08MIAR-Aprendizaje por refuerzo

Sesión 10 – Algoritmos basados en modelo

Universidad Internacional de Valencia

De:
Planeta Formación y Universidades

Índice

Índice

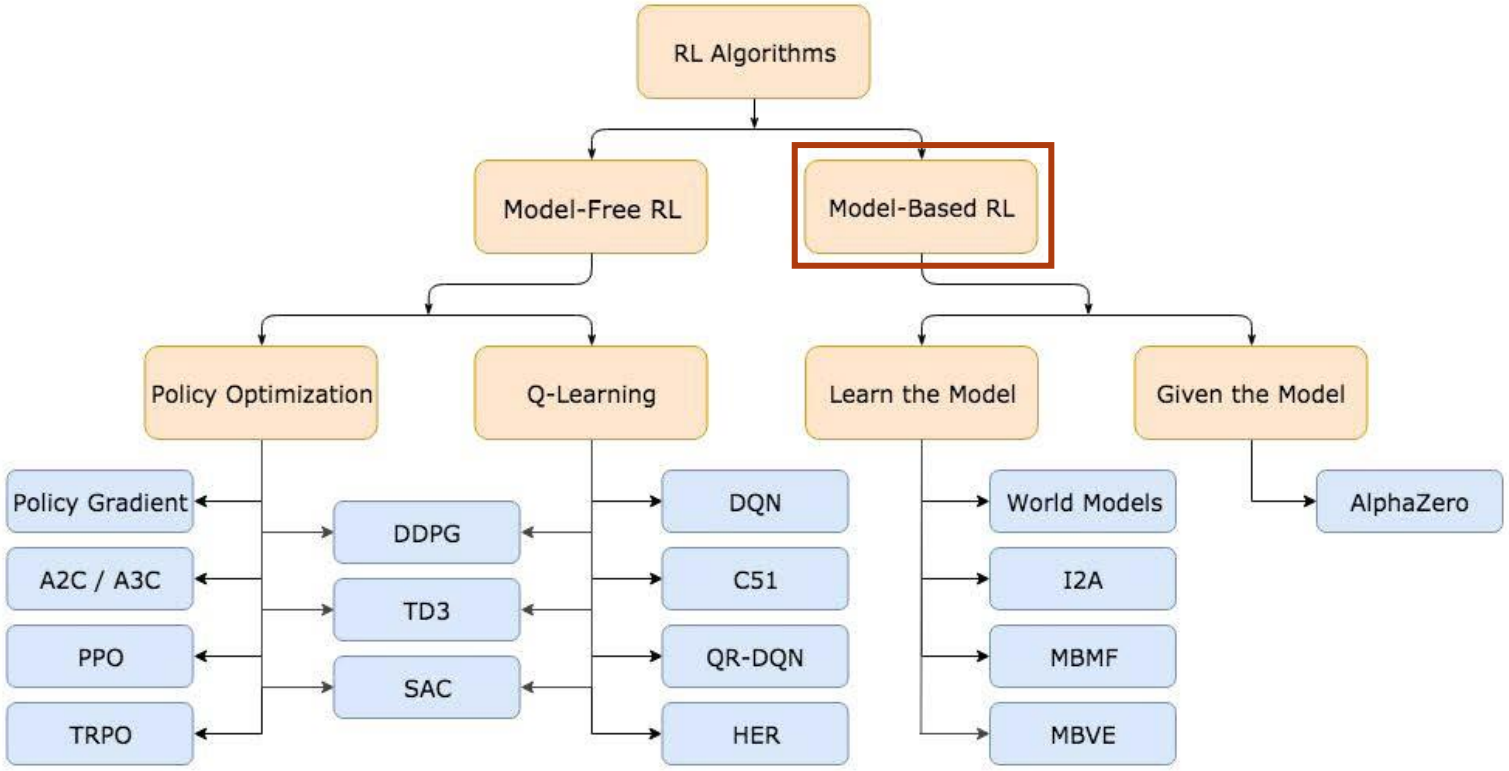
Introducción

The next big step forward in AI will be systems that actually understand their worlds. The world is only accessed through the lens of experience, so to understand the world means to be able to ***predict and control your experience****, your sense data, with some accuracy and flexibility. In other words, understanding means forming a predictive model of the world and using it to get what you want.* ***This is model-based reinforcement learning****.*

