

Técnicas de Inteligencia Artificial Aplicadas a Problemas Científico-Tecnológicos

Nilda M. Pérez Otero, Javier Izetta Riera, Abigail Verazay, Virginia Battezzati, Juan Salinas, Susana G. Pérez Ibarra, Alejandra Arias

GIDIA / Facultad de Ingeniería / Universidad Nacional de Jujuy

Ítalo Palanca 10, +54 (388) 4221587

{nilperez, javierizetta, abigailrn, virginiavir, juansalinas90, gperezunju, alejandraarias8099 }@gmail.com

RESUMEN

Cada vez es más intensivo el uso de técnicas de inteligencia artificial en la mayoría de las actividades humanas y en particular, en el área científico-tecnológico. Algunas de las áreas, que muestran tanto la diversidad de campos a los que pueden aplicarse como la utilidad que pueden alcanzar, son: Bioinformática, Robótica, Tecnología de los alimentos y Recuperación de información.

Una de las tareas más requeridas es la generación automática de modelos que resuman las características que debe cumplir un elemento para pertenecer a una determinada categoría, tarea que se conoce como clasificación automática. Los métodos computacionales desarrollados para tal fin forman parte de lo que se conoce como Aprendizaje Automatizado (*Machine Learning*). Otra tarea importante es la optimización de funciones, en las últimas décadas diferentes tipos de algoritmos de

resolver una gran cantidad de problemas. El principal desafío radica en problemas que presentan una función objetivo altamente no lineal y no convexa, esto dificulta garantizar la localización del mínimo global. Por lo tanto, la necesidad de encontrar nuevas técnicas que proporcionen un mejor desempeño en este tipo de problemas sigue aún vigente y hacen de esto un campo excitante para trabajar.

Palabras clave: Metaheurísticas,

Aprendizaje Automatizado, Inteligencia Artificial.

CONTEXTO

La línea de investigación aquí presentada se encuentra inserta en el proyecto *Técnicas de Inteligencia Artificial aplicadas a problemas científico-tecnológicos*, ejecutado a partir de 2016 por el Grupo de Investigación y Desarrollo en Informática Aplicada (GIDIA) de la Facultad de Ingeniería de la Universidad Nacional de Jujuy.

El proyecto se encuentra acreditado y financiado por la Secretaría de Ciencia y Técnica y Estudios Regionales de la Universidad Nacional de Jujuy (Cód. D/0150) y se encuentra bajo el Programa de Incentivos.

1. INTRODUCCIÓN

La Inteligencia Artificial (AI) propone computacionales que brindan solución a muchos problemas complejos del mundo real. Estas herramientas se utilizan en una variedad de aplicaciones que van desde el modelado, la clasificación y el reconocimiento de patrones hasta el análisis de datos. Las técnicas de inteligencia artificial se vienen utilizando en la ciencia y la tecnología de los alimentos para la clasificación, el modelado y la optimización de procesos, el control de calidad de los alimentos, la

predicción de propiedades reológicas de masa, la clasificación de vinos en función del contenido de antocianinas, entre otros [1]. Asimismo en el área de la ingeniería industrial se comenzaron a incorporar herramientas de inteligencia artificial, a tal punto que, en el año 2015, se realizó la Conferencia Internacional de Inteligencia Artificial e Ingeniería Industrial (AEII2015) [2].

Particularmente, una rama de la inteligencia artificial que en los últimos años ha tenido gran crecimiento es la del Aprendizaje Automatizado (*Machine Learning*), la cual propone nuevas herramientas computacionales que sirven de apoyo para tareas de investigación en distintas áreas.

