

Preservación de la diversidad en algoritmos evolutivos mediante la asignación de género

RESUMEN: Uno de los principales desafíos en los algoritmos evolutivos es la pérdida de diversidad genética a lo largo de las generaciones, lo cual puede conducir a una convergencia prematura y a soluciones subóptimas.

En este trabajo se propone una estrategia basada en la asignación de género a los individuos en la Programación Genética, con el objetivo de preservar la diversidad poblacional mediante una reproducción estructurada y equilibrada. A diferencia de la Programación Genética convencional, en la que los individuos se cruzan sin restricciones de apareamiento, nuestra propuesta empareja exclusivamente a los de género masculino con los de género femenino.

Se evaluó la estrategia propuesta en comparación con la Programación Genética tradicional en problemas de clasificación y regresión. Los resultados mostraron un rendimiento superior de la estrategia propuesta, lo que sugiere que la asignación de géneros y el emparejamiento controlado ayudan a preservar la diversidad genética. No obstante, serán necesarios más experimentos para confirmar estos resultados.

PALABRAS CLAVE: Algoritmos Evolutivos, Asignación de género, Convergencia Prematura, Diversidad, Programación Genética.



Colaboración

Luis Antonio Ibarra Flores; Lino Alberto Rodríguez Coayahuítl, Centro de Investigación Científica y Educación Superior de Ensenada / Unidad Académica Tepic.

Fecha de recepción: 07 de noviembre 2025

Fecha de aceptación: 12 de noviembre de 2025

ABSTRACT: One of the main challenges in evolutionary algorithms is the loss of diversity across generations, which can lead to premature convergence and suboptimal solutions.

This work proposes a strategy based on assigning genders to individuals in Genetic Programming, with the goal of preserving population diversity through structured and balanced reproduction. Unlike conventional Genetic Programming, in which individuals mate without restrictions, the proposed approach pairs exclusively male and female individuals.

The proposed strategy was evaluated in comparison with traditional Genetic Programming on classification and regression problems. The results showed superior performance of the proposed strategy, suggesting that gender assignment and controlled mating help preserve genetic diversity. Nevertheless, further experiments will be necessary to confirm these results

KEYWORDS: Assignment of genders, Evolutionary Algorithms, Genetic Diversity, Genetic Programming, Premature Convergence.

INTRODUCCIÓN

Los algoritmos bio-inspirados son técnicas computacionales inspiradas en fenómenos naturales, diseñadas para resolver problemas complejos mediante la imitación de los principios de adaptación, cooperación, autoorganización y evolución presentes en la naturaleza. Se agrupan en familias de acuerdo con el fenómeno que emulan; entre las más comunes se encuentran los algoritmos evolutivos, métodos de inteligencia de enjambre y algoritmos inspirados en el cerebro humano [1].

Los algoritmos evolutivos (EA) conforman una familia de técnicas de optimización estocástica inspiradas en los principios de la evolución biológica, como la selección natural y la reproducción, son útiles en problemas complejos, como aquellos de gran dimensionalidad o con espacios de búsqueda no lineales, donde los métodos tradicionales pueden fallar o resultar ineficientes.

Entre los EA existen múltiples variantes, entre las que destacan los algoritmos genéticos (GA), las estrategias evolutivas (ES), evolución diferencial (DE) y la programación genética (GP) [2].

La GP fue introducida por Koza y se caracteriza por representar soluciones como árboles sintácticos que codifican funciones matemáticas o programas simples. Esta estructura la hace especialmente adecuada para tareas como regresión simbólica y clasificación de datos [3].

Un problema común en GP es la pérdida de diversidad genética a medida que la población converge, que a su vez puede provocar convergencia prematura hacia soluciones subóptimas, limitando la capacidad de exploración del algoritmo [4].

En este trabajo se plantea la hipótesis de que la asignación de género a los individuos en la GP contribuye a preservar la diversidad genética y a mitigar la convergencia prematura, lo que puede conducir a un mejor rendimiento del algoritmo. Finalmente, el objetivo de este estudio es implementar y evaluar esta estrategia en problemas de clasificación y regresión, comparándola con la GP tradicional para determinar su eficacia en la preservación de la diversidad y en la calidad de las soluciones obtenidas.

