

Sistemas de Inteligencia Artificial

TRABAJO PRÁCTICO ESPECIAL 3 - INFORME

ALGORITMOS GENÉTICOS

GRUPO 1 - 1C 2017

DOCENTES

- Parpaglione, María Cristina
- Pierri, Alan

INTEGRANTES

- Comercio Vázquez, Matías Nicolás - 55309
- Ibars Ingman, Gonzalo - 54126
- Mercado, Matías - 55019
- Moreno, Juan - 54182

Introducción	2
Herramientas de análisis	2
Casos de análisis	2
Resultados	3
Conclusiones	5
Anexo	6
Configuración preliminar a los casos de prueba	6
Configuración inicial	6
Configuración final	7
Posible variante	7
Gráficos	8
Caso 1	8
Caso 2	11
Caso 3	12
Caso 4	13

Introducción

El presente trabajo tiene como objetivo la utilización de algoritmos genéticos para maximizar el desempeño (fitness) óptimo del personaje asignado al grupo: *Asesino 3*.

Herramientas de análisis

En cada generación se actualiza un gráfico del fitness máximo, promedio y mínimo, junto con una serie de datos estadísticos de relevancia para el análisis de la evolución de la población inicial y el desempeño de la misma.

La distancia entre el fitness máximo y el medio permite determinar si el fitness de la población ha convergido. La distancia entre el fitness medio y el mínimo permite estipular acerca de la tolerancia sobre la presencia de individuos no aptos en la población.

La variabilidad del fitness máximo permite identificar si al menos se está tomado o no el individuo más apto en cada generación para pasarlo a la siguiente. La variabilidad del fitness medio indica cuán laxo es el algoritmo a la hora de permitir la selección de individuos más o menos aptos para pasar a la siguiente generación. La variabilidad del fitness mínimo permite analizar, en líneas generales, cuán tolerante es el algoritmo para la selección de individuos no tan aptos en el reemplazo a lo largo de las generaciones.

Se muestran a su vez los genes del individuo más apto de la población. También se muestran, numéricamente, el máximo fitness de dicho individuo, y el fitness medio y mínimo de la población.

Para un seguimiento más detallado, se exhibe el número de generación actual, la última generación en la que aumentó el máximo fitness y de cuánto fue dicho aumento.

Asimismo, se muestran la cantidad de individuos que tienen fitness máximo y la cantidad que tiene su fitness por debajo de la media para poder apreciar, de manera aproximada, cuál es la distribución de fitness de la población a lo largo de las generaciones.

Finalmente, se muestran la cantidad de individuos únicos en la población actual (el parámetro se identifica en el gráfico como *lds*) y la cantidad de alelos únicos presentes entre todos los individuos de la población, lo que permite medir la diversidad presente en la generación actual.

Casos de análisis

En el primer caso de análisis se decidió variar los métodos de selección para analizar la mejora del fitness a través de las generaciones según el tipo de padres que se seleccionaron para la cruce.

En el segundo caso de análisis se varió la probabilidad de mutación para ver cómo la misma afectaba la evolución del fitness de la población en general. También se realizó una comparación entre los métodos de reemplazo 2 y brecha generacional.

En el tercer caso de análisis se decidió comparar dos funciones de mutación: una donde la probabilidad de mutación se aplica a cada individuo y otra donde dicha probabilidad se aplica a cada gen de cada individuo.

En el cuarto y último caso de análisis se decidió evaluar el porcentaje de selección de elite a utilizar para la selección de individuos previo a la cruce. Se analizó cómo dicha selección afectaba a las curvas de fitness y a la evolución de la población.

Resultados

Aclaraciones preliminares

- Para evitar que disminuya el máximo fitness a lo largo de las generaciones, el algoritmo de selección para el reemplazo realiza siempre un porcentaje de la selección utilizando la selección de élite.
- El método de reemplazo 1 también introduce la posibilidad de pérdida del máximo fitness, con lo que, si bien fue evaluado, no se utilizó para las pruebas.
- Salvo especificado lo contrario, los parámetros **iniciales** utilizados son los detallados en la sección [Anexo](#).
- A excepción del caso de análisis 4 se utilizará el mismo método de selección estocástico para elegir los padres para la cruce y para acompañar a la selección élite en el reemplazo.
- Los gráficos de cada caso analizado se encuentran en la sección [Anexo](#), identificados con el número y letra de caso analizado.
- Al hablar de la cantidad de puntos de distancia entre el fitness máximo y el medio, debe entenderse que $\#puntos \approx fitness\ máximo - fitness\ medio$.

Caso 1

1) Boltzmann: el fitness medio, con el correr de las generaciones, converge al fitness máximo de manera consistente a la función de temperatura elegida (en este caso, exponencial por épocas - ver código fuente), permitiendo una población con gran diversidad en las primeras generaciones que disminuye a medida que la temperatura disminuye. **Se obtuvo el mejor individuo de todas las corridas del algoritmo para el caso 1, cuyo fitness fue aproximadamente 34.7868.**

2) Ruleta: el fitness medio *copia* la curva de fitness máximo, aunque se encuentran distanciados constantemente (desde la primera hasta la última generación de la ejecución) por aproximadamente 6 puntos. Esto permite que haya diversidad en la población en todas las generaciones.

3) Universal: comportamiento de las curvas de fitness similar al obtenido con ruleta, aunque se obtuvieron fitness más altos, no sólo de manera más temprana, sino también en comparación de los fitness finales¹. Esto se debe a la media y constante diversidad de genes ofrecida por este método.

