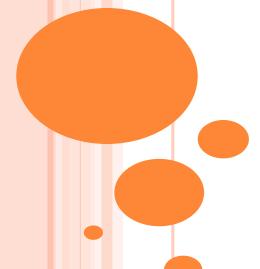
# EXTRACTING RULES FOR BLACKJACK USING MACHINE LEARNING AND FUZZY SYSTEMS

 $\textbf{Presentation} \cdot \mathsf{April}\, 2019$ CITATIONS READS 178 3 authors, including: Marcos Evandro Cintra Márcio Basgalupp Universidade Federal Rural do Semi-Árido - UFERSA Universidade Federal de São Paulo 53 PUBLICATIONS 806 CITATIONS 70 PUBLICATIONS 450 CITATIONS SEE PROFILE SEE PROFILE Some of the authors of this publication are also working on these related projects: National Seminar on Good Habits for Distance Education View project  $On the Generation of Fuzzy Classification Systems Using the Genetic Paradigm and Fuzzy Decision Trees for the Application on Large Datasets \\View Decision Trees for the Application on Fuzzy Classification Systems Using the Genetic Paradigm and Fuzzy Decision Trees for the Application on Fuzzy Classification Systems Using the Genetic Paradigm and Fuzzy Decision Trees for the Application on Fuzzy Classification Systems Using the Genetic Paradigm and Fuzzy Decision Trees for the Application on Fuzzy Classification Systems Using the Genetic Paradigm and Fuzzy Decision Trees for the Application on Fuzzy Classification Systems Using the Genetic Paradigm and Fuzzy Decision Trees for the Application on Fuzzy Decision Trees for the Application Systems Using Systems Usi$ project

## Federal University of the Semi Arid - UFERSA Federal University of São Paulo - UNIFESP



# EXTRACTING RULES FOR BLACKJACK USING MACHINE LEARNING AND FUZZY SYSTEMS

Karla R. Cardoso Marcos E. Cintra Marcio Basgalupp

## CONTENTS

- Introduction to Black Jack
- Existing Strategies
- Experiments
- Results
- Future Work

# THE BLACK JACK GAME



## BLACKJACK

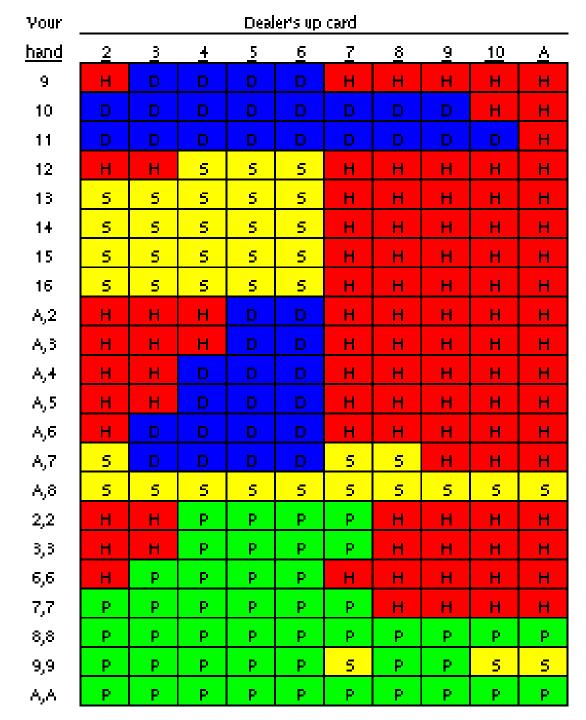
- Blackjack originated in French casinos around the 1700s where it was called "vingt-et-un" (twenty-and-one).
- There are approximately 140 countries in the world with casinos that offer blackjack.
- The object of the game is not to get 21; the real object is to beat the dealer.
- Mathematically, in 100 hands the dealer will win 48 hands, the player will win 44 and there will be eight ties.
- Why does the dealer always win more hands? Because the dealer plays their hand LAST.
- Casinos have the right to ask "card counters" and anyone else to leave the casino based on the old English law that states, "Management has the right to refuse service".
- House rules are important. They dictate what your advantage or disadvantage is.

