Métodos aplicados para la selección de variables de entrada en la predicción del consumo eléctrico en el corto plazo

Villacís Postigo, Fernando
Grupo de Investigación en Tecnologías Informáticas Avanzadas - U.T.N.-F.R.T.
Rivadavia 1050, Tucumán, Argentina
Universidad Nacional de Tucumán - Facultad de Ciencias Naturales e IML
Miguel Lillo 205, Tucumán, Argentina
fvillacis@csnat.unt.edu.ar

Jimenez, Victor Adrian
Grupo de Investigación en Tecnologías Informáticas Avanzadas - U.T.N.-F.R.T.
Rivadavia 1050, Tucumán, Argentina
adrian.jimenez@gitia.org

Will, Adrian
Grupo de Investigación en Tecnologías Informáticas Avanzadas - U.T.N.-F.R.T.
Rivadavia 1050, Tucumán, Argentina
Universidad Nacional de Tucumán - Facultad de Ciencias Exactas y Tecnología
Av. Independencia 1800, Tucumán, Argentina
adrian.will@gitia.org

Abstract

Short-Term Load Forecating (STLF), está tomando una mayor relevancia por el desarrollo de las microgrid y los sistemas Supervisory Control And Data Acquisition (SCADA). Existe una explosión de datos en esta área; para lograr que los sistemas sean prácticos y precisos se necesita seleccionar de manera adecuada las variables de entrada utilizadas por los sistemas de predicción. Se presenta en este trabajo un estado del arte de los métodos de selección de variables utilizados en STLF, orientado a permitir a usuarios del área obtener criterios para elegir los métodos de selección de variables más adecuados en cada caso.

1. Introducción

El objetivo de este trabajo es realizar un análisis de los diferentes métodos de selección de variables para *Short-Term Load Forecating (STLF)*, que son los encargados de elegir el mejor conjunto de variables, que luego serán utilizadas como entrada en los distintos métodos de cálculo. Sobre la base de los conocimientos de otros autores, se analizan los que pueden ser aplicados en regiones en donde existe variaciones climáticas, en estos lugares las personas suelen utilizan *Heating, Ventilating and Air Conditioning (HVAC)*.

La predicción del consumo eléctrico se puede dividir en líneas generales en tres categorías: ***corto***, ***mediano*** y ***largo plazo***. Podemos definir a la STLF o predicción del consumo eléctrico en el corto plazo, como las predicciones realizadas con un horizonte que va desde una hora hasta una semana en el futuro [3, 60]. La razón de su creciente importancia se relaciona, por un lado, con el surgimiento de los mercados de energía. Por otra parte surgimiento de normas ambientales y la necesidad de una mejor gestión de la energía. La gestión eficiente de la energía necesita de predicciones más precisas[3, 22, 28, 60].

También se debe tener en cuenta el surgimiento de las *Microgrids*, con sistemas *Supervisory Control And Data Acquisition (SCADA)*, que generan una gran cantidad de datos y variables que luego deben ser procesados por los sistemas de STLF. Estos sistemas suelen generar más datos que los sistemas tradicionales, que también están implementando sistemas SCADA.

Existen muchos estudios en la comunidad internacional relacionados con las técnicas o modelos que realizan STLF.

Estos modelos pueden ser clasificados en tres grandes grupos, de acuerdo a la técnica utilizada para generarlos. El primer grupo se basa en modelos estadísticos (*Regresión Lineal*, series de tiempo y/o econométricos). Algunos modelos de este grupo se han aplicado a la serie mensual de Chile [8] y serie anual de demanda de Ciprés [18]. Por otro lado, el segundo grupo reúne los modelos basados en inteligencia artificial. Las técnicas de inteligencia artificial han tomado una importante relevancia en los últimos años y se han elaborado estudios donde se comparan sus resultados con los obtenidos con métodos tradicionales. En el caso de Taylor, de Menezes, and McSharry [55], se pronostica la serie horaria de demanda de Río de Janeiro, Inglaterra y Gales mediante un modelo de redes neuronales artificiales y comparan los resultados con modelos estadísticos (*ARIMA, Suavizado Exponencial y Componentes Principales*). Dentro de las técnicas más usadas se encuentran las redes neuronales artificiales *Artificial Neural Networks (ANN)* [1, 7, 25, 30, 48], modelos *neuro-difusos* y modelos *híbridos* [7, 39, 43, 58]. El tercer grupo abarca modelos empíricos que dependen del juicio y la intuición humana. Existen otras clasificaciones en la literatura para los modelos que predicen el consumo eléctrico en el corto plazo propuestas en los siguientes trabajos: Alfares and Nazeeruddin [2], Hernandez, Baladron, Aguiar, Carro, Sanchez-Esguevillas, Lloret, and Massana [29], Hippert, Pedreira, and Souza [30], Singh, Ibraheem, and Muazzam [52].

2. Selección de Variables en STLF

El consumo eléctrico es la suma de todos los consumos eléctricos individuales de los nodos del sistema eléctrico, por este motivo los niveles de consumo eléctrico son un proceso no estacionario, aleatorio, compuesto de miles de componentes individuales. El comportamiento del consumo en el sistema eléctrico es influenciado por un número de factores, que se pueden clasificar según Alfares and Nazeeruddin [2] en *factores económicos*, *hora*, *día*, la *estación del año*, el *clima* y los *efectos aleatorios*. Gross and Galiana [26] (1987) proponen clasificar a los factores que influyen en el consumo eléctrico en cuatro grandes categorías: ***Económicos***, ***Temporales***, ***Climáticos*** y ***Acontecimientos aleatorios***.

La importancia de la selección de variables es evidente, ya que los métodos de predicción del consumo eléctrico aprenden de las relaciones entre las variables de entradas y de salida, basándose en los pares de datos de entrada-salida proporcionados. Si las variables de entrada no representan valores relevantes, no se puede esperar que estos métodos puedan predecir con precisión las variables dependientes. Por otra parte si se seleccionan variables de entradas fuertemente correlacionadas o insignificantes, entonces los recursos computacionales se desperdiciaran durante el entrenamiento, dando lugar a tiempos prolongados de entrenamiento y a veces a resultados con menos precisión [17].

La gran cantidad de características relacionadas con la STLF genera sistemas complejos y de largos períodos de entrenamiento [15, 32]. Por lo tanto, una de las tareas más importantes, en la construcción de un modelo de predicción eficiente basado ANN, como también en otros métodos , es la selección de las variables de entrada adecuadas [32, 60].

Con más variables de entrada, los modelos de predicción generalmente necesitan de más datos históricos para extraer la función de correspondencia entre las entradas y la salida, mientras que para los casos de *spike forcast*, se suele contar con una seria limitación en la cantidad de datos históricos disponibles. Por lo tanto, es necesario refinar el conjunto de variables candidatas, para seleccionar un subconjunto de entrada que sea eficaz para el modelo de predicción [5]. Por ejemplo, los resultados teóricos muestran que al centrarse en solo un pequeño subconjunto de características, un algoritmo puede reducir significativamente el número de hipótesis que debe considerar, entonces hay una correspondiente reducción en el tamaño de las muestras, suficiente para garantizar una buena generalización [9].

