Universidad Zaragoza
1542

Shamuel Manuel Manrrique Aquino
NIP: 802400

Busqueda de política

1. Question 1: Random and Naive Agents

En esta sección se implementó la función Random Agent con la cual se busca determinar una acción de forma aleatoria. Para lograrlo se crea un vector de tamaño dos, con valores aleatorios el cual es normalizado(se busca la distribución normal) y retornado para determinar una acción con base en el. La diferencia entre el Random y el Naive agente es que el primero produce una acción aleatoria sin tomar en cuenta el estado del sistema en lo absoluto. Por el contrario, la acción de la segunda heurística (naive) la acción va a ser determinada por los grado de la polea.

Para resaltar con mayor facilidad los resultados obtenidos con las dos aproximaciones se realizaron cinco pruebas en donde se extrajo el valor Min, Max y AVG del conjunto de episodios tanto en el entrenamiento como en el Test los mismos se encuentran en la Tabla 1. De la misma se puede deducir que en general los resultados del Agente ingenuo son mejores que el del random exceptuando que este último alcanza mejores valores de máximo que muchas veces no son llegados a explorar por el ingenuo. Como el random tiene su respaldo en la ley de número grandes se realizó pruebas incrementando el número de iteraciones a 10000 se obtuvo con Randon Training AVG 22.2505 , Test AVG 23.4 y con 15000 iteraciones Training AVG 22.398133333333334, Test AVG 25.2 mientras que con Naive Training AVG 42.04526666666667, Test AVG 43.0. En general los promedios de resultados del Training y Test tanto del Naive(40-50) como del Random(20-30) se mantienen en ese rango si se incrementa el número de iteraciones. Otro dato importante a notar es que en las cinco pruebas realizadas el en Naive en tres de ellas, arrojó el resultado promedio del Test fue mejor que el del entrenamiento. Mientras que en el Random cuatro de las pruebas arrojaron mejores resultados en el promedio del test. Es importante recalcar que estos resultados obtenidos y explicados anteriormente se refieren al valor de cada una de las heurísticas en las distintas pruebas(training-test), no a comparativas entre ellas.

