

APLICACIÓN DE REDES NEURONALES PARA LA PREDICCIÓN DE PROPIEDADES TERMODINÁMICAS

Micael Gerardo Bravo Sánchez

Instituto Tecnológico de Celaya
gerardo.bravo@itcelaya.edu.mx

Marco Carlo Guerrero Soto

Instituto Tecnológico de Celaya

Juan José Martínez Nolasco

Instituto Tecnológico de Celaya

Nallely Rodríguez Trejo

Instituto Tecnológico de Celaya

Resumen

En los últimos años la necesidad de contar con información experimental confiable para el diseño y la optimización de procesos industriales ha tenido una creciente demanda. Los equipos de medición son extremadamente caros y requieren de personal técnico calificado para su manipulación. Por lo cual, se han dirigido esfuerzos a la simulación de propiedades empleando diferentes metodologías y herramientas. En este trabajo se presenta la aplicación de las redes neuronales para la predicción de propiedades termodinámicas. Esta investigación en particular se centra en la predicción de equilibrio de fases. La red neuronal que se diseño es una red de retropropagación multicapa (backpropagation multilayer) programada en Matlab. El entrenamiento fue realizado por medio de Levenberg-Marquardt (`trainlm`), empleado datos de equilibrio de fases de mezclas binarias de n-alcanos + n-alcoholes, y evaluando la divergencia entre los datos

predecidos por la red neuronal y los datos experimentales. La propiedad termodinámica que se evaluó fue la fugacidad, calculada a partir de la ecuación de PC-SAFT. La divergencia entre las predicciones de la red neuronal y la ecuación de estado PC-SAFT fue de 0.05%. Y la divergencia entre las predicciones de la red neuronal y los datos experimentales fue menor al 0.05%.

Palabra(s) Clave(s): Sustancias, Redes, Error, Ecuación.

1. Introducción

Las redes de neuronas artificiales (denominadas habitualmente como RNA o en inglés como: "ANN") son un paradigma de aprendizaje y procesamiento automático inspirado en la forma en que funciona el sistema nervioso de los animales. Se trata de un sistema de interconexión de neuronas que colaboran entre sí para producir un estímulo de salida.

Las redes neuronales artificiales son un conjunto de técnicas pertenecientes al campo de la inteligencia artificial. Su estructura consiste en una red formada por nodos (o neuronas) y conexiones, razón por la cual se asemejan al cerebro de los seres humanos, del cual procede su nombre. Las redes neuronales son de aplicación en diversidad de problemas de reconocimiento de patrones y de aproximación de funciones, debido a su flexibilidad y facilidad de uso.

Funcionamiento de una red neuronal

Una red neuronal es capaz de detectar relaciones complejas y no lineales entre variables, a partir de unidades sencillas como las neuronas, al disponer muchas de estas unidades en paralelo. Las variables se dividen en variables de entrada y de salida, relacionadas por algún tipo de correlación o dependencia (no necesariamente causa-efecto). También

es posible que la salida sea la clasificación de las variables de entrada en diferentes grupos.

Las neuronas se pueden disponer en diferentes capas. Las redes neuronales más sencillas constan de una capa de entrada, una capa de neuronas o capa oculta, y una capa de salida (Figura 1).

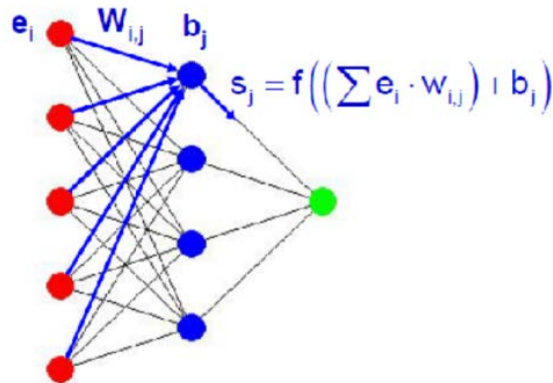


Figura 1. Ejemplo de red neuronal con una capa oculta.

El funcionamiento de una neurona consiste en la transformación de los valores de las entradas a través de las conexiones, en una salida. La salida se obtiene a partir de una función de propagación, una función de activación, y una función de transferencia (Figura 2).

