Sandia AESOP FABLES

PURDUE UNIVERSITY | The Data Mine

Matt Schnell, Ben Huffman, Max Kushner, Ian Clausen, Lillian Zhao, Grant McCord, Trevor Hoivik, Gavin Ma, Chelsea Lee, Ayush Thirumala, Aaryan Dave, Adam Ventura

About Sandia

Sandia National Laboratories is the engineering arm of the nation's nuclear weapons enterprise, working to ensure the reliability, safety, and security of the nuclear arsenal. To this end, Sandia seeks to better understand what makes strategic deterrence effective.

In collaboration with UC Berkeley, Sandia built a three-player, turn-based game called Strategic Interaction Game between Nuclear Armed Lands (SIGNAL) in order to study nuclear conflict escalation and the strategic stability of nations in an artificial world.

Project Background

- The AESOP project name stands for Analytically Extrapolating Strategies of Opponents from Play-data
- Our project scope had three deliverables:
  - Build an analytical toolkit in order to automatically analyze game data that would:
    - Generate basic game statistics of the data set such as number of games, length of the game
    - Generate win probability conditioned on first move, player type, or other specified conditionals
    - Cluster strategies employed by the players
    - Produce graphs to display computations
    - Accept new functionality
  - A general file format for the recording of wargaming data to be recommended for use by the broader wargaming community was also requested
  - Finally, we were tasked with exploring different means of generally extrapolating player strategy using only game data as a unit of analysis

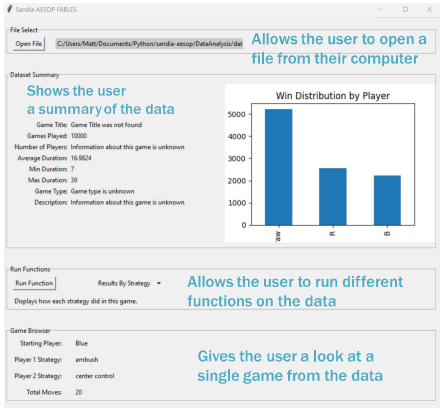


Figure 2: The analytical toolkit integrated into a GUI (Blue text denotes annotations)

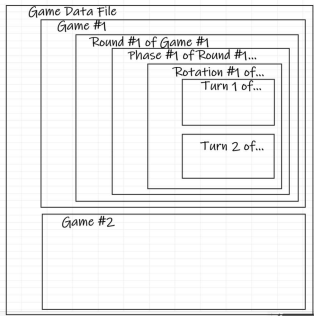


Figure 1: The hierarchal structure of game data

RESEARCH METHODOLOGY

- Fables and File Format:
  - In order to record the rich features at different points in the game, we developed a hierarchal structure to capture moves and other relevant attributes, such as player resources, on a player-turn basis, seen in Figure 1
  - Each layer is stored within a .json file to preserve the hierarchy
  - We use this hierarchal structure in our toolkit, which we named Framework for Analytics Based on Likelihood Estimation of Strategies (FABLES), seen in Figure 2
  - From this hierarchal structure, tables are generated, as they are needed, to perform the intended computation, an example of conditional win percentages by player strategies can be seen in Figure 3
  - Basic game summaries are generated when the file has been loaded, and more advanced functions can be specified by the user to compare different features within the data space
- RiskWorld
  - In the building of FABLES, we decided that a simple game would help us to test our hypotheses about how we might extract strategies and the toolkit itself
  - We developed RiskWorld, a grid-based, two-player board game with an objective of capturing the opponent's base or defeating all of the opponent's pieces, seen in Figure 4
  - Pieces can be moved horizontally or vertically, but not diagonally
  - Placing a piece onto an occupied square stacks the pieces if the square is occupied by allied pieces, or attacks the square if the square is occupied by enemy pieces
  - Each piece has a 50% chance of defeating an enemy piece; if there are multiple enemy pieces, the piece rolls again to attack the remaining pieces until all enemies are defeated or the attacker runs out of pieces
  - Built four different artificial intelligence (AI) strategies, offensive, defensive, ambush, and center control
  - We created data sets of 10,000 moves for each combination of pieces and board sizes from 3 to 6, inclusive
  - These data sets serve as reference data sets to classify games with random noise added to the moves
  - Additional data sets of 10,000 moves are created with designated noise added
  - Random forests and deep neural networks are used for classification

Figure 3: FABLES example output

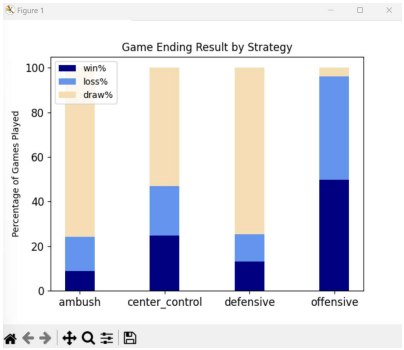
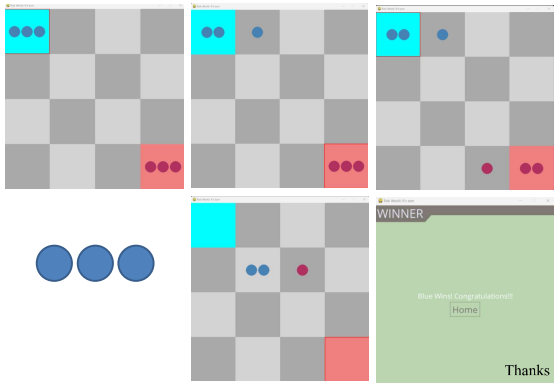


Figure 4: RiskWorld



CONCLUSIONS

- Fables and File Format:
  - We were able to come up with a hierarchy for the structure of game data and devised a means of encoding and decoding that hierarchy
  - We developed an analytical toolkit, FABLES, that would accept structured game data and provided requested outputs
  - FABLES generates an overall summary of a game data file and has prebuilt functions for data analysis that the user can request and customize
  - We found that maximizing customization options allows us to service a broader user base of varied backgrounds
  - The customization is also necessary to service different types of games, though we have primarily tested it on RiskWorld
- RiskWorld
  - Through the RiskWorld experiments, we found that the game design has a large effect on the ability to classify strategies
  - For example, the action space in a board size of 3x3 is not large enough to adequately distinguish all of the strategies
  - We also found that there are a lot of missing observations within our data structure, due to games ending at varied lengths
  - We found random forest classification to be the most resilient method to the variable observation length due to the ability of random forest to make splits based on the maximum entropy
  - Adding move-level noise to the play data linearly reduces the classification accuracy of the model
  - We have an article in draft summarizing all of our findings from the RiskWorld experiments

FUTURE GOALS

- We would like to test our framework on other types of games
- We plan to deliver the framework to Sandia to see it incorporated as a complement to their Conflict Analysis Strategy Testing Laboratory Environment (CASTLE) framework for building experimental wargames
- We plan to publish our findings from the RiskWorld experiments
- We additionally have a study currently under IRB review in order to capture human play data of RiskWorld in order to test our framework
- We also plan to publish our findings from the human play data at the conclusion of the study, if approved

SPECIAL THANKS

A special thanks to Dr. Jason Reinhardt, Grant McCord, Kevin Amstutz, Nick Lenfenstey, Dr. Mark Ward, and all the staff from Purdue's The Data Mine for enabling this project.