Shamuel Manuel Manrrique Aquino
NIP: 802400

Figura 1: Resultados obtenidos con Random

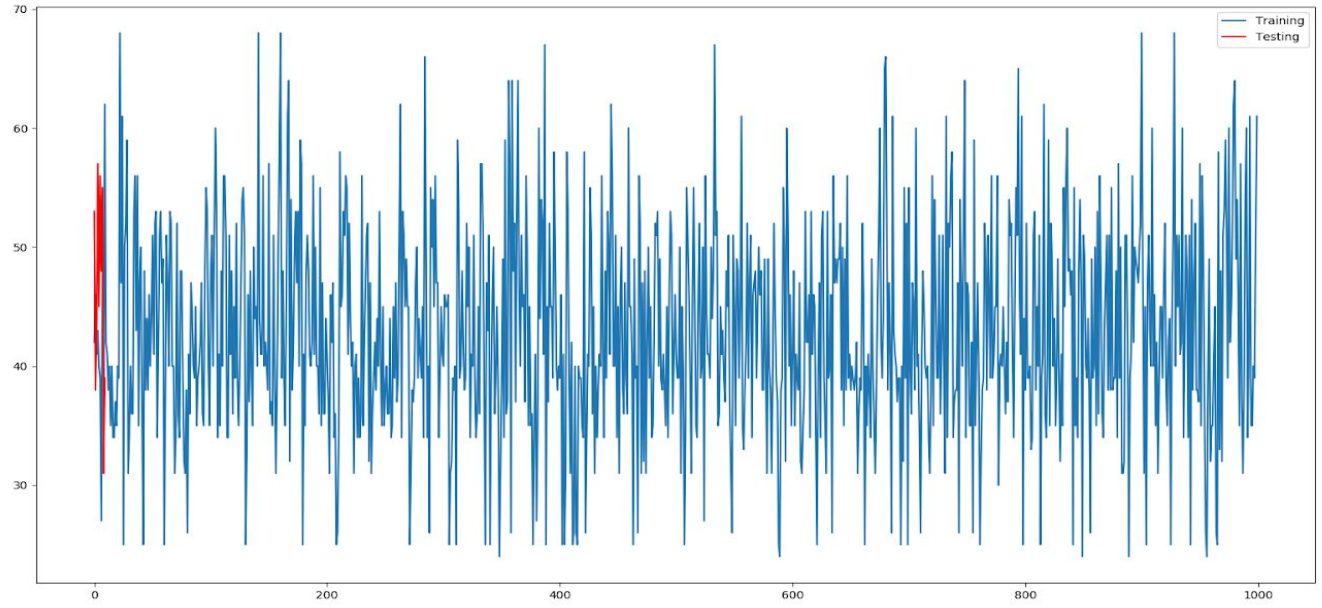
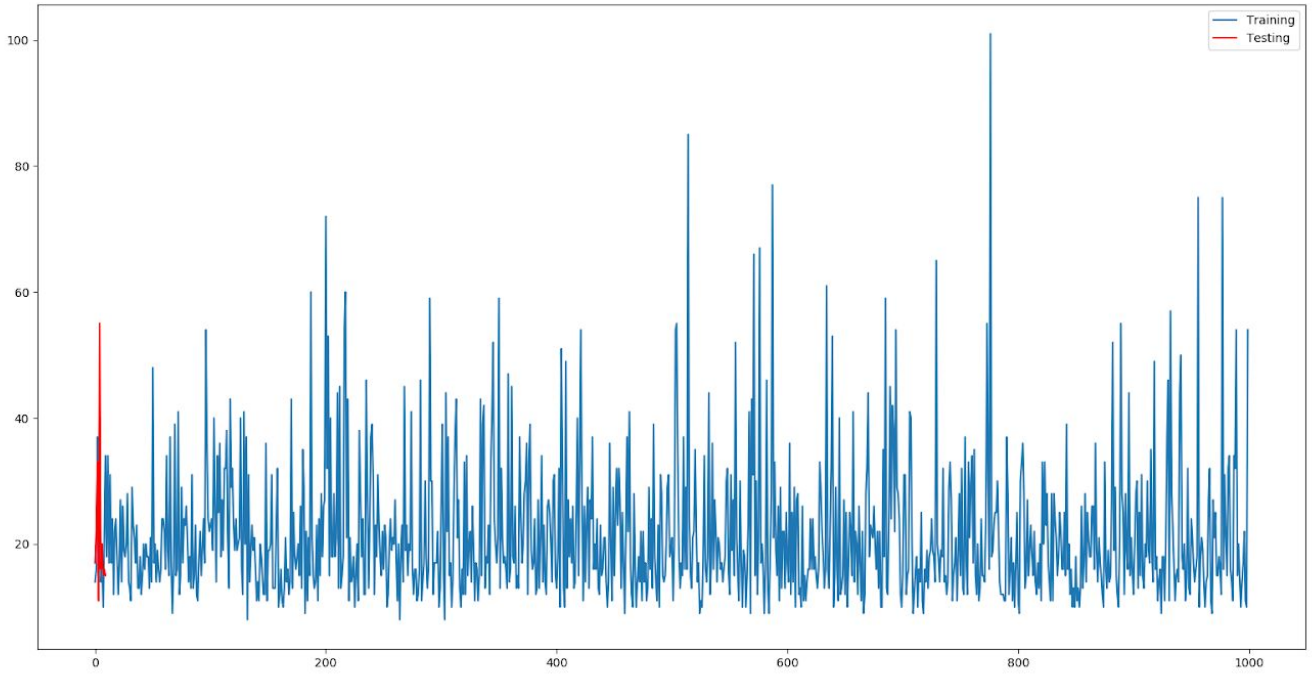


Figura 2: Resultados obtenidos con Naive



Universidad Zaragoza
1542

Shamuel Manuel Manrrique Aquino
NIP: 802400

Tabla 1: Resultados comparativo tras cinco pruebas con Naive Agent y Random Agent.

	Naive	Random
Prueba 1	Training Min 24.0	Training Min 8.0
	Training Max 71.0	Training Max 97.0
	Training AVG 42.004	Training AVG 22.63
	Test Min 38.0	Test Min 10.0
	Test Max 57.0	Test Max 69.0
	Test AVG 43.7	Test AVG 29.3
Prueba 2	Training Min 24.0	Training Min 8.0
	Training Max 71.0	Training Max 102.0
	Training AVG 42.111	Training AVG 22.328
	Test Min 32.0	Test Min 10.0
	Test Max 53.0	Test Max 41.0
	Test AVG 40.8	Test AVG 20.5
Prueba 3	Training Min 24.0	Training Min 8.0
	Training Max 68.0	Training Max 118.0
	Training AVG 42.804	Training AVG 22.53
	Test Min 31.0	Test Min 9.0
	Test Max 57.0	Test Max 54.0
	Test AVG 46.6	Test AVG 26.1
Prueba 4	Training Min 24.0	Training Min 8.0
	Training Max 72.0	Training Max 101.0
	Training AVG 42.227	Training AVG 21.659
	Test Min 32.0	Test Min 11.0
	Test Max 56.0	Test Max 55.0
	Test AVG 40.9	Test AVG 22.0
Prueba 5	Training Min 24.0	Training Min 9.0
	Training Max 68.0	Training Max 115.0
	Training AVG 41.925	Training AVG 22.004
	Test Min 25.0	Test Min 15.0
	Test Max 57.0	Test Max 52.0
	Test AVG 43.9	Test AVG 26.3