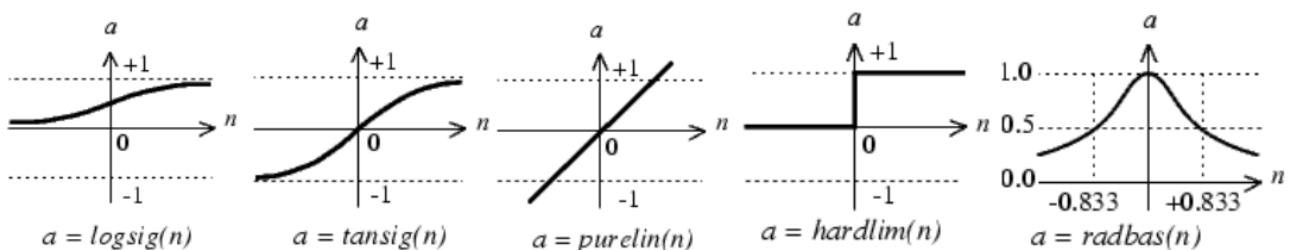


Figura 2. Funciones de transferencia.

- La función de propagación más común consiste en el sumatorio de todas las entradas multiplicadas por los pesos de las conexiones, más un valor de sesgo o “bias”.
- La función de activación, en caso de que exista, activa o desactiva la salida de esta neurona.
- La función de transferencia se aplica al resultado de la función de propagación y normalmente consiste en una función de salida acotada como la sigmoidea (logsig) [0,1], o la tangente hiperbólica (tansig) [-1,1]. Otras funciones de transferencia pueden ser una función lineal (purelin) $[-\infty, +\infty]$, base radial (radbas) [0,1] o una función de discriminación (hardlim) [0,1].

Tipos de redes neuronales

Los criterios más importantes para clasificar las redes neuronales son:

- Según el tipo de conexiones:
 - Redes de propagación hacia delante (feed-forward), donde las conexiones van en un solo sentido desde la capa de entrada hacia la capa de salida.
 - Redes recurrentes, donde las conexiones pueden realizar ciclos.
- Según el tipo de aprendizaje:
 - Aprendizaje supervisado. Los datos (o entradas) tienen una respuesta conocida (o salida), con la cual se ajusta o entrena la red neuronal.
 - Aprendizaje no supervisado o autoorganizado. Los datos son solamente entradas. Son redes empleadas fundamentalmente para clasificación y reconocimiento de patrones.

Estas páginas se centrarán en la aplicación de redes neuronales de propagación hacia adelante, con aprendizaje supervisado, empleadas en la aproximación de funciones.

Entrenamiento de la red neuronal

Dada una estructura y tamaño de la red neuronal, se procede al entrenamiento de la red. El entrenamiento o aprendizaje, cuyo objetivo es que la red neuronal sea capaz de reproducir el comportamiento subyacente en los datos aportados, consiste básicamente en la minimización de una función de coste o error, lo que equivale a que la salida de la red, se aproxima a la salida en los datos. La función de coste más común es la de promedio de errores al cuadrado (MSE).

Para la optimización de la red neuronal, se emplean diferentes métodos de ajuste de parámetros de la red (pesos de las conexiones y sesgo de las neuronas), a partir de unos valores o bien aleatorios, o bien predefinido (inicialización de la red). Algunos ejemplos de los métodos de ajuste son los de tipo gradiente o los algoritmos genéticos:

- Los métodos de tipo gradiente calculan la variación del error al variar cada uno de los parámetros (a modo de derivada multidimensional), y luego modifican todos los parámetros de la red neuronal obteniendo un error menor. Se puede decir que es una búsqueda en serie de la solución o mínimo global.
- Los métodos basados en algoritmos genéticos, consisten en la generación de un determinado número de redes o hijos a través de mutaciones en los parámetros, evaluando el error de la red para cada uno de ellos. Los hijos con menor error, tienen mayor probabilidad de convertirse en padres de nuevas redes, mientras que los hijos con mayor error desaparecen. Se trata de una búsqueda en paralelo de la solución.
- Ambos métodos son métodos iterativos, que se repiten hasta cumplir alguno de los diferentes criterios de parada. Algunos ejemplos de los criterios de parada son el número de iteraciones, la obtención de un error mínimo, o un tiempo de ejecución. En cualquier caso, generalmente es difícil asegurar que la solución obtenida no es un mínimo local.

Sobreaprendizaje

Un posible problema del proceso de entrenamiento, es el sobreaprendizaje o pérdida de generalización. Dado un conjunto de datos, es posible que la red neuronal reproduzca muy bien el comportamiento de dichos datos, pero no el de datos nuevos. Este problema se acentúa en el caso de que los datos tengan ruido o errores.

Otros tipos de aproximaciones de funciones como por ejemplo la interpolación con polinomios, también pueden aproximar correctamente los datos con los que se realiza el ajuste, mientras que aproximan de modo erróneo datos nuevos no empleados en el ajuste.

