

Aprendizaje mediante enlace Genético.

Felipe PADILLA

Departamento de Sistemas Electrónicos
Universidad Autónoma de Aguascalientes
Aguascalientes, México

Alejandro PADILLA

Departamento de Sistemas Electrónicos
Universidad Autónoma de Aguascalientes
Aguascalientes, México

Aurora TORRES

Departamento de Sistemas Electrónicos
Universidad Autónoma de Aguascalientes
Aguascalientes, México

y

Eunice PONCE DE LEON

Departamento de Sistemas Electrónicos
Universidad Autónoma de Aguascalientes
Aguascalientes, México

RESUMEN

En este artículo mostraremos la conveniencia de emplear aprendizaje mediante enlace genético como alternativa en mejorar los recorridos y trayectorias de búsquedas de los cromosomas. Mostraremos las distintas formas de representación que se han estudiado, haciendo un énfasis en los mecanismos de evolución de los mismos.

Palabras Claves: Algoritmos Genéticos, Algoritmos desordenados.

1. TÉCNICAS DE APRENDIZAJE DE ENLACE GENÉTICO

Existen diferentes técnicas dependiendo de la forma de detectar el proceso del enlace genético. Se clasifican de la siguiente forma, en tres categorías propuestas en [7] [8]:

- Métodos basados en la Perturbación.
- Técnicas de adaptación de enlace.
- Constructores de modelos probabilísticos.

Métodos basados en la Perturbación.

En ésta técnica, se detecta el enlace genético existente entre los genes mediante la observación de la diferencia del fitness causada por la perturbación.

El algoritmo genético desordenado (The messy genetic algorithm) codifica los genes en pares y corta a los cromosomas y los empalma; a ésta operación se le llama codificación desordenada. Con la iteración de la operación anterior y por medio de una fase de heterogenosis fue posible el aprendizaje del enlace genético por medio de la perturbación. Posteriormente, se implementó una mejora en el algoritmo el cual tenía una convergencia mas rápida, podía resolver problemas en tiempos menores al anterior, utilizaba bloques constructores de filtrado y utilizando enumeración de probabilidades.

Técnicas de adaptación de enlace.

Esta técnica emplea específicamente la representación de diseño de operadores y mecanismos de adaptabilidad de enlace genético a través de la evolución de todo el proceso. Schaffer et al. [1] [2] agregaron las marcas de puntuación, las

cuales eran bits extras que eran añadidos a la representación del cromosoma.

Las marcas de puntuación indican cuando una posición del cromosoma es un punto de cruce. En 1995 Levenick [9] propuso un nuevo control de cruzamiento, el cual utilizaba los metabits para implementar la probabilidad de cruzamiento diferencial.

La diferencia entre los metabits y las marcas de puntuación es que los metabits indican la probabilidad de una posición a ser escogida como punto de cruzamiento, mientras que las marcas de puntuación son usadas para determinar cuando una posición es o no un punto de cruce.

Posteriormente, Thierens y Goldberg [12] propusieron el operador genético de deducción de enlace (the linkage evolving genetic operator). En éste operador hay dos banderas booleanas las cuales afectan a cada gen para indicar cuando un gen será enlazado al cromosoma por la izquierda o por la derecha. En caso de que las dos banderas de dos genes tengan el valor de verdadero entonces se considera que estos genes se enlazan. Cuando existen varios genes enlazados en el cromosoma se consideran como bloques constructores y a la población entera se le denomina como agrupación de bloques constructores.

Constructores de modelos probabilísticos.

Para ésta técnica, Harik et al. [2][4] y [5] propusieron un algoritmo llamado: Algoritmo genético de extensión compacta (ECGA) el cual capturaba las relaciones que engloban a uno o más genes en la construcción del modelo de producto marginal, basado en la descripción de longitud mínima (MDL). En ésta técnica, una o más variables son incluidas en una unión de distribución de probabilidad para minimizar al MDL de todo el modelo.

