Predicción a corto plazo de la demanda de energía eléctrica con SICONEL

Fernando Pavón P., Manuel R. Arahal y Eduardo F. Camacho
Departamento de Ingeniería de Sistemas y Automática.
Escuela Superior de Ingenieros.
Camino de Los Descubrimientos s/n. 41092 Sevilla
arahal@esi.us.es

Resumen

En este artículo se muestra el desarrollo de un programa para la generación, explotación y mantenimiento de modelos de predicción a corto plazo del consumo de energía eléctrica. El programa, de nombre SICONEL, ha sido probado con los datos suministrados por la Compañía Sevillana de Electricidad. Los modelos que el programa utiliza son redes de neuronas acíclicas; es decir, no recurrentes. Se ha puesto especial interés en el problema de determinar el conjunto de variables de entrada del modelo. Los resultados obtenidos con diversos modelos generados por SICONEL son analizados.

Palabras clave: Predicción de la demanda eléctrica. Predicción a corto plazo. Redes de Neuronas. Selección de modelos.

1 Introducción

La predicción de la demanda de energía eléctrica tiene una gran importancia para la operación de las compañías productoras y distribuidoras. La reciente de-regularización del sector ha contribuido a aumentar dicha importancia. De hecho, la aparición del mercado nacional de energía obliga a las empresas a generar predicciones acertadas para participar en el mismo.

La predicción de la demanda suele llevarse a cabo mediante un experto humano ayudado por modelos lineales, neuronales, borrosos o de otro tipo [2]. En el proyecto SICONEL se atacó el problema de la construcción semi-automática de modelos para predicción. El programa resultante es capaz de crear modelos neuronales para la predicción a corto plazo utilizando como entradas los consumos pasados, la predicción meteorológica y los factores sociales. En este artículo se presentan los resultados logrados con este programa poniendo especial interés en el problema de determinar el conjunto de variables de entrada del modelo .

El texto se ha organizado de la siguiente forma: en la sección siguiente se presenta con más detalle el problema de la predicción a corto plazo, posteriormente se presenta la estructura del programa SICONEL. En la sección 4 se presenta el modelo neuronal utilizado. A continuación se muestran los resultados obtenidos con modelos SICONEL. La búsqueda del vector de entrada al modelo que produce mejores resultados se presenta en la sección 6. Finalmente se muestran las conclusiones obtenidas.

2 Predicción de carga corto plazo

Consiste en predecir el consumo horario en las siguientes 24, 48 ó 72 horas. El problema suele abordarse en dos pasos: predecir el consumo total diario y la forma horaria. La predicción horaria se obtiene repartiendo el consumo total diario predicho entre las 24 horas del día usando la forma horaria predicha. En este artículo se analiza únicamente el problema de la predicción del consumo total diario que es el que ha planteado mayores dificultades.

Los modelos creados con SICONEL para predecir el consumo total son redes de neuronas de dos capas ocultas con un nodo de salida lineal y con un vector de entradas definible por el usuario. El programa permite entrenar redes de neuronas a partir de datos históricos y también usar las redes construidas para generar predicciones.

La predicción del consumo total de la demanda eléctrica consiste en proporcionar una estimación de dicho consumo con 24 horas de antelación. Denotando al consumo habido en el día i como $y(i)$ y al día actual como $k-1$ el problema consiste en hallar $\hat{y}(k)$; es decir, estimar el valor de y para el día de mañana k.

Habitualmente se supone que el consumo total de un día cualquiera depende de factores como los que se listan a continuación.

- Autoregresivos. El consumo total depende de consumos en días pasados.
- Ambientales. Factores atmosféricos como la temperatura, humedad, etc.
- Sociales. Como el tipo de día (festivo o laborable), el día de la semana, etc.

- Temporales. Como la época del año.

Estos factores se usan para predecir la demanda por medio de un modelo. El modelo es una relación de la forma:

$$\hat{y}(k) = f(\mathbf{x}(k)) \tag{1}$$

donde la función f puede ser lineal, polinómica, o con cualquier otra estructura (por ejemplo una red de neuronas). El vector $\mathbf{x}$ es llamado vector de entrada del modelo y contiene medidas de los factores anteriormente citados.

El programa SICONEL permite usar cierto número de factores combinados como componentes del vector de entrada. De entre los factores disponibles se han resaltado algunos por ser los utilizados en las pruebas de este artículo. Son los siguientes:

- Consumo del día anterior del mismo tipo. Los días se clasifican en cuatro grupos: sábados, domingos, lunes y el resto. Esta variable se denota como $y(k-d)$, donde k representa al día a predecir y d es el retraso necesario para obtener un día del mismo tipo. Este retraso es variable según el tipo de día. Por ejemplo, si k es un viernes, entonces $d = 2$, puesto que $k-2$ es miércoles y pertenece al mismo grupo que el viernes. Si el día a predecir es lunes entonces $d = 7$.

- Medias de consumo de días pasados. Por ejemplo, la media de consumos del mes anterior.

- Temperatura media prevista para el día de la predicción. Se denota por $\hat{t}(k)$.

- Tipo de día, $td(k)$. Los valores asignados a cada grupo son $td(k) = 0.6$ si el día de la predicción es lunes, $td(k) = 0.8$ para días de martes a viernes, $td(k) = 0.4$ para los sábados y 0.2 para los domingos.

3 Desarrollo de los programas para SICONEL

El proceso de creación de modelos para la previsión de la demanda tiene cuatro partes bien diferenciadas:

1. Análisis del problema.

2. Determinación de la estructura del modelo.

3. Ajuste de parámetros del modelo y validación del mismo.

4. Explotación del modelo con posible adaptación.

Cada una de estas tareas puede estar asistida por la herramienta de predicción SICONEL. Los módulos que las llevan a cabo son:

1. Generación del modelo. Permite obtener modelos a partir de medidas de variables de interés. Para ello ha de estimar los parámetros ajustables del mismo. Es preciso indicar que el módulo necesita el concurso de un operador cualificado para decidir las variables que se han de incluir en el vector de entrada al modelo.

2. Uso del modelo. Una vez obtenido el modelo de predicción es necesario alimentarlo con las entradas convenientes para obtener como resultado la predicción de la demanda del día deseado.

3. Adaptación del modelo. Cada cierto tiempo es preciso retocar el modelo para refinarlo o simplemente para tomar en consideración nuevas medidas hechas con posterioridad a su construcción.