Existen diferentes modos de evitar el sobreaprendizaje. El primero de ellos sería obtener más datos para el entrenamiento, aunque esto no es siempre posible. Otra posibilidad es reducir el tamaño de la red (menor número de parámetros), de modo que la red neuronal es menos flexible y más robusta frente al ruido, aunque si se reduce demasiado, puede que no sea capaz de aprender o aproximar la función objetivo. Dotar a la red neuronal de suficientes parámetros para que sea capaz de aprender y evitar el sobreaprendizaje, es el principal aspecto a tener en cuenta en el dimensionamiento de una red neuronal.

Datos de entrenamiento, validación y prueba

Para controlar si una red neuronal ha sobreaprendido, se dividen los datos en diferentes grupos:

- Datos de entrenamiento. Son los datos empleados en el ajuste de los parámetros de la red neuronal. Han de ser representativos del total de datos, por lo que normalmente se seleccionan aleatoriamente.
- Datos de validación. Se emplean después de cada iteración en el proceso de entrenamiento, para comprobar si se produce el sobreaprendizaje.
- Datos de test. Sólo se emplean una vez finalizado el entrenamiento.

La división de los datos es normalmente un 80% de datos de entrenamiento, un 10% de validación y un 10% de prueba, aunque la elección de dichos porcentajes depende del

número de datos disponible y de su distribución. Dicha división se puede realizar con algún criterio de modo que los datos de cada grupo sean representativos, o de modo aleatorio.

Técnicas para evitar la pérdida de generalización (Early-stopping)

A partir del empleo de la división de datos en los grupos mencionados, es posible aplicar una técnica para evitar el sobreaprendizaje: early-stopping. Durante el proceso iterativo de optimización de los parámetros de la red, se comparan los errores obtenidos con los datos de entrenamiento y con los datos de validación. En el caso de que durante sucesivas iteraciones, el error con los datos de entrenamientos disminuya, mientras que el error con los datos de validación aumente, se detiene el proceso de ajuste, como un criterio de parada adicional (Figura 3).

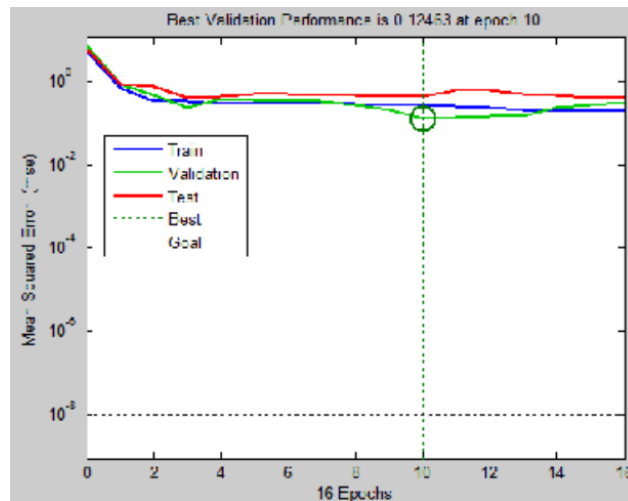


Figura 3. Representación de la evolución de errores de entrenamiento, validación y prueba. Parada por early-stopping.

Simulación

Una vez entrenada una red neuronal, se puede comprobar el funcionamiento de la misma, aportando datos de entrada y obteniendo datos de salida. Este proceso se llama simulación, ya que los datos de entrada pueden ser datos empleados en el entrenamiento, o datos nuevos de los cuales se desea tener una predicción.

Un primer paso consiste en la comparación de los datos de salida (o targets) empleados en el entrenamiento, con los datos simulados por la red neuronal. Esta comparación se llama validación cruzada, y permite representar gráficamente la bondad del ajuste de la red neuronal. Si el ajuste fuera perfecto, la representación de dichos valores se situaría sobre la recta $y = x$.

La validación cruzada se puede realizar para los datos de entrenamiento, validación, prueba o todos los datos, de modo que se puede comprobar si se ha producido sobreaprendizaje o no (Figura 4).