## BLACKJACK

- Existing strategies: systems for mentally counting cards in order to decide the next move.
- Strategies defined in the 1960's by mathematicians using probability.
- In this study, we aimed at finding new strategies using machine learning techniques applied to large sets containing information on possible hands of Blackjack.
- Hypothesis: it is possible to find new Blackjack strategies in the form of rule sets that offer advantage to the player over the dealer

# EXISTING STRATEGIES

# THORP STRATEGY



## THE HIGH-LOW STRATEGY

- One of the most popular and easy strategies to count cards
- Each card receives a value: -1, +1, or 0. The sum of all cards is zero.
- The values of each card are:
  - Cards from 2 to 6 = +1
  - Cards from 7 to 9 = 0
  - Cards from 10 to Ace = -1
- The bigger the sum of all cards, the bigger the odds of big cards in the deck
- On the other hand, if the sum is negative and the player has to take a decision, the player can Hit, knowing that small cards are more probable.
- The more cards are distributed and counted, the more precise the Hi-Lo method is.

- Classification, different approaches:
  - Artificial neural networks
  - Support vector machines
  - Rule-based approaches:
    - Decision trees
    - Genetic Fuzzy Systems

- Fuzzy systems:
  - Fuzzy rule base
  - Fuzzy data base
  - Inference Mechanism
- Fuzzy systems can be defined by experts or automatically learned from data

- Genetic fuzzy systems:
  - Global search of Genetic Algorithms
  - Fuzzy set theory
  - Provide accurate and interpretable rule-based systems

## ADOPTED ALGORITHMS FOR RULE EXTRACTION

- We adopted:
  - C4.5
  - FuzzyFCA
  - FuzzyDT
  - PART
  - Ripper
- Empirically selected:
  - Computational cost
  - Ability to deal with large datasets
  - Produce interpretable models (rule sets)

- Two datasets:
  - 10,000 and 1,000,000 Blackjack hands randomly created.
  - ARFF format, in order to be used with WEKA's algorithms.
- In order to define the entries of the datasets, we considered the following restrictions:
  - The use of a single deck of cards;
  - The entries represent hands between a single player and the dealer;
- Maximum number of cards a player or dealer can get not exceeding 21 is 11 cards, considering the drawn cards are all four ace cards, four 2 cards, and three 3 cards, totalling 21 points.
- Due to the low probability of a player getting the 11 smaller cards, and due to processing and memory restrictions, we empirically defined 8 as the maximum number of cards the player or dealer can have in a single hand for the 10,000 hand dataset.

#### • Entries format:

• CoP1, CoP2, CoP3, CoP4, CoP5, CoP6, CoP7, CoP8, CoP9, CoP10, SoP, TCoP, CoD1, CoD2, CoD3, CoD4, CoD5, CoD6, CoD7, CoD8, CoD9, CoD10, SoD, TCoD, W;

#### • Where:

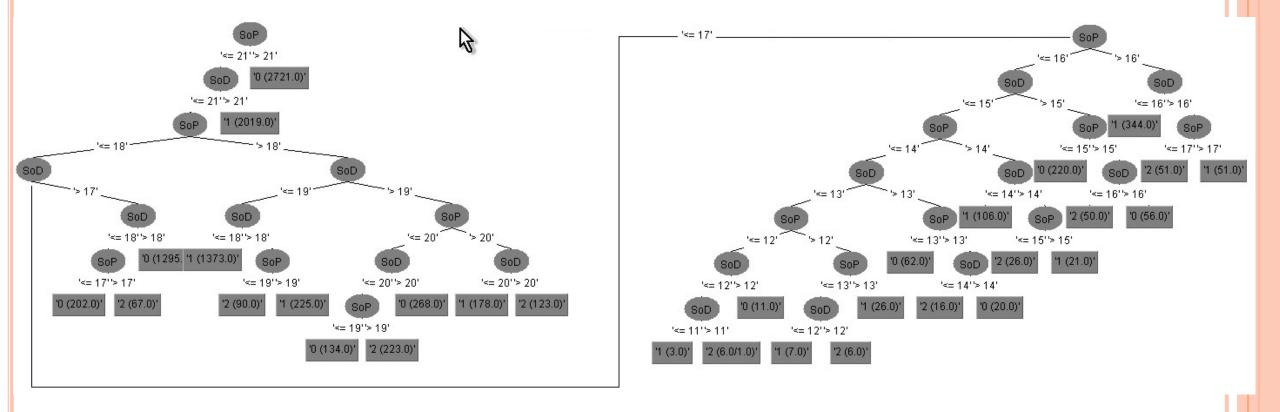
- CoP1 to CoP10: value of the 1st to 10th player's cards;
- SoP: sum of all player's cards when he/she stands;
- TCoP: total number of cards of the player in the hand;
- CoD1 to CoD10: value of the 1st to 10th dealer's cards;
- SoD: sum of all dealer's cards at the end of the hand;
- TCoD: total number of cards of the dealer in the hand;
- W: winner. W is set to 0 (zero) if the dealer wins, 1 (one) if the player wins, and 2 (two), if there is a tie.

