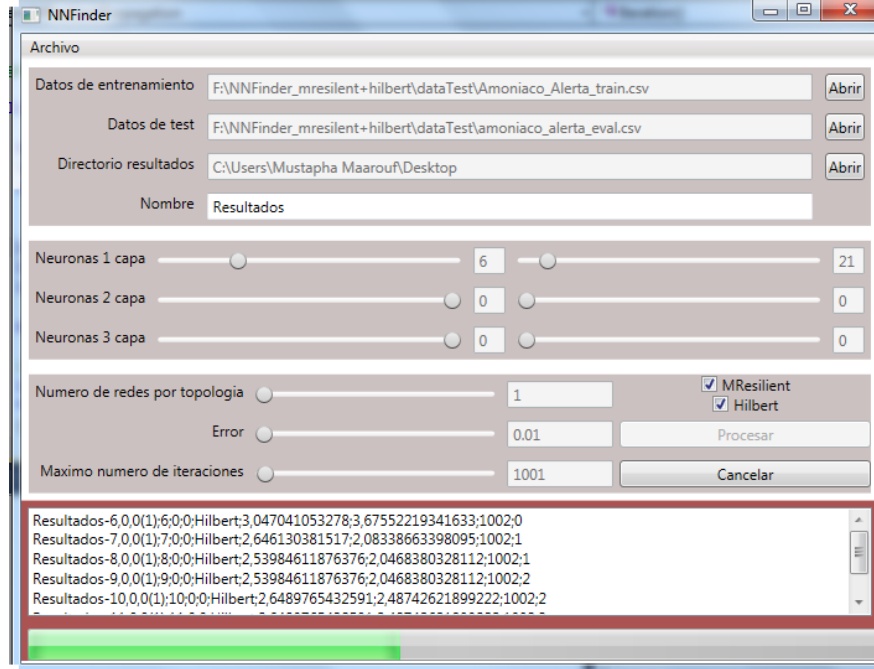


Fig. 3.2: Programa NNFINDER

siguiente manera:

$\beta = 0.7 \times n^{1/m}$ Siendo n el número de las neuronas ocultas, y m el número de las entradas.

Después se calcula la distancia euclidiana de todos los pesos:

$$D = \sqrt{\sum_{i=1}^n w_i^2}$$

Una vez calculado β y D , se modifican los pesos de la siguiente manera:

$$w_i = \frac{\beta w_i}{D}$$

Se ha demostrado empíricamente que la técnica de “Fan-In Weight Randomization” es una técnica eficaz, pero generalmente no proporciona tan buenos resultados como la técnica “Nguyen Widrow Weight Initialization” (170).

Ahora bien, las dos técnicas son mejores que la inicialización mediante un generador pseudoaleatorio de números (GPAN), aunque lo usan. Sin embargo el método de Nguyen sigue presentando la deficiencia de que también depende de la inicialización aleatoria de los pesos además de requerir más cálculos, aunque no es una tarea compleja o que necesite mucho tiempo de cálculo.

3. CONTRIBUCIONES

En 2.1.11.5, se han explicado algunas recomendaciones encontradas en la literatura para la inicialización de los pesos. En general se trata de inicializar los valores de los pesos dentro de unos rangos muy pequeños, pero la pregunta que cabe hacer es: ¿hasta qué punto este enfoque general es fiable y permite concluir que esta inicialización es buena y no?, ¿hay mejores?

Para responder a estas preguntas, es importante tener presente que a priori se desconoce la arquitectura adecuada de la red neuronal más adecuada para resolver un problema concreto. No existe ningún teorema universal que defina el número de neuronas ocultas que debe tener una red neuronal, siendo por tanto frecuente que los investigadores y aplicadores utilicen el error global de entrenamiento de la RNA como indicador para decidir el número de neuronas ocultas. También hay que tener presente, como se ha señalado en apartados anteriores, que cada vez que se entrena una red neuronal se generan los pesos iniciales (matriz de pesos de la red) de forma aleatoria. Por consiguiente en cada entrenamiento, se generan pesos iniciales distintos y se obtiene un error diferente a los del resto de entrenamientos, aunque se esté usando la misma arquitectura de la RNA (mismo número de neuronas ocultas, mismo número de capas, mismos datos de entrenamiento, mismas funciones de activación, mismo criterio de parada).

Como ilustración a continuación, se han generado dos redes neuronales con datos de entrenamiento de dos problemas distintos. Para el conjunto 1, los datos utilizados son datos reales del monitoreo de las actividades físicas de nueve individuos, para 18 diferentes actividades. Las mediciones de monitoreo han sido realizadas a través de sensores. El conjunto de datos se pueden descargar en “PAMAP project’s” <http://www.pamap.org/demo.html>. Para el conjunto 2, los datos completos y una descripción completa se pueden encontrar en el repositorio UCI de bases de datos de aprendizaje automático <http://kdd.ics.uci.edu/databases/coverttype/coverttype.html>. Los datos son de un problema real que describe el área quemada de los incendios forestales ocurridos, obtenidos del Servicio Forestal de Estados Unidos (USFS) usando sistema de Información de Recursos (RIS) de datos, se trata de predecir tipo de bosque a partir de variables cartográficas. Y se ha usado el mismo algoritmo de entrenamiento (RPROP), definido en 2.1.10.4.

La figura 3.3 corresponde a la red 1 con datos de entrenamiento (datos 1) del problema propuesto en 4.2) y la figura 3.4 corresponde a la red 2 con datos de entrenamiento (datos 2) del problema propuesto en 4.3). Se muestra la evolución del error de entrenamiento para la misma configuración de la red neuronal (número de entradas / salidas, neuronas ocultas y capas del modelo, son fijos) pero en diferentes sesiones de entrenamiento. Como se puede observar la evolución del entrenamiento es diferente en cada caso.

3.2 Contribuciones en las Redes Neuronales Artificiales

Nota 8 En las figuras, el símbolo $E - i$, $i \in \{1, 2, \dots\}$, se refiere a la ejecución i .

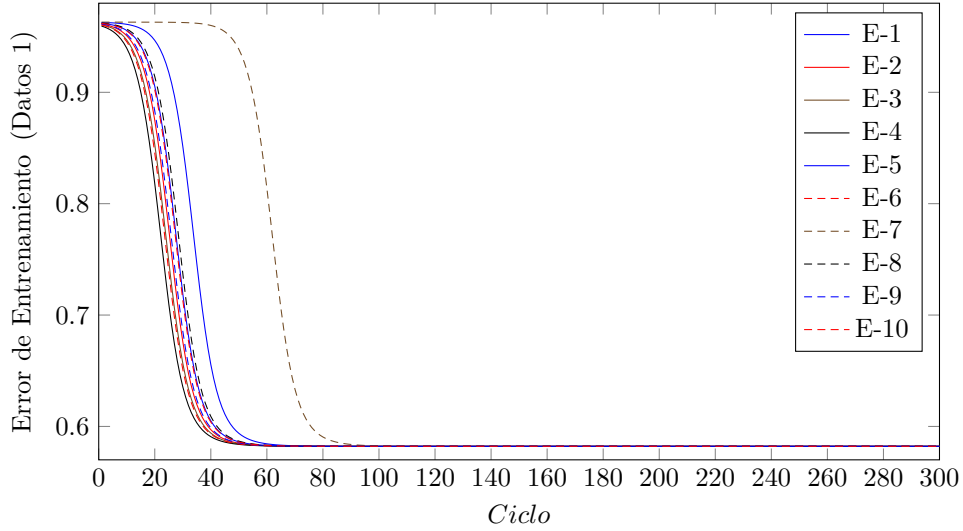


Fig. 3.3: Evolución del error (Datos 1)

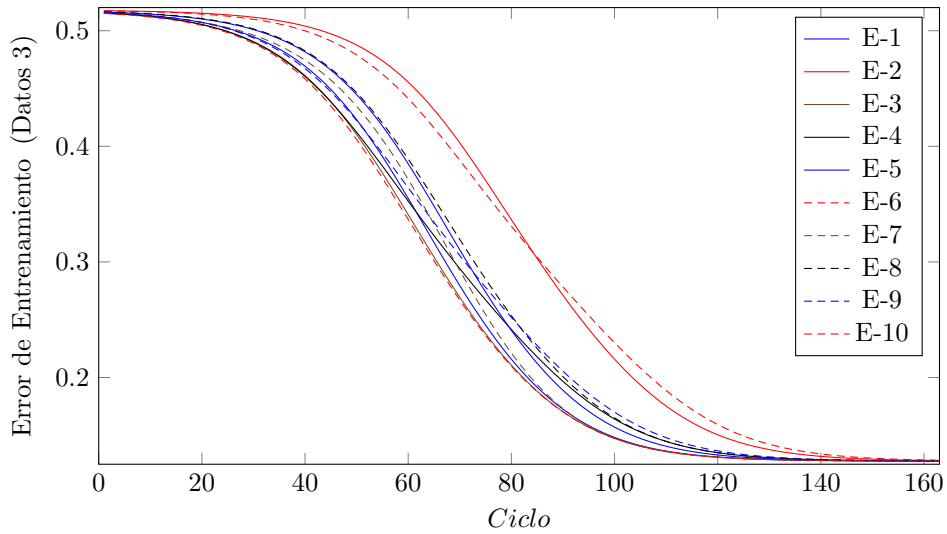


Fig. 3.4: Evolución del error (Datos 2)

La aportación de esta tesis pretende encontrar una forma determinista para la inicialización de los pesos, que satisfaga los siguientes requisitos:

- 1. Garantizar la convergencia durante el entrenamiento
- 2. Garantizar el mismo resultado.

3. CONTRIBUCIONES

- 3. Ofrecer buen resultado

En las redes unidireccionales, ante un patrón de entrada, las neuronas responden proporcionando directamente, la salida del sistema. Al no existir bucles de realimentación no existe problema en relación a su estabilidad. Por el contrario, las redes con realimentación (las RNA empleadas en esta tesis) son sistemas dinámicos no lineales, que requieren ciertas condiciones para que su respuesta *acabe convergiendo* a un estado estable o un punto fijo (23, 171).

Una serie de teoremas generales (teorema de Cohen-Grossberg (172), teorema de Cohen-Grossberg-Kosko (173), etc.) indican las condiciones que se deben cumplir para asegurar la estabilidad y la convergencia. Básicamente, el teorema de Cohen-Grossberg (172) establece las condiciones que garantizan la estabilidad asintótica global de un sistema dinámico no-lineal. Este teorema se basa en el concepto de la función de energía de Lyapunov.

El método de Lyapunov establece que para un sistema dinámico, como puede ser una RNA, de variables de entrada $x_i, i \in \{1, \dots, n\}$, si el sistema está en reposo solamente en el origen, entonces existen derivadas de las ecuaciones que lo describen en todo el dominio, las variables están acotadas y se puede encontrar una función, llamada función de energía de Lyapunov, de las variables (x_1, \dots, x_n) , tal que:

$$\dot{L} = \sum_{i=1}^n \frac{dL}{dx_i} \leq 0, \forall \dot{x}_i$$

Entonces el sistema converge para todas las posibles entradas (x_1, \dots, x_n) .

En esencia, se trata de ser capaz de encontrar una cierta función energía, que cumpla con las propiedades anteriormente descritas, lo que se puede interpretar en el campo de las RNA como que la **función de minimización del error** es asociable a esa **función de energía**.

Considérese a continuación el **teorema de Cohen-Gorssberg** (172), que establece lo siguiente:

Para cualquier sistema dinámico no lineal que se pueda describir de la forma siguiente:

$$\frac{dx_i}{dt} = \alpha_i(x_i) \left(\beta_i(x_i) - \sum_{j=1}^n w_{ij} S_j(x_j) \right), i = 1, 2, \dots, n$$

Tal que:

- La matriz $\|m_{ij}\|$ es simétrica y $m_{ij} \geq 0, \forall i, j$
- La función $\alpha_i(s)$ es continua $\forall s \geq 0$
- $\alpha_i(s) \geq 0, \forall s \geq 0; S_i(s), \forall s \geq 0$.

3.2 Contribuciones en las Redes Neuronales Artificiales

Entonces la función:

$$E = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n w_{ij} S_i(x_i) S_j(x_j) - \sum_{i=1}^n \int_0^{x_i} S'_i(\theta_j) \beta_i(\theta_i) d\theta_i$$

Es una función de energía de Lyapunov para el sistema, y el sistema es estable.

Por lo tanto en el caso de una red neuronal, considerando la función de minimización del error E , y la matriz de pesos w_{ij} .

Entonces $\dot{E} = \sum_{i=1}^n \frac{dE}{dx_i} \leq 0, \forall x_i$ siempre y cuando se cumplen las siguientes condiciones:

- Simetría y no-negatividad de los w_{ij} , es decir $w_{ij} = w_{ji} \geq 0, \forall i, j$
- S_i es una función diferenciable, no-negativa y no decreciente.
- α_i y β_i son funciones continuas, y $\alpha_i(t) \geq 0, \forall t \geq 0$

Entonces el sistema es estable y converge a un equilibrio en finitos puntos o finitas iteraciones.

Grossberg y Cross aplicaron el mismo concepto para demostrar la estabilidad de función de energía Lyapunov en diversas redes neuronales feedforward (174). Por otra parte, Hopfield demostró que su modelo neuronal completamente interconectado era estable en el caso de que la matriz de pesos fuese simétrica y de diagonal nula (171).

La aportación de esta tesis, consiste en considerar la inicialización de los pesos de una RNA feedforward mediante la matriz de Hilbert.

Esta matriz fue introducida por Hilbert en 1894 (175), que es una matriz simétrica definida positiva, definida en cada unidad según la fórmula siguiente: $H_{ij} = \frac{1}{i+j-1}$ En la matriz siguiente, un ejemplo de una matriz de Hilbert de orden $(n \times m)$ es:

$$\begin{array}{ccccccc} 1 & \frac{1}{2} & \frac{1}{3} & \cdot & \cdot & \cdot & \frac{1}{n-2} & \frac{1}{n-1} & \frac{1}{n} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \cdot & \cdot & \cdot & \frac{1}{n-1} & \frac{1}{n} & \frac{1}{n+1} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \cdot & \cdot & \cdot & \frac{1}{n} & \frac{1}{n+1} & \frac{1}{n+2} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \frac{1}{m-2} & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \frac{1}{m+n-3} \\ \frac{1}{m-1} & \frac{1}{m} & \frac{1}{m+1} & \cdot & \cdot & \cdot & \cdot & \frac{1}{m+n-2} & \frac{1}{m+n-1} \\ \frac{1}{m} & \frac{1}{m+1} & \frac{1}{m+2} & \cdot & \cdot & \cdot & \frac{1}{m+n-3} & \frac{1}{m+n-2} & \frac{1}{m+n-1} \end{array}$$

Por lo tanto, aplicar esta matriz a una red neuronal, que a su vez satisface las condiciones del teorema Cogen-Gorssberg, garantiza la convergencia de la red neuronal, algo primordial que en los métodos que se basan en la inicialización aleatoria no siempre ocurre.

3. CONTRIBUCIONES

Por otra parte, al tratarse de un método de inicialización determinista se garantiza la obtención del mismo resultado si se vuelve a repetir el proceso de entrenamiento con la misma arquitectura. Esta característica (que podría constituir una ventaja frente a los métodos existentes), debería permitir la resolución de algunos de los inconvenientes mencionados anteriormente para aplicar una RNA a determinados problemas. Citamos el hecho de que obtener el mismo resultado cada vez que se repite el entrenamiento, permite comparar si al añadir/eliminar más neuronas ocultas mejora en términos absolutos el resultado del entrenamiento de la RNA.

El algoritmo para la inicialización de los pesos se define de la manera siguiente en 3: El

Algorithm 3 Inicialización de pesos

```
1: Inicialización de los pesos
2:  $m$  número de entradas
3:  $n$  número de neuronas ocultas
4:  $l$  número de salidas
5: for  $i = 1, 2, \dots, m$  do
6:   for  $j = 1, 2, \dots, n$  do
7:      $w_{ij} = \frac{1}{i+j-1}$ 
8:   end for
9: end for
10: for  $j = 1, 2, \dots, n$  do
11:   for  $k = 1, 2, \dots, l$  do
12:      $w_{jk} = \frac{1}{k+j-1}$ 
13:   end for
14: end for
```

planteamiento anterior satisface los dos primeros requisitos mencionados al principio de esta sección. Queda pendiente el tercer requisito "*ofrecer buenos resultados*", que se analizará en el capítulo dedicado a la experimentación de esta tesis. 4

Los resultados de la evolución del error durante el entrenamiento, junto con las evoluciones del error mediante la inicialización de los pesos aleatoriamente, se presentan en 4.1, para poder realizar una comparación con otras técnicas de inicialización bien conocidas y aplicadas.

3.2.2.1 Aportación sobre las funciones de activación

Para los retos $I - RAMS$ abordados en este trabajo de investigación es importante la velocidad de todo el proceso de obtención de la red neuronal entrenada, por ello se ha analizado la posibilidad de acelerar el proceso de entrenamiento. Tras varias formulaciones se ha desarrollado el concepto de "suavizado" de las funciones de activación que les confiere la propiedad de

3.2 Contribuciones en las Redes Neuronales Artificiales

activación para valores menores que el estándar en RNA. La forma escogida de realizar este “suavizado” ha sido la introducción de un factor en la formulación que permite además el escalado de las activaciones para diferentes valores del mismo. Dado que para problemas no lineales se usan generalmente la función sigmoideal “Sigm” o la función tangente hiperbólica “Tanh”, la formulación del “suavizado” se ha centrado en estas dos funciones. Por otra parte, las modificaciones de “suavizado” llevadas a cabo en las funciones de activación cumplen los requisitos establecidos en 2.1.5.5, lo que garantiza que se pueden implementar para la resolución de problemas no lineales.

Suavizado de la función sigmoideal Se ha definido la función sigmoideal anteriormente en 7, y la idea es usar la forma general de las funciones sigmoideales, que añaden un factor α de suavizado en 3.1, de la siguiente manera:

$$y = 1/(1 + e^{-\alpha x}) = f(x)$$

$$y = 1/(1 + e^{-\alpha x}) = f(x) \quad (3.1)$$

Donde $\alpha \in \mathbb{R}$

La ecuación 3.1 también define una función continua, que tiende a 1 cuando $x \gg 0$, y tiende a 0 cuando $x \ll 0$, igual a $\frac{1}{2}$ cuando $x = 0$.

. Y calculando su derivada se obtiene:

$$\begin{aligned} \frac{dy}{dx} &= \frac{d}{dx} \left(\frac{1}{1 + e^{-\alpha x}} \right) \\ \frac{dy}{dx} &= \frac{d}{dx} (1 + e^{-\alpha x})^{-1} \\ \frac{dy}{dx} &= -1 \times (1 + e^{-\alpha x})^{-2} \times e^{-\alpha x} \times -\alpha \\ \frac{dy}{dx} &= \frac{\alpha}{1 + e^{-\alpha x}} \times \frac{e^{-\alpha x}}{1 + e^{-\alpha x}} \\ \frac{dy}{dx} &= \frac{\alpha}{1 + e^{-\alpha x}} \times \frac{1 + e^{-\alpha x} - 1}{1 + e^{-\alpha x}} \\ \frac{dy}{dx} &= \frac{\alpha}{1 + e^{-\alpha x}} \times \left(\frac{1 + e^{-\alpha x}}{1 + e^{-\alpha x}} - \frac{1}{1 + e^{-\alpha x}} \right) \end{aligned}$$

Por lo tanto:

$$\frac{dy}{dx} = \alpha y(1 - y) \quad (3.2)$$

En la figura 3.5, se muestra la presentación grafica de la función α -Sigm, para: $\alpha = 1, 2, 3$.

3. CONTRIBUCIONES

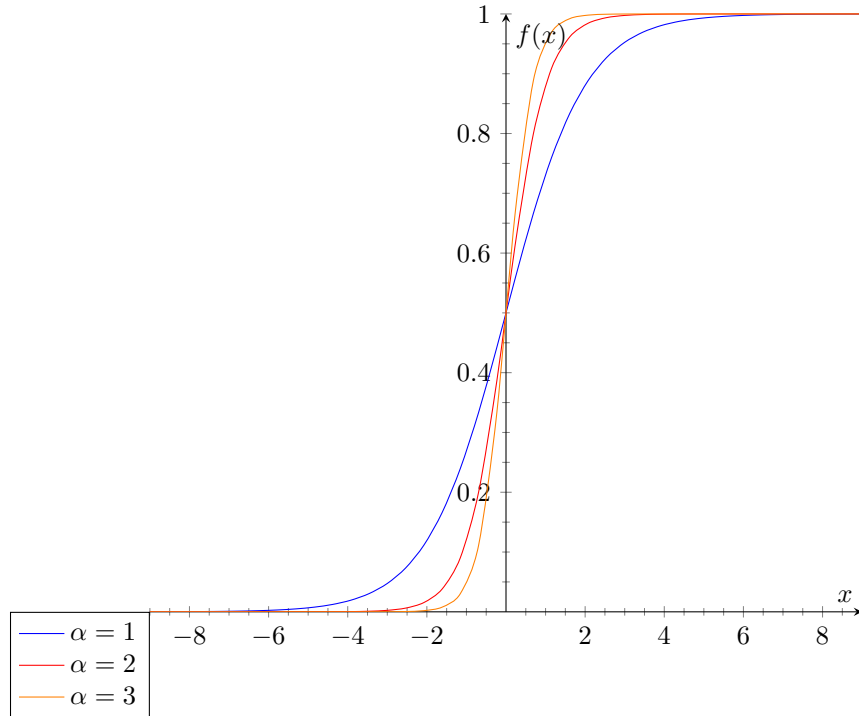


Fig. 3.5: Función α -Sigm

Suavizado de la función tangente hiperbólica La función tangente hiperbólica anteriormente definida en 2.4, también se usara en su forma general, que añade el factor α de suavizado de la manera siguiente:

$$y = \tanh\left(\frac{\alpha x}{2}\right) \text{ o:}$$

$$y = \frac{1 - e^{-\alpha x}}{1 + e^{-\alpha x}} \quad (3.3)$$

Es decir: $y = \frac{2}{1 + e^{-\alpha x}} - 1$

$y = 2f(x) - 1$, siendo f la función α -sigmoidal anterior.

Calculando la derivada de la α -Tanh se obtiene:

$$\begin{aligned} \frac{dy}{dx} &= 2 \frac{df(x)}{dx} \\ \frac{dy}{dx} &= 2\alpha f(x)(1 - f(x)) \end{aligned} \quad (3.4)$$

En la figura 3.6, se muestra la presentación grafica de la función α -Tanh, para: $\alpha = 1, 2, 3$

La verificación y comparación con las funciones de activación sin suavizado, se analizará mediante la experimentación detallada en el capítulo 4, que es el siguiente de esta memoria .

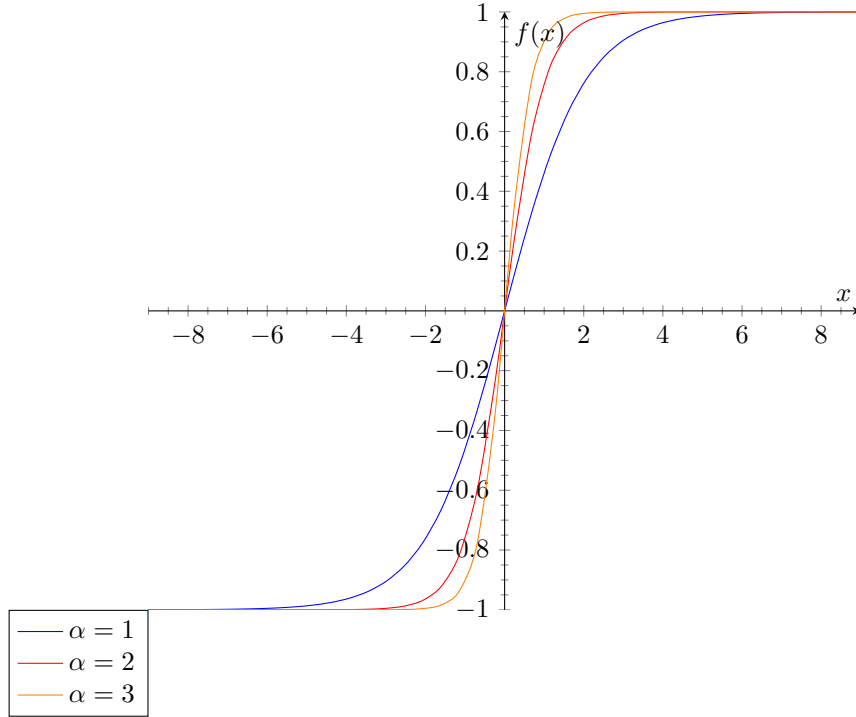


Fig. 3.6: Función α -Tanh

3.2.2.2 Aportación sobre la fiabilidad

Para el cálculo de cuán fiable o asumible es la respuesta de una red neuronal, se añade a la salida de la red neuronal el cálculo de un nuevo intervalo de confianza.

La posibilidad de definir intervalos de confianza consistentes en la salida de las RNA es, a priori, muy importante ante ciertas aplicaciones *I-RAMS* de las RNA, como se ha mencionado en 2.1. En este trabajo para la definición y el cálculo del intervalo de confianza se han considerado dos conceptos. El primero está basado en la información obtenida de las conexiones de las neuronas de la red neuronal una vez entrenada. Que a su vez se modifica con un segundo concepto, que añade información obtenida del conjunto de entrenamiento (la densidad del dato dentro del conjunto de entrenamiento).

Estos dos conceptos han sido definidos en 2.1.12 y se ha desarrollado un algoritmo llamado NNCI (Neural Network Confidence Interval), que permite calcular el intervalo de confianza, como se ve en la figura 3.7. El algoritmo 4 explica los pasos a seguir: El algoritmo “NNCI” ha sido aplicado en todos los experimentos que describe esta memoria, evaluando el impacto de ello en los resultados obtenidos.

3. CONTRIBUCIONES

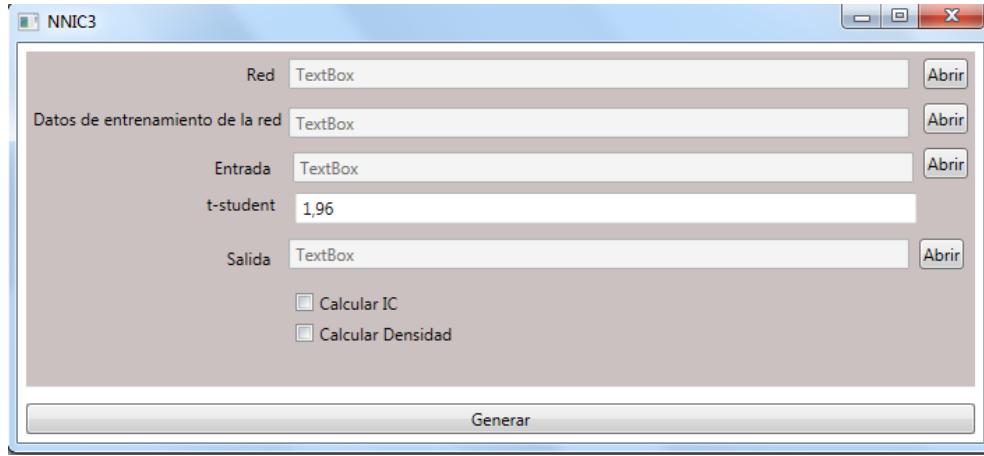


Fig. 3.7: Programa NNIC

Algorithm 4 Intervalos de Confianza

- 1: **for** Cada Entrada **do**
 - 2: $c = t_{n-p}^{\frac{\alpha}{2}} s \left(1 + f_0^T (F^T F)^{-1} f_0 \right)^{\frac{1}{2}}$ (ver 2.1.12.2)
 - 3: $C_s = \frac{2c}{1 + \frac{\rho}{\rho_{\max}}}$ (ver 2.1.12.2)
 - 4: $y = y \pm C_s$
 - 5: **end for**
-

3.3 Contribuciones en la Aplicación de RNA en la Predicción de la Vida Útil Remanente en Equipos Industriales

La Vida Útil Remanente (VR, 2.3.1) es el tiempo que queda antes de se produzca un fallo o avería debido a la edad, el estado actual del equipo o su perfil de funcionamiento. En algunas situaciones, especialmente en caso de fallo o avería catastrófica, es muy importante predecir el tiempo de funcionamiento adecuado que le queda al equipo para compararlo con alguna magnitud de interés (por ejemplo el tiempo hasta la siguiente inspección por parte del personal de mantenimiento). La estimación de la VR es una tarea importante como parte de las labores de Pronóstico que se incluyen en el proceso de determinación del estado de salud de equipos industriales. Con frecuencia se usa el denominado Índice de Salud (IS), una variable definida en el rango $[0, 1]$, como indicador de la salud.

En esta investigación, se analiza la aplicación de las RNA para la estimación de la VR en el caso de información incompleta (falta de información sobre fallo o sobre su forma de evolución), como es común en los problemas reales. Para ello se han definido tres etapas: Conceptualización de la VR, Caracterización del caso de Información Incompleta y Características de las RNA

propuestas para estimar la VR.

3.3.1 Caracterización de la VR

La VR de un equipo o sistema técnico es una variable que es función entre otros de la edad actual, el entorno operacional y de los parámetros del equipo observados mediante el monitoreo de la condición. El resultado de la estimación de VR está inevitablemente afectado por la incertidumbre sobre el futuro, por lo que es razonable plantearse la estimación en términos probabilistas y el uso de intervalos de confianza.

En el capítulo 2 se han sido citadas las técnicas y las metodologías que se suelen usar para la estimación de la VR. En general se requiere tanto información histórica (sobre fallos o eventos de interés) como información actual (variables medidas en el equipo por sistemas SCADA o por sensores, etc.). Aún así el conocimiento y las metodologías para esta estimación demandan considerable trabajo de investigación actualmente.

3.3.2 Caracterización del caso de Información Incompleta

Después de un proceso de análisis de las posibles fuentes de datos que permitan caracterizar el caso de Información Incompleta se ha optado por el problema conocido como “PHM08 Prognostics Data Challenge Database”, seleccionando dos enfoques distintos. Sabiendo que no se dispone de ninguna información acerca del fallo, el primero enfoque consiste en suponer en primer lugar que la curva de VR sigue una degradación lineal, en segundo lugar se modifica esta hipótesis según los resultados obtenidos mediante la red neuronal, dado que durante su entrenamiento, la RNA adquiere conocimiento de la información disponible en los datos, este enfoque ha sido utilizado aplicado también en (176). El segundo enfoque utiliza únicamente la información disponible; el componente funciona bien al principio y falla al final, entonces la red neuronal se entrena muy pocos datos, los primeros se les asigna 1 (dado que el componente funciona bien al principio) como salida y los últimos se les asigna 0 (porque el componente falla) por lo tanto se espera que la red neuronal determina por sí misma el proceso de degradación oculto. Analizar estos dos enfoques distintos nos permite evaluar y comparar los resultados a un mismo problema.

Un ejemplo de la no linealidad de las mediciones de los sensores en los datos de este problema, se muestra en la figura 3.8. La no linealidad y la dificultad de encontrar patrones no se puede notar a simple vista.

3. CONTRIBUCIONES

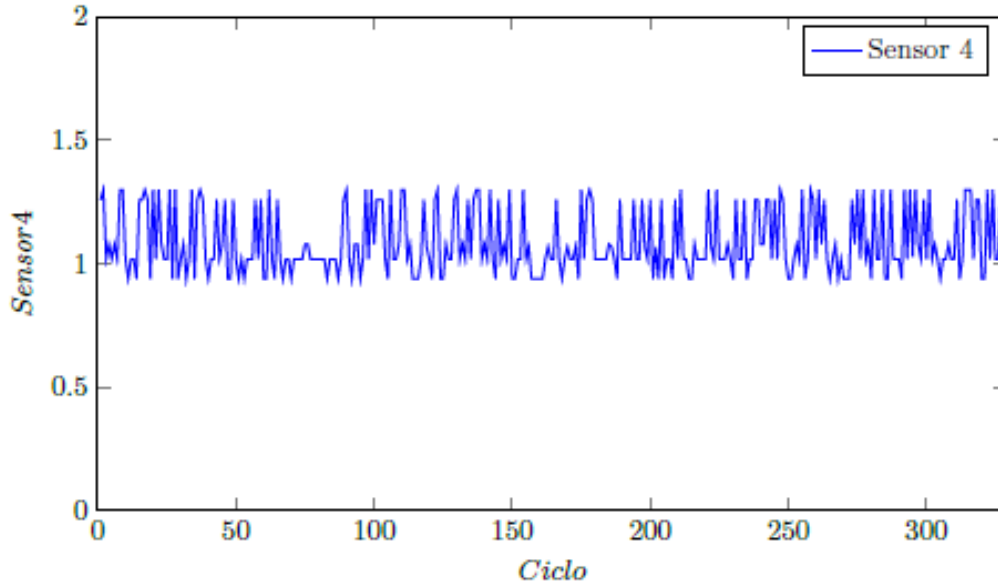


Fig. 3.8: Sensor 4

Una característica significativa del conjunto de datos del desafío de PHM es que contiene series temporales y por lo tanto la posibilidad de consideración del tiempo (pasado y presente) es una característica importante que debe permitir la metodología que se emplee para analizar los datos. El método más rápido y más simple para representar el tiempo es considerar cada punto de tiempo de forma independiente y crear una predicción en cada paso de tiempo. Una representación alternativa sería considerar el uso de una representación en espacios temporales, en el que se genera una secuencia de instancias utilizando una ventana fija de deslizamiento. Sin embargo utilizar espacios temporales, tiene el inconveniente de aumentar el número de dimensiones proporcionalmente con cada paso de tiempo representado, dando lugar a los problemas asociados con la “maldición de la dimensionalidad”. Leto Peel en (177) comprobaron que el rendimiento de la predicción no mejora significativamente cuando se usan ventanas temporales, debido principalmente al aumento de las demandas computacionales. Por siguiente, la representación elegida fue predecir la VR mediante puntos de tiempo individuales.