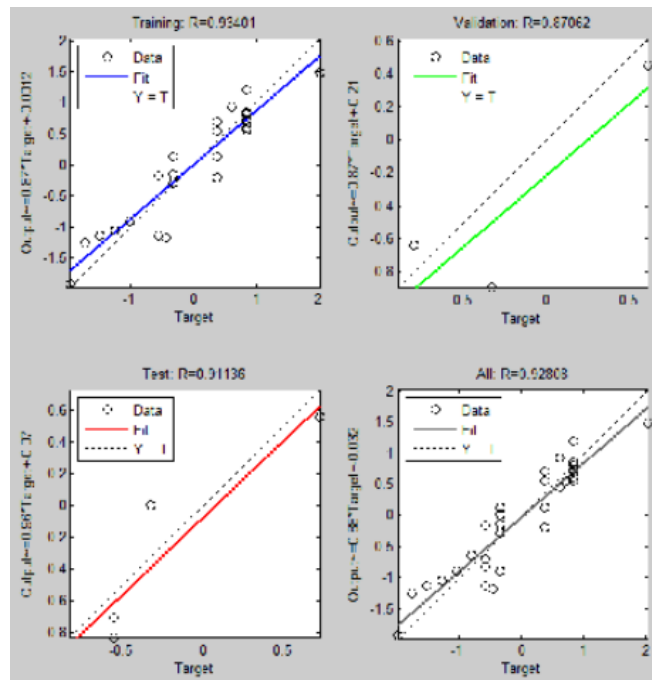


Figura 4. Validación cruzada de datos de entrenamiento, validación, test y total de datos.

Si en lugar de emplear datos existentes, se aplica la red neuronal a datos nuevos, se dispone de un laboratorio virtual con el que predecir el comportamiento del proceso aprendido por la red neuronal.

Ventajas de las redes neuronales

1. Robustez frente al ruido en datos de entrada y salida.
2. Independencia entre complejidad del problema y dimensionamiento de la red.
3. Rapidez de ajuste y simulación.
4. Laboratorio virtual, curvas de diseño o dimensionamiento.

Desventajas de las redes neuronales

1. Mínimos locales.
2. Sobreaprendizaje o pérdida de generalización.
3. Criterios de dimensionamiento de la red neuronal arbitrarios.
4. Comportamiento de caja negra.
5. La inicialización aleatoria de los pesos y bias, y la división aleatoria de datos en entrenamiento, validación y prueba, dan lugar a soluciones diferentes (relacionado con los mínimos locales).
6. No permiten la extrapolación [1].

Ecuación PC-SAFT

Los datos obtenidos con la simulación de la red neuronal de la propiedad termodinámica de equilibrio de fases calculada empleando la ecuación PC-SAFT a partir de datos experimentales de equilibrio de fases.

Las propiedades de saturación necesarias para estimar las propiedades termodinámicas derivativas se obtuvieron empleando la ecuación de estado PC-SAFT. Se utiliza esta ecuación de estado para reducir las inconsistencias termodinámicas presentes en el uso de otras metodologías para la predicción de propiedades de saturación. La ecuación PC-SAFT, tiene dos términos, uno para la contribución de la cadena de esfera dura de referencia y otra relacionada con la perturbación o dispersión [2].

$$\check{\alpha}^{res} = \check{\alpha}^{hc} + \check{\alpha}^{pert} \quad (1)$$

Dónde $\check{\alpha} = A/NkT$ y T , A , N y k son: la temperatura, energía libre de Helmholtz, número total de moléculas y la constante de Boltzmann, respectivamente. La contribución de la esfera dura se basa en la teoría de la perturbación termodinámica de primer orden y es definida como:

$$\check{\alpha}^{hc} - \check{\alpha}^{ideal} = \bar{m}\check{\alpha}^{hs} - \sum_i x_i(m_i - 1) \ln(g_{ii}^{hs})(\sigma_{ii}) \quad (2)$$

dónde x , m y g^{hs} son la fracción molar de las cadenas, número de segmentos en la cadena y la función de distribución par radial para segmentos. Los efectos del sistema de esfera dura incluyen el número de densidad total de moléculas y el diámetro de segmento dependiente de la temperatura. La contribución debido a la perturbación está representada por:

$$\check{\alpha}^{pert} = \check{\alpha}_1 + \check{\alpha}_2 \quad (3)$$

Dónde $\tilde{\alpha}_1, \tilde{\alpha}_2$, contienen las relaciones adoptadas para las reglas de mezcla de un fluido puro y extienden los términos de la perturbación a mezclas. Reglas de combinación convencionales son también usadas para determinar los términos de cruzamiento:

$$\sigma_{ij} = (1/2)(\sigma_{ii} + \sigma_{jj}) \quad (4)$$

$$\varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}}(1 + k_{ij}) \quad (5)$$

Dónde σ , ε y k_{ij} son el diámetro del segmento, parámetro de atracción y el parámetros de interacción binario, respectivamente. Los parámetros del componente puro de la EDO PC-SAFT son m , σ y ε [2].

Por lo cual, en esta investigación se estudiará el uso de redes neuronales para la predicción de propiedades termodinámica en mezclas binarias de n-Alcano + n-alcohol.

