CS221 Proposal: (Simplified) Hanabi Game Player

Qian Lin (linqian@stanford.edu)
Shanshan Xu (xuss@stanford.edu)

Game description: Hanabi is a cooperative card game, where the players try to reach a success state collectively. Players, aware of other players' cards but not their own, attempt to play a series of cards in a specific order to reach the success state. Players are limited in the types of information they may give to other players, and in the total amount of information that can be given during the game.

In a Hanabi game, there are five colors (pink, yellow, green, blue, white). For each color, there are three 1s, two 2s, two 3s, two 4s, and one 5. In addition, the group collectedly have 7 bitcoins. The goal state is where cards played on the table are ordered by number from 1 to 5, in all five colors.

During the game, each player will have 4 cards in his hand. All players take terms to act, and at each term he can decide between (1) play a card on his hand, in which case the predecessor card of the same color must have been played. Otherwise the card is discarded and a bitcoin is taken as punishment. (2) discard a card on his hand, which earns back a bitcoin. A new card is drawn after both (1) and (2). (3) tell one other play which of his cards are of a particular color or number.

After the first attempt to draw from an empty deck, each player gets another term to act, after which the result is evaluated to see if the goal state is reached.

In our project we will model the game as a multiplayer MDP process. To reduce the state space we will simplify the setup of the game: the number of colors Nc<=5, the deck size 0<=[N1,N2,N3,N4,N5], N1<=3, N2,N3,N4<=2 and N5<=3, and bitcoins Nb>=0.

Input: Nc, Ni (i=1-5), Nb, a randomly generated deck sequency (not known to the players). number of players Np, and playing conventions if any (for example always discard from the oldest card...)
Output: score of the final card state on the table (graded by how close to the success state)

Elevation Metric for Success:
The average score of multiple repeated games

Challenges:
Reduce state representation to a computation size. Design a strategy that could achieve a high success rate. Compare hard-code strategy, Q-learning. Investigate multi-player effects (known and unknown strategy of other players).

The baseline algorithm:
No information exchange. Each player makes decision randomly or deterministically based on the known cards.
The oracle:

More information are allowed. For example, the color and number information can be told at the same time by other players.
Or even with full information, that is, each player knows his cards.

Topics:
MDP.
The state is the known cards including the cards discarded, on the desk and in other player's hand.
The actions are the possible decisions of the player according to the game rule. The action's dependence on the state is the strategy we try to design.

Previous work:
A Hanabi variant, informationless Hanabi, which is exactly like Hanabi except you can only play or discard on your turn; you aren't allowed to spend your information tokens. We know of a deterministic strategy which perform reasonably, from a private friend. To our best knowledge, there isn't other public domain knowledge of a Hanabi game player is available.

Simulator:
The game is simulated by a random sampled sequence of cards. As a check, if each player knows his own cards, the simulator is expected to give a 100% success rate.