2. REPRESENTACIÓN DE CROMOSOMAS

La representación de los cromosomas en el aprendizaje de enlace genético esta compuesta de lo siguiente:

- Genes movibles.
- Segmentos sin código.

- Expresiones probabilísticas.
- Promotores.

Los cromosomas en el aprendizaje de enlace genético consisten en genes movibles compuestos en pares y se consideran o se conceptualizan alrededor de una circunferencia, éstos pueden estar en cualquier lugar y en cualquier orden en el cromosoma. Este análisis como se puede apreciar en la figura 1 difiere de las técnicas pasadas pues antes se concebían a los genes como fijos en el cromosoma.

Los segmentos sin código han sido ampliamente usados en el estudio de algoritmos genéticos, actúan como elementos no funcionales en el cromosoma, los cuales no tienen efecto en el fitness. Estos al igual que los elementos funcionales también pueden ser movibles en el cromosoma

Las expresiones probabilísticas se muestran en la figura anterior, teniendo un 50% para cada gen en el primer cromosoma, en el segundo se muestra la siguiente expresión de probabilidad $\ell/1$ para un gen y $1-\ell/1$ como complemento para el otro gen del cromosoma.

3. DIFICULTAD DEL APRENDIZAJE DE ENLACE GENÉTICO

Cuando un problema está compuesto de un corto y sencillo bloque constructor, el trabajo resulta satisfactorio. Sin embargo, cuando el problema consiste de múltiples bloques constructores, el éxito o la falla de la solución dependen de cómo estén estos bloques escalados.

Cuando los bloques constructores de un problema son exponencialmente escalados se resuelven en función del tiempo lineal del número de bloques. Pero cuando los bloques son uniformemente escalados, se necesita conocer el tamaño de la población que crece exponencialmente, lo que hace que la dificultad se incremente también exponencialmente como se aprecia en la figura 2.

4. TIEMPO DE CONVERGENCIA PARA EL APRENDIZAJE DE ENLACE GENÉTICO

Para desarrollar un modelo para determinar el tiempo de convergencia se tienen los siguientes pasos:

1. El primer paso es identificar empíricamente la secuencia del comportamiento del aprendizaje.
2. Lo siguiente es introducirse al modelo de tiempo limitado de un solo bloque constructor desarrollado con el trabajo previo basado en los dos mecanismos del aprendizaje de enlace.
3. Finalmente, después de haber extendido el modelo de tiempo limitado, establecer la conexión entre los modelos de las macro-vistas y las micro-vistas.

5. FUNCIÓN DE PRUEBA

Esta función está definida como unidad; esto significa: *el número de unos que tiene una cadena binaria de entrada.*

Esta función divide el dominio en dos secciones:

- Una de ellas es la que tendrá al óptimo global y
- La otra que tendrá el óptimo local.

En las Ecuaciones 1 y 2 podemos apreciar un ejemplo de la función, teniendo como radio propuesto el valor de la ecuación 1 y en la Ecuación 2 mostramos el radio calculado por la fórmula propuesta.

$$\text{radio } r = 3/4 = 0.75. \text{ Ec. (1)}$$

$$r = 0.75 \geq r_{\min} = (k-1)/(2k-3) = (4-1) / (8-3) = 0.6 \text{ Ec. (2)}$$

Donde k que es el orden de la función es igual a 4.