En los diferentes trabajos analizados se utilizaron varios métodos para la identificación de variables de entradas redundantes. La mayoría de ellos utilizan técnicas de análisis de correlación, por lo general acompañado de heurísticas y de la experiencia del operador. Una característica común de los métodos estadísticos de selección de variables, utilizados hasta ahora, es que tratan de identificar casi exclusivamente relaciones lineales en los conjuntos de candidatos [17]. Wang et al. [60] utiliza ***fuzzy-rough attribute reduction algorithm***, para seleccionar las variables más significativas. Emplea una clasificación en grupos con límites **blandos** en base a similitudes con otros, sin transiciones bruscas entre agrupaciones. Esta propiedad permite que un elemento pueda pertenecer a más de un grupo. Estos conjuntos son los que se utilizan como entradas para el sistema de predicción. Por su parte, en He et al. [28] para seleccionar las variables de entrada utilizan la ***maximum conditional entropy***, que determina la relevancia de cada variable en la predicción del consumo eléctrico, seleccionando las mejores. No obstante, en Drezga and Rahman [17] se emplea el método de ***phase-space embedding*** para identificar un conjunto de variables **independientes** que están relacionadas con la serie de tiempo del consumo eléctrico. En los casos mencionados, los métodos fueron utilizados para determinar las variables de entradas para sistemas basados en ANN. En Guo [27], se propone un sistema híbrido *PSO-SVR*, utilizando un algoritmo de *Particle Swarm Optimization (PSO)*, para seleccionar las variables óptimas de entrada, limitando el número de variables de entradas en *Support Vector Regression Machines (SVR)* para STLF. Por otra parte, Ceperic et al. [13] utiliza *Forward Selection with Stepwise Regression (SteepwiseFS)* para realizar la selección de varia-

bles, esta técnica inicia sin variables y se van agregando una por vez de acuerdo a un criterio especifico, hasta que todas están incluidas o hasta que se alcanza el criterio de parada [31]. A su vez, Cheng, Chan, and Qiu [15], plantea que en la selección de variables, las m mejores características no conducen necesariamente a una buena selección, y que parte de la información proporcionada por las variables eliminadas se pierde. Por lo tanto, Cheng et al. [15] emplean *Ramdon Forest (RF)* para reducir el conjunto de variables de entrada.

En la selección de características o variables se ha reconocido que las combinaciones de forma individual de buenas características no conducen necesariamente a un buen rendimiento en la predicción. En otras palabras,"las m mejores características no son las mejores m características" [5]. Por ejemplo entre las m mejores características puede haber características redundantes. En sistemas de predicción en donde se limitan a un número reducido de variables de entrada, al elegir variables redundantes, estas pueden dejar afuera de la elección, a variables que aportan distintos conocimientos al sistema. Al no ser incluidas estas variables como entrada del sistema de predicción, la información o conocimiento que aportarían no es tenida en cuenta y se pierde.

3. Variables utilizadas en STLF

La variable más importante en STLF, es la que representa los valores de consumo eléctrico en instantes anteriores de tiempo [55]. Pero existen otras variables o características exógenas que también son utilizadas en los distintos trabajos consultados. Se enuncian a continuación las más comúnmente utilizadas:

- Reactive power [46, 47]
- Consumo eléctrico en otras subestaciones [42].
- Clima (temperatura, velocidad del viento, precipitación, nubosidad, humedad)[6, 13, 19, 21, 24, 27, 28, 33, 34, 40, 42, 46, 53, 54, 60].
- Variables civiles [13, 38].
- Variables temporales [13, 38, 53, 54].
- Variables derivadas[46, 53].
- Predicciones de variables exógenas [22].

De las variables climáticas la más importante es la temperatura [19, 42]. En la Figura 1, se observa cómo los modelos que usaron variables climáticas en sus entradas logran mejores predicciones [53, 54]. Esto ocurre principalmente en ciudades donde los cambios en el consumo eléctrico tiene una estrecha relación con los cambios de la temperatura. En Shanghái, una ciudad importante de China, según Friedrich and Afshari [24], se analizo el consumo eléctrico desde el año 2003 a mayo en 2006, y llegaron a la conclusión de que existe una gran correlación entre ellos.

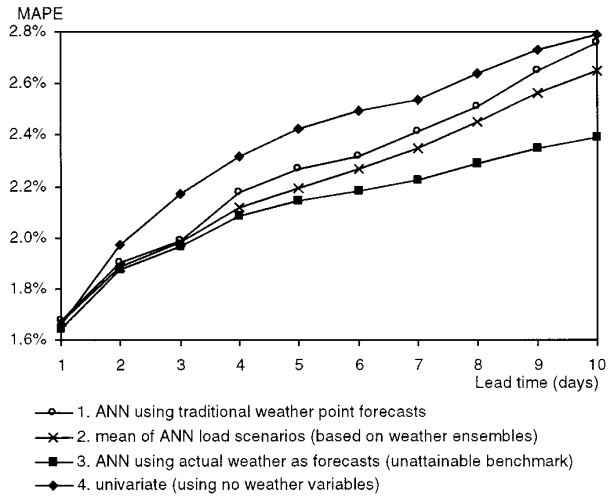

Figura 1. Error MAPE de predicciones del consumo eléctrico utilizando variables climáticas, para distintos horizontes (1 a 10 días).

4. Métodos de Selección de Variables

Existen muchos métodos para realizar STLF, sin importar cual sea el método que se elija, una de las cuestiones importantes en el proceso de predicción, es seleccionar las variables de entrada de un conjunto grande de características [32]. En los distintos trabajos analizados, se utilizaron variables como los consumos en momentos anteriores, distintos factores climáticos, información del calendario, variables derivadas, además de predicciones de algunas variables exógenas, ver Variables utilizadas en STLF. Por lo general el número de variables de entrada se establece de manera discrecional, la mayoría de las veces a partir de la experiencia adquirida, basados en análisis de auto-correlación [46] o con la ayuda de los métodos de selección como ser *Ramdon Forest (RF), Mutual Information (MI), Partial Mutual Information (PMI), two-stage feature selection, Gibbs measure (GM), Firefly Algorithm (FA)*, entre otro se puede lograr una selección más automatizada.

Chandrashekar and Sahin [14] divide los métodos de selección de variables en dos clases *Filter* o *Wrapper methods*. Los algoritmos de la primer clase establecen un ranking de variables de acuerdo a un algoritmo o criterio en particular, y se establece un valor de umbral debajo del cual

las variables son ignoradas. Según los autores, estos métodos se reportan como exitosos en aplicaciones prácticas y relativamente simples de aplicar. Los métodos *Wrapper* por otro lado, ven el predictor como caja negra, y usan alguna variable relacionada a la performance del predictor como función objetivo. En el caso de codificación binaria, elegir un subconjunto adecuado es un problema *NP-Hard*, lo que justifica el uso de métodos heurísticos.

Entre los métodos de filtrado más conocidos podemos citar a *Mutual Information (MI)* y métodos basados en entropía que utilizan conceptos de teoría de la información, teoría ergódica, y entropía, para establecer una medida de dependencia estadística entre dos variables [14, 56]. En general éstos analizan la relación de información mutua entre pares de variables, por lo que suelen producir resultados pobres. Fleuret [23] propuso una mejora llamada *Conditional Mutual Information*, que es un método iterativo que evita seleccionar variables similares a algunas ya seleccionadas. Esto tiene en cuenta mejor la capacidad predictiva de las variables agregadas y su relación con todas las variables puestas hasta el momento, lo cual mejora los resultados [14, 23]. Una de las ventajas de los métodos de filtrado es que suelen ser computacionalmente muy eficientes [32].