- Consider the following example of the dataset with 10,000 entries (up to 8 cards for the player and dealer):
  - 4, 6, 7, 3, 0, 0, 0, 0, 20, 8, 8, 2, 0, 0, 0, 0, 0, 18, 3, 1;
- The values of the entry correspond to:
  - 4, 6, 7, 3, 0, 0, and 0 are the values of the cards of the player;
  - 20: sum of the values of all the cards of the player
  - 4: number of cards the player received;
  - 8, 8, 2, 0, 0, 0, and 0 are the values of the cards of the dealer;
  - 18: sum of the values of all the cards of the dealer;
  - 3: number of cards the dealer received;
  - 1: the player won the hand.

## EXPERIMENTS -10,000 HANDS

- Experiments using C4.5 (3 models generated)
  - Using all attributes (21 attributes);
  - Using only 5 attributes: the sum of the cards of the player and dealer (SoP and SoD), the number of cards used by the player and the dealer (TCoP and TCoD), and the winner;
  - Using all attributes, except the SoP and SoD.
- In spite of the difference in the number of attributes, both datasets (containing 10,000 and 1,000,000 randomly generated Blackjack hands) produced the same model

## GENERATED TREE FOR C4.5



## GENERATED TREE FOR C4.5

- When considering only the branches that has the Player as the winner, the decision tree can be seen as a set of 10 rules.
- Some rules randomly selected:
  - 1. If SoP > 21 then Winner = Dealer
  - 2. If  $SoP \le 21$  and SoD > 21 then Winner = Player
  - 3. If  $SoP \le 18$  and  $18 < SoD \le 21$  then Winner = Player

#### EXPERIMENTS USING PART

- PART produced 25 rules.
  - If SoD > 21 then Winner = Player (2019.0)
  - If SoP > 19 and  $SoD \le 19$  then Winner = Player (1254.0)
  - If  $SoD \le 16$  and SoP > 16 then Winner = Player (559.0)
- The rules extracted by PART, as with C4.5, only include the sum of the cards of the player and of the dealer
- In an attempt to extract relevant rules, we removed the SoP and SoD attributes and obtained a large set of 398 rules
- However, the rule set has two important characteristics:
  - Rules present only one condition for each attribute (C4.5)
  - The rule set: indicates that when reaching 17 points, the player should Stand.