Revisión de la literatura

Investigaciones previas han explorado múltiples enfoques para mantener la diversidad en algoritmos evolutivos; sin embargo, existen pocas propuestas recientes que introduzcan explícitamente la asignación de género en los individuos. La idea se ha abordado como parte de la selección, de la reproducción o como rasgo diferenciador dentro de la población.

Uno de los trabajos más conocidos se propone un esquema de selección sexual basado en la elección femenina. En su enfoque, las hembras eligen a los machos en función de su aptitud, lo que introduce un tipo de presión evolutiva diferente [5]. Por su parte, [6] presenta un método eficiente para incorporar el género como propiedad adicional de los individuos y reporta mejoras de rendimiento en diversos problemas de optimización.

Asimismo, [7] estudia algoritmos genéticos específicos por género, comparando variantes neutrales y diferenciadas y mostrando mejoras de desempeño. De manera complementaria, [8] combina la selección sexual

con operadores distintos según el género, generando dinámicas de competencia y cooperación que incrementan la diversidad. Finalmente, [9] ofrece un marco de referencia sobre algoritmos evolutivos, incluyendo conceptos de diversidad y estructuras poblacionales relevantes para diseñar estrategias de preservación de variedad genética.

Las principales contribuciones de la presente investigación son:

- Planteamiento de un esquema de emparejamiento directo y balanceado por género en GP, ordenado por rangos de fitness.
- Evaluación comparativa frente a la GP tradicional en problemas de clasificación binaria y regresión, con el objetivo de analizar su eficacia en la preservación de la diversidad y en la calidad de las soluciones obtenidas.

MATERIAL Y MÉTODOS

El método propuesto se basa en la incorporación de género en GP. El proceso general del algoritmo sigue el esquema clásico de GP tradicional, pero con modificaciones clave para la asignación de género, la selección de padres y la preservación de la diversidad. La Figura 1 muestra el diagrama del proceso evolutivo propuesto.

El proceso evolutivo propuesto se organiza en siete etapas: (i) inicialización de la población con asignación balanceada de género, (ii) evaluación de fitness mediante métricas específicas según el tipo de problema, (iii) ordenamiento de los individuos dentro de cada género, (iv) emparejamiento estructurado hombre-mujer según posición de ranking, (v) aplicación de operadores genéticos de cruce y mutación, (vi) asignación de género a la descendencia para mantener el equilibrio, y (vii) generación de la nueva población mediante estrategia elitista balanceada.

Este proceso se repite hasta alcanzar el número máximo de generaciones indicadas.

A continuación, se describen detalladamente cada una de estas etapas.

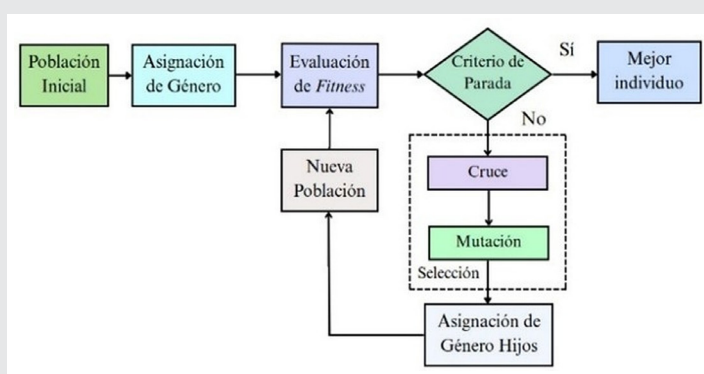


Figura 1. Diagrama general del método propuesto de GP con asignación de género.

Fuente: Elaboración propia.

Inicialización de la población con asignación de género

La población inicial se genera con un tamaño n , asegurando un balance de género desde el comienzo. Este proceso no depende de la aleatoriedad, sino de un mecanismo determinista: se establece de manera fija que el 50% de los individuos sean hombres y el 50% mujeres. De esta forma, no se deja al azar la proporción de géneros, sino que se controla desde el inicio para garantizar el equilibrio.

El criterio seguido consiste en aplicar esta asignación de manera directa y sistemática, dividiendo la población en dos grupos iguales: uno de hombres y otro de mujeres. Cada individuo, independientemente de su posición, se representa como un árbol sintáctico, siguiendo la estructura clásica de la GP tradicional.