Los métodos tipo *Wrapper* a su vez se pueden dividir en algoritmos secuenciales y algoritmos heurísticos. En los algoritmos secuenciales, una variable es elegida por iteración para incorporarse al conjunto de variables elegidas. Entre estos métodos podemos encontrar *Sequential Feature Selection*, un algoritmo *greedy* que comienza con el conjunto vacío y agrega cada nueva variable de acuerdo a su performance en la función objetivo. Similarmente, existe *Sequential Backward Selection* que elimina una variable por vez, eligiendo la que produzca un menor descenso en la performance del predictor. Finalmente una variante llamada *Sequential Floating Forward Selection (SFSS)* combina ambos métodos evitando introducir variables similares entre sí. Entre los algoritmos heurísticos se cuentan PSO, *Genetic Algorithm (GA)*, y otros. Los métodos se completan con los *Embedded Methods*, que combinan varios de los métodos anteriores en un esfuerzo por mejorar los resultados y bajar la cantidad de procesamiento necesario. Entre ellos se pueden contar métodos que utilizan MI para eliminar variables de entrada a una ANN; *Maximum Relevancy Minimun Redundancy (mRMR)* que combina MI con *Cross Validación* para elegir las k variables que produzcan el menor error de *Cross validación*, para luego aplicar un *wrapper method* para elegir el mejor subconjunto de k variables a incorporar.

Existen trabajos en la literatura que usan un clasificador como ANN o *Support Vector Machines (SVM)* para la clasificación, y luego basado en un análisis de los pesos conseguidos, realizan un ranking de las variables más importantes o directamente eliminan las de menor importancia. Entre estos podemos contar a SVM *Recursive Feature Elimination* [37, 59]. En esos casos se utiliza SVM como clasificador y se elimina recursivamente usando como ranking el cuadrado del peso encontrado por SVM. En todos los casos se realiza un preprocesamiento basado en algún procedimiento estadístico como MI o *distribución CHi cuadrado* en el caso de Li et al. [35] y Lin et al. [37], debido al alto número de variables (genética y bioquímica). De manera similar, en Romero and Sopena [45], se utiliza una red neuronal tipo *Multilayer Perceptrón* para la clasificación, y se eliminan las variables de manera recursiva procediendo a reentrenar la red luego de cada eliminación. Este procedimiento se conoce como *Sequential Backward Selection o Network Pruning* [49].

También hay autores que utilizan vectores de entrada que fueron definidos por otros autores. En varios trabajos analizados, muchas veces no se especifica como se eligieron las variables de entrada y en el peor de los casos ni se especifican de manera detallada, que variables de entrada se usaron, lo que dificultaría realizar los experimentos allí propuestos.

4.1. Empírico

En el trabajo de Fidalgo and Lopes [22], se plantea un método empírico para la selección de las variables del vector de entrada de las ANN. Los autores basaron su elección en el siguiente criterio:

1. Análisis de correlación entre el valor a ser pronosticado y los valores de las series temporales anteriores;
2. Análisis de sensibilidad de salida ANN con respecto a las entradas potenciales;
3. Resultado de los rendimientos de ANN;
4. El buen criterio del ingeniero. Los vectores de entrada elegidos se obtuvieron después de varios experimentos de ensayo y error con diferentes arquitecturas.

Pahasa and Theera-Umpon [42] en su trabajo tiene un apartado en donde especifican que es importante la selección de características, mencionando que la más importante es la de consumos en instantes de tiempo anteriores, y en el caso de Thailandia la temperatura tiene un gran impacto en los patrones de consumo. Además, dichos autores especifican el conjunto de variables de entrada ($L_{(0h)}$, $L_{(-24h)}$, $L_{(-144h)}$, $T_{(+24)}$), sin mencionar como hicieron esa selección.

4.2. Correlación Lineal

En la bibliografía examinada se utilizan métodos estadísticos, como los que consisten en analizar la correlación entre los valores que se deben predecir y los consumos

anteriores [22, 46]. También se realizan estudios de correlación con variables exógenas, como ser el clima [46] y los días feriados. En el artículo de Santos et al. [46], se analiza la correlación entre la predicción del consumo de *Active Power (P)*, con los consumos anteriores, además realiza un análisis de correlación de variables climáticas:

- Temperatura máxima y los picos de consumo eléctricos diarios.
- Temperatura promedio diaria y el promedio del consumo promedio diario de P.

Se observa que los valores de los coeficientes de autocorrelación, disminuyen rápidamente, en las horas contiguas. Esto es una consecuencia de la interdependencia entre los diferentes períodos del día. Paralelamente con el efecto anteriormente descripto también decrece el coeficiente de auto-correlación como consecuencia de diferentes situaciones climáticas. Todos estos efectos le permiten a los autores tomar la decisión de elegir solo incluir la información de las dos últimas semanas, en el vector de entrada de la ANN [46].

4.3. Filter

Los métodos de filtrado no tienen en cuenta el modelo utilizado para realizar STLF, estos métodos seleccionan el mejor subconjunto de entrada siguiendo un criterio independiente predefinido, que mide la relación entre las variables de entrada y la salida [32]. Tienen la ventaja de utilizar menos procesamiento que los métodos *wrapper*, pero suelen ser menos precisos.

4.3.1. Gibbs measure

En el artículo de Santos, Martins, and Pires [46], con el fin de evitar el exceso de parametrización del vector de entrada y elegir un número de variables continuas, los autores basaron su decisión en el concepto de *Gibbs measure (GM)*, para buscar valores de secuencias adyacentes y en un análisis de la entropía relativa.

El concepto de entropía que fue usado en la Equation 1, en donde μ, corresponde a la probabilidad de encontrar una secuencia de valores contiguos $p_1 \dots p_k$.

$$H_n = -\sum_{p_1 \dots p_k} \mu_{p_1 \dots p_k} \log(\mu_{p_1 \dots p_k}) \qquad (1)$$

Los autores realizaron el análisis de Gibbs measure (GM), en tres subestaciones con resultados muy similares. La series de la *Active Power (P)* puede ser descripta como un alfabeto finito S que consiste en secuencias contiguas infinitas:

$$? = p_1, p_2, \dots, p_k, \dots, p_k$$

La frecuencia del muestreo de los datos de los consumos eléctricos fue de una hora. Para identificar la eventual gramática se usaron los datos del período que va desde el 21/12/1998 hasta el 15/3/2001, que corresponde a una muestra de $N = 19584$ valores.

La señal se dividió de acuerdo al alfabeto: $= -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5$, en once niveles. Toda la señal es transformada al alfabeto.

El máximo número de secuencias diferentes en este ejemplo es $n = 4(11^4 N)$. Del análisis de GM los autores incluyeron la P de las dos horas anteriores.

4.3.2. Mutual Information

Mutual Information (MI) es una técnica desarrollada recientemente para la selección de variables o características. Esta técnica también es utilizada para seleccionar características en problemas de clasificación, como el reconocimiento de patrones, clasificación de los tipos de cáncer, procesamiento de imágenes médicas, clasificar conjuntos de muestras para entrenamiento que contienen ruido y valores atípicos, etc. [4, 10]. Es utilizada comúnmente para medir la dependencia entre dos variables aleatorias de tal manera que no requiere de conocimientos acerca de la naturaleza de sus relaciones subyacentes. Por lo tanto, MI es más potente, en algunos casos, que los estimadores que sólo consideran las relaciones lineales entre las variables [10]. De hecho, la MI entre dos variables aleatorias, X e Y, puede ser considerado como la medida de la cantidad de conocimiento sobre Y que es proporcionado por X (o la inversa de la cantidad sobre X proporcionado por Y). Si X e Y son independientes, por lo tanto X no contiene ninguna información sobre Y e viceversa; entonces la MI entre ellos es cero [57].