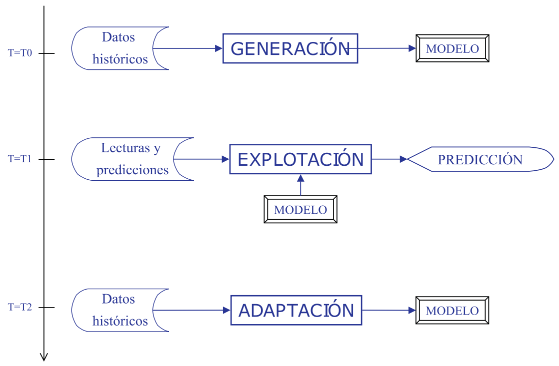


Figura 1: Secuencia de uso de las herramientas a desarrollar

La figura 1 ilustra el proceso de uso de las herramientas que se han de desarrollar. En la parte superior se tiene el módulo de generación, que usa datos para crear el modelo. Esta es la primera tarea a realizar por parte del usuario del programa SICONEL. El objetivo es obtener un fichero que contiene, en cierto formato de datos, el modelo de predicción. En un segundo paso el modelo se usa para predecir la demanda. El módulo que se usa en esta segunda tarea es el de explotación, que produce predicciones a corto plazo de la demanda eléctrica a partir de medidas de consumo y de predicciones de temperatura. Finalmente, en la parte inferior se muestra el módulo de adaptación que afina los parámetros del mismo para acomodarlo a situaciones cambiantes. La adaptación es una tarea posterior a las dos anteriores. Su objetivo es modificar el modelo de forma periódica, para tomar en consideración las últimas medidas.

3.1 Generación del modelo

Se entiende por generación del modelo el proceso a seguir partiendo del análisis previo del problema hasta obtener la estructura y los valores de los parámetros del modelo para la predicción de la demanda.

El predictor incluye herramientas para la construcción de los varios módulos de que consta el modelo de predicción. En la figura 2 se muestra el proceso de creación de uno cualquiera de los módulos.

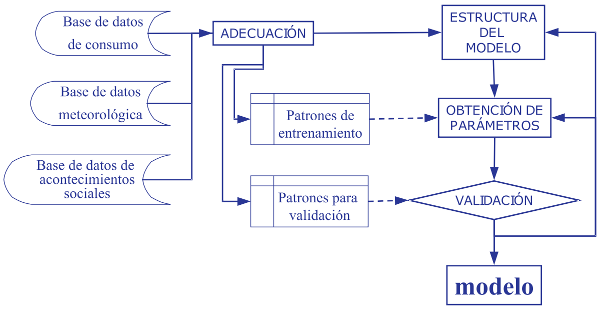


Figura 2: Proceso de creación de un módulo

La adecuación es el proceso mediante el cual los datos históricos se agrupan en vectores de entrada y salida. De este modo se pasa a un problema en el cual el tiempo no se considera explícitamente pues está implícito en las variables de entrada consideradas. Recuérdese que los vectores de e/s consisten en valores retrasados de las variables de interés. Gracias a que se incluyen valores pasados es posible reconstruir la dinámica observada en los datos históricos.

Se observa que la obtención del modelo conlleva un proceso iterativo hasta que se produzca la validación. En cada iteración puede ser necesario modificar la estructura del modelo o sólo obtener un nuevo conjunto de parámetros del mismo.

3.2 Datos

Los datos para los experimentos descritos en este documento han sido proporcionados por la Compañía Sevillana de electricidad. Consisten en medidas horarias de la potencia puesta a disposición de la red. Por motivos de confidencialidad se muestran sólo gráficas con valores normalizados en periodos sin identificar.

Además se dispone de medias diarias de temperaturas representativas de la región.

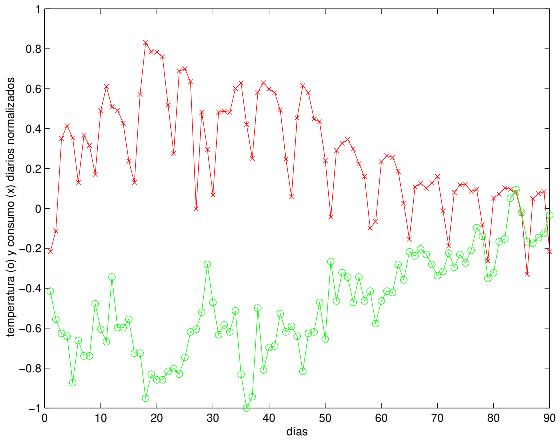


Figura 3: Consumo total (x) y temperatura media (o) diarios durante tres meses.

La figura 3 muestra la evolución del consumo total y de la temperatura media diaria durante tres meses. Es posible observar las variaciones semanales del consumo y cierto cambio de unas semanas a otras.

El efecto de la temperatura sobre el consumo total puede ponerse de manifiesto fácilmente dibujando la gráfica de una variable frente a la primera. Como se observa en la figura 4 las temperaturas extremas inducen un consumo alto.

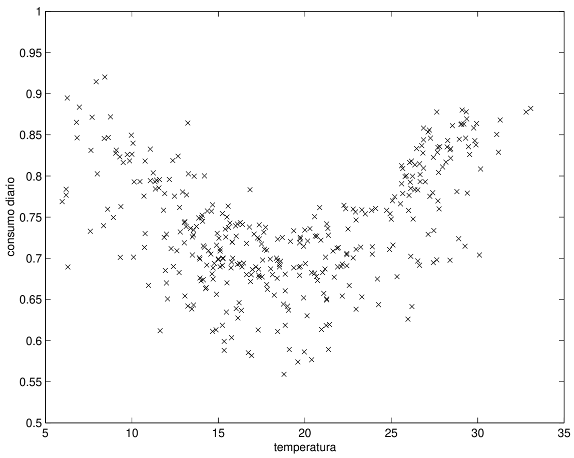


Figura 4: Consumo total frente a la temperatura media diaria en la región.

Este hecho advierte contra el uso de modelos lineales simples para la predicción. El efecto de la temperatura es claramente no lineal y esto ha de ser tenido en cuenta por el modelo. La figura 4 muestra además que el efecto de la temperatura sobre el consumo no es el mismo para cada año. Este hecho es facilmente explicable en términos económicos. La mejoría de la economía media en la región permite la instalación de nuevos equipos de refrigeración o calefacción, por lo que el consumo debido a estos aparatos es mayor.

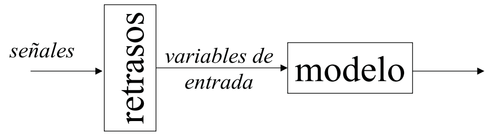
3.3 Filtrado y acondicionamiento

Los datos tal y como son aportados por la compañía eléctrica pueden no ser aptos para su uso en el programa SICONEL. Es preciso realizar el acondiciona-miento de los mismos en la forma en que se detalla a continuación.