Richard Sutton

https://medium.com/the-official-integrate-ai-blog/understanding-reinforcement-learning-93d4e34e5698

Introducción



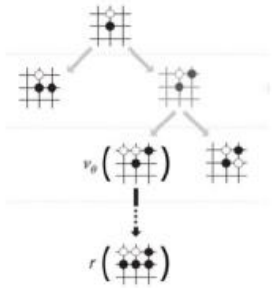
Introducción

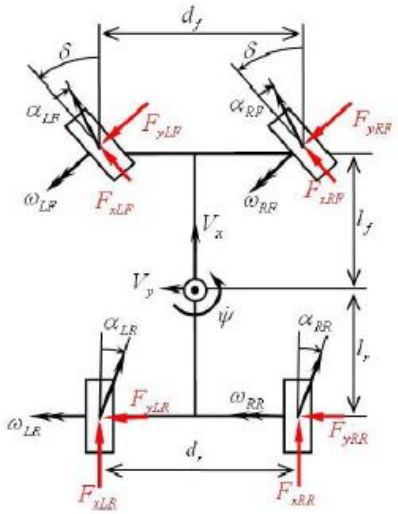
La principal diferencia entre soluciones *model free* y ***model based*** es el **conocimiento a priori de las dinámicas del entorno**.

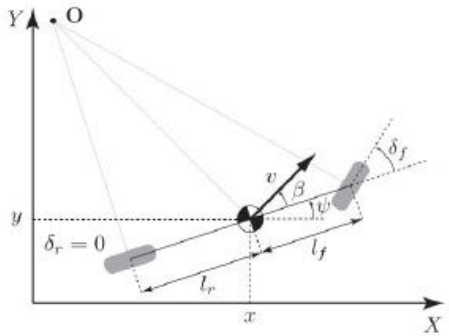
Si se conocen las dinámicas del entorno, se pueden **estimar las transiciones que el agente puede ejecutar** desde el estado actual para **valorar cuál es la mejor acción a tomar**.

Actualmente, uno de los **retos** principales aparece con entornos en los que el **espacio de acciones es muy elevado** y, por tanto, esta estimación es muy demandante en cuanto a **recursos computacionales** necesarios.

Introducción







https://jonathan-hui.medium.com/rl-model-based-reinforcement-learning-3c2b6f0aa323

Introducción

En esta asignatura nos centraremos en un tipo específico de implementación: **acompañar el aprendizaje del modelo del agente con *metaheurísticas* u otras técnicas de optimización**, que sirvan de **apoyo** al proceso de aprendizaje.

Dentro de las opciones que podemos encontrar, las más utilizadas son la **búsqueda en árbol o algoritmos basados en población (genéticos).** Todos ellos basados a su vez en simulaciones de **Montecarlo** para poder analizar los distintos caminos disponibles.

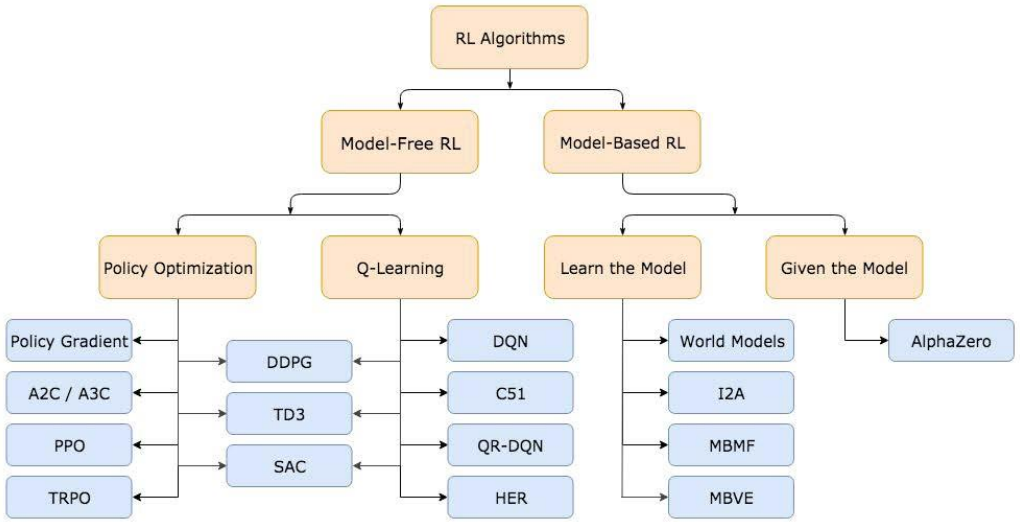
Introducción

Además, otro punto de vista de ***Model based*** es su conexión directa con **problemas de control** por computador. En este sentido, es común ver el **modelo como un controlador** y que el problema se transforme en "**minimizar el error cometido**" en vez de "maximizar la recompensa esperada".

