

Entrenamiento de CNN asistido por Algoritmos Evolutivos

Miguel Langone - Posgrado en Computación
mlangone13@gmail.com - 4.651.185-1

Abstract—Este documento detalla el uso de algoritmos evolutivos como alternativa para el entrenamiento de redes neuronales convolucionales. Asimismo se presenta una mutación basada en gradiente por descendiente estocástico aplicada en un problema de clasificación de imágenes.

Index Terms—Algoritmo, Gen, Genoma, Pipeline, Cruzamiento, Mutación, Red neuronal, Backpropagation, Entrenamiento.

I. INTRODUCCIÓN AL PROBLEMA

HOY en día uno de los recursos más utilizados para el procesamiento de imágenes son las redes neuronales convolucionales. Esto se debe mayormente a la gran capacidad de generalización para distintos tipos de problemas como detección de objetos, segmentación y clasificación de imágenes, entre otros. El entrenamiento de las mismas se lleva a cabo de una manera muy similar a las redes tradicionales aplicando el algoritmo de Backpropagation como método de optimización sobre una función de costo (loss function).

En este informe veremos una alternativa a dicho método de optimización la cual consiste en aplicar técnicas de algoritmos evolutivos para encontrar la mejor configuración de pesos (w_i) en un tiempo razonable y típicamente menor que el del algoritmo basado en Backpropagation. Además se propone una nueva técnica de mutación llamada Mutación SGD basada en aplicar uno o más pasos del descendiente por gradiente estocástico bajo determinadas condiciones.

A. ¿Como funciona Backpropagation?

El algoritmo *Backpropagation* se utiliza para aprender los pesos de una red multicapa de tipo *feed-forward*. Dada una cantidad arbitraria de salidas, se puede considerar la siguiente medida de error sobre el conjunto de entrenamiento:

$$E(\vec{w}) = \frac{1}{2} \sum_{d \in D} \sum_{k \in S} (t_{kd} - o_{kd})^2 \quad (1)$$

donde:

- S es el conjunto de unidades (neuronas) de salida.
- D es el conjunto de ejemplos de entrenamiento.
- t_{kd} es el valor objetivo para el ejemplo d y para la salida k .
- o_{kd} es el valor devuelto por una unidad, aplicado al ejemplo d , para la salida k .

Backpropagation busca en todo el espacio de hipótesis definido los posibles valores para los pesos en cada una de las unidades de la red. La idea es mediante el uso del algoritmo

del descendiente por gradiente estocástico modificar el vector de pesos en la dirección que produce el mayor decremento del error hasta llegar al mínimo posible.

B. ¿Qué limitaciones tiene Backpropagation?

Al utilizar el descendiente por gradiente estocástico en busca de minimizar la función de error antes descrita, el algoritmo Backpropagation tiene el problema en caer en mínimos locales y no asegurar el mínimo global de la función objetivo. Además es necesario que las funciones de activación que se utilizan en cada capa sean derivables para que el algoritmo funcione correctamente.

C. Propuesta

Dada una red para clasificar imágenes de colores de tamaño 32X32 en 10 clases distintas (CIFAR10[1]) con 4 capas convolucionales y 2 capas completamente conectadas, se busca encontrar los pesos $W(ij)$ que maximicen la accuracy para dicha tarea.

La propuesta consiste en sustituir la optimización basada en Backpropagation por un algoritmo evolutivo capaz de obtener resultados similares en un tiempo razonable con menor o igual capacidad de cómputo.

La comparación se realizará contra la misma red convolucional entrenada tradicionalmente en un ambiente de software y hardware idéntico.

II. DISEÑO DE LA SOLUCIÓN

En esta sección se detallan todos los aspectos del diseño de la solución del problema, desde las herramientas utilizadas hasta las decisiones de diseño tomadas.

A. Biblioteca utilizada

Se utilizará la biblioteca DEAP[3] desarrollada en Python para el entrenamiento del algoritmo genético. Esto se debe a que permite una integración muy fácil con Keras[4] para la utilización y evaluación de redes neuronales. Todo el pipeline se desarrollo en Python 3.6.

B. Representación de la solución

La solución del problema esta dada por un array de tamaño $P=1250858$, donde P es la cantidad de parámetros entrenables de todas las capas convolucionales y capas densas. Cada Gen es de tipo float de 32 bits.

La solución tiene la forma:

$$ind(i) = |a_{11} \ a_{12} \ a_{13} \ \dots \ a_{i1} \ a_{i2} \ a_{i3} \ \dots \ a_{ij}|$$

donde:

$$length(ind(i)) = P$$

Cada individuo se recompone en la arquitectura de red de la figura 1.

