

Tarea 10: Algoritmos genéticos

Simulación de sistemas

17 de octubre de 2017

1. Introducción

Los algoritmos genéticos fueron inventados por John Holland en los años de 1960. Dichos algoritmos se inspiran en la evolución biológica para resolver problemas ya sea de búsqueda o de optimización.

Los algoritmos genéticos trabajan con una población de individuos, cada uno de ellos representa una solución para un problema determinado. Las normas de la evolución, son sencillas, las especies evolucionan por medio de variaciones aleatorias (en este caso: mutaciones y recombinaciones) seguidas por la selección natural, donde el mejor se inclina a sobrevivir y reproducirse, propagando así su material genético a las siguientes generaciones.

En esta práctica se realizó un algoritmo genético para resolver el problema de la mochila y se comparó la mejor solución alcanzada con la solución óptima obtenida con un método exacto.

2. Especificaciones computacionales

La presente tarea se realizó en una máquina con las siguientes especificaciones: procesador Intel(R)Core(TM) i5-6200U CPU 2.30 GHz 2.40 GHz con 8GB en memoria RAM y sistema operativo Windows 10 Home. Se emplearon tres de los cuatro núcleos.

3. Tarea

En esta tarea se realizó una versión paralela para el algoritmo genético de la práctica, además se comparó los tiempos de ejecución obtenidos con esta versión paralela y la versión secuencial.

Para realizar la versión paralela se crearon cuatro funciones: `fun.mutar`, `fun.reproducción`, `val.objetivo` y `fun.factibilidad`, las cuales fueron paralelizadas. Cada una de ellas realiza una fase fundamental del algoritmo genético.

Se ejecutaron las versiones secuencial y paralela para comparar los tiempos de ejecución para veinte réplicas con cincuenta objetos, treinta generaciones y población inicial de 50, 100, 200 y 400 individuos. Los resultados se muestran en la figura 1.

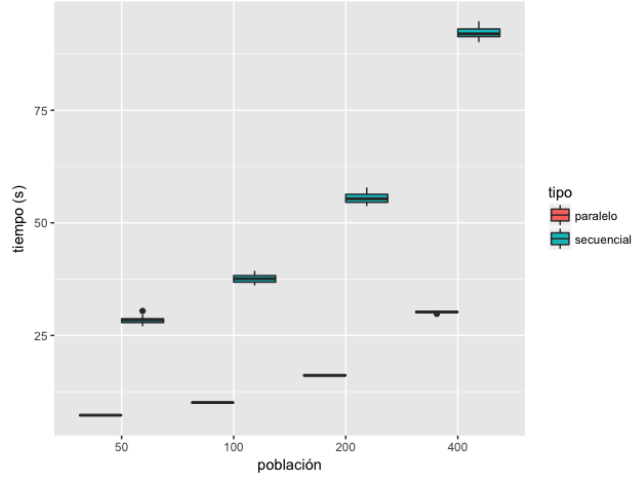


Figura 1: Tiempo de ejecución de veinte replicas con población inicial de 50, 100, 200 y 400 individuos.

En la figura 1 podemos observar que la versión paralela toma menores tiempos de ejecución para cualquier valor de la población inicial que tomemos, incluso cuando la población inicial es pequeña la versión secuencial se demora más.

4. Reto 1

El primer reto consiste en modificar la fase de reproducción, haciendo la selección de los padres por medio de ruleta y estudiar si esta modificación produce una mejora estadísticamente significativa en la calidad de la solución. Para poder realizar esto se calculó la probabilidad de cada individuo para ser seleccionado como padre, dicha probabilidad es directamente proporcional a su valor de función objetivo y esta información se guardó en un vector de probabilidades, luego, con ayuda del parámetro `prob` en la función `sample` se seleccionaron pares de individuos para reproducirse.

Para esta nueva versión se ejecutaron veinte réplicas con cincuenta objetos, treinta generaciones y población inicial de doscientos individuos. En la figura 2 se muestran los resultados.

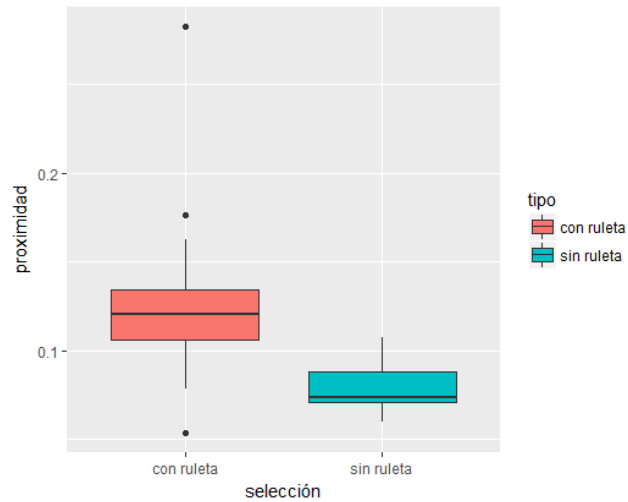


Figura 2: Proximidad al óptimo de veinte réplicas con selección de padres con: ruleta y sin ruleta.

