

INTELIGENCIA ARTIFICIAL PARA LA PREDICCIÓN DE PROPIEDADES TERMOQUÍMICAS: UNA APLICACIÓN

ARTIFICIAL INTELLIGENCE FOR THE PREDICTION OF
THERMOCHEMICAL PROPERTIES: AN APPLICATION

Fernanda Saviñon-Flores
Miguel Ángel García-Castro
Esmeralda Vidal Robles
*Jesús Andrés Arzola Flores

ISSN 2448-5829

Año 10, No. 30, 2024, pp. 31 - 41

RD-ICUAP

<https://orcid.org/0009-0005-9242-0655>
<https://orcid.org/0000-0003-4459-873X>
<https://orcid.org/0000-0001-8143-0878>
<https://orcid.org/0000-0001-9839-982X>

Año 10 No. 30
Recibido: 04/julio/2023
Aprobado: 12/marzo/2024
Publicado: 10/septiembre/2024

Facultad de Ingeniería Química, BUAP. Av. San Claudio S/N,
San Manuel, C.P. 72560. Puebla, México.
maria.savinson@alumno.buap.mx
miguel.garciacastro@correo.buap.mx
esmeralda.vidal@correo.buap.mx
*jesus.arzolaflares@correo.buap.mx
*Tel. (221) 203 8067.

Resumen

La inteligencia artificial (IA) está cambiando el modo de ver la tecnología a nivel mundial. Desde su aplicación en los diversos asistentes de voz hasta su uso en apps en teléfonos celulares, la IA es un tema que está tomando gran relevancia debido a los beneficios que ha mostrado en el avance tecnológico de diversos sectores de la industria y la ciencia. Además, del procesamiento de lenguaje natural, la visión computacional y la robótica, dos de las ramas más relevantes de la IA son el machine learning o aprendizaje automático y el deep learning o aprendizaje profundo. En general, el objetivo común es la predicción de valores u objetos a partir del uso de datos. Estas dos ramas pueden abordar problemas de regresión, clasificación y agrupamiento, es decir, son capaces de predecir variables cuantitativas y cualitativas, así como agruparlas de acuerdo a características similares. Dentro de las variables cuantitativas de interés predictivo en el área de la Termodinámica (de manera directa) y de la ingeniería química (como aplicación), están las propiedades termoquímicas, de las cuales una de las que ha destacado por su importancia es la entalpía de combustión. En la actualidad, es posible obtener valores de entalpía de combustión con diferentes técnicas tales como la calorimetría de combustión (técnica experimental), métodos de estimación (técnica semiempírica) y, en la última década, el machine learning (técnica teórica). Siendo esta última la que ha logrado captar el interés en la comunidad científica debido a su impacto en diversas áreas del conocimiento, incluyendo la Termodinámica, ya que permite predecir propiedades termoquímicas como la entalpía de combustión de diversos compuestos químicos.

Palabras Clave: Inteligencia artificial, conjunto de datos, machine learning, predicción, propiedades termoquímicas

Abstract

Artificial intelligence (AI) is changing the way we view technology worldwide. From its application in the various voice assistants until to its use in mobile phone apps, AI is a topic that is gaining great relevance due to the benefits it has shown in technological advances in various sectors of industry and science. In addition to natural language processing, computer vision or virtual digital assistants, two of the relevant branches of AI are machine learning or automatic learning and profound learning or deep learning. In both, the common objective is the prediction of values or objects from the use of data. These two branches can address two types of problems, regression and classification, that is, they are capable of predicting quantitative and qualitative variables, respectively. Among the quantitative variables of predictive interest in the areas of thermodynamics (directly) and chemical engineering (as an application) are the thermochemical properties of materials, of which one of the most prominent is the enthalpy of combustion. Although there are different ways to obtain the enthalpy of combustion values, some techniques with which it can be obtained include combustion calorimetry (experimental technique), estimation methods (semi-empirical technique), and, in recent decades, machine learning (theoretical technique). The last one is the one that has managed to capture interest in the scientific community. Also, the use of tools such as the Python programming language to implement machine learning models to obtain predictive values of thermochemical properties such as combustion enthalpy for a better analysis and study of various chemical compounds.

