Predicting Texas Hold'em Hand Strength

Jigar Gada, Akash Metawala
jgada@usc.edu, metawala@usc.edu

Problem Statement:

Estimating the hand strength of other players in a game of Texas Hold'em Poker, based on their betting patterns.

Description of dataset:

The dataset that we are using is an available dataset with the Poker Research Group at the University of Alberta. The link to the available dataset is: http://poker.cs.ualberta.ca/irc_poker_database.html. The dataset is of poker games played over Internet Relay Chat (IRC (by Humans)) poker server and each player has a limited amount of chips. There are a number of different types of hand databases stored overall. We are choosing to learn on hold'em tournament data. There are approximately 10,000 data samples (poker games). The dataset contains decisions made by a player (calls, raise, checks) at every instance of the game (pre-flop, flop, turn, river). These details are real-valued.

Description of approach:

Data Extraction:

We extract the datasets from the link available to us. The datasets are initially stored as individual databases. We plan to organize the data as a relational model, with a table of hand, table of roster and a table of player actions for each player.

Proposed Feature set:

- Some of the feature sets that might be used for each player are:
  - Number of Checks before the river
  - Number of Bets before the river
  - Number of Calls before the river
  - Number of Raises before the river
- We might use more robust feature sets like:
  - Value of bet compared to the current pot amount
  - Number of bets in certain ranges compared to the current pot
  - Ratio of aggressive (bet/raise) moves to passive (check/call) moves
  - Average bet in the round compared to the pot size
  - Number of actions in the round

Model Selection:

Once we have our features selected and our data organized, we plan to use various machine learning algorithms to train on the data. We plan to use Linear Regression as base line followed by SVR, Naïve Bayes and Random Forest. Later, if enough time, we plan on clustering similar types of users and use above mentioned machine learning models for improving accuracy.

Model Validation:

- We are going to divide the dataset into 3 parts:

- Training (Cross-Validation for selecting the best feature-set and dimension for a model)
- Test-Validation (Evaluation the best model)
- Final-Test-set (Final Results on the best model)

Work Segregation:

The downloaded data needs to be arranged in form of a proper database. So creating the database using SQL will be done by Akash and integrating it with Python will be done by Jigar. Given the time frame, we plan to collaborate on Model selection. As a team, we can play around with many different Machine Learning models to see which one would work the best.

Reference:

1. James Bensson, Alex Eckert, Maxwell Wu, December 13, 2013, Predicting Texas Holdem Hand Strength.
2. Computer Poker Research Group. University of Alberta, 2013. Web. 5 Dec. 2013. <http://poker.cs.ualberta.ca/>.