CSCI 491, Project P2

Group 8

Tao Huang and James Soddy

March 1, 2016

Our group has identified a dataset and operations on that dataset which we believe will find new information which has not been explored in the past, while also being both within reach of our knowledge and inside the scope of this project.

Overview

For our project, we will be performing a topological analysis of a large database of poker hands. These hands will be sourced both from hands played personally by James, and from datasets which have been made available for such research [1]. We will define a topology on this data, and attempt to find a pattern in its shape.

We considered several angles for this, but many of them seemed to broad, or too poorly defined. We decided on a tack which we feel will provide interesting results in a way that is more clearly defined. Rather than focusing on the shape of the game of poker, or of the space of poker hands, we will be looking at the shape of poker play of an individual player. We will attempt to find similarities and differences between the characteristics of the shape of play between different players.

Goals

In poker, every player makes many different decisions throughout the course of each play session. A player chooses call, raise or fold at their turn during every betting round. They choose how much money to keep in front of them and how long to play. In some types of poker a player chooses how much to bet when they bet.

If each player's decisions were random, or if all players played the game rationally and with complete information, we would expect that every player would exhibit a very similar pattern through these variables. However, that is clearly not the case. Observation of poker players shows that there are a great variety of different play styles. One player might never raise, another might raise 20 to 30 percent of the time, while another might raise every time the action is on them. Some players buy in for the maximum allowed, others for the minimum, while yet others buy in for some random amount which is hard to understand. It is our belief that if we define a topology over the decisions made in poker hands, then we will see a unique shape appear for players with a certain style of play.

We hope to be able to complete at least several of the following objectives:

- 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

Techniques

Our initial plan, which may change as we proceed, is to plot the hands played by each player based on the categories we have defined, and use persistent homology to determine the shape for each player. Due to the high dimensionality of the data we will be working with, we are likely to attempt to apply Principle Component Analysis to the datasets in order to reduce the number of factors we need to consider. After we have found a way to define the shape of each individual player, we will then proceed to define the distance between any two of those shapes.

Of primary importance to us is that we develop a technique which produces reliable results. To this end, we will break the hand sets of our various players into multiple samples. We will then attempt to refine our techniques until the shape of each subset of a single player's hands will be recognized as having a similar shape. A technique which can give wildly different results to samples from the same player can easily be discarded from consideration as a method to categorize different players.

Conclusion

In our decision process, we considered quite a few different data sets as candidates for study. What finally pushed us to choose this set is the fact that it is a large, complex dataset from an area in which topology has not yet been extensively applied, but we have found some way in which to make it approachable. The sheer scope of the data was intimidating at first, but the idea of looking at the shape of players makes it feel much more reasonable. Paradoxically, the decision to analyse at the player level both reduces the amount of data which we will need to process at a time and also makes the data groupings we will create bigger and more meaningful. We are excited to see what shapes emerge as we start to apply topological techniques.

References

[1] http://web.archive.org/web/20110205042259/http://www.outflopped.com/questions/286/obfuscated-datamined-hand-histories