Keywords: Artificial intelligence, data set, machine learning, prediction, thermochemical properties

¿Inteligencia artificial?

Sin duda la inteligencia artificial (IA), se está convirtiendo en una herramienta muy popular y controversial. Hay quienes están a favor de su uso y otros tantos que ponen en tela de juicio su aplicación. ¿Ha escuchado que, lejos de ser un beneficio, la IA es un peligro? Si al escuchar este tipo de preguntas se siente preocupado, es muy posible que usted sea considerado un neoludita. El neoludismo o los neoluditas son aquellas personas que creen que la IA nos llevará a un futuro no favorable o casi destructivo (Berzal, 2015; Downing, 2015).

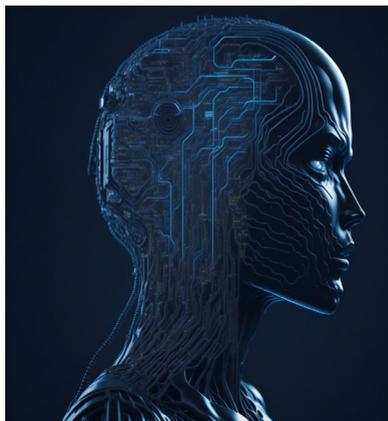


Figura 1. Inteligencia artificial. Imágenes creadas con inteligencia artificial Leonardo AI.2023. <https://leonardo.ai/ai-generations>

Lo cierto es que aún es un tema de poco dominio público. Se sabe que en el sector industrial la IA ha sido aplicada en la automatización de máquinas (industria 4.0), en el funcionamiento de smartphones, en aplicaciones móviles, en efectos artísticos cinematográficos, en la educación, y por supuesto, en la ciencia (Arzola y col., 2020; Berzal, 2015; López y col., 2021). Se torna algo un poco difícil imaginar cómo es que se aplica. La IA se puede llegar a tomar como uno de esos temas de los que creemos tener una ligera idea de lo que es, pero, si nos lo preguntamos, es muy probable que nos cueste definirlo, tal como pasa con temas relacionados con la física cuántica y sus aplicaciones.

Aunque lo anterior suena un poco burdo e incluso algo irónico, saber qué es la IA

abre un panorama para entender los avances que nuestro mundo está viviendo y, sí, eventualmente, comprender un poco más su aplicación. Un ejemplo de esto son las apps de IA, como lo es la IA Leonardo (<https://app.leonardo.ai>), que es una plataforma en línea que al insertar palabras clave o 'prompts', genera imágenes de autor, es decir, que al utilizarlas se vuelven únicas y no es necesario referenciarlas. Las imágenes mostradas (figuras 1 y 2), fueron generadas con esta IA y el 'prompt' para obtenerlas fue: 'applications of artificial intelligence to predict physicochemical properties of chemical compounds'.

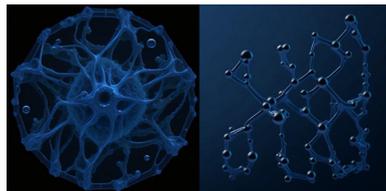


Figura 2. Aplicación de la inteligencia artificial. Imágenes creadas con la inteligencia artificial Leonardo AI. 2023. <https://app.leonardo.ai/ai-generations>

Si se pregunta sobre ¿cómo? o ¿cuándo? surgió la IA. El origen de la IA data de mediados de la década de 1950, cuando un grupo de investigadores de la Dartmouth College en New Hampshire, Estados Unidos, entre los cuales destacaron personajes de la talla de Marvin Minsky (cofundador del laboratorio de IA del MIT), Nathaniel Rochester (diseñador del IBM 701), Claude Shannon (padre de la Teoría de la Información), Arthur Samuel (autor del primer programa de ordenador capaz de aprender) o Herbert Simon (premio Nobel de Economía en 1978). Desde el nacimiento de la Inteligencia Artificial, el entusiasmo de este puñado de investigadores marcó un antes y un después en el conocimiento y aplicación del concepto de IA (Berzal, 2015; Hansen, 2019).