- Históricos de consumo. Los datos de consumo pueden incluir valores incorrectos que es preciso eliminar. Especial atención hay que prestar a días con incidencias tales como: cambios de horario, el 29 de Febrero, festivos especiales, etc.
- Históricos de temperaturas y fenómenos atmosféricos. Las temperaturas horarias, máximas y mínimas de la región en la que se va a aplicar el modelo de predicción han de obtenerse como media ponderada de las temperaturas correspondiente a zonas más o menos pobladas. Igualmente con los demás factores meteorológicos.
- Históricos de acontecimientos sociales. Los acontecimientos sociales han de clasificarse en varias categorías de acuerdo con la influencia que puedan tener en el consumo. Entre otros cabe citar los siguientes acontecimientos: vacaciones, puente, festivos, ferias.
- Históricos de tipo de día. Es importante saber el tipo de día correspondiente a cada fecha del calendario, pues la demanda de los Lunes es distinta a la de los fines de semana o la de los Martes.

3.4 Selección de variables de entrada

Como ya se ha indicado, el modelo usa información contenida en el vector de entrada para producir las predicciones. Las variables de entrada son las componentes del vector de entrada del modelo x. En dicho vector se incluyen valores pasados de las variables de interés, como son: la demanda del día anterior, la del día equivalente en la semana pasada, la del día equivalente el año pasado, la temperatura del día anterior, la predicción de la temperatura, el día de la semana, el mes, el año, y los acontecimientos sociales previstos.



$$\mathbf{x} = [t_1 \quad c_7 \quad h_9 \quad \cdots]$$

Figura 5: Construcción del vector de entrada del modelo.

Es preciso determinar qué variables se van a usar de entre todas las que se disponen, a este proceso se le llama selección de variables de entrada. Una elección inadecuada puede conducir a resultados sub-óptimos. Introducir menos información de la necesaria impide construir buenos predictores. En el otro extremo, incluir más variables de las necesarias puede producir sobreajuste y como consecuencia malos resultados en la predicción.

El programa ha de permitir al usuario la definición de las variables que se van a usar como entrada al modelo. Esta información será almacenada en un fichero de configuración que identifica cada coeficiente del vector de entrada al modelo con el valor de una señal en un instante de tiempo relativo al momento en que se realiza la predicción.

3.5 Pasado y futuro

Previamente a la realización de cada experimento los datos se dividen en los siguientes conjuntos disjuntos:

- Conjunto Histórico (CH). Incluye consumos y temperaturas en un intervalo temporal dado.
- Conjunto Futuro (CF) o conjunto nuevo (CN). Incluye consumos y temperaturas de un intervalo posterior a CH y no solapante.

Esta división sirve para recrear la situación real que encuentra un predictor. Por una parte posee datos pasados que puede usar para construir modelos. Por otra parte, el modelo será usado en datos que desconoce pues son obviamente futuros. Esto quiere decir que el predictor tendrá que realizar la evaluación del modelo se realizará con los datos del CF y estos datos no estarán disponibles hasta después de que el modelo ha sido totalmente construido.

En los experimentos mostrados en este documento se ha realizado la anterior división manteniéndose el CF oculto hasta el final. De este modo se puede comprobar si los modelos desarrollados pueden enfrentarse con datos no conocidos previamente.

El CH se subdivide a su vez en:

- Conjunto de Entrenamiento (CE)
- Conjunto de Prueba (CP)

Los datos del CE son los usados para obtener los parámetros ajustables del modelo (pesos de las redes de neuronas) los datos del CP se usan para decidir qué modelo es mejor entre dos dados. Por ejemplo, entre dos redes de neuronas con distinto número de nodos y que proporcionan errores parecidos en CE se prefiere aquella que proporciona menor error en CP. Esta técnica se conoce con el nombre de validación cruzada.

El CE contiene en los experimentos realizados un 80% del total de datos de CH, el resto van a parar al CE. La división se realiza de forma aleatoria, intentando que la distribución de vectores $\mathbf{x}$ sea parecida en ambos conjuntos.

Para poder comparar modelos es preciso definir índices que midan la bondad de los mismos. El elemento fundamental es el error de predicción, que usualmente se mide en términos porcentuales. El error porcentual se calcula como

$$ep(k) = 100\frac{y(k) - \hat{y}(k)}{y(k)} \tag{2}$$

donde k es el día de la predicción, $y(k)$ el consumo diario verificado el día k e $\hat{y}(k)$ es la predicción realizada.

El error porcentual cuadrático medio es una medida del error de predicción durante cierto periodo. Se define como

$$\|Ep\|_C^2 = \sqrt{\frac{\sum_{i=1}^{N_C} ep^2(i)}{N_C}} \tag{3}$$

donde C indica el conjunto de puntos usados para realizar la media.

El histograma es una representación gráfica que da una idea de la distribución del error a lo largo de una serie de días. En el eje horizontal se indica el valor del error en tramos, por ejemplo, de 0 a 0.5%. En el eje horizontal se indica el número de veces en que se produjeron errores dentro de dicho rango.

4 Modelo neuronal

El programa SICONEL permite usar redes de neuronas con la siguientes características:

- Una capa de entrada de dimensión elegida por el usuario.
- Dos capas intermedias con función de activación sigmoidal. El número de nodos en estas capas es indicado también por el usuario.
- Un nodo de salida lineal.
- Las conexiones son siempre de todos los nodos de cualquier capa con todos los demás de la capa siguiente. Además existen conexiones entre la capa de entrada y el nodo lineal, para facilitar la aproximación de la parte lineal del problema.
- Nodos de continua. Son nodos cuya entrada es constantemente igual a 1 y que tienen conexiones con cada nodo de la capa siguiente. En cada capa oculta y en la capa de entrada se ha dispuesto un nodo de continua.

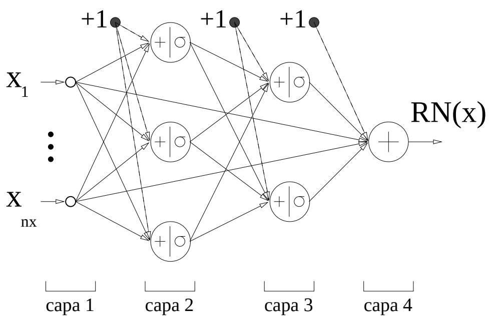


Figura 6: Estructura de la red de neuronas empleada en SICONEL.

En la figura 6 se muestra el esquema de la red de neuronas que utiliza SICONEL. La capa que está situada más a la izquierda es la capa de entrada. Las señales que se aplican a los nodos de esta capa son las componentes del vector de entrada $\mathbf{x}$. Las dos capas intermedias reciben el nombre de capas ocultas.