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 32, 32, 32)	896
activation_1 (Activation)	(None, 32, 32, 32)	0
conv2d_2 (Conv2D)	(None, 30, 30, 32)	9248
activation_2 (Activation)	(None, 30, 30, 32)	0
max_pooling2d_1 (MaxPooling2)	(None, 15, 15, 32)	0
dropout_1 (Dropout)	(None, 15, 15, 32)	0
conv2d_3 (Conv2D)	(None, 15, 15, 64)	18496
activation_3 (Activation)	(None, 15, 15, 64)	0
conv2d_4 (Conv2D)	(None, 13, 13, 64)	36928
activation_4 (Activation)	(None, 13, 13, 64)	0
max_pooling2d_2 (MaxPooling2)	(None, 6, 6, 64)	0
dropout_2 (Dropout)	(None, 6, 6, 64)	0
flatten_1 (Flatten)	(None, 2304)	0
dense_1 (Dense)	(None, 512)	1180160
activation_5 (Activation)	(None, 512)	0
dropout_3 (Dropout)	(None, 512)	0
dense_2 (Dense)	(None, 10)	5130
activation_6 (Activation)	(None, 10)	0
Total params: 1,250,858		
Trainable params: 1,250,858		
Non-trainable params: 0		

Fig. 1. Arquitectura de red convolucional.

C. Función de fitness

La función de fitness a maximizar es la accuracy de la red CIFAR10 utilizando los pesos del individuo i:

$$f(ind(i)) = accuracy(CIFAR10(ind(i)))$$

siendo la función de accuracy de la siguiente manera:

$$accuracy = \frac{tp + tn}{tp + fp + tn + fn}$$

siendo

- **tp**: Cantidad de verdaderos positivos
- **tn**: Cantidad de verdaderos negativos
- **fp**: Cantidad de falsos positivos
- **fn**: Cantidad de falsos negativos

D. Estrategia de inicialización de individuos

La inicialización de los pesos de la red neuronal se realiza utilizando la inicialización de Xavier[7].

La misma consiste en elegir números aleatorios entre:

$$+ - \frac{\sqrt{6}}{\sqrt{n_i + n_{i+1}}}$$

siendo n_i la cantidad de entradas de la capa y n_{i+1} la cantidad de salidas de la misma.

Luego de realizar esa selección se transforma la arquitectura de red en un vector de tamaño 1250858, siendo así en uno de los individuos del algoritmo evolutivo.

E. Estrategia Evolutiva

La técnica utilizada realizar el algoritmo evolutivo fue (μ, λ) con μ la cantidad de padres que se eligen, siendo éste un parámetro configurable (**POP**), y $\lambda = 50$ (fijo) la cantidad de hijos que se crean en cada generación de dicho algoritmo.

F. Operadores evolutivos

A continuación se detallan los operadores evolutivos utilizados para la realización del algoritmo evolutivo en su totalidad.

1) *Operadores de selección*: La estrategia de selección elegida fue selección por torneo con un tamaño 3 fijo. No se detalla más sobre este operador dado que no fue probado exhaustivamente como si lo fueron los demás operadores.

2) *Operadores de cruzamiento*: Luego de aplicar los operadores de selección se aplica un operador de cruzamiento de un punto con la probabilidad **CROSS_PROB** la cual será un parámetro a determinar en la sección III.

3) *Operadores de mutación*: Como operador de mutación se presenta un nuevo tipo de mutación específico llamado *Mutación SGD* el cual se detalla en II-G. La probabilidad de realizar dicha mutación esta dada por **MUT_PROB** la cual será otro parámetro más a determinar en la sección III.

G. Mutación SGD

Llamamos mutación SGD a la aplicación del descendiente por gradiente estocástico sobre un individuo y un subconjunto de elementos del entrenamiento (en este caso un subconjunto de imágenes). En pocas palabras, la mutación SGD es un entrenamiento parcial y acotado a un par de épocas con la finalidad de aproximar al individuo a un mínimo local.

Esta mutación tiene 3 parámetros configurables los cuales son:

- **SAMPLE_RATIO**: porcentaje del conjunto de entrenamiento elegido de manera aleatoria para realizar el entrenamiento (entre 0.0 y 1.0).
- **BATCH_SIZE**: tamaño del batch para ajustar durante el entrenamiento parcial. Se elige 64 elementos de batch.
- **N_ITER**: cantidad de épocas las cuales se aplica el algoritmo de SGD.

Como se puede ver, si se utiliza un **SAMPLE_RATIO** muy alto y un **N_ITER** también alto, la mutación de un individuo se decanta en el algoritmo de Backpropagation. Además, el uso de un **N_ITER** muy alto puede producir el fenómeno conocido como *overfitting* debido a la pérdida de generalización del algoritmo.