2. Desarrollo

Una red neuronal se compone de neuronas (elemento básico) que se agrupan en varios niveles o capas, las neuronas se encuentran interconectadas lo que por analogía del sistema nervioso sería la sinapsis. Esta estructura cuenta con diversas entradas y salidas, las neuronas son entrenadas para proporcionar valores de salida de manera deseada a partir de los datos o estímulos introducidos a la red. Las interconexiones entre neuronas están relacionadas por los pesos, umbrales y señales de activación. Los pesos representan la experiencia que va adquiriendo la red, esto permite a la red identificar patrones que no tienen relevancia dentro del comportamiento global que está capturando la red. Los umbrales representan el aporte de cada neurona a las neuronas de la siguiente

capa y de esta forma a toda la red. El umbral es independiente de neuronas de capas anteriores ya que no está conectado a ninguna neurona y da estabilidad a la red en el proceso de entrenamiento.

Las neuronas se encuentran distribuidas en una red neuronal por medio de las capas o niveles, estas capas pueden ser de alimentación donde es introducida la información a la red, capas ocultas donde la red captura el comportamiento no lineal y realiza la optimización del sistema y de salida que es hacia donde converge la solución del sistema. La conectividad entre neuronas es un aspecto importante en el procesamiento de la información que realizan las redes neuronales, es establecida de manera que cada neurona en una capa distribuye su respuesta a todas las neuronas de la capa superior inmediata.

En una RNA con alimentación hacia adelante la información fluye desde la capa de alimentación a través de las capas de mando superior hasta la capa de salida. La función de activación para neuronas en las capas de entrada y de salida es lineal y para capas ocultas se ocupa una función no lineal.

Base de datos experimentales

Simulación compara con información experimental de equilibrio de fases obtenida de artículos científicos [3].

Tipología del modelo de red neuronal

Se diseñó un modelo neuronal para la obtención de fugacidades. El cual consiste en un conjunto de elementos de cálculo llamados neuronas (por su similitud con las neuronas biológicas) conectadas en serie y paralelo. La conexión de varias neuronas en paralelo conforman una capa y varias de estas últimas puede conectarse en serie para formar una RNA. Las RNA realizan sus cálculos empleando funciones no lineales y factores simples

de multiplicación, llamados pesos, los que están asociados con un enlace entre dos neuronas.

Estas estructuras tienen la capacidad de “aprender” relaciones complejas no lineales entre entradas y salidas a partir de la experiencia vivida mediante un proceso denominado entrenamiento, durante el cual son ajustados los pesos hasta que el conjunto de entrada produzca las salidas deseadas. Existen varios tipos de RNA adecuadas para diferentes aplicaciones. Los modelos desarrollados se basan en una red multicapa con propagación hacia adelante con algoritmo de aprendizaje con retropropagación. Este tipo de red fue escogida dadas las posibilidades que brinda y su amplio uso en un gran número de aplicaciones. Para la modelación se realizó por medio de programación de código dentro del software Matlab. La simulación se realizó por medio de Simulink dentro del software Matlab (Figura 5).

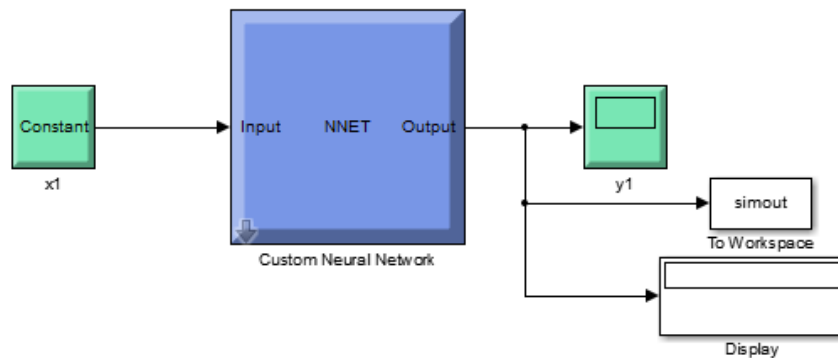


Figura 5. Red neuronal creada por Simulink.

3. Resultados

A continuación se presentan los resultados de fugacidades en las siguientes mezclas binarias n-Alcano + n-Alcohol (Figuras 6 – 12).