De la discusión anterior se concluye el interés de usar las RNA ya que han demostrado su eficiencia resolviendo problemas no lineales de la tipología caracterizada para la estimación de la VR.

3.3.3 Características de las RNA propuestas para estimar la VR

Tras una serie de análisis y algunos experimentos se han identificado las características de las RNA que esta contribución propone. Estas son:

- Aplicación del NNIC descrito anteriormente 4, para proporcionar un indicador del resultado de la RNA, basado en la información disponible en los datos, y en los pesos de la RNA una vez entrenada.
- Las métricas adicionales que se describirán a continuación.

Se van a definir nuevas métricas que incluyen algunas métricas tradicionales, por ello se comienza describiendo las métricas tradicionales, estas son:

- La exactitud (EXAC)cuantifica cuanto más las predicciones del modelo son buenas, cerca a cero significa que las predicciones no son buenas, mientras que una exactitud cerca de uno corresponde a buenas predicciones.

$$EXAC = \frac{1}{T} \sum_{t=1}^T e^{-\frac{|IS_{real}(t) - IS(t)|^\alpha}{IS_{real}(t)}} \quad (3.5)$$

Donde: $\alpha \in [1, 2]$.

- La precisión (PREC)es una medida que cuantifica la dispersión del error de la predicción alrededor de su media.

$$PREC = \sqrt{\frac{\sum_{t=1}^T (\varepsilon(t) - \bar{\varepsilon})^2}{T}} \quad (3.6)$$

Donde:

$$\bar{\varepsilon} = \frac{\sum_{t=1}^T \varepsilon(t)}{T} = \frac{\sum_{t=1}^T (IS_{real}(t) - IS(t))}{T}$$

- El error cuadrático medio (MSE) cuantifica la diferencia entre los valores implícitos en un estimador y los verdaderos valores de la cantidad que se estima.

$$MSE = \frac{\sum_{t=1}^T (IS_{real}(t) - IS(t))^2}{T} \quad (3.7)$$

3. CONTRIBUCIONES

- El porcentaje medio de error absoluto (MAPER): esta medida cuantifica el error medio del modelo en porcentaje:

$$MAPER = \frac{1}{T} \sum_{t=1}^T \left| \frac{100 \cdot \varepsilon(t)}{IS_{real}(t)} \right| \quad (3.8)$$

- La exactitud relativa “Relative accuracy measure (RA)”: es una medida que permite evaluar la exactitud de la estimación de la RUL en diferentes momentos:

$$RA = 1 - \frac{|IS_{real}(t) - IS(t)|}{IS_{real}(t)} \quad (3.9)$$

Estás métricas son útiles y han sido ampliamente utilizadas en aplicaciones científicas. Sin embargo proponen información acerca de todo el conjunto de la muestra, pero no sobre un punto/dato individual. Por consiguiente, su uso es bueno para comparar las redes neuronales una vez entrenadas con objeto de seleccionar la más adecuada (en general la de menor error).

En esta tesis se proponen como novedosos los siguientes indicadores:

Un indicador que proporciona el porcentaje de error en cada punto, que viene definido así:

$$Ind1_i = 100 \cdot \exp \left(\frac{-(real_i - estimado_i)^2}{MSE} \right), \forall i \quad (3.10)$$

Se ha elegido esta fórmula por las siguientes razones:

- En primer lugar por utilizar la función exponencial que es una función estrictamente creciente, por lo tanto se incrementa el contraste entre el error cometido en los diferentes puntos.
- La pendiente del gráfico en cualquier punto será la altura de la función en ese punto.
- Siendo

$$MSE = \frac{\sum_{i=1}^n (x_i - y_i)^2}{n}$$

, se ha usado $\frac{(x_i - y_i)^2}{MSE}$ para que muestre lo que representa el error cometido en un punto respecto al conjunto global.

- El signo negativo en la exponencial es debido a que $0 < \exp(x) \leq 1, \forall -\infty < x \leq 0$, y se tiene que:

$$0 > \frac{-(x_i - y_i)^2}{MSE} = \frac{-n(x_i - y_i)^2}{\sum_{i=1}^n (x_i - y_i)^2} > -n$$

Por lo tanto, $0 < \exp \left(\frac{-(x_i - y_i)^2}{MSE} \right) < 1$, es decir: $0 < Ind1_i < 100, \forall i$

3.3 Contribuciones en la Aplicación de RNA en la Predicción de la Vida Útil Remanente en Equipos Industriales

- Este indicador permite dar un valor que se interpreta en término de porcentaje ($\times 100$).

Un indicador basado en el indicador anterior y el intervalo de confianza Como se ha descrito anteriormente en refaportacionconfianza, se dispone del intervalo de confianza de forma individual, (se le va denotar por Ind_1), denotando al indicador anterior por Ind_2 , se propone otro indicador que refleja tanto la información dada por Ind_1 y Ind_2 , de la forma siguiente:

$$Ind_3 = (Ind_1 + Ind_2)/2$$

El comportamiento de estos indicadores se analizará en el capítulo dedicado a experimentación de esta tesis.

3. CONTRIBUCIONES

3.4 Contribuciones en la Aplicación de las RNA en el Comportamiento de Incendios Forestales

La investigación del comportamiento de Incendios Forestales se realiza desde muy diversos enfoques y la ciencia ha contribuido con grandes avances en la modelización, como se ha comentado en 2. Entre las últimas tendencias cabe citar la fuerte orientación hacia la ayuda a la toma de decisiones por parte de los responsables de extinción. Esta toma de decisiones se caracteriza por la necesidad de adoptar medidas en cortos espacios de tiempo y en base a información de campo, experiencia previa y resultados de los modelos. Entre otras los Directores de Extinción consideran muy valiosas las respuestas a las siguientes preguntas:

- ¿Cuándo alcanzará el incendio zonas críticas?
- ¿En qué dirección o direcciones se propagará el incendio?
- ¿Cuánta superficie puede quemar un incendio?
- ¿Cuál será la velocidad de propagación del incendio?
- ¿Cuáles serán el área, el perímetro y la distancia recorrida por el incendio en la dirección de propagación principal después de 1 hora o más?
- ¿Será un incendio de alta intensidad o de baja intensidad?
- ¿Será un incendio continuo o discontinuo?
- ¿Será un incendio de superficie o de copas de árboles?
- ¿Qué tan difícil será controlar la extinción?
- ¿Cuál es el riesgo del incendio?

Con frecuencia los modelos tradicionales son costosos en términos de tiempo de CPU y por tanto pueden resultar útiles para la planificación de las emergencias (Definir Planes de Extinción), pero poco útiles para responder cada vez que se requiera a las preguntas anteriormente citadas y, por ende, poco útiles para la ayuda a la toma de decisiones en tiempo real. Como resultado de sus trabajos Gimblett y Ball (178) señalaron que la toma de decisiones en los recursos naturales a menudo conduce complejidad más allá del alcance de las técnicas estadísticas empíricas, y requiere de enfoques a veces más heurísticos que algorítmicos, conclusión que sigue

3.4 Contribuciones en la Aplicación de las RNA en el Comportamiento de Incendios Forestales

siendo totalmente válida hoy en día. Otros autores han analizado las diferentes fuentes de error presentes en los métodos clásicos (7, 8, 9, 10, 11), identificando entre otros la falta de aplicabilidad, inexactitud interna y errores en los datos de entrada (datos anómalos y/o inexistentes). De entre los errores en los datos de entrada cabe destacar el asociado a la inexactitud sobre el tipo de combustible asociable a la masa forestal, debido a la no uniformidad, la discontinuidad, la heterogeneidad y la presencia de múltiples capas de vegetales.

La aplicación propuesta en esta investigación consiste en usar las RNA como método alternativo a las metodologías tradicionales (2.2), que no pretenden sustituirlos pero si complementarlos, especialmente en el apoyo a la toma de decisiones aportando respuestas rápidas y suficientemente precisas a algunas de las preguntas antes citadas.

El modelo neuronal buscado se inspira en las mismas entradas y salidas que los modelos convencionales, pero pretende crear y encontrar arquitecturas neuronales que tengan en cuenta solo algunas entradas y no todas, con la condición de conservar una calidad similar en la predicción de las salidas (parámetros) que los métodos tradicionales. La arquitectura del modelo propuesto, denominado ANNFBP, consiste de una red neuronal feedforward de tres capas. La primera y la tercera son reservadas para las entradas y las salidas deseadas sucesivamente. El algoritmo de entrenamiento usado está definido en definido en 2.1.10.4. Respecto a las funciones de activación tanto para la segunda capa como para la última, se ha utilizado la tangente hiperbólica definida en la ecuación 2.4. Respecto a las entradas del modelo neuronal, han sido las misma que las de los modelos convencionales (detallados en 2). La arquitectura obtenida del modelo neuronal es la siguiente: Para disminuir la influencia de los errores, antes mencionados, asociados a los datos se propone optimizar las entradas analizando si existen subconjuntos de las mismas que conducen a resultados aceptables. El algoritmo utilizado para optimizar (reducir) las entradas se detalla en la siguiente figura 3.10. Tras la aplicación de este algoritmo las entradas elegidas para la red neuronal se ilustran en la figura 3.11 y 3.12, con una estrella, para obtener las salidas deseadas a su vez marcadas con dos estrellas. Las figuras reflejan respectivamente el modelo forestal canadiense definido en 2.14 y la estructura del sistema canadiense de predicción del comportamiento de los incendios forestales definido en 2.12.

En el capítulo 4, se analiza el rendimiento de este modelo con dos caso de estudio, y se compararán los resultados del modelo neuronal con los resultados de otros modelos.

3. CONTRIBUCIONES

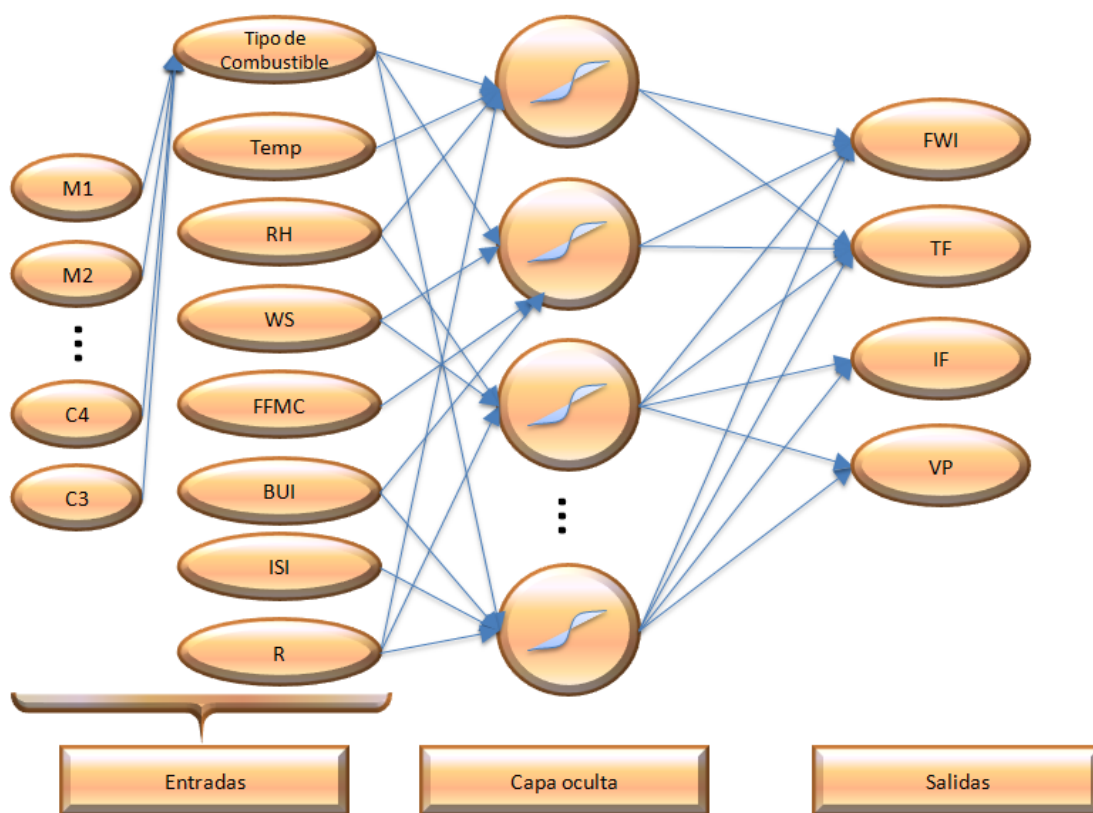


Fig. 3.9: Estructura del modelo ANNFBP

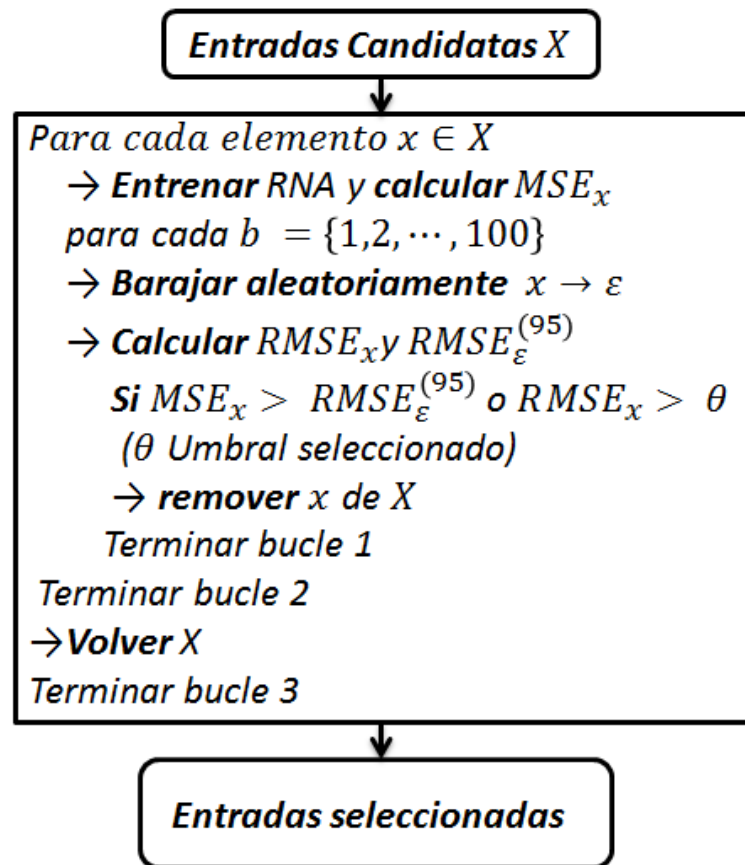


Fig. 3.10: Algoritmo para la reducción de entradas

3. CONTRIBUCIONES

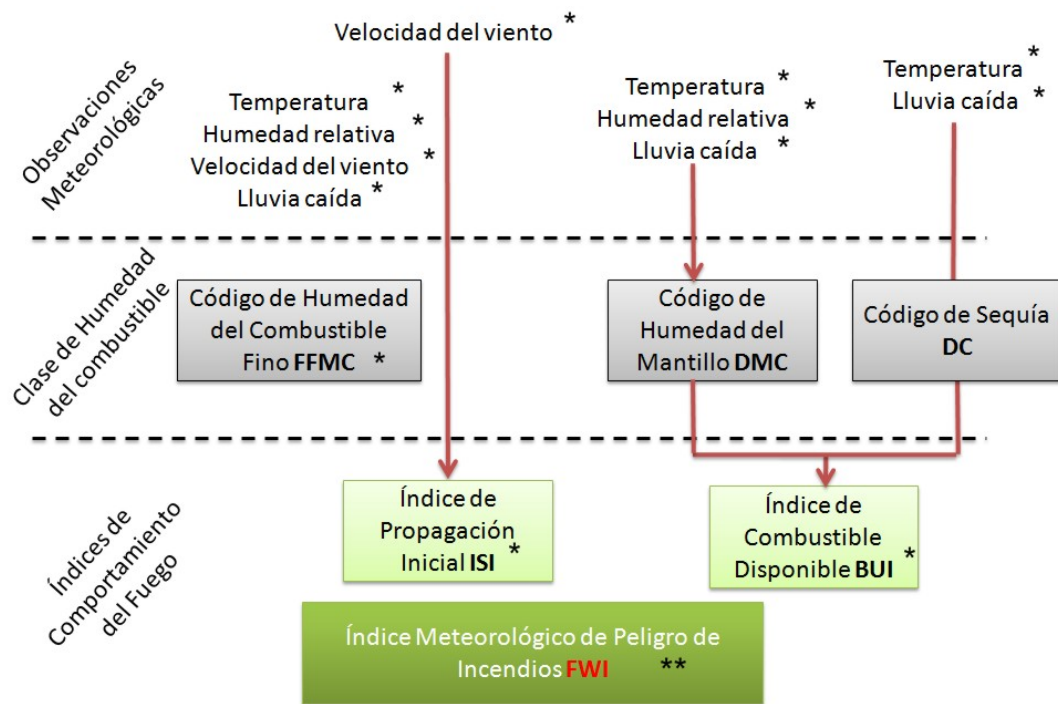


Fig. 3.11: El índice de FWI

3.5 Contribuciones en la Aplicación de las RNA en la estimación de la demanda de energía eléctrica

La estimación de la demanda de energía eléctrica (DEE) es muy importante para lograr una operación segura y económica de los sistemas eléctricos. Sobra indicar la importancia de la electricidad, y el alcance de su empleo en el mundo actual. La planificación del sistema de energía eléctrica abarca un conjunto de problemas de optimización interrelacionados entre ellos, los cuales requieren de una predicción eficiente de la demanda. Los factores económicos, los técnicos y la gestión de la demanda son claves en la comprensión de las tendencias a largo plazo.

Las tendencias futuras de energía son fundamentales para la formulación de políticas que permitan el desarrollo de nuestra sociedad. La aproximación de las técnicas de predicción es importante para intentar reducir la incertidumbre de la demanda y así obtener una programación óptima y realista. La estimación de la DEE es un problema de interés en la comunidad científica. En la actualidad nos encontramos frente a unos nuevos acontecimientos climáticos, sociales, económicos y tecnológicos, que dan lugar a nuevos conceptos y situaciones que generan incertidumbre y dificultan la estimación.

3.5 Contribuciones en la Aplicación de las RNA en la estimación de la demanda de energía eléctrica

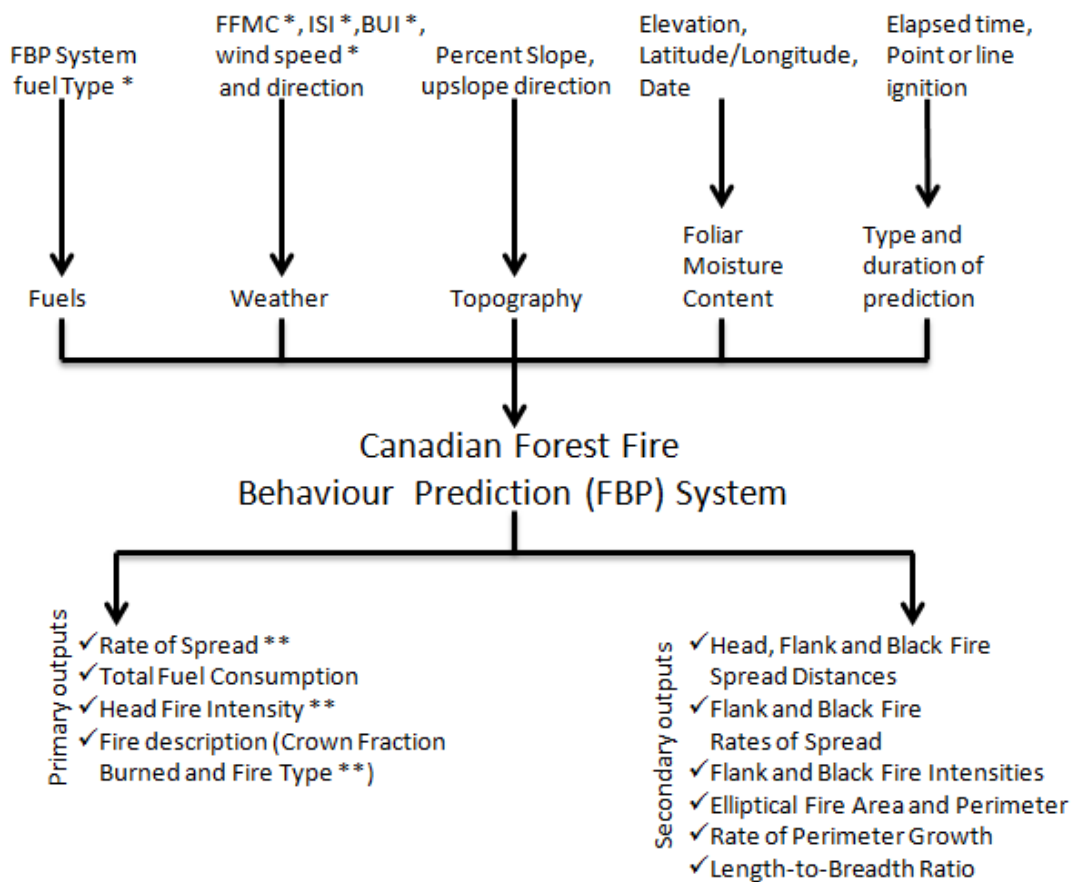


Fig. 3.12: Estructura del sistema FBP

3. CONTRIBUCIONES

Disponer de nuevas herramientas/modelos que pueden hacer frente a estos cambios, es indudablemente necesario, para evitar pérdidas económicas y no abusar de los recursos naturales sin necesidad. Por consiguiente, se ha propuesto comparar la predicción de la actual DEE usando RNA con las predicciones de otros métodos tanto lineales como no lineales. De entre los métodos lineales por su amplio uso y aceptación, se han elegido los modelos de regresión logarítmica y lineal múltiple. Estos se compararan con otros basados en Inteligencia Artificial, de entre los que por los mismo criterios se han seleccionado los siguientes: Máquina de Soporte Vectorial (con núcleo lineal y con núcleo gaussiano), Algoritmo Genético y una RNA.

Se ha creado un modelo neuronal, con una estructura básica (una capa oculta), teniendo en cuenta solo tres variables (entradas del modelo); Población (POB), Índice de Precios al Consumidor (IPC) y Producto Interior Bruto (PIB) (ver figura 3.13). Se añadirá al modelo la consideración de intervalos de confianza.

Las propuestas anteriormente mencionadas en este capítulo serán validadas o rechazadas mediante el diseño experimental detallado en el capítulo 4 de esta memoria.

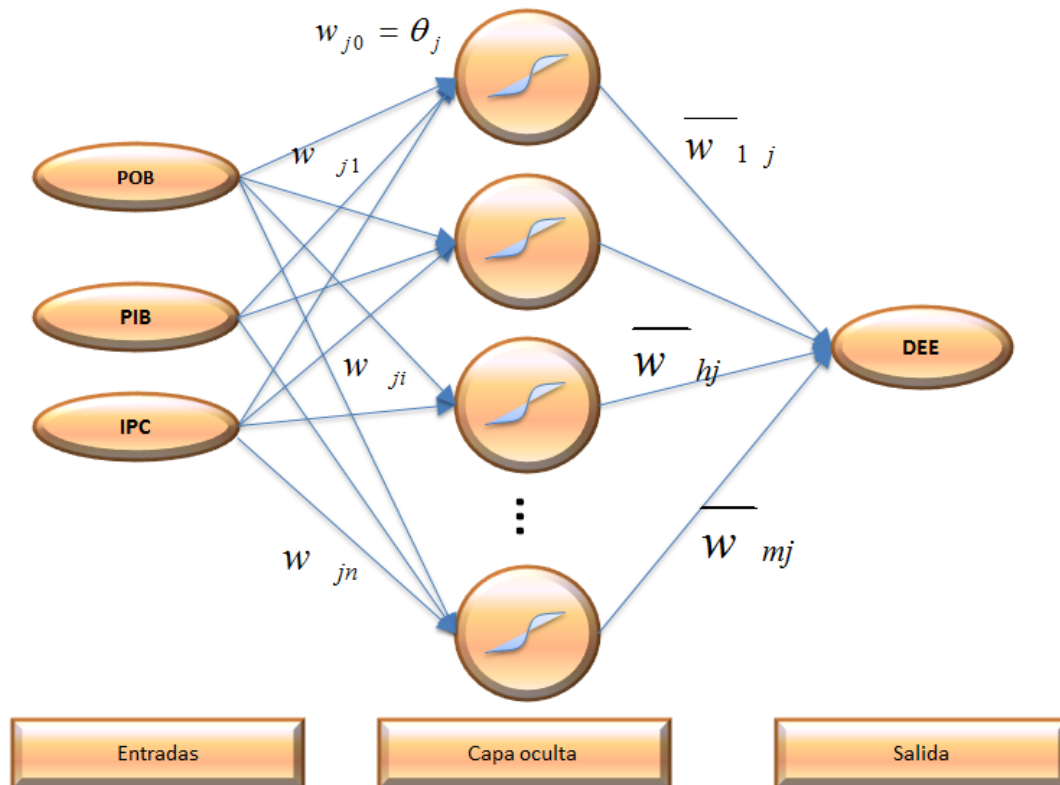


Fig. 3.13: El modelo neuronal para la estimación de la DEE

CAPÍTULO 4

EXPERIMENTACIÓN Y APLICACIONES

4. EXPERIMENTACIÓN Y APLICACIONES

Este capítulo contiene el diseño experimental, los resultados de los experimentos y las conclusiones de detalle de cada uno de ellos.

4.1 Experimentación en la Mejora del diseño y del uso de las Redes Neuronales Artificiales

En esta sección, se explican los experimentos y los resultados obtenidos para validar o rechazar las hipótesis sobre posibles mejoras mencionadas en 3.2.

4.1.1 Experimentación para la Inicialización de los pesos de una RNA

A continuación, se muestran los resultados de la aplicación de la matriz de Hilbert definida en 4.1.1 para generar los pesos iniciales de la RNA para los dos casos explicados en 3.2.2.

Los resultados de la evolución del error durante el entrenamiento, se observan en las figuras 4.1, 4.2, junto con las evoluciones del error mediante la inicialización de los pesos aleatoriamente, que se han obtenido antes, con la misma configuración de la red con objeto de realizar una comparación entre ambas técnicas.

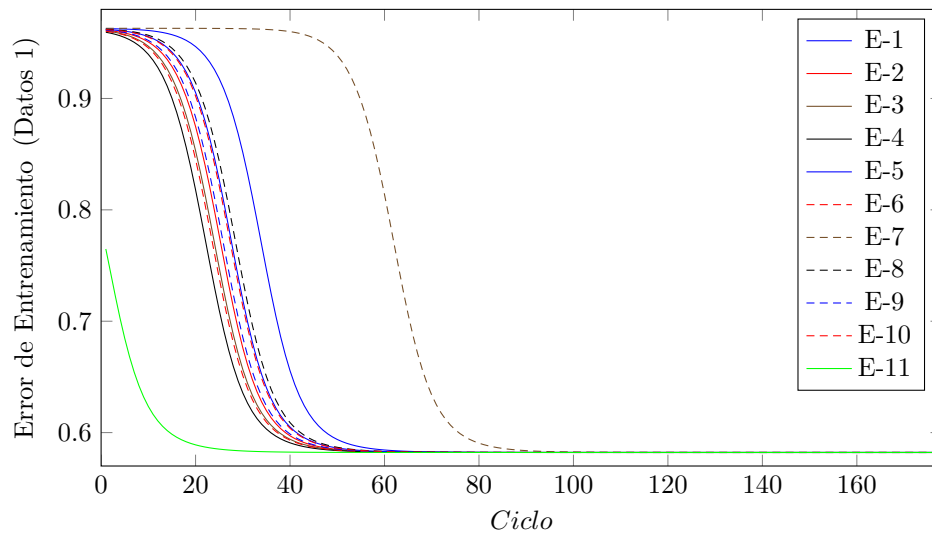


Fig. 4.1: Evolución del error con Matriz de Hilbert y diferentes ejecuciones (Datos 1)

Se puede observar claramente en las figuras 3.3 y 3.4 que la evolución del error no es la misma, si se ejecuta el entrenamiento en diferentes momentos, y con la misma arquitectura (mismo datos entrenamientos, ciclos de parada, mismo algoritmo de entrenamiento, ...). Sin

4.1 Experimentación en la Mejora del diseño y del uso de las Redes Neuronales Artificiales

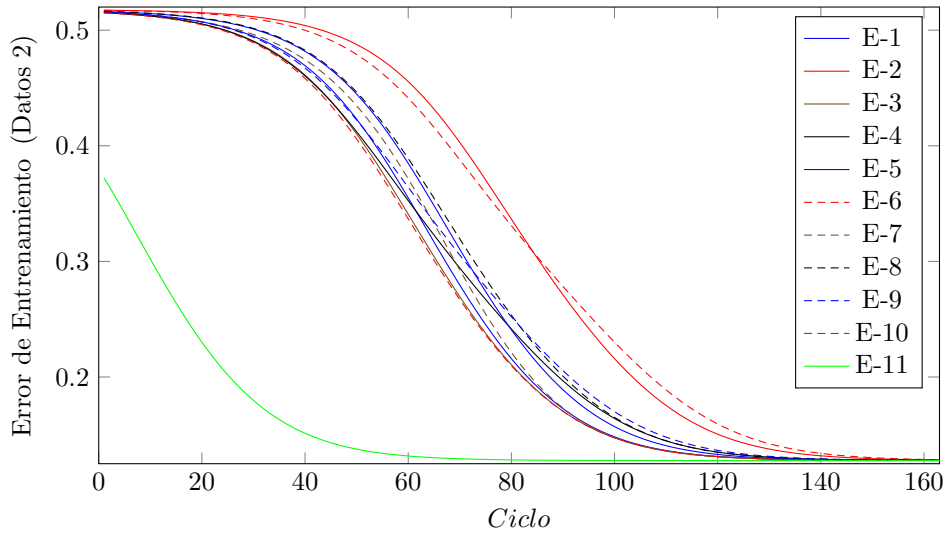


Fig. 4.2: Evolución del error con matriz de Hilbert y diferentes ejecuciones (Datos 2)

embargo al iniciar los pesos con la matriz de Hilbert, se puede observar en las figuras 4.1 y 4.2 que la minimización del error es mucho más rápida. Por otra parte, se obtiene la misma evolución del error cada vez que se ejecuta el entrenamiento a través de la matriz de Hilbert, y por lo tanto se tiene un criterio sólido para que cuando se desea comparar la mejora del entrenamiento al decidir añadir o eliminar más neuronas ocultas o capas ocultas, ... etc.

4.1.2 Experimentación en las funciones de activación

Se ha mencionado anteriormente en 4 que para resolver problemas no lineales, se recomienda el uso de la función sigmoideal (Sigm), definida en la ecuación 2.2, o la tangente hiperbólica (Tanh), definida en la ecuación 2.4. Y se ha mencionado también que no hay gran diferencia en general, cuando se usa una de una ellas en vez de la otra, excepto que si las entradas están normalizadas en un rango de $[0, 1]$, se usa la función sigmoideal, sin embargo si las entradas han sido normalizadas en un rango de $[-1, 1]$, se recomienda la función Tanh por sus ventajas de continuidad y derivabilidad.

En esta sección, se analiza el efecto del suavizado propuesto en esta tesis de la Sigm y la Tanh, aplicado a un problema no lineal, en este caso el problema PHM08 detallado en rephmproblem.

En 3.1 y 3.3, se han definido la función α -sigmoideal y la función α -Tanh. Una vez incorporada la función de activación α -sigmoideal a la red neuronal (respectivamente α -Tanh) con su derivada

4. EXPERIMENTACIÓN Y APLICACIONES

en el algoritmo RPROP, se han elegido valores naturales para α positivos, ya que se obtiene el mismo resultado eligiendo valores negativos debido a que tanto la función Tanh y Sigm son estrictamente crecientes, en el proceso de la minimización del error del entrenamiento.

En la figura 4.3 (respectivamente 4.5), se muestran los resultados de la evolución del error del entrenamiento con los mismos datos de entrenamiento1, y en la figura 4.4 (respectivamente 4.6), con los mismo datos de entrenamiento2, aplicados en el apartado anterior, la inicialización de los pesos se hizo mediante la técnica indicada en el apartado anterior (es decir usando la matriz de Hilbert, para garantizar la comparación y por otro lado una mejor inicialización de minimización), el criterio de parada ha sido $\varepsilon = 10^{-5}$.

