

Algoritmos Genéticos: ¿decepcionantes o no decepcionantes?

María Teresa Iglesias Otero

Departamento de Matemáticas. Universidade da Coruña

Los Algoritmos Genéticos están inspirados en la evolución de las especies y la mayoría de los organismos evolucionan a través de dos procesos primarios: la selección natural y la reproducción. El origen de las variaciones en la herencia se explican por la existencia de ciertos cambios (*mutaciones*) en el texto genético, pudiendo sufrir mutación todos los organismos de forma aleatoria. Pero además, en aquellos individuos que se reproducen sexualmente se puede considerar el proceso por el que las características de los organismos se mezclan al combinar su ADN (*cruce*). Este proceso de evolución natural es la fuente de inspiración de los Algoritmos Genéticos.

Los primeros hechos relacionados con este tipo de algoritmos surgieron en 1932 cuando Cannon interpreta la evolución natural como un proceso de aprendizaje muy similar al proceso mediante el cual una persona aprende por ensayo y error. También en 1950 Turing reconoce una conexión entre la evolución y el aprendizaje de una máquina. Pero los primeros intentos serios de relacionar la informática y la evolución surgieron a principios de los años sesenta del siglo pasado, cuando varios biólogos comenzaron a experimentar con simulaciones de sistemas genéticos. Éstos tuvieron poco éxito porque usaban la mutación como operador fundamental para producir nuevas combinaciones de genes. Sin embargo, John Holland, matemático de la Universidad de Michigan, estaba convencido de que era la *recombinación* de grupos de genes, que se hace mediante el cruce, la parte más importante de la evolución. A mediados de la década 1960-70 desarrolla una técnica de programación, *el Algoritmo Genético*, que se adapta a la evolución tanto por el cruce como por la mutación. Durante la década siguiente trabajó para ampliar el alcance de estos algoritmos y fruto de ese trabajo publica en 1975 la primera monografía sobre el tema, *Adaptation in Natural and Artificial Systems* ([4]), en la que se sientan las bases teóricas que fundamentan el desarrollo de los Algoritmos Genéticos desde el punto de vista computacional, abstrae los conceptos de la Genética natural, y los aplica a la economía y al reconocimiento de patrones. Desde entonces el campo de aplicación de este tipo de algoritmos no ha dejado de crecer (Medicina, Biología, Economía, Informática, Matemáticas, ...)

El funcionamiento de los Algoritmos Genéticos comienza generando una *población inicial* aleatoria de candidatos a ser el óptimo, y usan la información contenida en ésta para producir iterativamente nuevas poblaciones. Una vez calculada la idoneidad –a través de la función objetivo– de los individuos de la población inicial, se aplican de forma secuencial los diferentes *operadores genéticos* (*selección*, *cruce* y *mutación*) con el fin de crear una nueva y “mejor” población. El siguiente paso consiste en evaluar la nueva población y, a la vista de los resultados, decidir si se detiene el algoritmo o se ejecuta una nueva iteración. El algoritmo finaliza cuando se ha ejecutado un número determinado

de generaciones fijado de antemano, cuando se ha encontrado el óptimo, o bien cuando se ha obtenido en la población un nivel de idoneidad medio superior a un cierto nivel de control prefijado por el analista.

Estos algoritmos son métodos globales y robustos de búsqueda de soluciones. Su principal ventaja es el equilibrio alcanzado entre la eficiencia y eficacia para resolver diferentes y complejos problemas –como el clásico *problema del viajante* (*Travelling Salesman Problem*)– que no tienen soluciones tradicionales y únicamente habían sido abordados por métodos heurísticos.

El llamado *Teorema de los Esquemas* o *Teorema Fundamental de los Algoritmos Genéticos* recoge la importancia de estas estructuras en el proceso de convergencia y es la justificación teórica más importante del funcionamiento de este tipo de algoritmos. Asegura este teorema que el número de *esquemas* (presenten en una generación) *cortos*, de *orden bajo* y con idoneidad superior a la media poblacional –conocidos como *bloques constructivos* (*building blocks*) – aumenta en la siguiente generación.

Esta conclusión del teorema no se puede considerar más que una explicación cualitativa del proceso, pues el teorema no precisa cómo se comporta el algoritmo frente a esquemas sin representantes (“todavía”) en la población y, en consecuencia, no informa de la dirección de búsqueda en las iteraciones sucesivas. De hecho, en ocasiones, el algoritmo sucumbe a un *engaño* (*deception*) que le conduce, por ejemplo, a converger a un subóptimo.

Durante algún tiempo se consideró *engaño* (o “decepción”) = *dificultad de optimización*. Goldberg en 1989, en [3], presenta un sencillo problema –(*Minimal Deceptive Problem* o MDP)– que, aunque se plantea introduciendo un engaño, converge al óptimo deseado, en un tiempo razonable (i.e., no es difícil para el algoritmo). En 1993, Forrest y Mitchell ([2]) proporcionan ejemplos de funciones (*Royal Road functions*) que deberían ser “fáciles” –según la interpretación de los bloques constructivos– y, además, no sufren engaño. Sin embargo son difíciles para los Algoritmos Genéticos. Queda así claro que no se puede identificar dificultad con “decepción”. En 1991, Davidor, –inspirándose en una idea intuitiva de Rawlins ([6])– propone (en [1]) el concepto de *epistasis*¹ como estimador de la mayor o menor dificultad de una función ante un Algoritmo Genético. Se trata como en otros muchos casos de un término extraído de la Genética y hace referencia a la relación de dependencia entre los genes de un cromosoma.

Aunque la epistasis por sí sola no es, en general, suficiente para predecir la dificultad (funciones con igual epistasis pueden presentar comportamientos completamente diferentes frente a la convergencia al óptimo), se ha comprobado que en muchos casos particulares su valor está fuertemente relacionado con la dificultad que el algoritmo tiene para optimizar una función. De hecho, aunque estas dos propiedades son esencialmente independientes pueden reforzarse mutuamente, como sucede en el MDP (ver [5] para detalles) en el que es la epistasis la que explica el comportamiento en relación a la convergencia.

¹Un estudio detallado del concepto de epistasis se puede encontrar en [5] y [7], por ejemplo.

Referencias

- [1] Y. Davidor, *Epistasis Variance: A viewpoint on GA-hardness*, in: Foundations of Genetic Algorithms 1, pp. 23-35. Ed. G. J. E. Rawlins, Morgan Kaufmann Publishers, San Mateo, 1991.
- [2] S. Forrest and M. Mitchell, *Relative Building-Block Fitness and the Building-Block Hypothesis*, in: Foundations of Genetic Algorithms 2, pp. 109-126. Ed. L. D. Witley, Morgan Kaufmann Publishers, San Mateo, 1993.
- [3] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, 1989.
- [4] J. H. Holland, *Adaptation in Natural and Artificial Systems*, 2^a The University of Michigan Press, 1992.
- [5] B. Naudts and A. Verschoren, *Epistasis and deceptivity*, *Simon Stevin* **6**, pp.147-154, 1999.
- [6] G. J. E. Rawlins, *Foundations of Genetic Algorithms*, Morgan Kaufmann Publishers, San Mateo, 1991.
- [7] D. Suys and A. Verschoren, *Extreme Epistasis*, International Conference on Intelligent Technologies in Human-Related Sciences (ITHURS'96) Vol. II, pp.251-258, León, 1996.