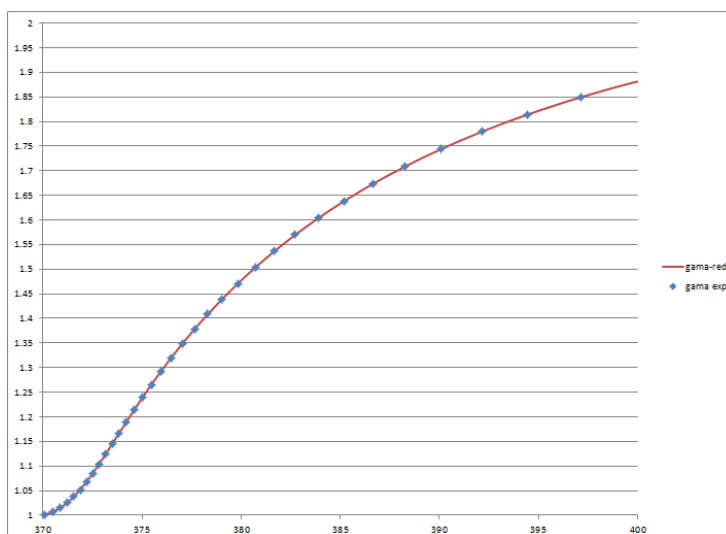


Figura 6. Comparación de fugacidades entre la red neuronal vs datos experimentales n-Decano + n-Decanol.

A continuación se analiza el error porcentual relativo entre la red neuronal y los datos experimentales.

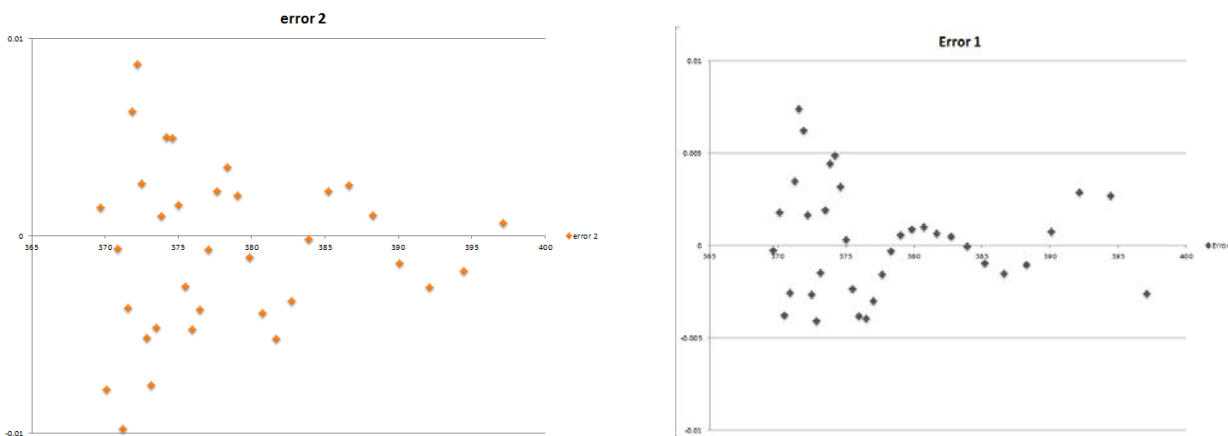


Figura 7. Error relativo neuronal vs datos experimentales n-Decano + n-Decanol.

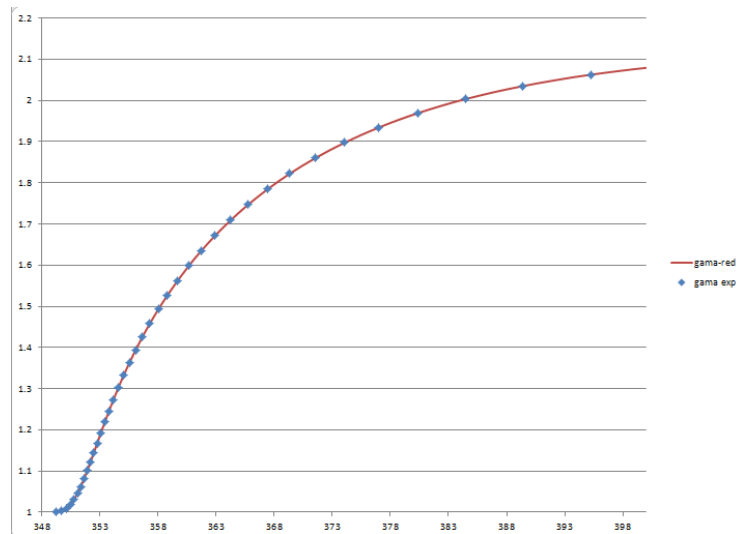


Figura 8. Comparación de fugacidades entre la red neuronal vs datos experimentales n-Nonano + n-Nonanol.

A continuación se analiza el error porcentual relativo entre la red neuronal y los datos experimentales.