Dicho esto es que se elige **SAMPLE_RATIO** = 0.05 y **N_ITER** = 3.

H. Hardware utilizado

En cuanto al hardware utilizado para el entrenamiento, se utilizó el ClusterUY [2] el cual permite ejecutar distribuido utilizando varias CPUs de alta capacidad de procesamiento.

- **Procesadores** Se utilizaron 30 procesadores Xeon Gold 6138 con 3GB de memoria RAM cada uno.
- **Almacenamiento** El almacenamiento provisto por ClusterUY es de 300GB de disco de estado solido.
- **Sistema Operativo** El sistema operativo es Linux CentOS 7

III. CONFIGURACIÓN DE PARÁMETROS

En esta sección se detallan las pruebas realizadas para buscar la mejor configuración de parámetros del algoritmo evolutivo. Los parámetros que se van a probar junto con sus valores posibles son los siguientes:

Tipos de algoritmos			
# Alg	POP	MUT PROB	CROSS PROB
1	30	0.05	0.5
2	30	0.05	0.8
3	30	0.1	0.5
4	30	0.1	0.8
5	50	0.05	0.5
6	50	0.05	0.8
7	50	0.1	0.5
8	50	0.1	0.8

TABLE I

DISTINTAS CONFIGURACIONES DE ALGORITMOS SEGÚN POBLACIÓN, PROBABILIDAD DE CRUZAMIENTO Y PROBABILIDAD DE MUTACIÓN.

A. Test de normalidad sobre las muestras

En todos los casos se aplicó el test de Anderson - Darling[8] para comprobar la normalidad de los datos obtenidos y así poder compararlos. Esto se realizó con la biblioteca de Python, SciPy [6], con el paquete *stats.anderson* que provee el test directamente y retorna si la distribución es o no normal.

Las tablas 2 3 y muestran los resultados de 30 ejecuciones para cada algoritmo en 3 instancias distintas. Para acortar el tiempo insumido se realizaron 20 generaciones (pasos de evolución) de cada combinación y cada instancia.

#POP	Mut Prob	Cross Prob	Instancia 1			Instancia 2			Instancia 3		
			Max Acc	Mean Acc	STD Dev Acc	Max Acc	Mean Acc	STD Dev Acc	Max Acc	Mean Acc	STD Dev Acc
30	0.05	0.5	0.4710	0.441591	0.022300	0.46350	0.415400	0.027207	0.50425	0.416908	0.027535
		0.8	0.4794	0.431781	0.023874	0.45525	0.412758	0.025097	0.47250	0.414067	0.024606
		0.1	0.5248	0.481869	0.024796	0.50225	0.459350	0.019188	0.51425	0.460742	0.028091
50	0.05	0.5	0.5342	0.470868	0.026259	0.48125	0.444333	0.017767	0.50050	0.450358	0.019624
		0.8	0.5018	0.444756	0.023529	0.47850	0.421708	0.023854	0.47850	0.415117	0.025236
		0.1	0.4908	0.443823	0.030307	0.47575	0.412842	0.026183	0.47375	0.409283	0.025215
50	0.1	0.5	0.5108	0.478201	0.016102	0.49100	0.455683	0.016560	0.48475	0.451908	0.019722
		0.8	0.5042	0.463756	0.019866	0.48700	0.443692	0.026341	0.48700	0.446500	0.019980
		0.1	0.5042	0.463756	0.019866	0.48700	0.443692	0.026341	0.48700	0.446500	0.019980

Fig. 2. Resultados estadísticos de cada algoritmo probado en 3 instancias distintas para la función de fitness. De izquierda a derecha se tienen el máximo fitness (accuracy), fitness medio, y la desviación estándar de dicho algoritmo para esa instancia.