Como vemos en la figura 2 el modo de selección sin ruleta da soluciones más cercanas al óptimo, ahora verificaremos si esta mejora es significativa. Examinamos primero si nuestros datos son normales utilizando la prueba de *Shapiro-Wilk*. Los resultados son mostrados en seguida.

Shapiro–Wilk normality test

```
data:  reto1$proximidad
W = 0.81201, p-value = 1.218e-05
```

Como el p -valor es menor a 0.05 se asume que los datos no cumplen normalidad, por tanto, es necesario utilizar una prueba no paramétrica. La prueba estadística no paramétrica por la que se optó es *Kruskal-Wallis*, cuyos contrastes de hipótesis son los siguientes:

- H_0 : Las medias son iguales.
- H_1 : Las medias son distintas.

Los resultados de la prueba los podemos ver a continuación.

Kruskal–Wallis rank sum test

```
data:  sin.ruleta by con.ruleta
Kruskal–Wallis chi-squared = 19, df = 19,
p-value = 0.4568
```

Como podemos observar, el p -valor es mayor que 0.05 se acepta la hipótesis nula, lo que da lugar a concluir que no existen diferencias significativas.

5. Reto2

El segundo reto consiste en realizar el método de la ruleta en la fase de supervivencia, es decir, cada individuo podrá pasar a la siguiente generación con una probabilidad que es directamente proporcional a su valor en la función objetivo. También se requiere estudiar el efecto de esta modificación en la calidad de la solución. Los cambios realizados para esta versión fueron similares a los del reto 1. Para esta versión se ejecutaron veinte réplicas con cincuenta objetos, treinta generaciones y población inicial de doscientos individuos. En la figura 3 se muestran los resultados obtenidos.

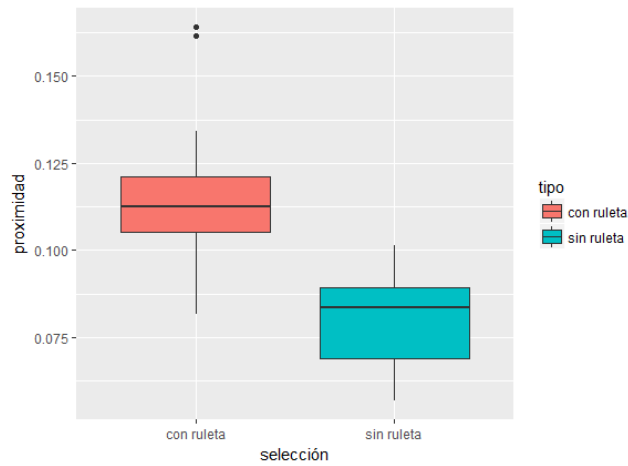


Figura 3: Proximidad al óptimo de veinte réplicas con selección en la fase de supervivencia con: ruleta y sin ruleta.

Observamos en la figura 3 que el modo de selección con ruleta en la fase de supervivencia da soluciones más alejadas al óptimo, ahora verificaremos si este cambio es significativo. Primero examinamos si los datos son normales utilizando la prueba de *Shapiro-Wilk*. Los resultados obtenidos son los siguientes.

Shapiro–Wilk normality test	
data:	datos2
W =	0.95918, p-value = 0.157

El p -valor es mayor a 0.05 se asume que los datos son normales. Como los datos son normales podemos realizar un análisis de varianza (ANOVA). Recordando que los contrastes de hipótesis de ANOVA son los siguientes:

- H_0 : No existen diferencias entre los tratamientos.
- H_1 : Existen diferencias entre los tratamientos.

El resumen de la tabla ANOVA se muestra a continuación.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
ruleta	1	0.01309	0.013088	41.92	1.27e-07 ***
Residuals	38	0.01186	0.000312		
<hr/>					
Signif. codes:	0	***	0.001	**	0.01 * 0.05 . 0.1 1

Ahora obtenemos el valor crítico de F bajo la hipótesis nula con un nivel de significancia $\alpha = 0.05$.

```
qf(0.05, 2-1, 20-2, lower.tail = F)
[1] 4.413873
```

Los valores del estadístico mayores a 4.413873 pertenecerán a la región de rechazo. En nuestro caso 41.92 es mucho mayor que el valor crítico obtenido por tanto se rechaza la hipótesis nula. Concluyendo de esta manera que existen diferencias significativas.

6. Conclusiones

Hemos observado que es conveniente paralelizar el algoritmo genético sin importar el tamaño de la población inicial, puesto que los tiempos de ejecución son mucho menores comparados con los tiempos obtenidos de manera secuencial. Por otro lado, si deseamos tener mejores soluciones, es decir, soluciones de calidad, la selección de padres por medio de ruleta no siempre es la mejor opción ya que los resultados obtenidos en el reto 1 nos muestran que la selección de padres mediante este método no proporciona soluciones con mayor proximidad al óptimo que al realizar la selección de padres sin la intervención de ruleta. De la misma manera, intentar tener soluciones de mejor calidad utilizando el método de la ruleta en la fase de supervivencia tampoco funcionó para nuestro caso ya que en el reto 2 pudimos verificar que la supervivencia con ruleta proporciona soluciones más alejadas del óptimo que la versión de supervivencia sin ruleta.