Pero, a todo esto, ¿Qué es la IA? Bueno, resulta que por separado sabemos qué son ambos términos. Inteligencia: capacidad de entender o comprender y, Artificial: que ha sido hecho por el ser humano y no por la naturaleza. Entonces, ¿qué es la IA?, una de las definiciones más concisas es la del autor Jerry Kaplan, de Stanford, quien además propone llamarla computa-

ción antrópica: término que engloba los esfuerzos dirigidos al diseño de sistemas informáticos de inspiración biológica, máquinas que imitan formas y capacidades humanas y sistemas que interactúan con personas de forma natural (Berzal, 2015; Jensen, 2017).

Ahora, una pauta muy importante a considerar en la IA es la información que se utiliza para poder aplicarla y lo que es conocido como conjunto de datos, bancos de datos, bases de datos, almacenamiento de datos (figura 3), es decir, datos. Además del procesamiento de lenguaje natural, la visión computacional o la robótica, dos de las ramas más relevantes de la IA son el machine learning o aprendizaje automático y el deep learning o aprendizaje profundo, los cuales de forma general tienen como objetivo principal la predicción o agrupación de valores u objetos. Tanto el machine learning como el deep learning, utilizan conjuntos de datos previamente obtenidos para realizar la predicción o agrupación de variables cuantitativas (por ejemplo, el precio de acciones en la bolsa de valores, el costo de un inmueble, etc., es decir, cantidades) o cualitativas (por ejemplo, un tipo de planta o flor, el sexo de una persona, etc., es decir, cualidades) (Berzal, 2015, Müller y col., 2016).



Figura 3. Almacenamiento de datos. 2023. <https://www.istockphoto.com/es/foto/analista>

Puente entre el machine learning y las propiedades termoquímicas

En general, en el machine learning existen tres tipos de problemas: clasificación, regresión y agrupación. Los problemas de clasificación consisten en predecir variables cualitativas, en los de regresión, variables cuantitativas y, en los de

agrupación, determinar características similares de las variables y formar grupos o “clusters”. Debido a que las propiedades termoquímicas de compuestos químicos son consideradas variables cuantitativas, es ahí donde surge el interés de implementar modelos de machine learning para predecir su valor. Pero ¿qué son las propiedades termoquímicas?, éstas son parte de la termodinámica y permiten estudiar las transformaciones energéticas que ocurren en los procesos químicos (Konkova y col., 2020). Estos procesos pueden liberar o absorber energía en forma de calor. Entre algunas propiedades de la termoquímica destacan la entalpía de combustión, la energía de Gibbs y la energía interna. La entalpía de combustión es la cantidad de energía liberada o absorbida en una reacción química a presión constante. Debido a que la entalpía de combustión es una variable cuantitativa ampliamente estudiada, ha reflejado ser una buena candidata para la aplicación de modelos de machine learning para su predicción (Faber y col., 2017; Székely y col., 2019). La obtención de esta propiedad ha tenido gran interés en áreas como la petroquímica, la ingeniería química, la industria automotriz, la electroquímica, la biomedicina, la ciencia de materiales y la ingeniería ambiental, en su mayoría, para su aplicación en compuestos orgánicos de interés.

Cálculo de propiedades termoquímicas: métodos convencionales

Por otra parte, la determinación experimental de propiedades termoquímicas de compuestos químicos orgánicos puros puede realizarse utilizando diversas técnicas y métodos, tales como calorimetría de combustión, método de efusión de Knudsen, termogravimetría, métodos por estimación por grupos funcionales, entre otras. Dentro de estas técnicas, la calorimetría es una técnica comúnmente utilizada para medir las propiedades termoquímicas de los compuestos orgánicos. Consiste en medir la cantidad de energía en forma de calor que se libera o absorbe durante una reacción química. Para ello, se utiliza un calorímetro (figura 4), que es un dispositivo que permite medir la transferencia de calor entre el sistema y

el entorno (Abdul y col., 2022; Kim y col., 2020; Miao y col., 2018).

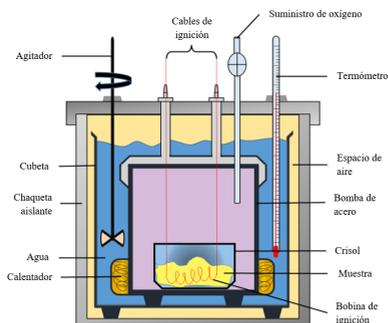


Figura 4. Sección transversal de una bomba calorimétrica típica. 2022. Tomada de Abdul y col., 2022.