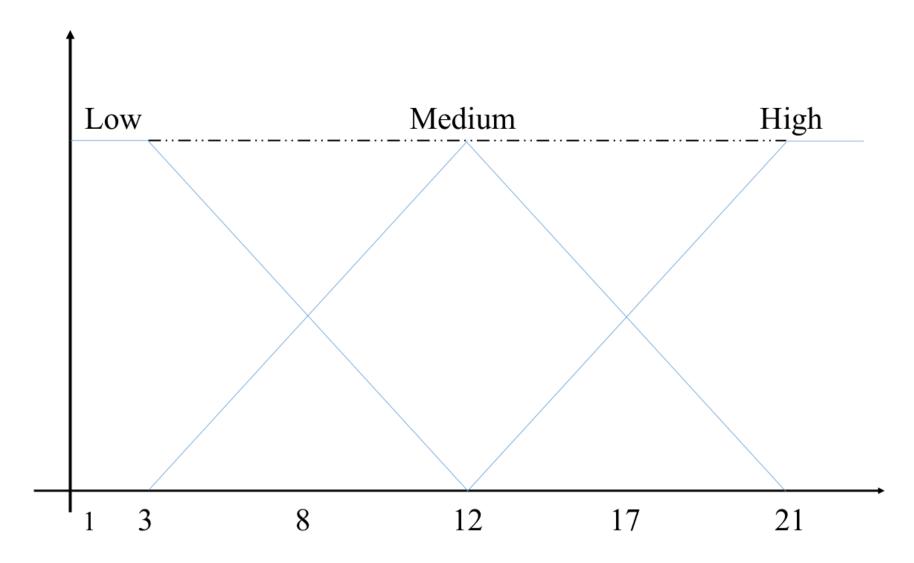
## EXPERIMENTS USING RIPPER

- Ripper extracted a set of 36 rules.
  - 1. If  $CoP1 \ge 10$  and  $CoP2 \ge 10$  and  $CoP3 \le 0$  and  $CoD1 \ge 10$  and  $CoD2 \ge 10$  and  $CoD3 \le 0$  then Winner is Tie
  - 2. If  $CoP3 \le 0$  and  $CoD2 \ge 8$  and  $CoD3 \ge 1$  and  $CoD4 \ge 6$  then Winner is Player
  - 3. If  $CoP3 \le 0$  and  $CoD1 \ge 8$  and  $CoD2 \ge 4$  and  $CoD3 \ge 8$  then Winner is Player

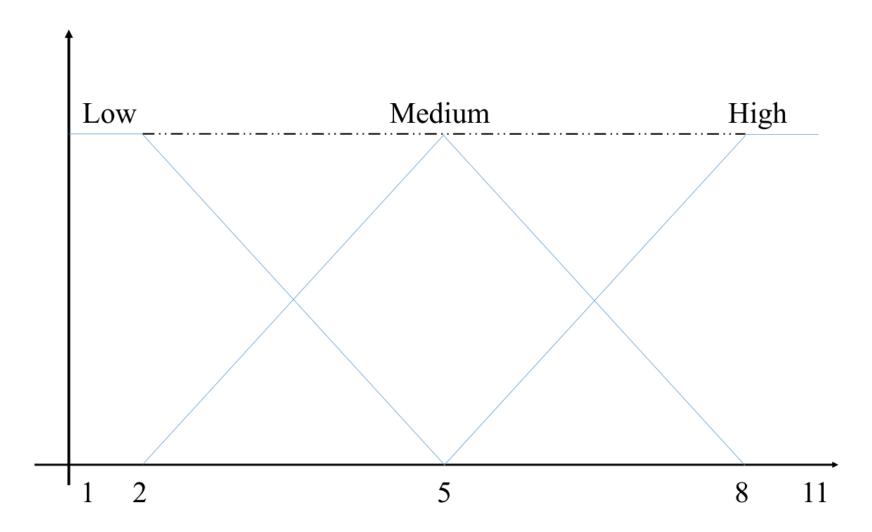
## EXPERIMENTS USING FUZZYDT

- FuzzyDT and FuzzyFCA
  - 3, 5, and 7 triangular sets evenly distributed in the domains.
- The definitions of the linguistic variables were done empirically.
- The total number of rules obtained by FuzzyDT:
  - 3 fuzzy sets: 319
  - 5 fuzzy sets: 517 rules
  - 7 fuzzy sets: 451 rules.

# DEFINITION OF THE SUM OF CARDS



# DEFINITION OF THE SUM OF CARDS



## EXAMPLE OF THE GENERATED RULES

- If CoP1 = low and CoP2 = low and SoP = low and CoD1 = low and CoD2 = medium and CoD3 = low and SoD = low and TCoD = low then  $Volume{T}$  where  $Volume{T}$  is a sum of  $Volume{T}$  and  $Volume{T}$  and  $Volume{T}$  and  $Volume{T}$  and  $Volume{T}$  is a sum of  $Volume{T}$  and  $Volume{T}$  and  $Volume{T}$  is a sum of  $Volume{T}$  and  $Volume{T}$  and  $Volume{T}$  and  $Volume{T}$  is a sum of  $Volume{T}$  and  $Volume{T}$  and  $Volume{T}$  is a sum of  $Volume{T}$  and  $Volume{T}$  is a sum of  $Volume{T}$  and  $Volume{T}$  and  $Volume{T}$  is a sum of  $Volume{T}$  and  $Volume{T}$  and  $Volume{T}$  is a sum of  $Volume{T}$  and  $Volume{T}$  and  $Volume{T}$  is a sum of  $Volume{T}$  and  $Volume{T}$  and  $Volume{T}$  is a sum of  $Volume{T}$  and  $Volume{T}$  and  $Volume{T}$  is a sum of  $Volume{T}$  and  $Volume{T}$
- If CoP1 = low and CoP2 = low and SoP = low and CoD1 = low and CoD2 high CoD3 = low and SoD = low and TCoD = low and then Winner = Player
- If CoP1 = medium and CoP2 = medium and SoP = low and CoD1 = low and CoD3 = low and SoD = low and TCoD = low then Winner = Player



