Etude de l’article “Mastering the Game of Go without Human Knowledge”

Application de l’algorithme AlphaZero aux échecs

Charly LAMOTHE
Encadré par François-Xavier DUPÉ

Sommaire

- Introduction
- AlphaZero
- Jeu d'échecs
- Implémentation
- Conclusion

Humain vs Machine (en 2013)

PUISSANCE 4
1995

DAMES*
2007

OTHELLO
1995,1997

SCRABBLE

ECHECS
1996 : Première victoire ordinateur
2005 : Dernière victoire humaine

BRIDGE

Résolus
Ordi >> Humain

Ordi > Humain

Humain > Ordi

source : slide cours MCTS de Liva Ralaivola

Humain vs Machine (en 2018)

PUISSANCE 4
1995

DAMES*
2007

OTHELLO
1995,1997

SCRABBLE

ECHECS
1996 : Première victoire ordinateur
2005 : Dernière victoire humaine

BRIDGE

Résolus
Ordi>>Humain

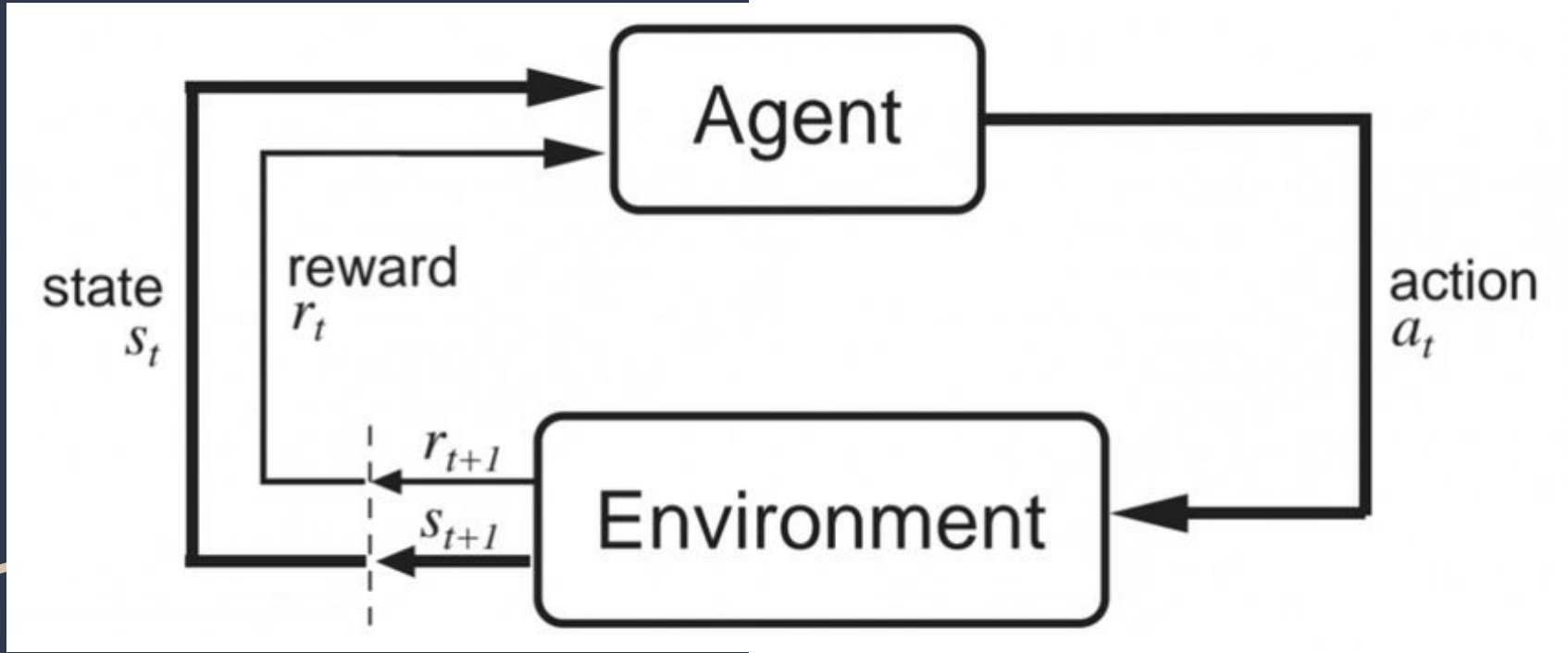
Ordi>>Humain

Ordi>Humain

Humain>Ordi

source : slide cours MCTS de Liva Ralaivola

L’Apprentissage par renforcement

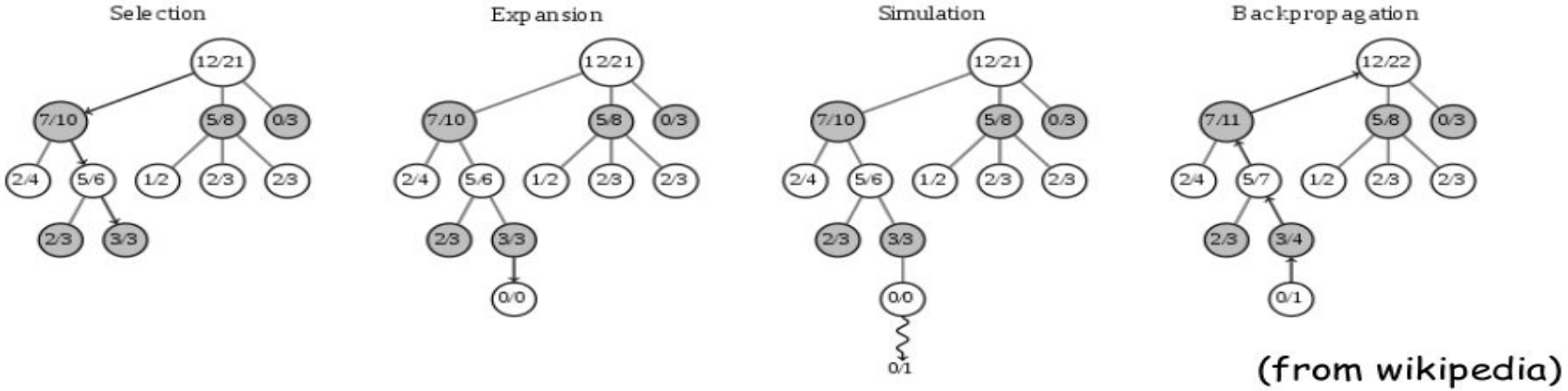


Algorithmes développés par Google DeepMind

AlphaGoFan < AlphaGoSedol < AlphaGoZero < AlphaZero

2015 **2016** **oct. 2017** **déc. 2017**

Monte Carlo Tree Search (MCTS)



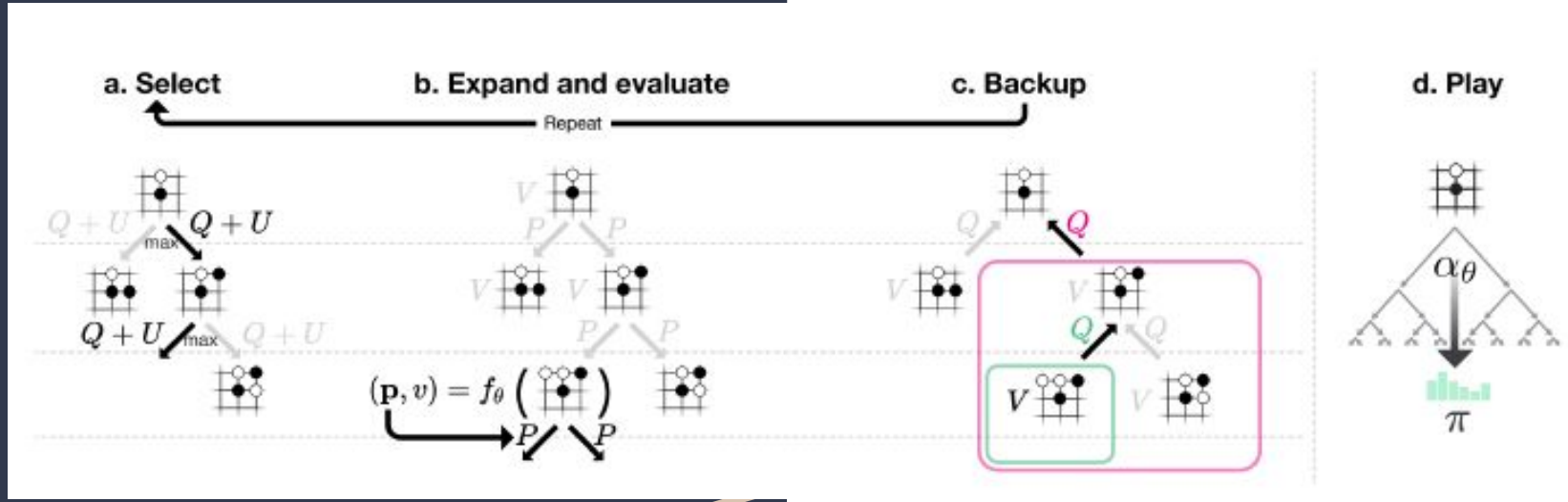
(from wikipedia)

Version « deep » des bandits et du dilemme exploration / exploitation

Exploitation vs. Exploration (credit: http://www.plexure.com/plexure-blog/2016/9/21/multiworld-testing)

source : slide cours MCTS de Liva Ralaivola

Monte Carlo Tree Search (MCTS)