Evaluación de fitness

Cada individuo es evaluado utilizando la función de fitness, basada en el error cuadrático medio (MSE) para problemas de regresión y en clasificación se utilizan otras métricas como exactitud.

Selección de padres por ordenamiento

A diferencia de métodos como la selección por torneo o ruleta, en este trabajo se emplea un esquema de selección de padres basado en el ordenamiento de los individuos según su fitness, con un enfoque separado por género.

La forma en que se ordenan depende del tipo de problema: El género masculino y femenino son ordenados de menor a mayor error en problemas de regresión. Por otra parte, en problemas de clasificación son ordenados de mayor a menor exactitud, lo cual es clave para el emparejamiento.

Cruza y Mutación (Emparejamiento estructurado y reproducción)

En esta etapa, se aplica la estrategia de emparejamiento balanceado, en la que los individuos con mejores fitness de cada género se reproducen de manera directa: el mejor hombre se empareja con la mejor mujer, el segundo mejor hombre con la segunda mejor mujer, y así sucesivamente hasta completar la lista.

A cada pareja se le aplica un operador genético, cruce o mutación, según una probabilidad. Estas probabilidades no son las mismas que en el esquema tradicional porque la dinámica evolutiva cambia: en el tradicional, la selección aleatoria de padres obliga a mantener un balance entre exploración (buscar soluciones nuevas mediante cruce) y explotación (aprovechar soluciones ya buenas mediante mutación). En cambio, en el enfoque propuesto se favorece el cruzamiento, para poder aprovechar de manera más amplia la dinámica poblacional propuesta con asignación de género.

Asignación de género a la descendencia

Una vez aplicados los operadores genéticos a los padres, la descendencia recibe un género asignado de forma alternada con el fin de preservar el balance de la población. Es

decir, los hijos se etiquetan siguiendo una secuencia hombre-mujer asegurando que ambos géneros se mantengan representados equitativamente en cada generación.

Esta alternancia, además de garantizar un equilibrio, mantiene una distribución ordenada dentro de la población y evita agrupamientos que podrían sesgar la dinámica evolutiva.

Generación de la nueva población

Para formar la nueva generación se implementa una estrategia elitista balanceada por género, diseñada para mantener tanto la diversidad genética como la proporción equitativa entre hombres y mujeres en cada iteración.

La técnica consiste en seleccionar el 25 % de los mejores hombres y el 25 % de las mejores mujeres de la población parental, y complementarlo con el 25 % de los mejores hijos hombres y el 25 % de las mejores hijas mujeres obtenidos en la etapa de reproducción. De esta manera, se conforma el 100 % de la nueva población, asegurando que ambos géneros estén representados equitativamente dentro de los individuos de mejor fitness.

Esta estrategia contribuye a la exploración de nuevas soluciones, ya que evita que un solo grupo genético domine la población. Al mantener diversidad en cada generación se generan combinaciones variadas que amplían la búsqueda y reducen el riesgo de convergencia prematura.

No obstante, existe la posibilidad de que los mejores individuos se crucen repetidamente, dado que la descendencia puede heredar una estructura muy similar al de sus padres. Este efecto se acentuaría en escenarios con un número excesivamente alto de generaciones.

Criterio de parada

El proceso evolutivo se repite hasta que se cumple el siguiente criterio de parada: Se alcanza un número máximo de generaciones y se obtiene al mejor individuo de todas las generaciones.

Herramienta de implementación

La implementación del método propuesto se realizó a partir de la biblioteca TurboGP [10], la cual fue modificada para incluir la asignación de género desde la población inicial y un esquema de emparejamiento estructurado. Con estas modificaciones, los individuos se organizan en grupos de hombres y mujeres, y su reproducción se define de forma determinista en función de su valor de fitness, a diferencia del mecanismo original donde los individuos podían ser seleccionados para cruzamiento con cualquier otro individuo (torneo binario).

Evaluaciones experimentales

En esta sección se describen los conjuntos de datos utilizados en las tareas de clasificación y regresión.