#POP	Mut Prob	Cross Prob	Instancia 1			Instancia 2			Instancia 3		
			Max Elapsed Time	Mean Elapsed Time	STD Dev Elapsed Time	Max Elapsed Time	Mean Elapsed Time	STD Dev Elapsed Time	Max Elapsed Time	Mean Elapsed Time	STD Dev Elapsed Time
30	0.05	0.5	585.430803	521.745615	24.437289	466.933791	422.403172	19.607599	452.778627	420.785072	13.812575
		0.8	746.316652	706.364144	24.789834	623.731419	568.553354	20.758944	613.912909	569.912134	20.594343
		0.1	768.895339	689.171115	30.563393	608.147165	557.844349	23.160279	617.549176	557.684428	27.490869
50	0.05	0.5	976.826695	903.189532	32.816633	745.428003	705.847191	23.098994	791.429846	706.900775	26.899257
		0.8	950.801160	939.820142	15.880484	459.499698	427.814481	16.078355	489.287759	432.397987	18.070713
		0.1	791.469037	734.047568	20.275110	614.341066	586.707987	18.940213	628.323254	588.960352	15.789168
50	0.1	0.5	754.845066	694.054882	38.824549	852.159874	822.846666	28.153603	641.246135	590.468143	27.769426
		0.8	966.439411	905.739424	26.349486	770.798395	725.155535	21.082239	769.059404	723.767458	23.029965
		0.1	966.439411	905.739424	26.349486	770.798395	725.155535	21.082239	769.059404	723.767458	23.029965

Fig. 3. Resultados estadísticos para cada algoritmo probado en 3 instancias distintas para el tiempo insumido de las ejecuciones (en segundos).

En las figuras 4, 5 y 6 se muestran los resultados de las medias y las desviaciones estándar para cada algoritmo en cada instancia de prueba.

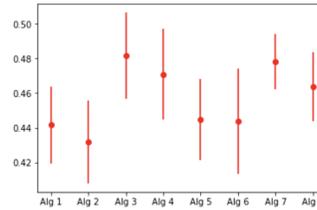


Fig. 4. Resultados de media y desviación estándar para la instancia 1 para cada algoritmo.

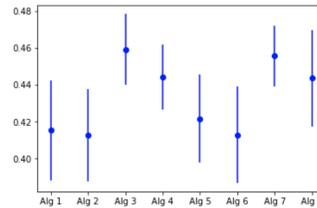


Fig. 5. Resultados de media y desviación estándar para la instancia 2 para cada algoritmo.

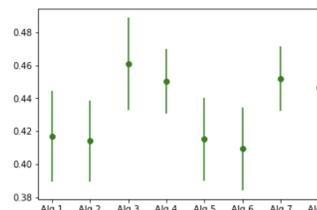


Fig. 6. Resultados de media y desviación estándar para la instancia 3 para cada algoritmo.

Luego de aplicar el test de normalidad de Anderson-Darling se comprueba que los valores de fitness siguen una distribución normal y se procede a comparar los algoritmos para determinar el mejor. Para esto se realiza un test de hipótesis de ANOVA (Analysis of Variance) para determinar si las medias de los conjuntos son similares o no.

B. Elección del mejor algoritmo

Una vez determinado que los valores de los mejores fitness siguen una distribución normal, es que podemos realizar un

test de hipótesis con la finalidad de determinar si los valores obtenidos provienen de una media similar o no.

Para ello se utiliza un test de hipótesis de una vía (1-way) ANOVA para descartar o confirmar si los valores de cada algoritmo provienen de una media similar o no. Por simplicidad se utiliza la biblioteca SciPy[6] de Python que provee el test antes mencionado [5] y realiza todos los cálculos necesarios.

Para ello se planten las siguientes hipótesis:

- Hipótesis nula: Los valores provienen de la misma media.
- Hipótesis alternativa: Los valores no provienen de la misma media.

Se utiliza $\alpha = 5\%$, buscando así tener un 95% de confianza para asegurar que la hipótesis nula no sea rechazada. Por tanto, si el p_valor para el test de ANOVA es menor que el nivel de significancia (0.05) entonces rechazamos la hipótesis nula, asegurando que las medias de los grupos difieren.

Se realizan todas las combinaciones entre los distintos algoritmos para (uno vs uno) y se obtienen los resultados de los $p_valores$ de la tabla 7.

	Alg1	Alg2	Alg3	Alg4	Alg5	Alg6	Alg7	Alg8
Alg1	1	2.24717e-08	0.111291	2.54474e-05	0.601053	1.53353e-09	0.772219	0.000183794
Alg2	2.24717e-08	1	1.3008e-10	0.112718	2.74316e-07	0.530117	2.40887e-06	0.00358646
Alg3	0.111291	1.3008e-10	1	1.77104e-07	0.0415278	4.47558e-12	0.103749	7.59027e-07
Alg4	2.54474e-05	0.112718	1.77104e-07	1	0.00018903	0.204929	0.000548638	0.249549
Alg5	0.601053	2.74316e-07	0.0415278	0.00018903	1	4.06977e-08	0.874206	0.00154833
Alg6	1.53353e-09	0.530117	4.47558e-12	0.204929	4.06977e-08	1	1.17876e-06	0.00352707
Alg7	0.772219	2.40887e-06	0.103749	0.000548638	0.874206	1.17876e-06	1	0.00406708
Alg8	0.000183794	0.00358646	7.59027e-07	0.249549	0.00154833	0.00352707	0.00406708	1

Fig. 7. Resultados del p_valor para los test de hipótesis ANOVA para todos las configuraciones de algoritmos. En rojo se encuentran aquellos algoritmos los cuales no podemos asegurar que las medias difieran.