Aprendizaje Automatizado

En las últimas décadas el incremento de métodos automáticos de medición y almacenamiento de datos produjo una revolución en las ciencias, donde la cantidad de nuevas observaciones disponibles supera ampliamente la capacidad actual de modelado. Esta situación originó la necesidad de desarrollar programas de computación que, a partir de un conjunto de datos, generen un modelo que produzca salidas aproximadas a las del proceso que está implícito en los datos observados. El aprendizaje automatizado intenta comprender los mecanismos por los cuales se adquiere el conocimiento a partir de los datos/experiencia y con ello lograr automatizar la etapa de modelado [3]. El conjunto de datos de entrada de un programa de aprendizaje automatizado consiste en descripciones de objetos de un universo acompañado de un valor de salida asociado. Cada objeto es considerado un ejemplo del universo que se desea conceptualizar. Generalmente estos ejemplos son tuplas de la forma (atributos; salida). El proceso de aprendizaje consiste entonces en la búsqueda, dentro de un espacio H , de

todas las hipótesis, la hipótesis f que mejor aproxime la función original F que dio origen al conjunto de datos de entrada.

Clasificación

Los problemas de clasificación son una parte del aprendizaje automatizado [4] en los que el objetivo es asignar a los datos $D = \mathbf{x}_1; \dots; \mathbf{x}_n$ un número de valores discretos llamados clases o categorías. Cada objeto tiene asignada una clase o categoría l \mathbf{x}_i conocida. Para un problema con c clases, l \mathbf{x}_i puede tomar c valores discretos distintos. El objetivo entonces es encontrar una función clasificadora f tal que para cada objeto \mathbf{x}_i sea $f \mathbf{x}_i = l \mathbf{x}_i$, donde $f \mathbf{x}_i$ es la clase o categoría asignada por f a un objeto.

Se han introducido muchos métodos de clasificación en los últimos años, algunos de éstos son: Redes Neuronales Artificiales [5], Máquinas de Vectores Soporte [6] y Random Forest [7] que se aplicaron a una amplia variedad de problemas científicos tecnológicos [8, 9, 10, 11, 12] demostrando buenos resultados. Esto lleva a proponer la utilización de estos métodos para resolver diferentes problemas en áreas como Bioinformática, Robótica, Recuperación de información y Tecnología de los alimentos, entre otros.

Metaheurísticas

Las metaheurísticas, en su definición original, son métodos de resolución que orquestan una interacción entre los procesos de mejora local y estrategias de mayor nivel para crear un proceso capaz de escapar de óptimos locales realizando una búsqueda robusta en el espacio de soluciones. Con el tiempo, estos métodos han llegado a incluir cualquier procedimiento que emplee estrategias para superar la trampa de la optimalidad local en espacios de soluciones complejos, especialmente aquellos procedimientos que utilizan una o más

estructuras locales como un medio para definir movimientos admisibles para la transición de una solución a otra, o para construir o destruir soluciones en procesos constructivos y destructivos [13].

Varias herramientas y mecanismos surgidos a partir de la creación de los métodos metaheurísticos demostraron ser eficaces, tanto es así que las metaheurísticas se convirtieron en el centro de atención en los últimos años como el enfoque preferido para resolver problemas de optimización complejos, en particular problemas de naturaleza combinatoria. Si bien, las metaheurísticas no son capaces de garantizar la optimalidad de las soluciones que encuentran, los procedimientos exactos o métodos de convergencia local (que en teoría podrían brindar tal garantía si se les permite correr el tiempo suficiente) a menudo son incapaces de encontrar soluciones cuya calidad sea similar a la obtenida por las principales metaheurísticas, en particular, en la resolución de problemas del mundo real donde se evidencia su eficiencia y eficacia para resolver problemas grandes y complejos. La aplicación de metaheurísticas comprende un gran número de áreas y campos disciplinarios, siendo algunos de ellos:

- Diseño de ingeniería, optimización de topologías y la optimización estructural en electrónica y VLSI, aerodinámica, dinámica de fluidos, telecomunicaciones y robótica.
- *Machine learning* y minería de datos en bioinformática y biología computacional, y finanzas.
- Modelado de sistemas, simulación e identificación en química, física y biología; control, señal, y procesamiento de imágenes.
- Planificación de problemas de enrutamiento, problemas de planificación, programación y

producción de robots, logística y transporte, gestión de la cadena de suministro, y otros.