## EXPERIMENTS USING C4.5

- C4.5 generate three models:
  - 1. Using all attributes (21 attributes);
  - 2. Using only 5 attributes: the sum of the cards of the player and dealer (SoP and SoD), the number of cards used by the player and the dealer (TCoP and TCoD), and the winner;
  - 3. Using all attributes, except SoP and SoD.
- The two first models are identical to the models obtained with the smaller dataset and present only the sum of the cards

## EXPERIMENTS USING FUZZYFCA

- 10-fold cross-validation strategy: 10 sets of rules (average of 320 rules each)
- We used only the rules present in all 10 rule sets.
- We found 9 rules in each the Dealer wins in all 10 rule sets, and 4 rules with ties in all 10 rule sets.
  - 1. If CoP1 = low and CoP2 = medium and CoP4 = low and CoP5 = low and SoP = low and CoD2 = high and CoD3 = low and TCoP = high and SoD = medium then Winner = Player
  - 2. If CoP3 = medium and TCoP = medium and CoD2 = medium and CoD3 = high then Winner = Player
  - 3. If CoP2 = high and CoP4 = low and CoP5 = low and TCoP = medium and SoP = low and CoD2 = low and CoD5 = low and TCoD = high then Winner = Player

## CONSIDERATIONS ON THE RULE SETS OBTAINED

- o C4.5: 398 rules;
- FuzzyFCA:  $\approx 320$  rules;
- FuzzyDT: 319 rules;
- PART: 398 rules;
- Ripper: 39 rules.

# EVALUATION OF THE OBTAINED MODELS

	Player	Dealer	Tie
C4.5	42	46	12
FuzzyFCA	71	21	8
FuzzyDT	40	58	2
PART	67	24	9
Ripper	40	46	14

## CONSIDERATIONS

- Blackjack is one of a few card games played in casinos that gives the players some advantage over the dealer.
- Strategies are based on card counting from studies of mathematicians who analysed large sets of examples of the game and applied probability to devise counting strategies.
- Experiments carried out using machine learning algorithms and rule-based fuzzy systems to extract rules from two sets of 1,000,000 and 10,000 Blackjack hands.
- o C4.5, PART, Ripper, FuzzyDT and FuzzyFCA

## CONSIDERATIONS

- All 4 decision trees induced by C4.5 only used the attributes containing the sum of the cards of the player and dealer.
- In an attempt to obtain relevant rules, the C4.5 algorithm was executed removing the SoP and SoD attributes, generating a large set of rules.
- Large rule sets: FuzzyDT, FuzzyFCA, and PART.
- Ripper produced a smaller rule set, but with many conditions.
- For FuzzyFCA, a small rule set was selected contemplating only the rules found in all 10 rule sets.

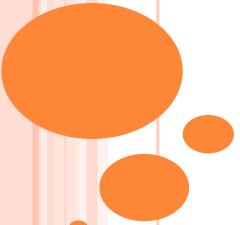
## FUTURE WORK

- Condensation of the rule sets in order to define table representations of strategies for each rule set, similar to the table defined by Thorp
- Compare the obtained models with Thorp.
- Definition of the linguistic variables used by FuzzyFCA and FuzzyDT with two fuzzy sets.
- Extract rules using other machine learning and fuzzy algorithms.
- Play more hands with the extracted models.
- Use more than one deck of cards and more than one player in the game at the same time.

## ACKNOLEDGEMENTS

- CAPES
- o São Paulo Research Foundation (FAPESP), grant 2016/02870-0

# Thank you!



EXTRACTING RULES FOR BLACKJACK USING MACHINE LEARNING AND FUZZY SYSTEMS

Marcos E. Cintra – mecintra@gmail.com