

# Research Proposal on Genetic Algorithm and Poker Rule Induction

Wenjie Xu

January 26, 2015

## 1 Overview

Genetic algorithm represents a set of very useful general methods in finding the best answer(s) to a problem. The fundamental idea behind genetic algorithm is not too hard to understand and shown to be powerful in many fields. However, it is relatively difficult to give a rigorous proof of the success of the algorithm. Thus, instead of explaining why the algorithm works, this research project would focus on introducing the general procedure and one of its interesting applications, which is poker rule induction. This comes from an open question in one of the academic competitions on kaggle.com (<http://www.kaggle.com/c/poker-rule-induction>). It is suggested that genetic algorithm can be used to derive the underlying poker rules out of the available poker hands. Thus, it is possible to embed the idea of genetic algorithm into machine learning in order to build models that predict the category of an unknown hand.

## 2 Methodology

### 2.1 References

I would refer to relevant books to acquire basic knowledge of genetic algorithm. The list of books that I have in mind for now includes *An Introduction to Genetic Algorithms for Scientists and Engineers* by David A. Coley, *Genetic Algorithms in Search, Optimization, and Machine Learning* by David E. Goldberg, *Programming Collective Intelligence* by Toby Segaran, and probably some book on rule induction in general.

### 2.2 Data

The data include a training set and a test set. The training set consists of 25,010 five-card poker hands with the corresponding class label. The test set consists of 1,000,000 poker hands without any class label. The class label represents the specific ordinal category of a specific hand. The data are all stored in

csv files and available at <http://www.kaggle.com/c/poker-rule-induction/data>.

## **2.3 Machine Learning**

This would be the most challenging part of the whole project, applying the methods to real data. Ideally, I would build classification models in Python, for example, decision trees, integrating ideas of genetic algorithm. Different models would be assessed, based on training errors and other model selection criteria. The goal is to predict the class label for poker hands in the test set. Some of the difficulties lie in the relative large size of the test set compared to the training set and the lack of representativeness of the training set; for example, the straight flush is 14.43 times more likely to occur in the training set, while the royal flush is 129.82 times more likely.

## **3 Conclusion**

In general, I regard this research topic as an opportunity to learn about a new idea, integrate it with my knowledge of machine learning, and get some practice in solving real-world questions. Indeed, this proposal only covers some of my original, partial thoughts about the topic. I'm certain that new challenges and questions will arise; those will only make the project more meaningful.