La definición de MI tiene sus orígenes en la Entropía de Shannon [50] en la Teoría de la Información [16].

Si MI entre dos variables aleatorias es grande, las dos variables están estrechamente relacionadas, y viceversa. Si MI se convierte en cero, las dos variables aleatorias no están relacionadas o las dos variables son independientes [4].

Mutual Information y la entropía tienen las siguientes relaciones, como se muestra en la Figura 2 [4].

Božić, Stojanović, Stajić, and Floranović [10], para realizar sus pruebas, utiliza la base de datos de *Elian Company*, esta base de datos está disponible para su descarga en ELIA [20]. En este trabajo se utilizaron los datos de Septiembre de 2011 para realizar la predicciones y como conjunto de entrenamiento los meses de Septiembre del los años 2008, 2009 y 2010. El vector de entrada de características candidatas está formado por los consumos de las anteriores 24 horas junto con la hora del día (numérico de 1 a 24) y el día de la semana (numérico de 1 a 7). En la arquitectura

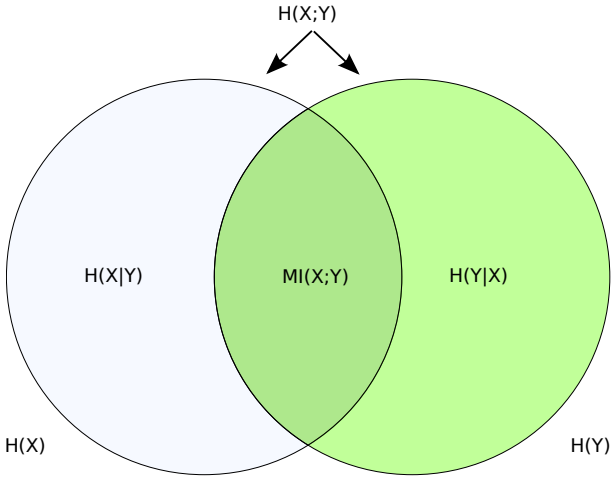

Figura 2. Representación de Mutual Information y las diferentes entropías

propuesta del algoritmo de selección de variables y la estrategia de predicción. El modelo propuesto permite definir dos parámetros: **MI theshold:** α y el **número de entradas:** r. Cuando se elige **MI theshold**, se debe asignar un valor al parámetro α. El valor de α determina la sensibilidad del algoritmo. Por otro lado si se elige la opción **número de entradas**, se debe asignar un valor al parámetro r. El valor de r define la cantidad total de muestras se usaran como conjunto de entrenamiento. En este trabajo tanto α como r fueron definidos de manera manual.

Utilizando el esquema planteado, definiendo o no cada parámetro, se plantean 4 modelos y 5 escenarios. Todos los modelos para realizar el cálculo de la predicción utilizan *Least Squares Support Vector Machines (LS-SVM)*, que es una reformulación de SVM. Las comparaciones se realizan utilizando el error MAPE. Se observa que el modelo M1 obtuvo el mejor promedio MAPE, en el conjunto de pruebas, en los escenarios NI_1, TH_1 y TH_2 mientras que el modelo M3 obtuvo los mejores resultados en el escenario NI_2. El autor concluye que comparado con otros método de selección, como ser el MI con estimador *kernel* y *Pearson correlation coefficient*, el método basado en MI con un estimador *k-Nearest Neighbor (kNN)*, logra una mejor exactitud en las predicciones, para la mayoría de los escenarios planteados [10].

Amjady, Keynia, and Zareipour [6] para predecir STLF en *microgrids*, utiliza un nueva estrategia de predicción en dos niveles, en el nivel superior utiliza una nueva técnica *Enhanced Differential Evolution (EDE)*, y en el nivel inferior primero realiza una selección de variables utilizando *two-stage feature selection*, de este proceso se obtiene las entradas para un sistema híbrido que realiza la predicción, compuesto de una ANN y un *Evolutionary Algorithm (EA)*. *Two-stage feature selection* utiliza dos filtros para seleccionar las variables, en la primera etapa filtra las características irrelevantes, quedando un conjunto de características relevantes, a este conjunto le aplica en la segunda etapa un filtro para eliminar las características redundantes. En Figura 3, se puede observar un diagrama de la estructura propuesta por Amjady and Keynia [4]. Los dos filtros utilizan *MI*, que se basa en el concepto de entropía, el método de selección es no-lineal y está desarrollado en un trabajo anterior del autor [4], en donde se presenta una nueva técnica para el pronóstico de la subida de los precios de los mercados eléctricos.

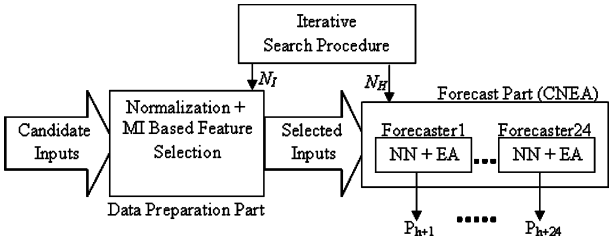

Figura 3. Estructura de la estrategia propuesta en Amjady and Keynia [4] en la Figura 2.

4.4. Wrapper

4.4.1. Random Forest

Según Lahouar and Slama [33], el modelo de *Ramdon Forest (RF)* utilizado para realizar STLF, no ha recibido mucha atención, a pesar de su exactitud, rapidez y sobre todo la facilidad a la hora de elegir sus parámetros. Este modelo es capaz de dar resultados tan buenos como una ANN optimizada y SVM, sin la necesidad de realizar un ajustes fino, e incluso una mala elección de sus parámetros no afecta drásticamente la precisión de la predicción. En este trabajo los autores lo utilizan para genera un modelo genérico, para predecir el consumo eléctrico, con un horizonte de un día, la frecuencia del muestreo es de una hora. Para realizar la predicción utilizaron los datos reales de una Compañía Eléctrica de Túnez, la elección de este conjuntos de datos se baso en que tiene una curva de consumo que es especifico de países cálidos en donde en el verano los consumos son elevados e inestables [33].

Ramdon Forest (RF) es una generalización de los árboles de decisión propuestos por Breiman en el 2001, el desarrollo matemático completo de *Ramdon Forest (RF)*, se puede consultar en Breiman [11] y en Breiman, Friedman, Stone, and Olshen [12].

En Cheng, Chan, and Qiu [15] los autores, plantean, que en la selección de variables, las mejores características no conducen necesariamente a una buena selección. Debido a

que la información proporcionada por las variables eliminadas se pierde. En este trabajo se aplica la teoría de *Ramdon Forest (RF)* combinada con *Ensemble System* para realizar la predicción del consumo eléctrico.

La ventaja de usar *Ensemble System* queda demostrada en los problemas que involucran un gran conjunto de clases y características de entradas complejas tales como STLF. Por este motivo Cheng, Chan, and Qiu [15] utilizan Ramdon Forest (RF) como método para reducir el conjunto de variables de entrada.

