UNIVERSITAT POLITÈCNICA DE VALÈNCIA

etsinf

Escola Tècnica Superior d'Enginyeria Informàtica

Escola Tècnica Superior d'Enginyeria Informàtica
Universitat Politècnica de València

Estrategias de aprendizaje automático aplicadas a videojuegos

TRABAJO FIN DE GRADO

Grado en Ingeniería Informática

Autor: Adrián Valero Gimeno

Tutor: Vicent Botti Navarro
Javier Palanca

Curso 2018-2019

Resum

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Maecenas semper facilisis rutrum. Nam ullamcorper orci id nisl euismod facilisis. Pellentesque condimentum orci placerat, pulvinar est ut, scelerisque lacus. Pellentesque magna augue, dignissim a tempor in, tincidunt eget mauris. Nunc at sodales nulla. Maecenas congue id sem sagittis mattis. Sed a vehicula justo. Sed rhoncus rutrum ipsum a dictum. Etiam luctus sodales aliquam. Praesent nisl justo, ullamcorper et luctus ut, facilisis vel nibh. Aliquam imperdiet finibus euismod. Donec id posuere libero, eu imperdiet eros. Sed commodo egestas dolor. Ut bibendum turpis mi, vitae iaculis ligula mattis sed. Aliquam dapibus augue et felis pulvinar condimentum vel id mauris.

Paraules clau: ??????????????????

Resumen

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Maecenas semper facilisis rutrum. Nam ullamcorper orci id nisl euismod facilisis. Pellentesque condimentum orci placerat, pulvinar est ut, scelerisque lacus. Pellentesque magna augue, dignissim a tempor in, tincidunt eget mauris. Nunc at sodales nulla. Maecenas congue id sem sagittis mattis. Sed a vehicula justo. Sed rhoncus rutrum ipsum a dictum. Etiam luctus sodales aliquam. Praesent nisl justo, ullamcorper et luctus ut, facilisis vel nibh. Aliquam imperdiet finibus euismod. Donec id posuere libero, eu imperdiet eros. Sed commodo egestas dolor. Ut bibendum turpis mi, vitae iaculis ligula mattis sed. Aliquam dapibus augue et felis pulvinar condimentum vel id mauris.

Palabras clave: Inteligencia artificial, aprendizaje, automatico, videojuegos, OpenAI, hiperparámetros

Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Maecenas semper facilisis rutrum. Nam ullamcorper orci id nisl euismod facilisis. Pellentesque condimentum orci placerat, pulvinar est ut, scelerisque lacus. Pellentesque magna augue, dignissim a tempor in, tincidunt eget mauris. Nunc at sodales nulla. Maecenas congue id sem sagittis mattis. Sed a vehicula justo. Sed rhoncus rutrum ipsum a dictum. Etiam luctus sodales aliquam. Praesent nisl justo, ullamcorper et luctus ut, facilisis vel nibh. Aliquam imperdiet finibus euismod. Donec id posuere libero, eu imperdiet eros. Sed commodo egestas dolor. Ut bibendum turpis mi, vitae iaculis ligula mattis sed. Aliquam dapibus augue et felis pulvinar condimentum vel id mauris.

Key words: ?????, ????? ?????, ?????????????

Índice general

Índice de figuras

Índice de tablas

CAPÍTULO 1
Introducción

El aprendizaje automático o "Machine Learning" ha sido durante los últimos años el foco de muchísima investigación, debido a su gran potencial en la aplicación a problemas del mundo moderno. En los últimos años, proyectos como AlphaGo o la supercomputadora de google Deep Mind han conseguido hacer grandes avances en juegos de gran dificultad, siendo los agentes desarrollados capaces de competir contra las mentes más experimentadas del tradicional juego de mesa Go.