Para la evaluación del método propuesto, se utilizaron tres conjuntos de datos comúnmente usados para evaluar el

desempeño de algoritmos de aprendizaje máquina del repositorio de aprendizaje automático de la Universidad de California Irvine (UCI). Se comparó el método propuesto contra un enfoque de GP tradicional en los siguientes problemas:

a) Conjunto de datos en masas de mamografía:

Este conjunto de datos contiene información clínica de masas detectadas en mamografías. El conjunto consta de 961 muestras y 5 atributos. La clase objetivo es la severidad de la masa obteniendo si es benigna igual a 0 o maligna igual a 1 [11].

b) WDBC (Wisconsin Diagnostic Breast Cancer):

Este conjunto de datos contiene información extraída de imágenes digitalizadas de células mamarias. Cada instancia representa una muestra de tejido, clasificada como benigna (clase 0) o maligna (clase 1). El conjunto incluye 569 muestras y 30 atributos utilizado comúnmente para tareas de clasificación binaria [12].

c) Spambase:

Este conjunto de datos se utilizó debido a que contiene correos electrónicos donde la clase objetivo indica si el correo es spam (1) o no (0) es spam. Consta de 4601 muestras, es un conjunto de datos ampliamente utilizado como pruebas comparativas en problemas de clasificación binaria [13].

Para el conjunto de datos utilizados en problemas de regresión se utilizaron tres funciones sintéticas también típicas en GP para evaluar la estrategia propuesta y el enfoque tradicional [14].

En la Ecuación 1 se presenta una función conocida como función sombrero esta ecuación genera una superficie ondulada en tres dimensiones frecuentemente utilizada como pruebas comparativas.

Función 1:

$$z_1 = \sin \left(\sqrt{x^2 + y^2} \right) \quad \text{Ec. (1)}$$

Se evaluó en un rango de, $[-5, 5]$ para ambas variables. Se utilizó para las muestras de entrenamiento 2000 y para prueba 400 muestras.

En la Ecuación 2 es una función univariada con comportamiento oscilante y decaimiento exponencial.

Función 2:

$$z_2 = x^3 e^{-x} \cos(x) \sin(x) (\sin^2(x) \cos(x) - 1) \quad \text{Ec. (2)}$$

Se generaron 1000 muestras aleatorias en el intervalo $[0, 2\pi]$ para entrenamiento y 200 muestras para prueba.

En la Ecuación 3 la función combina una oscilación sinusoidal con un crecimiento lineal en x, generando una señal modulada.

Función 3:

$$z_3 = 0.3x \sin(2\pi x) \quad \text{Ec. (3)}$$

Se utilizaron 1000 muestras aleatorias para el entrenamiento y 200 para la prueba, en el rango $[-10, 10]$ para entrenamiento y prueba.

Dentro de la configuración experimental se emplearon los mismos parámetros tanto para los problemas de clasificación y regresión, manteniendo constante la población, número de generaciones, profundidad máxima y el conjunto de operadores primitivos. En la Tabla 1 se presentan estos valores para ambos métodos. La única diferencia entre ambos métodos radica en las tasas de cruce y mutación. En particular, el enfoque propuesto, denominado AG/GP, emplea una mayor tasa de cruce, lo cual se relaciona directamente con el diseño específico del mecanismo de apareamiento basado en género y con las necesidades evolutivas particulares del modelo.

Tabla 1. Parámetros experimentales para clasificación y regresión.

Parámetro	GP. Trad	AG/GP
Población	1000	
Generaciones	50	
Cruce	0.5	0.7
Mutación	0.5	0.3
Profundidad	7	
Operadores	add, sub, mul, div, x^2 , max, min, sin, cos, sqrt	

Fuente: Elaboración propia.

Finalmente, todos los experimentos se realizaron en un equipo de cómputo ASUS TUF GAMING A15, con procesador AMD Ryzen 5 7535HS con Radeon Graphics, 16 GB de RAM y utilizando únicamente la CPU.

RESULTADOS

Cada experimento fue ejecutado 15 veces para analizar el comportamiento promedio del modelo. En cada ejecución se registró el mejor valor de fitness alcanzado, a partir del cual se calcularon el promedio y la desviación estándar para cada conjunto de datos. En problemas de clasificación, un valor promedio más alto indica mejor rendimiento, mientras que, en problemas de regresión, un valor más bajo del MSE refleja mejor rendimiento. En ambos casos, una desviación estándar menor sugiere mayor estabilidad en el comportamiento del modelo.