Shamuel Manuel Manrrique Aquino
NIP: 802400

2. Question 2: Linear Agent

En esta parte nos centramos en buscar la política óptima, es decir que nos de mayor recompensa(Maximización). Para lograrlo se buscan los mejores parámetros thetas, por lo que los para maximizar el alcance. mismos se van ajustando poco a poco hasta obtener los deseados. Como el sistema que queremos replicar nos gustaría al llegar a un estado terminal inicial en donde nos quedamos anteriormente, y esto es tan sencillo. Por ende se define una política que usa el valor promedio usando (softmax) ya que confiamos en el estado inicial. En la búsqueda de los parámetros theta entra en juego los gradientes con el que garantizamos que en el peor de los casos se converge a un un máximo local y máximo global en el mejor caso. Esto le permite al algoritmo comenzar aprender cuál será el máximo en lugar de calcularlo directamente. Por último se encuentra la tasa de aprendizaje que va incrementando en uno en cada iteración usando la fórmula de 1 entre k(número de iteración) más 100.

Shamuel Manuel Manrrique Aquino
NIP: 802400

Tabla 2: Resultados comparativo tras cinco pruebas con Naive Agent, Random Agent y Linear Agent.

	NaiveAgent	RandomAgent	LinearAgent
Prueba 1	Training Min 24.0	Training Min 8.0	Training Min 15.0
	Training Max 71.0	Training Max 97.0	Training Max 500.0
	Training AVG 42.004	Training AVG 22.63	Training AVG 269.914
	Test Min 38.0	Test Min 10.0	Test Min 500.0
	Test Max 57.0	Test Max 69.0	Test Max 500.0
	Test AVG 43.7	Test AVG 29.3	Test AVG 500.0
Prueba 2	Training Min 24.0	Training Min 8.0	Training Min 11.0
	Training Max 71.0	Training Max 102.0	Training Max 500.0
	Training AVG 42.111	Training AVG 22.328	Training AVG 316.041
	Test Min 32.0	Test Min 10.0	Test Min 500.0
	Test Max 53.0	Test Max 41.0	Test Max 500.0
	Test AVG 40.8	Test AVG 20.5	Test AVG 500.0
Prueba 3	Training Min 24.0	Training Min 8.0	Training Min 15.0
	Training Max 68.0	Training Max 118.0	Training Max 500.0
	Training AVG 42.804	Training AVG 22.53	Training AVG 269.914
	Test Min 31.0	Test Min 9.0	Test Min 500.0
	Test Max 57.0	Test Max 54.0	Test Max 500.0
	Test AVG 46.6	Test AVG 26.1	Test AVG 500.0
Prueba 4	Training Min 24.0	Training Min 8.0	Training Min 18.0
	Training Max 72.0	Training Max 101.0	Training Max 500.0
	Training AVG 42.227	Training AVG 21.659	Training AVG 466.7
	Test Min 32.0	Test Min 11.0	Test Min 500.0
	Test Max 56.0	Test Max 55.0	Test Max 500.0
	Test AVG 40.9	Test AVG 22.0	Test AVG 500.0
Prueba 5	Training Min 24.0	Training Min 9.0	Training Min 15.0
	Training Max 68.0	Training Max 115.0	Training Max 500.0
	Training AVG 41.925	Training AVG 22.004	Training AVG 288.14

	Test Min 25.0	Test Min 15.0	Test Min 500.0
	Test Max 57.0	Test Max 52.0	Test Max 500.0
	Test AVG 43.9	Test AVG 26.3	Test AVG 500.0

