Trabajo Final de Curso

Antonio Manuel Padilla Urrea

Curso IA para profesionales del sector TIC

2024-08-25

Contenidos

Topic

Antecedentes

- Aplicación del Aprendizaje por Refuerzo (2019-2020)
- Juego sencillo con información oculta (Dominó)
- Primera implementación con Perceptrón Multicapa
- Buenos resultados, superando a jugadores algorítmicos simples.

Objetivos

- El objetivo de este trabajo es extender el trabajo que hice allá en 2020 usando simplemente un modelo MLP sencillo con dos capas densas.
- Tras el curso y el análisis de los Transformes, se me ocurrió interpretar una partida de dominó como una frase en un lenguaje particular -la sucesión de jugadas realizadas por cada jugador-.
- Con este *lenguaje* se pueden entrenar transformers en la esperanza de superar el juego del jugador MLP.
- En este juego, puesto que hay información oculta, es importante prestar *atención* a todas las anteriores jugadas -> Transformers.

Jugadores *unidad de medida*

Para medir los resultados del jugador basado en IA se crean dos tipos de jugadores básicos contra los que comparar la calidad del juego.

Jugador Aleatorio: Elije una jugada aleatoria entre todas las posibles

Jugador Maximizador: Elije la ficha de mayor puntuación de entre todas las posibles.

Topic

Resumen muy resumido

1. Idea Principal.
   - Averiguar qué acción es mejor a largo plazo.
2. Solución:
   - Reproducir el juego muchas veces.
   - Registrar las acciones y el resultado.
   - Entrenar modelo supervisado con datos etiquetados.
3. Datos para el entrenamiento:
   - Cada acción modifica el estado del juego.
   - Cada partida tiene un ganador y varios perdedores.
   - Dataset está formado por:
     - Estado alcanzado tras la acción del jugador.
     - Resultado de la partida para este jugador.

Dataset

- Conjunto de datos casi infinito, no habrá sobreajuste.
- Los datos se generan durante el entrenamiento.
- Un objeto de tipo *generador* genera lotes de datos.
- Durante la generacion se usa el modelo que está siendo entrenado.

Topic

Modelo de Red Neuronal

- La red neuronal estima la probabilidad de ganar la partida.
- Durante el juego se evaluan las posibles acciones.
- Se elige la acción de mayor probabilidad estimada de ganar.
- Con modelos simples ya se obtienen buenos resultados.

Modelo con 1 capa oculta

Model: "MLP"

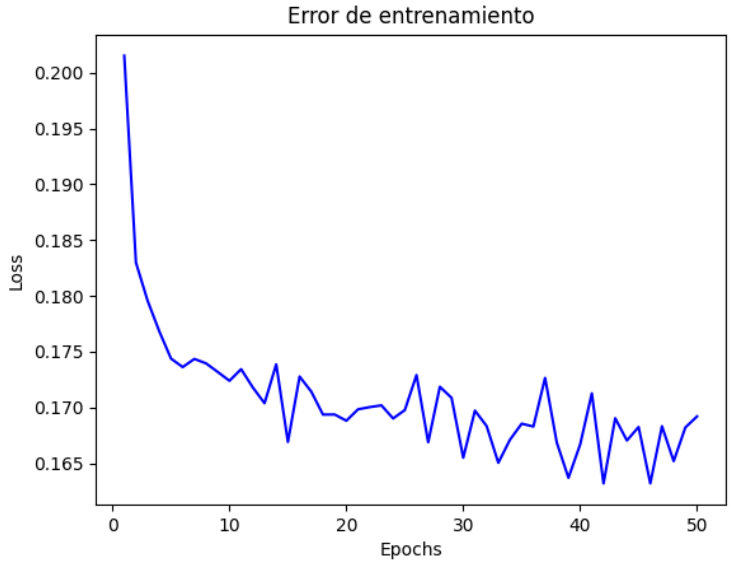
Layer (type)	Output Shape	Param #
dense (Dense)	(None, 100)	22,500
dense_1 (Dense)	(None, 60)	6,060
dense_2 (Dense)	(None, 1)	61