Por último, hay un aspecto muy importante a tener en cuenta en relación con la recompensa. Hasta ahora, la recompensa nos ha venido dada como valores que el entorno nos devuelve de manera arbitraria. **En soluciones basadas en modelo es común conocer la estructura o función de recompensa**, lo que complementa al uso de técnicas de optimización a la hora de estimar las acciones.

Índice

Model based: Modelo conocido vs. Modelo desconocido



Cuando hablamos de soluciones basadas en modelo podemos encontrarnos en dos situaciones diferentes: que el **modelo sea conocido o no**.

Si el modelo es **conocido**, podremos aplicar directamente las técnicas de **metaheurísticas** vistas anteriormente.

Si el modelo es **desconocido**, tendremos que llevar a cabo un paso previo en la ejecución para aprender o **aproximar** de la mejor manera posible el modelo del **entorno**.

Model based: Modelo conocido vs. Modelo desconocido

Como hemos comentado, si el **modelo es conocido** entonces podemos utilizar ese conocimiento para enriquecer el proceso de entrenamiento. Esta es la situación típica con **metaheurísticas, planificación y otras técnicas de optimización**. Este es también el tipo de ejecución en la que se basan soluciones como **Alphago o Alphazero**.

El elemento principal a tener en cuenta en este tipo de soluciones es la **capacidad computacional** necesaria para la implementación. Los ejemplos de Alphago y Alphazero muestran esta problemática, ya que aunque podamos estimar las transiciones del entorno de una manera precisa, es imposible poder llevar todas las estimaciones posibles en cada momento de la ejecución.

Model based: Modelo conocido vs. Modelo desconocido

Si el **modelo es desconocido**, tendremos que llevar a cabo una **fase previa de aprendizaje o aproximación**. Lo común es utilizar un **enfoque de aprendizaje supervisado/no supervisado** para crear el modelo del entorno **utilizando datos recolectados**. Una vez tengamos esta aproximación, podemos utilizarla en el algoritmo de aprendizaje por refuerzo que deseemos.

A diferencia del uso de metaheurísticas, este tipo de solución Model-based tiene su base en **algoritmos/procesos basados en gradientes**, debido a la necesidad de optimizar la representación que se va obteniendo del entorno.

Índice

Model learning: Aspectos principales

Model-based Reinforcement Learning: A Survey.

Thomas M. Moerland[1,2], Joost Broekens[2], and Catholijn M. Jonker[1,2]
[1] Interactive Intelligence, TU Delft, The Netherlands
[2] LIACS, Leiden University, The Netherlands

