CSCI 491/591, Project P5
Progress Report

Group 8
Tao Huang and James Soddy

April 19, 2016

For our project, our group has been investigating the shape of play by various poker players. We started with the intention of exploring the space of poker play, without any more specific aim than to find something interesting. As we have worked on our project, our goal has solidified to attempt to find a relationship between the shape of a player's poker hands and their success while playing those observed hands. We have had some degree of success in finding shape in the data we have looked at, but the shapes are not as strong or as uniform as we would like. There are still some challenges ahead in finishing our project, but we have a solid basis to present and write up if our remaining work does not come up with stronger results.

Goals

Our original goals, which we laid in our proposal, were to find a way to create a shape which represented the play style of an individual player and to use those shapes to find similarities and differences between players. In the proposal we listed several objectives, which were:

- Define a topology which will give meaningful shape to the overall play for a given player.
- Create a method which will judge the difference between the shapes of each player's play
- Apply James' knowledge of poker to verify that the groupings our method defines seem reasonable.
- Be able to positively identify an individual player based on the 'fingerprint' of the shape of their play
- Come up with some interesting and unexpected conclusions

We have made what we think is significant progress toward all of these objectives, with the exception of positively identifying a single player. We have set that goal aside, as our techniques have not come close to that level of precision.

Initial Investigations

Our initial work involved doing some research into poker, the types of different play styles, and find some concrete metric we could compare our hands to. We were able to find some information about the primary differences in poker play styles, as well as details about the value of different starting card combinations to the player holding them. We thought we could use the play styles to help us interpret any groupings among players that we might find in our analysis. We plotted the starting hand values, in the hope that the resulting shape might be useful as a comparison point for the shapes generated by actual players. (Shown in figure 1)

We also looked at academic literature to see if we could find work similar to ours which we could apply. While we found academic papers on some aspects of poker, they weren't topology related or particularly useful to our goals. We also found a number of papers related to topology of games. Most of the topology papers dealt with properties of special topological games, and of the remaining papers, there were
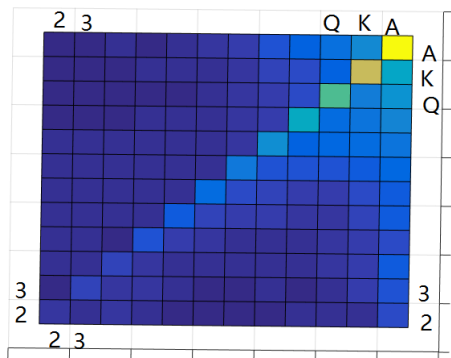
1

Figure 1: **Heatmap of Poker Hand Values.** All possible differently valued starting Hold'Em hand combinations aligned in a grid, with their color indicating profitability (lighter is better)


```
Stage #3019753234: Holdem  No Limit $10, $2.50 ante - 2009-07-01 20:10:57 (ET)
Table: PROSPECT DR (Real Money) Seat #4 is the dealer
Seat 4 - +38WdZ4L8J7gkKBgk8Y+aQ ($2,169.25 in chips)
Seat 5 - giyaeGLs57Tb7W9Ci0jVzA ($2,840.65 in chips)
Seat 1 - /P+7Z0P/b7YiK60FW9dRAQ ($1,111 in chips)
Seat 2 - 68rbdJHmi6SSDnEzLS6S8Q ($691.74 in chips)
Seat 3 - c2tiA/SMUK+T0PsP2rCOGA ($1,783.49 in chips)
/P+7Z0P/b7YiK60FW9dRAQ - Ante $2.50
68rbdJHmi6SSDnEzLS6S8Q - Ante $2.50
c2tiA/SMUK+T0PsP2rCOGA - Ante $2.50
+38WdZ4L8J7gkKBgk8Y+aQ - Ante $2.50
giyaeGLs57Tb7W9Ci0jVzA - Ante $2.50
giyaeGLs57Tb7W9Ci0jVzA - Posts small blind $5
/P+7Z0P/b7YiK60FW9dRAQ - Posts big blind $10
*** POCKET CARDS ***
68rbdJHmi6SSDnEzLS6S8Q - Folds
c2tiA/SMUK+T0PsP2rCOGA - Folds
+38WdZ4L8J7gkKBgk8Y+aQ - Raises $40 to $40
giyaeGLs57Tb7W9Ci0jVzA - Raises $135 to $140
/P+7Z0P/b7YiK60FW9dRAQ - Folds
+38WdZ4L8J7gkKBgk8Y+aQ - Raises $355 to $395
giyaeGLs57Tb7W9Ci0jVzA - Raises $900 to $1,040
+38WdZ4L8J7gkKBgk8Y+aQ - All-In(Raise) $1,771.75 to $2,166.75
giyaeGLs57Tb7W9Ci0jVzA - Calls $1,126.75
*** FLOP *** [Qh 7s 3s]
*** TURN *** [Qh 7s 3s] [10s]
*** RIVER *** [Qh 7s 3s 10s] [2d]
*** SHOW DOWN ***
+38WdZ4L8J7gkKBgk8Y+aQ - Shows [Kd Ks] (One pair, kings)
giyaeGLs57Tb7W9Ci0jVzA - Shows [Qs Qd] (Three of a kind, queens)
giyaeGLs57Tb7W9Ci0jVzA Collects $4,353 from main pot
*** SUMMARY ***
Total Pot($4,356) | Rake ($3)
Board [Qh 7s 3s 10s 2d]
Seat 1: /P+7Z0P/b7YiK60FW9dRAQ (big blind) Folded on the POCKET CARDS
Seat 2: 68rbdJHmi6SSDnEzLS6S8Q Folded on the POCKET CARDS
Seat 3: c2tiA/SMUK+T0PsP2rCOGA Folded on the POCKET CARDS
Seat 4: +38WdZ4L8J7gkKBgk8Y+aQ (dealer) HI:lost with One pair, kings [Kd Ks - P
Seat 5: giyaeGLs57Tb7W9Ci0jVzA (small blind) won Total ($4,353) HI:($4,353) wit
P:Qs,B:Qh,P:Qd,B:10s,B:7s]
```