El algoritmo con mejor fitness promedio y mejor fitness total es el número 3 el cual, según los test de hipótesis realizados, difiere su media a todos los algoritmos exceptuando el algoritmo 1 y el 7.

De todas formas, se entiende que el algoritmo 3 fue superior en generar al los demás algoritmos propuestos y proviene de una muestra con media distinta al resto.

Por esto el algoritmo seleccionado es el número 3 con la configuración:

- Cantidad de individuos: 30
- Probabilidad de mutación: 0.1
- Probabilidad de cruzamiento: 0.5

Este algoritmo será evaluado contra el entrenamiento utilizando Backpropagation en las siguientes secciones.

IV. EVALUACIÓN DE ALGORITMO FINAL

Una vez seleccionado la configuración de parámetros la cual dio mejores resultados es que se pasa a la fase final de evaluación. En esta etapa se busca comparar el típico entrenamiento basado en Backpropagation contra el algoritmo evolutivo recientemente elaborado.

La evaluación consiste en dos etapas, la primera busca evaluar cual algoritmo llega a las mejores soluciones y la segunda etapa busca determinar cual llega más rápido a su propia mejor solución.

Como se explicó anteriormente, ambos algoritmos se probaron utilizando el ClusterUY con 30 núcleos de

procesamiento en paralelo y con 3GB de memoria RAM cada uno.

En cuanto al algoritmo utilizando Backpropagation únicamente, se entrenó durante 25 épocas obteniendo el Benchmark de la figura 8 para superar con el algoritmo evolutivo.

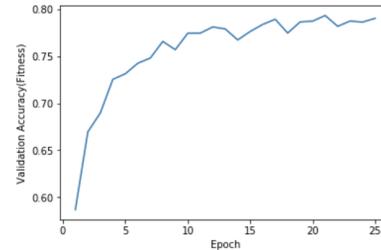


Fig. 8. Evolución de Accuracy de validación durante el entrenamiento de 25 épocas utilizando Backpropagation.

La demora promedio es de 1600 segundos (aprox. 27 minutos) para realizar 25 épocas de entrenamiento obteniendo como mejor accuracy de validación 79%.

A. Comparación de mejor solución

En la sección III-B se realizaron test de hipótesis para comparar las muestras de cada algoritmo y determinar que efectivamente tienen medias y varianzas distintas para elegir la mejor combinación de parámetros.

Gracias a eso se determinó que la mejor configuración de parámetros, la cual será tratada de ahora en adelante, es:

- **#Población:** 30
- **Probabilidad de Mutación:** 0.1
- **Probabilidad de Cruzamiento:** 0.5

Se ejecutó el algoritmo con dichos parámetros 8 veces, y un total de 200 generaciones cada uno y se obtienen los resultados de la figura 9. Cada algoritmo demora en promedio un total de 4300 segundos (aprox. 71 minutos).

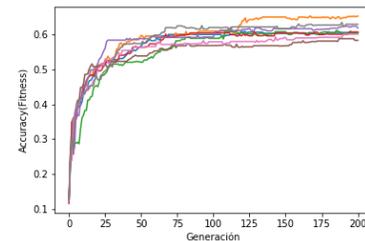


Fig. 9. Resultados para 8 ejecuciones del algoritmo evolutivo con la configuración encontrada.

Los resultados finales (tabla II) muestran que el algoritmo propuesto no fue superior al entrenamiento por Backpropagation en encontrar una mejor solución al cabo de varias generaciones.

Algoritmo	Mejor Accuracy	Accuracy Media	Tiempo insumido (seg)
Backpropagation	0.79	0.79	1600
AE(Mutación SGD)	0.65	0.62	4300

TABLE II
COMPARACIÓN DE RESULTADOS ENTRE EL ALGORITMO BACKPROPAGATION Y EL ALGORITMO EVOLUTIVO.

No obstante el sesgo introducido con la mutación SGD puede ser la clave para obtener mejores resultados, siendo que los parámetros intrínsecos a dicha mutación no fueron evaluados en este caso.

B. Velocidad de convergencia a mejor solución

Para analizar la velocidad de convergencia a la mejor solución debemos probar que tan rápido logramos llegar a la mejor solución encontrada, o en su defecto, a una solución casi similar. Para ello establecemos 4 percentiles los cuales vamos a evaluar, para encontrar en que generación logramos encontrar distintas soluciones.