En la validación de su propuesta, utilizaron dos conjuntos de datos reales, bien conocidos, NYISO [41] y PJM [44]. Este método es comparado con *Generalized Regression Neural Network (GRNN)* y *Back Propagation Neural Network (BPNN)*, usando dos modelos de selección de variables basados en *Mutual Information Theory*: *One-stage* y *two-stage feature selection*. Las comparaciones experimentales muestran que el método basado en RF, supera a los otros métodos en términos de precisión y estabilidad [15].

En Liaw and Wiener [36], se realiza una breve introducción en el uso del paquete de software *randomForest*, este software se encuentra en `http://www.stat.berkeley.edu/users/breiman/`, en este trabajo se presenta una descripción del algoritmo utilizado tanto para clasificación como para regresión.

Una ventaja importante que tiene este paquete de software es que permite obtener dos informaciones adicionales:

1. Mide que tan importante es la variable en la predicción, Figura 4.

2. Mide la estructura interna de los datos (que tan próximos están los diferentes datos entre sí).

Otra de las características importantes de este algoritmo que es altamente paralelizable, se pueden correr varios *Ramdon Forest (RF)*, en diferentes computadoras y luego agregar el componente de los votos para obtener el resultado final. Para futuros trabajos con el método *Ramdon Forest (RF)* de predicción en el artículo de Liaw and Wiener [36], presenta algunos consejos prácticos en el apartado Liaw and Wiener [36, Some notes for practical use].

4.5. Mixto

Hu, Bao, Xiong, and Chiong [32], proponen un método mixto. En primer lugar utilizan *Partial Mutual Information (PMI)*, que es un método de *filter*, para filtrar las características que son irrelevantes y redundantes, del conjunto original de variables, de esta manera logran una significativa reducción del conjunto inicial de variables candidatas. Entonces un proceso *wrapper* utilizando *Firefly Algorithm (FA)*, se realiza sobre el conjunto de características que fueron filtradas anteriormente, este conjunto de características

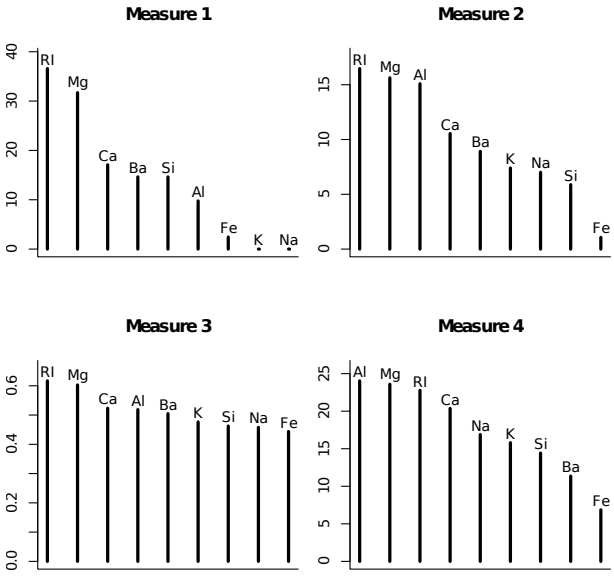

Figura 4. Importancia de las variables para el conjunto de datos *Forensic Glass*. Figura 1 del trabajo de Liaw and Wiener [36].

reducidas produce una fuerte disminución del costo computacional del algoritmo *wrapper*, permitiendo que se pueda aplicar de manera práctica. Ver Figura 5. *Firefly Algorithm (FA)* es un método *wrapper*, que imita el comportamiento social de las luciérnagas, estas se mueven y se comunican entre sí en función de sus características intermitentes como el brillo y la frecuencia.

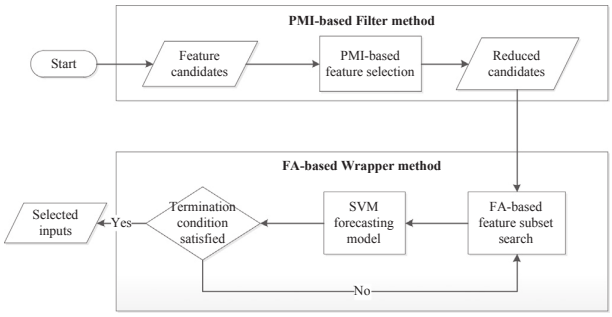

Figura 5. Diagrama de flujo del método de selección de variables *filter-wrapper* propuesto por Hu, Bao, Xiong, and Chiong [32]

Partial Mutual Information (PMI) es un implementación no paramétrica del criterio de MI, una forma de medir la dependencia entre variables. Este criterio se basa en una caracterización de la distribución de probabilidad conjunta. Este implementación utiliza un criterio de MI "parcial" como base para identificar más de un predictor de una ma-

nera escalonada. Utiliza métodos no paramétricos de *kernel* para caracterizar la distribución de probabilidad conjunta de las variables involucradas [51].

Partial Mutual Information (PMI), es nuevo en la STLF [32]. La PMI entre la variable de salida Y y la variable de entrada X, para un conjunto de predictores ya seleccionados Z, se puede definir como:

$$PMI(X,Y) = \int\int \mu^{X',Y'}(x',y') \log \frac{\mu^{X',Y'}(x',y')}{\mu^{X'}(x')\mu^{Y'}(y')} \mathrm{d}x'\mathrm{d}y' \tag{2}$$

con

$$x' = x - E[x|z] \;\; y' = y - E[y|z] \tag{3}$$

donde $E[\cdot]$ denota la operación de expectativa $\mu^{X',Y'}, \mu(x',y')$ y $\mu^{X,Y}$(x',y'), que son la respectiva probabilidad univariada y conjunta de las densidades estimadas en los puntos de los datos de la muestra. La función de densidad de probabilidad usada como estimador es *city block distance kernel*. Las variables x' e y', representan la información residual de las variables x y y una vez que el efecto de los predictores existentes z fue tenido en cuenta. Cuando se realiza una selección de variables basada en PMI, la variable de entrada que agrega el valor más alto de PMI, es agregada como un nuevo factor del predictor[32]. Más detalles del método PMI, se pude encontrar en Sharma [51].

5. Conclusiones

En la actualidad para realizar la gestión de la energía eléctrica se necesita de predicciones precisas y disponibles para su uso práctico. La cantidad de datos que generan los sistemas SCADA, aumenta considerablemente, tanto en la frecuencia como en cantidad de variables que manejan, por este motivo es necesario contar con sistemas de selección de variables para eliminar los datos redundantes, o que aportan poca información a los sistemas de predicción.

Algunos de los modelos de predicción, no pueden manejar de manera práctica muchas variables de entrada. En estos modelos es esencial solo seleccionar un reducido número de variables. Por otro lado se deben eliminar las variables que agreguen ruido para lograr mejores predicciones.

En las regiones cálidas, la curva del consumo eléctrico, presenta consumos excesivos e inestables en el verano [33]. En estas regiones se utilizan en gran escala sistemas de Heating, Ventilating and Air Conditioning (HVAC) o aires acondicionados, por este motivo es de esperar una alta correlación entre el consumo eléctrico y las variables exógenas climáticas, debido a la elevada humedad y temperatura.

