

Aplicación de autoencoders variacionales para mejorar los procesos de optimización evolutiva multiobjetivo

UTN-Regional Paraná

Maestría en Minería de Datos: Plan de Tesis

Catedra: Seminario 2

Alumno: Claudio Sebastián Castillo

Directores: Dr. Matías Gerard y Dr. Leandro Vignolo

Introducción

La optimización es un componente central del diseño de los algoritmos de aprendizaje automático, contribuyendo significativamente a su rendimiento y extendida aplicación. En contextos que involucran funciones objetivo de naturaleza simple y diferenciable, los problemas de optimización pueden abordarse de manera eficiente a través de métodos determinísticos, como por ejemplo el cálculo de gradientes. Sin embargo en problemas de optimización multimodales, altamente no lineales y complejos -i.e. un vasto campo de situaciones reales- las técnicas determinísticas resultan ineficaces, planteando la necesidad de disponer métodos diferentes.

Los algoritmos evolutivos (en adelante AE) son métodos de optimización inspirados en la evolución natural, diseñados para encontrar soluciones en espacios de búsqueda complejos (Jiao et al. 2023). A diferencia de los métodos de optimización determinísticos, los algoritmos evolutivos son particularmente efectivos en espacios de búsqueda discretos o cuando la función objetivo es desconocida o no diferenciable (Williams, 2018). Utilizando técnicas de selección y evolución estos algoritmos generan iterativamente soluciones a partir de una población de candidatos (Jones et al., 2020), de manera similar a cómo la evolución natural optimiza las características biológicas a lo largo del tiempo (Smith, 2015). Su resultado regularmente implica soluciones cercanas al óptimo, es decir soluciones que mantienen un buen compromiso en la satisfacción de los múltiples objetivos que puede presentar un problema.

Pero como sucede con gran parte de los algoritmos de aprendizaje automático, los AE también se enfrentan a problemas desafiantes cuando se aplican a datasets de alta dimensionalidad y bajo número de muestras. En el contexto de espacios de gran dimensionalidad, la cardinalidad del conjunto de soluciones candidatas se incrementa de manera exponencial, conduciendo virtualmente a problemas computacionalmente intratables. Por ejemplo, si un conjunto contiene n cantidad de características, eso implica un total de 2^n subconjuntos en el espacio de búsqueda. Así, con un $n = 20$ (que representa un valor moderado) el número total de subconjuntos a evaluar supera el millón de posibles candidatos (específicamente: 1.048.576). En este contexto, la escasa disponibilidad de datos representa un problema crítico en la optimización, ya que limita la capacidad informativa de la función objetivo y, por ende, incrementa la probabilidad de converger hacia soluciones subóptimas (Hastie, Tibshirani, and Friedman, Jerome 2009).

Por lo dicho hasta aquí, vemos que la disponibilidad de datos muestrales es un aspecto crítico de los algoritmos de aprendizaje automático en general, y su carencia ha disparado el surgimiento de diferentes estrategias de solución.

Entre esas estrategias la aumentación de datos mediante Autoencoders Variacionales (en adelante AV) ocupa un lugar prominente por sus propiedades y resultados. Estos modelos generativos son capaces de aprender una representación latente de los datos de entrada y generar nuevos datos que mantienen las mismas características fundamentales (i.e. similar distribución conjunta de probabilidad) que los datos originales. La posibilidad de expandir el conjunto de datos de esta manera, abre una alternativa para mejorar la robustez de los AE, permitiendo una exploración efectiva del espacio de soluciones y mitigando los riesgos de sobreajuste y convergencia prematura. De este modo, la combinación de AE con técnicas de aumentación de datos a partir de AV ofrece un enfoque prometedor para abordar problemas de optimización en escenarios de alta dimensionalidad y muestras escasas.

Definición del problema

La escasez de datos muestrales en procesos de optimización evolutiva multiobjetivo y su posible solución a partir de técnicas de aumentación de datos mediante autoencoders variacionales.

Objetivos

General:

1. Investigar la eficacia de los autoencoders variacionales en la aumentación de datos para mejorar los procesos de optimización evolutiva multiobjetivo.

Específicos:

1. Implementar una arquitectura que integre autoencoders variacionales en la etapa inicial de un algoritmo evolutivo para la generación aumentada de población en vistas a la optimización.
2. Implementar una arquitectura que integre autoencoders variacionales en el contexto de

optimización evolutiva, particularmente en la etapa de evolución de poblaciones, para favorecer el proceso de búsqueda de subconjuntos de características potencialmente valiosas.

Bibliografía

- Hastie, Trevor, Robert Tibshirani, and Friedman, Jerome. 2009. *The Element of Statistical Learning: : Data Mining, Inference, and Prediction*. Second Edition. Springer.
- Jiao, Ruwang, Bach Hoai Nguyen, Bing Xue, and Mengjie Zhang. 2023. “A Survey on Evolutionary Multiobjective Feature Selection in Classification: Approaches, Applications, and Challenges.” *IEEE Transactions on Evolutionary Computation*, 1–1. <https://doi.org/10.1109/TEVC.2023.3292527>.