https://arxiv.org/pdf/2006.16712.pdf

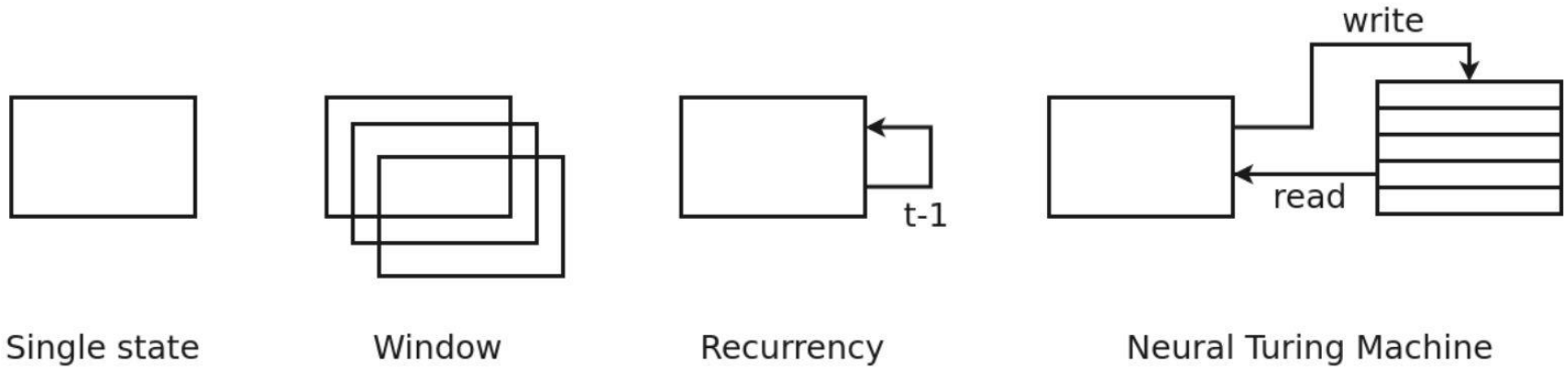
Model learning: Aspectos principales

El primer aspecto a analizar es el tipo de **modelo que se va a estimar a partir del entorno**. Podemos encontrar tres tipos principales:

1) Modelo forward $(s_t, a_t) \rightarrow s_{t+1}$
2) Modelo backward $s_{t+1} \rightarrow (s_t, a_t)$
3) Modelo inverso $(s_t, s_{t+1}) \rightarrow a_t$

Model learning: Aspectos principales

Las **observaciones parciales** se refieren a cuando la **información** disponible en una observación **no es completa**. Es una situación muy típica que puede ocurrir por la propia naturaleza del problema. Algunas formas de suavizar esta situación son:



Model learning: Aspectos principales

Uno de los **puntos más críticos** cuando se aproxima el modelo de un entorno es la **abstracción de los estados**. Debido a la alta carga de información de este tipo de problemas, **el modelo del entorno siempre va a ser una representación del mismo**.

Normalmente, esta representación va asociada a algún método o proceso de abstracción, desde una reducción de dimensionalidades hasta el uso de alguna **arquitectura de Deep Learning**.

Model learning: Aspectos principales

Otros aspectos analizados en el artículo de referencia son la **incertidumbre, la aleatoriedad o la estacionalidad de los datos**.

Incertidumbre y aleatoriedad son conceptos que están muy relacionados entre sí, ya que la incertidumbre trata la falta de información en los datos para aproximar un modelo preciso, mientras que la aleatoriedad trata la propia naturaleza estocástica del proceso que se está modelando. La aleatoriedad siempre va a estar presente mientras que la incertidumbre se podría disminuir teniendo más datos disponibles.

La estacionalidad se refiere a si cambian las funciones de transición o recompensa a lo largo del tiempo. Estos cambios, si no son percibidos por el agente, pueden conllevar a que la solución se desgaste con el paso del tiempo.

Índice

Planning & Learning

Model-based Reinforcement Learning: A Survey.

Thomas M. Moerland[1,2], Joost Broekens[2], and Catholijn M. Jonker[1,2]
[1] Interactive Intelligence, TU Delft, The Netherlands
[2] LIACS, Leiden University, The Netherlands

https://arxiv.org/pdf/2006.16712.pdf

Planning & Learning

A la hora de aplicar **planificación** junto a nuestra solución de Aprendizaje por refuerzo se nos presentan una serie de preguntas:

1) ¿En qué estado comenzamos la planificación?

2) ¿Al ejecutar simulaciones, cuánta performance reservamos para la planificación?

3) ¿Cómo llevamos a cabo la planificación?

4) ¿Cómo relacionamos la planificación con el proceso de aprendizaje y la toma de acciones?

Planning & Learning

A la hora de aplicar planificación junto a nuestra solución de Aprendizaje por refuerzo se nos presentan una serie de preguntas:

1) ¿En qué estado comenzamos la planificación?

2) ¿Al ejecutar simulaciones, cuánta performance reservamos para la planificación?

3) ¿Cómo llevamos a cabo la planificación?

4) ¿Cómo relacionamos la planificación con el proceso de aprendizaje y la toma de acciones?

- **Estado aleatorio**
- **Estado visitado**
- **Estado prioritario**
- **Estado actual**

Planning & Learning

A la hora de aplicar planificación junto a nuestra solución de Aprendizaje por refuerzo se nos presentan una serie de preguntas:

1) ¿En qué estado comenzamos la planificación?

2) ¿Al ejecutar simulaciones, cuánta performance reservamos para la planificación?

3) ¿Cómo llevamos a cabo la planificación?

4) ¿Cómo relacionamos la planificación con el proceso de aprendizaje y la toma de acciones?