4) Torneos determinísticos: el fitness medio converge rápidamente al máximo, y se mantiene distanciado por aproximadamente 2 puntos. Esto puede explicarse por el hecho de que se aumenta la probabilidad de tomar varias veces al mismo individuo para la cruce y para el reemplazo, agotando rápidamente la diversidad en la población que queda determinada, en mayor medida, por la probabilidad de mutación, lo que lleva a un rápido estancamiento del fitness máximo en caso de que la probabilidad de mutación no sea alta.

5) Torneos probabilísticos: al darle posibilidad de ganar al individuo con peor fitness, el fitness medio y máximo se mantienen distanciados ahora por 3 puntos en lugar de los 2 del caso anterior. El resto del análisis es similar al del método anterior.

¹ El criterio de corte utilizado es por estancamiento de fitness máximo luego de 500 generaciones, sin que el mismo aumente más de 0.001.

6) Ranking: la distancia entre el fitness máximo y medio se mantiene aproximadamente constante alrededor de 1.5 o 2 puntos. Incluso con el mismo criterio de corte que para el resto de los casos recién mencionados, el máximo fitness fue el peor de entre todos los obtenidos anteriormente.

7) Elite: se esperaba obtener un fitness medio mucho más cercano al máximo, que derivara en una convergencia prematura. Sin embargo, esto no sucedió, y puede deberse a dos motivos: al estar utilizando el método de reemplazo 2, se toman los hijos tal cual son (sean buenos o malos), con lo cual en cada generación habrá individuos por debajo del fitness medio; además, al ser media la probabilidad de mutación, había altas posibilidades de introducir un nuevo alelo que, combinado con la cruce de los mejores individuos en cada generación, provocara la aparición de un nuevo individuo más apto. Se aprende de este caso la importancia de la diversidad genética, de la mutabilidad y de la necesidad de tener una porción elitista de la población.

Caso 2

Se aumentó la probabilidad de mutación a 0.1 por gen.

1) Selección estocástica con Boltzmann (siguiendo los lineamientos del caso 1) con la temperatura que converge a partir de la generación 1500, para asegurar mayor diversidad en las primeras generaciones, favoreciendo la posibilidad de aparición de individuos más aptos, que sean cruce de individuos aptos y no tan aptos. Para el reemplazo se tomó una brecha generacional de 90%.

La alta probabilidad de mutación junto al hecho de que el método de selección para el reemplazo provocaba que se eligieran individuos, en general, repetidos (sobre todo a medida que disminuía la temperatura), por lo que la cantidad de individuos únicos en la población era menor que el 60%, lo que provocaba que la diversidad disponible para la cruce fuese muy baja (ver los ítems en la figura 8 en [Anexo](#)), y la generación de individuos mejores a los existentes dependiese fuertemente de que la mutación provocara un cambio *oportuno* en los genes de los individuos.

2) Se volvió a utilizar el método de reemplazo 2 para asegurar que todos los hijos sean pasados a la siguiente generación (y así asegurar un aumento de la diversidad). El resto de los parámetros se mantuvieron como en el caso anterior. **Con esta configuración, se obtuvo el mejor individuo de todas las pruebas realizadas (fitness = 34.899).** Esto puede explicarse tanto por la alta diversidad de alelos disponibles a la hora de la cruce (causada por el método de reemplazo 2) como por la alta probabilidad de mutación, que permite iterar sobre la vasta colección de alelos disponibles (en total 5 millones). Dado que la cantidad de ítems es tan alta, la introducción de esta cuota de mutabilidad puede ser de gran ayuda para escapar de máximos locales y para la exploración de la amplia gama de genes disponibles.

Caso 3

Se utilizó probabilidad de mutación por individuo.

Se puede notar que la probabilidad de mutación, aunque numéricamente fuese la misma, disminuía considerablemente con respecto a la implementación anterior. Se pudo observar que si bien cada mutación sucedía con una probabilidad menor, cuando ocurría elevaba el fitness, en general, más que cualquier mutación que ocurría con el método de mutación anterior. Lo que se está haciendo en definitiva con este método de mutación es fomentar la

explotación de los alelos presentes en la población (a través de las distintas combinaciones dadas por la cruza) por sobre la exploración de nuevos alelos.

Luego de la convergencia se modificaron varios parámetros de la configuración (entre ellos la temperatura, la probabilidad de mutación y el porcentaje de selección elitista) para disminuir el fitness promedio, aumentando así la diversidad y provocando el escape de máximos locales. El detalle de cómo se modificaron esos parámetros se omite por falta de espacio para la redacción de todos los casos.

Caso 4

Porcentaje de selección élite para la cruza (10%); el resto con Boltzmann. Se disminuyó la probabilidad de mutación a 0.05, y se utilizó la mutación por individuo (por los resultados obtenidos en el caso 3). Se cambió el criterio de corte: si el gen con mayor cantidad de alelos presentes en la población tiene una cantidad de alelos menor que un límite seteado (el valor de este límite se estableció en 15), se termina el algoritmo.

Se pudo notar que al hacer esto, el fitness medio converge más rápidamente al fitness máximo, lo cual era de esperarse por estar tomando más individuos aptos para la cruza. Además, se pudo observar que en pocas generaciones se obtienen individuos con un elevado fitness por el mismo motivo que el explicado recientemente.

En pruebas consecutivas se varió el porcentaje de elitismo de la selección para la cruza y para el reemplazo, y se pudo apreciar, como era de esperarse, que cuanto más bajo es el porcentaje de élite tomado, más lenta es la convergencia del fitness medio al fitness máximo a lo largo de las generaciones, favoreciendo la diversidad genética disponible para la cruza.