En el caso de que se tenga distintos sistemas para predecir determinadas horas o intervalos de tiempo, en los intervalos en donde no exista una incidencia de los sistemas de HVAC, se deberían excluir las variables exógenas climáticas ya que presentan una baja correlación [19, 46]. En Taylor, de Menezes, and McSharry [55], utilizan modelos univariables (solo el consumo eléctrico), para horizontes de predicción de hasta 6 horas, debido a que los datos climáticos para intervalos menores a 6 horas son difíciles de conseguir y costosos.

Es necesario incluir la mayor información específica y relevante, del instante que se intenta predecir, (como ser si es un fin de semana, si es un feriado, si se tiene una buena predicción del clima de ese instante, que día de la semana es, si es de día o de noche, etc.). En el trabajo de Taylor and Buizza [53], realizan una predicción del clima, de esta manera logran mejorar el MAPE de la predicción. Con el uso de variables climáticas en los dos trabajos [53, 54], los autores observaron que mejoraron los resultados de los modelos que utilizaban variables climáticas, pero los mejores resultados se obtuvieron utilizando la predicción del clima.

Utilizar variables binarias para el caso de días feriados [13]. En el caso de variable tipo tiempo como ser hora, semana, mes, año, intentar usar variables homólogas generadas de coseno o seno, aunque en los períodos iniciales del proyecto sea mejor el uso de los valores casi puros para que sea más fácil el análisis de los valores atípicos u *outliers*. También se debería intentar con el uso de variables binarias [13].

En el caso de modelos de predicción basados en ANN, es necesario reducir la cantidad de variables de entrada. Se debe tener en cuenta que las mejores variables individuales puede no ser las mejores como conjunto de entrada. En este conjunto pueden existir variables redundantes.

Por otro lado, en el trabajo de Taylor and Buizza [54], los autores hacen la observación que los resultados del error MAPE, en la estimación con un horizonte de un 1 día, existe muy poca diferencia entre los rendimientos de los distintos métodos utilizados, este comportamiento también se observa en su otra investigación Taylor and Buizza [53]).

Por otra parte, para saber cuánto se puede mejorar la predicción del consumo eléctrico, si se puede contar con una predicción de alguna variable exógena (por ejemplo variables climáticas), se debe utilizar el valor real la variable exógena de ese momento futuro, de esta manera podremos determinar, cuanto es lo máximo que puede ayudar al sistema de predicción si se cuenta con una predicción precisa de la misma.

En futuros trabajos, se investigará cuales son los métodos más precisos para realizar el cálculo de STLF, en regiones con marcadas variaciones climáticas, en donde los habitantes usan sistemas HVAC.

6. Agradecimientos

Este trabajo fue soportado parcialmente por subsidios asignados a los proyectos "Análisis Inteligente de Datos Aplicado a Gestión y Optimización de la Energía" (UTN3870) y "Gestión y Optimización Inteligente de la Energía II" (UTI4015).

7. Referencias

[1] T. Al-Saba and I. El-Amin. Artificial neural networks as applied to long-term demand forecasting. *Artificial Intelligence in Engineering*, 13(2):189–197, 1999. ISSN 0954-1810. doi: 10.1016/S0954-1810(98)00018-1. URL http://www.sciencedirect.com/science/article/pii/S0954181098000181.

[2] H. K. Alfares and M. Nazeeruddin. Electric load forecasting: Literature survey and classification of methods. *International Journal of Systems Science*, 33(1):23–34, 2002. ISSN 0020-7721. doi: 10.1080/00207720110067421. URL http://dx.doi.org/10.1080/00207720110067421.

[3] N. Amjady. Short-Term Bus Load Forecasting of Power Systems by a New Hybrid Method. *IEEE Transactions on Power Systems*, 22(1):333–341, Feb. 2007. ISSN 0885-8950. doi: 10.1109/TPWRS.2006.889130.

[4] N. Amjady and F. Keynia. Day-Ahead Price Forecasting of Electricity Markets by Mutual Information Technique and Cascaded Neuro-Evolutionary Algorithm. *IEEE Transactions on Power Systems*, 24(1):306–318, Feb. 2009. ISSN 0885-8950. doi: 10.1109/TPWRS.2008.2006997.

[5] N. Amjady and F. Keynia. Electricity market price spike analysis by a hybrid data model and feature selection technique. *Electric Power Systems Research*, 80(3):318–327, Mar. 2010. ISSN 0378-7796. doi: 10.1016/j.epsr.2009.09.015. URL http://www.sciencedirect.com/science/article/pii/S0378779609002260.

[6] N. Amjady, F. Keynia, and H. Zareipour. Short-Term Load Forecast of Microgrids by a New Bilevel Prediction Strategy. *IEEE Transactions on Smart Grid*, 1(3):286–294, 2010. ISSN 1949-3053. doi: 10.1109/TSG.2010.2078842.

[7] A. Azadeh, S. F. Ghaderi, S. Tarverdian, and M. Saberi. Integration of artificial neural networks and genetic algorithm to predict electrical energy consumption. *Applied Mathematics and Computation*, 186(2):1731–1741, Mar. 2007. ISSN 0096-3003. doi: 10.1016/j.amc.2006.08.093. URL http://www.sciencedirect.com/science/article/pii/S0096300306011088.

[8] J. M. Benavente, A. Galetovic, R. Sanhueza, and P. Serra. Estimando la Demanda Residencial por Electricidad en Chile: El Consumo es Sensible al Precio. *Cuadernos de economía*, 42(125):31–61, May 2005. ISSN 0717-6821. doi: 10.4067/S0717-68212005012500002. URL http://www.scielo.cl/scielo.php?script=sci_abstract&pid=S0717-68212005012500002&lng=es&nrm=iso&tlng=es.

[9] A. L. Blum and P. Langley. Selection of relevant features and examples in machine learning. *Artificial Intelligence*, 97(1–2):245–271, 1997. ISSN 0004-3702. doi: 10.1016/S0004-3702(97)00063-5. URL http://www.sciencedirect.com/science/article/pii/S0004370297000635.

[10] M. Božić, M. Stojanović, Z. Stajić, and N. Floranović. Mutual Information-Based Inputs Selection for Electric Load Time Series Forecasting. *Entropy*, 15(3):926–942, Feb. 2013. doi: 10.3390/e15030926. URL http://www.mdpi.com/1099-4300/15/3/926.

[11] L. Breiman. Random forests. *Machine learning*, 45 (1):5–32, 2001. URL http://link.springer.com/article/10.1023/A:1010933404324.

[12] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen. *Classification and Regression Trees*. Chapman and Hall/CRC, edición: new ed edition, 1984. ISBN 978-0-412-04841-8.

[13] E. Ceperic, V. Ceperic, and A. Baric. A Strategy for Short-Term Load Forecasting by Support Vector Regression Machines. *IEEE Transactions on Power Systems*, 28(4):4356–4364, Nov. 2013. ISSN 0885-8950. doi: 10.1109/TPWRS.2013.2269803.

[14] G. Chandrashekar and F. Sahin. A survey on feature selection methods. *Computers & Electrical Engineering*, 40(1):16–28, Jan. 2014. ISSN 00457906. doi: 10.1016/j.compeleceng.2013.11.024. URL http://linkinghub.elsevier.com/retrieve/pii/S0045790613003066.

[15] Y.-Y. Cheng, P. Chan, and Z.-W. Qiu. Random forest based ensemble system for short term load fo-

recasting. In *2012 International Conference on Machine Learning and Cybernetics (ICMLC)*, volume 1, pages 52–56, July 2012. doi: 10.1109/ICMLC.2012.6358885.