- **Después de cuántos steps comenzamos?**
- **Cuánto esfuerzo ejecutamos por iteración?**

(Alphago Zero, 1MCTS, 1600 trazas de profundidad 200)

Planning & Learning

A la hora de aplicar planificación junto a nuestra solución de Aprendizaje por refuerzo se nos presentan una serie de preguntas:

1) ¿En qué estado comenzamos la planificación?

2) ¿Al ejecutar simulaciones, cuánta performance reservamos para la planificación?

3) ¿Cómo llevamos a cabo la planificación?

4) ¿Cómo relacionamos la planificación con el proceso de aprendizaje y la toma de acciones?

- **De tipo discreto (árbol, tabla) o diferencial (modelos de transición y recompensa diferenciables)**
- **La dirección puede ser forward o backward**
- **Tenemos que definir los niveles de profundidad y anchura en la simulación**

Planning & Learning

A la hora de aplicar planificación junto a nuestra solución de Aprendizaje por refuerzo se nos presentan una serie de preguntas:

1) ¿En qué estado comenzamos la planificación?

2) ¿Al ejecutar simulaciones, cuánta performance reservamos para la planificación?

3) ¿Cómo llevamos a cabo la planificación?

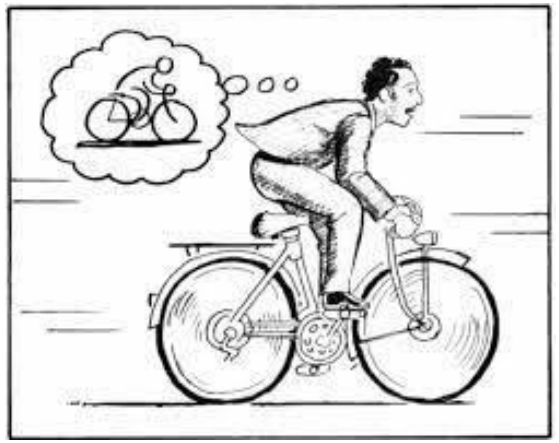
4) ¿Cómo relacionamos la planificación con el proceso de aprendizaje y la toma de acciones?

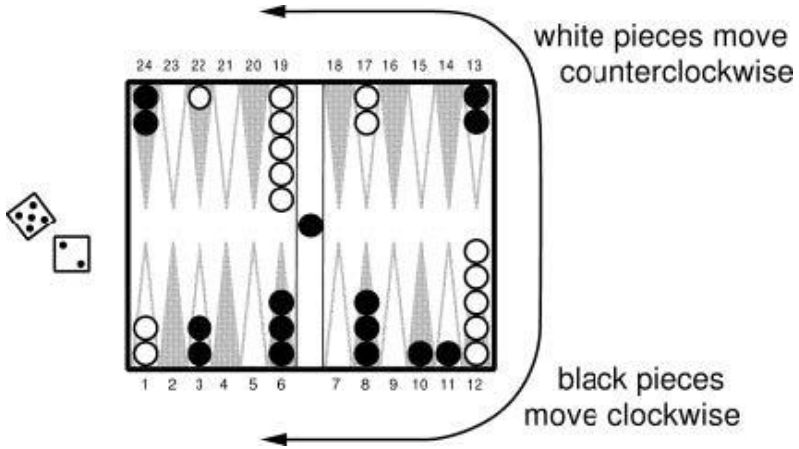
- **Usando la policy y el value para llevar a cabo la planificación.**
- **Usando el resultado de la planificación para actualizar la policy y el value.**
- **Usando directamente la planificación para seleccionar una acción.**