Conclusiones

En base a los casos analizados (mencionados y no en el presente informe) se puede concluir que, para que encontrar los individuos más aptos utilizando algoritmos genéticos, para el caso en estudio (configuración de *Asesino* 3), es necesario dejar abierta la posibilidad de exploración de nuevos genes y combinaciones de ellos en todo momento, que en el caso del presente trabajo se realiza través de la utilización del método de reemplazo 2 y del uso de una probabilidad de mutación media a alta.

La probabilidad de mutación puede aplicarse por individuo si se quieren introducir pocos alelos nuevos y explorar la combinación de los presentes en la población, o por atributo si lo que se quiere es darle prioridad a la exploración de la combinación de los alelos presentes con nuevos.²

Además, es recomendable que un porcentaje de la selección, tanto para la cruza como para el reemplazo, se realice de manera elitista, no sólo para asegurar la permanencia del máximo fitness a lo largo de las generaciones, sino también para aumentar la probabilidad de generar individuos por sobre el fitness promedio.

El mejor fitness obtenido fue 34.899 (ver individuo en la figura 9 en [Anexo](#)).

² Dada la amplia diversidad de alelos (1 millón por cada ítem + rango de variabilidad de la altura), se prefiere la mutación por atributos para favorecer la exploración de nuevos alelos.

Anexo

Configuración preliminar a los casos de prueba

Ésta fue la primera configuración que se probó y que se fue mejorando previo a los análisis, y con criterios que no se exponen en el informe por limitación de espacio.

- Tamaño de la población (N) = 100
- K de selección (K) = 90
- Probabilidad de cruza (pc) = .8
- Probabilidad de mutación (pm) = .003
- Método de mutación = mutación por atributos
- Función de la temperatura = $\exp(-\text{generation}) + 1$
- Selección para la cruza = 100% Boltzmann
- Selección para el reemplazo = 10% Élite + 90% Boltzmann
- Método de cruza = Cruza en un punto
- Método de reemplazo = Reemplazo por brecha generacional
- G = 0.8
- Método de finalización = Cantidad máxima de generaciones: 1000

Configuración inicial

A continuación se especifican los parámetros utilizados inicialmente para los casos que luego fueron modificándose según los resultados obtenidos para poder obtener mejores resultados en corridas sucesivas.

- Tamaño de la población (N) = 1000
- K de selección (K) = 900
- Probabilidad de cruza (pc) = .9
- Probabilidad de mutación (pm) = .05
- Método de mutación = mutación por atributos
- Función de la temperatura = $\exp(-\text{generation} * .005 + 5) + 1$
- Selección para la cruza = 100% Boltzmann
- Selección para el reemplazo = 10% Élite + 90% Boltzmann
- Método de cruza = Cruza en dos puntos³
- Método de reemplazo = Método de reemplazo 2
- Método de finalización = Máximo fitness sin variar por un delta mayor que .001 durante 500 generaciones

³ El resto de las funciones de cruza fueron evaluadas, pero dada la limitación en la extensión del trabajo no se hace mención de las mismas, debiéndose entender que esta función fue elegida por los buenos resultados que produjo al realizar las pruebas cuyos resultados fueron omitidos por lo recién mencionado.

Configuración final

En base a los lineamientos de las conclusiones, se encuentra que para comenzar a correr el algoritmo, la siguiente configuración ofrece altas probabilidades de encontrar individuos con un fitness cercano al máximo encontrado.

- Tamaño de la población (N) = 1000
- K de selección (K) = 900
- Probabilidad de cruza (pc) = .95
- Probabilidad de mutación (pm) = .05
- Método de mutación = mutación por atributo
- Función de la temperatura = $\exp(-\text{generation} * .005 + 5) + 1$
- Selección para la cruza = 5% Élite + 95% Boltzmann
- Selección para el reemplazo = 5% Élite + 95% Boltzmann
- Método de cruza = Cruza en dos puntos
- Método de reemplazo = Método de reemplazo 2
- Método de finalización = Máximo fitness sin variar por un delta mayor que .001 durante 500 generaciones⁴

Posible variante

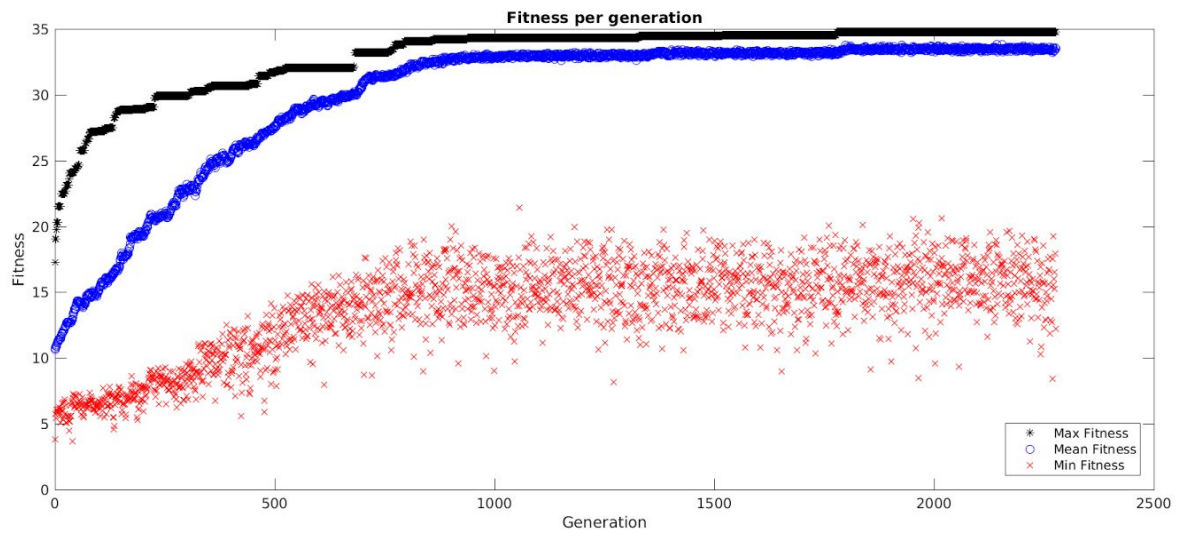
Como se explicó en el **Caso 3** y en la conclusión, según se deseen explorar nuevos alelos o explotar los actuales, se pueden modificar los parámetros de una u otra forma. Para favorecer la exploración de nuevos alelos, usar los parámetros mencionados anteriormente. Para favorecer la explotación de los actuales, cambiar los parámetros como se muestra a continuación:

- Probabilidad de mutación (pm) = .09
- Método de mutación = mutación por individuo

⁴ Este criterio de corte puede ser reemplazado por cualquiera de los otros, de manera indistinta.

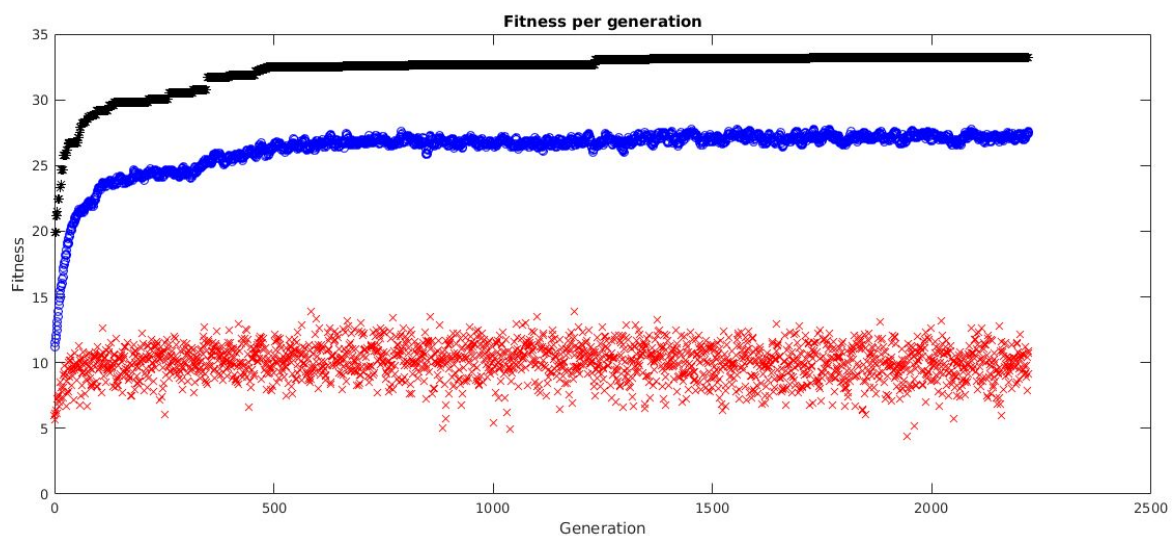
Gráficos

Caso 1



Character: Height: 1.915205, Boots: 921745, Chest: 86387, Gloves: 50682, Helmet: 81607, Weapon: 988523
Max fitness (current character): 34.786836; Mean fitness: 33.507433
Generation: 2277
Individuals with max fitness: 660
Unique: Ids = 982, Boots = 53, Helmets = 40, Gloves = 49, Chests = 41, Weapons = 41

Figura 1 - Método de selección Boltzmann



Character: Height: 1.915180, Boots: 793770, Chest: 960007, Gloves: 805631, Helmet: 199578, Weapon: 968658
Max fitness (current character): 33.200083; Mean fitness: 27.511847
Generation: 2221
Individuals with max fitness: 95
Unique: Ids = 985, Boots = 158, Helmets = 115, Gloves = 182, Chests = 108, Weapons = 76