Figura 3: Resultados obtenidos con Linear Agent

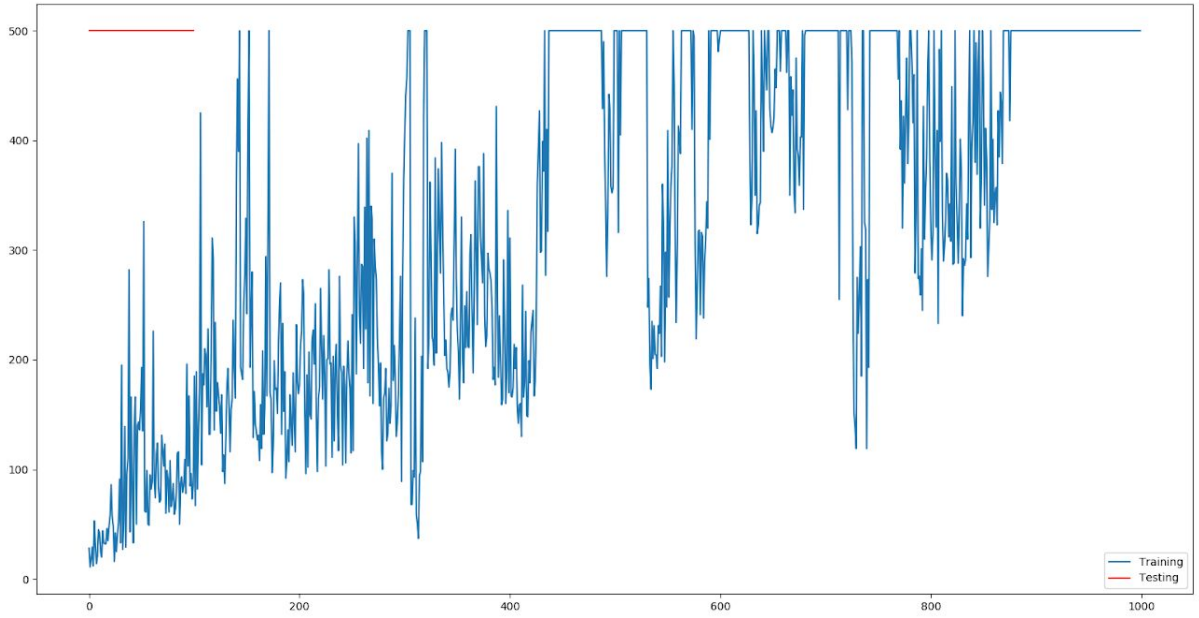
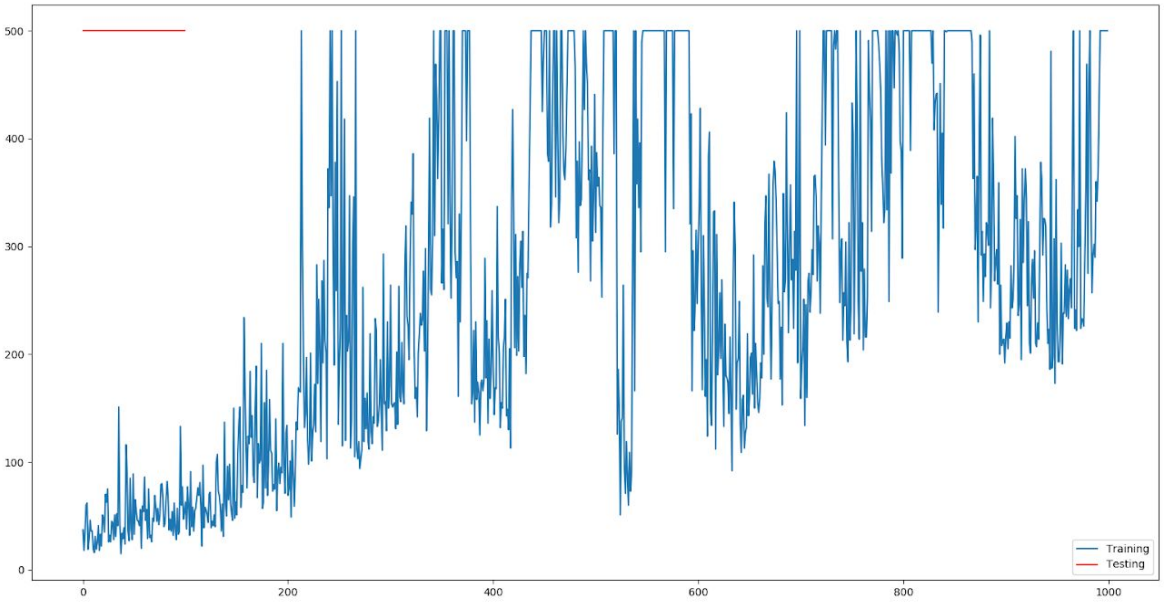


Figura 4: Resultados obtenidos con Linear Agent

Shamuel Manuel Manrrique Aquino
NIP: 802400



3. Posibles Mejoras

El algoritmo necesita un número considerable de prácticas para encontrar la política óptima por lo que esto podría tornarse lento en entornos con altas varianzas dado que estabilizar los parámetros no es trivial. Por lo que una solución sería asegurar que la desviación de la política nueva respecto a la de la anterior se mantenga relativamente pequeña. Para esto se introduce Baseline para corregir las variaciones que se generan por el gradiente.

Shamuel Manuel Manrrique Aquino
NIP: 802400