- **Percentil 99:** En que iteración encontramos la solución con 99% de precisión en base a la mejor encontrada.
- **Percentil 95:** En que iteración encontramos la solución con 95% de precisión en base a la mejor encontrada.
- **Percentil 90:** En que iteración encontramos la solución con 90% de precisión en base a la mejor encontrada.
- **Percentil 70:** En que iteración encontramos la solución con 70% de precisión en base a la mejor encontrada.

En las tablas III y IV se muestran las generaciones (o épocas dependiendo el algoritmo) necesarias para lograr cada uno de los percentiles de desempeño para cada algoritmo.

Percentil	Accuracy esperada	Época	Tiempo insumido (seg)
100	0.79	25	1600
99	0.78	16	1024
95	0.75	7	448
90	0.71	3	192
70	0.55	1	64

TABLE III
ÉPOCAS NECESARIAS PARA ENCONTRAR EL ACCURACY DETERMINADO PARA CADA PERCENTIL USANDO BACKPROPAGATION.

Percentil	Accuracy esperada	Generación	Tiempo insumido (seg)
100	0.65	200	4300
99	0.64	120	2580
95	0.61	66	1419
90	0.58	26	559
70	0.45	9	193

TABLE IV
GENERACIONES NECESARIAS PARA ENCONTRAR EL ACCURACY DETERMINADO PARA CADA PERCENTIL USANDO EL AE CON MUTACIÓN SGD.

Las tablas muestran comportamientos similares de ambos algoritmos para encontrar las soluciones en el percentil 70

y 90. En los percentiles más altos el algoritmo evolutivo se estancó en un posible mínimo local y por eso es que el percentil 99 esta cerca del 60% de las generaciones realizadas (versus el percentil 99 de Backpropagation que esta en el 65% de las generaciones).

Luego de estos analisis no se puede comprobar una mejora en la velocidad de convergencia por parte del algoritmo evolutivo contra el entrenamiento por Backpropagation.

V. CONCLUSIONES Y TRABAJOS A FUTURO

En este informe se trabajó en una alternativa para entrenar redes neuronales convolucionales a partir de algoritmos evolutivos. Se planteo una alternativa de mutación basada en el descendiente por gradiente la cual agregar otro punto más de aleatoriedad a cada generación del algoritmo y realiza una perturbación (mutación) completa del individuo en un momento dado.

Algunas de las conclusiones que se pueden obtener luego de finalizado todos los experimentos es que realizar un entrenamiento completo de una red neuronal utilizando algoritmos evolutivos no queda descartado sino que se puede decir que es posible. Por otro lado, obtener un desempeño similar al encontrado por Backpropagation puede ser muy costo y requiere de una búsqueda exhaustiva por muchos parámetros configurables.

A. Trabajos a futuro

Como trabajos a futuro se plantean:

- **Búsqueda exhaustiva de parámetros:** Uno de los puntos más importantes a trabajar es realizar un análisis sobre los parámetros de la mutación SGD. La misma trae consigo 3 nuevos parámetros que pueden ser configurables para obtener un desempeño mucho mayor al logrado.
- **Expansión en una red mayor:** Se podría realizar el mismo procedimiento pero en una red mucho más grande como puede ser una VGG16 (con 16 capas convolucionales y 135 millones de pesos).

APPENDIX

A. Instructivo de uso

Para utilizar el algoritmo propuesto basta con ejecutar el comando:

```
>> python SGDEvolve.py
```

En caso de querer cambiar la configuración del algoritmo evolutivo, se debe entrar en el módulo y editar aquellos parámetros que se prefiera cambiar.

B. Tablas de configuración paramétrica

Table with 13 columns: Gen, #POP, Mut Prob, Cross Prob, Instancia 1 (Best Acc, Elapsed Time), Instancia 2 (Best Acc, Elapsed Time), Instancia 3 (Best Acc, Elapsed Time). Rows 1-30 show evaluation results for algorithm 1.

Fig. 10. Resultados de evaluación para el algoritmo 1.

Table with 13 columns: Gen, #POP, Mut Prob, Cross Prob, Instancia 1 (Best Acc, Elapsed Time), Instancia 2 (Best Acc, Elapsed Time), Instancia 3 (Best Acc, Elapsed Time). Rows 1-30 show evaluation results for algorithm 2.

Fig. 11. Resultados de evaluación para el algoritmo 2.

Table with 13 columns: Gen, #POP, Mut Prob, Cross Prob, Instancia 1 (Best Acc, Elapsed Time), Instancia 2 (Best Acc, Elapsed Time), Instancia 3 (Best Acc, Elapsed Time). Rows 1-30 show evaluation results for algorithm 3.