Figura 2 - Método de selección Ruleta

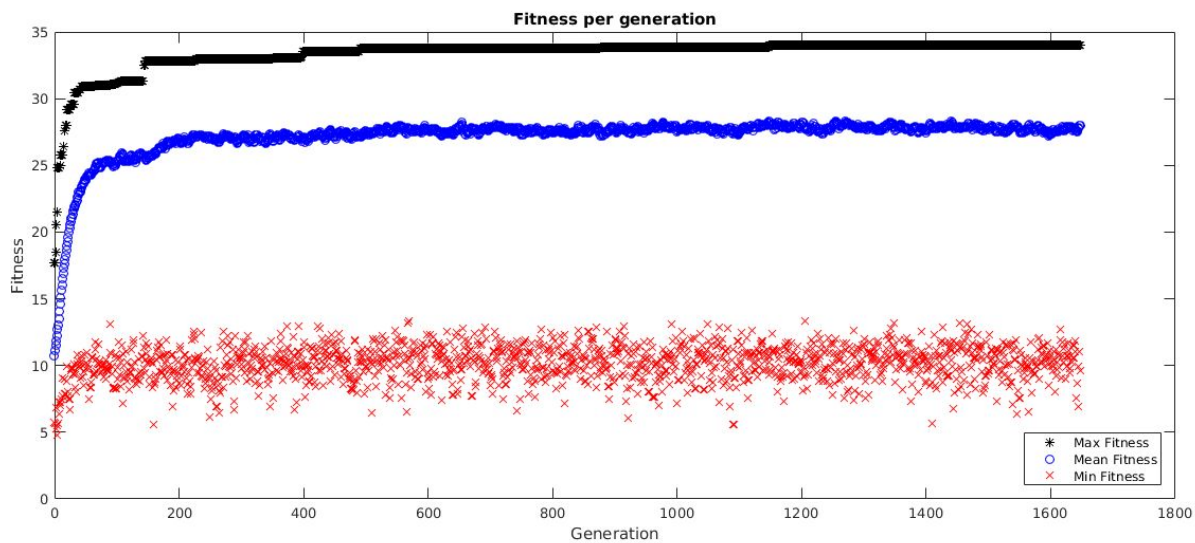


Figura 3 - Método de selección Universal

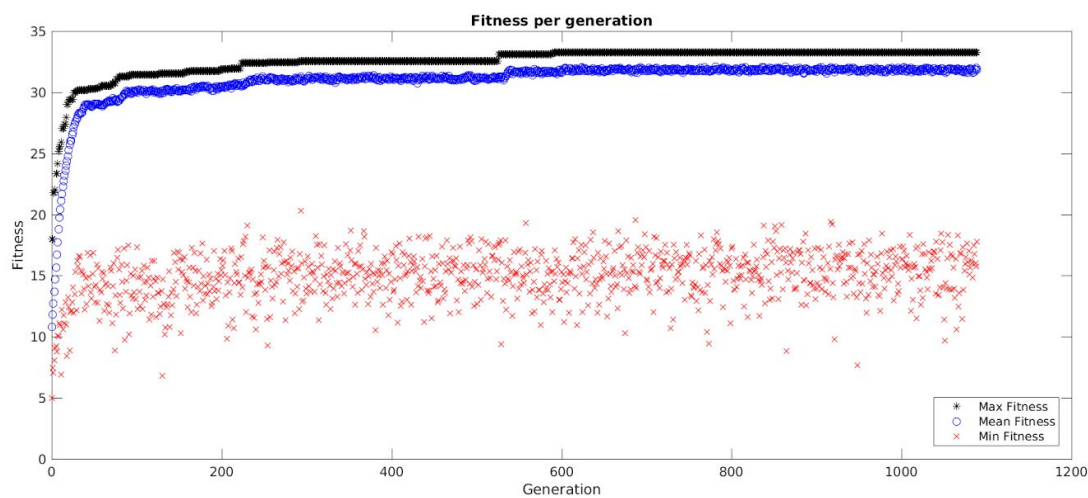
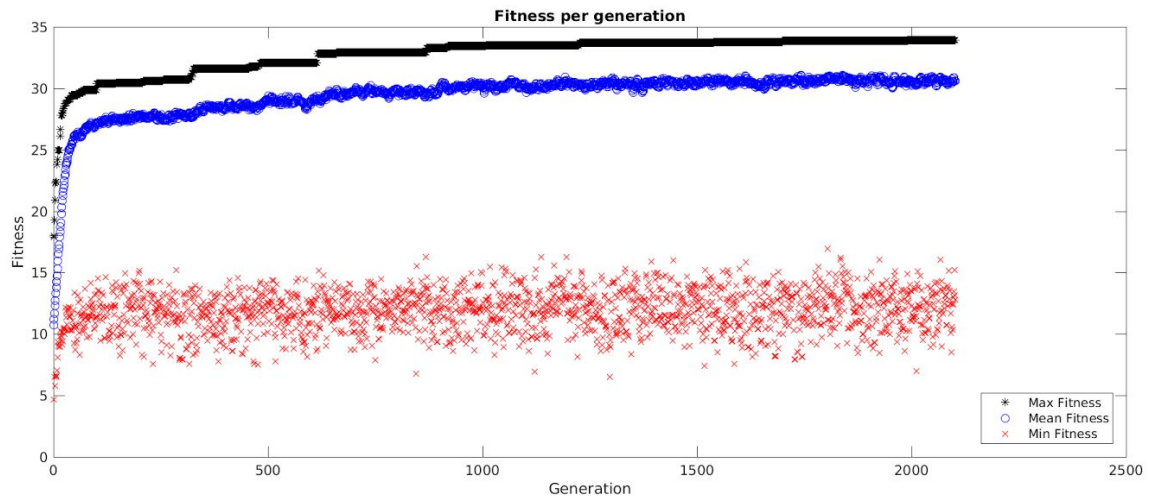
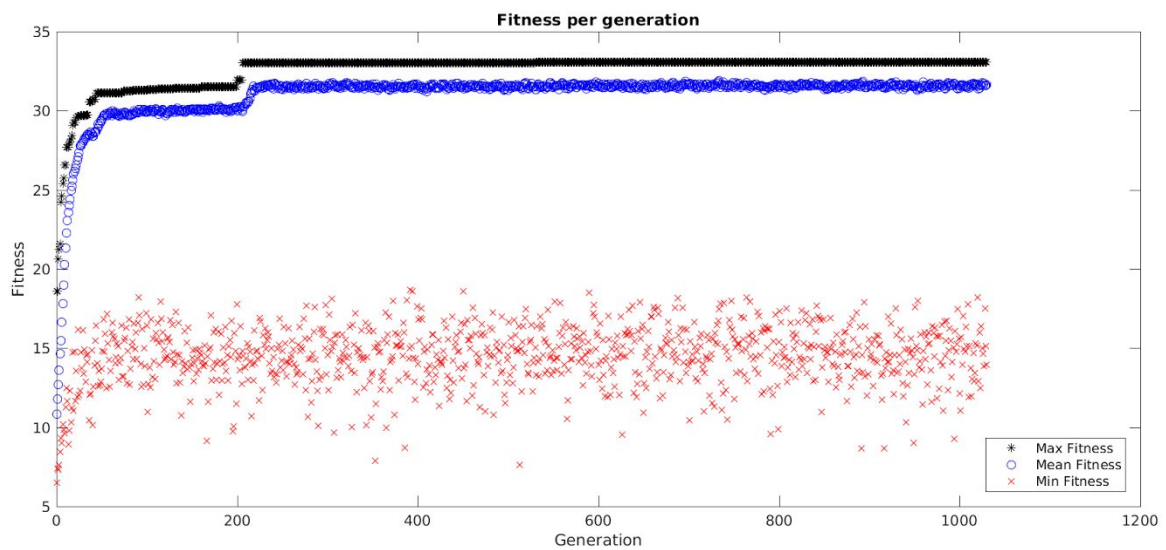