Cada nodo de las capas ocultas realiza una función no lineal de las señales que recibe. Esta función tiene dos partes: en primer lugar se calcula una combinación lineal o suma ponderada que puede escribirse como producto vectorial del vector de señales $\mathbf{n}$ y de un vector de pesos de ponderación $\mathbf{W}$, se obtiene de este modo:

$$s = \mathbf{W}^t\mathbf{n} = \sum_{i=1}^{N_n} W_i n_i \tag{4}$$

En segundo lugar se obtiene el valor de salida del nodo aplicando la función $\sigma(.)$ a la suma s

$$z = \sigma(s) = \frac{1 - e^{-s}}{1 + e^{-s}} \tag{5}$$

El nodo de salida únicamente realiza la suma ponderada de las señales que inciden sobre él. Los patrones de entrada a la red y la salida son normalizados al intervalo $[-1, 1]$ antes de ser introducidos en la misma. Estos cálculos no afectan a la salida del programa pues siempre se realiza el escalado inverso antes de escribir en los ficheros de resultados.

Este tipo de redes permiten aproximar arbitrariamente bien funciones no lineales sobre conjuntos compactos bajo leves asunciones acerca de la continuidad y derivabilidad de tales funciones supuesto que se dispone de un número infinito de nodos en las capas ocultas. En muchos casos prácticos es además posible obtener aproximaciones de utilidad con un número de nodos manejable.

El algoritmo de entrenamiento usado por SICONEL es Rprop [3], que ha demostrado su eficacia en bancos de prueba.

Un ejemplo de entrenamiento se muestra en la figura 7. Cada marca muestra los errores E_{CE} (cruces) y E_{CP} y (círculos) obtenidos durante el entrenamiento. El eje horizontal representa el número de pasadas o ciclos de entrenamiento usando el CE. Cada cuatro pasadas se detiene el entrenamiento para calcular E_{CE} y E_{CP}. El valor de los pesos se graba entonces para poder ser recuperados posteriormente. En la figura se han unido con segmentos los puntos calculados para ayudar a ver la evolución.

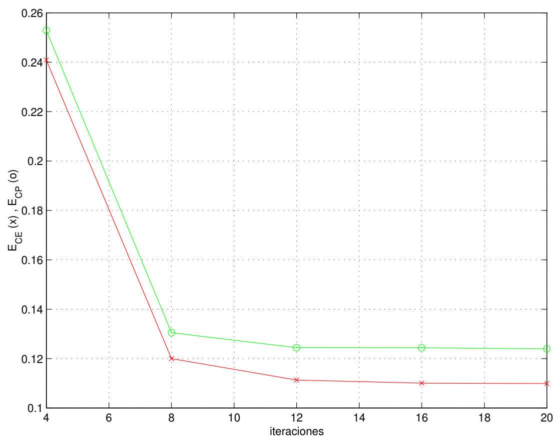


Figura 7: Evolución de los errores en CE y CP durante el entrenamiento.

Los resultados que se obtienen al entrenar una red varían de unos experimentos a otros pues el valor inicial de los pesos es aleatorio. Para remediar este problema se entrenan muchas redes y se almacenan los valores indicados anteriormente. Con estos valores se pueden obtener unos índices que ayuden a elegir la red:

$$
\begin{aligned}
J_1 &= \frac{1}{N_{CE}} \sum_{i=1}^{N_{CE}} e_i^2 + \frac{1}{N_{CP}} \sum_{i=1}^{N_{CP}} e_i^2 \qquad (6) \\
J_2 &= \frac{1}{N_{CN}} \sum_{i=1}^{N_{CN}} e_i^2
\end{aligned}
$$

que miden el error de predicción en los conjuntos CH y CF.

Por suposición los valores de J_2 pertenecen al futuro, por lo que no pueden usarse para elegir la red con la que efectuar predicciones. Sí puede usarse J_1 en la esperanza de que sea un buen indicador de J_2. Para ello se realizan numerosas pruebas y se escogen aquellas redes que proporcionan un mejor valor de J_1.

Un ejemplo de los resultados obtenidos puede verse en la tabla 1 (de la cual se han apartado los casos de sobreentrenamiento).

nn	*nc*	J_1	J_2	*ve*
7.0000	15.0000	0.0272	0.1858	0.6177
7.0000	40.0000	0.0273	0.1768	0.9961
6.0000	10.0000	0.0274	0.1753	0.6354
5.0000	60.0000	0.0275	0.1865	1.0121
6.0000	40.0000	0.0277	0.1774	1.8425

Tabla 1: Resultados de varios entrenamientos con el mismo CE y CP.

Hay que tomar ahora en consideración el problema de la elección de CE y CP. Estos conjuntos son el resultado de dividir aleatoriamente el conjunto CH en dos. Es previsible que los valores de J_1 y J_2 se vean afectados por esta división aleatoria. Para eliminar la influencia de este factor se repite la prueba anterior varias veces utilizando distintas divisiones en CE y CP. Los resultados finales se hallarán como media de todas las pruebas realizadas.

De estas y otras pruebas se obtiene como conclusiones:

- Que el valor de J_1 no es un buen indicador del valor de J_2 .
- Que la elección de CE y CP puede tener influencia sobre los resultados.

Por este motivo se va a realizar media sobre varias redes y sobre varias elecciones de CE y CP.

4.1 Uso del modelo

La explotación del modelo consiste en obtener predicciones a corto plazo de la demanda; es decir, el modelo ha de producir predicciones de demandas con distintos horizontes. Por ejemplo, con un día de antelación, con medio día, 48 horas, etc. Para ello ha de obtener datos de consumo pasado, las predicciones y medidas meteorológicas, la clasificación de factores sociales de diversa índole y los indicadores del día dentro del calendario. Estos valores forman lo que se ha denominado vector de entrada al modelo.

En la figura 8 se muestra el diagrama de bloques de la herramienta de explotación, en el cual se puede observar que el modelo es alimentado con un vector de entrada formado por variables de interés seleccionadas de los datos disponibles. Se observa que la influencia de los acontecimientos sociales en el modelo viene determinada por un bloque cuantificador. La tarea de este bloque consiste en asignar importancia a los acontecimientos sociales según categorías previamente analizadas.

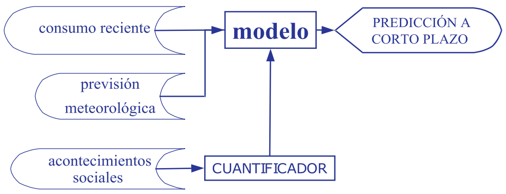


Figura 8: Bloques de la herramienta de explotación

A partir de los datos de consumo, temperaturas y del calendario se generarán ficheros que podrán ser usados por el programa. Además de estos ficheros de datos son necesarios otros de configuración y de resultados.