En el diseño de una metaheurística se deben tener en cuenta dos criterios contradictorios: la exploración del espacio de búsqueda (diversificación) y la explotación de las mejores soluciones encontradas (intensificación) [14]. Las regiones más promisorias están determinadas por las “buenas” soluciones encontradas. En la intensificación, estas regiones se exploran más a fondo con la esperanza de encontrar mejores soluciones. En la diversificación, las regiones no exploradas deben ser visitadas para asegurarse de que todas las regiones del espacio de búsqueda se exploran de manera uniforme y que la búsqueda no se limita a sólo un número reducido de regiones.

Las metaheurísticas pueden clasificarse desde varios puntos de vista, [14] presenta las siguientes alternativas de clasificación:

- Inspiradas o no en la naturaleza: varias metaheurísticas se inspiran en procesos naturales, por ejemplo de la biología o de la física.
- Con o sin uso de memoria: algunos algoritmos metaheurísticos no emplean elementos de memoria, es decir, no se utiliza información extraída de forma dinámica durante la búsqueda mientras que otros usan una memoria que contiene parte de la información extraída en línea durante la búsqueda.
- Deterministas o estocásticas: una metaheurística determinista resuelve un problema de optimización mediante la toma de decisiones deterministas. Las metaheurísticas estocásticas aplican algunas reglas aleatorias durante la búsqueda.
- Búsqueda poblacional o búsqueda de única solución: los algoritmos basados en una sola solución manipulan y transforman una única solución

durante la búsqueda, mientras que en los algoritmos poblacionales es toda una población de soluciones la que evoluciona.

- Iterativas o avaras: en los algoritmos iterativos, se parte de una solución completa (o población de soluciones) y se transforman en cada iteración utilizando algunos operadores de búsqueda. Los algoritmos avaros, parten de una solución vacía, y en cada paso se asigna una variable de decisión del problema hasta que se obtiene una solución completa.

2. LÍNEAS DE INVESTIGACIÓN, DESARROLLO E INNOVACIÓN

Esta línea de investigación consiste en analizar técnicas metaheurísticas y de aprendizaje automatizado y aplicarlas al desarrollo de algoritmos robustos para problemas de optimización global y clasificación, involucrados en tareas científicas tecnológicas de distintas áreas.

3. RESULTADOS OBTENIDOS/ESPERADOS

El proyecto, que se desarrolla en el bienio 2016-2017 tiene como objetivo general desarrollar algoritmos que implementen técnicas de inteligencia artificial, una combinación de las ya existentes u otras técnicas para resolver problemas científico-tecnológicos de interés. Actualmente se lograron los siguientes resultados:

- Clasificación Automática de Textos Periodísticos: en los últimos años el periodismo regional, al igual que en todas partes del mundo, pasó de su formato clásico de publicación al electrónico. Así las webs de noticias regionales también se ven obligadas a evolucionar y mejorar sus prestaciones a través de una mejor organización y categorización previa de toda la información disponible para el lector.

En este proyecto se propuso abordar la clasificación automática de textos periodísticos digitales a través del Aprendizaje Automatizado. Se desarrollaron cuatro clasificadores automáticos de textos periodísticos extraídos de páginas webs de noticias del NOA dos de ellos basados en *Random Forest* y los dos restantes basados en *Support Vector Machine*. Además se propusieron dos técnicas nuevas para la reducción de dimensionalidad del espacio de características. Estos clasificadores fueron evaluados con distintas colecciones de noticias, pudiéndose comprobar la robustez de los mismos al mantener un buen desempeño.