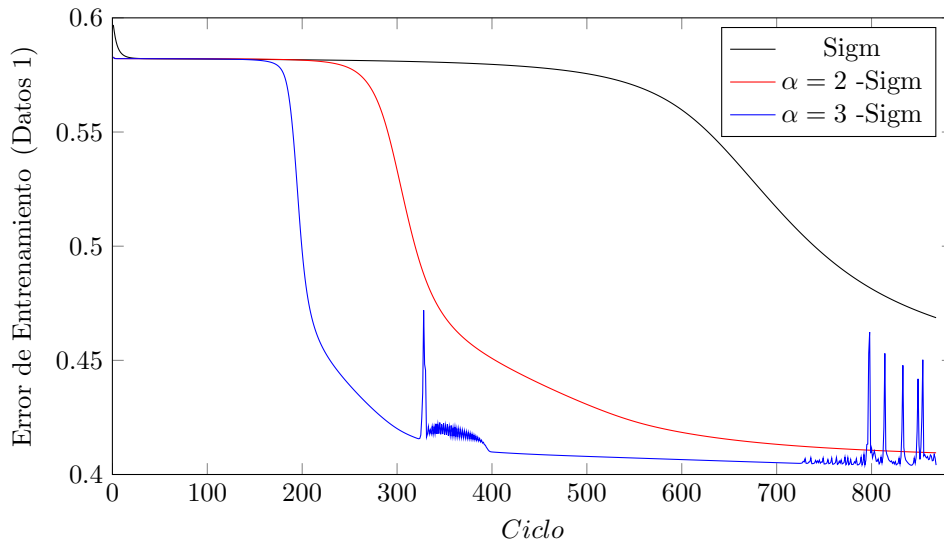


Fig. 4.3: Evolución del error con la función de activación sigmoidal y $\alpha = 2$ -Sigm y $\alpha = 3$ -Sigm (Datos 1)

Se observa claramente en las figuras 4.3 y 4.4, (respectivamente 4.5 y 4.6) que el uso de la $\alpha = 3$ -sigmo (Respectivamente $\alpha = 3$ -tanh) converge considerablemente antes, que mediante la $\alpha = 2$ -sigmo (Respectivamente $\alpha = 2$ -tanh), que a su vez converge mucho más rápido que la función sigmoidal (respectivamente con la tangente hiperbólica). Pero con $\alpha = 3$ -sigm (Respectivamente $\alpha = 3$ -tanh) derivan y convergen 3 veces el cambio de los pesos, dado que según la ecuación 3.4, teníamos: $\frac{dy}{dx} = 2\alpha y(1-y)$, siendo $y = \alpha - \text{Tanh}$ o: $y = \alpha - \text{sigm}$

Una ventaja que a veces podría ser una desventaja cuando ocurren oscilaciones, como se aprecia en la figura 4.5 y 4.6 por lo tanto se recomienda el uso de $\alpha = 2$ -sigm (Respectivamente $\alpha = 2$ -tanh).

4.1 Experimentación en la Mejora del diseño y del uso de las Redes Neuronales Artificiales

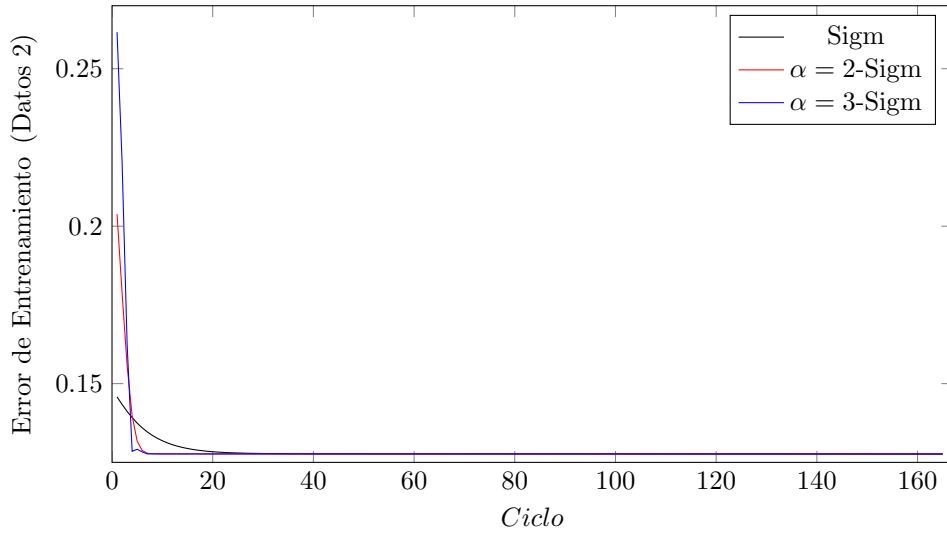


Fig. 4.4: Evolución del error con la función de activación sigmoidal y $\alpha = 2$ -Sigm y $\alpha = 3$ -Sigm (Datos 2)

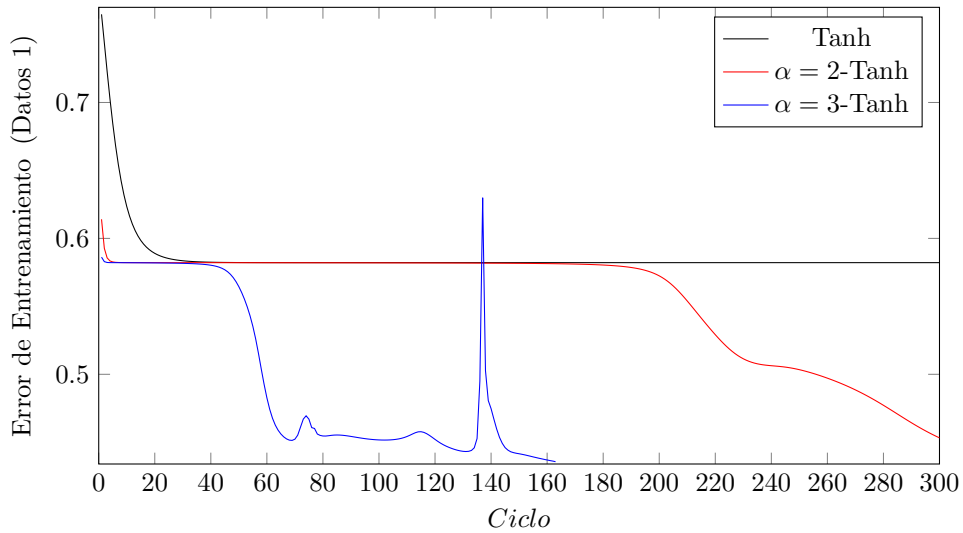


Fig. 4.5: Evolución del error con la función de activación Tanh y $\alpha = 2$ -Tanh y $\alpha = 3$ -Tanh (Datos 1)

4. EXPERIMENTACIÓN Y APLICACIONES

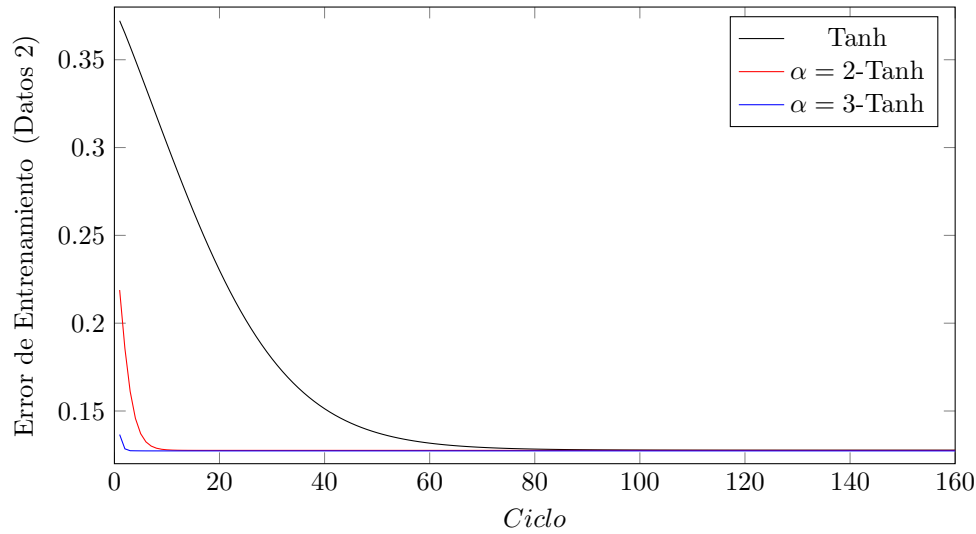


Fig. 4.6: Evolución del error con la función de activación Tanh y $\alpha = 2$ -Tanh y $\alpha = 3$ -Tanh (Datos 2)

4.1.3 Experimentación para la definición de la estructura de una RNA

En el capítulo anterior se ha mencionado que no existe ningún criterio para definir el número de neuronas / capas ocultas suficientes o necesarias para la convergencia. Por lo tanto se ha programado un programa nombrado “NNFINDER” (Neural Network Finder) definido y descrito en 3.2.1 que permite evaluar distintos rangos de neuronas ocultas, permite también variar el número de las capas ocultas.

Se ha incorporado a este programa el hecho de utilizar la matriz de Hilbert para generar los pesos inicialmente, de esta manera se puede elegir la mejor arquitectura dependiendo del criterio (en general el error global, y el tamaño de la red), sabiendo que la unificación en la inicialización de los pesos iniciales permite obtener el mismo resultado para la misma configuración.

A continuación en la tabla 4.1, se muestra el resultado tras ejecutar el programa para el caso 4.2.1:

El porcentaje del error figurado en la tabla, tanto para el conjunto de entrenamiento y el de validación de la red neuronal, nos permiten elegir con que red debemos elegir para su uso.

Otro criterio que nos permite el programa, es el número de iteración. En este ejemplo, tenemos 5000 iteración como número máximo de iteraciones.

El hecho de que figura el mismo número de iteraciones, permite concluir que todavía se puede minimizar aún el error. La ventaja de este programa radica en poder ejecutar varias

4.1 Experimentación en la Mejora del diseño y del uso de las Redes Neuronales Artificiales

Tabla 4.1: Resultado de NNFINDER ejemplo 1

Red	N° neur. ocultas	Inicialización	Error del entr. %	Error de valid. %	N° de iter.
RNA-1	1	Hilbert	5.86	7.11	5000
RNA-2	2	Hilbert	5.86	7.10	5000
RNA-3	3	Hilbert	6.25	7.27	5000
RNA-4	4	Hilbert	1.58	2.04	5000
RNA-5	5	Hilbert	1.93	2.05	5000
RNA-6	6	Hilbert	1.42	2.01	5000
RNA-7	7	Hilbert	1.25	1.89	5000
RNA-8	8	Hilbert	7.58	8.09	5000
RNA-9	9	Hilbert	3.14	3.89	5000
RNA-10	10	Hilbert	1.06	1.76	5000
RNA-11	11	Hilbert	1.70	1.44	5000
RNA-12	12	Hilbert	1.53	1.82	5000
RNA-13	13	Hilbert	1.28	1.68	5000
RNA-14	14	Hilbert	2.59	2.20	5000
RNA-15	15	Hilbert	4.81	4.19	5000
RNA-16	16	Hilbert	1.74	2.37	5000
RNA-17	17	Hilbert	2.38	2.57	5000
RNA-18	18	Hilbert	1.67	2.19	5000
RNA-19	19	Hilbert	1.73	2.18	5000
RNA-20	20	Hilbert	3.01	2.45	5000
RNA-21	21	Hilbert	2.36	2.61	5000
RNA-22	22	Hilbert	1.55	2.03	5000
RNA-23	23	Hilbert	1.25	1.66	5000
RNA-24	24	Hilbert	1.94	1.91	5000
RNA-25	25	Hilbert	1.41	1.71	5000
RNA-26	26	Hilbert	2.01	1.97	5000
RNA-27	27	Hilbert	2.30	2.23	5000
RNA-28	28	Hilbert	2.72	2.46	5000
RNA-29	29	Hilbert	2.28	2.09	5000
RNA-30	30	Hilbert	2.76	2.44	5000
RNA-31	31	Hilbert	2.34	1.92	5000
RNA-32	32	Hilbert	1.74	2.11	5000
RNA-33	33	Hilbert	1.71	2.17	5000
RNA-34	34	Hilbert	1.81	2.16	5000
RNA-35	35	Hilbert	1.24	1.62	5000
RNA-36	36	Hilbert	3.17	2.06	5000
RNA-37	37	Hilbert	3.75	1.41	5000
RNA-38	38	Hilbert	1.40	1.93	5000
RNA-39	39	Hilbert	1.38	1.74	5000
RNA-40	40	Hilbert	2.20	2.28	5000

4. EXPERIMENTACIÓN Y APLICACIONES

topologías y guardar en un archivo los resultados obtenidos durante el entrenamiento. Sin embargo, sufre de la desventaja de coste computacional, sobre todo cuando se entrena con un conjunto de datos de gran tamaño. Para evitar este inconveniente, se recomienda limitar el número de iteración en el criterio de parada, y así una vez obtenidos los resultados, se elige la red con mejor error obtenido y volvemos a entrenarla aumentando el número de iteraciones para minimizar aún más el error del entrenamiento.

4.1.4 Conclusiones

Disponer de herramientas como lo es el programa NNCI y NNFINDER, que permiten por un lado; encontrar la estructura de la RNA más adecuada para un problema específico, por otro lado; proporcionar medidas de confianza en las respuestas de la RNA. Ayuda a mejorar el empleo de las RNA y disponer de criterios de fiabilidad para la toma de decisión. Por otra parte, la inicialización de la RNA con la matriz de Hilbert, que ha mostrado buenos resultados en la minimización del error, y gracias a la justificación científica a base del teorema de Cohen-Grossberg, que garantiza la convergencia una característica importante. Debido a que el uso de este tipo de inicialización proporciona el mismo resultado al contrario de otras técnicas basadas en la inicialización aleatoria, permitirá no perder tiempo en caso de querer comparar otras arquitecturas, cambiando de número de neuronas ocultas por ejemplo, capas ocultas, funciones de activación, etc. El empleo del suavizado en las funciones de activación (Tanh y Sigm) permite acelerar también el proceso de aprendizaje.

Todo aquello, permite manejar las RNA en situaciones de alto riesgo o de emergencia que requieren soluciones en poco tiempo.

4.2 Experimentación en el uso de las RNAs para la Predicción de la Vida Remanente Útil en Equipos Industriales

En esta sección, se describen los experimentos desarrollados y las conclusiones extraídas sobre el uso de las redes neuronales para la predicción de la vida remanente útil.

4.2.1 Problema propuesto

El problema tratado, es reconocido como “PHM08 Prognostics Data Challenge Dataset”, representado por un conjunto de datos derivados de varios experimentos, que se utilizaron para el pronóstico en una competición en la Conferencia Internacional “Prognostics and Health Management (PHM08, 2008)”. La competición se planteó como reto hacia los diferentes desarrolladores con objeto de poder comparar los resultados estimulando la competitividad sobre la estimación de la Vida Remanente basada en datos y sin ningún conocimiento previo de la física del sistema.

El objetivo general es definir la curva que representa la vida útil remanente del estado de salud para cada unidad. Un ejemplo de tal curva se presenta en la figura 4.7.

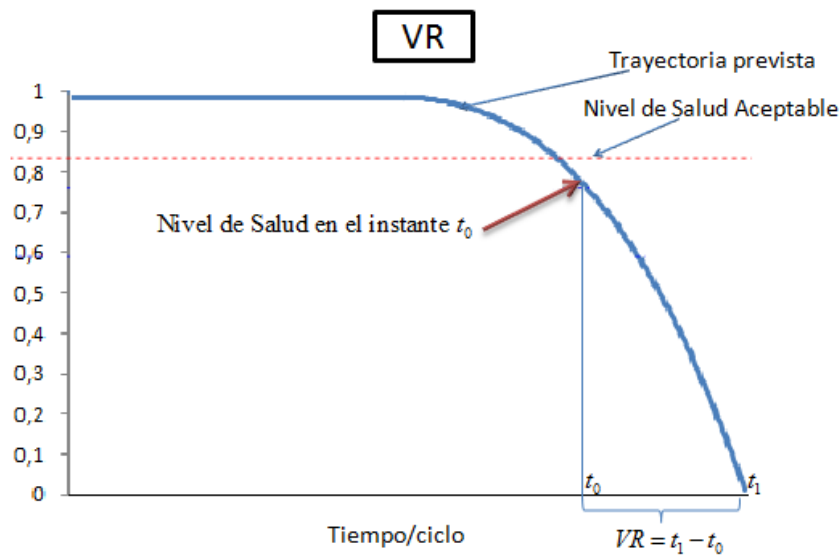


Fig. 4.7: Definición de la vida remanente útil

4. EXPERIMENTACIÓN Y APLICACIONES

4.2.1.1 Análisis de Datos y selección de Entradas para el Modelo

En: <http://ti.arc.nasa.gov/tech/dash/pcoe/prognostic-data-repository/>, se encuentran los conjuntos de datos disponibles, que constan de varias series temporales con múltiples variables. Cada conjunto de datos está dividido en dos sub-conjuntos uno para el entrenamiento y el otro para el test. Cada serie temporal proviene de un motor de avión tipo "turbofan" diferente, es decir, los datos pueden ser considerados como procedentes de una flota de motores del mismo tipo. Cada serie comienza a partir de una instancia diferente a las otras series (definido en los datos como unidad "unit"). Cada unidad comienza con diferentes grados de desgaste inicial y la variaciones posibles del proceso de fabricación son desconocidas para el usuario. La existencia de esas variaciones y del desgaste se consideran normales. Existen tres configuraciones operativas que tienen un efecto sustancial en el rendimiento de cada unidad (se denominan como ajuste operacional "operational setting"). Estos ajustes también se incluyen en los datos, considerándose también que los datos pueden estar contaminados por ruido procedente los sensores.

Se supone que cada motor está funcionando en modo normal al principio de cada serie temporal, y empieza a degradarse en algún momento durante la serie (este momento no está definido). En el conjunto de entrenamiento, la degradación crece en magnitud hasta que se alcanza un umbral predefinido más allá del cual es preferible no hacer funcionar el motor. El objetivo es predecir el número de ciclos operativos restantes.

El conjunto de entrenamiento incluye datos operativos de 218 motores diferentes. En cada conjunto de datos, el motor se hizo funcionar durante un número variable de ciclos hasta el fallo. Las longitudes de los ciclos para cada motor varían (Longitud mínima es de 127 ciclos y la longitud máxima de 356 ciclos).

En la figura 4.8 y 4.9, se muestran mediciones de dos sensores 1, 2, 7 y 8 de la unidad 1, en ciclos de tiempo.

En la figura 4.10 se muestra la dispersión del sensor 2 respecto al sensor 1 y del sensor 7 respecto al sensor 12, y se ve claramente los seis modos de operación que sufren los datos. Solo se han mostrado estas dos dispersiones (los mismos resultados se han obtenido para los demás sensores).

A partir del análisis de correlación entre variables, y sus dispersiones como se ha mostrado en las figuras: 4.8, 4.9 y 4.10, con el resto de sensores, se han descartado sensores cuyo valores no son consistentes y continuos. Y se han remplazado las tres primeras variables: "operational setting 1", "operational setting 2" y "operational setting 3", por solo una nueva variable que

4.2 Experimentación en el uso de las RNAs para la Predicción de la Vida Remanente Útil en Equipos Industriales

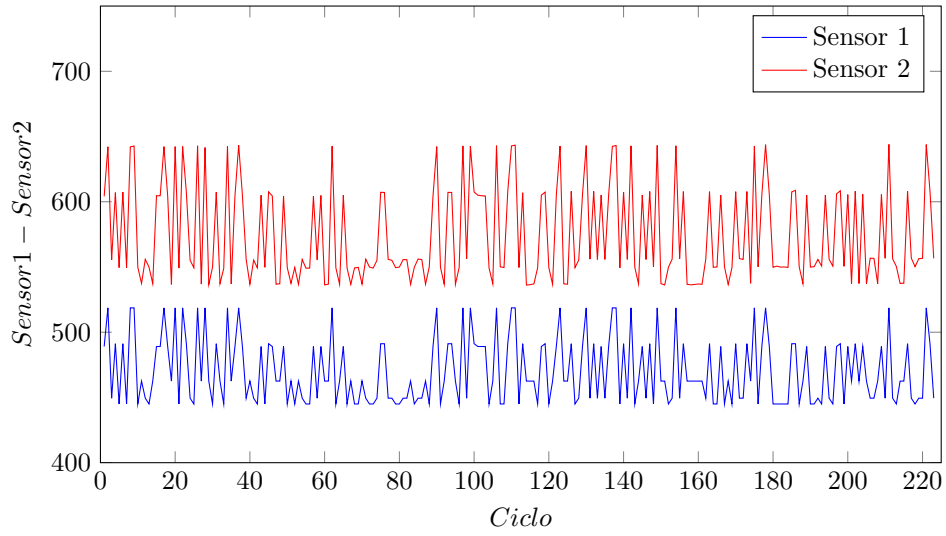


Fig. 4.8: Sensor 1 vs sensor 2

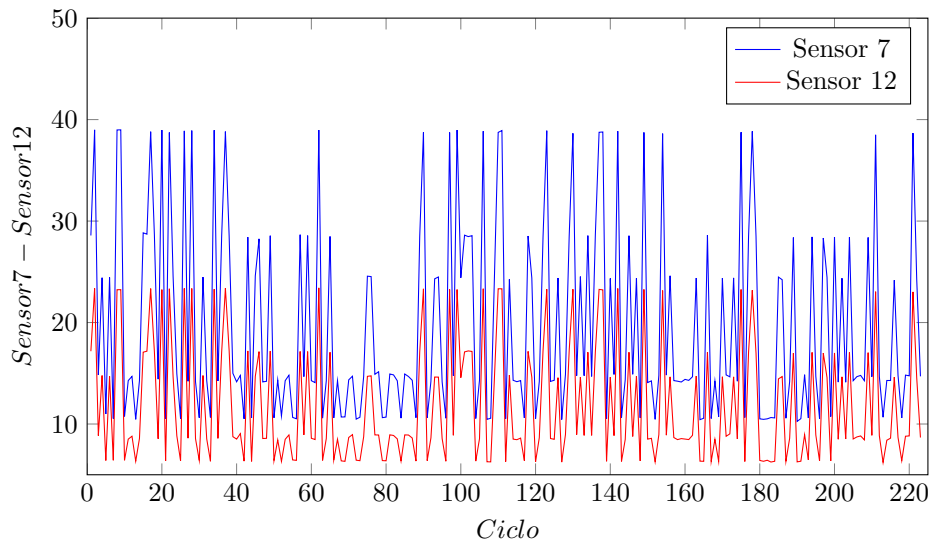


Fig. 4.9: Sensor 7 vs sensor 12

4. EXPERIMENTACIÓN Y APLICACIONES

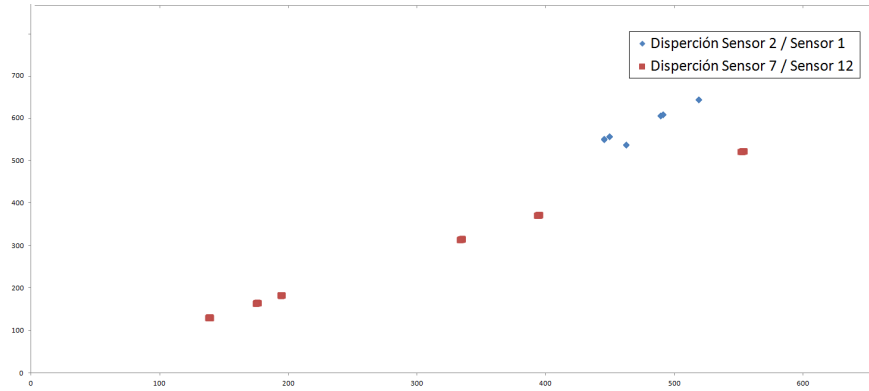


Fig. 4.10: Dispersión (sensor1 - sensor2) y (Sensor 7- sensor 12)

tiene como valores $\{1, 2, 3, 4, 5, 6\}$, representando así los seis modos de operación encontrados (nombrada: Modo.Oper).

Por otra parte, aplicando los pasos sugeridos en el apartado 2.1.11.3, se han seleccionado los siguientes sensores: $\{1, 2, 4, 7, 8, 11, 12, 15, 18, 21\}$, y la variable creada en la etapa anterior, denominada como: Modo.oper, como entradas para la red neuronal. Por lo tanto hay 11 variables de entrada y una variable de salida: Índice de Salud “IS”, que tendrá valores entre 0 y 1, donde: $I = 1$ el motor está funcionando en pleno rendimiento en modo normal, y cuando $I = 0$, el motor ha fallado y no está funcionando. Y Así, la función IS representa en cada ciclo dado el umbral de salud, en el cual se encuentra la maquina en este momento.

4.2.1.2 Selección del modelo neuronal

En esta fase, se han seguido los pasos y técnicas recomendadas en 2.1.11, tanto para la normalización de los datos como para los ciclos de entrenamiento y la validación cruzada. Se han creado para los datos seleccionados en el apartado anterior, varias redes neuronales. Cada red neuronal destinada para una unidad (motor) de las 218 presentes en los datos.

Cada red neuronal está compuesta de tres capas, la primera reservada para los datos de entrada (11 variables), en la segunda capa se alojan las neuronas ocultas, que se han calculado utilizando el NNFINDER, que sirven de conexión entre la primera capa y la última donde hay solo una neurona que representa la salida deseada, en este caso el índice de salud del motor/unidad, que representa el ciclo de vida remanente (ver 4.11).

Respecto a las funciones de activación de la segunda y la última capa, se ha utilizado la 2-Tanh definida en la ecuación 3.4 debido a la no linealidad del problema, y como algoritmo de

4.2 Experimentación en el uso de las RNAs para la Predicción de la Vida Remanente Útil en Equipos Industriales

entrenamiento el algoritmo definido en 1.

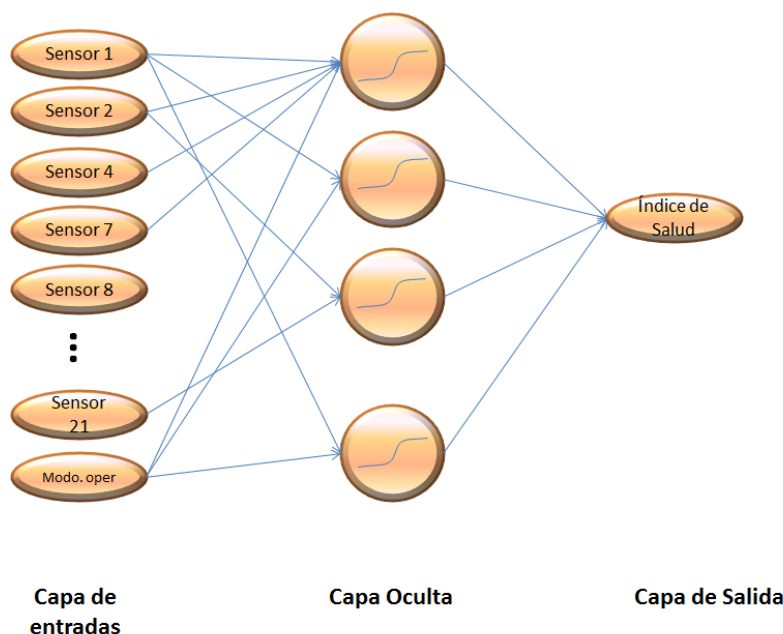


Fig. 4.11: Arquitectura de la red neuronal para la predicción de IS

No se dispone de ninguna información acerca del ciclo de la vida, solo se sabe que los motores al principio del ciclo de vida tienen buena salud y que sufren degradación hasta el último ciclo. Se aportan dos ideas/enfoques para estudiar y predecir cuantos ciclos de vida le quedan a cada motor mediante una curva del "índice de salud" de valor 1 al comienzo de la vida útil y valor 0 al final de la misma.

En las siguientes subsecciones, se presenta solo el análisis de tres unidades, la unidad 1 (223 ciclos), la unidad 6 (225 ciclos), y la unidad 147 (323 ciclos), con dos enfoques distintos. Se han seleccionado estas tres unidades, debido a que la unidad 1 y 6 tienen casi la misma longitud de ciclos, y así se podría verificar si el periodo de degradación es parecido o no. Y la unidad 147 porque tiene más ciclos hasta el fallo que la unidad 1 y 6.

4.2.1.3 Primer enfoque para la predicción del Índice de Salud

Este primer enfoque supone que la degradación de la unidad es lineal de pendiente -1 , en el ciclo 1: $IS = 1$, y en el último ciclo: $IS = 0$. Para cada unidad 1, 6 y 147 se ha creado la salida IS (salida deseada) con una función lineal $y = ax + b$. Los coeficientes a y b se calculan de la

4. EXPERIMENTACIÓN Y APLICACIONES

manera siguiente:

$$a = \frac{-1}{\max(\text{ciclo})}$$
$$b = \frac{1 + \max(\text{ciclo})}{\max(\text{ciclo})}$$

a y b varían de una unidad a otra.

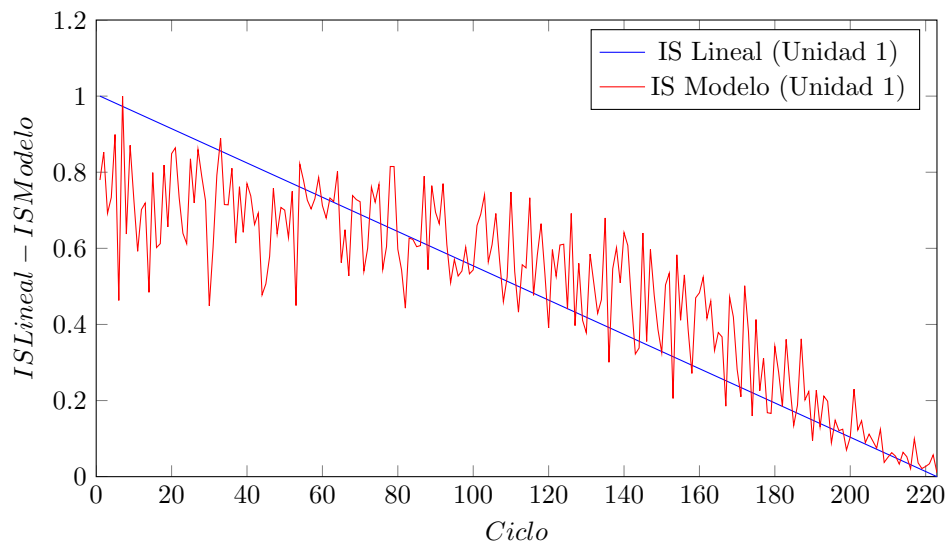


Fig. 4.12: IS modelo vs regresión lineal (Unidad 1)

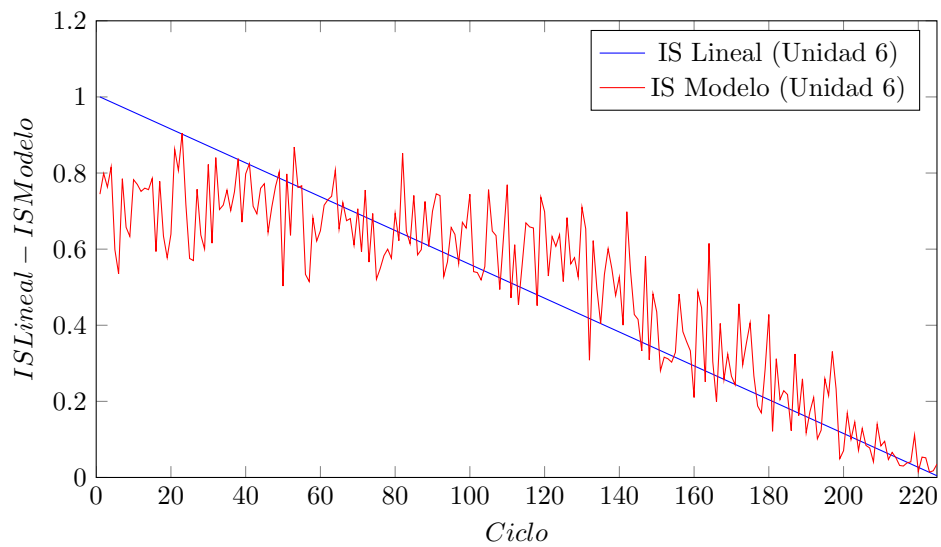


Fig. 4.13: IS modelo vs regresión lineal (Unidad 6)

4.2 Experimentación en el uso de las RNAs para la Predicción de la Vida Remanente Útil en Equipos Industriales

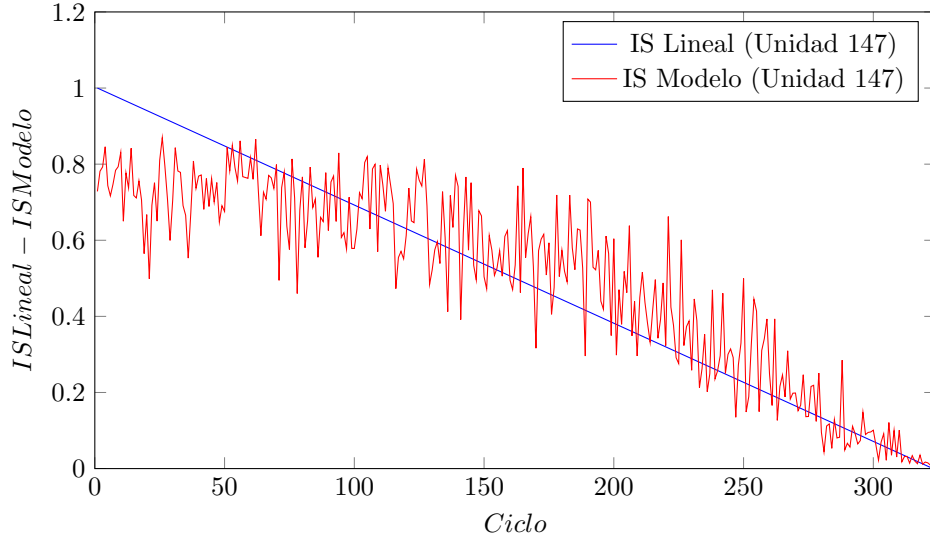


Fig. 4.14: IS modelo vs regresión lineal (Unidad 147)

Se puede apreciar en la predicción del Modelo Neuronal en las tres figuras: 4.12, 4.13 y 4.14, aparecen dos periodos. El primero sufre de una degradación pequeña casi constante, el segundo de una degradación más fuerte (fijándose en la pendiente de las curvas en diferentes ventanas durante los ciclos).