En la Tabla 2 se presentan los resultados obtenidos para los tres conjuntos de datos de clasificación, donde se observa que el enfoque propuesto supera al modelo

tradicional en dos de los tres conjuntos en términos de rendimiento promedio.

Tabla 2. Resultados de los experimentos de clasificación con 15 ejecuciones por conjunto de datos.

Conjunto de datos	GP. Trad	AG/GP
Mamografía	0.824 ± 0.01	0.890 ± 0.02
WDBC	0.933 ± 0.02	0.995 ± 0.01
Spambase	0.904 ± 0	0.897 ± 0.01

Fuente: Elaboración propia.

Para una mejor visualización de la dispersión y consistencia de los resultados, se incluye el gráfico de caja en la Figura 2, el cual muestra la distribución de la exactitud alcanzada por cada método.

Se observa que, en los casos de Mamografía y WDBC, el enfoque propuesto supera consistentemente al GP tradicional tanto en términos de rendimiento promedio como de menor dispersión en los resultados. En el caso de Spambase, ambos métodos presentan un desempeño similar, aunque el GP tradicional muestra una ligera ventaja.

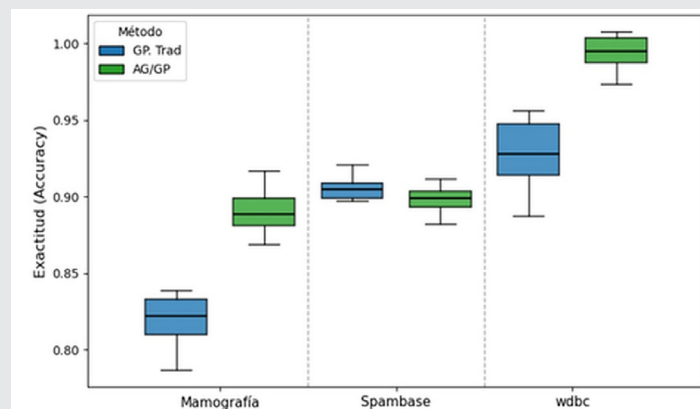


Figura 2: Gráfico de caja para la comparación de los resultados de clasificación.

Fuente: Elaboración propia.

En el caso de los problemas de regresión, los resultados se presentan en la Tabla 3, donde se reportan los valores promedio y desviación estándar del MSE obtenidos tras 15 ejecuciones por función.

Tabla 3. Resultados de los experimentos de regresión con 15 ejecuciones por función.

Conjunto de datos	GP. Trad	AG/GP
Función 1	0.075 ± 0.03	0.049 ± 0.03
Función 2	0.006 ± 0.001	0.007 ± 0.003
Función 3	1.074 ± 0.344	0.964 ± 0.179

Fuente: Elaboración propia.

En la Figura 3, se muestran los diagramas de caja correspondiente, los cuales permiten observar la dispersión de los errores y comparar gráficamente el comportamiento de ambos enfoques. Se puede observar que, en las funciones 1 y 3, el modelo con asignación de género obtiene un MSE promedio menor, lo cual indica una mejor capacidad de aproximación.

Además, en estas dos funciones, la dispersión es también más reducida, reflejando una mayor estabilidad evolutiva. En la función 2, en cambio, el GP tradicional obtiene un rendimiento ligeramente superior.

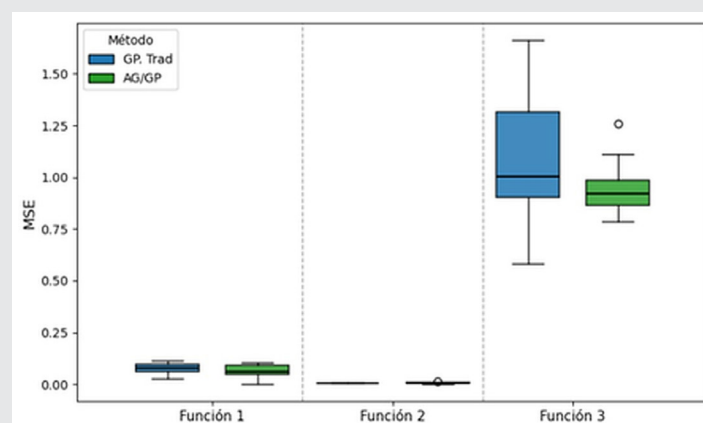


Figura 3: Gráfico de caja para la comparación de los resultados de regresión.