- Cálculos de equilibrio de fase: estos cálculos juegan un rol crucial en la simulación, diseño y optimización de procesos de separación. La dificultad de éstos radica en que la forma de la función objetivo, altamente no lineal y no convexa, no permite garantizar la localización del mínimo global. AEvol es un algoritmo evolutivo, con un operador genético sencillo, para la optimización de funciones continuas presentado por nuestro grupo de investigación en el año 2012. Una vez validado con funciones *benchmarks*, este algoritmo se aplicó en la minimización de la función de Gibbs, función termodinámica relacionada al equilibrio de fases. En los años siguientes se desarrollaron dos nuevas versiones de AEvol que agregaron a este algoritmo operadores de cruce con el fin de mejorar su desempeño. Estas nuevas versiones fueron validadas sobre funciones *benchmarks* logrando mejores resultados que AEvol. Esta mejora está relacionada a la ampliación de la búsqueda del espacio de soluciones, mejorando la tasa de éxitos al disminuir la probabilidad de caer en mínimos locales. Actualmente se pretende realizar un estudio comparativo del desempeño de AEvol

y sus distintas versiones utilizando tres sistemas que de acuerdo a la literatura son los más adecuados para la evaluación de nuevos algoritmos de optimización, debido a que presentan mínimos locales en el espacio de soluciones, situación que los hace desafiantes para el cómputo del equilibrio de fases.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo está integrado por docentes-investigadores, egresados y alumnos de la Facultad de Ingeniería de la Universidad Nacional de Jujuy. El equipo actual cuenta con 6 ingenieros en informática y 1 alumno. En el marco de este proyecto se finalizó una tesina de grado y los integrantes del equipo dirigen actualmente 3 tesinas más. Además se prevé la finalización de una tesis de doctorado y realización de una tesis de maestría.

5. REFERENCIAS

- [1] Lee, C. J. (Editor). International Conference on Artificial Intelligence and Industrial Engineering 2015 (AIIE 2015). Proceedings of a meeting held 26-27 July 2015, Phuket, Thailand. ISBN 9781510806450. Atlantis Press.
- [2] Sofu, B.A., Demir, N., Ekinci, F.Y., 2007. Gıda Bilimi Ve Teknolojisi Alanında Yapay Zeka Uygulamaları (Applications of Artificial Intelligence in Food Science and Technology Area). GIDA (2007) 32 (2): 93-99.
- [3] Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255-260.
- [4] Mitchell, T. Machine Learning. McGraw-Hill series in computer science. McGraw-Hill, 1997.
- [5] Hassoun, M. H. Fundamentals of artificial neural networks. MIT press, 1995.
- [6] Vapnik, V. Statistical learning theory. 1998, 1998.
- [7] Breiman, L. Random forests. *Machine learning*, 45 (1), 5-32, 2001.
- [8] Pavlidis, P., Wapinski, I., & Noble, W. S. (2004). Support vector machine classification on the web. *Bioinformatics*, 20(4), 586-587.
- [9] Byvatov, E., Fechner, U., Sadowski, J., & Schneider, G. (2003). Comparison of support vector machine and artificial neural network systems for drug/nondrug classification. *Journal of Chemical Information and Computer Sciences*, 43(6), 1882-1889.
- [10] Gardner, G. G., Keating, D., Williamson, T. H., & Elliott, A. T. (1996). Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool. *British journal of Ophthalmology*, 80(11), 940-944.
- [11] Lempitsky, V., Verhoek, M., Noble, J. A., & Blake, A. (2009). Random forest classification for automatic delineation of myocardium in real-time 3D echocardiography. In *Functional Imaging and Modeling of the Heart* (pp. 447-456). Springer Berlin Heidelberg.
- [12] Miller, K., Huettmann, F., Norcross, B., & Lorenz, M. (2014). Multivariate random forest models of estuarine-associated fish and invertebrate communities. *Mar Ecol Prog Ser*, 500, 159-174.
- [13] Glover, F. W. & Kochenberger, G. A., editores (2003). Handbook of Metaheuristics, volume 114 of International Series in Operations Research & Management Science. Springer.
- [14] Talbi, E.-G. (2009). Metaheuristics: From Design to Implementation. Wiley Publishing.