Observación

En la unidad 1, parece que en torno a los ciclos 100 y 130, empieza el segundo periodo de degradación más notable.

En la unidad 6, parece que en torno a los ciclos 120 y 140, empieza el segundo periodo de degradación más notable.

En la unidad 147, parece que en torno a los ciclos 160 y 190, empieza el segundo periodo de degradación más notable.

La etapa siguiente se ha diseñado para mejorar la predicción del modelo neuronal, empleando:

- La salida de la red neuronal obtenida en la etapa anterior, como salida deseada de nuevo, en vez de una salida lineal descendiente.
- Entrenamiento del modelo aplicando como entradas las 11 disponibles en C_t , añadiendo 11 del estado en el ciclo C_{t-2} y C_{t-1} , para predecir el IS en el C_t .

Con ello se mejora el ajuste de la predicción del IS, y la red neuronal se adapta a la naturaleza de la degradación de cada unidad. Los resultados obtenidos para la unidad 1,

4. EXPERIMENTACIÓN Y APLICACIONES

6 y 147, se puede ver en la curva IS Modelo (Unidad 1, 6 y 147) en la figura 4.15, 4.16 y 4.17 sucesivamente.

- Una vez terminada la etapa anterior, se usa un ajuste a la curva del IS predicha, usando una regresión lineal de orden 3 (ver la curva IS Modelo Fit (Unidad 147) en la figura 4.17).

Se observa en la gráfica 4.17, un índice de salud casi constante (una curva lineal con una pendiente de $-0,0005452$) hasta el ciclo 160, y después empieza una degradación más notable.

A continuación, se empleara otra técnica para ajustar mejor la curva de IS, suavizando las oscilaciones. Por ello según Leto en (177) ha considerado que una característica importante del conjunto de datos del Desafío PHM es que contiene series temporales y por lo tanto la representación del tiempo, o al menos la forma de representar y tener en cuenta las observaciones temporales anteriores es una consideración importante.

El método más rápido y más simple para la representación del tiempo es considerar cada punto de tiempo de forma independiente y crear una predicción en cada paso de tiempo. Una representación alternativa sería considerar el uso del espacio temporal, en el que se genera una secuencia de instancias utilizando una ventana deslizante de longitud fija. Sin embargo, los problemas asociados con la maldición de la dimensionalidad (23, 179), perjudican el uso de estas técnicas. Por ejemplo en los trabajos de (180) y (181), han usado redes neuronales recurrentes para este propósito.

En nuestro caso, al estado " t ", que ya estaba definido como entrada a la red neuronal, se añaden como entradas adicionales los estados " $t - 2$ " y " $t - 1$ ", para contribuir a la predicción IS en el instante t . Por lo tanto se tendrán 33 entradas para cada red neuronal. A continuación en las figuras 4.18, 4.19 y 4.20, se muestran los resultados obtenidos con dicha modificación.

4.2.1.4 Segundo enfoque para la predicción del Índice de Salud

El segundo enfoque, se basa en asignar para cada primer ciclo en los seis modos de operación: $IS = 1$, y para cada último ciclo de los seis modos: $IS = 0$. Entonces, los datos de entrenamiento consisten en solo las primeras y últimas medidas de los sensores para cada modo de operación de cada unidad y se pretende predecir todos los ciclos de la misma unidad.

A continuación, se obtienen los resultados mostrados en la figura 4.21 para la unidad 1, la figura 4.22 para la unidad 6, la figura 4.23 para la unidad 147. En estas gráficas, se ha realizado también un ajuste lineal de orden 3 de las predicciones obtenidas por la red neuronal.

4.2 Experimentación en el uso de las RNAs para la Predicción de la Vida Remanente Útil en Equipos Industriales

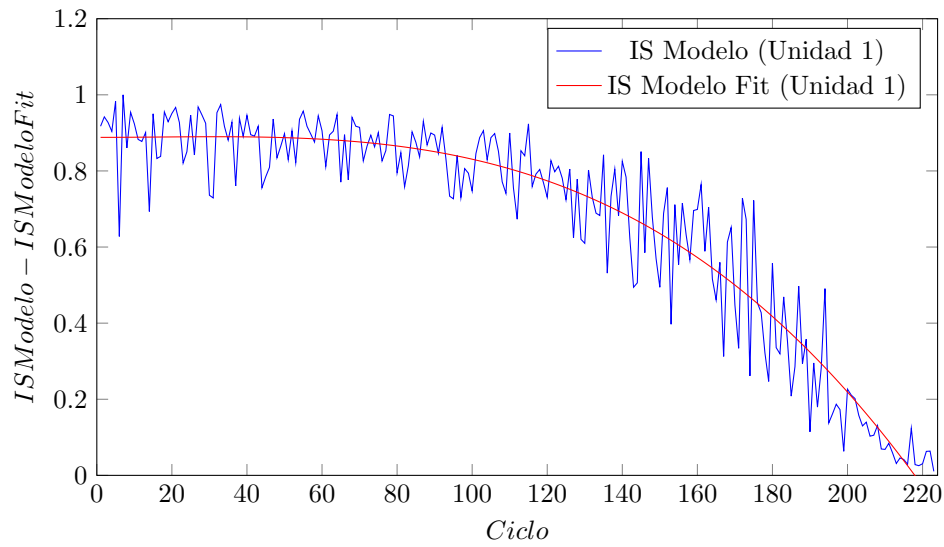


Fig. 4.15: IS Modelo vs Modelo Fit regresión (Unidad 1)

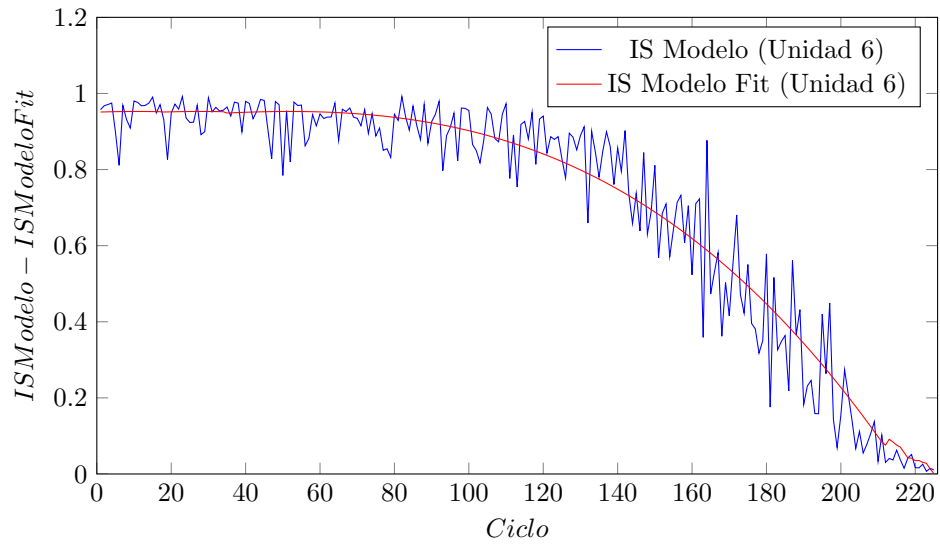


Fig. 4.16: IS Modelo vs Modelo Fit regresión (Unidad 6)

4. EXPERIMENTACIÓN Y APLICACIONES

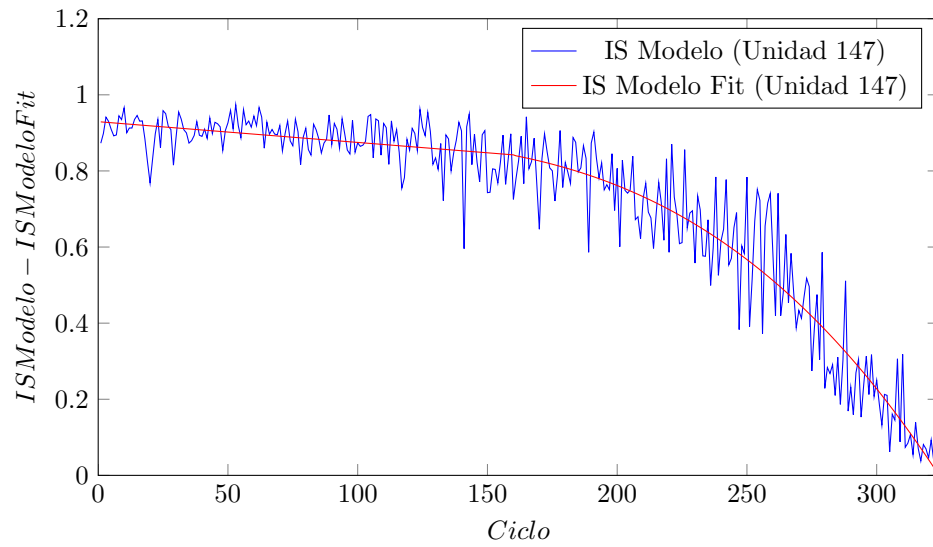


Fig. 4.17: IS Modelo vs Modelo Fit regresión (Unidad 147)

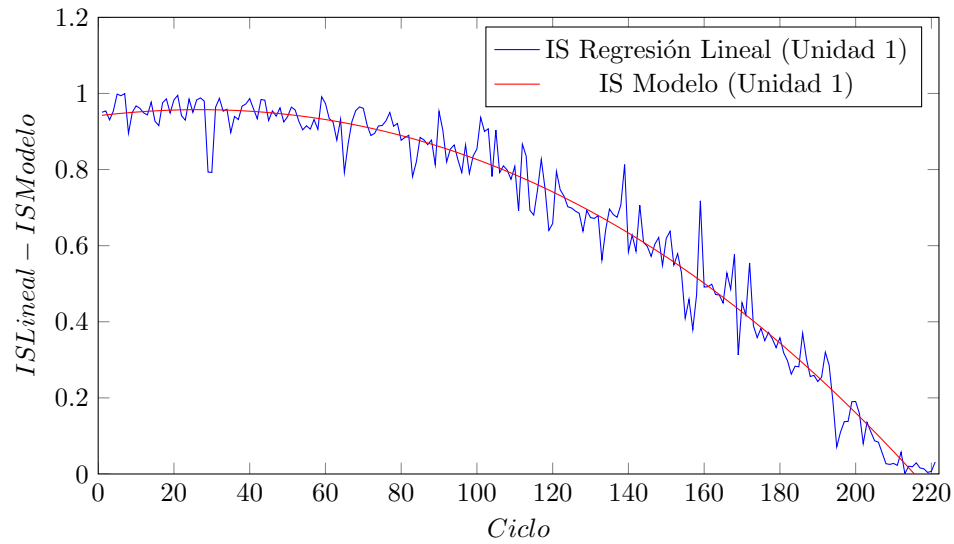


Fig. 4.18: IS modelo vs regresión lineal (considerando $t - 2$, $t - 1$ y t) (Unidad 1)

4.2 Experimentación en el uso de las RNAs para la Predicción de la Vida Remanente Útil en Equipos Industriales

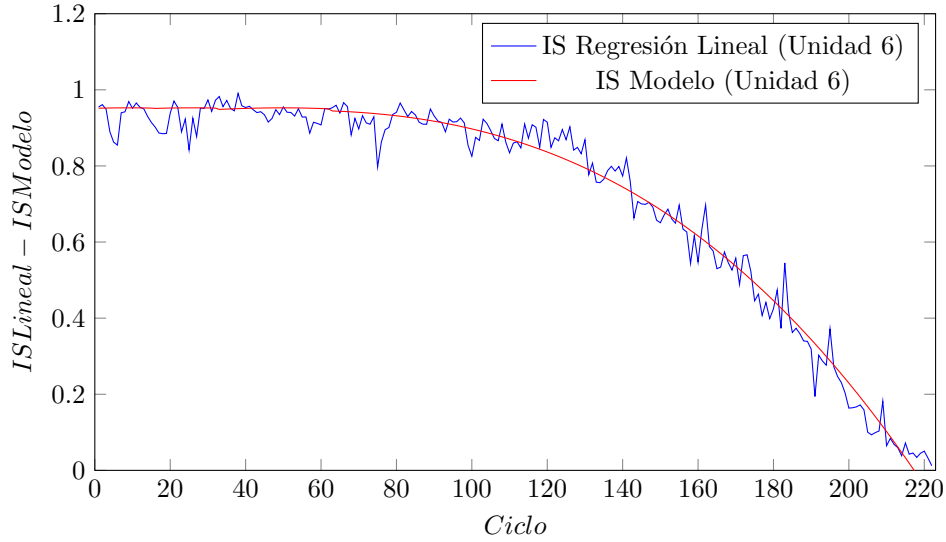


Fig. 4.19: IS modelo vs regresión lineal (considerando $t - 2$, $t - 1$ y t) (Unidad 6)

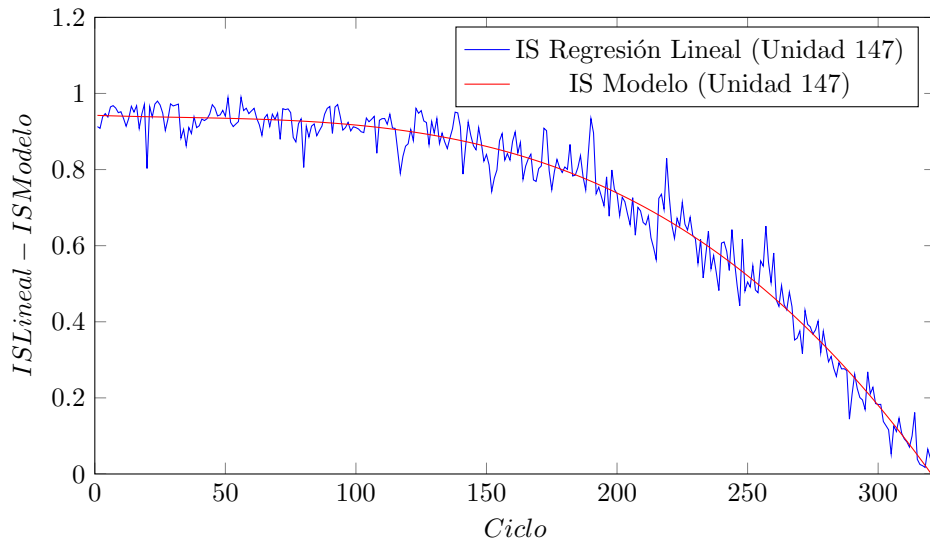


Fig. 4.20: IS modelo vs regresión lineal (considerando $t - 2$, $t - 1$ y t) (Unidad 147)

4. EXPERIMENTACIÓN Y APLICACIONES

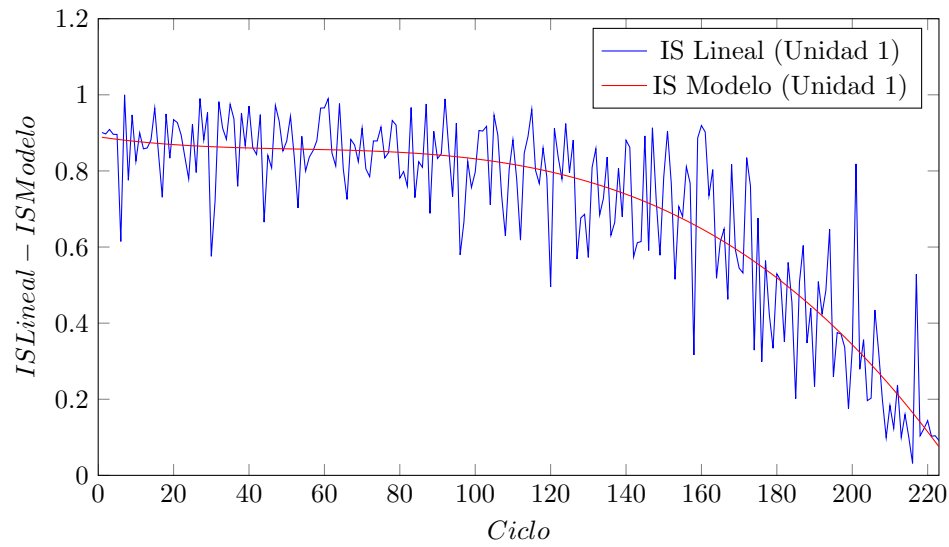


Fig. 4.21: IS modelo vs regresión lineal (Unidad 1)

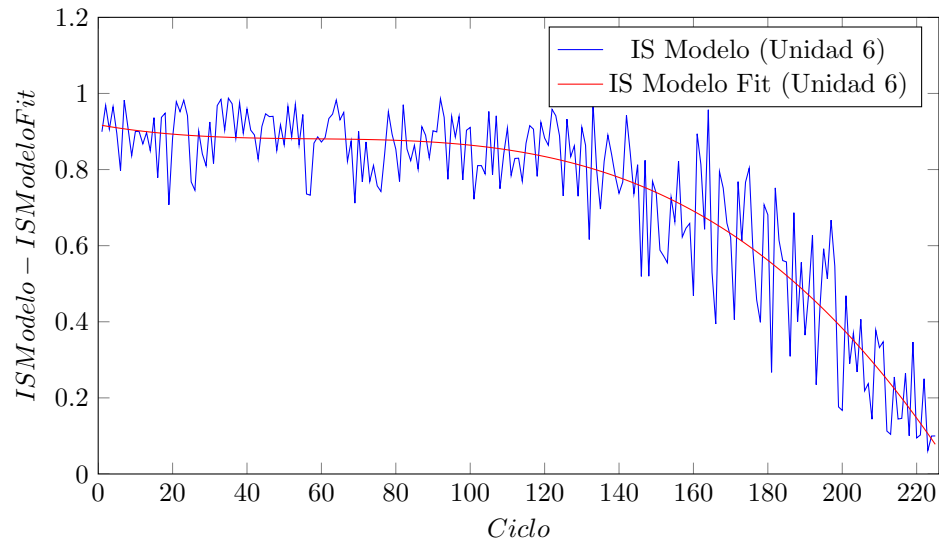


Fig. 4.22: IS Modelo vs Modelo regresión lineal ajustada (Fit) (Unidad 6)

4.2 Experimentación en el uso de las RNAs para la Predicción de la Vida Remanente Útil en Equipos Industriales

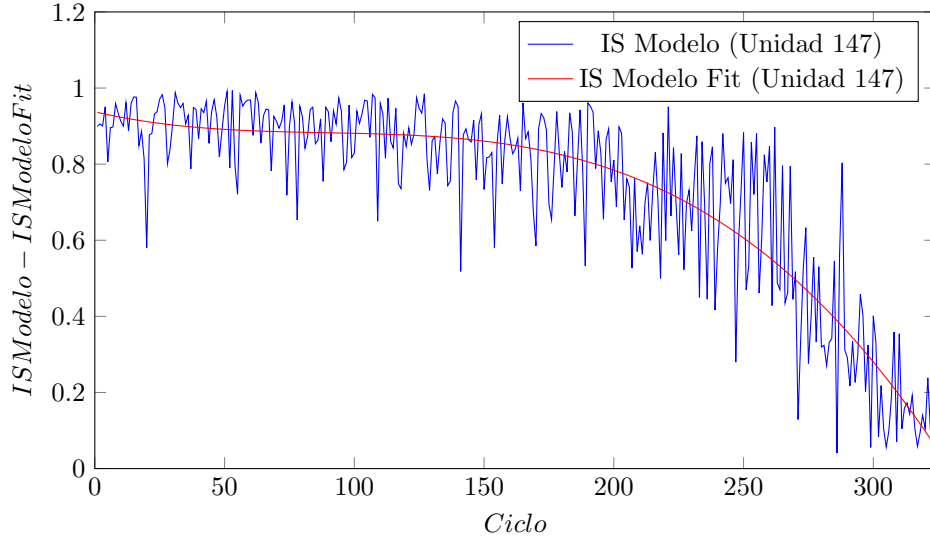


Fig. 4.23: IS Modelo vs Modelo regresión lineal ajustada (Fit) (Unidad 147)

Llegando a este punto, no se puede aplicar la misma modificación que se ha realizado con el primer enfoque (es decir añadir la información de la mediciones de los sensores en los estados $t - 2$ y $t - 1$), ya que solo se han considerado las mediciones del primer ciclo y el último ciclo de cada unidad. Por lo tanto o se aceptan los resultados obtenidos, o se puede aplicar los filtros de Kalman para eliminar el ruido, como se ha hecho en (177).

4.2.2 Comparación de los dos enfoques

En esta sección, se comparan los resultados obtenidos por los dos enfoques explicados en la sub-sección anterior. Los experimentos se han realizado en el mismo ordenador, el tiempo de entrenamiento empleado para el entrenamiento de los datos en el primer enfoque para cada unidad, en la primera etapa ha sido de 40 segundos, más otros 30 segundos cuando se han añadido los datos del retardo ($t - 2$ y $t - 1$) 2. Sin embargo en el segundo enfoque solo se ha necesitado menos de 4 segundos (esto es normal ya que el grupo de entrenamiento solo tiene 12 entradas frente a más de 100 en el primer enfoque).

En las gráficas 4.24, 4.25 y 4.26, se compara los ajustes de la curvas obtenidas por la red neuronal de los dos enfoques.

Se deduce claramente que se parecen mucho, y la diferencia entre ellos no supera el 2%, esta similitud entre los resultados de los dos enfoques (que tienen datos de entrenamientos

4. EXPERIMENTACIÓN Y APLICACIONES

distintos), nos puede permitir llegar a la conclusión de que los dos son correctos hasta cierto punto. Teniendo en cuenta el criterio del tiempo, el segundo enfoque supera con mucha ventaja el primer enfoque, sin embargo el primer enfoque supera el segundo en el ajuste. Como los dos enfoques han sido de sencilla implementación y no requieren mucho tiempo de cálculo ni memoria, se pueden emplear conjuntamente para que el uno valide el resultado obtenido mediante el otro enfoque.

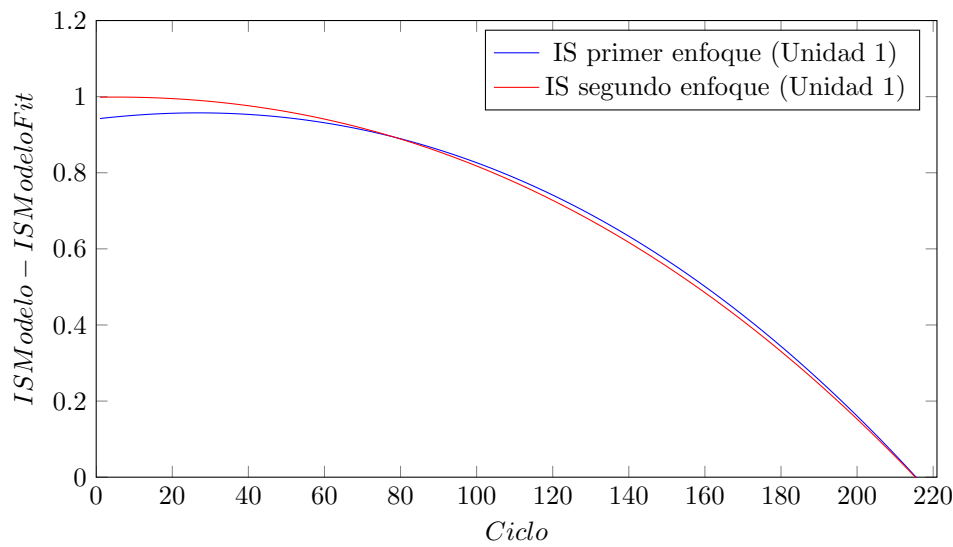


Fig. 4.24: IS: primer enfoque vs segundo enfoque (Unidad 1)

4.2.3 Evaluación de la Fiabilidad - Rendimiento de los Modelos

El rendimiento del método propuesto para la predicción del IS, se evalúa a través de la aplicación de métricas propuestas en (182, 183, 184).

Se calculan los indicadores más utilizados, es decir, la exactitud “Accuracy”, la precisión “Precision”, el error cuadrático medio “Mean Squared Error (MSE)”, y la media del porcentaje de error absoluto “Mean Absolute Percentage Error (MAPE)”. Estas métricas son útiles para la evaluación del modelo frente a los valores reales, pero no pueden capturar diferentes aspectos temporales en las predicciones del IS. Por lo tanto otras referencias anteriores añaden métricas que evalúan la fiabilidad de los modelos. Pero en este caso de estudio, no se dispone de información relativa a la curva real del IS, por lo tanto se analizan las predicciones de los modelos propuestos en esta tesis contra las curvas modeladas vía la regresión lineal de orden 3.

4.2 Experimentación en el uso de las RNAs para la Predicción de la Vida Remanente Útil en Equipos Industriales

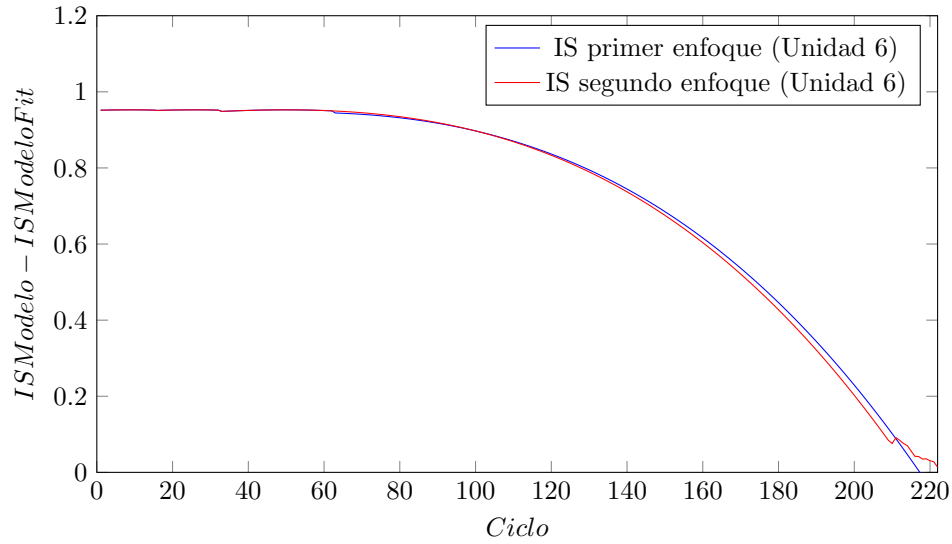


Fig. 4.25: IS: primer enfoque vs segundo enfoque (Unidad 6)

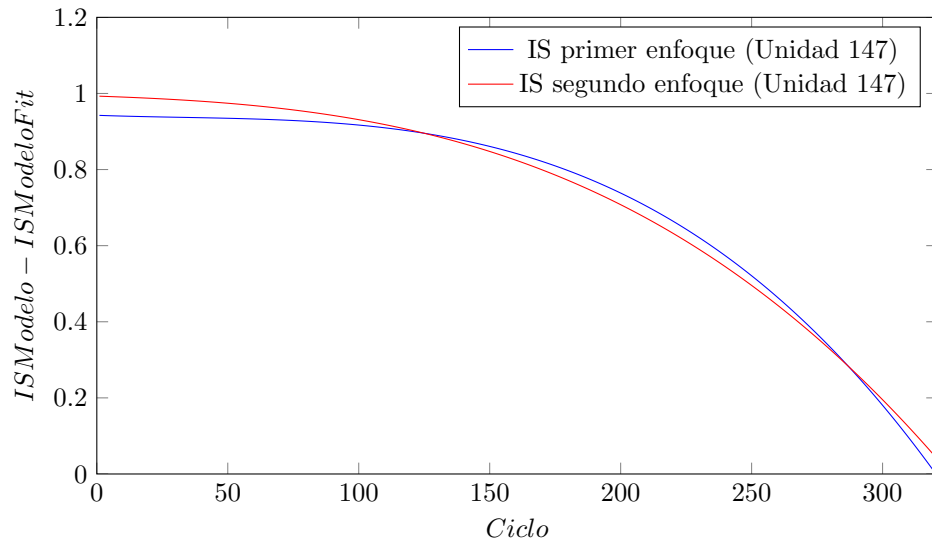


Fig. 4.26: IS: primer enfoque vs segundo enfoque (Unidad 147)

4. EXPERIMENTACIÓN Y APLICACIONES

En la tabla 4.2 se muestran algunas métricas mencionadas anteriormente en 3.3.3 comparando el rendimiento del modelo propuesto obtenido mediante la red neuronal en los enfoques 1 y 2, para las unidades 1, 6, y 147:

Tabla 4.2: Evaluación Enfoque 1

Enfoque 1					
Unidad	EXAC ($\alpha = 1$)	EXAC ($\alpha = 2$)	PREC	MSE	MAPER
1	0,905844039	0,971228627	0,05396473	0,002989507	13,2531039
6	0,944797	0,9999875	0,03814809	4,1833E-07	6,6172791
147	0,92814335	0,994940243	0,0526161	0,002784573	7,99333295

Tabla 4.3: Evaluación Enfoque 2

Enfoque 2					
Unidad	EXAC ($\alpha = 1$)	EXAC ($\alpha = 2$)	PREC	MSE	MAPER
1	0,751918999	0,903087621	0,16381335	0,028623061	59,1327657
6	0,819823304	0,955466941	0,12981545	1,34876E-08	33,1942476
147	0,824414442	0,953504853	0,14474878	0,021566449	25,8771422

En las tablas 4.2 y 4.3, se puede observar que los resultados obtenidos mediante el primer enfoque son mejores que los resultados obtenidos mediante el segundo enfoque. También es importante señalar que el error cuadrático medio del segundo enfoque para la unidad 6 es mejor que el del primer enfoque, sin embargo el resto de las métricas confirman lo contrario, de todo ello se deduce la importancia de disponer de varias métricas para concluir. En lo siguiente, se añaden indicadores que proporcionan más información sobre la fiabilidad del modelo.

Teniendo en cuenta la importancia del intervalo de confianza “IC” de la red neuronal a la hora de la validación, como se ha explicado en 2.1.12.2, se ha calculado el IC utilizando el NNCL. En las figuras 4.27, 4.31, 4.35, se muestra el IS ajustado para el enfoque 1, y las predicciones de la red neuronal (IS_Modelo), para cada unidad 1, 6, 147. Las curvas “IS_Modelo + IC” y “IS_Modelo - IC”, representan el ancho del intervalo de confianza de la red neuronal con un nivel de confianza del 95%. Se nota que es más estrecho en los dos extremos de la curva.

En las figuras 4.28, 4.32, 4.36, se muestra el porcentaje de confianza de la predicción de la red neuronal para cada unidad, en cada ciclo, como primer indicador definido en el tercer capítulo (ver 4). Se observa que el nivel de confianza es alto y casi constante en los extremos del ciclo de vida, excepto en la parte donde se produce una degradación más notable en la curva del

4.2 Experimentación en el uso de las RNAs para la Predicción de la Vida Remanente Útil en Equipos Industriales

IS. Eso es debido a la falta de información (datos de entrenamiento) en este intervalo de tiempo, lo que confirma los conocimientos adquiridos del problema sobre cómo se está produciendo el fallo.

Las figuras 4.29, 4.33, 4.37, muestran el resultado con el segundo indicador utilizado, que refleja el porcentaje relativo al error del modelo, obtenido mediante la ecuación 3.3.3. Este segundo indicador, refleja un punto importante, que cada vez que el IS predicho mediante el modelo se aleja de la curva, el porcentaje es muy bajo, y viceversa.

Teniendo en cuenta, estos dos indicadores, se ha calculado un tercer indicador 3.3.3, (El promedio de los dos indicadores anteriores), para tener en cuenta tanto la confianza de la red neuronal a la hora de su predicción como el error relativo al valor real. El resultado se muestra en las figuras 4.30, 4.34, 4.38.

Estos indicadores que dan indicaciones a la fiabilidad del modelo propuesto, no se pueden observar a primera vista, si solo se tienen en cuenta las métricas calculadas como el MSE o la exactitud y la precisión que superan el 98%, por eso dar un porcentaje de fiabilidad del modelo en cada ciclo es un indicador importante para la toma de decisión del estado de salud de la unidad.

