**DESCENT Wins Five Gold Medals at the Computer Olympiad**

Quentin Cohen-Solal a, Tristan Cazenave a

a LAMSADE, University Paris-Dauphine, PSL, CNRS, France
E-mail: quentin.cohen-solal@dauphine.psl.eu

Descent (Cohen-Solal, 2020; Cohen-Solal and Cazenave, 2021) is a zero knowledge Deep Reinforcement Learning algorithm that has learned to play many games. It won five gold medals at the 2020 Computer Olympiad.

Unlike AlphaZero-like algorithms (Silver et al., 2018), the Descent framework uses a variant of Unbounded Minimax (Korf and Chickering, 1996), instead of Monte Carlo Tree Search, to construct the partial game tree used to determine the best action to play and to collect data for learning. During training, at each move, the best sequences of moves are iteratively extended until terminal states. During evaluations, the safest action is chosen (after that the best sequences of moves are iteratively extended each until a leaf state is reached). Moreover, it also does not use a policy network, only a value network. The actions therefore do not need to be encoded. Unlike the AlphaZero paradigm, with Descent all data generated during the searches to determine the best actions to play is used for learning. As a result, much more data is generated per game, and thus the training is done more quickly and does not require a (massive) parallelization to give good results (contrary to AlphaZero). It can use end-of-game heuristic evaluation to improve its level of play faster, such as game score or game length (in order to win quickly and lose slowly).

Five gold medals were won by our programs based on these algorithms for the following games: Othello 10x10, Breakthrough, Surakarta, Amazons, and Clobber. A silver medal was won by our programs at Othello 8x8.

The other competitors for each game were:

- **Breakthrough**: DaSoJai (author: Wei-Lin Wu and Shun-Shii Lin) and Polygames (Facebook NDHU). Polygames (Cazenave et al., 2020) is a reimplementation of AlphaZero.
- **Amazons**: 8QP (Johan de Koning) and SherlockGo (Liang Shuang, Liang Tailin, Wang Jilong, Li Xiaorui, and Zhou Ke).
- **Othello 10x10**: Polygames (Facebook NDHU) and Persona (Surag Nair, Nai-Yuan Chang, Shun-Shii Lin).
- **Othello 8x8**: Polygames (Facebook NDHU) and Maverick (Yen-Chi Chen and Shun-Shii Lin).
- **Clobber**: Pan.exe (Johan de Koning), Klopper (Johannes Schwagereit), and Calpurnia (Christian Jans). Calpurnia uses an AlphaZero-like approach and an endgame solver based on the Combinatorial Game Theory.
- **Surakarta**: CZF_Surakarta (Liang-Fu Liu), FuChou (Jia-Fong Yeh, Yen-Chi Chen, Shun-Shii Lin), and VSSurakarta (Zhang Yunpeng, Li Wei, Zhang Yuxuan, Zhang Pei, and Zhou Ke). CZF_Surakarta is trained based on AlphaZero.

Some details of the matches performed by the programs created by the Descent framework are described in Table 1.
The Descent framework could struggle with connection and alignment games. It notably lost against Polygames at Hex and Havannah. However, Polygames used much more computing power for their training (100 GPU against 1 GPU) and used very deep networks, so it is possible that this is the cause of this difference.

Descent thus constitutes an alternative to AlphaZero, in particular under resource constraints.

REFERENCES


### Table 1

Details of matches played during the first phase et second phase (playoffs) of the 2020 Computer Olympiad on Ludii: Ludii game ID, Ludii version, game name, first player program, second player program, and the winner (in bold: program based on the Descent framework).