5 Resultados con modelos creados con SICONEL

En esta sección se presentan los resultados obtenidos al realizar predicciones sobre días futuros (días no utilizados en la construcción del modelo) utilizando diferentes entradas a los modelos neuronales con el fin de comparar los resultados obtenidos con los diferentes modelos. Se han utilizado cuatro modelos que experimentalmente dieron los mejores resultados. Persiguiendo hacer una comparativa entre ellos con el fin de hallar un buen modelo para predecir correctamente la demanda eléctrica en el caso de la compañía Sevillana de Electricidad.

Para realizar las predicciones se han utilizado los días correspondientes al periodo de tiempo comprendido entre un 1 de Enero y un 31 de Marzo (estos días son consecutivos a los utilizados en la construcción del modelo). Teniendo en las bases de datos los datos correspondientes hasta el día k-1 (ayer) se intenta predecir el día k (mañana). La curva predicha se compara con la curva real, observándose las similitudes entre ellas y calculándose el error producido en la predicción del consumo total. Para entrenar la red se han utilizado los datos recogidos durante 4 años completos. No se ha vuelto a reentrenar la red conforme se ha ido avanzando en las predicciones.

5.1 Modelos considerados

Los modelos que se van a ir cambiando son los correspondientes a la red de predicción del consumo total ya que es la más crítica debido a los grandes errores que pueden presentar las predicciones del consumo total para un día. El modelo correspondiente a la forma horaria se mantiene invariante.

Las variables consideradas para construir el vector de entrada para la red de la forma horaria son las siguientes:

- La forma horaria del día anterior del mismo tipo.
- La forma horaria del día similar del año pasado.
- La temperatura media en la región prevista para el día de la predicción.
- El tipo de día.

Para la red de consumo total se han considerado cuatro grupos de diferentes entradas. Estos grupos de variables de entrada para el consumo total son los que nos van a diferenciar un modelo de otro.

1. Variables de entrada para la red del consumo total en el modelo 1
    - El consumo total del día anterior del mismo tipo.
    - La media de consumos de los siete días anteriores.
    - La media de consumos de 31 días centrada en tres meses atrás.
    - La temperatura media en la región prevista para el día de la predicción.
    - El tipo de día, es decir, si es festivo, lunes, sábado, domingo, ...

2. Variables de entrada para la red del consumo total en el modelo 2
   - El consumo total del día anterior del mismo tipo.
   - La media de consumos de los siete días anteriores.
   - La media de consumos de 31 días centrada en tres meses atrás.
   - La media de consumos de 31 días centrada en seis meses atrás.
   - La temperatura media en la región prevista para el día de la predicción.
   - El tipo de día, es decir, si es festivo, lunes, sábado, domingo, ...
3. Variables de entrada para la red del consumo total en el modelo 3
   - El consumo total del día anterior del mismo tipo.
   - La media de consumos de los siete días anteriores.
   - La media de consumos de 31 días centrada en dos meses atrás.
   - La media de consumos de 31 días centrada en cuatro meses atrás.
   - La temperatura media en la región prevista para el día de la predicción.
   - El tipo de día, es decir, si es festivo, lunes, sábado, domingo, ...
4. Variables de entrada para la red del consumo total en el modelo 4
   - El consumo total del día anterior del mismo tipo.
   - La media de consumos de los siete días anteriores.
   - La media de consumos de 31 días centrada en tres meses atrás.
   - Consumo total medio de 7 días centrado en un año anterior.
   - La temperatura media en la región prevista para el día de la predicción.
   - El tipo de día, es decir, si es festivo, lunes, sábado, domingo, ...

5.2 Obtención de los parámetros del modelo

El entrenamiento de las redes neuronales se ha realizado mediante el algoritmo "Rprop" ya que es más rápido que el algoritmo "Backpropagation" en el el entrenamiento de redes basadas en el "perceptron multicapa".

Modelo 1	Modelo 2	Modelo 3	Modelo 4
0.184	0.130	0.124	0.128
0.202	0.116	0.132	0.149

Tabla 2: Errores cuadráticos medios en los datos del conjunto de entrenamiento (primera fila) y en el de prueba (segunda fila) para la red del consumo total.

Modelo 1	Modelo 2	Modelo 3	Modelo 4
0.0014	0.0011	0.0011	0.0011
0.0013	0.0011	0.0011	0.0011

Tabla 3: Errores cuadráticos medios en los datos del conjunto de entrenamiento (primera fila) y en el de prueba (segunda fila) para la red de la forma horaria.

5.3 Resultados obtenidos en la predicción de días futuros

Para probar la bondad del modelo obtenido se predijeron "días futuros", como ya se ha indicado, para la construcción del modelo se utilizaron datos correspondientes a cuatro años. A continuación se presentarán los resultados obtenidos al utilizar los modelos para la predicción de días de los meses siguientes a los utilizados en el entrenamiento.

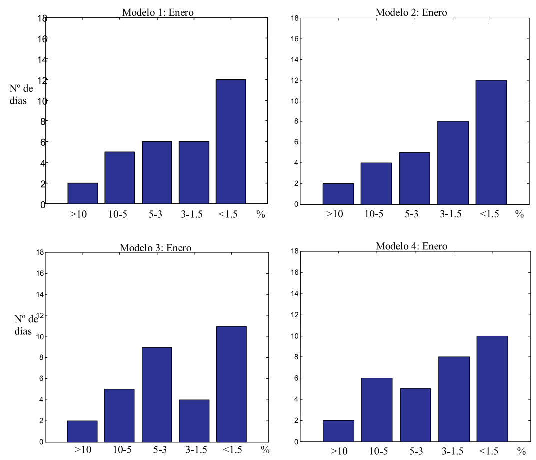


Figura 9: Distribución de los errores obtenidos en un mes de enero para cada uno de los cuatro modelos analizados. En el eje de las y se indica el número de días, y en el eje de las x cada uno de los márgenes de error considerados

En las siguientes figuras se muestran los resultados obtenidos en la predicción de los días de un mes de Enero, no hay que olvidar todas las dificultades que presentan la predicción de los días de este mes por los abundantes festivos y puentes (tanto oficiales como no) que en él suelen concurrir. El otro mes del cual se presentaran las predicciones obtenidas es el mes de marzo, mes que en el año 98 no tenía ninguna característica especial, ya que la Semana Santa ese año fue en abril.

Las gráficas de barras indican el número de días cuyo error entre el valor del consumo real ocurrido ese día y el predicho están dentro de unos determinados márgenes porcentuales. Ese error se calcula según la siguiente expresión:

$$e_{CT} = \frac{CT_Real - CT_Predicho}{CT_Real} * 100 \qquad (7)$$

Donde:
CT: Es el Consumo Total (real o predicho) de un día determinado.