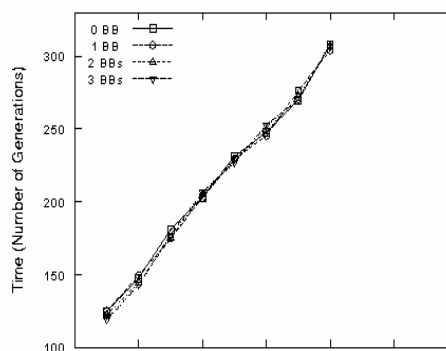
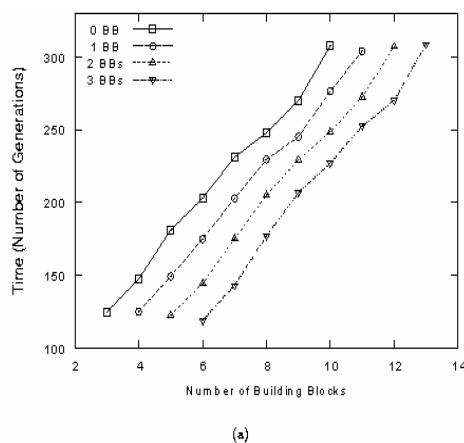
Otro parámetro esencial para los algoritmos genéticos es el tamaño de la población. Usando un tamaño de población fija para manejar problemas de distintas magnitudes es inapropiado por que la dificultad del problema usualmente se incrementa con el tamaño de la población.

En la Ecuación 3 podemos apreciar la expresión para el tamaño de la población (T.P.):

$$\text{T.P.: } n = -2\exp(k-1)\ln(\hat{\alpha}) * ((\gamma * \text{sqr}(\pi(\mu-1))/d)) \text{ Ec. (3)}$$

Donde:

- k = Longitud de un solo bloque constructor.
- $\hat{\alpha}$ = Probabilidad de fallo (Margen de error).
- γ = Desviación estándar del fitness de un bloque constructor.
- μ = Total de constructores.



6. CONCLUSIONES

Se tienen diferentes técnicas para resolver problemas de aprendizaje de enlace genético, estos métodos están clasificados en tres grupos:

- Métodos basados en la Perturbación.
- Técnicas de adaptación de enlace.
- Constructores de modelos probabilísticos.

Por otro lado se tienen algunos elementos para el aprendizaje de enlace como la representación de cromosomas, puntos de

cruzamiento, el comportamiento observado cuando se resuelven problemas con escalamiento uniforme y exponencial, una función de prueba para el tiempo de convergencia y tamaños de población y por último estas pruebas las hemos hecho en aplicaciones de optimización

combinatoria arrojando resultados muy favorables en comparación de emplear los conceptos anteriores en los tamaños de la población principalmente.