Figure 2: **Example hand.** One hand taken from the hand history files we used in our project

none which dealt with the topology of a player space or games with incomplete information. So we needed to find our own techniques.

Hand Histories

Another part of our work was dealing with the large number of hand history files that we needed to extract data from. Our first step was to write a small perl script which could parse the raw text files (example in figure 2) and retrieve all of the information that we needed. The script created a file for each player the first time they were encountered, and appended a new line for each hand the player played in. On this line was each decision they made in the hand, as well as what cards they had if that information was available. Once in this format, the data would be easy to manipulate in any way we needed.

Player Shapes

Once we had our data in a manageable form, we needed to visualize it. We created a second perl script which read the player files created by the first script and calculated the average amount of money they were willing to invest during the first betting round with each possible starting hand. We then used MATLAB

to visualize these data in 2-dimensional and 3-dimensional pictures, similar to the one we created for the starting hand values.

Looking at these charts, we were able to see clear patterns, and make some conclusions about the players the charts had come from. Some of the charts which were visually fairly similar to the starting hand value graphic belonged, as expected, to the players who were winning at the highest rates. Also, some of the shapes which bore little or no resemblance to the hand value shape belonged to players with the heaviest losses. However, there was a great deal of noise in the shapes, which we attributed mostly to the number of hands for each player with hole card data available being very limited.

Distance

Our next step was to attempt to find a useful way to measure the distance between the shapes generated by different players. We partitioned our data into two sets of hands for each player, so that we would have a way to compare check that the shapes and distances actually said something about the players. Our first attempt was to use

a Hausdorff distance calculation package for R on our data. Unfortunately, we found that the distances calculated were not any closer for our partitioned data than for shapes generated by completely different players. This contradicted our principle that the techniques we use should return consistent results.

We attempted several other metrics, and found our best results using a type of L^1 norm on the data. Unfortunately, the average distance between the shape of two partitioned sets from one player was still only about 25% closer than the distance between shapes from two unrelated players. The fact that we can see a definite difference leads us to believe that our metric can be useful, but the small magnitude of the difference makes it unhelpful for performing calculations or coming to a better understanding of the data. We believe that if we had a larger set of data with hole cards then we would be able to clean up some of the noise and achieve more meaningful results.

Future Work

Currently, we are working to create a space of hands played based only on the decisions made during the hand, and without any player card information included. This is helpful because without the need to have players' hidden card information, the number of hands we can use increases significantly. However, the reason we have avoided this path is because it is much less clear how the space should be defined. We hope to have this work done by the end of the week, leaving us enough time to consider our results and include them in our final report.

Some other work we still need to complete:

- Attempt a few more variations on our hidden-card shapes, such as removing zero points from the calculations, converting to a logarithmic scale, and working from a point cloud of all hands observed, rather than averaging them before doing calculations. This should be fairly straightforward and we should be complete on Saturday.

- We need to interpret our results, whether or not we are able to find any improved metrics before doing so. By Monday, April 25th we should have our analysis completed and know what we will be including in our reports.

- We will need to complete our presentation and final paper, detailing our experience and any results. This portion should be completed by Wednesday, April 27th at the latest.

Conclusion

We have tried several different techniques to analyse our poker hand history data, but have been somewhat disappointed with the results. We have been able to see some interesting patterns and draw some conclusions, but we are still holding out hope that we will have something stronger and more computable by the time we are done. While the tasks we still have to complete are significant, they are also manageable and we feel good about our prospects for turning in an interesting presentation and a quality paper.