Fig. 12. Resultados de evaluación para el algoritmo 3.

Table with 13 columns: Gen, #POP, Mut Prob, Cross Prob, Instancia 1 (Best Acc, Elapsed Time), Instancia 2 (Best Acc, Elapsed Time), Instancia 3 (Best Acc, Elapsed Time). Rows 1-30 show evaluation results for algorithm 4.

Fig. 13. Resultados de evaluación para el algoritmo 4.

Table with 13 columns: Gen, #POP, Mut Prob, Cross Prob, Instancia 1 (Best Acc, Elapsed Time), Instancia 2 (Best Acc, Elapsed Time), Instancia 3 (Best Acc, Elapsed Time). Rows 1-30 show evaluation results for algorithm 5.

Fig. 14. Resultados de evaluación para el algoritmo 5.

Table with 13 columns: Gen, #POP, Mut Prob, Cross Prob, Instancia 1 (Best Acc, Elapsed Time), Instancia 2 (Best Acc, Elapsed Time), Instancia 3 (Best Acc, Elapsed Time). Rows 1-30 show evaluation results for algorithm 6.

Fig. 15. Resultados de evaluación para el algoritmo 6.

Gen	#POP	Mut Prob	Crossover Prob	Instancia 1		Instancia 2		Instancia 3	
				Best Acc.	Elapsed Time	Best Acc.	Elapsed Time	Best Acc.	Elapsed Time
1	50	0.1	0.8	0.417400	897.028134	0.44475	732.070208	0.42100	699.899725
2	50	0.1	0.8	0.420200	919.484040	0.42475	717.207859	0.44860	742.021883
3	50	0.1	0.8	0.524200	884.456411	0.38875	682.966888	0.44100	742.561314
4	50	0.1	0.8	0.477200	838.654787	0.44275	748.602381	0.42825	714.310848
5	50	0.1	0.8	0.424000	880.881833	0.48725	716.427820	0.48400	720.338833
6	50	0.1	0.8	0.488800	933.038826	0.47825	739.024910	0.48550	738.247910
7	50	0.1	0.8	0.550400	842.317901	0.48200	770.786936	0.46775	720.848633
8	50	0.1	0.8	0.487600	886.118387	0.41300	703.286781	0.44725	683.838887
9	50	0.1	0.8	0.481800	888.028489	0.46400	737.778157	0.44860	704.794142
10	50	0.1	0.8	0.469600	874.428789	0.44000	714.146458	0.42500	711.541383
11	50	0.1	0.8	0.464000	880.218621	0.43875	731.182700	0.48700	738.668334
12	50	0.1	0.8	0.488800	842.534339	0.38775	711.106776	0.48200	715.868847
13	50	0.1	0.8	0.471600	885.794302	0.42300	722.020200	0.44700	699.676153
14	50	0.1	0.8	0.472800	830.783020	0.43775	720.966056	0.48925	734.838869
15	50	0.1	0.8	0.427000	889.312271	0.43300	737.282086	0.47725	736.136024
16	50	0.1	0.8	0.432000	931.232377	0.48775	734.879617	0.44025	680.843882
17	50	0.1	0.8	0.481800	888.167183	0.48875	757.708257	0.42600	701.421882
18	50	0.1	0.8	0.452000	803.515882	0.48025	718.764319	0.42800	730.346038
19	50	0.1	0.8	0.470200	878.131482	0.41875	748.720880	0.43275	708.670880
20	50	0.1	0.8	0.482000	913.251238	0.44400	720.088869	0.48225	738.887000
21	50	0.1	0.8	0.412200	887.237804	0.48025	780.060221	0.41200	789.006824
22	50	0.1	0.8	0.447400	808.086160	0.41100	706.828680	0.43800	707.838612
23	50	0.1	0.8	0.487039	881.886796	0.44875	718.089735	0.44400	678.887583
24	50	0.1	0.8	0.476939	881.328283	0.43875	670.163020	0.43175	734.288819
25	50	0.1	0.8	0.477800	877.163280	0.38800	680.002780	0.48275	702.002880
26	50	0.1	0.8	0.481191	810.343872	0.48750	720.200144	0.41100	732.227887
27	50	0.1	0.8	0.433803	861.188251	0.44825	723.717309	0.48975	753.500075
28	50	0.1	0.8	0.482000	882.412741	0.43700	730.391453	0.43600	708.887022
29	50	0.1	0.8	0.455851	888.397426	0.48700	720.818182	0.47100	750.270283
30	50	0.1	0.8	0.478850	819.378508	0.48300	744.530736	0.48150	728.987842

Fig. 16. Resultados de evaluación para el algoritmo 7.