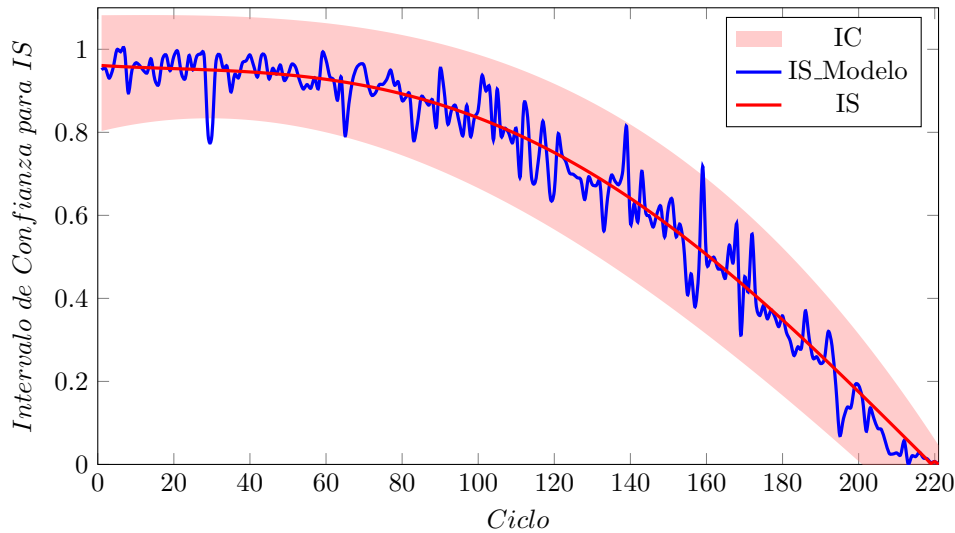


Fig. 4.27: IS vs. IS_Modelo y el intervalo de confianza del modelo (Unidad 1)

4. EXPERIMENTACIÓN Y APLICACIONES

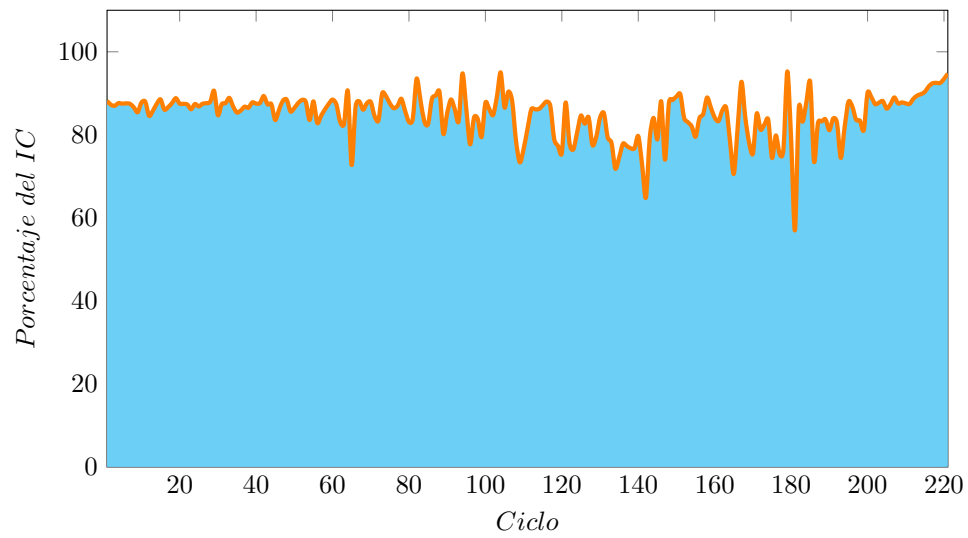


Fig. 4.28: Indicador 1 (Unidad 1)

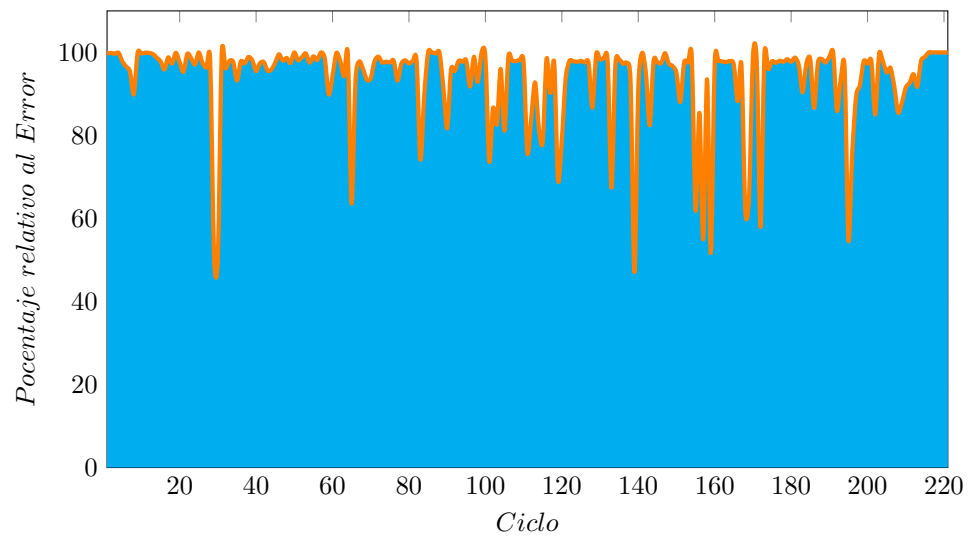


Fig. 4.29: Indicador2 (Unidad 1)

4.2 Experimentación en el uso de las RNAs para la Predicción de la Vida Remanente Útil en Equipos Industriales

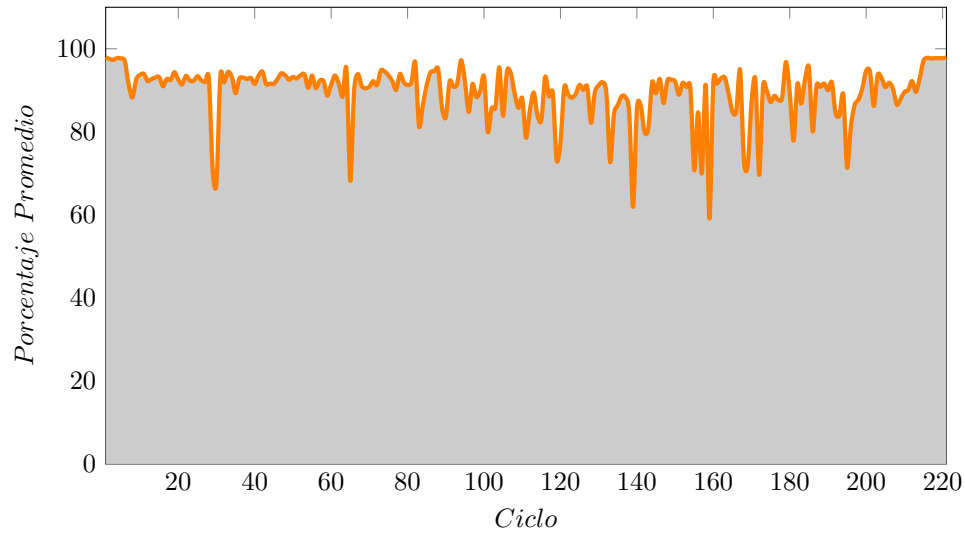


Fig. 4.30: Indicador 3 (Unidad 1)

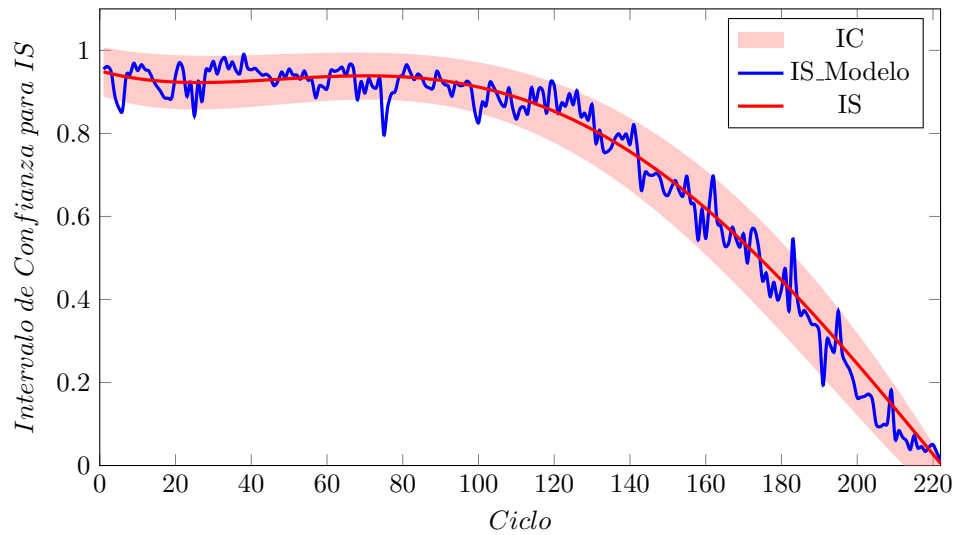


Fig. 4.31: IS vs. IS_Modelo y el intervalo de confianza del modelo (Unidad 6)

4. EXPERIMENTACIÓN Y APLICACIONES

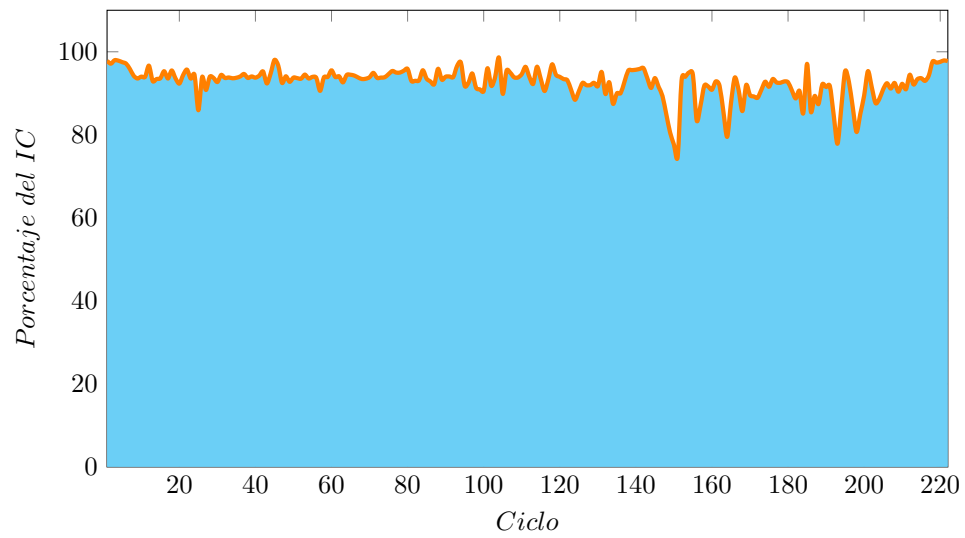


Fig. 4.32: Indicador 1 (Unidad 6)

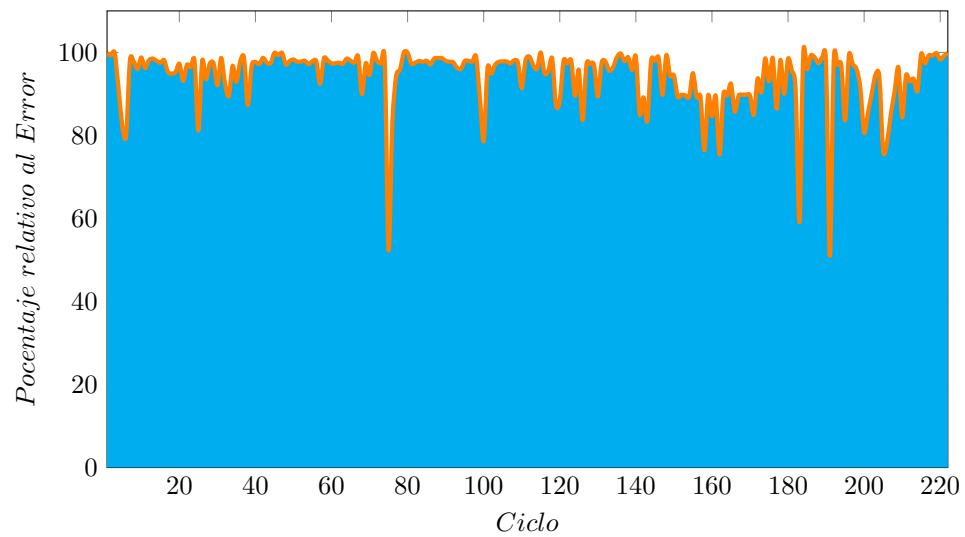


Fig. 4.33: Indicador 2 (Unidad 6)

4.2 Experimentación en el uso de las RNAs para la Predicción de la Vida Remanente Útil en Equipos Industriales

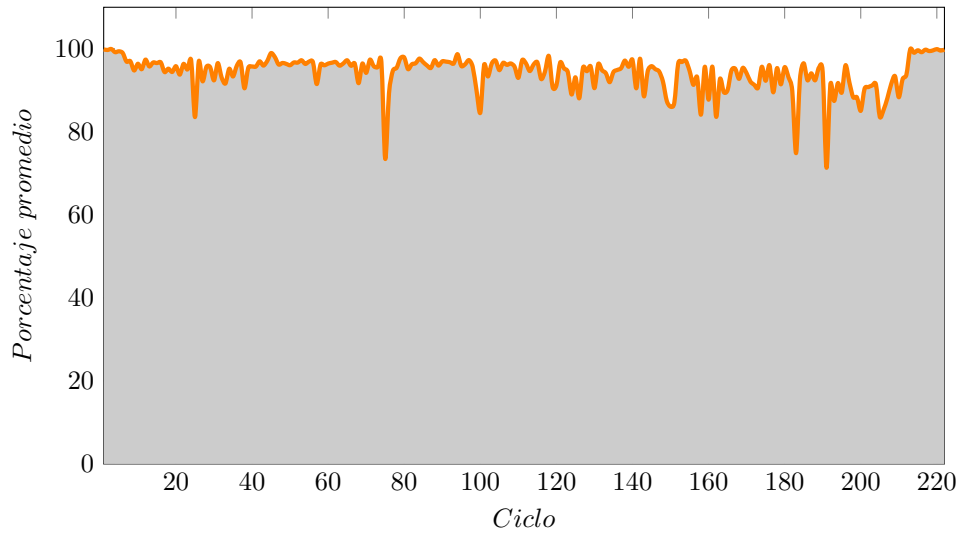


Fig. 4.34: Indicador 3 (Unidad 6)

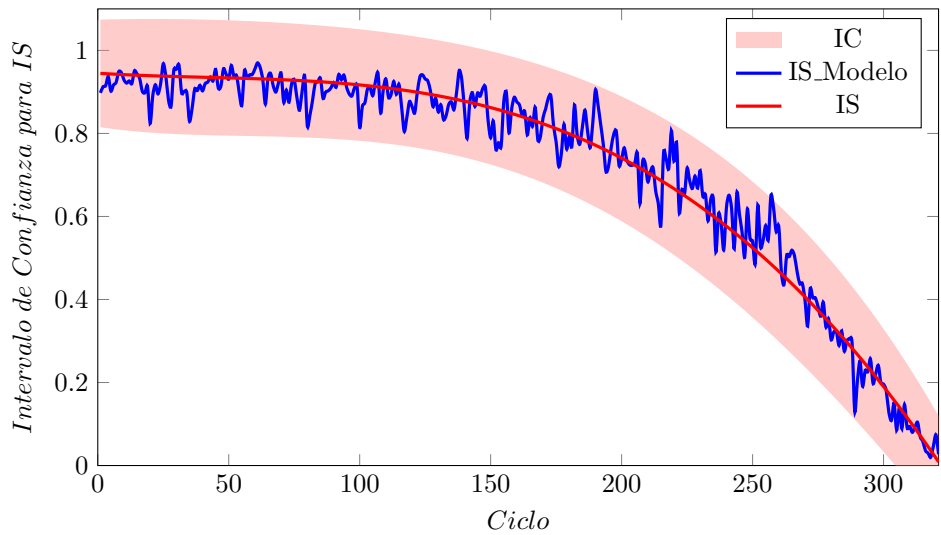


Fig. 4.35: IS vs. IS.Modelo y el intervalo de confianza del modelo (Unidad 147)

4. EXPERIMENTACIÓN Y APLICACIONES

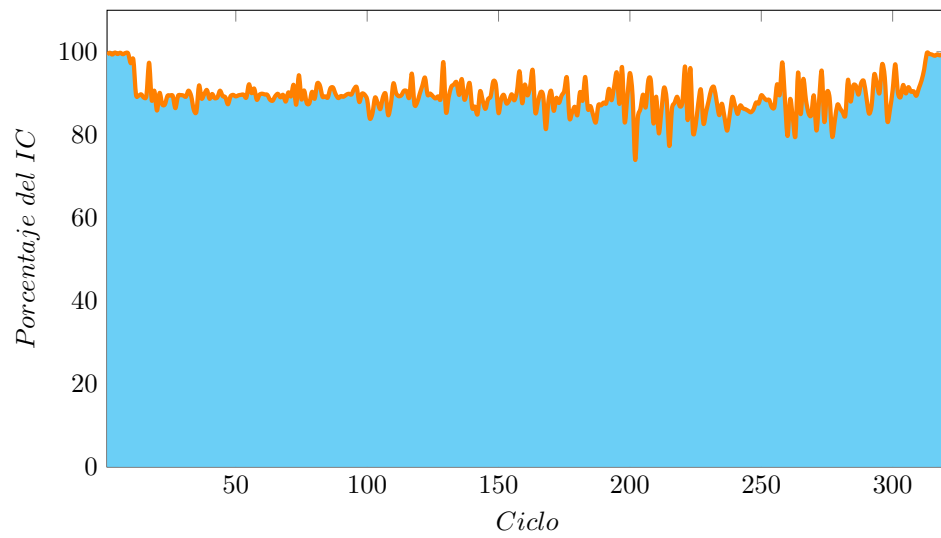


Fig. 4.36: Indicador 1 (Unidad 147)

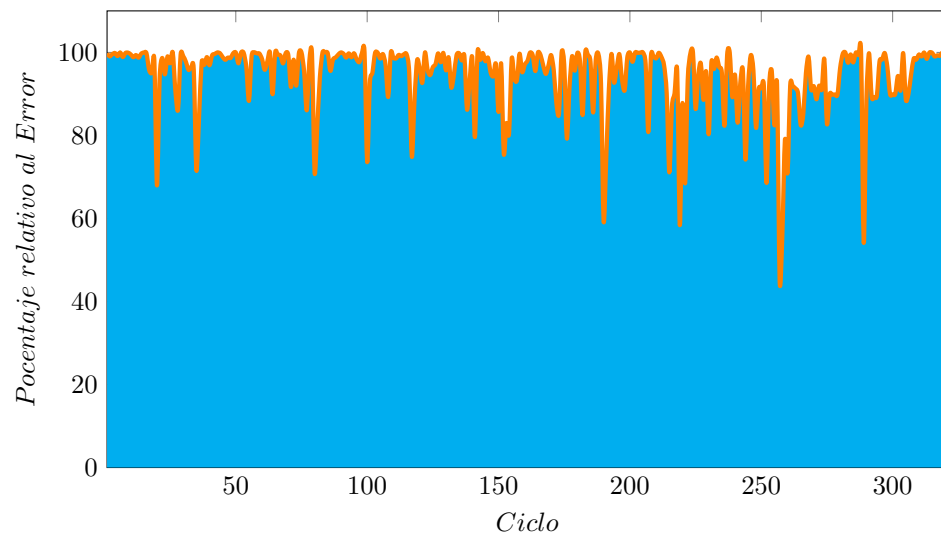


Fig. 4.37: Indicador 2 (Unidad 147)

4.2 Experimentación en el uso de las RNAs para la Predicción de la Vida Remanente Útil en Equipos Industriales

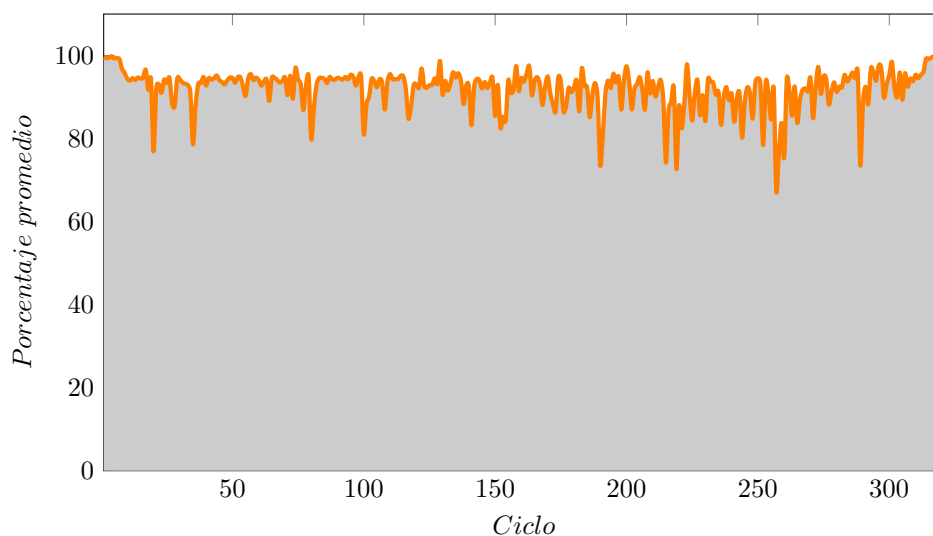


Fig. 4.38: Indicador 3 (Unidad 147)

4.2.4 Conclusiones

Los resultados obtenidos pueden permitir llegar a la conclusión de que los dos distintos enfoques, empleados para estimar la curva de VR, son correctos ya que proporcionan el mismo resultado. En términos de tiempo de cálculo, el segundo enfoque supera con ventaja al primer enfoque, sin embargo el primer enfoque supera al segundo en el ajuste. Como la implementación de los dos enfoques ha sido sencilla y no requieren mucho tiempo de cálculo ni memoria, se pueden emplear conjuntamente para que el uno valide el resultado obtenido mediante el otro enfoque.

Por otra parte, añadir un criterio de fiabilidad basado en el intervalo de confianza es de gran ayuda para saber la exactitud de la estimación del modelo neuronal, y permite definir el rango / intervalo de donde debe estar la respuesta correcta.

Por último el indicador (Ind3), que tiene en cuenta el nivel de confianza dado por la red neuronal y el error cometido a nivel individual (en cada punto) y no a nivel global (todo el conjunto) es de ayuda para la toma de decisión en ese mismo punto .

4. EXPERIMENTACIÓN Y APLICACIONES

4.3 Experimentación en el uso de las RNAs para la Predicción del Comportamiento de Incendios Forestales

Se ha mencionado anteriormente en [2.2.2](#), durante más de cuatro décadas, los investigadores de los Servicios Forestales han estado desarrollando y refinando varios sistemas nacionales para identificar (dónde, cuándo y cómo) el comportamiento de incendios forestales. Esta sección analiza la eficiencia de las redes neuronales como herramienta de inteligencia artificial, aprendiendo a través de datos históricos, el comportamiento de la propagación de los incendios forestales. Remplazar los modelos existentes o de apoyarlos con modelos basados en redes neuronales (Por ejemplo cuando falta algún dato o en caso de tener una mala medición, impide al modelo realizar sus cálculos), puede jugar papel crucial en la predicción rápida y segura de variables críticas en los futuros incendios forestales.

En esta sección, se muestran los resultados obtenidos con la red neuronal propuesta entrenada con los datos disponibles en un caso de un incendio forestal conocido como “SAU13”. También se muestran los resultados para otros casos reales de incendios forestales y se comparan los resultados con los que proceden de los informes oficiales de ellos. Finalmente se presentan las conclusiones alcanzadas.

4.3.1 Estructura del modelo ANNFBP

for Fire Behaviour Predictions (ANNFBP) tiene como objetivo fundamental la predicción de algunos parámetros principales que determinan la propagación de un incendio forestal. La estructura de esta red neuronal está inspirada y basada en la estructura del Subsistema de Predicción de Comportamiento de los Incendios (FBP) y el Índice Meteorológico de Peligro de Incendios (FWI) (ver el apartado [2.2.4.1](#)). Las entradas de la red neuronal, han sido elegidas cuidadosamente (desde el Modelo conceptual de la gestión de incendios forestales y el Índice Meteorológico de Peligro de Incendios (FWI), ver [2.12](#) y [2.13](#)). Como resultado se obtuvo la estructura del modelo neuronal propuesto que se muestra en la figura [3.9](#).

4.3.2 Arquitectura del Modelo ANNFBP

La arquitectura del modelo ANNFBP consiste de una red neuronal feedforward con tres capas, con la primera y la tercera reservadas para las entradas y las salidas deseadas. El algoritmo de entrenamiento usado está definido en [1](#). Respeto a las funciones de activación tanto para la segunda capa como para la última, se ha utilizado la tangente hiperbólica definida en

4.3 Experimentación en el uso de las RNAs para la Predicción del Comportamiento de Incendios Forestales

la ecuación 2.4. Y en la tabla 4.4, se muestra las arquitecturas obtenidas después de aplicar el algoritmo de reducción de variables mencionado en 3.4.

Tabla 4.4: Arquitectura del ANNFBP

	1ª Capa (Entradas)	2ª Capa (Neuronas Ocultas)	3ª Capa (Salidas)
ANNFBP-1	TC, Temp, RH, R, WS, FPMC, DC, BUI, ISI	27	FWI, TF, IF, VP
ANNFBP-2	TC, Temp, RH, R, WS, FPMC, ISI	23	FWI, TF, IF, VP
ANNFBP-3	Temp, RH, WS, R, FPMC, DMC, DC, ISI ISI, BUI	19	FWI
ANNFBP-4	Temp, RH, WS, R, FPMC, ISI	18	FWI
ANNFBP-5	Temp, RH, WS, R, FPMC, DC, ISI	18	Area

4.3.3 Entrenamiento y Validación del ANNFBP

El 13 de mayo de 2007 fue autorizado un incendio en Sharpsand para una investigación federal, manejado por el servicio forestal de recursos naturales de Canadá de la región de Ontario, en cooperación con la Unidad del Ministerio de Ciencias Naturales y Tecnología. El incendio fue iniciado a las 13 : 00 horas para producir y obtener información para su investigación, mediante la validación de los procedimientos que pueden ser utilizados con el fin de estimar el consumo de combustible de los incendios forestales de alta intensidad. Para obtener las mediciones, se utilizó la tecnología de control remoto infrarrojo, evaluación espacial y propagación temporal del fuego en unos terrenos complejos, que contienen como combustible los árboles de Pino de Banks (con 60 años de edad). La intención de provocar este incendio era también desarrollar técnicas de prueba mediante el uso de la teledetección infrarroja para ampliar la recopilación de datos relacionados con la velocidad de propagación y la intensidad, para mejorar la predicción del comportamiento del fuego en incendios forestales. El fuego se escapó de los límites del área establecida y se convirtió en un incendio forestal incontrolable, llamado SAU13, que consumió

4. EXPERIMENTACIÓN Y APLICACIONES

aproximadamente 1500 hectáreas. A base de los datos recogidos y validados en este incendio, ha sido entrenado nuestro modelo neuronal. El 70% de los datos se usaron para el entrenamiento, y el resto para la validación.

En la tabla 4.5, se evalúa el rendimiento de la red neuronal ANNFBP, con las mismas métricas definidas en el apartado 3.3.3.

A base de los datos recogidos y validados en este incendio, ha sido entrenado el modelo neuronal propuesto en este trabajo. El 70% de los datos se usarán para el entrenamiento y el resto para la validación.

En la tabla 4.5, se evalúa el rendimiento de la red neuronal ANNFBP, con las mismas métricas RMSE, MSE, EXAC PREC definidas en el apartado 3.3.3, para cuatro redes neuronales.

La red ANNFBP-2 tiene las mismas salidas que la ANNFBP-1, pero con menos entradas. Este proceso de reducción de variables se ha realizado mediante el algoritmo detallado en 3.4. Lo mismo para la red ANNFBP-4 que calcula el FWI con menos entradas que la ANNFBP-3.

Se observa que la ANNFBP 2 y 4 tienen similitud en términos de error con las métricas empleadas, que la ANNFBP 1 y 3, con una diferencia máxima de del orden de 0.07%.

4.3.4 Test del modelo ANNFBP con otros casos de estudio

En este apartado, se muestra la aplicación de las redes neuronales ANNFBP 1-4 en otros tres casos distintos del SAU13. Estos casos fueron elegidos porque están bien documentados y contienen datos similares (entradas y salidas) a los incluidos en el modelo ANNFBP, permitiendo en consecuencia la posibilidad de comparación de los resultados obtenidos con los resultados publicados en estos casos.

1. Caso del Incendio Mt-Cook:

El 16 de enero de 2008, el incendio Monte Cook (Mt Cook) fue iniciado en tierras de cultivo, situadas en el lado oriental del lago Pukaki en South Canterbury. El incendio fue muy significativo y ha sido el mayor incendio forestal que se ha producido en el distrito de incendios rurales del Sur de Canterbury en más de una década. Los vientos fuertes obstaculizaron los esfuerzos para controlar el fuego bajo. La topografía de la zona del incendio se caracterizó como "terreno moderadamente empinado" con pendientes de 15° a 20° (185). La agencia de control calculó el comportamiento del incendio para los tipos de combustible existentes, utilizando los modelos que se basan tanto en el sistema de

4.3 Experimentación en el uso de las RNAs para la Predicción del Comportamiento de Incendios Forestales

Tabla 4.5: Rendimiento del Modelo ANNFBP

Red Neuronal	Etapas	Salida	RMSE	MSE	EXAC	PREC
ANNFBP-1	Entrenamiento	FWI	0,00069	0,00048	0,99999	0,00069
		VP	0,00072	0,00054	0,99999	0,00072
		IF	0,00075	0,00056	0,99997	0,00075
		TF	0,00116	0,00056	0,99999	0,00116
	Validación	FWI	0,00778	0,00691	0,99984	0,00599
		VP	0,01702	0,01306	0,99497	0,01165
		IF	0,02053	0,01366	0,99601	0,0156
		TF	0,06634	0,01366	0,98299	0,06558
ANNFBP-2	Entrenamiento	FWI	0,00128	0,00097	0,9999	0,00128
		VP	0,00252	0,00159	0,99982	0,00252
		IF	0,00197	0,00166	0,99985	0,0019
		TF	0,00139	0,00166	0,99999	0,00139
	Validación	FWI	0,00358	0,00256	0,99996	0,003
		VP	0,01614	0,01219	0,99578	0,01479
		IF	0,01472	0,00926	0,99731	0,01252
		TF	0,05994	0,00926	0,98633	0,05994
ANNFBP-3	Entrenamiento	FWI	0,00037	0,00033	0,99878	0,00084
	Validación	FWI	0,00037	0,00033	0,99896	0,00085
ANNFBP-4	Entrenamiento	FWI	0,01974	0,01444	0,99374	0,01974
	Validación	FWI	0,04589	0,03336	0,93456	0,04585

4. EXPERIMENTACIÓN Y APLICACIONES

predicción de comportamiento de incendios de Nueva Zelanda (186) y el sistema FBP canadiense (187).

Usando la red ANNFBP-4 (ver 4.5), se obtuvo la predicción del parámetro FWI y se comparó con los resultados obtenidos de la interpretación diaria y los valores del sistema FWI de la estación Mt Cook que a su vez se basó en los datos recogidos de la estación meteorológica de Tekapo y las precipitaciones de Braemar (185). Los resultados se muestran en la figura 4.39.

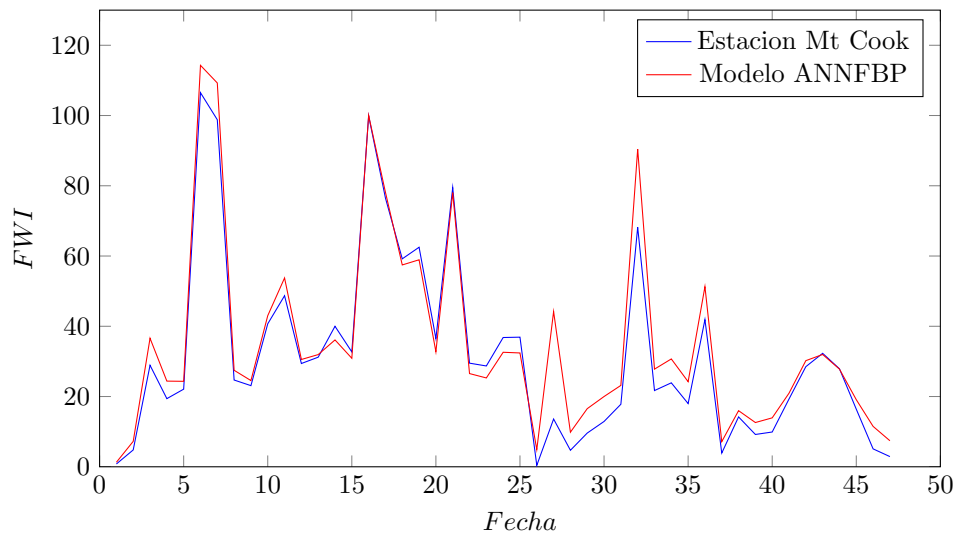


Fig. 4.39: Comparación de la predicción de FWI: ANNFBP vs. Estación Mt Cook

En la figura 4.39, la predicción de FWI fue prácticamente la misma que la obtenida en (185) con una precisión del 96%. Por lo tanto se demuestra que las redes neuronales pueden aprender la relación entre la temperatura, la humedad relativa, la velocidad del viento y la precipitación de las 24h anteriores, así como la FFMC e ISI (Entradas de la RNA) para predecir el FWI.