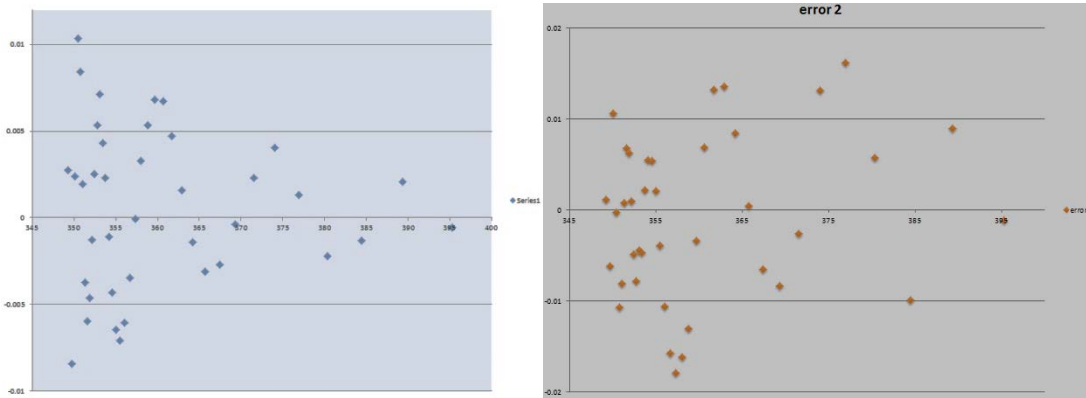


Figura 9. Error relativo neuronal vs datos experimentales n-Nonano + n-Nonanol.

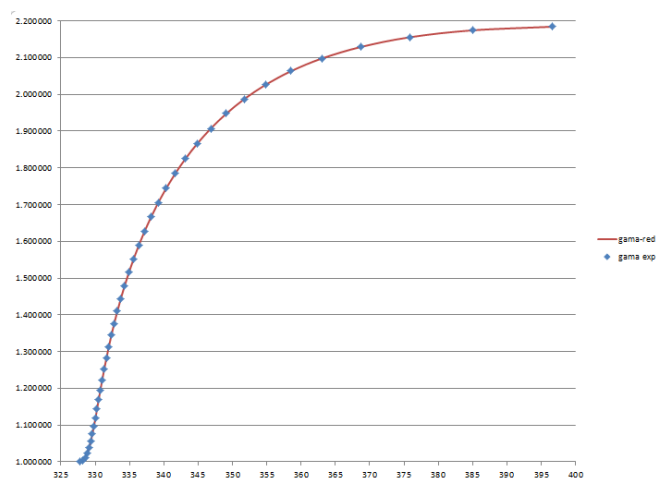


Figura 10. Comparación de fugacidades entre la red neuronal vs datos experimentales n-Octano + n-Octanol.

A continuación se analiza el error porcentual relativo entre la red neuronal y los datos experimentales.

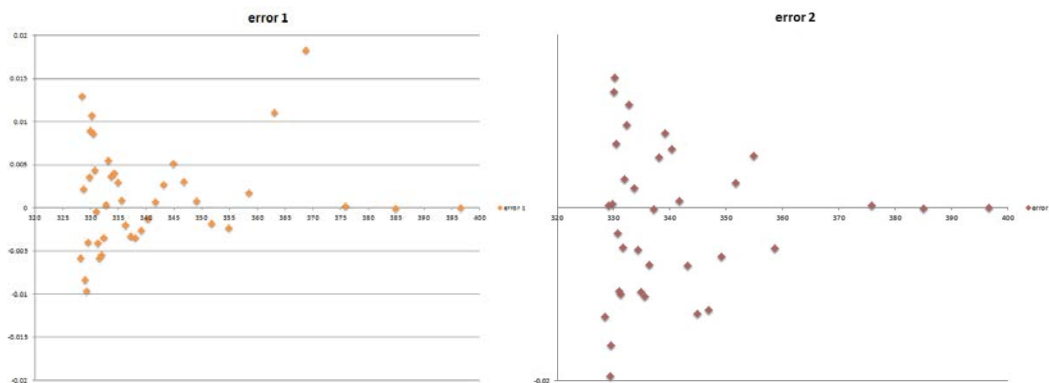


Figura 11. Error relativo neuronal vs datos experimentales n-Octano + n-Octanol.