La determinación de la entalpía de combustión de los compuestos orgánicos es una propiedad requerida para el cálculo de las entalpías de formación en las fases líquida y cristalina. Con todo lo anterior, es claro que surge de forma natural la pregunta ¿qué es la entalpía de combustión? Bueno, es la cantidad de energía liberada durante la combustión o quema completa de un compuesto bajo ciertas condiciones de presión y temperatura. Pero, no solo la determinación experimental de la entalpía de combustión ha sido empleada. Adicionalmente, están los métodos de estimación de esta propiedad termoquímica para compuestos químicos orgánicos puros, que ha sido un método alternativo a la determinación experimental. Estos métodos se basan en la utilización de relaciones empíricas y modelos teóricos para estimar dichas propiedades. Estas estimaciones tienen como base las propiedades estructurales de los compuestos, tales como la conectividad de los átomos y la presencia de grupos funcionales específicos. Entre los métodos más utilizados para la estimación de estas propiedades está el método de Benson o el método de Gani, los cuales se basan en la aditividad de la contribución por grupos funcionales, es decir, utilizan una serie de valores de incremento asociados entre sí para estimar la entalpía de combustión y otros parámetros termoquímicos (Constantinou y col., 1994; Holmes y col., 2012; Johnson, 2003). Otro método comúnmente utilizado es la química computacional. El programa de cómputo más usado se

conoce como Gaussian y se emplea para realizar cálculos de mecánica cuántica en compuestos químicos, incluyendo la predicción de entalpías de combustión de compuestos orgánicos puros. El programa Gaussian permite realizar cálculos de alta precisión utilizando métodos de cálculo de alta calidad, como la teoría del funcional de la densidad o DFT (por sus siglas en inglés). Estos métodos pueden proporcionar valores precisos de energía de enlace, entalpía de formación, entalpía de combustión, entre otras propiedades termoquímicas, lo que permite el cálculo de estas propiedades para una amplia variedad de compuestos orgánicos. No obstante, aunque el programa de computación Gaussian es ampliamente utilizado para realizar cálculos computacionales de propiedades termoquímicas de compuestos químicos orgánicos puros, tiene algunas desventajas que es importante tener en cuenta. Una de las principales es que los cálculos realizados con Gaussian pueden ser computacionalmente intensivos, lo que significa que pueden requerir una gran cantidad de tiempo y recursos computacionales. Además, los resultados obtenidos de los cálculos realizados con Gaussian pueden ser altamente dependientes de la elección del método de cálculo, los parámetros de configuración y la calidad de los datos de entrada utilizados, el estado de agregación de los compuestos de estudio, lo que puede afectar la precisión del cálculo de propiedades termoquímicas (Ramakrishnan y col., 2015).

Predicción de propiedades termoquímicas: nuevas metodologías

La adaptación a los avances tecnológicos es crucial para actualizar o renovar las metodologías convencionales, incluso tomarlas como punto de partida para mejorarlas. Como lo mencionamos anteriormente, una aplicación del machine learning y que ha sido poco abordada en la última década es la predicción de propiedades termoquímicas de compuestos químicos puros que no pueden ser fácilmente sintetizados, la cual puede ser complemento tanto en la estimación de entalpías de combustión mediante métodos de esti-

mación, como para la comprobación de los valores previamente obtenidos experimentalmente por técnicas calorimétricas. Adicionalmente, cabe señalar que actualmente existen múltiples lenguajes de programación que permiten implementar modelos de machine learning con cierta facilidad, tal es el caso del lenguaje de programación Python, el cual hoy en día es considerado el lenguaje de programación más popular por sus diversas aplicaciones y su gran comunidad de desarrolladores (figura 5). Además, es importante señalar, que Python posee una de las librerías o paquetes más utilizados para machine learning, scikit-learn (<https://scikit-learn.org/stable/>).

lizadas en la predicción de propiedades termoquímicas de compuestos químicos orgánicos puros, incluyendo la regresión lineal, la regresión logística, las redes neuronales artificiales y los métodos basados en árboles de decisión. Estos modelos pueden ser entrenados utilizando diversas variables moleculares, como la estructura molecular, las energías de enlace, la electronegatividad, la carga y el tamaño de la molécula. Es importante tener en cuenta que la precisión de estas predicciones depende en gran medida de la calidad y la cantidad de los datos de entrenamiento utilizados (Müller y col., 2016).