Figura 4 - Método de selección Torneos Determinísticos



Character: Height: 1.915181, Boots: 356572, Chest: 848491, Gloves: 352353, Helmet: 283046, Weapon: 572499
 Max fitness (current character): 33.920144; Mean fitness: 30.582982
 Generation: 2102
 Individuals with max fitness: 436
 Unique: Ids = 976, Boots = 106, Helmets = 106, Gloves = 116, Chests = 87, Weapons = 83

Figura 5 - Método de selección Torneos Probabilísticos



Character: Height: 1.915202, Boots: 40222, Chest: 210915, Gloves: 562257, Helmet: 279951, Weapon: 583288
 Max fitness (current character): 33.079297; Mean fitness: 31.651830
 Generation: 1030
 Individuals with max fitness: 734
 Unique: Ids = 975, Boots = 65, Helmets = 54, Gloves = 60, Chests = 35, Weapons = 39

Figura 6 - Método de selección Ranking

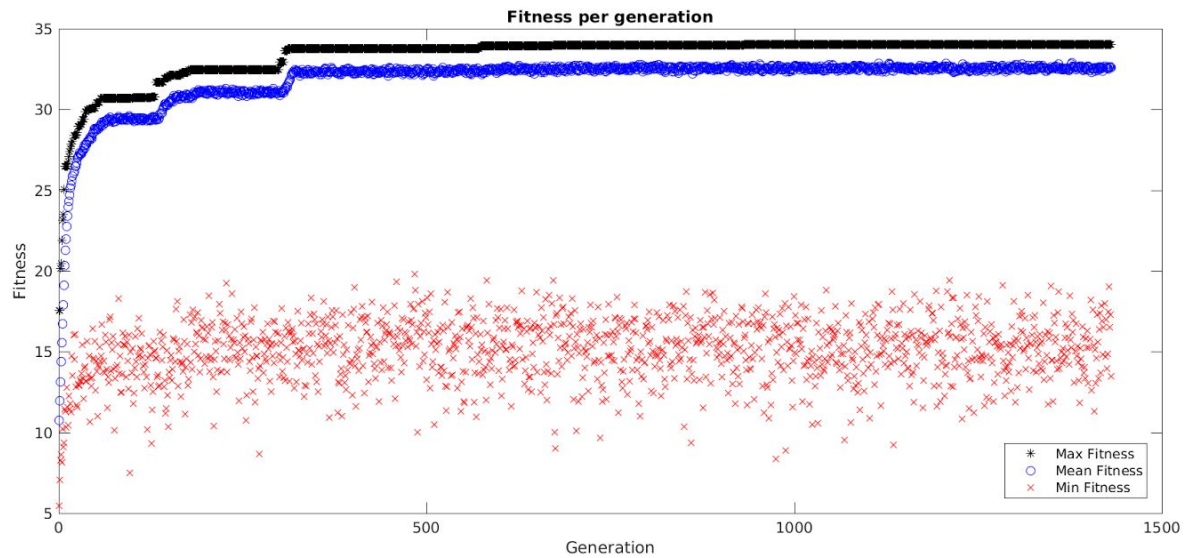


Figura 7 - Método de selección Élite

Caso 2

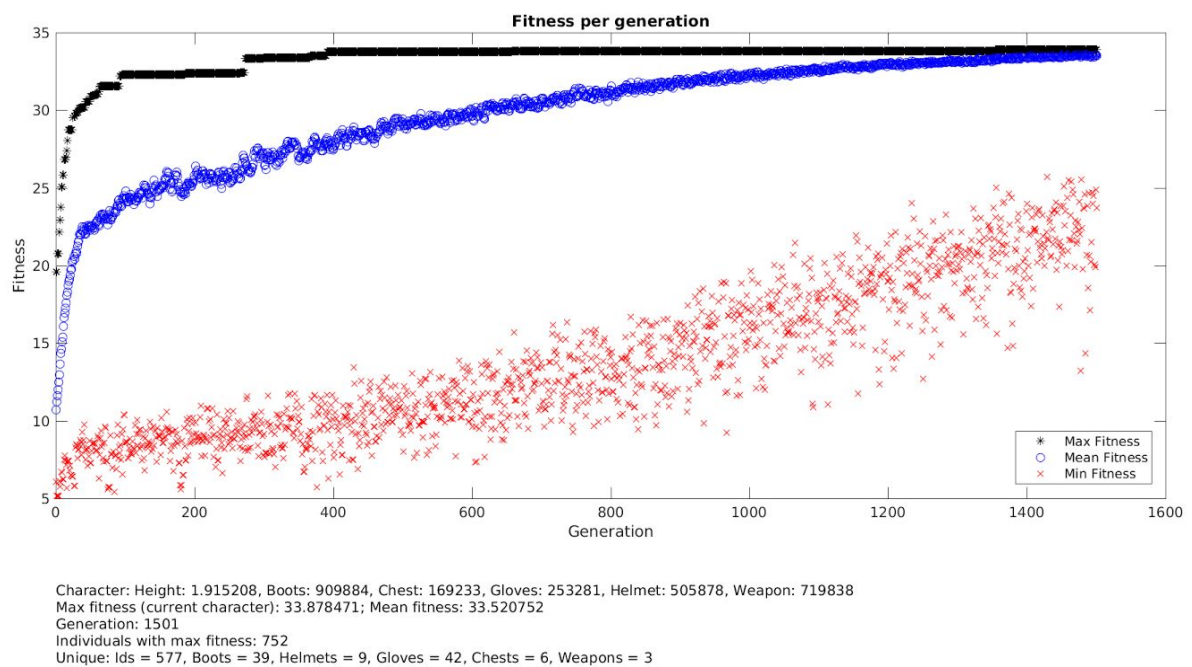


Figura 8 - Método de selección Boltzmann - Probabilidad de mutación 0.1 - Reemplazo por Brecha Generacional

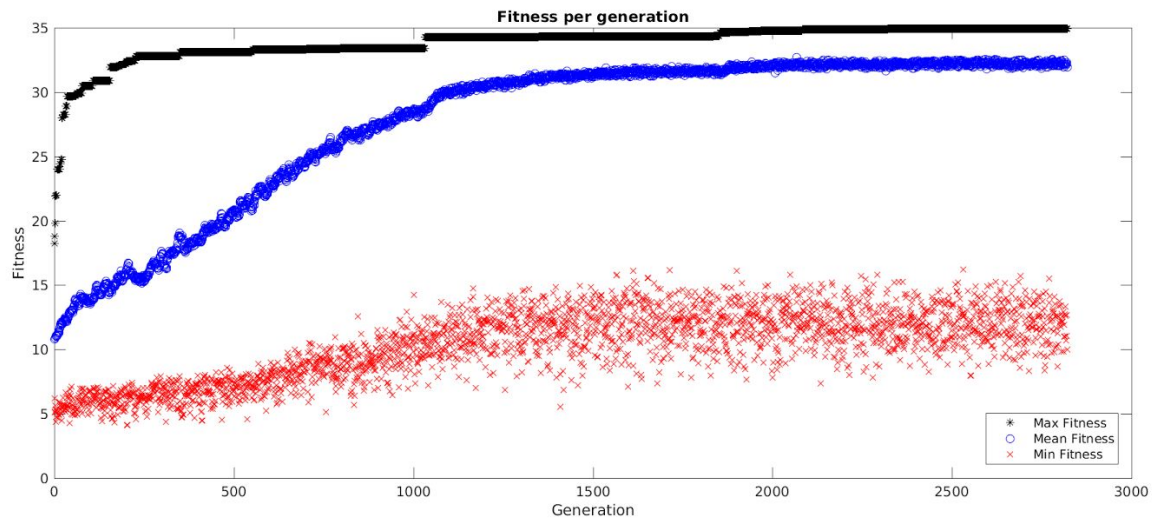


Figura 9 - Método de selección Boltzmann - Probabilidad de mutación 0.1 - Método de Reemplazo 2 - **Máximo fitness de todas las pruebas**

Caso 3

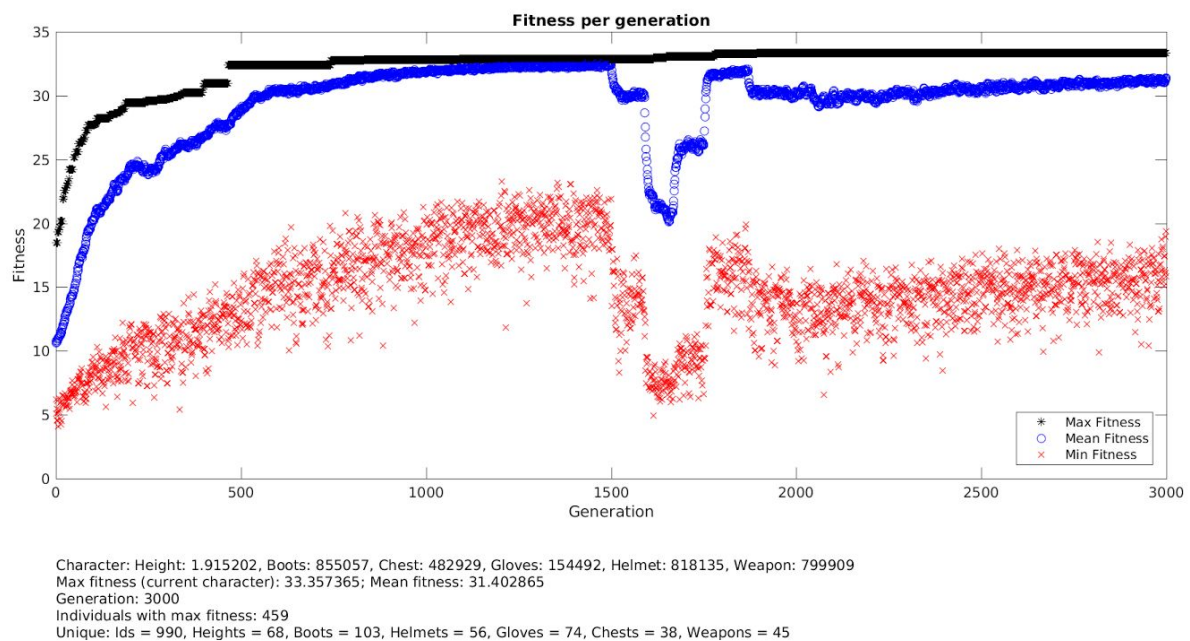


Figura 10 - Probabilidad de mutación por individuo - Luego de la convergencia, se modificaron varios parámetros de la configuración (entre ellos, la temperatura, la probabilidad de mutación y el porcentaje de selección elitista), para disminuir el fitness promedio, aumentando así la diversidad y provocando el escape de máximos locales.