Hoy en día se están consiguiendo hacer grandes avances en el campo, debido a la investigación en nuevas técnicas como la combinación de redes neuronales tradicionales con otros métodos como el Deep Learning. Por el momento, Google está liderando este nuevo movimiento, con su supercomputadora DeepMind. Esta computadora fue capaz de desarrollar AlphaGo, una inteligencia artificial capaz de ganar a las mentes más experimentadas del juego tradicional chino Go. Éste juego, considerado uno de los más difíciles del mundo, se calcula que tiene sobre 10^{172} configuraciones posibles de las piezas sobre el tablero - un número superior al número estimado de átomos existentes en el universo -, haciéndolo extraordinariamente más complejo que juegos como el ajedrez.

1.1 Motivación

Se ha elegido un trabajo de estas características debido a una motivación personal hacia los campos de la inteligencia artificial, así como una manera de extender el conocimiento en ciertos ámbitos del mundo de la computación los cuales no se encuentran necesariamente presentes en un plan de estudios tradicional en el campo de la Ingeniería Informática. De esta forma se busca además un desarrollo personal y profesional en el campo de la computación, que permita el acceso a un futuro laboral en éste campo.

Se partía además de una necesidad personal de realizar un trabajo propio y, de cierta manera, novedoso con respecto a lo que podría conllevar la realización de un proyecto de fin de carrera típico. En definitiva, realizar un trabajo unos objetivos y una metodología a seguir bien definidos y, por encima de todo, que contribuyera al desarrollo de las competencias necesarias para la formación en el campo de la investigación y el desarrollo de sistemas de estas características.

Inicialmente, la idea para realizar el trabajo surgió meses antes de comenzar el trabajo per se, y consistía en la realización de algoritmos especializados para el aprendizaje de Cuphead, un videojuego del año 2017 con características similares a los juegos en 2D de antaño. Después de realizar una investigación exhaustiva de las maneras que tendríamos que enfocar el problema, observamos que dado que no se trata de un juego con un entorno accesible para la realización de nuestro trabajo. Esto es debido a restricciones

como el código propietario, el alto nivel de dificultad de la mayoría de sus niveles y el hecho de que requiere unas grandes cantidades de tiempo y recursos de entrenamiento, al no ser posible de ejecutar fácilmente de manera asíncrona sin un alto coste de recursos computacionales.

1.2 Objetivos

El propósito de este TFG consistirá en el estudio de las diferentes técnicas existentes de aprendizaje automático aplicadas a sencillos videojuegos en 2 dimensiones, los cuales tendrán un número limitado de acciones a realizar. Se pretende además, realizar implementaciones propias de las diferentes técnicas descritas durante el resto del documento.

Otro de los objetivos consistirá en el estudio y realización de pruebas sobre distintos entornos predefinidos de juegos arcade, haciendo uso de librerías que permiten el acceso a dicho tipo de juegos como la herramienta OpenAI desarrollada por Google, o las distintas librerías de Python relacionadas con la creación de entornos de redes neuronales y, más concretamente, la manipulación de hiperparámetros para encontrar los mejores resultados para cada uno de los entornos estudiados.

Por último, resultaría interesante aplicar estas nuevas técnicas estudiadas en la resolución de algún videojuego de mayor complejidad. Se plantea, por lo tanto, un estudio de los algoritmos más destacables estudiados en algún juego complejo como puede ser Montezuma's Revenge, un juego de la atari 2600 estrenado en el año 1983.

1.3 Metodología

Para la realización de este trabajo se pretende realizar una investigación en el campo de la inteligencia artificial, concretamente en el desarrollo de algoritmos de aprendizaje aplicados a entornos estocásticos. Para ello, se investigarán técnicas como el aprendizaje por refuerzo o Q-Learning, algoritmos evolutivos, o el uso de redes neuronales. Adicionalmente, se ha seguido un curso externo enfocado a la teoría de inteligencia artificial aplicada a distintos videojuegos, en los cuales se introducen los conceptos principales en los cuales se basa todo el campo del aprendizaje por refuerzo, desde la ecuación de Bellman hasta el diseño de redes neuronales convolucionales como método para acceder a la información de nuestros entornos.

Para complementar nuestra base de conocimientos, nos basaremos en artículos de investigación publicados por entidades como Google DeepMind, o el Massachussets Institute of Technology, en busca de ideas y nuevas técnicas para su posterior aplicación en nuestro trabajo.