Figura 5. Python. 2023. <https://www.pexels.com/es-es/buscar/machine%20learning%20/>

¿Te has preguntado cómo es que un niño aprende a caminar? La respuesta es, entrenando. La implementación de un modelo de machine learning no es muy diferente, es decir, el modelo tiene que ser sometido a un entrenamiento para que pueda posteriormente ser utilizado para realizar predicciones.

Python es una herramienta que a través de sus librerías permite realizar el entrenamiento y prueba de modelos de machine learning utilizando conjuntos de datos, por ejemplo, datos relacionados con propiedades termoquímicas de compuestos químicos orgánicos puros conocidos, tanto en fase gaseosa como en fase condensada, es decir, es posible entrenar modelos de machine learning con datos de compuestos orgánicos conocidos, para predecir, por ejemplo, propiedades de compuestos orgánicos aún no sintetizados. Existen varias técnicas de machine learning uti-

En general, el procedimiento para la implementación de modelos de machine learning es: 1. La obtención de un conjunto de datos, ya sea de forma experimental o teórica. 2. La limpieza del conjunto de datos, para eliminar datos nulos, incorrectos, etc. 3. El análisis del conjunto de datos o ingeniería de características, para determinar las variables de mayor relevancia para realizar la predicción de la variable de interés. 4. La elección del método, el cual depende en primer lugar del tipo de problema (regresión, clasificación o agrupamiento). 5. La selección de las métricas de evaluación, para determinar que tan bien ha aprendido el modelo. 6. Entrenamiento, validación y prueba del modelo de machine learning y 7. La presentación de los resultados obtenidos. (Berzal, 2015).

Sin embargo, algunas pautas claves para la correcta aplicación del machine learning son:

a) saber si la variable a predecir es cualitativa o cuantitativa, eso definirá si el problema es de clasificación o regresión.

b) conocer las variables predictoras, es decir, conocer la información del conjunto de datos para poder hacer una correcta limpieza y un buen análisis de datos.

c) tener un buen entendimiento de los métodos y métricas de evaluación, ya que es lo que permite conocer la calidad y el rendimiento del modelo de machine learning aplicado.

En adición, la aplicación del machine learning en la predicción de propiedades termoquímicas de compuestos químicos orgánicos puros presenta varias ventajas, tales como mejorar significativamente la precisión de las predicciones de propiedades termoquímicas en comparación con los métodos de estimación tradicionales; ahorro significativo de tiempo de cómputo y recursos, ya que los modelos de machine learning pueden predecir rápidamente las propiedades termoquímicas de un gran número de compuestos sin necesidad de llevar a cabo experimentos costosos o cálculos computacionales extensos; identificación de relaciones no lineales, puesto que los modelos de machine learning permiten identificar relaciones no lineales entre las variables moleculares y las propiedades termoquímicas, lo que puede mejorar las predicciones.



Perspectivas

Algunas perspectivas respecto al uso del machine learning en la predicción de propiedades termoquímicas de compuestos químicos orgánicos puros incluyen: el desarrollo de modelos predictivos más robustos, precisos y adaptables a diferentes tipos de compuestos, la identificación de nuevas relaciones entre las variables moleculares y las propiedades termoquímicas, la optimización y diseño de nuevos compuestos químicos, su aplicación a grandes conjuntos de datos de compuestos químicos para la identificación de patrones, y, la predicción de valores de propiedades termoquímicas para compuestos químicos orgánicos puros en distintas fases.

Declaración de no Conflicto de intereses

Los autores declaran que no existe conflicto de interés alguno

Declaración de privacidad

Los datos de este artículo, así como los detalles técnicos para la realización del experimento, se pueden compartir a solicitud directa con el autor de correspondencia.