$z : [-1, 0, +1]$

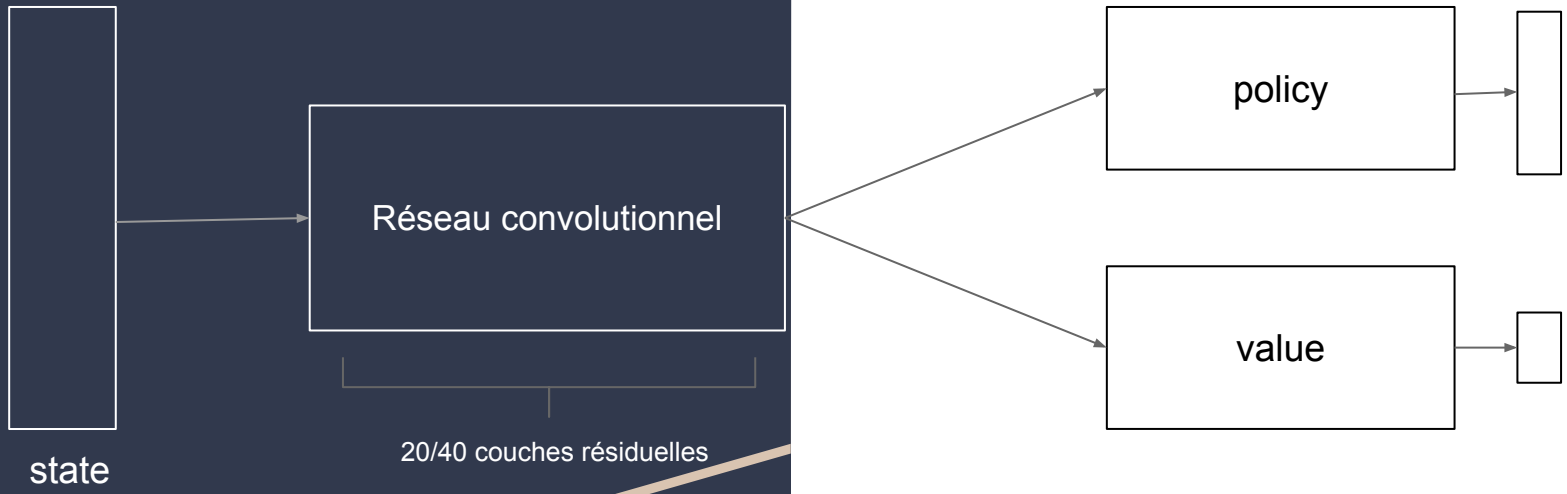
$Q(s, a)$: valeur moyenne du prochaine coût

$N(s, a)$: nombre de coût faisable

$$U(s, a) = C_{puct} P(s, a) \frac{\sqrt{\sum_b N(s,b)}}{1+N(s,a)}$$

$\pi : p$ optimal

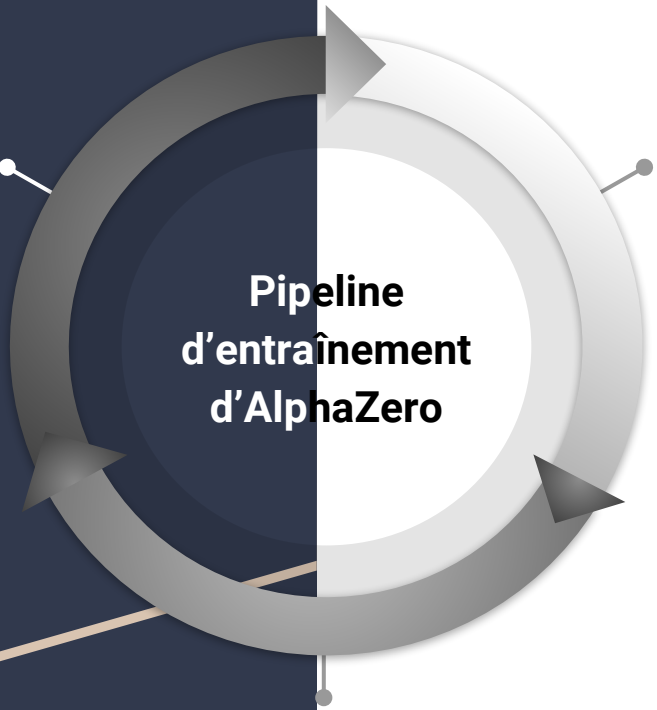
Le réseau de neurones profond (DNN)



source : AlphaGoZero Cheatcheet de applied-data.science

Evaluator

Sélectionne le meilleur réseau



Self Play

Génère le dataset

Optimisation

Entraîne le réseau en utilisant le dataset

AlphaGo vs Alpha(Go)Zero

- **Resources :** Moins de ressources hardware nécessaires
- **Généralisation :** Aucune connaissance expert, hormis les règles du jeu
- **MCTS :** 800 itérations au lieu de 1600
- **Convergence :** Learning rate évolutif
- **Modèle :** Un seul modèle continuellement entraîné au lieu de un par MCTS

AlphaZero > AlphaGo & cie

Chronologie des programmes d'échecs

1956, invention de l'algorithme *alpha-bêta* par *John McCarthy*.