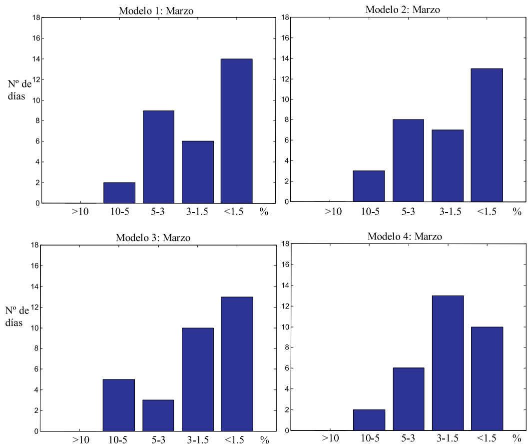


Figura 10: Distribución de los errores obtenidos en un mes de marzo para cada uno de los cuatro modelos analizados. En el eje de las y se indica el número de días, y en el eje de las x cada uno de los márgenes de error considerados

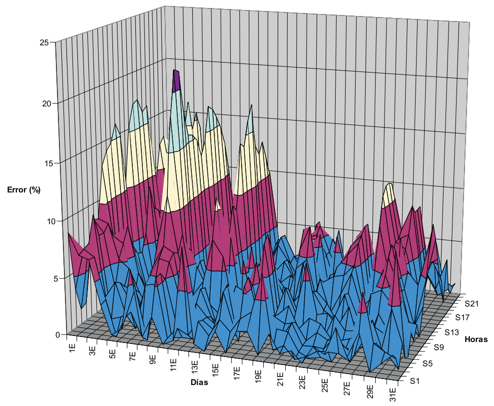


Figura 11: Error porcentual del consumo total a lo largo de un mes de Enero

En los gráficos tridimensionales (figuras 12 y 13) se han representado los errores cometidos en la estimación de consumos horarios para todos los días de los meses de enero y marzo utilizando el modelo 1. Del estudio de estos resultados se pueden sacar algunas conclusiones acerca del desempeo de las herramientas de SICONEL, indicando posibles vías para mejorar, por ejemplo, una mejor consideración de los días festivos y algunas predicciones de días influenciados por ellos.

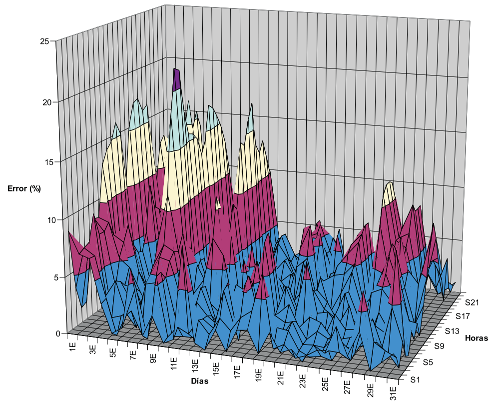


Figura 12: Error porcentual del consumo total a lo largo de un mes de Enero

Por ejemplo los días 8 y 12 presentan errores elevados de estimación, cuando, en principio, estos días no tienen ninguna particularidad especial. La explicación se encuentra en que los días en que el modelo se ha basado para estimar el consumo son especiales. Recordemos que una de las entradas usadas por el modelo 1 es el consumo total del día anterior del mismo tipo; para el caso del jueves día 8 el día anterior del mismo tipo es el viernes día 2. Pero este día está entre un festivo y un sábado por lo que cabe de suponer que muchas personas y empresas hicieron puente por lo que el consumo no fue el de un viernes normal y esto repercutió en la predicción del día 8. En la predicción del día 12 influyó el comportamiento del lunes anterior, en este caso el día 5; día cuyo consumo no corresponde a un lunes cualquiera. Se puede observar como las horas con errores mayores en las predicciones del día 5 y 12 prácticamente son las mismas en los días.

6 Selección de entradas

Ya se ha indicado que el programa SICONEL permite entrenar redes de neuronas cuya entrada son vectores de dimensión seleccionable por el usuario. Las variables que pueden formar parte del vector

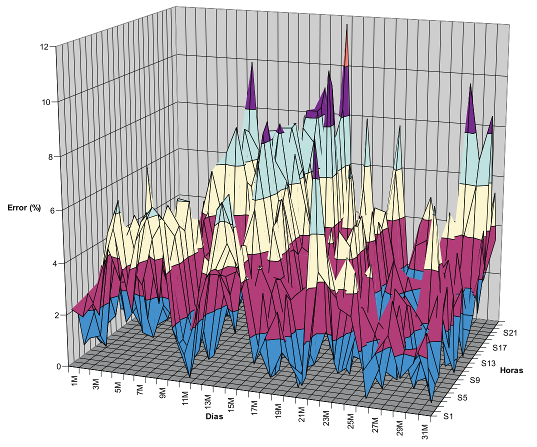


Figura 13: Error porcentual del consumo total a lo largo de un mes de Marzo

de entrada se han definido con anterioridad e incluyen: valores pasados de consumo, factores atmosféricos, factores sociales e indicadores.

Se plantea el problema de decidir cuales de estas variables se van a usar en el modelo de predicción. Resulta evidente que las componentes más idóneas del vector de entrada pueden variar de unas compañías eléctricas a otras, en función de las características del consumo. La elección por tanto ha de realizarse en base a los datos históricos disponibles. En [1] puede hallarse un estudio más profundo de este problema.

6.1 Método de selección

Pudiera pensarse que ante la duda de que variables incluir en $\mathbf{x}$ la mejor opción es incluirlas todas. Esta no es una opción recomendable pues el número de parámetros de la red depende del número de conexiones y éste del número de entradas. No es aconsejable usar redes con un alto número de parámetros pues se corre el peligro de que se produzca sobreentrenamiento.

En los problemas de ajuste de parámetros, como es el caso del entrenamiento de redes de neuronas, es aconsejable que el número de parámetros ajustables sea mucho menor que el de datos usables como ejemplos. Los datos de que se dispone son $nx = 365 \cdot 3 = 1095$ y el número de conexiones es $ne + 1 + (ne + 1) \cdot nn + nn$, que depende, como se ve, de ne, número de entradas. Si se toma como tope para el número de parámetros $nx/10 \approx 100$ se tiene se usan todas las variables de entrada $ne = 11$, por lo que el número de nodos de la capa oculta ha de ser tal que $12 + 12 \cdot nn + nn < 100$, por lo que $nn < 6$.