Los datos personales facilitados por los autores a RD-ICUAP se usarán exclusivamente para los fines declarados por la misma, no estando disponibles para ningún otro propósito ni proporcionados a terceros.

Agradecimientos

Los autores agradecen al CONAHCYT por el financiamiento a la estudiante de posgrado María Fernanda Saviñon Flores (2023-00002-01NACF-00835).



https://yandex.com/images/search?img_url=https%3A%2F%2Fpics14.pikabu.ru%2Fpost_img%2F2023%2F03%2F17%2F11%2Fog_og_1679081870260831982.jpg&lr=20272&pos=1&rpt=simage&text=Artificial%20intelligence

Referencias

- Abdul Jameel, A. G., Al-Muslem, A., Ahmad, N., Alquaity, A. B., Zahid, U., Ahmed, U. (2022). Predicting Enthalpy of Combustion Using Machine Learning. *Processes*, 10(11), 2384.
- Arzola-Flores, J. A., González, A. L. (2020). Machine learning for predicting the surface plasmon resonance of perfect and concave gold nanocubes. *The Journal of Physical Chemistry C*, 124(46), 25447-25454.
- Berzal, F. (2018). *Redes neuronales & deep learning: Volumen I*. Independently published.
- Constantinou, L., Gani, R. (1994). New group contribution method for estimating properties of pure compounds. *AIChE Journal*, 40(10), 1697-1710.
- Downing, K. L. (2015). *Intelligence emerging: adaptivity and search in evolving neural systems*. MIT Press.
- Faber, F. A., Hutchison, L., Huang, B., Gilmer, J., Schoenholz, S. S., Dahl, G. E., von Lilienfeld, O. A. (2017). Prediction errors of molecular machine learning models lower than hybrid DFT error. *The Journal of chemical physics*, 147(3), 1-18.
- Hansen, K., Montavon, G. (2019). *Machine learning for the prediction of molecular properties*. John Wiley & Sons.
- Holmes, J. L., Aubry, C. (2012). Group additivity values for estimating the enthalpy of formation of organic compounds: an update and reappraisal. 2. C, H, N, O, S, and halogens. *The Journal of Physical Chemistry A*, 116(26), 7196-7209.
- Jensen, F. (2017). *Introduction to computational chemistry*. John Wiley & Sons.
- Johnson, R. P. (2003). Estimating thermodynamic properties of organic compounds. *Journal of Chemical Education*, 80(11), 1277.
- Kim, M. J., Kim, J., Kim, K. (2020). The effect of molecular structure on the glass transition temperature and enthalpy of organic compounds. *Journal of Molecular Liquids*, 304, 112783.
- Konkova, T. S., Matyushin, Y. N., Miroshnichenko, E. A., Vorobev, A. B., Palysaeva, N. V., Sheremetev, A. B. (2020). Thermochemical Properties of [1, 2, 4] Triazolo [4, 3-b]-[1, 2, 4, 5] tetrazine Derivatives. *Russian Journal of Physical Chemistry B*, 14(1), 69-72.
- López-Badillo, M., García-Castro, M. A., Galicia-Aguilar, J. A., Aranda-García, R. J., Galicia-Hernández, E., Velasco-Hernández, M. A. (2021). Experimental Standard Enthalpies of Formation of 4, 4'-Methylenedi (phenylene isocyanate) and Polyamide-imides. *Russian Journal of Physical Chemistry B*, 15(2), S201-S208.

- Miao, X., Zhang, J. (2018). Thermodynamic properties of several organic compounds by calorimetry and quantum chemical calculations. *Journal of Chemical Thermodynamics*, 126, 151-156.
- Müller, A. C., Guido, S. (2016). Introduction to machine learning with Python: a guide for data scientists. "O'Reilly Media, Inc."
- Ramakrishnan, R., Dral, P. O., Rupp, M., von Lilienfeld, O. A. (2015). Quantum chemistry structures and properties of 134 kilo molecules. *Scientific data*, 2(1), 1-12.
- Székely, G., Papp, F. (2019). Determination of the enthalpy of combustion of some organic compounds by calorimetry. *Journal of Chemical Education*, 96(6), 1242-1246.