Total params: 28,621 (111.80 KB)
Trainable params: 28,621 (111.80 KB)
Non-trainable params: 0 (0.00 B)

Resultados del entrenamiento

Configuración

Loss	"mse"
Optimizer	"adam"
epochs	50
steps	10
batch-size	1000



Topic

Modelo de Transformer

Model: "Transformer"

Layer (type)	Output Shape	Param #
input_layer_1 (InputLayer)	(None, 50)	0
embedding (Embedding)	(None, 50, 64)	3,200
transformer_encoder (TransformerEncoder)	(None, 50, 64)	70,816
global_average_pooling1d (GlobalAveragePooling1D)	(None, 64)	0
dense_5 (Dense)	(None, 20)	1,300
dense_6 (Dense)	(None, 2)	42

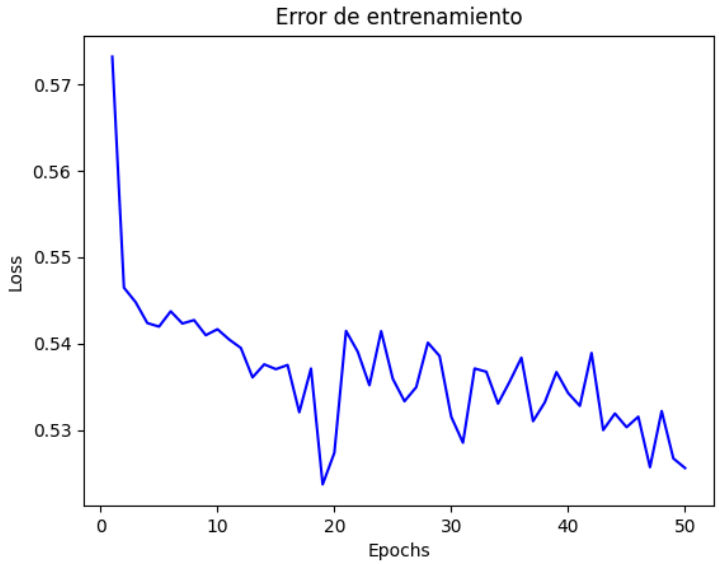
Total params: 75,358 (294.37 KB)
Trainable params: 75,358 (294.37 KB)
Non-trainable params: 0 (0.00 B)

- La salida representa las probabilidades de Ganar/Perder.
- Triple de coeficientes que el MLP

Resultados del entrenamiento

Configuracion.

Vocabulario	de 32 palabras.
Frases	50 palabras de longitud máxima
Embeddings	64 dimensiones.
Loss	Binary Cross Entropy
Optimizer	"adam"



Topic

Partidas todos contra todos

Se enfrentan los 4 tipos de jugadores entre si
Jugadas: 4000 partidas

Resultados	Ganadas	Ratio (%)
Jugador 1 : Aleatorio	936	19%
Jugador 2 : Maximizador	1277	26%
Jugador 3 : MLP	1670	34%
Jugador 4 : Transformer	1020	21%
Total	4903	100%

Hay más partidas ganadas que jugadas debido a que se producen empates y se consideran ganadores ambos.

Topic

Análisis de resultados

- En el Dominó, el azar juega un papel importante. El mejor jugador pierde muchas veces debido al reparto de fichas.
- El mejor jugador resulta ser el basado en MLP.
- El basado en transformer es solo un poco mejor que el jugador aleatório. Lo que indica que ha conseguido aprender algo durante el entrenamiento.
- Es sorprendente que con tan pocos parámetros, el MLP se haya proclamado campeón.

Trabajos futuros

- Realizar entrenamientos mas prolongados para el Transformer.
- Probar con otras funciones de perdida y/o optimizadores.
- Diseñar un modelo de transformer mas complejo.
- Diseñar una Interfaz Gráfica para jugar contra la IA.

FIN

GITHUB : `https://github.com/AntManUPCT/valgrai`
Esto es todo.