[16] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Nov. 2012. ISBN 978-1-118-58577-1.

[17] I. Drezga and S. Rahman. Input variable selection for ANN-based short-term load forecasting. *IEEE Transactions on Power Systems*, 13(4):1238–1244, Nov. 1998. ISSN 0885-8950. doi: 10.1109/59.736244.

[18] F. Egelioglu, A. A. Mohamad, and H. Guven. Economic variables and electricity consumption in Northern Cyprus. *Energy*, 26 (4):355–362, 2001. ISSN 0360-5442. doi: 10.1016/S0360-5442(01)00008-1. URL http://www.sciencedirect.com/science/article/pii/S0360544201000081.

[19] A. A. El-Keib, X. Ma, and H. Ma. Advancement of statistical based modeling techniques for short-term load forecasting. *Electric Power Systems Research*, 35(1):51–58, Oct. 1995. ISSN 0378-7796. doi: 10.1016/0378-7796(95)00987-6. URL http://www.sciencedirect.com/science/article/pii/0378779695009876.

[20] ELIA. Elia, 0. URL http://www.elia.be/en/grid-data/data-download.

[21] J. Y. Fan and J. D. McDonald. A real-time implementation of short-term load forecasting for distribution power systems. *Power Systems, IEEE Transactions on*, 9(2):988–994, 1994. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=317646.

[22] J. N. Fidalgo and J. Lopes. Load forecasting performance enhancement when facing anomalous events. *IEEE Transactions on Power Systems*, 20(1):408–415, Feb. 2005. ISSN 0885-8950. doi: 10.1109/TPWRS.2004.840439.

[23] F. Fleuret. Fast Binary Feature Selection with Conditional Mutual Information. *J. Mach. Learn. Res.*, 5:1531–1555, Dec. 2004. ISSN 1532-4435. URL http://dl.acm.org/citation.cfm?id=1005332.1044711.

[24] L. Friedrich and A. Afshari. Short-term Forecasting of the Abu Dhabi Electricity Load Using Multiple Weather Variables. *Energy Procedia*, 75:3014–3026, 2015. ISSN 1876-6102. doi: 10.1016/j.egypro.2015.07.616. URL http://www.sciencedirect.com/science/article/pii/S1876610215013843.

[25] M. Ghiassi, D. K. Zimbra, and H. Saidane. Medium term system load forecasting with a dynamic artificial neural network model. *Electric Power Systems Research*, 76(5):302–316, Mar. 2006. ISSN 0378-7796. doi: 10.1016/j.epsr.2005.06.010. URL http://www.sciencedirect.com/science/article/pii/S0378779605001951.

[26] G. Gross and F. Galiana. Short-term load forecasting. *Proceedings of the IEEE*, 75(12):1558–1573, 1987. ISSN 0018-9219. doi: 10.1109/PROC.1987.13927. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1458194.

[27] Y.-C. Guo. An integrated PSO for parameter determination and feature selection of SVR and its application in STLF. In *2009 International Conference on Machine Learning and Cybernetics*, volume 1, pages 359–364, July 2009. doi: 10.1109/ICMLC.2009.5212569.

[28] Y.-J. He, Y.-C. Zhu, D.-X. Duan, and W. Sun. Application of Neural Network Model Based on Combination of Fuzzy Classification and Input Selection in Short Term Load Forecasting. In *2006 International Conference on Machine Learning and Cybernetics*, pages 3152–3156, 2006. doi: 10.1109/ICMLC.2006.258409.

[29] L. Hernandez, C. Baladron, J. M. Aguiar, B. Carro, A. J. Sanchez-Esguevillas, J. Lloret, and J. Massana. A Survey on Electric Power Demand Forecasting: Future Trends in Smart Grids, Microgrids and Smart Buildings. *IEEE Communications Surveys Tutorials*, 16(3):1460–1495, 2014. ISSN 1553-877X. doi: 10.1109/SURV.2014.032014.00094.

[30] H. Hippert, C. Pedreira, and R. Souza. Neural networks for short-term load forecasting: a review and evaluation. *IEEE Transactions on Power Systems*, 16 (1):44–55, Feb. 2001. ISSN 0885-8950. doi: 10.1109/59.910780.

[31] R. R. Hocking. A Biometrics Invited Paper. The Analysis and Selection of Variables in Linear Regression. *Biometrics*, 32(1):1–49, 1976. ISSN 0006-341X. doi: 10.2307/2529336. URL http://www.jstor.org/stable/2529336.

[32] Z. Hu, Y. Bao, T. Xiong, and R. Chiong. Hybrid filter–wrapper feature selection for short-term load forecasting. *Engineering Applications of Artificial Intelligence*, 40:17–27, 2015. ISSN 0952-1976.

doi: 10.1016/j.engappai.2014.12.014. URL http://www.sciencedirect.com/science/article/pii/S0952197614003066.

[33] A. Lahouar and J. B. H. Slama. Random forests model for one day ahead load forecasting. In *Renewable Energy Congress (IREC), 2015 6th International*, pages 1–6, Mar. 2015. doi: 10.1109/IREC.2015.7110975.

[34] K. Li, N. Tai, and S. Zhang. Research and application of climatic sensitive short - term load forecasting. In *2015 IEEE Power Energy Society General Meeting*, pages 1–5, July 2015. doi: 10.1109/PESGM.2015.7286315.

[35] Z. Li, J. Liu, Y. Yang, X. Zhou, and H. Lu. Clustering-Guided Sparse Structural Learning for Unsupervised Feature Selection. *IEEE Transactions on Knowledge and Data Engineering*, 26(9):2138–2150, Sept. 2014. ISSN 1041-4347. doi: 10.1109/TKDE.2013.65.

[36] A. Liaw and M. Wiener. Classification and Regression by randomForest. *R News*, 2(3):18–22, 2002. URL ftp://ftp.stat.math.ethz.ch/CRAN/doc/Rnews/Rnews_2002-3.pdf.

[37] X. Lin, F. Yang, L. Zhou, P. Yin, H. Kong, W. Xing, X. Lu, L. Jia, Q. Wang, and G. Xu. A support vector machine-recursive feature elimination feature selection method based on artificial contrast variables and mutual information. *Journal of Chromatography B*, 910:149–155, Dec. 2012. ISSN 1570-0232. doi: 10.1016/j.jchromb.2012.05.020. URL http://www.sciencedirect.com/science/article/pii/S1570023212002929.

[38] D. Lizondo, V. A. Jimenez, F. Villacis Postigo, A. Will, and S. Rodriguez. Análisis de Variables Temporales para la Predicción del Consumo Eléctrico. *Revista Técnica Energía*, 2015. ISSN 1390-5074.

[39] R. Mamlook, O. Badran, and E. Abdulhadi. A fuzzy inference model for short-term load forecasting. *Energy Policy*, 37(4):1239–1248, 2009. ISSN 0301-4215. doi: 10.1016/j.enpol.2008.10.051. URL http://www.sciencedirect.com/science/article/pii/S0301421508006563.

[40] I. Moghram and S. Rahman. Analysis and evaluation of five short-term load forecasting techniques. *IEEE Transactions on Power Systems*, 4(4):1484–1491, Nov. 1989. ISSN 0885-8950. doi: 10.1109/59.41700.

[41] NYISO. NYISO, 0. URL http://www.nyiso.com.