Índice

Ejemplos de soluciones Model based

AlphaGo



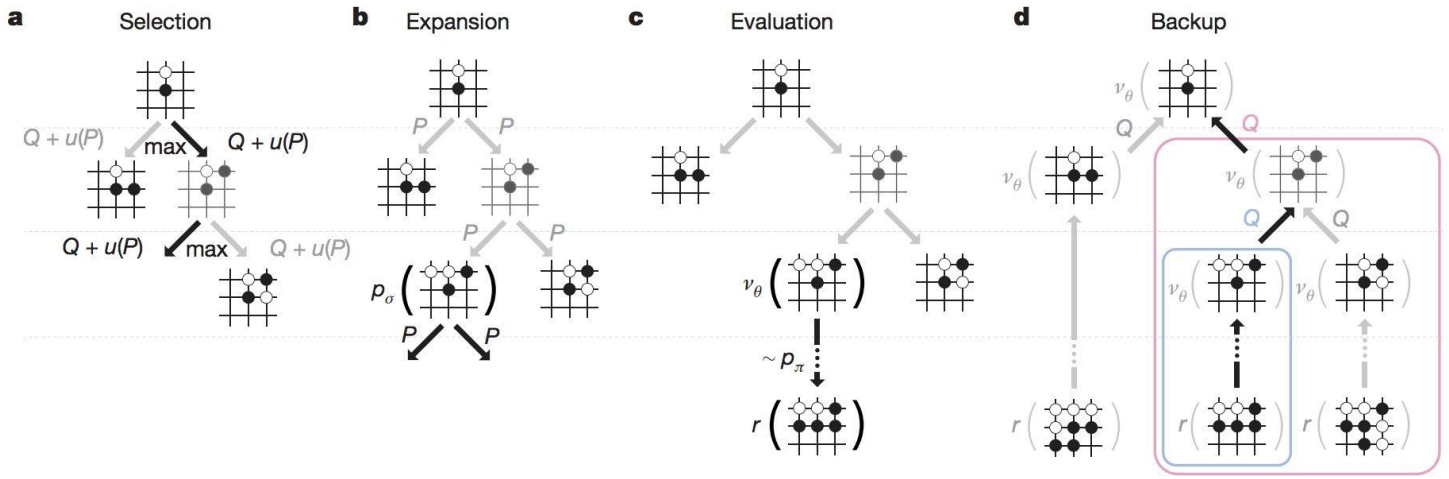


https://es.m.wikipedia.org/wiki/Archivo:Alphago_logo_Reversed.svg
http://incompleteideas.net/book/ebook/node108.html
https://arxiv.org/pdf/1803.10122.pdf

Ejemplos de soluciones Model based

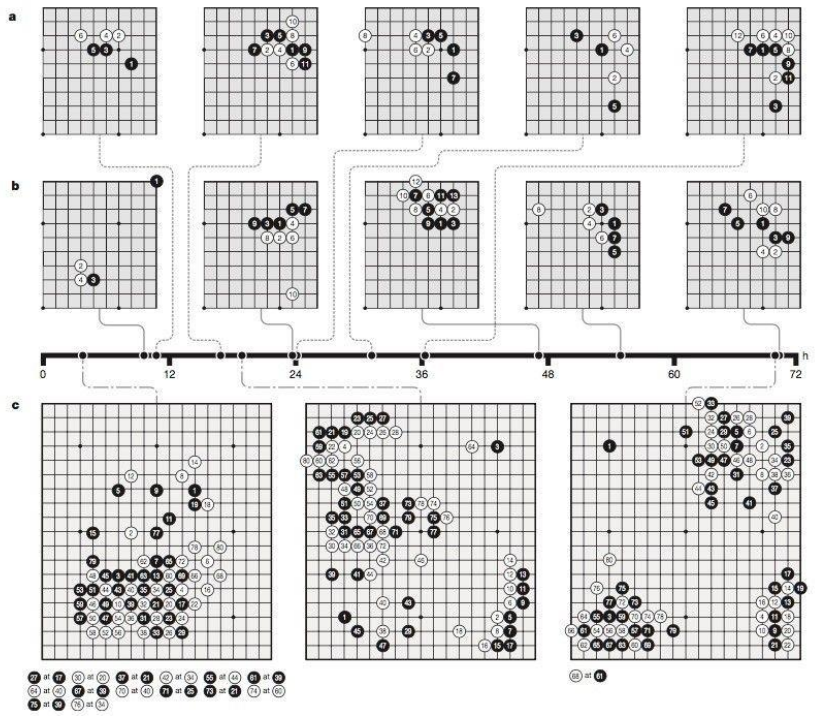
AlphaGo

- Solución entrenada con **aprendizaje supervisado y aprendizaje por refuerzo**
- Utiliza **MCTS** para la **estimación de qué acción escoger** en cada estado



https://es.m.wikipedia.org/wiki/Archivo:Alphago_logo_Reversed.svg
https://deepmind.com/research/case-studies/alphago-the-story-so-far