1.4 Estructura de la memoria

????? ????????????? ????????????? ????????????? ????????????? ?????????????

CAPÍTULO 2

Estado del arte. La situación del aprendizaje automático en la actualidad

2.1 Introducción al Q-Learning

El aprendizaje por refuerzo o *Q-Learning* es un principio altamente utilizado en la actualidad en la investigación del campo de la Inteligencia Artificial. Este enfoque se inspira en el campo de la psicología de comportamiento y pone el énfasis en cómo un agente independiente será capaz de tomar las acciones pertinentes basándose en un sistema de recompensas a partir de las acciones que toma.

Debido a la capacidad creciente de los ordenadores de realizar computaciones cada vez más complejas en cantidades de tiempo exponencialmente menores, así como la facilidad de acceso a la información que han permitido factores como internet o, en igual medida, el abaratamiento creciente de sistemas de almacenamiento con respecto a décadas anteriores, se plantea el problema de que disponemos de un acceso a la información demasiado grande para ser tratado por métodos tradicionales. En los últimos años, se ha podido observar un crecimiento notable en el uso de sistemas de inteligencia artificial en empresas de internet hasta el punto en el que se han convertido en un requisito casi indispensable para ser capaz de ofrecer un servicio específico; por ejemplo, los sistemas de recomendación en servicios de plataformas de Streaming o la publicidad personalizada que muchas páginas de compras por internet ofrecen basándose en las compras o búsquedas realizadas anteriormente. Estos sistemas se basan en el entrenamiento de redes neuronales que permiten crear una relación entre las selecciones anteriores y productos afines.

2.1.1. Cadenas de Markov

Este enfoque se basa en la consideración del problema a resolver como un proceso de decisión de Markov con recompensa (MRP), es decir, una tupla (S, P, R, γ) donde S_n denota el estado posible **S** en el que un agente se puede encontrar, P_n una función de transición **P** y R_n es la recompensa que un agente conseguirá al realizar una determinada acción. Por último, γ denotará el factor de descuento donde $\gamma \in [0, 1]$. Éste parámetro indica al agente cuanto peso debe darle a las recompensas que recibe para aplicar estos conocimientos en momentos posteriores del aprendizaje. Se ha observado en diferentes investigaciones que introducir un factor de descuento de 1 no garantizará necesariamente la convergencia, por lo que resulta recomendable dotar al agente de cierta autonomía

para la exploración, que podría resultar en el descubrimiento de acciones más ventajosas a la hora de resolver un problema determinado. De esta manera, la noción en la que el agente percibirá la señal de recompensa $R(s,a)$ vendrá dada por:

$$R(s,a) = R_{t+1} + \gamma^2 R_{t+2} + ... + \gamma^{n-t} R_n = \sum_{k=0}^{n} \gamma^k R_{t+k+1}$$

El objetivo de nuestro agente será en todo momento conseguir maximizar la recompensa total obtenida en un episodio[1], basándose en el principio de la ecuación anterior para determinar los posibles caminos a seguir. De esta manera, la función de recompensa final podrá ser expresada como:

$$V(s) = max_a[R(s,a) + \gamma \sum s' P(s,a,s') V(s')]$$

2.1.2. Redes neuronales

El concepto de red neuronal nace de la idea de conseguir un comportamiento autónomo similar al de un ser humano. De esta manera, las redes neuronales intentan replicar el razonamiento humano creando una relación entre, en nuestro caso, lo que el agente observa y las acciones que ha de realizar.

Una red neuronal se compone esencialmente de tres componentes: Una capa de entrada, una o varias capas ocultas, y una capa de salida.