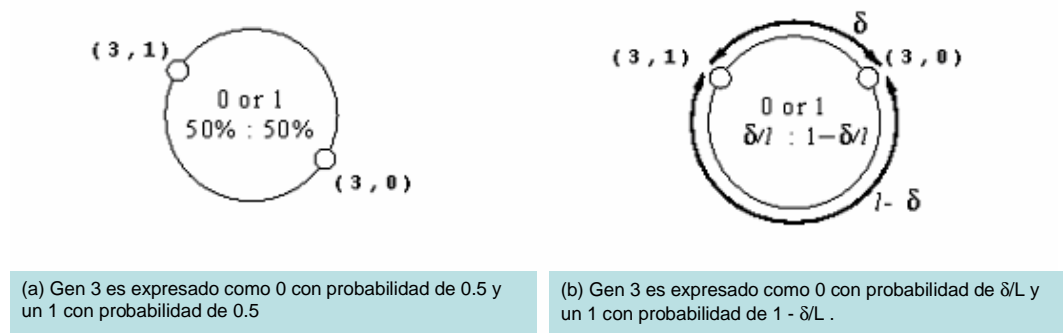


Figura 1: Representación del cromosoma

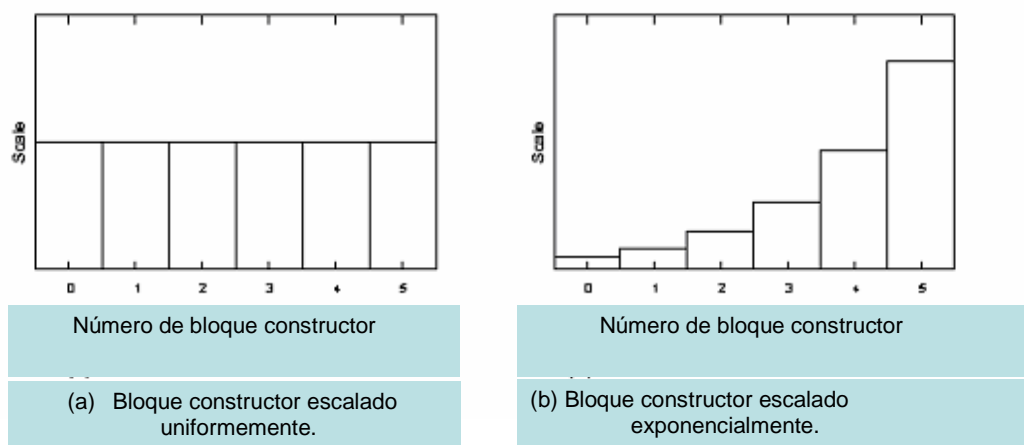


Figura 2: Esquema de la representación de los bloques constructores escalados.

7. REFERENCIAS.

[1] K.A. De Jong. **An analysis of the behavior of a class of genetic adaptive systems**. Doctoral dissertation, University of Michigan, Ann Arbor. (University Microfilms No. 76-9381), 1975.

[2] K. Deb and D. E. Goldberg. **Analyzing deception in trap functions**. In Foundations of Genetic Algorithms 2, 1993, (pp. 93--108).

[3] G. R. Harik. **Learning gene linkage to efficiently solve problems of bounded difficulty using genetic algorithms**. Unpublished doctoral dissertation, University of Michigan, Ann Arbor, 1997, Also IlliGAL Report No. 97005.

[4] G.R. Harik and D. E. Goldberg. **Learning linkage**. Foundations of Genetic Algorithms 4 , 1996, 247--262.

[5] G.R. Harik and D.E. Goldberg. **Learning linkage through probabilistic expression**. Computer Methods in Applied Mechanics and Engineering , 186 , 2000, 295--310.

[6] T. Haynes. **Duplication of coding segments in genetic programming**. In Proceedings of the Thirteenth National Conference on Artificial Intelligence, 1996, (pp. 344--349).

[7] J. H. Holland. **Adaptation in natural and artificial systems**. Ann Arbor, MI: University of Michigan Press, 1975.

[8] H. Iba and M. Terao. **Controlling effective introns for multi-agent learning by genetic programming**. In Proceedings of the Genetic and Evolutionary Computation Conference 2000 (pp. 419--426).

[9] J. R. Levenick. Swappers: **Introns promote flexibility, diversity and invention**. In Proceedings of the Genetic and Evolutionary Computation Conference 1999: Volume 1 (pp. 361--368).

- [10] R. Lindsay and A. S. Wu. **Testing the robustness of the genetic algorithm on the floating building block representation.** In AAAI/IAAI, 1996, Vol. 1 (pp. 793--798).
- [11] P. Nordin, F. Francone and W. Banzhaf. **Explicitly defined introns and destructive crossover in genetic programming.** In Proceedings of the Workshop on Genetic Programming: From Theory to Real-World Applications, 1999, (pp. 6--22).
- [12] D. Thierens and D. E. Goldberg. **Mixing in genetic algorithms.** Proceedings of the Fifth International Conference on Genetic Algorithms, 1993, 38--45.
- [13] M. Wineberg and F. Oppacher. **The benefits of computing with introns.** In Proceedings of the First Annual Conference on Genetic Programming, 1996, (pp. 410--415).
- [14] A.S. Wu and R. K. Lindsay. **A survey of intron research in genetics.** Parallel Problem Solving from Nature, 1996, PPSN IV , 101--110.
- [15] D.H. Ackley. **A connectionist machine for genetic hill climbing.** Boston: Kluwer Academic, 1987.