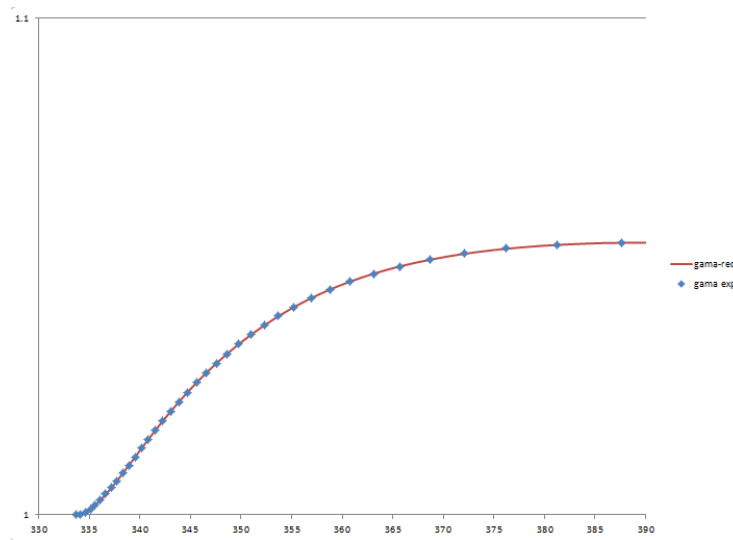


Figura 12. Comparación de fugacidades entre la red neuronal vs datos experimentales n-Butano + n-Butanol.

A continuación se analiza el error porcentual relativo entre la red neuronal y los datos experimentales.

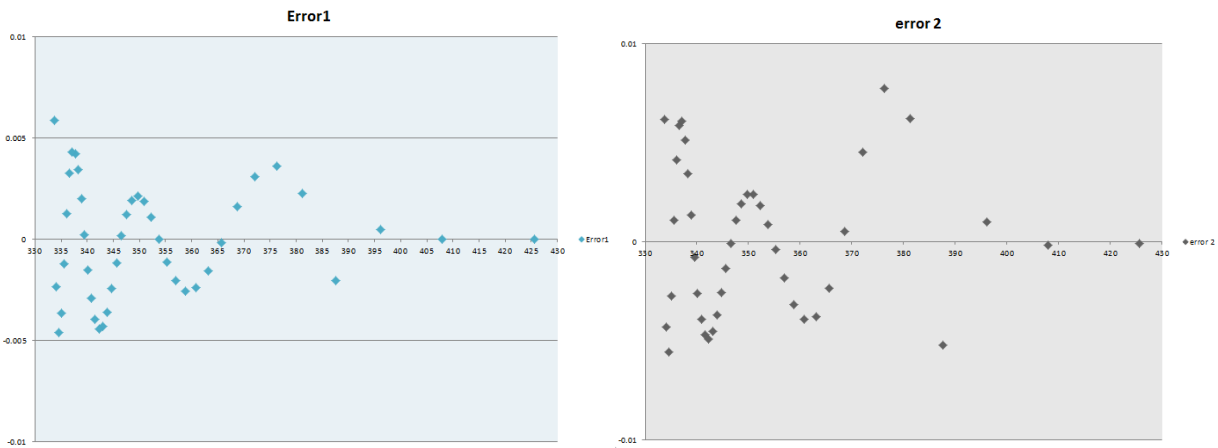


Figura 12. Error relativo neuronal vs datos experimentales n-Octano + n-Octanol.

4. Conclusiones

En este trabajo de investigación se desarrollo un código en Matlab que integra cada una de las zonas de las Redes Neuronales (RNA) para la simulación y evaluación de propiedades termodinámicas de sistemas binarios n-Alcano + n-Alcohol con una divergencia entre las predicciones de la red neuronal y la ecuación de estado PC-SAFT menor al 0.05%; y un divergencia entre las predicciones de la red neuronal y los datos experimentales menor al 0.05%. Lo cual demuestra la fiabilidad de usar redes neuronales para la predicción de propiedades termodinámicas.

5. Referencias

- [1] http://www.mathworks.com/access/helpdesk/help/pdf_doc/nnet/nnet.pdf MATLAB Neural Network Toolbox
- [2] Gross J.; Sadowski G.; "Application of the Perturbed-Chain SAFT Equation of State to Associating Systems" *Ind. Eng. Chem. Res.* 41, 5510, 2002.
- [3] Barker J.A.; Henderson D.; "Perturbation Theory and Equation of State for Fluids: The Square-Well Potential, *J. Chem. Phys.*, 47, 2856-2861, 1967.
- [4] Chapman, W.G.; Gubbins, K.E.; Jackson, G.; Radosz, M., "New Reference Equation of State for Associating Liquids.", *American Chemical Society*, 29, 1709-1721, 1990.
- [5] Currás M.; Vijande J.; Piñeiro M. M.; Salgado J.; García J.; "Behaviour of the Environmentally Compatible Absorbents bminBF₄ and eminBF₄ with TFE: Experimental Densities at High Pressures" 24th European Symposium on Applied Thermodynamics, Santiago de Compostela, 2009.
- [6] Gross J.; Sadowski G.; "Application of the Perturbed-Chain SAFT Equation of State to Associating Systems" *Ind. Eng. Chem. Res.* 41, 5510, 2002.