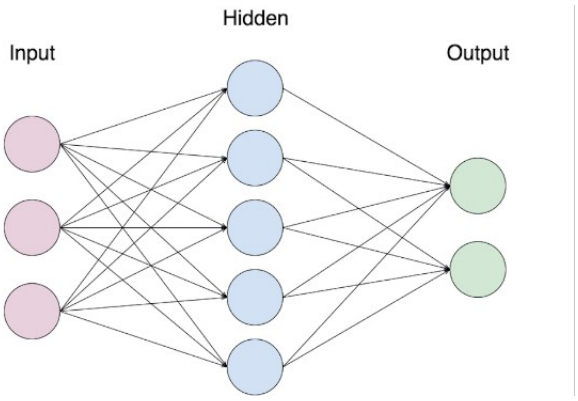

Figura 2.1: Ejemplo de una red neuronal con una capa oculta de 5 nodos (neuronas).

Este sistema toma una serie señales introducidas a través de la capa de entrada, normalizadas a unos pesos independientes entre si con valores entre 0 y 1, que a su vez emitirán estas señales a cada una de las neuronas de la primera capa oculta. Aquí, las señales tendrán que pasar por una **función de activación** que determinará si esa neurona se activa o no, para después pasar la respectiva señal a la siguiente capa. Finalmente, se obtendrá un valor de salida de las neuronas de la última capa correspondiente a 0 o 1, dependiendo de la clasificación que la red haya determinado. Este concepto se conoce como *forward-propagation*.

Dado que nuestro entorno se trata de un entrono de **aprendizaje supervisado**, se puede determinar el valor correcto que las capas exteriores deben tomar. El concepto del *backpropagation* utiliza esta ventaja que nos ofrecen los entornos deterministas para realizar así una correción de cómo el sistema maneja las señales de entrada. Así, la red neuronal es capaz de ajustar sus funciones de activación para poder realizar una mejor clasificación en pruebas posteriores.

[1] Al hablar de episodio, nos referimos a una determinada secuencia de estados dentro de un MRP hasta dar con un estado final.

2.2 Historia de la evolución tecnológica

2.3 Algoritmos existentes en la actualidad

CAPÍTULO 3
??? ???? ??????

????? ????????????? ????????????? ????????????? ????????????? ?????????????

3.1 ?? ???? ???? ? ?? ??

????? ????????????? ????????????? ????????????? ????????????? ?????????????

CAPÍTULO 4
Conclusions

????? ????????????? ????????????? ????????????? ????????????? ?????????????

Bibliografía

[1] Jennifer S. Light. When computers were women. *Technology and Culture*, 40:3:455–483, juliol, 1999.

[2] Georges Ifrah. *Historia universal de las cifras*. Espasa Calpe, S.A., Madrid, sisena edició, 2008.

[3] Comunicat de premsa del Departament de la Guerra, emés el 16 de febrer de 1946. Consultat a http://americanhistory.si.edu/comphist/pr1.pdf.

[4] Volodymyr Mnih, Adrià Puigdomènech Badia, Mehdi Mirza, Alex Graves, Tim Harley, Timothy P. Lillicrap, David Silver, Koray Kavukcuoglu. Asynchronous Methods for Deep Reinforcement Learning, febrero de 2016. *Google DeepMind, Montreal Institute for Learning Algorithms (MILA), University of Montreal*

[5] Michel Tokic Adaptive ε-greedy Exploration in Reinforcement Learning Based on Value Differences. *Institute of Applied Research, University of Applied Sciences Ravensburg-Weingarten, 88241 Weingarten, Germany*

[6] Jianxin Wu Introduction to Convolutional Neural Networks, mayo de 2017. *LAMDA Group National Key Lab for Novel Software Technology. Nanjing University, China*

[7] Tom Schaul, John Quan, Ioannis Antonoglou, David Silver Prioritized Experience Replay, febrero de 2016. *Google DeepMind*

APÉNDICE A
Configuració del sistema

????? ????????????? ????????????? ????????????? ????????????? ?????????????

A.1 Fase d'inicialització

????? ????????????? ????????????? ????????????? ????????????? ?????????????

A.2 Identificació de dispositius

????? ????????????? ????????????? ????????????? ????????????? ?????????????

APÉNDICE B
??? ????????????? ????

????? ????????????? ????????????? ????????????? ????????????? ?????????????