1962, le premier programme avec un jeu crédible est publié au *MIT*.

1977, *Chess* devient le premier ordinateur à remporter un tournoi d'échecs majeur.

2017, *AlphaZero* de *DeepMind* bat *Stockfish*, l'un des meilleurs programmes d'échecs existants.

1997, *Deep Blue* bat *Garry Kasparov* (champion du monde de l'époque).

source : Wikipédia

Représentation des entrées/sorties

Go		Chess		Shogi	
Feature	Planes	Feature	Planes	Feature	Planes
P1 stone	1	P1 piece	6	P1 piece	14
P2 stone	1	P2 piece	6	P2 piece	14
		Repetitions	2	Repetitions	3
				P1 prisoner count	7
				P2 prisoner count	7
Colour	1	Colour	1	Colour	1
		Total move count	1	Total move count	1
		P1 castling	2		
		P2 castling	2		
		No-progress count	1		
Total	17	Total	119	Total	362

Chess		Shogi	
Feature	Planes	Feature	Planes
Queen moves	56	Queen moves	64
Knight moves	8	Knight moves	2
Underpromotions	9	Promoting queen moves	64
		Promoting knight moves	2
		Drop	7
Total	73	Total	139

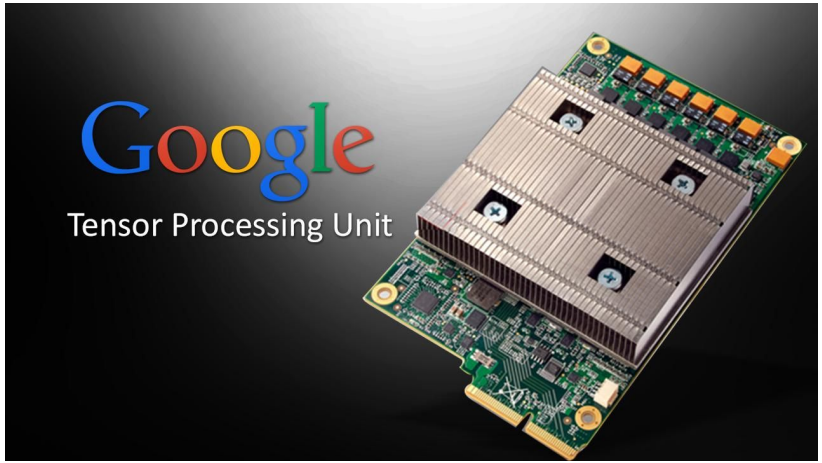
Implémentation

- Python
- Keras
- Utilisation de python-chess (https://pypi.org/project/python-chess/)
- Peu de simulations
- Pas de concurrence

URL repository
https://github.com/swasun/Yet-Another-AlphaZero

Conclusion : nuance des résultats

- Entraînement : 5 000 TPUs
- Evaluation : 4 TPUs
- 44 M de parties



Merci pour votre attention !

Bibliographie

[1] Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., van den Driessche, G., Graepel, T. & Hassabis, D. (2017). Mastering the game of Go without human knowledge. Nature, 550, 354--.

[2] Silver, David, Huang, Aja, Maddison, Chris J., Guez, Arthur, Sifre, Laurent, van den Driessche, George, Schrittwieser, Julian, Antonoglou, Ioannis, Panneershelvam, Veda, Lanctot, Marc, Dieleman, Sander, Grewe, Dominik, Nham, John, Kalchbrenner, Nal, Sutskever, Ilya, Lillicrap, Timothy, Leach, Madeleine, Kavukcuoglu, Koray, Graepel, Thore and Hassabis, Demis. "Mastering the Game of Go with Deep Neural Networks and Tree Search." Nature 529 , no. 7587 (2016): 484--489.

[3] Silver, David, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dharshan Kumaran, Thore Graepel, Timothy P. Lillicrap, Karen Simonyan and Demis Hassabis. “Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm.” CoRR abs/1712.01815 (2017): n. pag.

[4] C. B. Browne et al., "A Survey of Monte Carlo Tree Search Methods," in IEEE Transactions on Computational Intelligence and AI in Games, vol. 4, no. 1, pp. 1-43, March 2012. doi: 10.1109/TCIAIG.2012.2186810

[5] A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play, D Silver, T Hubert, J Schrittwieser, I Antonoglou, M Lai, A Guez, M Lanctot, L Sifre, D Kumaran, T Graepel, T Lillicrap, K Simonyan, D Hassabis, December 20181