En la tabla 4.6 se muestra el resultado de la red neuronal ANNFBP-2 (ver 4.5) de la predicción para las entradas disponibles en (185) cuyo objetivo es predecir la velocidad de propagación (VP).

En la tabla 4.6, las predicciones de la ANNFBP-2 son más mayores que los valores predichos por la estación MtCook usando (185). Por lo tanto el perímetro total quemado predicho por ANNFBP-2 es de 13.3 km, sin embargo con el otro método se obtuvo 9.3 km. El

4.3 Experimentación en el uso de las RNAs para la Predicción del Comportamiento de Incendios Forestales

área del incendio abarcó unas 756 hectáreas con unos 14 *km* de perímetro. Por la tanto, la predicción del modelo neuronal ANNFBP-2 se acercó al perímetro quemado.

Tabla 4.6: Predicción de VP mediante ANNFBP vs. Observado por La estación de Mt Cook basandose en modelos de Nueva Zelanda

Id	Intervalo de Tiempo (DST)	VP Observada (m/h)	Predicción de ANNFBP (m/h)
16 Enero 2008			
1	1820 - 1940	600	994,54
2	1940 - 2200	1400	1745,36
3	1820 - 2200	1700	2739,90
17 Enero 2008			
4	0220 - 0315	200	603,84
5	0220 - 0350	200	916,77
6	0350 - 0600	1100	578,08
7	0220 - 0600	1300	1494,85
8	0600 - 0630	310	354,39
9	0600 - 0630	400	354,39
10	0630 - 0700	600	354,39
11	0630 - 0800	640	1057,68
12	0630 - 0800	530	1057,68
13	0630 - 0800	400	1057,68
Suma del perímetro Quemada		9380	13309,57

2. Caso del Incendio Chisholm:

El incendio Chisholm, comenzó el 23 de mayo de 2001, el área quemada estaba a unos 16 km al norte de Edmonton, en Alberta, Canadá. Bajo condiciones extremas de sequía, variaciones en las precipitaciones y con vientos fuertes, los comportamientos del incendio mostraron tasas altas de propagación y el incendio fue de tipo ”*corona continua*”.

El 27, 28 de mayo de 2001, los servicios forestales pudieron controlar la extensión del incendio cerca de Chisholm, el 31 de mayo, el fuego ardía en Smith a 10 km de Chisholm. El incendio Chisholm fue el más grande entre los seis incendios forestales ocurridos en Alberta.

El resumen de los eventos del incendio Chisholm desde el momento en que se detectó el fuego, se puede consultar en (188). La figura 4.39 presenta la predicción de FWI y VP

4. EXPERIMENTACIÓN Y APLICACIONES

mediante el modelo ANNFBP vs la predicción del modelo utilizado por la División de Protección Forestal de Alberta (2001).

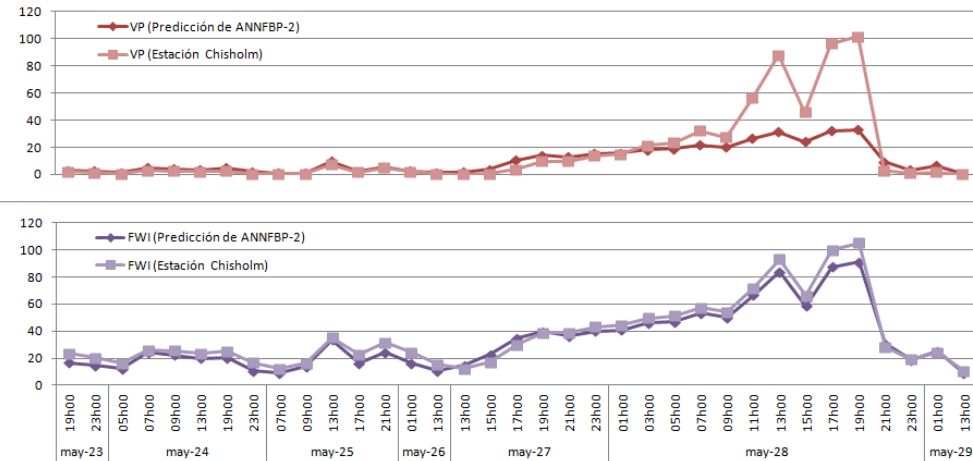


Fig. 4.40: Predicción de FWI y VP mediante ANNFBP-2 y Estación de Alberta

Tabla 4.7: Evaluación del modelo ANNFBP v. Estación de Alberta

	RMSE	MSE	EXAC ($\alpha = 1$)	EXAC ($\alpha = 2$)	PREC
FWI	5,509658	4,633841	0,83642	0,948103	4,094063
VP	19,84228	9,067916	0,582453	0,680973	18,86979

En la figura 4.40, la predicción de la FWI y VP por la ANNFBP-2 se presenta y muestra que la predicción de la FWI y VP son los mismos que los obtenidos mediante la predicción de la División de Protección Forestal del Desarrollo Sostenible de los Recursos de Alberta (188), a excepción de los valores mayores a 35m/min para la predicción de VP. Este fenómeno, se puede explicar por el conjunto de entrenamiento de la red neuronal, ya que los datos de entrenamiento disponibles para la red neuronal no contienen estos rangos. Es un defecto conocido que se produce en los enfoques de redes neuronales: “Las redes neuronales sólo aprenden de la información ya disponible en el conjunto de datos de entrenamiento”.

Para determinar las estimaciones realistas de diversos terrenos (combustible y topografía) y bajo una variedad de condiciones climáticas, se debe disponer de información completa y amplia sobre las distintas variedades de parámetros a afrontar. En la tabla 4.7 se

4.3 Experimentación en el uso de las RNAs para la Predicción del Comportamiento de Incendios Forestales

presentan la evaluación del modelo ANNFBP-2.

3. Caso del parque natural de Montesinho:

La detección rápida de un incendio forestal es un elemento clave para el control de este fenómeno. Para lograr este objetivo, una alternativa es el uso de herramientas automáticas basadas en sensores locales implementados por las estaciones meteorológicas que capturan las condiciones ambientales (temperatura, viento, humedad, etc.) que influyen principalmente en los incendios forestales.

En esta aplicación, se explora un enfoque de minería de datos (Data Mining) para predecir el área quemada de los incendios forestales. Los datos de entrenamiento, validación y de test se pueden consultar en:

<http://archive.ics.uci.edu/ml/machine-learning-databases/forest-fires/>.

Se trata de un histórico de datos meteorológicos recogidos en la región noreste de Portugal, precisamente del parque natural de Montesinho. Se pretende también analizar si las redes neuronales son capaces de predecir el área quemada en incendios pequeños, que son más frecuentes en dicha área y otras muchas. Tal conocimiento es particularmente útil para mejorar la gestión de los recursos locales de extinción de los incendios.

El conjunto de datos contiene las variables siguientes: fecha, coordenadas, temperatura, humedad relativa, viento, precipitaciones, FFMCI, DMC, DC, ISI (ver definiciones en el apartado 2.2.4.1), y la área quemada. La variable "fecha", no se ha tenido en cuenta en la elaboración de la red neuronal, dado que se pretende analizar la relación entre las variables independientemente del tiempo. En la figura 4.41, se aprecia la distribución de las variables disponibles, con sus histogramas en la diagonal de la figura. Diversos trabajos han sido publicados para resolver este problema, entre ellos el trabajo de Cortez in (95), (189), que usaron las redes neuronales.

En este ejemplo, se usa la red neuronal ANNFBP-5. La evaluación de las predicciones frente a las observaciones reales, se pueden apreciar en la tabla 4.8.

Tabla 4.8: Evaluación de la predicción del área quemada: ANNFBP-5 frente a las mediciones reales

	RMSE	MAE	RMSPE	PREC
Area	7,1266	3,7002	71,266	7,1197

4. EXPERIMENTACIÓN Y APLICACIONES

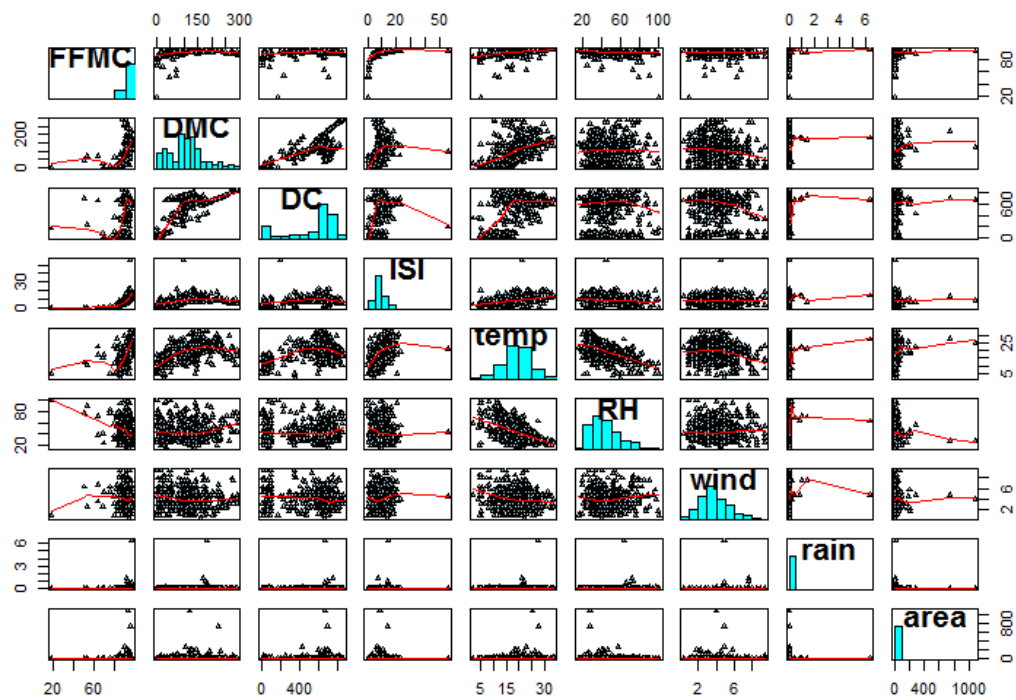


Fig. 4.41: Dispersión matricial de los datos disponibles

4.3.5 Conclusiones

Mediante el análisis de los resultados obtenidos en cada caso de estudio, se puede concluir que existen muy buenas aproximaciones para la predicción del índice meteorológico de peligro - FWI de los incendios (más de 96% de precisión). Este resultado confirma la posibilidad de uso del modelo ANNFBP (las redes 3 y 4) para predecir el índice FWI en diferentes lugares y que proporciona un mayor nivel de confianza a las predicciones que los otros métodos.

Sin embargo, en las predicciones de VP, IF y TI, los resultados son ligeramente diferentes, debido a la variación inherente en los combustibles y la topografía. Varias variables no están incorporadas en el modelo ANNFBP debido a la falta de más datos históricos, lo que podría haber facilitado el aprendizaje, aprendiendo con más información en la fase de entrenamiento del modelo (rango de datos, más tipos de combustible, ..., etc.).

A pesar de que se han podido obtener buenas aproximaciones con un modelo entrenado con pocos datos, se deben añadir más casos históricos con una amplia gama de datos de manera que se puede generalizar y proporcionar un mayor grado de confianza, y también aumentar su fiabilidad a la hora de tomar decisiones en tiempo real.

Respecto al caso de predecir el área quemada en el parque natural Montesinho, la precisión de la red neuronal ANNFBP-5 ha sido buena con tan solo 7.11% de error.

Para concluir, se puede decir que la primera ventaja que se puede observar es la capacidad de la red neuronal para predecir con pocas entradas algunos parámetros decisivos del comportamiento de los incendios forestales, ya que no se usaron todas las entradas que figuran en el modelo conceptual de la gestión de incendios forestales 2.12 y el Índice Meteorológico de Peligro de Incendios (FWI) 2.13. Además parece bastar con una sola capa oculta, en vez de tener varias capas de ellas, por lo tanto se ocupa menos espacio de almacenamiento en la memoria interna y, sobre todo, se emplea menor tiempo de entrenamiento. Otro punto a favor es la posibilidad predicción de varios parámetros al mismo tiempo, sabiendo que la mayoría de los modelos clásicos no disponen de esta posibilidad, además de que necesitan mucho más tiempo para realizar los cálculos y simulaciones.

4. EXPERIMENTACIÓN Y APLICACIONES

4.4 Experimentación de las RNA para la estimación de la Demanda de Energía Eléctrica

Este estudio de la Demanda de Energía Eléctrica (DEE) tiene por objeto estimar la cantidad de energía eléctrica hasta el año 2020 para la totalidad de Canarias. Para ello se usan métodos estadísticos clásicos (Regresión Lineal Múltiple RLM, Mínimos Cuadrados Parciales "MCP") y métodos de inteligencia artificial (algoritmos genéticos "AG", máquinas de soporte vectorial "MSV" y las RNAs), analizando sus eficiencias. Los valores históricos del consumo de energía eléctrica se han obtenido del Instituto Canario de Estadística, ISTAC, <http://www.gobiernodecanarias.org/istac/>. Sólo se dispone de datos a partir del año "1991". En especial se consultan y se analizan los valores de la demanda de energía eléctrica en el pasado, especialmente los efectos en la demanda de energía eléctrica dentro y fuera de épocas de crisis.

Los valores de las predicciones de la población se han tomado de la Proyección de la población española desde el año 2014 hasta 2064, elaborado por el INE. Respecto al IPC y el PIB, se ha empleado una estimación de la Planificación energética indicativa (Ley 2/2011, de 4 marzo de Economía Sostenible).

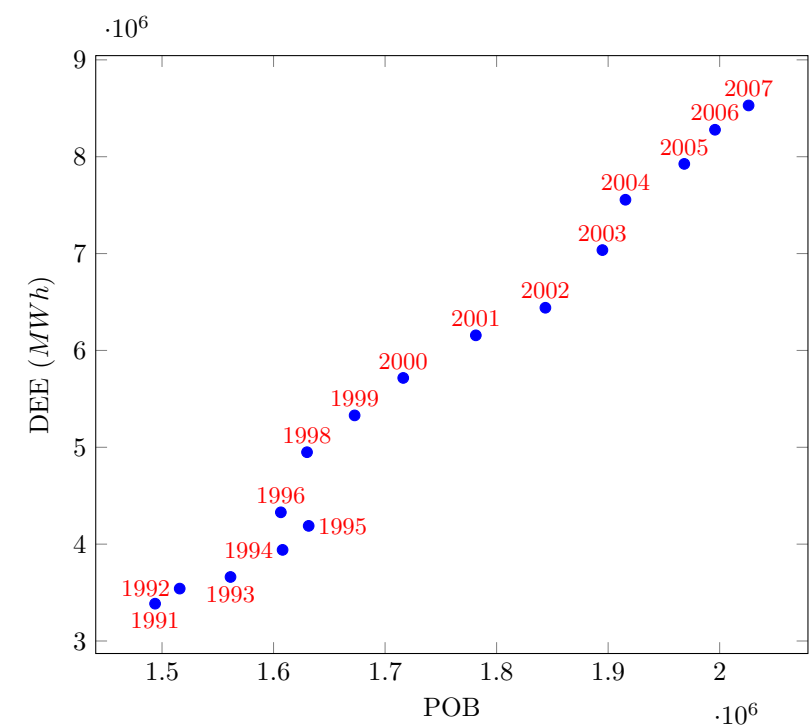
4.4.1 Análisis de datos

Los eventos económicos de los últimos años ponen de manifiesto que actualmente es más complejo predecir una evolución futura de la demanda eléctrica. Es evidente que se ha acentuado el comportamiento no lineal de la demanda, como se puede ver en la figura 4.43a, que se acentúa si en periodos sometidos a crisis económica sobre todo de notorio impacto después de 2008, que presenta importantes oscilaciones de las variables como el PIB y el IPC (ver 4.42b).

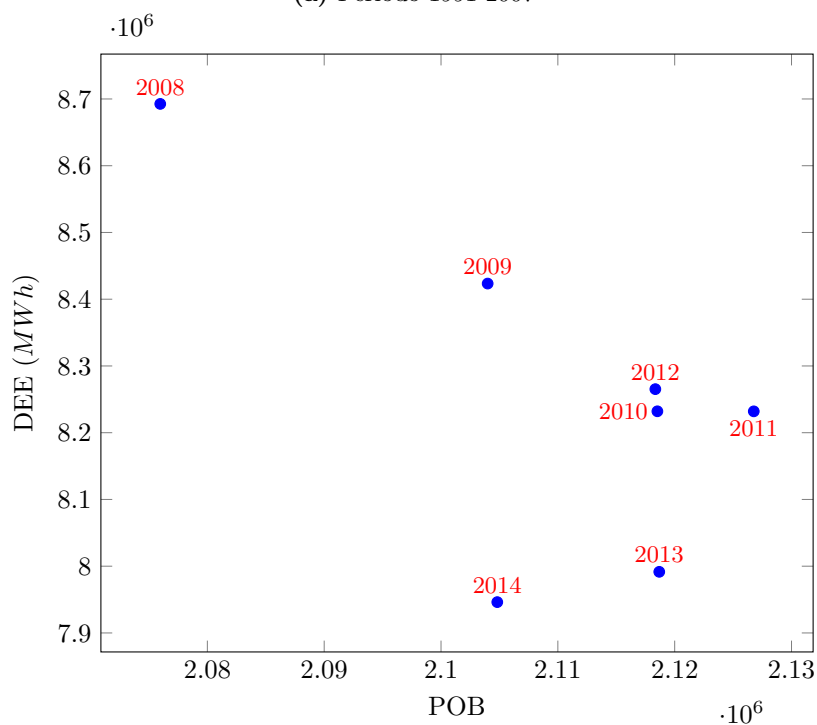
Las variables explicativas empleadas son las más habituales encontradas en la bibliografía especializada. Se tienen datos de la Población de cada isla, el PIB y del IPC referentes a toda Canarias, pero no se dispone a nivel insular de ninguna variable económica. El precio de la electricidad no se ha considerado por dos razones, la primera es que muchos autores la descartan por presentar una baja elasticidad con respecto a la demanda, la segunda es que no existen estimaciones de su evolución en el futuro.

A fin de poder realizar la comparación de los diferentes métodos, se emplean las mismas variables y de forma aleatoria se ha tomado los mismos años test, siendo estos 1992, 2001, 2004,

4.4 Experimentación de las RNA para la estimación de la Demanda de Energía Eléctrica



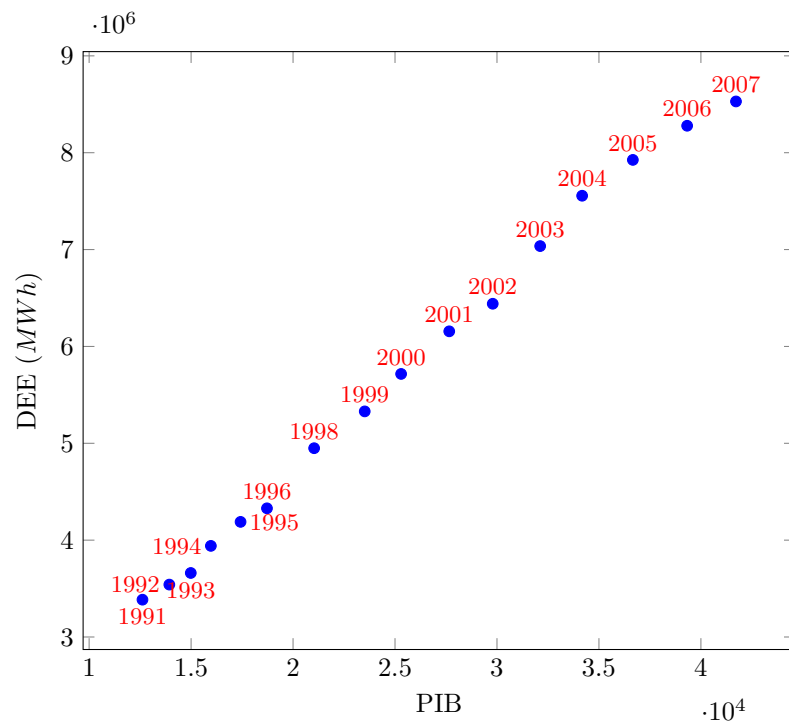
(a) Periodo 1991-2007



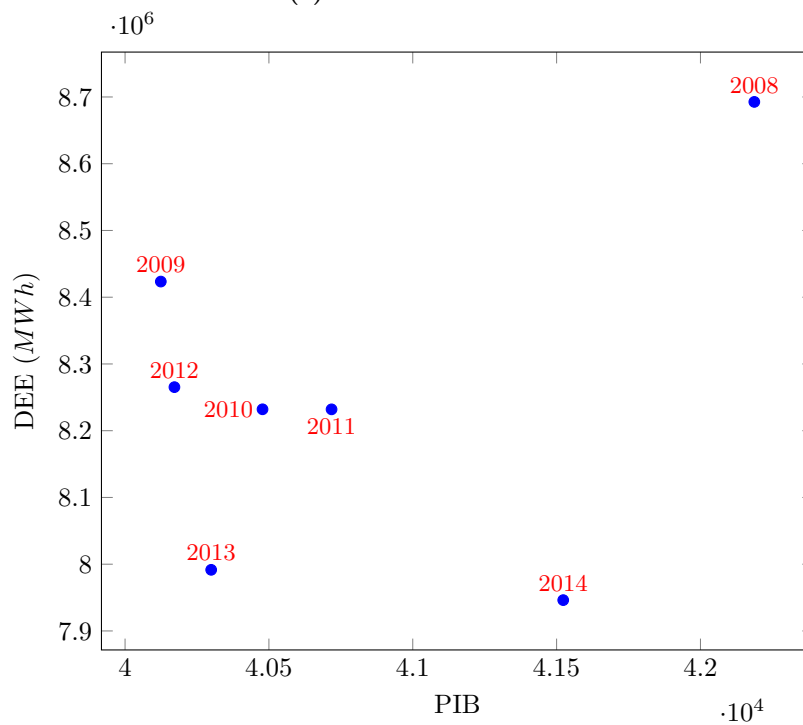
(b) Periodo 2008-2014

Fig. 4.42: DEE vs. POB, Islas Canarias

4. EXPERIMENTACIÓN Y APLICACIONES



(a) Periodo 1991-2007



(b) Periodo 2008-2014

Fig. 4.43: DEE vs. PIB, Islas Canarias

4.4 Experimentación de las RNA para la estimación de la Demanda de Energía Eléctrica

2009 y 2014. El resto de los años han sido utilizados para ajustar los modelos. A continuación, se presentan los resultados obtenidos.

4.4.2 Estimación de la DEE mediante regresión lineal y logarítmica

Los modelos econométricos usan, entre otros, métodos paramétricos: regresión múltiple, regresión logarítmica o bien series temporales. En este apartado se aplica regresión lineal múltiple, del tipo:

$$DEE = a_1POB + a_2PIB + a_3IPC + a_4$$

Y la regresión logarítmica, del tipo

$$\log(DEE) = a'_1\log(POB) + a'_2\log(PIB) + a'_3\log(IPC) + a'_4$$

Donde $a_i, a'_i, \forall i$ son coeficientes a ajustar.

A continuación se presenta en la figura 4.44b los resultados obtenidos mediante la regresión lineal, usando el software libre R (<https://www.r-project.org/>), y la función "lm".

En la figura 4.45b, se presentan los resultados de la estimación de la DEE mediante regresión logarítmica, usando el software libre R (<https://www.r-project.org/>), y la función glm.

Los detalles empleados para la regresión lineal múltiple y la logarítmica (190).

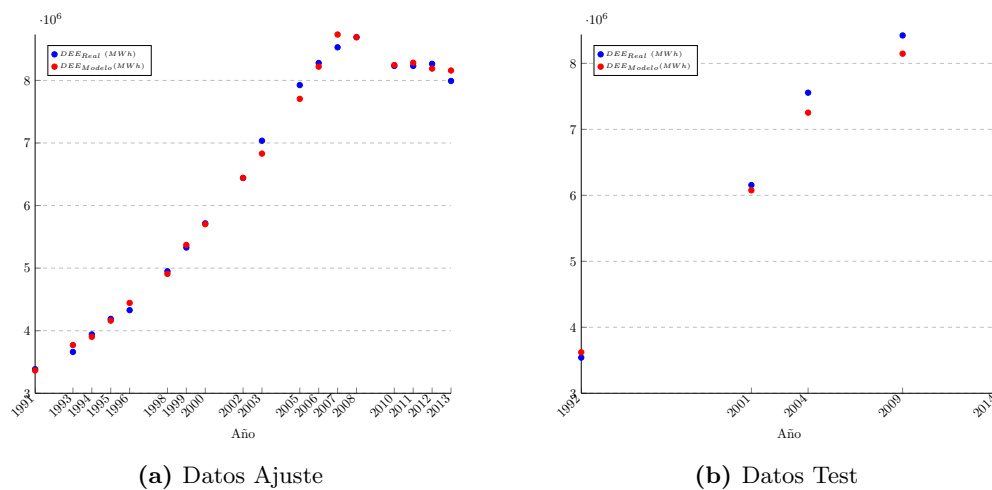


Fig. 4.44: Estimación de la DEE mediante RLM

4. EXPERIMENTACIÓN Y APLICACIONES

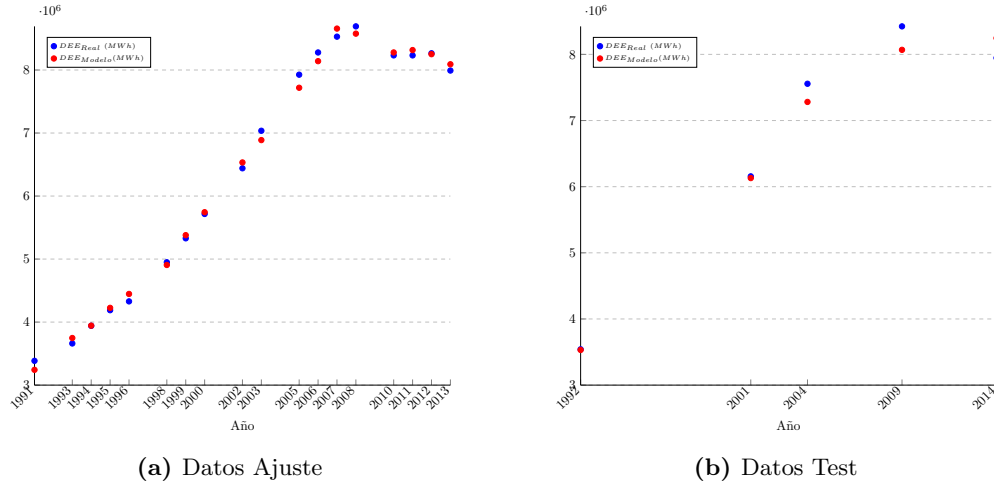


Fig. 4.45: Estimación de la DEE mediante RLog

4.4.3 Estimación de la DEE mediante Inteligencia Artificial

En este apartado, se presentan los resultados obtenidos con los tres métodos mencionados anteriormente: Máquina de Soporte Vectorial (MSV), Algoritmos Genéticos (AG) y Red Neuronal Artificial (RNA).

Las etapas y los detalles de implementación del MSV, y del AG para este caso de estudio, se pueden consultar en (190). Aquí solo se presentan los resultados obtenidos.

4.4.3.1 Mediante MSV

En la figura 4.46b, se presentan los resultados empleando el software R, con la función “ksvm”, de la librería “kernlab” para la regresión mediante MSV con núcleo lineal.

$$k(x_i, x) = x_j^T x_i$$

(Función Lineal)

En la figura 4.47b, corresponde a la estimación de la DEE con MSV con núcleo gaussiano, su función núcleo se define:

$$k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma^2}\right)$$

(Función Gaussiana; $\|\cdot\|$ es la norma euclídea)

Variando el parámetro σ y el coeficiente de regularización C , se muestra el mejor resultado obtenido. Se muestra el mejor resultado obtenido en la figura 4.47, correspondiente a $C = 10$ y $\sigma = 0.05$ para el MSV con núcleo gaussiano, y $C = 3$ para MSV con núcleo lineal.

4.4 Experimentación de las RNA para la estimación de la Demanda de Energía Eléctrica

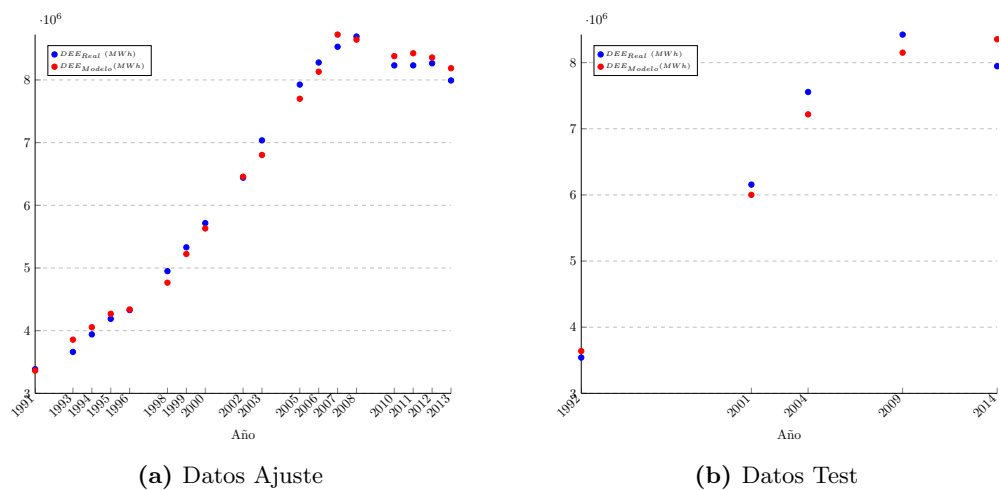


Fig. 4.46: Estimación de la DEE mediante MSV con núcleo lineal

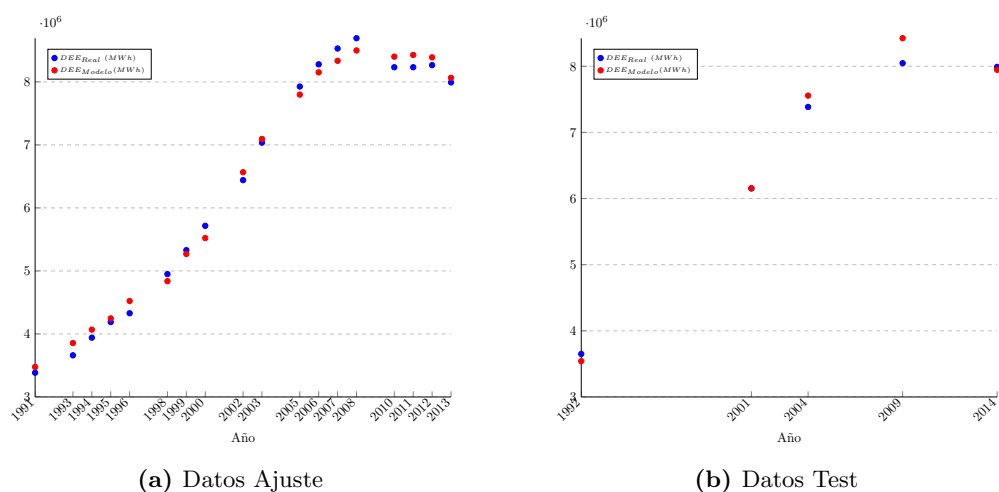


Fig. 4.47: Estimación de la DEE mediante MSV con núcleo gaussiano

4. EXPERIMENTACIÓN Y APLICACIONES

4.4.3.2 Mediante AG

La función objetivo del problema de optimización que se resuelve, se define:

$$DEE = e^{\frac{a_1 \log(POB)}{a_2 + \log(POB)} + \frac{a_3 \log(PIB)}{a_4 + \log(PIB)} + \frac{a_5 \log(IPC)}{a_6 + \log(IPC)} + a_7}$$

Donde $a_i, \forall i$ son los coeficientes a optimizar en la minimización de la función de la DEE.

Los detalles del algoritmo genético y su implementación en este caso, se pueden consultar en (190).

A continuación en la figura 4.48, se muestran los resultados obtenidos.