Gen	#POP	Mut Prob	Crossover Prob	Instancia 1		Instancia 2		Instancia 3	
				Best Acc.	Elapsed Time	Best Acc.	Elapsed Time	Best Acc.	Elapsed Time
1	50	0.05	0.8	0.428900	713.720044	0.38875	582.437858	0.41550	606.201102
2	50	0.05	0.8	0.410200	724.232886	0.48975	588.988807	0.42025	603.868887
3	50	0.05	0.8	0.448000	718.215912	0.42325	547.428208	0.41800	598.848840
4	50	0.05	0.8	0.422600	709.823130	0.41775	614.341856	0.41825	571.108444
5	50	0.05	0.8	0.486000	737.533391	0.38475	578.287882	0.38475	614.083778
6	50	0.05	0.8	0.469600	740.828911	0.42825	593.578833	0.42850	587.562714
7	50	0.05	0.8	0.448000	730.613795	0.43775	581.122183	0.38700	576.763711
8	50	0.05	0.8	0.438000	748.320037	0.43225	603.418861	0.38800	604.679382
9	50	0.05	0.8	0.449000	744.440273	0.38400	585.619188	0.42800	608.322054
10	50	0.05	0.8	0.423400	680.160284	0.38025	585.448885	0.42775	580.871808
11	50	0.05	0.8	0.386000	681.200848	0.37900	581.172887	0.41225	568.887476
12	50	0.05	0.8	0.473000	733.573828	0.44300	607.783882	0.44825	571.384470
13	50	0.05	0.8	0.458800	749.988078	0.44825	574.588180	0.44875	581.354883
14	50	0.05	0.8	0.462000	760.788800	0.48875	588.123800	0.38800	585.187142
15	50	0.05	0.8	0.467800	738.843885	0.41975	613.787861	0.40300	575.951601
16	50	0.05	0.8	0.418200	781.480887	0.40300	588.538842	0.38025	578.018008
17	50	0.05	0.8	0.487000	712.021800	0.43025	605.044881	0.37700	588.338838
18	50	0.05	0.8	0.426000	680.221787	0.42025	578.988842	0.38175	603.388886
19	50	0.05	0.8	0.428800	721.717731	0.44900	588.811854	0.42550	598.938804
20	50	0.05	0.8	0.386600	738.181774	0.41825	585.581041	0.40880	567.778847
21	50	0.05	0.8	0.488814	720.787899	0.42075	603.688200	0.38075	601.138809
22	50	0.05	0.8	0.478400	740.824284	0.41775	601.348288	0.38175	671.021786
23	50	0.05	0.8	0.471610	756.442818	0.38600	606.641316	0.41025	592.415787
24	50	0.05	0.8	0.480447	730.303898	0.38675	612.133671	0.41000	565.182840
25	50	0.05	0.8	0.448070	738.618888	0.42400	588.633100	0.38800	600.387148
26	50	0.05	0.8	0.384800	748.828820	0.37775	578.982884	0.41150	588.378103
27	50	0.05	0.8	0.447857	712.103164	0.40325	580.787788	0.44300	591.482881
28	50	0.05	0.8	0.438840	737.342000	0.37900	610.923375	0.38300	573.379139
29	50	0.05	0.8	0.488442	749.880108	0.42025	580.588728	0.41700	574.023887
30	50	0.05	0.8	0.413551	756.484828	0.37400	678.838848	0.40300	590.712576

Fig. 17. Resultados de evaluación para el algoritmo 8.

REFERENCES

- [1] Cifar10 net. https://keras.io/examples/cifar10_cnn/. Accessed: 2019-12-01.
- [2] Clusteruy. <https://clusteruy.com/>. Accessed: 2019-12-06.
- [3] Deap. <https://deap.readthedocs.io/en/master/>. Accessed: 2019-11-15.
- [4] Keras. <https://keras.io/>. Accessed: 2019-10-07.
- [5] One-way anova, [scipy. https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.f_oneway.html](https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.f_oneway.html). Accessed: 2019-12-01.
- [6] Scipy. <https://www.scipy.org/>. Accessed: 2019-12-01.
- [7] Weight initialization in neural networks: A journey from the basics to kaiming. <https://towardsdatascience.com/weight-initialization-in-neural-networks-a-journey-from-the-basics-to-kaiming-954fb9b47c79>. Accessed: 2019-10-06.
- [8] M. A. Stephens. Edf statistics for goodness of fit and some comparisons, journal of the american statistical association, 1974. pp. 730-737.