Ejemplos de soluciones Model based



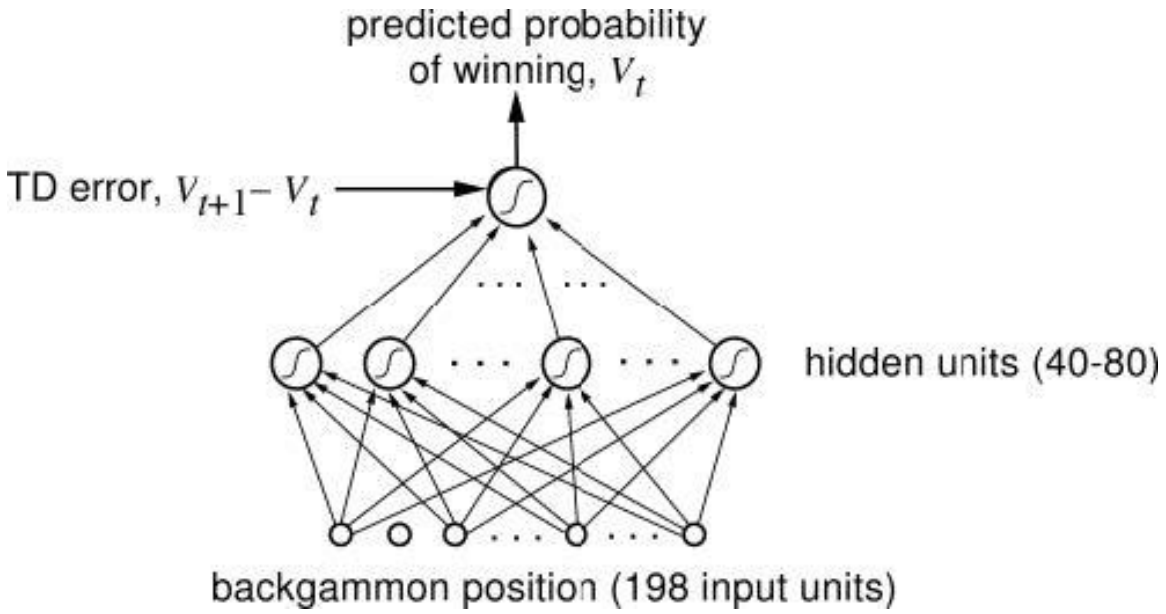
En su siguiente versión, Alphago Zero, el entrenamiento de la solución se produce mediante ***self-play***, sin incluir la primera fase de aprendizaje supervisado.

Como versión siguiente, Deepmind desarrolla AlphaZero, para generalizar a los juegos del ajedrez y Shogi además del Go.

Como curiosidad, se estima que **Alphago Zero costó unos 35M$ en recursos computacionales***

https://www.xataka.com/robotica-e-ia/la-nueva-alphago-esta-un-paso-mas-cerca-de-la-singularidad-aprende-de-si-misma-y-deja-en-ridiculo-a-la-anterior
*https://www.yuzeh.com/data/agz-cost.html

Ejemplos de soluciones Model based



- **TD-Gammon**, desarrollado por Gerry Tesauro (Tesauro, 1992, 1994, 1995).
- A diferencia de otras soluciones de la época, **utiliza como datos de entrada los datos en crudo del tablero**.
- Otro elemento diferencial es la **función de evaluación** utilizada, **basada en diferencias temporales** para optimizar la red neuronal con la que se aproxima la probabilidad de ganar.

https://en.wikipedia.org/wiki/TD-Gammon
http://incompleteideas.net/book/ebook/node108.html

Ejemplos de soluciones Model based

https://worldmodels.github.io/

Índice

Conclusiones

1) La principal diferencia entre model-free y model-based es que con **model-based conocemos las dinámicas del entorno**, por ejemplo en cuanto a transiciones y funciones de recompensa.

2) Dentro de model-based, podemos encontrarnos **con dos situaciones diferentes: que el modelo del entorno sea conocido o que sea desconocido** y, por tanto, tengamos que aproximarlo.

3) Dentro de las metaheurísticas usadas en estas soluciones, las implementaciones más comunes combinan técnicas de **planificación junto con simulaciones de Montecarlo**.

Índice

Bibliografía recomendada

- “Model-Based Reinforcement Learning: A survey”, Moerland, T. et al
https://arxiv.org/pdf/2006.16712.pdf

- “Alphago”, Google Deepmind
https://deepmind.com/research/case-studies/alphago-the-story-so-far

- “World Models”, David Ha, Jurgen Schmidhuber
https://arxiv.org/abs/1803.10122
https://worldmodels.github.io/

Universidad
Internacional
de Valencia

universidadviu.com

De:
Planeta Formación y Universidades