Fuente: Elaboración propia.

CONCLUSIONES

La incorporación de la asignación de género en Programación Genética permitió preservar mejor la diversidad poblacional y reducir la convergencia prematura. En los experimentos de clasificación, el enfoque propuesto alcanzó un buen rendimiento en los conjuntos de datos Mamografía y WDBC. En los problemas de regresión, obtuvo mejores resultados en dos de las tres funciones evaluadas. Estos resultados confirman que la asignación por géneros y el emparejamiento estructurado pueden mejorar tanto la calidad como la estabilidad de las soluciones evolutivas frente al enfoque tradicional.

AGRADECIMIENTOS

Los autores agradecen a la Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI) por su apoyo en el desarrollo de esta investigación, a través de las Becas Nacionales para estudios de Posgrado 2024-2. Asimismo, se agradece a los doctores Ansel Yoan Rodríguez González y José Agustín Almaraz Damián por sus valiosas revisiones y comentarios que contribuyeron al fortalecimiento de este trabajo.

BIBLIOGRAFÍA

[1] S. Somvanshi, M. M. Islam, S. A. Javed, G. Chhetri, K. S. Islam, T. I. Chowdhury, S. B. B. Polock, A. Dutta, and S. Das, "A Comprehensive Survey on Bio-Inspired Algorithms: Taxonomy, Applications,

and Future Directions,” 2025. [Online]. Available: <https://arxiv.org/abs/2506.04238>.

[2] C. A. Coello Coello, G. B. Lamont, and D. A. Van Veldhuizen, *Evolutionary Algorithms for Solving Multi-Objective Problems*. New York, NY, USA: Springer, 2002.

[3] J. R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. Cambridge, MA, USA: MIT Press, 1992.

[4] R. N. Abulail, “Premature Avoidance in Genetic Algorithm using Dynamic Mutation Operator,” *Journal of Optimization and Control*, vol. 10, no. 1, pp. 45–58, 2025.

[5] K. Goh, A. Lim, and L. Rodrigues, “Sexual selection for genetic algorithms,” *Artificial Intelligence Review*, vol. 19, pp. 49–66, 2003.

[6] C. Kasten, J. Fahr, and M. Klein, “An efficient way of introducing gender into evolutionary algorithms,” *IEEE Transactions on Evolutionary Computation*, vol. 27, no. 4, pp. 1005–1014, 2023, doi: 10.1109/TEVC.2022.3192481.

[7] Z. Drezner and T. Drezner, “Gender-specific genetic algorithms,” in *Proc. 34th Western Decision Sciences Institute (WDSI)*, Vancouver, Canada, Mar. 2005, pp. 411–413.

[8] J. Sánchez-Velazco and J. A. Bullinaria, “Sexual selection with competitive/co-operative operators for genetic algorithms,” in *Neural Networks and Computational Intelligence*, 2003, pp. 191–196.

[9] A. E. Eiben and J. E. Smith, *Introduction to Evolutionary Computing*. Berlin, Germany: Springer, 2003.

[10] L. Rodriguez-Coayahuitl, A. Morales-Reyes, and H. J. Escalante, “TurboGP: A flexible and advanced Python-based GP library,” 2023. [Online]. Available: <https://arxiv.org/abs/2309.00149>.

[11] “Mammographic Mass Data Set,” UCI Machine Learning Repository. [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/mammographic+mass>.

[12] “Breast Cancer Wisconsin (Diagnostic) Data Set,” UCI Machine Learning Repository. [Online]. Available: [https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+\(Diagnostic\)](https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)).

[13] “Spambase Data Set,” UCI Machine Learning Repository. [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/spambase>.

[14] M. Keijzer, “Improving symbolic regression with interval arithmetic and linear scaling,” in *Genetic Programming. EuroGP 2003*, C. Ryan, T. Soule, M. Keijzer, E. Tsang, R. Poli, and E. Costa, Eds., *Lecture Notes in Computer Science*, vol. 2610. Berlin, Germany: Springer, 2003, pp. 70–82. doi: 10.1007/3-540-36599-0_6.

