

# Project Proposal: Hanabi

Xiuhan Wang, Jing Wei

November 5, 2020

## 1 Problem Definition

*Hanabi* is a game for two to five players. There are 50 cards in the deck. Each card has a rank (from 1 to 5) and a color (red, green, blue, yellow and white). For each color, there are three 1's, two 2's, two 3's, two 4's and one 5. Each player holds four cards in their hands if there are at least four players or five cards otherwise. Each player can see all others' cards but can't see the cards of himself. The players are fully cooperative: they want to build five stacks, one for each color, beginning with a 1 and ending with a 5. The goal is to maximize the score.

The players take turns to play. In one's turn, he must take exactly one of the following actions:

- **Hint.** The player can give a hint to another player. The hint consists of choosing a color or rank, and telling the other player "these cards are all the cards of that color (or number) in your hand". Only ranks and colors that are present in the other player's hand can be chosen. To limit the number of hints, each player has initially eight information tokens and a token is consumed when giving a hint.
- **Discard.** When the player has less than eight information tokens, he can discard a card from his hand, draw a new card from the deck and recover an information token. The discarded card is visible to all players and the newly drawn card is visible only to all other players.
- **Play.** The player can choose one card from his hand and try to put it on the stacks. The operation is successful if the rank is exactly one larger than the top of the stack corresponding to the color. If it is successful, put the card on the top of the corresponding stack. The player recovers one information token if the rank of the card is 5 and he has less than eight tokens. If the operation is unsuccessful, discard the card and the group loses one life. There are three lives initially. If all three lives are lost, the game immediately ends. Finally, the player draws a new card from the deck.

The game ends in one of three ways: when the players have completed the stacks, when all three lives are lost, or when the last card in the deck is drawn and every player has taken one final turn. If all three lives are lost, the score is 0, otherwise the score is the number of cards in the stack, which is at most 25.

## 2 A Survey on Previous Works

Why is Hanabi so important? In [1] it is pointed out that Hanabi is a representative of fully cooperative game with imperfect information. A state-of-the-art technique which performs well on it may be generalizable to some other games.

There are a line of works on Hanabi. In [2] Hanabi game is proved to be NP-hard even all hidden information is revealed. In [3] a human-intelligence hat guessing strategy is applied to the Hanabi game and almost achieves the best performance for the five-player setting. By best performance, we mean a cheating strategy where each player sees his own cards and use the information strategy (details described in the paper). However, this performs badly when there are only two players as experimented in [4] and it is not generalizable to similar games. In [5] a Monte-Carlo tree search algorithm is given, but the performance is still not so good.

We mainly analyze (and try to improve) the methods two recent works: In [6] a method called Bayesian action decoder is invented. The general idea is to compute a joint public belief over the players' private features and to sample a deterministic partial policy to resolve the dilemma between informative actions and stochastic actions for exploration. This method works well in the two-player setting. In [7] a method called other-play is invented. The idea is to use known symmetries in the problem and find the other players' strategies over equivalence classes. This method performs much better than the same training method without other-play in cross-play settings.

However, the other-play method has some limitations: it relies on symmetries given in the problem. What if we don't know the symmetries? Can we learn the symmetries while learning the policies? Moreover, it seems the results of cross-play settings is not so satisfying. Can we develop better methods to improve the performance?

### 3 Our Goals and Ideas

Our first goal is to reproduce the results in [6] and [7]. Since our computation resources are limited, our results might be a little worse than the results in the paper.

Our second goal is trying to learn some symmetries if we don't know them initially. There are two possible solutions.

- **Compute approxiamte graph automorphisms.** Symmetries are actually graph automorphisms. During searching we can use some useful hash functions to calculate some approxiamte graph automorphisms and learn the symmetries. It might be useful when the symmetries are unknown.
- **Use symmetric parameters.** TODO.

Our third goal is to achieve better performance on cross-play settings. TODO.

### References

- [1] Bard, N., Foerster, J. N., Chandar, S., Burch, N., Lanctot, M., Song, H. F., ... & Dunning, I. (2020). The hanabi challenge: A new frontier for ai research. *Artificial Intelligence*, 280, 103216.
- [2] Baffier, J. F., Chiu, M. K., Diez, Y., Korman, M., Mitsou, V., van Renssen, A., ... & Uno, Y. (2017). Hanabi is np-hard, even for cheaters who look at their cards. *Theoretical Computer Science*, 675, 43-55.
- [3] Cox, C., De Silva, J., Deorsey, P., Kenter, F. H., Retter, T., & Tobin, J. (2015). How to make the perfect fireworks display: Two strategies for hanabi. *Mathematics Magazine*, 88(5), 323-336.

- [4] Bouzy, B. (2017, July). Playing hanabi near-optimally. In *Advances in Computer Games* (pp. 51-62). Springer, Cham.
- [5] Walton-Rivers, J., Williams, P. R., Bartle, R., Perez-Liebana, D., & Lucas, S. M. (2017, June). Evaluating and modelling hanabi-playing agents. In *2017 IEEE Congress on Evolutionary Computation (CEC)* (pp. 1382-1389). IEEE.
- [6] Foerster, J., Song, F., Hughes, E., Burch, N., Dunning, I., Whiteson, S., ... & Bowling, M. (2019, May). Bayesian action decoder for deep multi-agent reinforcement learning. In *International Conference on Machine Learning* (pp. 1942-1951).
- [7] Hu, H., Lerer, A., Peysakhovich, A., & Foerster, J. (2020). “Other-Play” for Zero-Shot Coordination. *arXiv preprint arXiv:2003.02979*.