[42] J. Pahasa and N. Theera-Umpon. Cross-substation short term load forecasting using support vector machine. In *5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, 2008. ECTI-CON 2008*, volume 2, pages 953–956. IEEE, May 2008. doi: 10.1109/ECTICON.2008.4600589. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4600589.

[43] P.-F. Pai and W.-C. Hong. Forecasting regional electricity load based on recurrent support vector machines with genetic algorithms. *Electric Power Systems Research*, 74(3):417–425, June 2005. ISSN 0378-7796. doi: 10.1016/j.epsr.2005.01.006. URL http://www.sciencedirect.com/science/article/pii/S0378779605000702.

[44] PJM. PJM, 0. URL http://www.pjm.com/.

[45] E. Romero and J. M. Sopena. Performing Feature Selection With Multilayer Perceptrons. *IEEE Transactions on Neural Networks*, 19(3):431–441, Mar. 2008. ISSN 1045-9227. doi: 10.1109/TNN.2007.909535.

[46] P. J. Santos, A. Martins, and A. Pires. Short-term load forecasting based on ANN applied to electrical distribution substations. In *Universities Power Engineering Conference, 2004. UPEC 2004. 39th International*, volume 1, pages 427–432 Vol. 1, Sept. 2004.

[47] P. J. Santos, A. G. Martins, and A. J. Pires. Designing the input vector to ANN-based models for short-term load forecast in electricity distribution systems. *International Journal of Electrical Power & Energy Systems*, 29(4):338 – 347, 2007. ISSN 0142-0615. doi: http://dx.doi.org/10.1016/j.ijepes.2006.09.002. URL http://www.sciencedirect.com/science/article/pii/S0142061506001670.

[48] T. Senjyu, P. Mandal, K. Uezato, and T. Funabashi. Next day load curve forecasting using recurrent neural network structure. *IEE Proceedings - Generation, Transmission and Distribution*, 151(3):388–394, May 2004. ISSN 13502360. doi: 10.1049/ip-gtd:20040356. URL http://digital-library.theiet.org/content/journals/10.1049/ip-gtd_20040356.

[49] R. Setiono and H. Liu. Neural-network feature selector. *IEEE Transactions on Neural Networks*, 8 (3):654–662, May 1997. ISSN 1045-9227. doi: 10.1109/72.572104.

[50] C. E. Shannon. Communication Theory of Secrecy Systems*. *Bell System Technical Journal*, 28(4):656–715, Oct. 1949. ISSN 1538-7305. doi: 10.1002/j.1538-7305.1949.tb00928.x. URL http://onlinelibrary.wiley.com/doi/10.1002/j.1538-7305.1949.tb00928.x/abstract.

[51] A. Sharma. Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 1 — A strategy for system predictor identification. *Journal of Hydrology*, 239 (1–4):232–239, 2000. ISSN 0022-1694. doi: 10.1016/S0022-1694(00)00346-2. URL http://www.sciencedirect.com/science/article/pii/S0022169400003462.

[52] A. K. Singh, S. K. Ibraheem, and M. Muazzam. An Overview of Electricity Demand Forecasting Techniques. *Network and Complex Systems*, 3(3):38–48, 2013. ISSN 2225-0603. URL http://iiste.org/Journals/index.php/NCS/article/view/6072.

[53] J. Taylor and R. Buizza. Neural network load forecasting with weather ensemble predictions. *IEEE Transactions on Power Systems*, 17(3):626–632, 2002. ISSN 0885-8950. doi: 10.1109/TPWRS.2002.800906.

[54] J. W. Taylor and R. Buizza. Using weather ensemble predictions in electricity demand forecasting. *International Journal of Forecasting*, 19(1):57–70, 2003. ISSN 0169-2070. doi: 10.1016/S0169-2070(01)00123-6. URL http://www.sciencedirect.com/science/article/pii/S0169207001001236.

[55] J. W. Taylor, L. M. de Menezes, and P. E. McSharry. A comparison of univariate methods for forecasting electricity demand up to a day ahead. *International Journal of Forecasting*, 22(1):1–16, 2006. ISSN 0169-2070. doi: 10.1016/j.ijforecast.2005.06.006. URL http://www.sciencedirect.com/science/article/pii/S0169207005000907.

[56] K. Torkkola. Feature Extraction by Non Parametric Mutual Information Maximization. *J. Mach. Learn. Res.*, 3:1415–1438, Mar. 2003. ISSN 1532-4435. URL http://dl.acm.org/citation.cfm?id=944919.944981.

[57] A. H. Vahabie, M. Yousefi, B. Araabi, C. Lucas, S. Barghinia, and P. Ansarimehr. Mutual Information Based Input Selection in Neuro-Fuzzy Modeling for Short Term Load Forecasting of Iran National Power System. In *IEEE International Conference on Control and Automation, 2007. ICCA 2007*, pages 2710–2715, May 2007. doi: 10.1109/ICCA.2007.4376854.

[58] M. M. B. R. Vellasco, M. A. C. Pacheco, L. S. R. Neto, and F. J. de Souza. Electric load forecasting: evaluating the novel hierarchical neuro-fuzzy BSP model. *International Journal of Electrical Power & Energy Systems*, 26(2):131–142, Feb. 2004. ISSN 0142-0615. doi: 10.1016/S0142-0615(03)00060-7. URL http://www.sciencedirect.com/science/article/pii/S0142061503000607.

[59] S. Wang, X. Chang, X. Li, Q. Z. Sheng, and W. Chen. Multi-task support vector machines for feature selection with shared knowledge discovery. *Signal Processing*, 120:746–753, Mar. 2016. ISSN 0165-1684. doi: 10.1016/j.sigpro.2014.12.012. URL http://www.sciencedirect.com/science/article/pii/S016516841400574X.

[60] Z. Wang, C. X. Guo, and Y. Cao. A new method for short-term load forecasting integrating fuzzy-rough sets with artificial neural network. In *Power Engineering Conference, 2005. IPEC 2005. The 7th International*, pages 1–173, Nov. 2005. doi: 10.1109/IPEC.2005.206900.

8. Glosario

Siglas

ANN Artificial Neural Networks. 2, 4–6, 8

BPNN Back Propagation Neural Network. 7

FA Firefly Algorithm. 3, 7

GA Genetic Algorithm. 4

GM Gibbs measure. 3, 5

GRNN Generalized Regression Neural Network. 7

HVAC Heating, Ventilating and Air Conditioning. 1, 8

kNN k-Nearest Neighbor. 6

LS-SVM Least Squares Support Vector Machines. 6

MAPE Maximal Absolute Percentage Error. 3, 6, 8

MI Mutual Information. 3–7

P Active Power. 5

PMI Partial Mutual Information. 3, 7, 8

PSO Particle Swarm Optimization. 2, 4

RF Ramdon Forest. 3, 6, 7

SCADA Supervisory Control And Data Acquisition. 1, 8

SteepwiseFS Forward Selection with Stepwise Regression. 2

STLF Short-Term Load Forecating. 1–3, 5–8

SVM Support Vector Machines. 4, 6

SVR Support Vector Regression Machines. 2

Glosario

microgrid A microgrid is an electrical system that includes multiple loads and distributed energy resources that can be operated in parallel with the broader utility grid or as an electrical island. 1, 6

two-stage feature selection Método de selección de variables que primero filtra las características menos relevantes y en la segunda etapa filtra las características similares. 3, 6, 7