En ocasiones las variables de entrada contienen información redundante, pues no son completamente independiente unas de otras. Por este motivo, el hecho de eliminar una de ellas no causa problemas para conseguir buenas predicciones. Se consigue además que el tamaño máximo de la red crezca, por lo que se pueden obtener menores errores en el entrenamiento.

El método usado para seleccionar variables de entrada consiste en explorar selectivamente las combinaciones posibles de variables en el vector de entrada. Se comienza obteniendo modelos que contienen todas las variables y modelos en los que falta una variable. En estos últimos se observa el aumento de error. Aquél que produce menor aumento indica qué variable es menos útil. Se elimina esta variable de la lista y se continua el proceso obteniendose modelos en los que falta una variable más.

La figura 14 muestra de forma gráfica el proceso de selección de variables por eliminación. Cada caja representa una combinación diferente de las posibles variables. La combinación se codifica con unos y ceros. La posición del uno o cero indica la variable. Un uno significa que la variable en cuestión es usadas, un cero indica la ausencia de la variable en el vector de entrada.

En cada etapa (fila horizontal) se elige la variable que se va a desechar en sucesivas etapas. Esta variable es indicada por el modelo que produce errores menores. Puede observarse que, como consecuencia de desechar progresivamente variables, no se explora todo el árbol, sino que en cada paso se selecciona una rama.

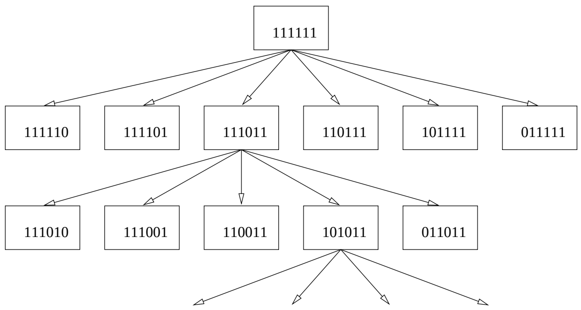


Figura 14: Exploración del árbol de selección de variables de entrada. Cada caja representa una combinación diferente de las posibles variables. La combinación se codifica con unos y ceros. La posición del uno o cero indica la variable. Un uno significa que la variable en cuestión es usadas, un cero indica la ausencia de la variable en el vector de entrada.

El proceso de selección de entradas ha de terminarse cuando los errores crezcan de forma significativa.

6.2 Notación

A efectos de simplificar la descripción de los experimentos y sus resultados se introduce una notación binaria. En primer lugar se asigna un orden a las variables susceptibles de ser usadas como entrada. El orden es el siguiente:

1. Consumo del día anterior del mismo tipo *ctdamt*.
2. Media de consumo de siete días anteriores $m7da$.
3. Media de consumo de catorce días anteriores $m14da$.
4. Temperatura media prevista para el día de la predicción $t(k)$.
5. Media de temperaturas en días pasados *mta*.
6. Media semanal de consumo centrada en un año atrás *mdaa*.
7. Media mensual de consumo centrada en dos meses $mmc2ma$.
8. Media mensual de consumo centrada en tres meses pasados $mmc3ma$.
9. Media mensual de consumo centrada en cuatro meses $mmc4ma$.
10. Media mensual de consumo centrada en seis meses $mmc6ma$.
11. Tipo de día, *td*.

La presencia o ausencia de cada una de estas variables en el vector de entrada se indicará mediante un número binario de 11 guarismos. Un uno en la posición j del número indica que la variable j-ésima forma parte del vector de entrada; un cero indica lo contrario. Así, el vector de entrada correspondiente a 01000010001 es el formado por las variables: $m7da$, $mmc2ma$ y *td*.

6.3 Exploración de combinaciones

Para poder realizar la selección de variables de entrada se han de entrenar numerosas redes. Para cada combinación explorada se entrenan 200 redes que solo tienen en común el vector de entrada. El tamaño de la red, el número de ciclos de entrenamiento y hasta la división de datos en CE y CP es distinto (y aleatorio) para cada una de las redes.

De las 200 redes para cada combinación se seleccionan las 20 mejores. La selección se hace de acuerdo con un índice que penaliza el error en CE y el error en CP. En las pruebas se desechan las redes con sobreentrenamiento; es decir, el error E_{CP} medido cada *nc* pasadas no puede ser creciente.

La comparación de redes pertenecientes a distintas combinaciones de variables de entrada se realiza atendiendo al índice $J = E_{CE}^2 + E_{CP}^2$.

En los experimentos realizados se han probado las combinaciones $1{*}{*}1{*}{*}{*}{*}{*}{*}1$ donde el símbolo $*$ significa cualquier valor uno o cero. Las variables primera, cuarta y undécima siempre se incluyen en el vector de entradas pues su eliminación no es lógica según el conocimiento que ya se tiene del problema.

La tabla 4 muestra los resultados de las distintas etapas de selección. Cada fila corresponde a una combinación de variables de entradas. Las columnas indican (para las 20 mejores redes) el mínimo valor del índice J, la media y la varianza.

combinación	min J	$\overline{J}$	var(J)
1 1 1 1 1 1 1 1 1 1 1	0.0197	0.0206	0.0005
1 0 1 1 1 1 1 1 1 1 1	0.0196	0.0207	0.0005
1 1 0 1 1 1 1 1 1 1 1	0.0190	0.0210	0.0008
1 1 1 1 0 1 1 1 1 1 1	0.0200	0.0213	0.0007
1 1 1 1 1 0 1 1 1 1 1	0.0184	0.0206	0.0009
1 1 1 1 1 1 0 1 1 1 1	0.0186	0.0207	0.0007
1 1 1 1 1 1 1 0 1 1 1	**0.0177**	**0.0203**	0.0011
1 1 1 1 1 1 1 1 0 1 1	0.0207	0.0219	0.0007
1 1 1 1 1 1 1 1 1 0 1	0.0201	0.0221	0.0007

Tabla 4: Resultados de la selección de entradas para la primera etapa.

De la tabla 4 puede inferirse que la eliminación de variables del vector de entrada puede incluso mejorar la capacidad de las redes para realizar la predicción de la demanda, basta para ello comparar la primera fila con la tercera o quinta. En la fila séptima se produce el menor aumento (en este caso mayor disminución) del error. Esto indica que la variable correspondiente ($mmc3ma$) puede desecharse sin peligro de que el modelo resultante empeore.

Continuando con el procedimiento se han de probar ahora las combinaciones $1{*}{*}1{*}{*}{*}0{*}{*}1$ conteniendo seis de las siete posible variables. De este modo se obtienen los resultados mostrados en la tabla 5.