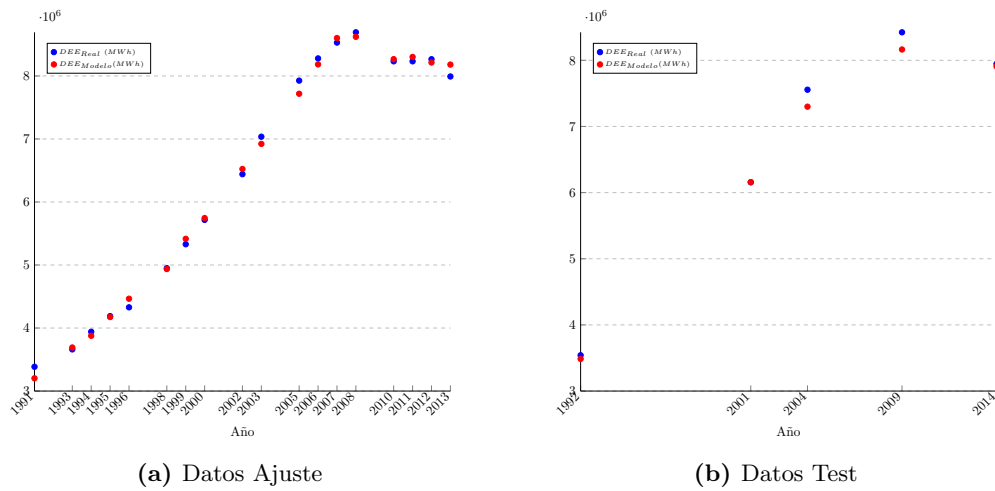


Fig. 4.48: Estimación de la DEE mediante AG

4.4.3.3 Mediante RNA

Dado que una RNA de única capa oculta puede aproximar hasta el nivel deseado para cualquier función continua en un intervalo (ver el teorema 14), en este trabajo, la RNA propuesta contiene una capa oculta (Ver figura 4.49).

4.4 Experimentación de las RNA para la estimación de la Demanda de Energía Eléctrica

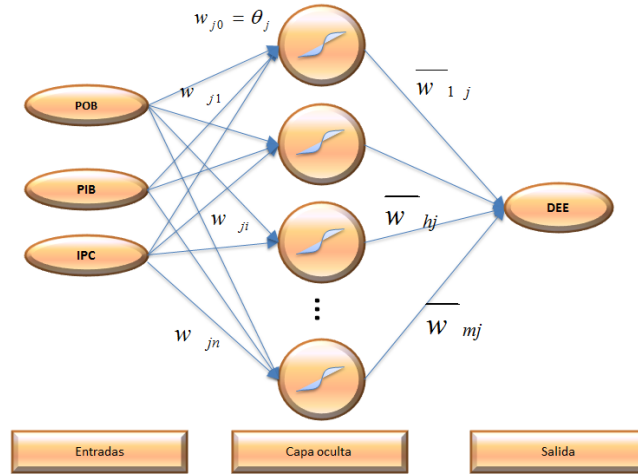


Fig. 4.49: Estructura de la RNA para predecir la DEE

Los datos se normalizan dentro del intervalo $[-1, 1]$. Aplicando el algoritmo RPROP definido en 2.1.10.4 Se entrena la red para un número limitado de episodios, variando en cada entrenamiento el número de neuronas ocultas. El número de neuronas con una convergencia satisfactoria se toma como tamaño de la capa oculta. La función de activación usada es la TANH definida en 8.

Para cada procedimiento de inicialización de la RNA, 30 redes diferentes (con distintos pesos aleatorios iniciales) son entrenadas durante un máximo de 100 iteraciones para evitar el sobreentrenamiento. La estructura de la mejor red neuronal corresponde a una red neuronal de 5 neuronas en la capa oculta. Y ha requerido tan solo 40 iteraciones para ajustar los pesos. En la figura 4.50, se muestran las predicciones de la RNA tanto para el conjunto de ajuste como para el de test.

4.4.4 Eficiencia de los métodos

Para una adecuada comparativa entre métodos, debemos considerar para todos la misma función de medida del error asociada al ajuste funcional, y también igual para todos los métodos la misma función de medida del error asociada a la evaluación de las predicciones de la demanda en años test.

En la tabla 4.9, se muestra la eficiencia de los métodos utilizados. En este caso se ha usado la métrica MAPE mencionada anteriormente en nota 3.3.3, dado que el número de datos es pequeño. Comparando el rendimiento de nuestro modelo obtenido mediante la red neuronal

4. EXPERIMENTACIÓN Y APLICACIONES

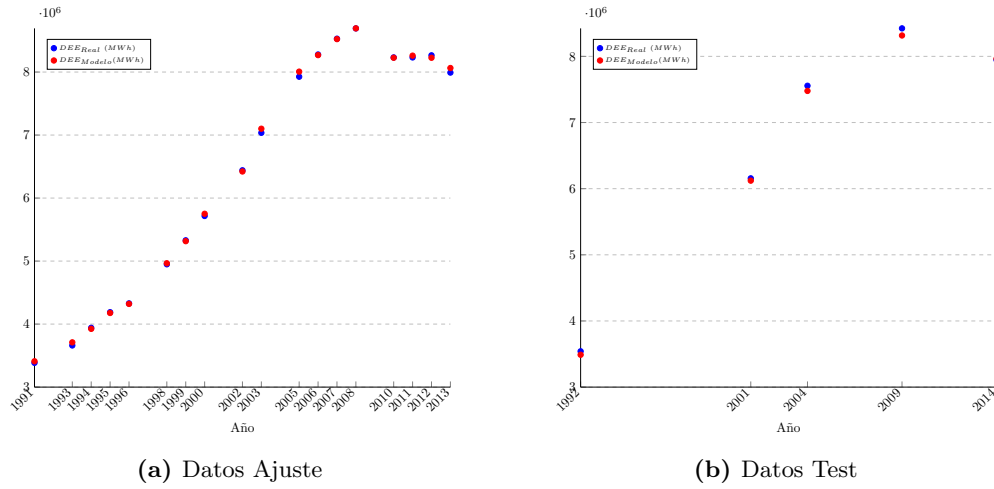


Fig. 4.50: Estimación de la DEE mediante RNA

y los demás métodos. Se puede observar que los métodos de inteligencia artificial superan los

MAPE %	Datos de ajuste	Datos de test
RLM	1.25	3.41
RLog	1.46	2.46
MSV (Lineal)	2.06	3.63
MSV (Gaussiana)	2.29	2.09
AG	1.46	1.71
RNA	0.65	1.22

Tabla 4.9: Evaluación del error de estimación de los métodos

modelos lineales, y en primera posición está el modelo neuronal que tiene el mejor ajuste tanto para el conjunto de ajuste y de test.

Otro punto importante, es el error obtenido mediante la RNA que es también pequeño. Pues ello indica que red neuronal ha capturado correctamente el *mapping* subyacente en los datos, lo cual permite su generalización.

4.4.5 Medida de confianza para la RNA

Utilizando el NNCI detallado en 3.2.2.2, que nos permite calcular el intervalo de confianza de la RNA, se obtiene el porcentaje para cada año, tanto del conjunto de datos de ajuste y de test.

Como se ha mencionado en 4.4.4, no es de extrañar que el nivel de confianza para la red neuronal sea muy alto. Más del 98% en el conjunto de ajuste (ver figura 4.51, y más del 97%

4.4 Experimentación de las RNA para la estimación de la Demanda de Energía Eléctrica

para el conjunto de test (ver figura 4.52)

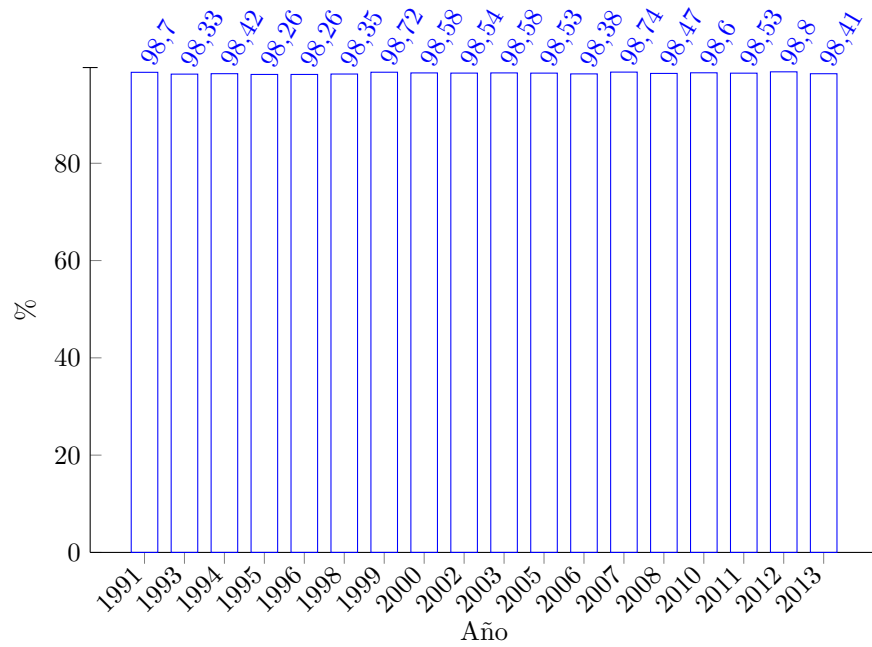


Fig. 4.51: Porcentaje de precisión para datos de ajuste

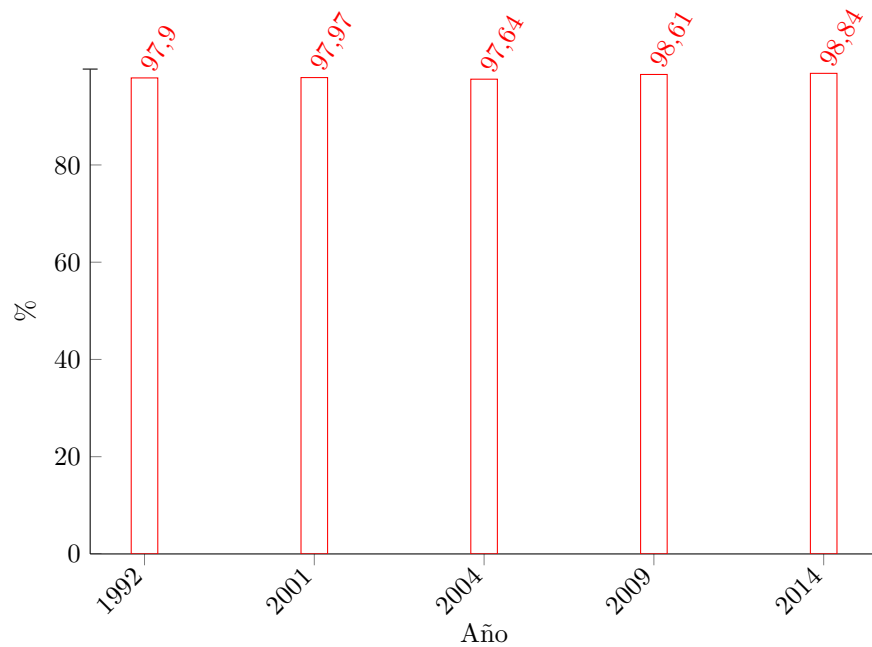


Fig. 4.52: Porcentaje de precisión para datos de test

4. EXPERIMENTACIÓN Y APLICACIONES

4.4.6 Conclusiones

Motivado por las grandes oscilaciones del PIB e IPC durante la actual crisis económica, en este trabajo se analiza la mejora de la predicción de la demanda de energía eléctrica, en la Región Canaria. Los datos recopilados se ajustan con métodos de predicción de distinta naturaleza. Los métodos no lineales o de inteligencia artificial ofrecen mejoras y buenas precisiones respecto a los resultados de predicción de los métodos econométricos lineales.

Las RNA mostraron capacidad de emular y aprender, en este caso con pocos datos, la relación entre el PIB, POB y el IPC, para predecir la DEE. Las estimaciones de las RNA han superado los demás métodos citados. Además la inclusión de los intervalos de confianza a la respuesta de las RNA, les permite añadir otro criterio de fiabilidad a la estimación. El grado de confianza permitirá disponer de una información adicional, necesaria y útil a la hora de estimar la DEE.

Los modelos de estimación de la DEE basados en inteligencia artificial presentan proyecciones más realistas en comparación con los modelos econométricos- lineales, que todavía sufren de deficiencias de no linealidad, especialmente en los últimos años.

Tras la experimentación realizada y de todas formas se consideran requisitos clave para obtener estimaciones más precisas de la DEE los siguientes:

- El desarrollo y mantenimiento de bases de datos de energía más detalladas,
- Un mayor desarrollo de modelos eficientes para reflejar mejor los factores e incertidumbres de las situaciones sociales y económicas

CAPÍTULO 5

CONCLUSIONES Y LÍNEAS FUTURAS

5. CONCLUSIONES Y LÍNEAS FUTURAS

En los últimos años, se ha emergido un nuevo concepto dentro del campo de la computación, que abarca un conjunto de metodologías inspiradas en los sistemas biológicos, para resolver problemas complejos de I-RAMS y del mundo real, ofreciendo soluciones robustas y de fácil implementación.

En este conjunto de metodologías emergentes se encuentran la computación evolutiva, la lógica difusa, las redes neuronales artificiales, los algoritmos genéticos y la teoría del caos y del aprendizaje, entre otras. Actualmente, las Redes Neuronales Artificiales están causando un mayor impacto, debido a su extraordinaria aplicabilidad en la práctica, causada por su flexibilidad para modelar fenómenos altamente no lineales.

En esta tesis, se han descrito diversas líneas de investigación en torno a la mejora del uso de RNA mediante técnicas que mejoran ciertas deficiencias de las mismas, experimentadas en aplicaciones del ámbito industrial y medioambiental. Cada una de ellas ha operado en un nivel de análisis distinto. Por un lado, se ha estudiado el uso de las RNA mejorando el entrenamiento y conceptos relacionados con su eficiencia mediante métricas de confianza. Por otro lado, se han aplicado las RNA en los campos de pronóstico, en la predicción asociada a incendios forestales y por último en la demanda de energía eléctrica.

Como fruto de esta investigación desarrollada, se han publicado artículos en diversas revistas científicas y existen otros en proceso de publicación. Por otra parte se han presentado trabajos en diversos congresos internacionales, que han sido publicados en libros. Esta tesis trata de aglutinar los logros alcanzados en este conjunto de trabajos, en los que se pone de manifiesto la utilidad de las redes neuronales para resolver problemas complejos. A continuación, se presenta una discusión global de los resultados obtenidos.

Actualmente, existen más de 40 paradigmas de RNA usados en diversos campos de aplicación/investigación. Las RNA poseen una serie de características interesantes, como la habilidad de procesamiento de datos con ruido o incompletos, su alta tolerancia a fallos, permitiendo operar satisfactoriamente con pocas neuronas/conexiones y su capacidad para responder en tiempo real debido a su paralelismo inherente.

Se ha destacado en los trabajos descritos en esta memoria, la red feedforward con un esquema de aprendizaje supervisado, que ha sido utilizada satisfactoriamente en problemas de diversa naturaleza.

La descripción de este tipo de red ha sido detallada en el capítulo 2, donde se han presentado los fundamentos esenciales con los cuales han sido creadas las redes neuronales artificiales. Entre ellos se han descrito los teoremas (como el Teorema de Convergencia de LMS, Teorema de

Convergencia del Perceptrón Simple, Teorema de Kolmogorov, Teorema de Stone-Weierstrass, Teorema de Cohen-Grossberg ...), que establecen los requerimientos fundamentales para la convergencia de las RNA.

En teoría, las redes neuronales pueden sintetizar y aproximar cualquier función no lineal. Sin embargo, no existe una metodología concreta para la selección de la red neuronal (estructura, número de capas y de neuronas ocultas y tipo de funciones de activación). El problema del aprendizaje, en particular en aplicaciones de ingeniería y problemas reales de riesgo, también suele ser complejo. Por otra parte, es importante señalar que las redes neuronales no son capaces de aprender propiedades que no están contenidas o reflejadas en los datos disponibles durante el entrenamiento.

Después de presentar en general las redes neuronales artificiales, esta tesis se centra en las contribuciones fruto de la investigación realizada, siguiendo cuatro ejes fundamentales:

- La mejora del aprendizaje de la red neuronal acelerando su entrenamiento, estableciendo algoritmos y técnicas, demostrando su eficiencia en ayuda a la construcción de una RNA adecuada, tales como la inicialización de los pesos de forma determinista eficiente, que garantiza la convergencia siempre. El análisis de las funciones de activación para la aceleración del entrenamiento. Una metodología para establecer la estructura adecuada de la RNA. Se ha añadido un criterio de fiabilidad a sus respuestas basado en intervalos de confianza, definidos en base a la información de los datos de entrenamiento y la información (pesos y conexiones) de la red neuronal una vez entrenada.
- Mejoras en el problema del pronóstico de fallos en los sistemas industriales, calculando la vida útil remanente de los componentes/equipos, sin ninguna información sobre la evolución del desgaste/degradación, o del fallo. Y añadiendo métricas e indicadores de confianza, dando un criterio de fiabilidad a las respuestas del modelo neuronal.
- La predicción del comportamiento de los incendios forestales, donde se han creado modelos de redes neuronales que pueden predecir algunos parámetros importantes que ayudan a tomar decisiones en casos reales.
- La estimación de la demanda de energía eléctrica en situaciones de incertidumbre económica.

En el primer eje, se ha mejorado el uso de las redes neuronales definiendo algunos pasos iniciales en la construcción de la red neuronal, esenciales para garantizar su buen uso, como lo es la inicialización de los pesos de la red neuronal. Esta cuestión ha sido resuelta reemplazando

5. CONCLUSIONES Y LÍNEAS FUTURAS

el concepto de la inicialización aleatoria con un enfoque determinista que garantiza buenos resultados y sobre todo la convergencia del entrenamiento. La eficiencia de la matriz de Hilbert propuesta para la inicialización de los pesos, ha quedado demostrada tras la experimentación realizada, a pesar de ser una matriz mal condicionada, pero que satisface las condiciones de convergencia.

Por otra parte, el empleo de la matriz de Hilbert no garantiza que sea la mejor forma de inicializar los pesos, pero se ha comprobado que es una buena forma para su inicialización. Esta matriz permite la comparación de los resultados que se obtienen al cambiar parámetros, como se suele hacer, cambiando el moméntum, la tasa de aprendizaje o cambiando el número de neuronas ocultas, sabiendo que siempre parte de la misma inicialización.

También se ha planteado acelerar el entrenamiento mediante el suavizado de las funciones de activación. En este trabajo, solo se han estudiado las funciones de activación sigmoideal y tangente hiperbólica, debido a la recomendación de su uso para problemas no lineales. La implementación del parámetro α para suavizar la sigmoideal o la tangente hiperbólica, ha permitido acelerar el entrenamiento, como lo justifican los resultados experimentales obtenidos.

Se han implementado dos metodologías. La primera sirve para encontrar el número de neuronas y capas ocultas. La segunda permite calcular dos criterios de fiabilidad. El primer criterio se basa en el intervalo de confianza que tiene en cuenta la información disponible dentro del conjunto de entrenamiento (en general la densidad), y el segundo criterio se basa en los valores de los pesos de la red neuronal una vez entrenada. Esta información es relevante, ya que permite tener un criterio de fiabilidad de la respuesta de la red neuronal entrenada, que es de gran utilidad en la ayuda a la toma de decisiones.

Lo demás ejes, representan mejoras en la aplicación de los resultados obtenidos para resolver diversos problemas, que se centran en el ámbito de la ingeniería de fiabilidad y el análisis de riesgos. El primer problema aborda la predicción de la vida remanente. El problema elegido para su estudio ha sido el PHM08 (ver 4.2), su dificultad radica en la ausencia de información acerca de los componentes así como del tipo de fallo o degradación. Lo que ha conducido a la propuesta de dos soluciones con enfoques distintos utilizando las redes neuronales, que han permitido comparar hasta qué punto la red neuronal pudo obtener la información disponible en el conjunto de entrenamiento. Los dos enfoques, han mostrado casi las mismas conclusiones. Por otra parte, se han añadido también métricas para la evaluación de la fiabilidad de la respuesta del modelo neuronal, y se ha llegado a la conclusión de que el hecho de combinar

las métricas aporta más información para la toma de decisiones, una conclusión derivada del resultado obtenido.

El segundo problema trata de un problema del ámbito del riesgo medio ambiental. Se ha analizado el problema de la predicción de los incendios forestales. Para este propósito, se han utilizado las redes neuronales artificiales, presentando como contribución original que, por primera vez, en este trabajo se utilizan las RNA para la predicción de parámetros de su comportamiento. El modelo neuronal creado ha sido comparado con otros modelos bien conocidos. Un avance que se ha podido obtener de la red neuronal desarrollada, ha sido la posibilidad de crear modelos con solo algunos parámetros de entrada. Este aspecto es primordial cuando no se dispone de algún parámetro o cuando está mal medido. Otro avance es el tiempo requerido para su uso, son instantáneas, y se pueden usar en tiempo real durante la evolución del incendio.

Otra aplicación ha sido estudiar la estimación de la demanda de energía eléctrica. Dadas las incertidumbres actuales económicas y sociales, existe la necesidad de disponer de modelos más precisos. Se ha realizado una comparativa de la eficiencia de la estimación entre métodos tradicionales y de inteligencia artificial. Los resultados experimentales confirman la utilidad de las RNA, su alta precisión en las estimaciones futuras y la mayor fiabilidad aportada por el intervalo de confianza.

Se puede deducir del trabajo realizado la utilidad y la gran eficiencia de utilizar las redes neuronales artificiales para resolver determinados problemas complejos de confiabilidad y riesgo. Quedan muchos posibles trabajos futuros para mejorar el uso de las redes neuronales. Entre ellos, aplicar y mejorar los resultados obtenidos en otros problemas emergentes de la I-RAMS, y buscar alternativas para la mejora del entrenamiento y la optimización de la estructura y arquitectura de las RNAs para ser usadas en tiempo real. Profundizar con mayor grado de detalle en los nuevos algoritmos de aprendizaje, y analizar el potencial de las nuevas técnicas de optimización para diseño óptimo de modelos neuronales.

Esta tesis ha propuesto una forma sencilla la inicialización de los pesos. Sin embargo la tarea de encontrar la estructura optimizada no está resuelta, el enfoque empleado en esta tesis, ha sido comprobar distintas estructuras, cambiando de numero de neuronas ocultas. En la actualidad se está trabajando en crear un modelo neuronal robusto. En el sentido de que durante el entrenamiento, se van añadiendo neuronas ocultas o se van eliminando, dependiendo de la saturación en dichas neuronas.

5. CONCLUSIONES Y LÍNEAS FUTURAS

No obstante, En cuanto al futuro más inmediato se pretende aumentar las capacidades del modelo ANNFBP, intentando adquirir más datos históricos de otros incendios forestales, a así incorporar más salidas. Por otra parte, se pretende incorporar el parámetro del tiempo al modelo neuronal, para obtener un modelo predictivo a tiempo real.

Respecto a I-RAMS, existen muchos retos a estudiar y resolver. Se pretende continuar la búsqueda de modelos neuronales para estos retos pero de una manera más sencilla, fiable y con menos requerimiento de tiempo o de procesamiento.

- [1] ABBAS B.; OVE T. G.; JAVAD B. **RAMS data collection under Arctic conditions.** *Reliability Engineering & System Safety*, 2015. [2](#)
- [2] ZHAOGUANG P.; ALICE A. M.; TINGDI Z.; CHRISTOPHER W. J. YU L. **How reliable is satellite navigation for aviation? Checking availability properties with probabilistic verification.** *Reliability Engineering & System Safety*, Vol.144, pp. 95-116, 2015. [2](#)
- [3] S. MARTORELL S.; VILLAMIZAR M.; MARTÓN I.; VILLANUEVA J. F.; CARLOS S.; SÁNCHEZ A. I. **Evaluation of risk impact of changes to surveillance requirements addressing model and parameter uncertainties.** *Reliability Engineering & System Safety*, Vol. 126, pp. 153-165, 2014. [2](#)
- [4] JI-MIN L.; XIAO-YUE W.; YILIU L.; MARY A. L. **Reliability analysis of large phased-mission systems with repairable components based on success-state sampling.** *Reliability Engineering & System Safety*, Vol. 142, pp. 123-133, 2015. [2](#)
- [5] NICHOLAS J. BAHR. **System Safety Engineering and Risk Assessment: A Practical Approach, Second Edition.** Editor CRC Press, ISBN 1466551607, 2014. [3](#)
- [6] XIAOMIN X.; SHENG H.; YAOLIANG C.; BROWNY K.; HALILOVICY I.; WEI L. **TSAAA: Time Series Analytics as a Service on IoT,”** in *Web Services (ICWS)*. 2014 *IEEE International Conference on*, pp. 249-256, 2014. [3](#)
- [7] ALBINI F. A. **Estimating wildfire behavior and effects.** *Gen. Tech. Rep. INT-GTR-30*. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station, 92 p, 1976. [4](#), [123](#)
- [8] FINNEY M.A.; GRENFELL I.C.; MCHUGH C.W.; SELI R.C.; TRETHEWEY D.; STRATTON R.D.; BRITTAIN S. **A method for ensemble wildland fire simulation.** *Environmental Modeling and Assessment*, Vol. 16, pp. 153–167, 2011. [4](#), [123](#)
- [9] LAUTENBERGER C. **Wildland fire modeling with an Eulerian level set method and automated calibration.** *Fire Safety Journal* 62, Part C, 289-298, ISSN 0379-7112, 2013. [4](#), [123](#)

BIBLIOGRAFÍA

- [10] ROCHOUX M.C.; RICCI S.; LUCOR D.; CUENOT B.; TROUVÉ A. **Towards predictive data-driven simulations of wildfire spread & ndash; Part I: Reduced-cost Ensemble Kalman Filter based on a Polynomial Chaos surrogate model for parameter estimation.** *Natural Hazards and Earth System Sciences Discussions* 2, doi: 10.5194/nhessd-2-3289-2014, pp. 3289-3349, 2014. [4](#), [123](#)
- [11] ROCHOUX M.C.; RICCI S.; LUCOR D.; CUENOT B.; TROUVÉ A. **Towards predictive data-driven simulations of wildfire spread & ndash; Part 2: Ensemble Kalman Filter for the state estimation of a front-tracking simulator of wildfire spread.** *Natural Hazards and Earth System Sciences Discussions* 2, doi: 10.5194/nhessd-2-3769-2014, pp. 3769-3820, 2014. [4](#), [123](#)
- [12] LOLAS S.; OLATUNBOSUN O.A. **Prediction of vehicle reliability performance using artificial neural networks.** *Expert Systems with Applications*, Vol. 34 (4), pp. 2360–2369, 2008. [5](#)
- [13] K.; XIE M.; TANG L. C. XU. **Application of neural networks in forecasting engine systems reliability.** *Application of Soft Computation Journal*, Vol. 2 (4), pp. 255–268, 2003. [5](#)
- [14] YI-HUI L. **Evolutionary neural network modeling for forecasting the field failure data of repairable systems.** *Expert Systems with Applications*, Vol. 33 (4), pp.1090–1096, 2007. [5](#)
- [15] WERBOS P. **Beyond regression: New tools for prediction and analysis in the behavioral sciences.** *Ph.D. dissertation, Committee on Appl. Math., Harvard Univ, Cambridge, MA.,* 1974. [6](#), [13](#)
- [16] PARKER D. **Learning logic.** *Technical Report TR-87, Cambridge, MA: Center for Computational Research in Economics and Management Science, MIT,* 1985. [6](#), [41](#)
- [17] RUMELHART D. E.; HINTON G. E.; WILLIAMS R. J. **Learning internal representations by error propagation.** *Parallel Distributed Processing: Exploration in the Microstructure of Cognition*, Vol. 1, MIT Press, Cambridge, MA, 1986. [6](#), [41](#), [42](#)
- [18] HUAGUANG Z.; ZHANSHAN W.; DERONG L. **A Comprehensive Review of Stability Analysis of Continuous-Time Recurrent Neural Networks.** *Neural Networks and Learning Systems, IEEE Transactions on* , Vol. 25 (7), pp.1229-1262, 2014. [6](#)
- [19] KIAKOJOORI S.; KHORASANI K. **Dynamie neural networks for jet engine degradation prediction and prognosis.** *Neural Networks (IJCNN), 2014 International Joint Conference on*, pp.2531-2538, 2014. [6](#)
- [20] CABESSA J.; VILLA A.E.P. **Computational capabilities of recurrent neural networks based on their attractor dynamics.** *Neural Networks (IJCNN), 2015 International Joint Conference on*, pp.1-8, 2015. [6](#)
- [21] POPA C.-A. **Lie algebra-valued neural networks.** *Neural Networks (IJCNN), 2015 International Joint Conference on*, pp.1-6, 2015. [6](#)
- [22] HORNIK K.; STINCHCOMBE M.; WHITE H. **Multilayer feedforward networks are universal approximators.** *Neural Networks*. Vol. 2, pp. 359-366, 1993. [6](#), [27](#)

-
- [23] BISHOP C. M. **Neural Networks for Pattern Recognition**. *Oxford University Press*, 1995. [6](#), [35](#), [68](#), [110](#), [144](#)
- [24] MINSKY M.; PAPERT S. **Perceptrons**. *MIT Press, Cambridge, MA*, 1969. [13](#), [23](#), [41](#)
- [25] CARPENTER G. A.; STEPHEN G. **The ART of Adaptive Pattern Recognition by a Self-Organizing Neural Network**. *Special Issue on Artificial Neural Systems. Vol. 21, pp. 77-88*, 1988. [13](#)
- [26] KUNIIHIKO F. **A Neural Network for Visual Pattern Recognition**. *IEEE Computer, Vol.21, pp. 65-75*, 1988. [13](#)
- [27] CHRISTOS S.; DIMITRIOS S. **Neural Networks**. *Computer Science Deptt. University of U.K.; Journal, Vol. 4*, 1996. [13](#)
- [28] ROBERT SCHALKOFF J. **Artificial Neural Networks**. *McGraw-Hill International Editions*, 1997. [13](#)
- [29] JORDAN POLLACK B. **Connectionism: Past, Present and Future**. *Computer and Information Science Department, The Ohio State University*, 1998. [13](#)
- [30] FRANK R. **The Perceptron, A Perceiving and Recognizing Automaton**. *Project Para Report No. 85-460-1, Cornell Aeronautical Laboratory (CAL)*, 1957. [23](#)
- [31] HASSOUN M. H.; SONG J. **Adaptive Ho-Kashyap Rules for Perceptron Training, Vol. 3, N. 1**. *IEEE Transactions on Neural Networks*, 1992. [26](#)
- [32] NERRAND O.; ROUSSEL-RAGOT P.; PERSONNAZ L.; DREYFUS G. **Neural Networks and Non-linear Adaptive Filtering: Unifying Concepts and New Algorithms**. *Neural Computation Vol. 5, pp. 165-197*, 1993. [27](#)
- [33] BARRON A. R. **Universal Approximation Bounds for Superpositions of a Sigmoidal Function**. *IEEE Transactions on Information Theory, Vol.39, N. 3*, 1993. [28](#)
- [34] KALMAN B.L.; KWASNY S.C. **Why tanh: choosing a sigmoidal function**. *Neural Networks, 1992. IJCNN., International Joint Conference on, Vol.4, pp. 578-581*, 1992. [31](#)
- [35] MCCULLOCH W. S.; PITTS W. **A logical calculus of the ideas immanent in nervous activity**. *Bulletin of Mathematical Biophysics, Vol. 5, pp. 115-133*, 1943. [33](#)
- [36] STOPPIGLIA H. **Méthodes statistiques de sélection de modèles neuronaux; applications financières et bancaires**. *Thèse de Doctorat de l'Université Pierre et Marie Curie, Paris*, 1997. [41](#)
- [37] LECUN Y. **Learning processes in an asymmetric threshold network**. *Disordered Systems and Biological Organization, Berlin: Springer Verlag, pp. 233-240*, 1986. [41](#)
- [38] YU X. H.; CHEN G. A.; CHENG S. X. **Dynamic learning rate optimization of the back-propagation algorithm**. *IEEE Transactions on Neural Networks, Vol. 6 (3), pp. 669-77*, 1995. [42](#)

BIBLIOGRAFÍA

- [39] RIEDMILLER M.; BRAUN H. **A Direct Adaptive Method for Faster Backpropagation Learning: The RPROP Algorithm.** *IEEE International Conference on Neural Networks*, 1993. [47](#)
- [40] SCHIFFMANN W.; JOOST M.; WERNER R. **Optimization of the backpropagation algorithm for training multilayer perceptrons.** *Technical report, University of Koblenz, Institute of Physics*, 1993. [47](#)
- [41] JACOBS R. **Increased rates of convergence through learning rate adaptation.** *Neural Networks, Vol. 1 (4)*, 1988. [47](#)
- [42] TOLLENAERE T. **Supersab: Fast adaptive backpropagation with good scaling properties.** *Neural Networks, Vol. 3 (5)*, 1990. [47](#)
- [43] MARTIN R.; HEINRICH B. **A direct adaptive method for faster backpropagation learning: the RPROP algorithm.** *International Symposium on Neural Networks*, 1993. [47](#)
- [44] IGEL C.; HUSKEN M. **Improving the Rprop Learning Algorithm.** *ICSC Symposium on Neural Computation*, 2000. [47](#)
- [45] POWELL M. J. D. **Radial basis functions for multivariable interpolation: A review.** *IMA conference on "Algorithms for the Approximation of Functions and Data", RMCS Sllfivenha*, 1985. [51](#)
- [46] YAO S.J.; SONG Y. H.; ZHANG L. Z.; CHENG X. Y. **Wavelet transform and neural networks for short-term electrical load forecasting.** *Energy Conversion and Management, Vol. 41 (18), pp. 1975-1988*, 2000. [51](#)
- [47] OUSSAR Y.; DREYFUS G. **Initialization by selection for wavelet network training.** *Neurocomputing, Vol. 34, pp 131-143*, 2002. [51](#)
- [48] GEMAN S.; BENENSTOCK E.; DOURSAT R. **Neural networks and the bias/variance dilemma.** *Neural Computation, Vol.4, pp. 1-58*, 1992. [52](#)
- [49] HASSABIS D.; CHU C.; REES G.; WEISKOPF N.; MOLYNEUX P. D.; MAGUIRE E. A. **Decoding Neuronal Ensembles in the Human Hippocampus.** *Current Biology, Vol. 19 (7), pp. 546-554*, 2009. [53](#)
- [50] STONE M. **Cross-validatory choice and assessment of statistical predictions.** *Journal of the Royal Statistical Society, B 36, pp. 111-147*, 1974. [54](#)
- [51] KULLBACK S. **Information theory and statistics.** *John Wiley and Sons, NY*, 1959. [54](#)
- [52] VAPNIK V. **The Nature of Statistical Learning Theory.** *Springer*, 1995. [56](#), [67](#)
- [53] CICERO A. M.; MAURO R.; GUENTHER S. N. **A method to estimate prediction intervals for artificial neural networks that is sensitive to the noise distribution in the output.** *Proceedings of International Joint Conference on Neural Networks, Atlanta, Georgia, USA*, 2009. [56](#)