Caso 4

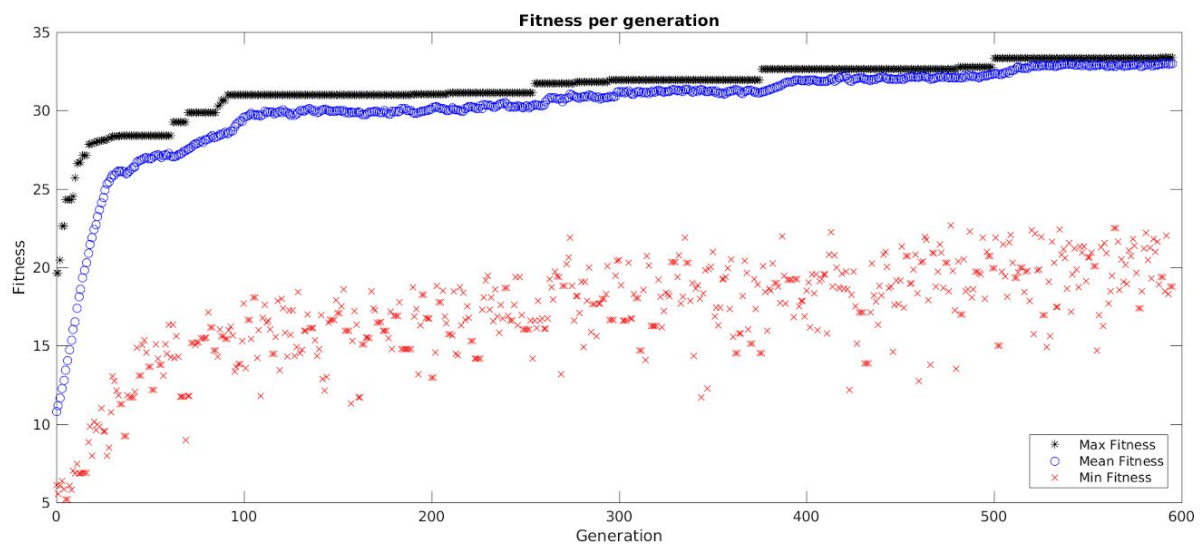


Figura 11 - Porcentaje 10% de Selección Élite para la cruza

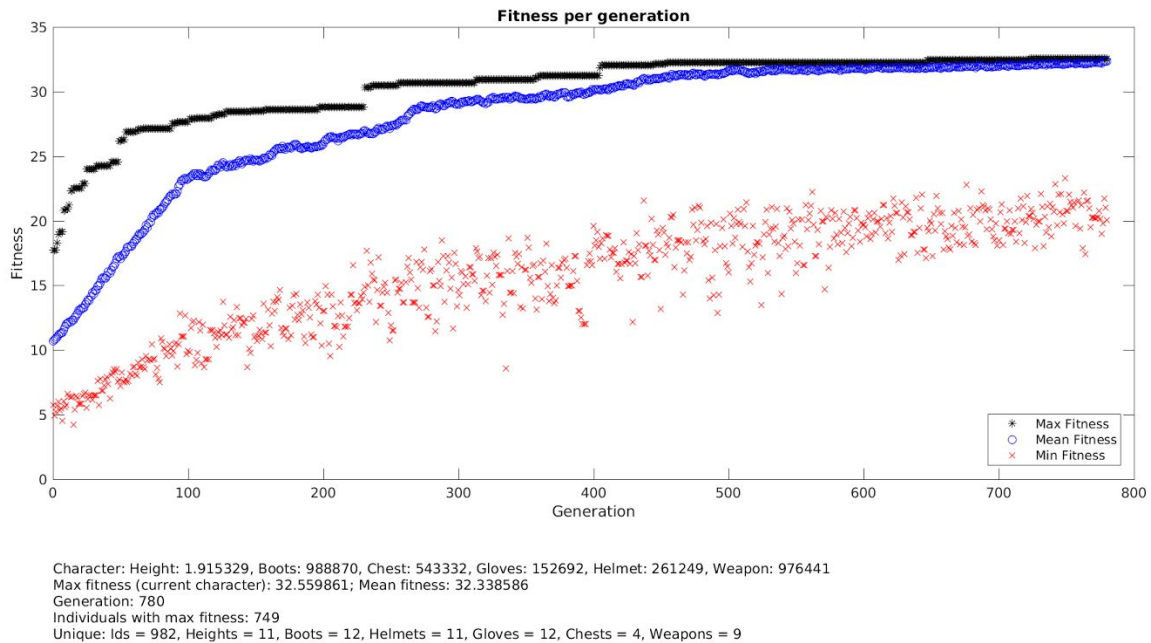


Figura 12 - Porcentaje 0% de Selección Élite para la cruza - Se puede apreciar que el máximo fitness alcanzado es menor que en la figura anterior, pero que se necesitan más generaciones para alcanzar el mismo nivel de baja diversidad.