En la fila cuarta de la tabla 5 se produce el menor aumento (en este caso mayor disminución) del error (indicado en negrita). Esto indica que la variable correspondiente (*mdaa*) puede desecharse sin peligro de que el modelo resultante empeore. Continuando con el procedimiento se han de probar ahora las combinaciones $1{*}{*}1{*}0{*}0{*}{*}1$ conte-

combinación	min J	$\overline{J}$	var(J)
1 1 1 1 1 1 1 0 1 1 1	0.0177	0.0203	0.0011
1 0 1 1 1 1 1 0 1 1 1	0.0193	0.0208	0.0007
1 1 0 1 1 1 1 0 1 1 1	0.0196	0.0213	0.0006
1 1 1 1 0 1 1 0 1 1 1	0.0199	0.0218	0.0008
1 1 1 1 1 0 1 0 1 1 1	**0.0177**	**0.0206**	0.0010
1 1 1 1 1 1 0 0 1 1 1	0.0201	0.0220	0.0007
1 1 1 1 1 1 1 0 0 1 1	0.0210	0.0251	0.0012
1 1 1 1 1 1 1 0 1 0 1	0.0207	0.0221	0.0006

Tabla 5: Resultados de la selección de entradas para la segunda etapa.

niendo cinco de las seis posible variables elegibles. De este modo se obtienen los resultados mostrados en la tabla 6.

combinación	min J	$\overline{J}$	var(J)
1 1 1 1 1 0 1 0 1 1 1	0.0177	0.0206	0.0010
1 0 1 1 1 0 1 0 1 1 1	**0.0193**	**0.0210**	0.0006
1 1 0 1 1 0 1 0 1 1 1	0.0197	0.0214	0.0007
1 1 1 1 0 0 1 0 1 1 1	0.0194	0.0214	0.0009
1 1 1 1 1 0 0 0 1 1 1	0.0210	0.0226	0.0006
1 1 1 1 1 0 1 0 0 1 1	0.0248	0.0265	0.0008
1 1 1 1 1 0 1 0 1 0 1	0.0194	0.0222	0.0009

Tabla 6: Resultados de la selección de entradas para la tercera etapa.

En la fila primera de la tabla 6 se produce el menor aumento del error (indicado en negrita). Esto indica que la variable correspondiente ($m7da$) puede desecharse sin peligro de que el modelo resultante empeore. Continuando con el procedimiento se han de probar ahora las combinaciones $10*1*0*0**1$ conteniendo cuatro de las cinco posibles variables elegibles. De este modo se obtienen los resultados mostrados en la tabla 7.

combinación	min J	$\overline{J}$	var(J)
1 0 1 1 1 0 1 0 1 1 1	0.0193	0.0210	0.0006
1 0 0 1 1 0 1 0 1 1 1	0.0215	0.0228	0.0005
1 0 1 1 0 0 1 0 1 1 1	0.0206	**0.0223**	0.0008
1 0 1 1 1 0 0 0 1 1 1	0.0202	0.0224	0.0009
1 0 1 1 1 0 1 0 0 1 1	0.0235	0.0273	0.0013
1 0 1 1 1 0 1 0 1 0 1	0.0215	0.0227	0.0006

Tabla 7: Resultados de la selección de entradas para la cuarta etapa.

En la fila segunda de la tabla 7 se produce el menor aumento del error medio (indicado en negrita), por lo que se va a desechar la variable correspondiente (mta). Continuando con el procedimiento se han de probar ahora las combinaciones $10*100*0**1$ conteniendo tres de las cuatro posibles variables elegibles. De este modo se obtienen los resultados mostrados en la tabla 8.

combinación	min J	$\overline{J}$	var(J)
1 0 1 1 0 0 1 0 1 1 1	0.0206	0.0223	0.0008
1 0 0 1 0 0 1 0 1 1 1	0.0234	0.0240	0.0003
1 0 1 1 0 0 0 0 1 1 1	**0.0213**	**0.0230**	0.0010
1 0 1 1 0 0 1 0 0 1 1	0.0264	0.0281	0.0006
1 0 1 1 0 0 1 0 1 0 1	0.0225	0.0236	0.0005

Tabla 8: Resultados de la selección de entradas para la quinta etapa.

En esta ocasión, los errores han aumentado considerablemente con respecto a los modelos anteriores. Se va a continuar no obstante una etapa más para verificar este aumento. La variable que se desecha es $mc2ma$ (fila segunda de la tabla 8). Se continua con las combinaciones $10*10000**1$. Los resultados obtenidos se muestran en la tabla 9.

combinación	min J	$\overline{J}$	var(J)
1 0 1 1 0 0 0 0 1 1 1	0.0213	0.0230	0.0010
1 0 0 1 0 0 0 0 1 1 1	0.0240	0.0259	0.0007
1 0 1 1 0 0 0 0 0 1 1	0.0278	0.0290	0.0007
1 0 1 1 0 0 0 0 1 0 1	**0.0233**	0.0248	0.0005

Tabla 9: Resultados de la selección de entradas para la sexta etapa.

A la vista de estos últimos resultados se elige finalmente como entrada para el modelo las variables $ctdamt$, $m14da$, t, $mc4ma$ y td.

7 Conclusiones

Durante el desarrollo de las pruebas mostradas en este artículo se ha puesto de manifiesto la dificultad que plantea la elección de modelos adecuados para la predicción. En efecto, los resultados se ven afectados por:

- La cantidad y calidad de los datos disponibles.
- Los pesos iniciales dados a la red de neuronas.
- El tipo y duración del entrenamiento de la red.
- Las componentes del vector de entrada utilizado.

Todos estos factores tienen en muchos casos interdependencias, por lo que es difícil sentenciar qué modelo es más adecuado. En lugar de esto se han obtenido diversos modelos que producen resultados aceptables en el CN.

Quedan abiertas cuestiones tales como:

- Inclusión de medidas en tiempo real
- Utilización de otras variables meteorológicas, como la nubosidad o el viento.
- Uso de medidas geográficamente distribuidas por la región de interés.
- Posibilidad de automatizar el proceso de prueba y selección de modelos.

Estas tareas serán objeto de estudios futuros.

Agradecimientos

Los autores agradecen a la compañía Sevillana de Electricidad la cesión de los datos usados en el desarrollo del proyecto SICONEL.

Referencias

[1] Arahal, M.R., (1996) Identificación y control mediante redes de neuronas, Tesis Doctoral, Universidad de Sevilla.

[2] Chow, T.W.S. y Leung, C.T. (1997) *Neural Network based short-term load forecasting usign weather compensation.* IEEE Transactions on Power Systems, vol. 11, no. 4, pp 1736–1742.

[3] Riedmiller, M. y Braun, H., (1993) *A direct adaptive method for faster backpropagation learning: the RPROP algorithm.* Actas de la IEEE ICNN, pp 586–591.