-
- [54] CHRYSSOLOURIS G.; LEE M.; RAMSEY A. **Confidence interval prediction for neural network models.** *IEEE Transaction on Neural Networks*, pp.229-232, 1996. [56](#), [57](#), [65](#)
 - [55] GEORGIOS P.; PETER J. E.; ALAN F. M. **Confidence Estimation Methods for Neural Networks: A Practical Comparison.** *IEEE Transaction on Neural Networks*, Vol.12, 2001. [56](#)
 - [56] SHAO R.; MARTIN E. B.; ZHANG J.; MORRIS A. J. **Confidence Bounds for Neural Network Representations.** *Elsevier Science Ltd. Vol. 21*, pp.1173-1178, 1997. [56](#), [65](#)
 - [57] LEONARD J. A.; KRAMER M. A.; UNGAR L. H. **Using radial basis functions to approximate a function and its error bounds.** *IEEE Transaction on Neural Networks*, Vol. 3, pp. 624-627, 1992. [57](#)
 - [58] FIX E.; HODGES J. L. **Discriminatory analysis Nonparametric discrimination consistency properties.** *Report Number 4, Project Number 21-49-004, USAF School of Aviation Medicine, Randolph Field, Texas. Reprinted as pp. 261-279 of Agrawala, 1977, 1951.* [57](#)
 - [59] ROSENBLATT M. **Remarks on some nonparametric estimates of a density function.** *The Annals of Mathematical Statistics.*, Vol. 27, pp. 832-837, 1956. [57](#)
 - [60] PARZEN E. **On estimation of a probability density function and mode.** *Ann. of Math. Stat. Vol. 33*, pp. 1065-1076, 1962. [57](#)
 - [61] CENCOV N. N. **Evaluation of an Unknown Distribution Density from Observations.** *Soviet Mathematics*, Vol. 3, pp. 1559-1562, 1962. [57](#)
 - [62] WATSON G. S.; LEADBETTER M. R. **On the estimation of the probability density.** *I Ann Math. Statist. Vol. 34*, pp. 480-491, 1963. [57](#)
 - [63] STUART C. S. **Estimation of Probability Density by an Orthogonal Series.** *Ann. Math. Statist. Vol. 38 (4)*, pp. 977-1322, 1967. [57](#)
 - [64] EPANECHNIKOV V. A. **Nonparametric estimation of a multidimensional probability density.** *Teor. Veroyatnost. i Primenen.*, Vol. 14 (1), pp. 156-161, 1969. [57](#)
 - [65] MICHAEL T.; RICHARD K. **On Multivariate Density Estimates Based on Orthogonal Expansions.** *Ann. Math. Statist. Vol. 41 (2)*, pp. 341-764, 1970. [57](#)
 - [66] GRACE W. **A Polynomial Algorithm for Density Estimation.** *Ann. Math. Statist. Vol. 42 (6)*, pp. 1815-2213, 1971. [57](#)
 - [67] CACOULOS T. **Estimation of a multivariate density.** *Ann Inst Statist. Math. Vol.18*, pp. 178-189, 1967. [57](#)
 - [68] SILVERMAN B. W. **Density Estimation for Statistic and data Analysis.** *Chapman and Hall London Great Britain*, 1986. [57](#), [58](#), [61](#)
 - [69] SAIN S. R. **Adaptive Kernel Estimation.** *Phd thesis , rice University.*, 1994. [57](#), [58](#)
 - [70] WALTER G.; BLUM J. **Probability Density Estimation Using Delta Sequences.** *Ann. Statist. Vol. 7 (2)*, pp. 237-466, 1979. [57](#)

BIBLIOGRAFÍA

- [71] DAVID W. S. **Multivariate Density Estimation. Theory, Practice and Visualization.** *New York: Wiley.*, 1992. [57](#)
- [72] JONES M. P. **Kernel Smoothing.** *Champan and Hall, London , Great Britain*, 1995. [58](#)
- [73] MEER D. C. **Distribution free decomposition of Multivariate data.** *Workshop on Statistical in Pattern Recognition*, 1998. [58](#)
- [74] FUKUNAGA K. **Introduction to Statistical Pattern Recognition.** *New Yrok Academic*, 1972. [60](#)
- [75] SEBER G. A. F.; WILD C. J. **Nonlinear Regression.** *Hoboken, New Jersey: John Wiley & Sons*, 2003. [62](#), [63](#)
- [76] BURROWS N.D. **Experimental development of a fire management model for Jarrah (*Eucalyptus marginata*) forest [Ph.D. thesis].** *Department of Forestry, Australian National University, Canberra, Australia*, 1994. [70](#)
- [77] SHARPLES J. J.; MCRAE R. H. D.; WEBER R. O.; GILL A. M. **A simple index for assessing fuel moisture content.** *Environmental Modelling & Software*, Vol. 24, pp. 637-646, 2009. [73](#)
- [78] SKVARENINA J.; MINDAS J.; HOLECY J.; TUCEK J. **Analysis of the natural and meteorological conditions during two largest forest fire events in the Slovak Paradise National Park.** *Forest Fire in the Wildland-Urban Interface and Rural Areas in Europe: An Integral Planning and Management Challenge*, *Institute of Mediterranean Forest Ecosystems and Forest Products Technology, Athens, Greece*, 2003. [73](#)
- [79] LUKE R. H.; MCARTHUR A. G. **Bushfires in Australia.** *Australian Government Publishing Service, Canberra*, 1978. [73](#)
- [80] THORNTHWAITTE C. W. **An Approach toward a Rational Classification of Climate.** *Geographical Review*, vol 38, 1948. [74](#)
- [81] MARTELL R. F.; HEILMAN M. E.; BLOCK C. J.; SIMON M. **Has anything changed? Current characterizations of men, women and managers.** *Journal of Applied Psychology*, Vol. 74, pp. 935-942, 1989. [75](#)
- [82] TODD J. B.; KOURTZ P. H. **Prédiction quotidienne des incendies de forêts causés par négligence.** *Forêts Canada, Institut forestier national de Petawawa, Chalk River (Ontario). Rapport d'information PI-X-103F. 20 p*, 1992. [75](#)
- [83] LAWSON A. J.; TURNER M. J. L. **GINGA observations of the X-ray spectra of quasars.** *Monthly Notices of the Royal Astronomical Society*, Vol. 288 (4), pp. 920-944, 1997. [75](#)
- [84] HAWKES B.; BECK J. **A wildfire threat rating system.** *Fire Manage, Notes 59*, pp. 25-33, 1999. [75](#)
- [85] VAN WAGNER C. E.; PICKETT T. L. **Equation and FORTRAN program for the Canadian Forest Fire Weather Index System.** *Canadian Forest Service Forestry Technical Report 33*, 1985. [76](#), [77](#)

-
- [86] NWCG FIRE DANGER WORKING TEAM. **Gaining a Basic Understanding of the National Fire Danger Rating System A Self-Study Reading Course.** *National Wildfire Coordinating Group*, 2002. [78](#)
 - [87] KOURTZ P. **Artificial intelligence: a new tool for forest management.** *Canadian Journal of Forest Research-revue Canadienne de Recherche Forestiere*, Vol. 20 (4), pp.428-437, 1990. [80](#)
 - [88] AMPARO A.; OSCAR F.; BERTHA G.; ELENA H.; MARIA INMACULADA P. A.; EULOGIO J.; JOSE LUIS L. S.; TARSY C. **An intelligent system for forest fire risk prediction and fire fighting management in Galicia.** *Expert Systems With Applications*, Vol. 25 (4), pp. 545-554, 2002. [80](#)
 - [89] SAKR G. E.; ELHAJJ I. H.; MITRI G. **Efficient forest fire occurrence prediction for developing countries using two weather parameters.** *Engineering Applications of Artificial Intelligence*, Vol. 24 (5), pp. 888–894, 2011. [80](#), [81](#)
 - [90] ÇETIN E.; YUSUF S. **A data fusion framework with novel hybrid algorithm for multi-agent Decision Support System for Forest Fire.** *Expert Syst. Appl.* Vol.38 (8), pp. 9225-9236, 2011. [81](#)
 - [91] LI Z.; KHANANIAN A.; FRASER R. H.; CIHLAR J. **Automatic detection of fire smoke using artificial neural networks and threshold approaches applied to AVHRR imagery.** *IEEE Transactions on Geoscience and Remote Sensing* Vol. 39 (9), pp. 1859-1870, 2001. [81](#)
 - [92] ARMANDO M. F.; ANDREI B. U.; ALEXANDER V. L.; RUI M. V. **Development of neural network committee machines for automatic forest fire detection using lidar.** *Pattern Recognition*, Vol. 37 (10), pp. 2039-2047, 2004. [81](#)
 - [93] WU J.; RULE H.; XU Z.; HAN N. **Forest fire smog feature extraction based on Pulse-Coupled neural network.** *Information Technology and Artificial Intelligence Conference (ITAIC), 2011 6th IEEE Joint International*, Vol. 1, pp. 186-189, 2011. [81](#)
 - [94] CHENG T.; WANG J. **Integrated Spatio-temporal Data Mining for Forest Fire Prediction.** *Transactions in Gis*, Vol. 12 (5), pp. 591-611, 2008. [81](#)
 - [95] CORTEZ P.; MORAIS A. **A Data Mining Approach to Predict Forest Fires using Meteorological Data.** In J. Neves, M. F. Santos, and J. Machado, editors, *Proc. EPIA 2007*, pp. 512–523, 2007. [81](#), [167](#)
 - [96] DENG YI Z.; SHIZHONG H.; JIANHUI Z.; ZHONG Z.; CHENGZHANG Q.; YOUWANG K.; XIANG C. **Image Based Forest Fire Detection Using Dynamic Characteristics with Artificial Neural Networks.** *Artificial Intelligence, 2009. JCAI '09 International Joint Conference on*, pp. 290-293, 2009. [81](#)
 - [97] CHENG C.; SUN F.; ZHOU X. **One Fire Detection Method Using Neural Networks.** *Tsinghua Science & Technology*. Vol. 16 (1), pp. 31-35, 2011. [81](#)
 - [98] ANGAYARKKANI K.; RADHAKRISHNAN N. **An Intelligent System For Effective Forest Fire Detection Using Spatial Data.** *Computing Research Repository*, abs/1002.2, 2010. [81](#)

BIBLIOGRAFÍA

- [99] DENG C.; WU L.; FAN W. **Neural network approach for RHR calculation and prediction in fire science.** *International Conference on Signal Processing Proceedings*, masid: 50063695, doi:10.1109/ICSIGP.1996.571152, 1996. [81](#)
- [100] VAKALIS D.; SARIMVEIS H.; KIRANOUDIS C. T.; ALEXANDRIDIS A.; BAFAS G. **A GIS based operational system for wildland fire crisis management I.** *Mathematical modelling and simulation, Applied Mathematical Modelling*, Vol. 28, pp. 389–410, 2004. [81](#)
- [101] VAKALIS D.; SARIMVEIS H.; KIRANOUDIS C. T.; ALEXANDRIDIS A.; BAFAS G. **A GIS based operational system for wildland fire crisis management II.** *Mathematical modelling and simulation, Applied Mathematical Modelling*, Vol. 28, pp. 389–410, 2004. [81](#)
- [102] OKAYAMA Y. **A primitive study of a fire detection method controlled by artificial neural net.** *Fire Safety Journal - FIRE SAFETY*, Vol. 17 (6), pp. 535–553, 1991. [81](#)
- [103] ANDREW K. S. J.; DAMING L.; DRAGAN B. **A review on machinery diagnostics and prognostics implementing condition-based maintenance.** *Mechanical Systems and Signal Processing*, Vol. 20 (7), pp. 1483-1510, 2006. [82](#)
- [104] MAYNARD K.; BYINGTON C. S.; NICKERSON G. W.; VAN DYKE M. **Validation of Helicopter Nominal and Faulted Conditions Using Fleet Data Sets.** *Proc. of the Intl. Conf. on Condition Monitoring, UK*, pp. 129-141, 1999. [82](#)
- [105] AIWINA H.; SHENG Z.; ANDY C. C.; JOSEPH M. **Rotating machinery prognostics: State of the art, challenges and opportunities.** *Mechanical Systems and Signal Processing*, Vol. 23 (3), pp. 724-739, 2009. [83](#), [86](#)
- [106] JIHONG Y.; MUAMMER K.; JAY L. **A prognostic algorithm for machine performance assessment and its application.** *Production Planning & Control*, Vol. 15 (8), pp. 796-801, 2004. [83](#)
- [107] VAN T. T.; BO-SUK Y. **Data-driven approach to machine condition prognosis using least square regression tree.** *Journal of Mechanical Science and Technology*, Vol. 23 (5), pp. 1468-1475, 2009. [83](#)
- [108] MING D.; DAVID H. **Hidden semi-Markov model-based methodology for multi-sensor equipment health diagnosis and prognosis.** *European Journal of Operational Research*, Vol. 178 (3), pp. 858-878, 2007. [83](#)
- [109] ALAA H. E.; NAGI Z. G. **Sensor-driven prognostic models for equipment replacement and spare parts inventory.** *IEEE Transactions*, Vol. 40 (7), pp. 629-639, 2008. [83](#)
- [110] SHOJI H.; TOSHIYUKI A.; SHENG Z. **Study of machine fault diagnosis system using neural networks.** *International Symposium on Neural Networks*, Vol. 1, pp. 956-961, 2002. [84](#)
- [111] HU J.; KUMAMARU K.; INOUE K.; HIRASAWA K. **KDI-Based Robust Fault Detection Scheme for Nearly Linear Systems.** *Preprints of 29th ISCIE International Symposium on Stochastic Theory and Its Application, Tokyo*, pp. 3-4, 1997. [85](#)

-
- [112] NAKAMURA M.; KATAFUCHI T.; YOSHIDA A.; HATAZAKI H. **Optimum Decision Making of Maintenance Scheduling of Boilers in Thermal Power Stations with Few Data Set.** *Preprints of 29th ISCIE International Symposium on Stochastic Theory and Its Application, Tokyo, pp.7-8, 1997.* [85](#)
- [113] EDWIN L.; KLEMENT E. P.; LUJAN J. M.; GUARDIOLA C. **Model-based fault detection in multi-sensor measurement systems.** *Intelligent Systems, IS, International IEEE Conference, 2004.* [85](#)
- [114] JIAH SHING C.; SARGUR N. S. **A probabilistic theory of model-based diagnosis.** *International Journal of Human-computer Studies / International Journal of Man-machine Studies, Vol. 40 (6), pp. 933-963, 1994.* [85](#)
- [115] CHANTLER M. J.; WYATT T. K.; ALDEA A.; MCGUNNIGLE G. **A taxonomy of user requirements for model-based diagnostic systems.** *In: Proc. 14th IFAC World Congress. Beijing, 1999.* [85](#)
- [116] CARL S. B.; MICHAEL I. R.; THOMAS G. **A probabilistic theory of model-based diagnosis.** *Pain Medicine, 2002.* [86](#), [88](#)
- [117] ISO 13381-1. **Condition monitoring and diagnostics of machines -prognostics - Part1: General guidelines.** *Int. Standard, ISO, 2004.* [86](#)
- [118] AIWINA SOONG Y. H. **Intelligent prognostics of machinery health utilising suspended condition monitoring data.** *Queensland University of Technology, 2009.* [86](#)
- [119] ANDREW K. S. J.; DAMING L.; DRAGAN B. **A review on machinery diagnostics and prognostics implementing condition-based maintenance.** *Mechanical Systems and Signal Processing, Vol. 20 (7), pp. 1483-1510, 2006.* [86](#)
- [120] VACHTSEVANOS G.; LEWIS F. L.; ROEMER M.; HESS A. **Intelligent Fault Diagnosis and Prognosis for Engineering Systems.** *New Jersey Hoboken: Wiley & Sons, 2006.* [86](#)
- [121] EL-KOUJOK M.; GOURIVEAU R.; ZERHOUNI N. **Reducing arbitrary choices in model building for prognostics: An approach by applying parsimony principle on an evolving neuro-fuzzy system.** *Microelectronics Reliability, Vol. 51 (2), pp. 310-320, 2011.* [88](#)
- [122] DAS S.; HALL R.; HERZOG S.; HARRISON G.; BODKIN M. **Essential steps in prognostic health management.** *International Conference on Prognostics and Health Management, 2011.* [88](#)
- [123] THURSTON M.; LEBOLD M. **Open Standards for Condition-Based Maintenance and Prognostic Systems.** *Reliability Engineering and System Safety, Vol.77, pp. 151-166, 2001.* [88](#)
- [124] MULLER A.; SUHNER M.; IUNG B. **Formalisation of a new prognosis model for supporting proactive maintenance implementation on industrial system.** *Reliability Engineering & System Safety, Vol. 93 (2), pp. 234-253, 2008.* [88](#)
- [125] TOBON-MEJIA D. A.; MEDJAHED K.; ZERHOUNI N. **The ISO 13381-1 standard's failure prognostics process through an example.** *Prognostics and Health Management Conference, 2010. PHM '10., pp. 1-12, 2010.* [88](#)

BIBLIOGRAFÍA

- [126] PETER J. F. L. **Bayesian model-based diagnosis.** *International Journal of Approximate Reasoning*. Vol. 27 (2), pp. 99-119, 2001. [xvii](#), [89](#)
- [127] BYINGTON C.S.; WATSON M.; ROEMER M. J.; GALIC T. R.; MCGROARTY J. J. **Prognostic enhancements to gas turbine diagnostic systems.** *Aerospace Conference. Proceedings. IEEE*, Vol. 7, pp. 3247-3255, 2003. [89](#)
- [128] VAN T. T.; BO SUK Y.; ANDY C. C. T. **Multi-step ahead direct prediction for the machine condition prognosis using regression trees and neuro-fuzzy systems.** *Expert Systems With Applications*, Vol. 36 (5), pp. 9378-9387, 2009. [89](#)
- [129] VAN T. T.; BO-SUK Y.; MYUNG-SUCK O.; ANDY C. C. T. **Machine condition prognosis based on regression trees and one-step-ahead prediction.** *Mechanical Systems and Signal Processing*, Vol. 22 (5), pp. 1179-1193, 2008. [89](#)
- [130] SCHWABACHER M. A. **A Survey of Data Driven Prognostics.** *Proceedings of the AIAA Infotech@Aerospace Conference*. Reston, VA: American Institute for Aeronautics and Astronautics, Inc., 2005. [89](#)
- [131] LIN D.; MAKIS V. **Recursive filters for a partially observable system subject to random failure.** *Advances in Applied Probability*, Vol. 35, pp. 207-227, 2003. [90](#)
- [132] FARRAR C. R.; HEMEZ F.; PARK G.; ROBERTSON A. N.; SOHN H.; WILLIAMS T. O. **A coupled approach to developing damage prognosis solutions, in: Damage Assessment of Structures.** *The Fifth International Conference on Damage Assessment of Structures (DAMAS 2003)*, Southampton, UK, 2003. [90](#)
- [133] GOEBEL K.; SABA B.; SAXENA A. **A comparison of three data-driven techniques for prognostics.** *Failure prevention for system availability, 62th meeting of the MFPT Society*, pp. 119-131, 2008. [90](#), [91](#)
- [134] TANG L. C.; LU Y.; CHEW E. P. **Mean residual life of lifetime distributions.** *IEEE Trans. Reliability*, Vol. 48 (1), pp. 73-78, 1999. [91](#)
- [135] LIM J.; PARK D. H. **Trend change in mean residual life.** *IEEE Trans. Reliability*, Vol. 44 (2), pp. 291-296, 1995. [91](#)
- [136] SIDDIQUI M. M.; CAGLAR M. **Residual Lifetime Distribution and Its Applications.** *Microelectronics Reliability*, Vol. 34 (2), pp. 211-227, 1994. [91](#)
- [137] BRADLEY D. M.; GUPTA R. C. **Limiting behavior of the mean residual life.** *Ann. Inst. Statist. Math.*, Vol. 55 (1), pp. 217-226, 2003. [91](#)
- [138] YOGESH G. B.; IBRAHIM Z.; SAGAR V. K. **A Framework for Prognostics and Health Management of Electronic Systems.** *IEEE Aerospace Conference*, 2008. [91](#)
- [139] GEBRAEEL N.; LAWLEY M.; LIU R.; PARMESHWARAN V. **Residual life predictions from vibration-based degradation signals: a neural network approach.** *Industrial Electronics, IEEE Transactions on*, Vol. 51 (3), pp. 694-700, 2004. [91](#), [93](#)

-
- [140] GEBRAEEL N.Z.; LAWLEY M.A. **A neural network degradation model for computing and updating residual life distributions.** *IEEE Transactions on Automation Science and Engineering*, Vol. 5 (1), pp. 154-163, 2008. [91](#), [93](#)
- [141] SHAO Y.; NEZU K. **Prognosis of remaining bearing life using neural networks.** *Proceedings of the Institution of Mechanical Engineers. Part I: Journal of Systems and Control Engineering*. Vol. 214 (3), pp. 217-230, 2000. [91](#)
- [142] RUNQING H.; LIFENG X.; XINGLIN L.; RICHARD L. C.; HAI Q.; JAY L. **Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods.** *Mechanical Systems and Signal Processing*, Vol. 21 (1), pp. 193-207, 2007. [93](#)
- [143] ESSAWY M. A. **Methods to Estimate Machine Remaining Useful Life Using Artificial Neural Networks.** *Defense Technical Information Center*, 2001. [94](#)
- [144] PINDYCK R. C. **The structure of world energy demand.** *The MIT Press, Cambridge, Massachusetts*, 1979. [96](#)
- [145] MARTINEZ M.; A. PULIDO A.; ROMERO J.; QUINTANA J.C.; CRUZ J. **Process of a Territorial Energy Plan in an Isle.** *En International Conference on Renewable Energies and Power Quality (ICREPQ'10), Granada, Spain*, 2010. [98](#)
- [146] HARRIS J. L.; LIU L. M. **Dynamic structural analysis and forecasting of residential electricity consumption.** *Int. J. Forecast.* Vol. 9, pp. 437-455, 2007. [99](#)
- [147] EDIGERA V.; AKARB S. **ARIMA forecasting of primary energy demand by fuel in Turkey.** *Energy Policy*, Vol. 35, pp. 1701-1708, 2007. [99](#)
- [148] BIANCOA V.; MANCA O.; NARDINIA S. **Electricity consumption forecasting in Italy using linear regression models.** *Energy*, Vol. 34, pp. 1413-1421, 2009. [99](#)
- [149] SHAKOURI H.; NADIMIA G.; GHADERIA F. **A hybrid TSK-FR model to study short-term variations of the electricity demand versus the temperature changes.** *Expert Syst. Appl.*, Vol. 36, pp. 1765-1772, 2008. [100](#)
- [150] PADMAKUMARI K.; MOHANDAS K. P.; THIRUVENGADAM S. **Long term distribution demand forecasting using neuro fuzzy computations.** *Int. J. Electr. Power Energy Syst.*, Vol. 21, pp. 315-322, 1999. [100](#)
- [151] BANSAL R. C.; PANDEY J. C. **Load forecasting using artificial intelligence techniques: A literature survey.** *Int. J. Comput. Appl. Technol.*, Vol. 22, pp. 109-119, 2005. [100](#)
- [152] HAHN H.; MEYER-NIEBERG S.; PICKL S. **Electric load forecasting methods: Tools for decision making.** *Eur. J. Oper. Res.*, Vol. 199, pp. 902-907, 2009. [100](#)
- [153] SUGANTHI L.; SAMUEL A. **Energy models for demand forecasting: A review.** *Renew. Sustain. Energy Rev.*, vol. 10, pp. 281-311, 2011. [100](#)

BIBLIOGRAFÍA

- [154] CRAIG P. P.; GADGIL A.; KOOMEY J. G. **What can history teach us? A retrospective examination of long-term energy forecasts for the United States.** *Ann. Rev. Ener Envi.*, Vol. 27, pp. 83-11, 2002. [100](#)
- [155] ARMSTRONG J. S. **Principles of Forecasting: A handbook for researchers and practitioners.** Norwell, MA: Kluwer Academic, 2001. [101](#)
- [156] KOOMEY J. G. **From my perspective: Avoiding “the Big Mistake” in forecasting technology adoption.** *Technological Forecasting and Social Change*, Vol. 69, pp. 511-18, 2002. [101](#)
- [157] URBAN F.; BENDERS R. J. M.; MOLL H. C. **Modelling energy systems for developing countries.** *Energy Policy*, Vol. 35, pp. 3473-3482, 2007. [101](#)
- [158] TIRADO HERRERO S.; JIMÉNEZ MENESES L.; LÓPEZ FERNÁNDEZ J. L.; MARTIN GARCIA J. **Pobreza energética en España. Análisis de tendencias.** *Asociación de Ciencias Ambientales*, Madrid, 2014. [101](#)
- [159] RODILLA VICENTE. **Gestión localizada de la demanda eléctrica asistida por ordenador.** *DYNA*, vol. 81, no. 1, p.31, 2006. [102](#)
- [160] BLOKHUIS E. G. J.; BROUWERS B. W.; VAN DER PUTTEN E. W. F. M.; SCHAEFER W. F. **Peak loads and network investments in sustainable energy transitions.** *Energy Policy*, Vol. 39 (10), pp. 6220-6233, 2011. [102](#)
- [161] BILLING S. A.; VOON W. S.. F. **Aprediction-error and stepwise-regression estimation algorithm for non-linear systems.** *International Journal ofControl*. Vol. 44, pp.803-822, 1986. [105](#)
- [162] JIE Z.; MORRIS A. J. **A Sequential Learning Approach for Single Hidden Layer Neural Networks.** *Neural Networks*. Vol. 1, pp. 65-80, 1998. [105](#)
- [163] BJORCK A. **Solving linear least squares problems by Gram-Schmidt orthogonalization.** *Nordisk Tidshrift for Information behadtung*, Vol. 7, pp. 1-21, 1967. [105](#)
- [164] CHNG S. C.; MULGREW B. **Efficient Computation Schemes for the Orthogonal Least Squares Algorithm.** *IEEE Transaction on Signal Processing*. Vol. 40, pp. 2904-2920, 1995. [105](#)
- [165] HIROAKI KITANO. **Neurogenetic learning: an integrated method of designing and training neural networks using genetic algorithms.** *Physica D-nonlinear Phenomena*. Vol. 75, pp. 225-238, 1994. [105](#)
- [166] XIUCHENG D.; SHOUCHUN W.; RENJIN S.; SUOQI Z. **Design of artificial neural networks using a genetic algorithm to predict saturates of vacuum gas oil.** *Petroleum Science*. Vol. 7, pp. 118-122, 2010. [105](#)
- [167] EBERHART R. C.; DOBBINS R. W. **Designing neural network explanation facilities using genetic algorithms.** *International Symposium on Neural Networks*, 1991. [105](#)

-
- [168] HAYKIN S. **Neural networks a comprehensive foundation.** 2nd edn. Prentice Hall, New York, 1998. [106](#)
 - [169] NGUYEN D.; WIDROW B. **Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights.** *International Symposium on Neural Networks*, 1990. [106](#)
 - [170] THIMM G.; FIESLER E. **Neural network initialization.** *From Natural to Artificial Neural Computation Lecture Notes in Computer Science*, Vol. 930, pp. 535-542, 1995. [107](#)
 - [171] HOPFIELD J. J. **Neurons with graded response have collective computational properties like those of two-state neurons.** *Proc. of the National Academy of Sciences*, Vol. 79, pp. 3088-3092, 1984. [110](#), [111](#)
 - [172] COHEN M.A.; GROSSBERG S. **Absolute stability of global pattern formation and parallel memory storage by competitive neural networks.** *IEEE Transactions on Systems, Man, and Cybernetics*, 1983. *SMC-13*(5), pp. 8-15, 1983. [110](#)
 - [173] KOSKO B. **Neural Networks and Fuzzy Systems— A Dynamical System Approach Machine Intelligence.** Prentice Hall, Englewood Cliffs, NJ, pp. 38-108, 1992. [110](#)
 - [174] KHALIL H. K. **Nonlinear Systems.** Prentice Hall, 3rd edition. 2002 . ISBN 0-13-067389-7, 2002. [111](#)
 - [175] HILBERT D. **Ein Beitrag zur Theorie des Legendre'schen Polynoms.** *Acta Mathematica*, (Springer Netherlands) Vol. 18, pp. 155-159, 1894. [111](#)
 - [176] HEIMES F. O. **Recurrent neural networks for remaining useful life estimation.** *Prognostics and Health Management. PHM 2008. International Conference on*, pp.1-6, 2008. [117](#)
 - [177] LETO P. **Data driven prognostics using a Kalman filter ensemble of neural network models.** *International Conference on Prognostics and Health Management*, 2008. [118](#), [144](#), [149](#)
 - [178] GIMBLETT R.H.; BALL G. L. **Neural network architectures for monitoring and simulating changes in forest resources management.** *AI Applications Vol. 9*, pp. 103-123, 1995. [122](#)
 - [179] JAYNE CHRISINA (EDS.) ILIADIS, LAZAROS. **Engineering Applications of Neural Networks.** *16th International Conference, EANN 2015, Rhodes, Greece, September 25-28*, 2015. [144](#)
 - [180] JIE L.; ABHINAV S.; KAI G.; BHASKAR S.; WILSON W. **An adaptive recurrent neural network for remaining useful life prediction of lithium-ion batteries.** *Annual conference of the Prognostics and Health Management Society*, 2010. [144](#)
 - [181] FELIX O. H.; BAE SYSTEMS. **Recurrent Neural Networks for Remaining Useful Life Estimation.** *2008 international conference on prognostics and health management*, 2008. [144](#)
 - [182] SAXENA A.; CELAYA J.; SAHA B.; SAHA S.; GOEBEL K. **Evaluating algorithm performance metrics tailored for prognostics.** *Aerospace conference, 2009 IEEE*, 2009. [150](#)

BIBLIOGRAFÍA

- [183] SAXENA A.; CELAYA J.; SAHA B.; SAHA S.; GOEBEL K. **Metrics for offline evaluation of prognostic performance.** *International Journal of Prognostics and Health Management*, 2010. [150](#)
- [184] KAMAL M.; DIEGO A. T.; NOUREDDINE Z. **Remaining Useful Life Estimation of Critical Components With Application to Bearings.** *IEEE Transactions on Reliability*, 2012. [150](#)
- [185] VERONICA R.C.; STUART A. J. A. **Validating the moisture codes of the FWI System in New Zealand pine plantation fuels.** *SCION , Report No. 44922*, 2009. [162](#), [164](#)
- [186] PEARCE H. G.; ANDERSON S. A. J. **A manual for predicting fire behaviour in New Zealand fuels.** *Scion Rural Fire Research Group, Christchurch, N.Z.*, 2008. [164](#)
- [187] FORESTRY CANADA FIRE DANGER GROUP. **Development and Structure of the Canadian Forest Fire Behavior Prediction System.** *Forestry Canada Information Report ST-X-3, Ottawa*, 1992. [164](#)
- [188] ALBERTA SUSTAINABLE RESOURCE DEVELOPMENT FOREST PROTECTION DIVISION. **Final Documentation Report Chisholm Fire (LWF-063).** *ISBN: 0-7785-1841-8 (on-line edition)*, 2001. [165](#), [166](#)
- [189] YOUSSEF S.; ABDELAZIZ B. **A Neural Network Approach for Predicting Forest Fires.** *Multimedia Computing and Systems (ICMCS), 2011 International Conference on*, pp.1-5, 2010. [167](#)
- [190] WINTER-ALTHAUS G.; GONZALEZ-LANDIN B.; PULIDO-ALONSO A.; GALVAN-GONZALEZ B.; MAAROUF M. **Forecasting Electricity Consumption Including Data Of The Present Economic And Financial Crisis. Application To Canary Islands.** *DYNA Energia y Sostenibilidad, Vol. 4 (1), pp. 1-13*, 2015. [173](#), [174](#), [176](#)